Econometric Theory, **39**, 2023, 582–622.
doi:10.1017/S0266466622000147

TESTING FOR UNOBSERVED HETEROGENEOUS TREATMENT EFFECTS WITH OBSERVATIONAL DATA

YU-CHIN HSU
Academia Sinica, National Central University, National Chengchi University, and National Taiwan University

TA-CHENG HUANG
National University of Singapore

HAIQING XU
University of Texas at Austin

Unobserved heterogeneous treatment effects have been emphasized in the recent policy evaluation literature (see, e.g., Heckman and Vytlacil (2005, *Econometrica* 73, 669–738)). This paper proposes a nonparametric test for unobserved heterogeneous treatment effects in a treatment effect model with a binary treatment assignment, allowing for individuals' self-selection to the treatment. Under the standard *local average treatment effects* assumptions, i.e., the no defiers condition, we derive testable model restrictions for the hypothesis of unobserved heterogeneous treatment effects. Furthermore, we show that if the treatment outcomes satisfy a monotonicity assumption, these model restrictions are also sufficient. Then, we propose a modified Kolmogorov–Smirnov-type test which is consistent and simple to implement. Monte Carlo simulations show that our test performs well in finite samples. For illustration, we apply our test to study heterogeneous treatment effects of the Job Training Partnership Act on earnings and the impacts of fertility on family income, where the null hypothesis of homogeneous treatment effects gets rejected in the second case but fails to be rejected in the first application.

The authors are grateful to the Editor (Peter C.B. Phillips), the Co-Editor (Yoon-Jae Whang), three anonymous referees, Jason Abrevaya, Qi Li, Xiaojun Song, and Quang Vuong for valuable comments and suggestions, which have considerably improved the presentation of the paper. Yu-Chin Hsu gratefully acknowledges the research support from the Ministry of Science and Technology of Taiwan (MOST2628-H-001-007, MOST110-2634-F-002-045), Academia Sinica Investigator Award of Academia Sinica (AS-IA-110-H01), and Center for Research in Econometric Theory and Applications (110L9002) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education of Taiwan. Ta-Cheng Huang is indebted to Qi Li for his continued inspiration, guidance, and support. Haiqing Xu would like to dedicate this paper to the memory of Professor Halbert White. Address correspondence to Ta-Cheng Huang, Global Asia Institute, National University of Singapore, Singapore; e-mail: tchuang@nus.edu.sg.

1. INTRODUCTION

Heterogeneous treatment effects due to unobserved latent variables have been emphasized in the policy evaluation literature. See, e.g., Imbens and Angrist (1994), Heckman, Smith, and Clements (1997), Heckman and Vytlacil (2001, 2005), Abadie, Angrist, and Imbens (2002), Abadie (2002, 2003), Blundell and Powell (2003), Matzkin (2003), Chesher (2003, 2005), Chernozhukov and Hansen (2005), Florens et al. (2008), Imbens and Newey (2009), Frölich and Melly (2013), D'Haultfœuille and Février (2015), and Torgovitsky (2015), among many others. In the empirical study of treatment effects using observational data, the interpretation of the widely used instrumental variable (IV) estimation relies on the key assumption that after we control for covariates, treatment effects are homogeneous across individuals. In the presence of unobserved heterogeneous treatment effects, the standard IV approach only estimates the *local average treatment effects* (LATEs), rather than the *average treatment effects* (ATEs); see Imbens and Angrist (1994) and Imbens (2010).

In this paper, we develop a nonparametric test for the (unobserved) heterogeneous treatment effects. We model the unobserved heterogeneous treatment effects by a nonparametric and nonseparable model, i.e., the error terms are not additively separable from the treatment indicator. Together with the endogeneity issue introduced due to individuals' self-selection to treatment, it is well known in the literature that identification and estimation of nonseparable models are challenging. On the other hand, the homogeneous treatment effects assumption substantially simplifies econometrics analysis of treatment effects, since it implies that ATE is the same as LATE, after controlling for observed heterogeneity (i.e., covariates). For instance, Angrist and Krueger (1991) use a two-stage least-squares approach to estimate treatment effects of compulsory schooling on earnings. Therefore, if ATE is the main object of research interest, by testing for heterogeneous treatment effects, our method can assess whether the complicated nonparametric and nonseparable treatment effect model is more appropriate (than, e.g., the two-stage least-squares approach) for a program evaluation assignment.

Though important, there are only a handful of papers on testing for such unobserved heterogeneity.[1] In the context of ideal social experiment data, i.e., lack of endogeneity, Heckman et al. (1997) develop a lower bound for the variance of heterogeneous treatment effects, thereby providing a test for whether the data are consistent with the identical treatment effects model. Moreover, Hoderlein and Mammen (2009) discuss specification tests for endogeneity as well as unobserved heterogeneity in nonseparable triangular models. Recently, Lu and White (2014) and Su, Tu, and Ullah (2015) establish nonparametric tests for unobserved heterogeneous treatment effects under the unconfoundedness assumption. In particular,

[1] There exists a substantial literature for testing observed heterogeneity, i.e., whether (conditional) ATEs vary across different subpopulations defined by observed covariates. For example, see Heckman et al. (1997), Crump et al. (2008), Chang, Lee, and Whang (2015), Abrevaya, Hsu, and Lieli (2015), Athey and Imbens (2016), Hsu (2017), and Lee, Okui, and Whang (2017), among many others.

Lu and White (2014) test unobserved heterogeneity in treatments effects via testing an equivalent independence condition on observables. Another closely related paper is by Heckman, Schmierer, and Urzua (2010), who test the absence of self-selection on the gain to treatment in the generalized Roy model framework, allowing for unobserved heterogeneous treatment effects. Furthermore, our paper is also related to Heckman and Vytlacil (2005), who suggest an approach to test heterogeneity of the *marginal treatment effects* (MTEs). Our test object of interest, however, focuses on whether there exists individual-level unobserved heterogeneity in treatment effects, rather than group-level (defined by a margin) unobserved heterogeneity, i.e., whether MTE varies across margins.

Motivated by Lu and White (2014), we show that in the presence of endogeneity, model restrictions arising from the homogeneous treatment effects hypothesis can also be characterized by a set of independence conditions that involve LATE. These testable implications are related to important literature on testing whether the distributions of potential outcomes are affected by the treatment. In the LATE framework, Abadie et al. (2002) considers the null hypothesis of the equality between distribution functions of the potential outcomes among the compliers in treatment and control groups, and also first-order and second-order stochastic dominance of the two distribution functions. Lee and Whang (2009) and Chang et al. (2015) generalize Abadie et al.'s (2002) test for conditional distributional effects by allowing for observed treatment effects heterogeneity.[2] Our test problem differs from that literature in that we investigate whether the distribution function of the treatment group is an (unknown) constant shift from the control group's distribution. Specifically, the equality hypothesis on the two distribution functions is a special case of our test implication.

Nonparametric tests for conditional independence restrictions have been well studied in different contexts. See, e.g., Andrews (1997), Dauxois and Nkiet (1998), Su and White (2007, 2008, 2014), Huang (2010), Bouezmarni and Taamouti (2014), Hoderlein and White (2012), Linton and Gozalo (2014), and Huang, Sun, and White (2016), among many others. When one tests independence restrictions of variables that are nonparametrically constructed, a key technical issue arises in the case, in particular, in which the nonparametric components are functions of continuous covariates (see, e.g., Lu and White, 2014). Motivated by Stinchcombe and White (1998), we modify the classic Kolmogorov–Smirnov tests by using the primitive function of cumulative distributive functions. Such a modification is novel and plays a key role in our approach. Moreover, we establish the asymptotic properties of the proposed tests under the null and alternative hypotheses.

The remainder of this paper is organized as follows. In Section 2, we introduce the model and derive testable model restrictions. Section 3 discusses our test statistics and their asymptotic results. We distinguish whether the covariates include continuous variables. In Section 4, we conduct Monte Carlo experiments to study the finite-sample performance of the proposed test. Section 5 illustrates

[2]See also, e.g., Jun, Lee, and Shin (2016) and Hsu (2017) for further extensions, and the references therein.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

our testing approach by two empirical applications. All proofs are collected in Appendix A.

2. MODEL AND TESTABLE RESTRICTIONS

We consider the following nonseparable treatment effect model:

$$Y = g(D, X, \epsilon), \tag{1}$$

where $Y \in \mathrm{R}$ is the outcome variable, $D \in \{0,1\}$ denotes the treatment status, $X \in \mathrm{R}^{d_X}$ is a vector of covariates, ϵ is an unobserved random disturbance of general form (e.g., without invoking any restriction on the dimensionality of ϵ), and g is an unknown but smooth function defined on $\{0,1\} \times \mathcal{S}_{X\epsilon}$.[3] In particular, the treatment variable D is allowed to be correlated with ϵ to allow for selection to the treatment; see, e.g., Heckman et al. (1997). To deal with endogeneity, we introduce a binary IV $Z \in \{0,1\}$. Throughout the paper, we use uppercase letters to denote random variables, and their corresponding lowercase letters to stand for realizations of random variables.

As is motivated in the seminal paper by Matzkin (2003), the nonadditivity of the structural relationship g in ϵ captures the idea of unobserved heterogeneous treatment effects in that the individual treatment effect, $g(1,X,\epsilon) - g(0,X,\epsilon)$, would depend on the unobserved individual heterogeneity ϵ, even after controlling for covariates X. Therefore, we have the following proposition.

PROPOSITION 1. *Suppose (1) holds, then the homogeneous treatment effects hypothesis, i.e., for some measurable function* $\delta(\cdot) : \mathcal{S}_X \mapsto \mathrm{R}$,

$$\mathcal{H}_0 : \; g(1,X,\cdot) - g(0,X,\cdot) = \delta(X) \tag{2}$$

holds if and only if the structural relationship g is additively separable in ϵ (w.r.t. D), i.e.,

$$g(D,X,\epsilon) = m(D,X) + \nu(X,\epsilon), \tag{3}$$

where $m : \mathcal{S}_{DX} \mapsto \mathrm{R}$ *and* $\nu : \mathcal{S}_{X\epsilon} \mapsto \mathrm{R}$.

Proposition 1 directly follows Lu and White (2014). Note that if (3) holds, $\delta(x) = m(1,x) - m(0,x)$ in (2), which is the homogenous individual treatment effects across individuals with covariates $X = x$.

The key insight in Lu and White (2014) is that they further show the equivalence between the additive separability hypothesis and a conditional independence restriction on observables. In the presence of treatment endogeneity, we derive a similar result. Let $p(x,z) = \mathbf{P}(D = 1|X = x, Z = z)$ be the *propensity score* for each $x \in \mathcal{S}_X$ and $z \in \{0,1\}$.

[3]For a generic random vector A, $\mathcal{S}_A$ denotes the support of A.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Assumption 1. Suppose $Z \perp\!\!\!\perp \epsilon|X$. For all $x \in \mathcal{S}_X$, $\mathbf{P}(Z=1|X=x)$ is bounded away from zero and one, and $p(x,0) \neq p(x,1)$.

Assumption 1 is standard in the literature and requires the IV Z to be (conditionally) exogenous and relevant. See, e.g., Imbens and Angrist (1994) and Chernozhukov and Hansen (2005). Throughout, we maintain Assumption 1. Moreover, let $\mu(x,z) = \mathbf{E}(Y|X=x, Z=z)$. Under $\mathcal{H}_0$ and Assumption 1, we have

$$\begin{aligned}\mu(x,z) &= \mathbf{E}[g(0,X,\epsilon) + \delta(X) \times D|X=x, Z=z] \\ &= \mathbf{E}[g(0,X,\epsilon)|X=x] + \delta(x)p(x,z), \quad \text{for } z=0,1.\end{aligned}$$

In the above system of equations, we treat $\mathbf{E}[g(0,X,\epsilon)|X=x]$ and $\delta(x)$ as two unknowns. Solving the equations, we then identify LATE $\delta(x)$ as follows:

$$\delta(x) = \frac{\mu(x,1)-\mu(x,0)}{p(x,1)-p(x,0)} = \frac{\mathrm{Cov}(Y,Z|X=x)}{\mathrm{Cov}(D,Z|X=x)}. \tag{4}$$

See Imbens and Angrist (1994) for the LATE interpretation of (4). Note that $\delta(x)$ is well defined given $p(x,0) \neq p(x,1)$ under Assumption 1 and identified as well directly from the data regardless of the monotonicity of the selection. Furthermore, let

$$W \equiv Y + (1-D) \times \frac{\mathrm{Cov}(Y,Z|X)}{\mathrm{Cov}(D,Z|X)}. \tag{5}$$

Under the null hypothesis $\mathcal{H}_0$, we have

$$W = g(D,X,\epsilon) + (1-D) \times [g(1,X,\epsilon) - g(0,X,\epsilon)] = g(1,X,\epsilon),$$

which implies that W is conditionally independent of Z given X under Assumption 1. Therefore, we obtain the following lemma.

LEMMA 1. *Suppose (1) and Assumption 1 hold. Then, $\mathcal{H}_0$ implies that $W \perp\!\!\!\perp Z|X$. On the other hand, if $W \perp\!\!\!\perp Z|X$, then the observed data can be rationalized by a structure that satisfies $\mathcal{H}_0$.*

Lemma 1 shows that the conditional independence condition is all the testable restrictions of $\mathcal{H}_0$, i.e., it is sharp in the sense of Definition 1 of Hsu, Liu, and Shi (2019). Regarding the first part of Lemma 1, intuitively, if treatment effects are homogeneous, we can estimate them by the IV method, and further construct potential outcomes that are independent of the IV.[4] Note that the conditional independence condition in Lemma 1 can be equivalently rewritten as

$$\begin{aligned}&\frac{\mathbf{P}(Y \le y; D=1|X,Z=1) - \mathbf{P}(Y \le y; D=1|X,Z=0)}{p(X,1)-p(X,0)} \\ &\quad = \frac{\mathbf{P}(Y \le y-\delta(X); D=0|X,Z=0) - \mathbf{P}(Y \le y-\delta(X); D=0|X,Z=1)}{p(X,1)-p(X,0)},\end{aligned}$$

[4]Note that one could also define $W^a = Y - D \times \delta(X)$, which is equal to $g(0,X,\epsilon)$ under Assumption 1.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

provided $p(X,1) \neq p(X,0)$ almost surely. Under the additional monotonicity condition on the selection, both sides in the above equation can be interpreted as the conditional distribution of "potential outcomes" given the compliers group in Imbens and Rubin (1997).

According to Lemma 1, rejecting $W \perp\!\!\!\perp Z|X$ suggests the presence of unobserved heterogeneous treatment effects. It is worth pointing out, however, that the other direction of the above statement is also true under additional assumptions given below. These additional assumptions have been widely used for obtaining identification of quantile treatment effects and LATEs in the IV literature (see, e.g., Imbens and Angrist, 1994; Chernozhukov and Hansen, 2005).

Assumption 2 (Single-index error term). There exists a measurable function $\tilde{g} : \mathcal{S}_{DX} \times \mathrm{R} \mapsto \mathrm{R}$ and $\nu : \mathcal{S}_{X\epsilon} \mapsto \mathrm{R}$ such that

$$g(D,X,\epsilon) = \tilde{g}(D,X,\nu(X,\epsilon)).$$

Moreover, $\tilde{g}(d,x,\cdot)$ is strictly increasing in the scalar-valued index ν.

Assumption 2 imposes the monotonicity of the single-index error term, of which various simplified assumptions have also been made in the literature for identification and estimation of nonseparable functions. For instance, among many others, Matzkin (2003) and Chesher (2003) assume that the structural function g is strictly increasing in the scalar-valued error term ϵ. Note that Assumption 2 holds under the null hypothesis $\mathcal{H}_0$ because (3) would hold under $\mathcal{H}_0$. Assumption 2 narrows down the space of alternatives such that the model restrictions derived in Lemma 1 are also sufficient to distinguish the null and alternative hypotheses.

Assumption 3 (Monotone selection). The selection to the treatment is given by

$$D = \mathbb{1}\left[\theta(X,Z) - \eta \geq 0\right], \tag{6}$$

where θ is an unknown function, and $\eta \in \mathrm{R}$ is an error term satisfying $Z \perp\!\!\!\perp (\epsilon,\eta)|X$.

Imbens and Angrist (1994) first introduce the monotone selection assumption, which is essentially the "no defier" condition. Moreover, Vytlacil (2002) shows that such a monotonicity condition is observationally equivalent to the weak monotonicity of (6) in the error term η.

For any $x \in \mathcal{S}_X$, let $\mathcal{C}_x \equiv \{\eta \in \mathrm{R} : \min\{\theta(x,0),\theta(x,1)\} < \eta \leq \max\{\theta(x,0),\theta(x,1)\}\}$. Note that $\mathcal{C}_x$ is called the "complier group" if $p(x,0) < p(x,1)$ (see Imbens and Angrist, 1994, for the concept of the "complier group.")

Assumption 4. The support of $g(d,x,\epsilon)$ given $X = x$ and the complier group $\mathcal{C}_x$ is equal to the support of $g(d,x,\epsilon)$ given $X = x$, i.e., $\mathcal{S}_{g(d,x,\epsilon)|X=x,\eta\in\mathcal{C}_x} = \mathcal{S}_{g(d,x,\epsilon)|X=x}$.

Assumption 4 is a support condition, first introduced by Vuong and Xu (2017) as the effectiveness of the IV. It implies that $\mathcal{S}_{g(d,x,\epsilon)|X=x,\eta\in\mathcal{C}_x} = \mathcal{S}_{Y|D=d,X=x}$.[5] Note that the distribution of $g(d,x,\epsilon)$ given $X = x$ and $\eta \in \mathcal{C}_x$ can be identified; see,

[5]To see this, note that $\mathcal{S}_{g(d,x,\epsilon)|X=x,\eta\in\mathcal{C}_x} \subseteq \mathcal{S}_{g(d,x,\epsilon)|D=d,X=x} \subseteq \mathcal{S}_{g(d,x,\epsilon)|X=x}$.

e.g., Imbens and Rubin (1997). Thus, Assumption 4 is testable. Specifically, for all $t \in \mathrm{R}$,

$$\mathbf{P}[g(d,x,\epsilon) \le t | X = x, \eta \in \mathcal{C}_x] = \frac{\mathbf{P}(Y \le t, D = d | X = x, Z = 1) - \mathbf{P}(Y \le t, D = d | X = x, Z = 0)}{\mathbf{P}(D = d | X = x, Z = 1) - \mathbf{P}(D = d | X = x, Z = 0)},$$

from which we can identify the support $\mathcal{S}_{g(d,x,\epsilon)|X=x,\eta\in\mathcal{C}_x}$.

Assumption 4 allows one to use the data to address questions involving counterfactuals of outcomes of the "always-takers" and the "never-takers" groups. It is possible to provide sufficient primitive conditions for Assumption 4. For instance, if one assumes $\mathcal{S}_{\epsilon|X=x,\eta\in\mathcal{C}_x} = \mathcal{S}_{\epsilon|X=x}$, or even a stronger condition that (ϵ, η) has a rectangular support conditional on $X = x$, then Assumption 4 holds.

THEOREM 1. *Suppose that (1) and Assumptions 1–4 hold. Then, $\mathcal{H}_0$ holds if and only if $W \perp\!\!\!\perp Z|X$.*

Recall the definition of W in (5). Theorem 1 shows that testing the null hypothesis $\mathcal{H}_0$ should just rely on the information from the compliers group. It is worth pointing out that Theorem 1 is related to Lu and White (2014), who show that $\mathcal{H}_0$ holds if and only if $Y - \mathbf{E}(Y|D,X) \perp\!\!\!\perp D|X$ under the unconfoundedness condition (i.e., $D \perp\!\!\!\perp \epsilon|X$) and Assumption 2.

It should also be noted that Assumption 2 is a crucial condition for the equivalence between the null hypothesis $\mathcal{H}_0$ and the testable model restrictions $W \perp\!\!\!\perp Z|X$. Chernozhukov and Hansen (2005) show that this assumption is observationally equivalent to the rank similarity assumption. In the current literature, Assumption 2 (or the rank similarity assumption) has been widely used for identifying heterogeneous treatment effects. See, e.g., Chernozhukov and Hansen (2005) and Vuong and Xu (2017).

We also note that throughout, we maintain the validity of the instrument, i.e., Assumption 1. If this assumption is questionable, then our test should be more carefully interpreted as a joint test of Assumption 1 and the homogeneity of treatment effects.

3. CONSISTENT TESTS

Based on Theorem 1, we now propose tests for unobserved treatment effect heterogeneity via testing the conditional independence restriction. Because Z is binary, the conditional independence restriction in Theorem 1 is equivalent to

$$F_{W|XZ}(\cdot|x,0) = F_{W|XZ}(\cdot|x,1), \ \forall\, x \in \mathcal{S}_X.$$

Note that the variable W needs to be nonparametrically constructed from the data. In the following discussion, we distinguish the cases where the covariates X are continuous random variables because the continuous-covariates case is more difficult to deal with due to the nonparametric function $\delta(\cdot)$ in the construction of W.

3.1. Case 1: Discrete Covariates

We first discuss the case where X takes only a finite number of values. Let $\{(Y_i, D_i, X_i, Z_i) : i \le n\}$ be a random sample of (Y, D, X, Z). By Theorem 1, we test $\mathcal{H}_0$ via the following model restrictions:

$$F_{W|XZ}(\cdot\,|x,0) = F_{W|XZ}(\cdot\,|x,1),\ \forall\, x \in \mathcal{S}_X,$$

where $W = Y + (1-D) \times \delta(X)$ is generated from the observables.

For a generic k-dimensional random vector $(A_1, \ldots, A_k)$, we let $\mathbb{1}_{A_1 \cdots A_k}(a_1, \ldots, a_k) \equiv \mathbb{1}(A_1 = a_1, \ldots, A_k = a_k)$, and let $\mathbb{1}(\cdot)$ be the indicator function. We estimate $\delta(X_i)$ as follows:

$$\hat{\delta}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{X_i Z_i}(x,1) \times \sum_{i=1}^n \mathbb{1}_{X_i}(x) - \sum_{i=1}^n Y_i \mathbb{1}_{X_i}(x) \times \sum_{i=1}^n \mathbb{1}_{X_i Z_i}(x,1)}{\sum_{i=1}^n D_i \mathbb{1}_{X_i Z_i}(x,1) \times \sum_{i=1}^n \mathbb{1}_{X_i}(x) - \sum_{i=1}^n D_i \mathbb{1}_{X_i}(x) \times \sum_{i=1}^n \mathbb{1}_{X_i Z_i}(x,1)},$$

and further let $\widehat{W}_i = Y_i + (1-D_i) \times \hat{\delta}(X_i)$. We now define our test statistic:

$$\widehat{\mathcal{T}}_n = \sup_{(w,x)\in\mathcal{S}_{WX}} \sqrt{n}\, \left|\widehat{F}_{W|XZ}(w|x,0) - \widehat{F}_{W|XZ}(w|x,1)\right|,$$

where $\widehat{F}_{W|XZ}(w|x,z) = \frac{\sum_{i=1}^n \mathbb{1}(\widehat{W}_i \le w) Z_i \mathbb{1}_{X_i}(x)}{\sum_{i=1}^n Z_i \mathbb{1}_{X_i}(x)}$.

Next, we establish the asymptotic properties of the test statistic $\widehat{\mathcal{T}}_n$. Let

$$f_{WD|XZ}(w,d|x,z) \equiv f_{W|DXZ}(w|d,x,z) \times \mathbf{P}(D = d|X = x, Z = z)$$

and

$$\kappa(w,x) \equiv -\frac{f_{WD|XZ}(w,0|x,1) - f_{WD|XZ}(w,0|x,0)}{p(x,1) - p(x,0)}.$$

Note that, under Assumptions 1 and 3, $\kappa(w,x) \ge 0$ since it becomes the conditional density of $g(0,x,\epsilon)$ given the complier group and $X = x$.

Assumption 5. Assume that

(i) X is discrete and takes a finite number of values and $P(X = x, Z = z) > 0$, for all $(x,z) \in \mathcal{S}_{XZ}$;
(ii) $p(x,1) - p(x,0) > 0$, for all $x \in \mathcal{S}_X$;
(iii) W has a compact support, and $\partial f_{WD|XZ}(w,0|x,z)/\partial w$ is bounded above uniformly over $(w,x,z) \in \mathcal{S}_{WXZ}$.

Moreover, let

$$\psi_{wx} \equiv \left[\mathbb{1}(W \le w) - F_{W|XZ}(w|x,1)\right] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X = x, Z = 1)} - \left[\mathbb{1}(W \le w) - F_{W|XZ}(w|x,0)\right] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X = x, Z = 0)}, \quad (7)$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

$$\phi_{wx} \equiv \kappa(w,x)\Big[W - \mathbf{E}(W|X=x,Z=0)\Big] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x,Z=1)} \\ -\kappa(w,x)\Big[W - \mathbf{E}(W|X=x,Z=1)\Big] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x,Z=0)}. \quad \textbf{(8)}$$

We now derive the asymptotic behavior of the test statistic.

THEOREM 2. *Suppose that (1) and Assumptions 1–5 hold. Then, under $\mathcal{H}_0$,*

$$\widehat{\mathcal{T}}_n \xrightarrow{d} \sup_{(w,x)\in\mathcal{S}_{WX}} |\mathcal{Z}(w,x)|,$$

where $\mathcal{Z}(\cdot,\cdot)$ is a mean-zero Gaussian process with a covariance kernel:

$$Cov\left[\mathcal{Z}(w,x),\mathcal{Z}(w',x')\right] = \mathrm{E}\left[(\psi_{wx}+\phi_{wx})(\psi_{w'x'}+\phi_{w'x'})\right],\ \forall (w,x),(w',x') \in \mathcal{S}_{WX}.$$

Moreover, under $\mathcal{H}_1$, we have

$$n^{-\frac{1}{2}}\widehat{\mathcal{T}}_n \xrightarrow{p} \sup_{(w,x)\in\mathcal{S}_{WX}} \left|F_{W|XZ}(w|x,0) - F_{W|XZ}(w|x,1)\right| > 0.$$

In Appendix A, we show that the influence function for $\sqrt{n}[\widehat{F}_{W|XZ}(w|x,0) - \widehat{F}_{W|XZ}(w|x,1) - F_{W|XZ}(w|x,0) + F_{W|XZ}(w|x,1)]$is $\psi_{wx}+\phi_{wx}$, in which ψ_{wx} is the influence function when $\delta(x)$ is known, and ϕ_{wx} accounts for the estimation effect of estimating $\delta(x)$. By Theorem 2, our test is one-sided: reject $\mathcal{H}_0$ at significance level α if and only if $\widehat{\mathcal{T}}_n > c_\alpha$, where c_α is the $(1-\alpha)$th quantile of $\sup_{(w,x)\in\mathcal{S}_{WX}}|\mathcal{Z}(w,x)|$.

Since the asymptotic distribution of $\sup_{(w,x)\in\mathcal{S}_{WX}}|\mathcal{Z}(w,x)|$ is nonpivotal and complicated, we apply the multiplier bootstrap method to approximate the entire process to construct the critical value. See, e.g., van der Vaart and Wellner (1996), Delgado and Manteiga (2001), Barrett and Donald (2003), and Donald and Hsu (2014). Specifically, we simulate a sequence of i.i.d. pseudorandom variables $\{U_i : i = 1,\ldots,n\}$ that is independent of the random sample path $\{(Y_i,X_i,D_i,Z_i) : i = 1,2,\ldots\}$ with $\mathrm{E}[U] = 0$, $\mathrm{E}[U^2] = 1$, and $\mathrm{E}[U^4] < +\infty$. Then, we obtain the following simulated empirical process:

$$\widehat{\mathcal{Z}}^u(w,x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \times (\hat{\psi}_{wx,i} + \hat{\phi}_{wx,i}),$$

where $\hat{\psi}_{wx,i}+\hat{\phi}_{wx,i}$ is the estimated influence function. Namely,

$$\hat{\psi}_{wx,i} = \left[\mathbb{1}(\widehat{W}_i \le w) - \frac{\sum_{j=1}^n \mathbb{1}(\widehat{W}_i \le w)\mathbb{1}_{X_jZ_j}(x,1)}{\sum_{j=1}^n \mathbb{1}_{X_jZ_j}(x,1)}\right] \times \frac{\mathbb{1}_{X_iZ_i}(x,1)}{\widehat{\mathbf{P}}(X=x,Z=1)} \\ -\left[\mathbb{1}(\widehat{W}_i \le w) - \frac{\sum_{j=1}^n \mathbb{1}(\widehat{W}_i \le w)\mathbb{1}_{X_jZ_j}(x,0)}{\sum_{j=1}^n \mathbb{1}_{X_jZ_j}(x,0)}\right] \times \frac{\mathbb{1}_{X_iZ_i}(z,0)}{\widehat{\mathbf{P}}(X=x,Z=0)},$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

$$\hat{\phi}_{wx,i} = \hat{\kappa}(w,x)\left[\widehat{W}_i - \frac{\sum_{j=1}^n \widehat{W}_j \mathbb{1}_{X_jZ_j}(x,0)}{\sum_{j=1}^n \mathbb{1}_{X_jZ_j}(x,0)}\right] \times \frac{\mathbb{1}_{X_iZ_i}(x,1)}{\widehat{\mathbf{P}}(X=x,Z=1)}$$
$$- \hat{\kappa}(w,x)\left[\widehat{W}_i - \frac{\sum_{j=1}^n \widehat{W}_j \mathbb{1}_{X_jZ_j}(x,1)}{\sum_{j=1}^n \mathbb{1}_{X_jZ_j}(x,1)}\right] \times \frac{\mathbb{1}_{X_iZ_i}(x,0)}{\widehat{\mathbf{P}}(X=x,Z=0)},$$

where $\widehat{\mathbf{P}}(X=x,Z=z)$ and $\hat{\kappa}(w,x) = -\frac{\hat{f}_{WD|XZ}(w,0|x,1)-\hat{f}_{WD|XZ}(w,0|x,0)}{\hat{p}(x,1)-\hat{p}(x,0)}$ are uniformly consistent nonparametric estimators for $\mathbf{P}(X=x,Z=z)$ and $\kappa(w,x)$, respectively. For a given significance level α, the critical value $\hat{c}_n(\alpha)$ is obtained as the $(1-\alpha)$-quantile of the simulated distribution of $\sup_{(w,x)\in\mathcal{S}_{WX}}\left|\widehat{\mathcal{Z}}^u(w,x)\right|$.

Now, we give additional conditions for the validity of the multiplier bootstrap critical value.

Assumption 6. Assume that $\{U_i : i = 1,\ldots,n\}$ is a sequence of i.i.d. pseudorandom variables that is independent of the random sample path $\{(Y_i, X_i, D_i, Z_i) : i = 1,2,\ldots\}$ with $\mathrm{E}[U] = 0$, $\mathrm{E}[U^2] = 1$, and $\mathrm{E}[U^4] < \infty$.

In simulations and empirical studies, we set U_i's as standard normals so that Assumption 6 is satisfied.

Assumption 7. Assume that, for $z = 0,1$,

(i) $\sup_{x\in\mathcal{S}_X}|\widehat{\mathbf{P}}(X=x,Z=z) - \mathbf{P}(X=x,Z=z)| \xrightarrow{p} 0$;
(ii) $\sup_{x\in\mathcal{S}_X}|\hat{p}(x,z) - p(x,z)| \xrightarrow{p} 0$;
(iii) $\hat{f}_{WD|XZ}(w,0|x,z)$ is continuous in w for all $x \in \mathcal{S}_X$ and $\sup_{(w,x)\in\mathcal{S}_{WX}}|\hat{f}_{WD|XZ}(w, 0|x,z) - f_{WD|XZ}(w,0|x,z)| \xrightarrow{p} 0$;
(iv) $\sup_{x\in\mathcal{S}_X}|\hat{\delta}(x) - \delta(x)| \xrightarrow{p} 0$.

THEOREM 3. *Suppose Assumptions 1–7 hold. Then,*

(a) *under* H_0, $\lim_{n\to\infty} P(\widehat{\mathcal{T}}_n \geq \hat{c}_n(\alpha)) = \alpha$;
(b) *under* H_1, $\lim_{n\to\infty} P(\widehat{\mathcal{T}}_n \geq \hat{c}_n(\alpha)) = 1$.

Theorem 3 shows the size and power of our test for the discrete case. The proof of Theorem 3 follows standard arguments once we establish the validity of the multiplier bootstrap for the processes in that $\widehat{\mathcal{Z}}^u(\cdot,\cdot) \Rightarrow \mathcal{Z}(\cdot,\cdot)$ conditional on the sample path $\{(Y_i, X_i, D_i, Z_i) : i = 1,2,\ldots\}$ with probability approaching one. Assumption 7 contains high-level conditions, and we provide estimators and give low-level conditions in Appendix A. Please see the discussion after the proof of Theorem 3.

By Assumption 5, X is assumed to be a discrete random variable taking a finite number of values. In this case, we literally conduct the test by sample splitting in that we test the equality of conditional distributions over all subpopulations defined by covariate value x. As a result, it is straightforward to extend our test to the case of discrete random vector X. Therefore, we omit the details for brevity.

3.2. Case 2: Continuous Covariates

We now consider the case where X is a vector of continuous covariates. To extend the empirical process argument used in the proof of Theorem 2 to this case, we propose a modified Kolmogorov–Smirnov test statistic. Such a modification allows us to account for the estimation effects from the generated variable W, which is constructed from the unknown function $\delta(\cdot)$ as an infinite-dimensional parameter.

Let $\lambda(t) = -t \times 1(t \leq 0)$ and $\Pi(w|x,z) = \mathrm{E}[\lambda(W-w)|X = x, Z = z]$. Note that $\lambda(\cdot)$ is continuous and has a directional derivative. By definition, $\Pi(\cdot|x,z)$ is the primitive function of the $F_{W|XZ}(\cdot|x,z)$, i.e.,

$$\frac{\partial}{\partial w}\Pi(w|x,z) = F_{W|XZ}(w|x,z).$$

Thus, the model restriction $W \perp\!\!\!\perp Z|X$ can be characterized as follows:

$$\frac{\partial}{\partial w}\Pi(w|x,0) = \frac{\partial}{\partial w}\Pi(w|x,1), \ \ \forall (w,x) \in \mathcal{S}_{WX},$$

which holds if and only if $\Pi(w|x,0) = \Pi(w|x,1)$, for all $(w,x) \in \mathcal{S}_{WX}$.[6]

In terms of the probability distribution of W given $(X,Z) = (x,z)$, both $F_{W|XZ}(\cdot|x,z)$ and $\Pi(\cdot|x,z)$ contain the exact same amount of information. The latter, however, allows us to derive a test statistic and establish its limiting distribution when W has to be nonparametrically generated. When covariates X are continuously distributed, the generated sample $\{\widehat{W}_i : i \leq n\}$ involves the nonparametric component $\hat{\delta}(X_i)$. We exploit the smoothness of $\lambda(\cdot)$ and show that this first-stage estimation error can be further aggregated out at the $\sqrt{n}$-rate in our test statistic defined on $\{\widehat{W}_i : i \leq n\}$. It should be noted that when covariates X are discrete as discussed in the last subsection, we can also apply a similar testing procedure via testing $\Pi(\cdot|x,1) = \Pi(\cdot|x,0)$. Moreover, we assume that $\mathcal{S}_W$ is compact for expositional simplicity.

We denote $f_{XZ}(x,z) \equiv f_{X|Z}(x|z) \times \mathbf{P}(Z = z)$. We also let $\mathbb{1}^*_A(a) \equiv \mathbb{1}(A \leq a)$. For $z \in \{0,1\}$, let $z' = 1 - z$ and

$$G(w,x,z) = \mathbf{E}\left[\lambda(W-w)\mathbb{1}^*_X(x)\mathbb{1}_Z(z)f_{XZ}(X,z')\right].$$

Motivated by Stinchcombe and White (1998), we rewrite the above conditional expectation restrictions by the following unconditional ones:

$$G(w,x,0) = G(w,x,1), \ \ \forall (w,x) \in \mathcal{S}_{WX}. \tag{9}$$

To see the equivalence, first note that

$$G(w,x,z) = \mathbf{E}\left[\lambda(W-w)\mathbb{1}(X \leq x)f_{X|Z}(X|z')|Z = z\right]\mathbf{P}(Z = 0)\mathbf{P}(Z = 1).$$

[6]The equivalence holds by the fact that for a continuous function $f(t)$, $f(t) = 0$, for all $t \in [0,1]$, if and only if $\int_0^t f(s)ds = 0$, for all $t \in [0,1]$.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Moreover, by the law of iterated expectation,

$$\frac{\partial}{\partial x}\mathbf{E}[\lambda(W-w)\mathbb{1}(X\le x)f_{X|Z}(X|z')|Z=z]=\Pi(w|x,z)f_{X|Z}(x|0)f_{X|Z}(x|1).$$

Therefore, we obtain the conditional expectation restrictions as the derivatives of (9). Note that the estimation of $G(w,x,z)$ avoids any denominator issues, which thereafter simplifies our asymptotic analysis.

For a random variable A and a value a, let $K_{A,h}(a)\equiv K((A-a)/h)/h,$ where K and h are a kernel function and a smoothing bandwidth, respectively. For a d_X-dimensional random vector A, we let $K_{A,h}(a)=\Pi_{j=1}^{d_X}K_{A_j,h}(a_j)=h^{-d_X}K((A-a)/h)$. We estimate $\delta(X_i)$ by

$$\hat{\delta}(X_i)=\frac{\sum_{j\neq i}Y_jZ_jK_{X_j,h}(X_i)\sum_{j\neq i}K_{X_j,h}(X_i)-\sum_{j\neq i}Y_jK_{X_j,h}(X_i)\sum_{j\neq i}Z_jK_{X_j,h}(X_i)}{\sum_{j\neq i}D_jZ_jK_{X_j,h}(X_i)\sum_{j\neq i}K_{X_j,h}(X_i)-\sum_{j\neq i}D_jK_{X_j,h}(X_i)\sum_{j\neq i}Z_jK_{X_j,h}(X_i)}.$$

Note that in this paper, we consider a kernel estimator for the nonparametric components of the test. To avoid the boundary issue, we follow the literature to trim the support of X. To be specific, we will assume that the support of covariates X is a Cartesian product of compact intervals, $\mathcal{S}_X=\prod_{j=1}^{d_X}[x_{\ell j},x_{uj}],$ and $\mathcal{S}_X^{\xi}=\prod_{j=1}^{d_X}[x_{\ell j}+\xi,x_{uj}-\xi],$ where $\xi>0$ is a small positive number. Moreover, let $\widehat{W}_i=Y_i+(1-D_i)\times\hat{\delta}(X_i),$ and

$$\widehat{G}(w,x,z)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in\mathcal{S}_X^{\xi})\lambda(\widehat{W}_i-w)\mathbb{1}_{X_i}^{*}(x)\mathbb{1}_{Z_i}(z)\hat{f}_{XZ}(X_i,z'),$$
$$\hat{f}_{XZ}(X_i,z)=\frac{1}{n}\sum_{j\neq i}K_{X_j,h}(X_i)\mathbb{1}_{Z_j}(z).$$

Thus, we define our test statistic as follows:

$$\widehat{\mathcal{T}}_n^c=\sup_{w\in\mathcal{S}_W,x\in\mathcal{S}_X^{\xi}}\sqrt{n}\left|\widehat{G}(w,x,0)-\widehat{G}(w,x,1)\right|.$$

In the above definition, the supports of W and X are assumed to be known for simplicity. In practice, this assumption can be relaxed by using consistent set estimators $\widehat{\mathcal{S}}_W$ and $\widehat{\mathcal{S}}_X$.

We show that the proposed test statistic $\widehat{\mathcal{T}}_n^c$ converges in distribution at the regular parametric rate under the null. The key step of our proof is to show that

$$\sup_{w\in\mathcal{S}_W,x\in\mathcal{S}_X^{\xi}}\left|\widehat{G}(w,x,z)-\widetilde{G}(w,x,z)\right|=o_p\left(n^{-1/2}\right), \qquad \textbf{(10)}$$

where $\widetilde{G}(w,x,z)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in\mathcal{S}_X^{\xi})(\widehat{W}_i-w)\mathbb{1}(W_i\le w)\mathbb{1}_{X_i}^{*}(x)\mathbb{1}_{Z_i}(z)\hat{f}_{XZ}(X_i,z')$. The above result requires that the nonparametric elements in the estimation of $\hat{\delta}(\cdot)$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

should converge to the corresponding true values uniformly at a rate faster than $n^{-1/4}$.

Assumption 8. Assume that

(i) the support of the d_X-dimensional covariates X is a Cartesian product of compact intervals, $\mathcal{S}_X = \prod_{j=1}^{d_X}[x_{\ell j}, x_{uj}]$;
(ii) for $z = 0, 1$, $\inf_{x\in\mathcal{S}_X} f_{X|Z}(x|z) > 0$, $\sup_{x\in\mathcal{S}_X} f_{X|Z}(x|z) < \infty$, and $\inf_{x\in\mathcal{S}_X} |p(x,1) - p(x,0)| > 0$.

Assumption 9. For $z = 0, 1$, $f_{X|Z}(x|z)$, $p(x,z)$, and $\mu(x,z)$ are continuous in $x \in \mathcal{S}_X$.

Assumption 10. For some $\iota > \frac{1}{4}$, we have $h \to 0$, and $n^{\iota}/\sqrt{nh^{d_X}} \to 0$ as $n \to \infty$. Moreover, the first-stage estimators satisfy the condition that, for $z = 0, 1$,

$$\sup_{x\in\mathcal{S}_X^{\xi}} \left| \mathbf{E}\left[\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{Z_j}(z) K_{X_j,h}(x)\right] - f_{XZ}(x,z) \right| = O(n^{-\iota}),$$

$$\sup_{x\in\mathcal{S}_X^{\xi}} \left| \mathbf{E}\left[\frac{1}{n}\sum_{j=1}^{n} D_j \mathbb{1}_{Z_j}(z) K_{X_j,h}(x)\right] - p(x,z) f_{XZ}(x,z) \right| = O(n^{-\iota}),$$

$$\sup_{x\in\mathcal{S}_X^{\xi}} \left| \mathbf{E}\left[\frac{1}{n}\sum_{j=1}^{n} Y_j \mathbb{1}_{Z_j}(z) K_{X_j,h}(x)\right] - \mathbf{E}(Y|X=x, Z=z) f_{XZ}(x,z) \right| = O(n^{-\iota}).$$

Assumptions 8 and 9 are standard in the nonparametric estimation literature. Assumption 10 is a high-level condition that requires the nonparametric estimation bias to diminish uniformly at a rate faster than $n^{1/4}$. Such a condition on the bias term can be satisfied under additional primitive conditions on the kernel function $K(\cdot)$ and the bandwidth h, respectively, as well as the smoothness of the underlying structural functions. See, e.g., Pagan and Ullah (1999).

LEMMA 2. *Suppose that Assumptions 1–4 and 8–10 hold. Then, (10) holds for* $z = 0, 1$.

By Lemma 2, it suffices to establish the limiting distribution of $\widetilde{G}(w,x,1) - \widetilde{G}(w,x,0)$ for the asymptotic properties of our test statistics. Note that in the definition of $\widetilde{G}(w,x,z)$, there are no nonparametric elements estimated in the indicator function.

To establish asymptotic properties for our test, we make the following assumption.

Assumption 11. Assume that, for $z = 0, 1$,

$$\sup_{x\in\mathcal{S}_X^{\xi}} \left|\mathbf{E}[\hat{\delta}(x)] - \delta(x)\right| = o(n^{-1/2}) \text{ and } \sup_{x\in\mathcal{S}_X^{\xi}} \left|\mathbf{E}[\hat{f}_{XZ}(x,z)] - f_{XZ}(x,z)\right| = o(n^{-1/2}).$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Assumption 11 strengthens Assumption 10 by requiring the bias term in the first-stage nonparametric estimation to be smaller than $o_p(n^{-1/2})$, which can be established by using higher-order kernels (see, e.g., Powell, Stock, and Stoker, 1989).

Next, we establish the asymptotic properties of the test statistic $\widehat{\mathcal{T}}_n^c$. Let $F^*_{WD|XZ}(w,d|x,z) \equiv F_{W|DXZ}(w|d,x,z) \times \mathbf{P}(D=d|X=x,Z=z)$ and

$$\kappa^c(w,x) = -\frac{F^*_{WD|XZ}(w,0|x,1) - F^*_{WD|XZ}(w,0|x,0)}{p(x,1)-p(x,0)}.$$

Moreover, we define

$$\begin{aligned}\psi^c_{wx} = \mathbb{1}(X \in \mathcal{S}_X^{\xi})\Bigg\{&\Big[\lambda(W-w) - \mathbf{E}[\lambda(W-w)|X,Z=1]\Big] \times \frac{\mathbb{1}_X^*(x)\mathbb{1}_Z(0)}{f_{XZ}(X,0)} \\ &- \Big[\lambda(W-w) - \mathbf{E}[\lambda(W-w)|X,Z=0]\Big] \times \frac{\mathbb{1}_X^*(x)\mathbb{1}_Z(1)}{f_{XZ}(X,1)}\Bigg\} f_{XZ}(X,0)f_{XZ}(X,1),\end{aligned} \tag{11}$$

$$\begin{aligned}\phi^c_{wx} = \mathbb{1}(X \in \mathcal{S}_X^{\xi})\kappa^c(w,X)\Bigg\{&\Big[W - \mathbf{E}(W|X,Z=0)\Big] \times \frac{\mathbb{1}_X^*(x)\mathbb{1}_Z(1)}{f_{XZ}(X,1)} \\ &- \Big[W - \mathbf{E}(W|X,Z=1)\Big] \times \frac{\mathbb{1}_X^*(x)\mathbb{1}_Z(0)}{f_{XZ}(X,0)}\Bigg\} f_{XZ}(X,0)f_{XZ}(X,1).\end{aligned} \tag{12}$$

THEOREM 4. *Suppose that Assumptions 1–4 and 8–11 hold. Then, under $\mathcal{H}_0$,*

$$\widehat{\mathcal{T}}_n^c \xrightarrow{d} \sup_{w \in \mathcal{S}_W, x \in \mathcal{S}_X^{\xi}} |\mathcal{Z}^c(w,x)|,$$

where $\mathcal{Z}^c(\cdot,\cdot)$ is a mean-zero Gaussian process with covariance kernel

$$Cov\left[\mathcal{Z}^c(w,x), \mathcal{Z}^c(w',x')\right] = \mathbf{E}\left[(\psi^c_{wx} + \phi^c_{wx})(\psi^c_{w'x'} + \phi^c_{w'x'})\right],\ \forall w,w' \in \mathcal{W},\ x,x' \in \mathcal{S}_X^{\xi}.$$

Moreover, under $\mathcal{H}_1$, we have

$$n^{-\frac{1}{2}}\widehat{\mathcal{T}}_n^c \xrightarrow{p} \sup_{w \in \mathcal{S}_W, x \in \mathcal{S}_X^{\xi}} |G(w,x,0) - G(w,x,1)| > 0.$$

We also show in Appendix A that the influence function for $\sqrt{n}(\widehat{G}(w,x,0) - \widehat{G}(w,x,1) - G(w,x,0) + G(w,x,1))$ is $\psi^c_{wx} + \phi^c_{wx}$, in which ψ^c_{wx} is the influence function when $\delta(x)$ and $f_{XZ}(x,z)$ are known, and ϕ^c_{wx} accounts for the estimation effect of $\delta(x)$ and $f_{XZ}(x,z)$. Note that the $\mathbb{1}(X \in \mathcal{S}_X^{\xi})$ term in the influence function accounts for the fact that we consider a trimmed support of X.

Let the simulated empirical process for the continuous case be

$$\widehat{\mathcal{Z}}^{c,u}(w,x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \times (\hat{\psi}^c_{wx,i} + \hat{\phi}^c_{wx,i}),$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

where $\hat{\psi}^c_{wx,i} + \hat{\phi}^c_{wx,i}$ is the estimated influence function such that

$$\hat{\psi}^c_{wx,i} = \mathbb{1}(X_i \in \mathcal{S}_X^\xi)\Bigg\{\Big[\lambda(\widehat{W}_i - w) - \widehat{\mathbf{E}}[\lambda(W-w)|X_i, Z_i = 1]\Big] \times \frac{\mathbb{1}^*_{X_i}(x)\mathbb{1}_{Z_i}(0)}{\hat{f}_{XZ}(X_i,0)}$$
$$- \Big[\lambda(\widehat{W}_i - w) - \widehat{\mathbf{E}}[\lambda(W-w)|X_i, Z_i = 0]\Big] \times \frac{\mathbb{1}^*_{X_i}(x)\mathbb{1}_{Z_i}(1)}{\hat{f}_{XZ}(X_i,1)}\Bigg\}\hat{f}_{XZ}(X_i,0)\hat{f}_{XZ}(X_i,1),$$
$$\hat{\phi}^c_{wx,i} = \mathbb{1}(X_i \in \mathcal{S}_X^\xi)\hat{\kappa}^c(w,X_i)\Bigg\{\Big[\widehat{W}_i - \widehat{\mathbf{E}}[W|X_i, Z=0]\Big] \times \frac{\mathbb{1}^*_{X_i}(x)\mathbb{1}_{Z_i}(1)}{\hat{f}_{XZ}(X_i,1)}$$
$$- \Big[\widehat{W}_i - \widehat{\mathbf{E}}[W|X_i, Z_i=1]\Big] \times \frac{\mathbb{1}^*_{X_i}(x)\mathbb{1}_{Z_i}(0)}{\hat{f}_{XZ}(X_i,0)}\Bigg\}\hat{f}_{XZ}(X_i,0)\hat{f}_{XZ}(X_i,1),$$

where $\hat{f}_{XZ}(X_i,z)$, $\widehat{\mathbf{E}}[W|X_i, Z=z]$, $\widehat{F}_{W|DXZ}(w|0,X_i,z)$, and $\hat{p}(X_i,z)$ are uniformly consistent nonparametric estimators for $f_{XZ}(X,z)$, $\mathbf{E}[W|X,Z=z]$, $F_{W|DXZ}(w|0,X,z)$, and $p(X,z)$, respectively. For a given significance level α, the critical value $\hat{c}^c_n(\alpha)$ is obtained as the $(1-\alpha)$-quantile of the simulated distribution of $\sup_{w\in\mathcal{S}_W, x\in\mathcal{S}_X^\xi}\left|\widehat{\mathcal{Z}}^{u,c}(w,x)\right|$. We would reject $\mathcal{H}_0$ at significance level α when $\widehat{\mathcal{T}}^c_n > \hat{c}^c_\alpha$.

Now, we give additional conditions for the validity of the multiplier bootstrap critical value.

Assumption 12. Assume that, for $z = 0, 1$,

(i) $\sup_{x\in\mathcal{S}_X^\xi} |\hat{f}_{XZ}(x,z) - f_{XZ}(x,z)| \xrightarrow{p} 0$;
(ii) $\sup_{x\in\mathcal{S}_X^\xi} |\widehat{\mathbf{E}}[W|X=x] - \mathbf{E}[W|X=x]| \xrightarrow{p} 0$;
(iii) $\sup_{x\in\mathcal{S}_X^\xi} |\hat{p}(x,z) - p(x,z)| \xrightarrow{p} 0$;
(iv) $\sup_{x\in\mathcal{S}_X^\xi, w\in\mathcal{S}_W} |\widehat{F}_{W|DXZ}(w|0,x,z) - F_{W|DXZ}(w|0,x,z)| \xrightarrow{p} 0$, and $\widehat{F}_{W|DXZ}(w|0,x,z)$ is nondecreasing in w for all x and z;
(v) $\sup_{x\in\mathcal{S}_X^\xi} |\hat{\delta}(x) - \delta(x)| \xrightarrow{p} 0$.

THEOREM 5. *Suppose Assumptions 1–4, 6, and 8–12 hold. Then,*

(a) *under* H_0, $\lim_{n\to\infty} P(\widehat{\mathcal{T}}^c_n \geq \hat{c}^c_n(\alpha)) = \alpha$;
(b) *under* H_1, $\lim_{n\to\infty} P(\widehat{\mathcal{T}}^c_n \geq \hat{c}^c_n(\alpha)) = 1$.

Theorem 5 shows the size and power of our test for the continuous case. The proof of Theorem 5 is similar to that of the discrete case. Note that Assumptions 10–12 are high-level conditions, and we provide estimators and give low-level conditions in Appendix A. Please see the discussion after the proof of Theorem 5.

We can extend our test to cases where the covariate vector contains both discrete and continuous variables as in our second empirical study. We leave the details to Appendix C.

3.3. Computational Issue

When the dimension of the covariates X is large, there will be a computational issue. For example, suppose that $X = (X_1, X_2, X_3)$ contains three continuous variables. If we use 100 grids in W and 100 grids in each element of X, it will result in 100^4 points of (w,x) when we calculate the test statistic and the simulated critical value. Therefore, the computation burden can be too heavy to be practical when the number of grids we use in each dimension is large, or the dimension of covariates is large. We suggest conducting the test based on the test statistic calculated by all combinations of two covariates only to reduce the burden. Specifically, we calculate the test statistic based on the supremum of 100^3 grids in (W,X_1,X_2), (W,X_1,X_3), and (W,X_2,X_3), respectively, so that we just need to use 3×100^3 grids to calculate the test statistic and the simulated critical value. The procedure is similar to the one recommended by Andrews and Shi (2013).

4. MONTE CARLO SIMULATIONS

In this section, we investigate the finite-sample performance of our tests with a simulation study. The data are simulated as follows:

$$
\begin{aligned}
Y &= D + X + [\gamma + (1-\gamma)D] \times \epsilon_1,\\
D &= \mathbb{1}[\Phi(\eta) \le 0.5 \times Z],\\
\eta &= \rho \times \epsilon + \sqrt{1-\rho^2} \times \epsilon_2,
\end{aligned}
$$

where both ϵ_1 and ϵ_2 conform to a uniform distribution on $[-2,2]$, $\rho = 0.7$, and $Z \sim Bernoulli(p)$ with $p = 0.25$, 0.5, and 0.75, respectively.[7] For simplicity, X, Z, and (ϵ, η) are mutually independent. Moreover, X is uniformly distributed on $\{1,2,3,4\}$ and on $[0,1]$ in the discrete covariates and the continuous covariates cases, respectively. Furthermore, $\gamma \in [0,1]$ describes the degree of unobserved heterogeneous treatment effects in our specification. In particular, $\mathcal{H}_0$ holds if and only if $\gamma = 1$. Intuitively, the smaller γ is, the more power we expect from our tests. To investigate the size and power of our tests, we choose $\gamma \in \{1, 0.75, 0.5\}$.

We consider sample size $n = 1{,}000, 2{,}000, 4{,}000$, a nominal level of $\alpha = 5\%$, and $1{,}000$ Monte Carlo repetitions. Given $X_i = x$, to compute the suprema of the simulated stochastic processes, we use $n/20$ grids on the support of $[\min(\widehat{W}_{i:X_i=x}), \max(\widehat{W}_{i:X_i=x})]$. Moreover, we use $1{,}000$ multiplier bootstrap samples to simulate the p-values. Regarding the estimation of $\kappa(w,x)$, we choose the second-order Epanechnikov kernel function with the bandwidth $h_x = c \times \mathrm{std}(\widehat{W}_{i:X_i=x}) \times n^{-1/4.5}$, and we set $c \in \{1.7, 2, 2.34, 2.6\}$ to study the sensitivity of the test to the bandwidth.

Table 1 reports rejection probabilities of our simulations in the discrete-covariates case under the null hypothesis (i.e., $\gamma = 1$) and alternative hypotheses

[7]Note that we also try different values for the correlation coefficient, and all the results are qualitatively similar. Additional Monte Carlo simulation results are available upon request.

TABLE 1. Rejection probabilities ($\alpha = 5\%$) in the discrete-covariates case.

p	n	$c = 1.7$	$c = 2$	$c = 2.34$	$c = 2.6$
		Panel A: at null hypothesis with $\gamma = 1$			
	1,000	0.0040	0.0060	0.0150	0.0150
0.25	2,000	0.0190	0.0280	0.0270	0.0460
	4,000	0.0420	0.0360	0.0440	0.0470
	1,000	0.0140	0.0210	0.0200	0.0350
0.5	2,000	0.0260	0.0330	0.0500	0.0410
	4,000	0.0300	0.0560	0.0430	0.0630
	1,000	0.0050	0.0130	0.0150	0.0310
0.75	2,000	0.0220	0.0180	0.0290	0.0260
	4,000	0.0390	0.0350	0.0420	0.0430
		Panel B: at alternative hypothesis with $\gamma = 0.75$			
	1,000	0.0230	0.0290	0.0480	0.0410
0.25	2,000	0.2070	0.2750	0.3070	0.3300
	4,000	0.7720	0.7980	0.8030	0.8190
	1,000	0.0700	0.1140	0.1550	0.1620
0.5	2,000	0.4740	0.5350	0.5590	0.5870
	4,000	0.9470	0.9500	0.9670	0.9520
	1,000	0.0480	0.0640	0.1140	0.1160
0.75	2,000	0.2580	0.3490	0.4020	0.4040
	4,000	0.8080	0.8340	0.8600	0.8670
		Panel C: at alternative hypothesis with $\gamma = 0.50$			
	1,000	0.2560	0.3750	0.4900	0.6010
0.25	2,000	0.9620	0.9880	0.9950	0.9970
	4,000	1.0000	1.0000	1.0000	1.0000
	1,000	0.8010	0.8840	0.9470	0.9620
0.5	2,000	1.0000	1.0000	1.0000	1.0000
	4,000	1.0000	1.0000	1.0000	1.0000
	1,000	0.5100	0.6890	0.8010	0.8600
0.75	2,000	0.9990	0.9990	1.0000	1.0000
	4,000	1.0000	1.0000	1.0000	1.0000

(i.e., $\gamma = 0.75, 0.5$). From Panel A, the level of our test is fairly well behaved: It gets closer to the nominal level as the sample size increases, and the rejection probabilities are not sensitive to the constant c for the bandwidth choice. Panels B and C show that the power of the test is reasonable. In particular, when γ is closer to 1, it is more difficult to detect such a "local" alternative. Therefore, we obtain relatively low power even when the sample size reaches $n = 2{,}000$ in

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Panel B. For relatively "small" sample size, e.g., $n = 1,000$, our results show that our test performs better with a larger bandwidth choice. Moreover, when p (i.e., the probability of $Z = 1$) is 0.5, all the results for size and power dominate the other two cases with $p = 0.25, 0.75$, which is expected by our asymptotic theory.

Next, we evaluate the performance of our tests in the case where the covariates X are continuous. To compute the suprema, we calculate the test statistic by using $n/20$ grid points in the support $[\min_{i=1}^{n}(X_i), \max_{i=1}^{n}(X_i)]$, as well as in the support $[Q_{\widehat{W}_i}(0.05), Q_{\widehat{W}_i}(0.95)]$, where $Q_{\widehat{W}_i}(\cdot)$ is the quantile function of $\widehat{W}_i$. In the estimation of $\delta(X_i)$ and $G(w,x)$, we choose the fourth-order Epanechnikov kernel function with the bandwidth $h_n = c \times \text{std}(X_i) \times n^{-1/3}$ to reduce the bias. Furthermore, we study the sensitivity of the test to the bandwidth with $c \in \{1.7, 2, 2.34, 2.6\}$. Table 2 reports the size and power properties of our test, which are qualitatively similar to those in the discrete-covariates case.

5. EMPIRICAL APPLICATIONS

5.1. Effect of Job Training Program on Earnings

We now apply our tests to study the effects of a job training program on earnings, i.e., the *National Job Training Partnership Act* (JTPA), commissioned by the Department of Labor of the United States. This program funded training from 1983 to the late 1990s to increase employment and earnings for participants. The major component of JTPA aims to support training for the economically disadvantaged. The effects of JTPA training programs on earnings have also been studied by, e.g., Heckman et al. (1997) and Abadie et al. (2002) under a general framework allowing for unobserved heterogeneous treatment effects.

Our sample consists of 11,204 observations from the JTPA, a survey dataset from over 20,000 adults and out-of-school youths who applied for the JTPA in 16 local areas across the country between 1987 and 1989.[8] ,[9] Each participant was randomly assigned to either a program group or a control group (one out of three on average). Members of the program group were eligible to participate in JTPA services, including classroom training, on-the-job training or job search assistance, and other services, whereas members of the control group were not eligible for JTPA services for 18 months. Following the literature (see, e.g., Bloom et al., 1997), we use the program eligibility as an IV for the endogenous individual participation decision.

The outcome variable is individual earnings, measured by the sum of earnings in the 30-month period following the offer. The observed covariates include a set of dummies for race, high-school graduate or the general educational development (GED) holder, and marriage, whether the applicant worked at least 12 weeks in the

[8]The data are publicly available at https://www.upjohn.org/data-tools/employment-research-data-center/national-jtpa-study.

[9]JTPA services are provided at 649 sites, which might not be randomly chosen. For a given site, the applicants were randomly selected for the JTPA dataset.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Table 2. Rejection probabilities ($\alpha = 5\%$) in the continuous-covariates case.

p	n	$c = 1.7$	$c = 2$	$c = 2.34$	$c = 2.6$
		Panel A: at null hypothesis with $\gamma = 1$			
	1,000	0.0140	0.0500	0.0580	0.0940
0.25	2,000	0.0940	0.1180	0.1120	0.1200
	4,000	0.1400	0.1240	0.0940	0.0720
	1,000	0.0320	0.0280	0.0360	0.0620
0.5	2,000	0.0048	0.0440	0.0620	0.0360
	4,000	0.0460	0.0380	0.0660	0.0440
	1,000	0.0100	0.0060	0.0060	0.0260
0.75	2,000	0.0340	0.0300	0.0200	0.0220
	4,000	0.0700	0.0580	0.0320	0.0480
		Panel B: at alternative hypothesis with $\gamma = 0.75$			
	1,000	0.0420	0.0420	0.1060	0.1120
0.25	2,000	0.1520	0.2980	0.3540	0.4380
	4,000	0.7420	0.8220	0.8700	0.8620
	1,000	0.0220	0.0440	0.0880	0.1060
0.5	2,000	0.2960	0.3820	0.4740	0.5320
	4,000	0.8780	0.9100	0.9480	0.9460
	1,000	0.0020	0.0000	0.0140	0.0120
0.75	2,000	0.0240	0.0000	0.0920	0.1180
	4,000	0.3980	0.5260	0.5980	0.7000
		Panel C: at alternative hypothesis with $\gamma = 0.50$			
	1,000	0.1780	0.3680	0.5980	0.7280
0.25	2,000	0.8360	0.9480	0.9860	0.9960
	4,000	1.0000	1.0000	0.9980	1.0000
	1,000	0.8500	0.8460	0.9780	0.9940
0.5	2,000	1.0000	1.0000	1.0000	1.0000
	4,000	1.0000	1.0000	1.0000	1.0000
	1,000	0.2660	0.5300	0.6900	0.8240
0.75	2,000	0.9960	1.0000	1.0000	1.0000
	4,000	1.0000	1.0000	1.0000	1.0000

12 months preceding random assignment, and also five age-group dummies (22–24, 25–29, 30–35, 36–44, and 45–54), among others. Descriptive statistics can be found in Table 3. For simplicity, we group all applicants into three age categories (22–29, 30–35, and 36 and above), and pool all non-White applicants as minority applicants.

To implement the test, given $X_i = x$, we use the second-order Epanechnikov kernel and set the smoothing parameter to $2.34 \times \text{std}(\widehat{W}_{i:X_i=x}) \times n^{-1/4.5}$ when we

Table 3. Descriptive statistics for the National JTPA Study.

	All	$Z = 1$ (eligible)	$Z = 0$ (not eligible)
Men			
Number of observations	5,102	3,399	1,703
Training ($D = 1$)	41.87%	62.28%	1.12%
High school or GED	69.32%	69.26%	69.43%
Married	35.26%	36.01%	33.75%
Minorities	38.38%	38.69%	37.76%
Worked less than 13 weeks in the past year	40.02%	40.28%	39.05%
30-month earnings	19,147	19,520	18,404
Women			
Number of observations	6,102	4,088	2,014
Training ($D = 1$)	44.61%	65.73%	1.74%
High school or GED	72.06%	72.85%	70.45%
Married	21.93%	22.48%	20.82%
Minorities	40.41%	40.58%	51.86%
Worked less than 13 weeks in the past year	51.79%	51.75%	51.86%
30-month earnings	13,029	13,439	12,197

Note: Means are reported in this table for the National JTPA Study 30-month earnings sample.

estimate $\kappa(x, w)$. For the critical value, we use 10,000 multiplier bootstrap samples and search for the suprema by using 5,000 grid points. The p-value of our test is 0.5732. Therefore, the null hypothesis (i.e., no unobserved heterogenous treatment effects) cannot be rejected at the 10% significance level. Our results are robust to the size of bootstrap samples, the number of grid points, and the choices of bandwidth. Note that our results are consistent with Abadie et al. (2002), who estimate quantile treatment effects under a linear specification. In particular, one cannot reject the null hypothesis that quantile treatment effects are invariant across different quantile levels.[10]

5.2. The Impact of Fertility on Family Income

The second empirical illustration considers the heterogeneous impacts of children on parents' labor supply and income. Recently, Frölich and Melly (2013) studied the heterogeneous effects of fertility on family income within the general LATE

[10] We obtain a pointwise 95% confidence interval for each of the quantile treatment effects from Table III of Abadie et al. (2002) and find that these confidence intervals overlap. We can conclude no evidence against homogeneous treatment effects because a joint confidence band is, in general, wider than a pointwise confidence band.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

framework. To deal with the endogeneity of fertility decisions, Rosenzweig and Wolpin (1980), Angrist and Evans (1998), Bronars and Grogger (1994), and Jacobsen, Pearce, and Rosenbloom (1999), among many others, suggest using twin births as an IV.

Our data use the 1% and 5% Census Public Use Micro Sample (PUMS) from 1990 and 2000 censuses, consisting of 602,767 and 573,437 observations, respectively.[11] Similar to Frölich and Melly (2013), our sample is restricted to 21- to 35-year-old married mothers with at least one child, since we use twin birth as an instrument for fertility. The outcome variable of interest is the family's annual labor income.[12] The treatment variable is a dummy variable that takes the value 1 to indicate when a mother has two or more children. The IV is also a dummy variable and it equals 1 if the first birth is a twin. The covariates include mother's age, race, and educational level. Some covariates, i.e., age and years in education, are treated as continuous variables. Summary statistics can be found in Table 4.

Similar to the previous empirical illustration, we use the fourth-order Epanechnikov kernel and set the smoothing parameter to $2.34 \times \text{std}(X_i) \times n^{-1/3}$ when we estimate the influence function. For the critical value, we use $5,000$ bootstrapped samples and search for the suprema by using $1,000$ grids for each of the supports of both W and X's. The bandwidths are selected in the same manner as those in the JTPA case. The p-values of our tests are 0.0013 and 0.0007 for the 1990 and 2000 censuses, respectively. These results suggest that the null hypothesis, i.e., homogeneous treatment effects, should be rejected at all usual significance levels.

5.3. Extensions

When Z takes multiple values rather than being binary, one could extend our approach of testing for unobserved heterogeneous treatment effects. Namely, let $W = Y + (1-D)\delta(X)$, where $\delta(x) = \frac{\text{Cov}(Y,Z|X=x)}{\text{Cov}(D,Z|X=x)}$. Then, we test $\mathcal{H}_0$ by testing $W \perp\!\!\!\perp Z|X$. Since Z takes more than binary values in its support, this model restriction can be equivalently written as

$$F_{W|XZ}(\cdot|x,z) = F_{W|X}(\cdot|x), \ \forall \ (x,z) \in \mathcal{S}_{XZ}$$

or

$$\Pi(\cdot|x,z) = \mathbf{E}(\lambda(W-\cdot)|X=x), \forall \ (x,z) \in \mathcal{S}_{WXZ}$$

depending on whether covariates X or instruments Z contain any continuously distributed components or not.

Such a test, however, does not exploit model restrictions arising from multiplicity of Z. For instance, suppose that $\mathcal{S}_Z = \{0,1,2\}$. Under H_0 and Assumption 1, we have

[11] The data are publicly available at https://www.census.gov/programs-surveys/acs/microdata.html.

[12] It includes wages, salary, armed forces pay, commissions, tips, piece-rate payments, cash bonuses earned before deductions were made for taxes, bonds, pensions, union dues, etc. See Frölich and Melly (2013) for more details.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

TABLE 4. Descriptive statistics for the 1999 and 2000 censuses.

	1990			2000		
	All	$Z=1$ (twin birth)	$Z=0$ (no twin birth)	All	$Z=1$ (twin birth)	$Z=0$ (no twin birth)
Observations	602,767	6,524	596,243	573,437	8,569	564,868
Number of children	1.9276	2.5318	1.9209	1.8833	2.5196	1.8734
At least two children ($D=1$)	0.6500	1.0000	0.6461	0.6163	1.0000	0.6104
Mother						
Age in years	29.7894	29.9530	29.7876	30.0562	30.3943	30.0510
Years of education	12.9196	12.9623	12.9191	13.1131	13.2615	13.1108
Black	0.0637	0.0757	0.0636	0.0724	0.0816	0.07228
Asian	0.0326	0.0321	0.0326	0.0447	0.0335	0.0448
Other races	0.0537	0.0592	0.0536	0.0912	0.0806	0.0914
Currently at work	0.5781	0.5444	0.5785	0.5629	0.5132	0.5637
Usual hours per week	24.5660	23.3537	24.5795	25.1400	23.0491	25.1723
Wage or salary income last year	8,942	8,593	8,946	14,200	13,757	14,206

TABLE 4. Continued.

	1990			2000		
	All	$Z=1$ (twin birth)	$Z=0$ (no twin birth)	All	$Z=1$ (twin birth)	$Z=0$ (no twin birth)
Father						
Age in years	32.5358	32.7534	32.5333	32.9291	33.3102	32.9232
Years of education	13.0436	13.0748	13.0432	13.0331	13.1806	13.0308
Black	0.0671	0.0796	0.0670	0.0800	0.0945	0.0798
Asian	0.0291	0.0263	0.0292	0.0402	0.0318	0.0403
Other races	0.0488	0.0529	0.0488	0.0919	0.0802	0.0921
Currently at work	0.8973	0.8922	0.8974	0.8512	0.8584	0.8511
Usual hours per week	42.7636	42.7704	42.7635	43.8805	43.8789	43.8805
Wage or salary income last year	27,020	28,039	27,010	38,041	41,584	37,987
Parents						
Wage or salary income last year	35,963	36,632	35,956	52,241	55,342	52,193

Note: Data from the 1% and 5% PUMS in 1990 and 2000. Own calculations using the PUMS sample weights. The sample consists of married mothers between 21 and 35years of age with at least one child.

$$\frac{\mathbf{E}(Y|X=x,Z=z)-\mathbf{E}(Y|X=x,Z=z)}{p(x,z)-\mathbf{E}(D|X=x)}=\delta(x),\ \forall x.$$

As a matter of fact, our test does not exploit such a model restriction.

Our analysis naturally extends the case where the treatment variable D takes multiple values. For illustration, suppose $\mathcal{S}_D=\{0,1,2\}$. Under the homogeneous treatment effects hypothesis, denote $\delta_1(x)\equiv g(1,x,\cdot)-g(0,x,\cdot)$ and $\delta_2(x)\equiv g(2,x,\cdot)-g(0,x,\cdot)$. For $d=1,2$, let $p_d(X,Z)=P(D=d|X,Z)$ and $W_d\equiv\delta_d(X)+Y-\sum_{d'=1}^{2}\mathbb{1}(D=d')\times\delta_{d'}(X)$. Note that under H_0, i.e., $g(d,x,\cdot)-g(0,x,\cdot)=\delta_d(x)$, we have

$$W_d=g(d,X,\epsilon).$$

By a similar argument, we test for unobserved heterogeneous treatment effects by testing $W_d \perp\!\!\!\perp Z|X$, for $d=1,2$. To complete our analysis, it suffices to establish the identification of $\delta_d(x)$. Note that under $\mathcal{H}_0$ and Assumption 1, we have

$$\mathbf{E}(Y|X=x,Z=z)=\mathbf{E}[g(0,X,\epsilon)|X=x]+\delta_1(x)\times p_1(x,z)+\delta_2(x)\times p_2(x,z),\ \forall z.$$

Therefore, δ_d is identified if $\{(p_1(x,z),p_2(x,z))':z\in\mathcal{S}_{Z|X=x}\}$ has the full rank. Note that such a rank condition requires $\mathcal{S}_{Z|X=x}$ to contain at least three values.

APPENDIX

A. Proofs

A.1. *Proof of Proposition 1.* **Proof:** For the "if" part, under (3), we have

$$g(1,x,\epsilon)-g(0,x,\epsilon)=m(1,x)-m(0,x)\equiv\delta(x),\quad \forall x\in\mathcal{S}_X.$$

For the "only if" part, (2) implies

$$g(d,x,\epsilon)=d\times[g(1,x,\epsilon)-g(0,x,\epsilon)]+g(0,x,\epsilon)=d\times\delta(x)+g(0,x,\epsilon).$$

Therefore, (3) holds in the sense $m(d,x)=d\times\delta(x)$ and $\nu(x,\epsilon)=g(0,x,\epsilon)$.

A.2. *Proof of Lemma 1.* **Proof:** The first part of Lemma 1 is straightforward given the discussion before Lemma 1. We now show the second part. It suffices to construct a structure that can rationalize the data and also satisfy (1), Assumption 1, and $\mathcal{H}_0$. Given the observed data, denoted by F^*_{YDXZ}, we now construct a data generating structure for it. In the following proof, we use $Q^*_{W|X}$ to denote the quantile function of W given X, obtained from F^*_{YDXZ}. Similarly, we define $\delta^*(x)$ and $p^*(x,z)$. To begin with our construction, let $\epsilon\sim U[0,1]$, $X\sim F^*_X$, and $Z\sim F^*_Z$. Moreover, let X, ϵ, and Z be mutually independent for $F_{XZ\epsilon}$. To complete our construction, it suffices to define the probability distribution $\mathbf{P}(D=1|X,Z,\epsilon)$ and the function g for Y. Let

$$g(d,x,\tau)=Q^*_{W|X}(\tau|x)-(1-d)\times\delta^*(x),$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

and $Y = g(D, X, \epsilon)$. Regarding $\mathbf{P}(D = 1|X, Z, \epsilon)$, given that we have constructed $F_{\epsilon|XZ}$, we can equivalently define the joint distribution of (D, ϵ) given X and Z. Furthermore, let

$$\mathbf{P}(D = 1; \epsilon \le \tau|X = x, Z = z) = p^*(x, z) \times F^*_{Y|DXZ}(Q^*_{W|X}(\tau|x)|1, x, z), \ \ \forall\, x, z.$$

By construction, Assumption 1 and $\mathcal{H}_0$ are satisfied. Thus, it suffices to show the observational equivalence. First, let $\tau = 1$ in the construction of $P(D = 1; \epsilon \le \tau|X, Z)$, then it follows that $\mathbf{P}(D = 1|X = x, Z = z) = p^*(x, z)$. Moreover, note that

$$\begin{aligned}\mathbf{P}(Y \le y|D = 1, X = x, Z = z) &= \frac{\mathbf{P}(g(1, x, \epsilon) \le y; D = 1|X = x, Z = z)}{\mathbf{P}(D = 1|X = x, Z = z)}\\ &= \frac{\mathbf{P}(Q^*_{W|X}(\epsilon|x) \le y; D = 1|X = x, Z = z)}{\mathbf{P}(D = 1|X = x, Z = z)}\\ &= F^*_{Y|DXZ}(y|1, x, z).\end{aligned}$$

A.3. *Proof of Theorem 1.* **Proof:** Because Proposition 1 provides the "only if" part, then it suffices to show the "if" part. Suppose $W \perp\!\!\!\perp Z|X$. Let $\tilde{\delta}(X) = \frac{\mathrm{Cov}(Y, Z|X)}{\mathrm{Cov}(D, Z|X)}$. By the definition of W,

$$W \equiv Y + (1 - D) \times \frac{\mathrm{Cov}(Y, Z|X)}{\mathrm{Cov}(D, Z|X)} = Y + (1 - D)\tilde{\delta}(X).$$

Thus, for any $y \in \mathrm{R}$,

$$\begin{aligned}&\mathbf{P}(Y \le y, D = 1|X, Z = 1) + \mathbf{P}(Y + \tilde{\delta}(X) \le y, D = 0|X, Z = 1)\\ &\quad = \mathbf{P}(Y \le y, D = 1|X, Z = 0) + \mathbf{P}(Y + \tilde{\delta}(X) \le y, D = 0|X, Z = 0).\end{aligned}$$

It follows that

$$\begin{aligned}&\mathbf{P}(Y \le y, D = 1|X, Z = 1) - \mathbf{P}(Y \le y, D = 1|X, Z = 0)\\ &\quad = \mathbf{P}(Y \le y - \tilde{\delta}(X), D = 0|X, Z = 0) - \mathbf{P}(Y \le y - \tilde{\delta}(X), D = 0|X, Z = 1).\end{aligned} \tag{13}$$

Denote

$$\begin{aligned}\Delta_0(\tau, x) &\equiv \mathbf{P}(\nu(X, \epsilon) \le \tau, D = 0|X = x, Z = 1) - \mathbf{P}(\nu(X, \epsilon) \le \tau, D = 0|X = x, Z = 0),\\ \Delta_1(\tau, x) &\equiv \mathbf{P}(\nu(X, \epsilon) \le \tau, D = 1|X = x, Z = 0) - \mathbf{P}(\nu(X, \epsilon) \le \tau, D = 1|X = x, Z = 1).\end{aligned}$$

By Assumptions 1 and 3, we have

$$\Delta_0(\tau, x) = \mathbf{P}(\nu(X, \epsilon) \le \tau, \eta \in \mathcal{C}_x|X = x) = \Delta_1(\tau, x),$$

which is strictly monotone in $\tau \in \mathcal{S}_{\nu(X,\epsilon)|X=x,\ \eta\in\mathcal{C}_x}$. Moreover, there is $\mathcal{S}_{\nu(X,\epsilon)|X=x,\ \eta\in\mathcal{C}_x} = \mathcal{S}_{\nu(X,\epsilon)|X=x}$ under Assumptions 2 and 4.

Therefore, we have

$$\begin{aligned}&\mathbf{P}(Y \le y, D = 1|X = x, Z = 0) - \mathbf{P}(Y \le y, D = 1|X = x, Z = 1)\\ &\quad = \Delta_1(\tilde{g}^{-1}(1, x, y), x)\\ &\quad = \Delta_0(\tilde{g}^{-1}(1, x, y), x)\end{aligned}$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

$$= \mathbf{P}(Y \le \tilde{g}(0,x,\tilde{g}^{-1}(1,x,y)), D=0|X=x, Z=1)$$
$$- \mathbf{P}(Y \le \tilde{g}(0,x,\tilde{g}^{-1}(1,x,y)), D=0|X=x, Z=0),$$

where $\tilde{g}^{-1}(1,x,\cdot)$ is the inverse function of $\tilde{g}(1,x,\cdot)$, and $\tilde{g}$ is a monotone function introduced in Assumption 2. Note that both sides are strictly monotone in $y \in \mathcal{S}_{\tilde{g}(1,X,V)|X=x}$ since $\Delta_d(\cdot,x)$ is strictly monotone on $\mathcal{S}_{\nu(X,\epsilon)|X=x}$ under Assumption 4.

Combine the above result with (13), then we have

$$\tilde{g}(0,x,\tilde{g}^{-1}(1,x,y)) = y - \tilde{\delta}(x), \quad \forall x \in \mathcal{S}_X,\ y \in \mathcal{S}_{\tilde{g}(1,x,\nu(X,\epsilon))|X=x}.$$

Let $y = \tilde{g}(1,x,\tau)$ for some $\tau \in \mathcal{S}_{\nu(X,\epsilon)|X=x}$. Then, the above equation becomes

$$\tilde{g}(0,x,\tau) = \tilde{g}(1,x,\tau) - \tilde{\delta}(x).$$

A.4. *Proof of Theorem 2.* **Proof:** Let $\mathbb{1}^*_{WXZ}(w,x,z) = \mathbb{1}(W \le w) \times \mathbb{1}_{XZ}(x,z)$ and $\mathbb{1}^*_{\widehat{W}XZ}(w,x,z) = \mathbb{1}(\widehat{W} \le w) \times \mathbb{1}_{XZ}(x,z)$. Furthermore, let $\mathbb{1}^*_{W(\tilde{\delta})XZ}(w,x,z) = \mathbb{1}(W(\tilde{\delta}) \le w) \times \mathbb{1}_{XZ}(x,z)$, where $W(\tilde{\delta}) = Y + (1-D)\tilde{\delta}(X)$, be a function indexed by $\tilde{\delta}(\cdot) \in \mathrm{R}^{\mathcal{S}_X}$. By the definition, $\mathbb{1}^*_{W(\delta)XZ}(w,x,z) = \mathbb{1}^*_{WXZ}(w,x,z)$, and $\mathbb{1}^*_{W(\hat{\delta})XZ}(w,x,z) = \mathbb{1}^*_{\widehat{W}XZ}(w,x,z)$.

We first derive the asymptotic of $\sqrt{n}[\widehat{F}_{W|XZ}(w|x,z) - F_{W|XZ}(w|x,z)]$. By the definition,

$$F_{W|XZ}(w|x,z) = \frac{\mathbf{E}[\mathbb{1}^*_{WXZ}(w,x,z)]}{\mathbf{E}[\mathbb{1}_{XZ}(x,z)]} \text{ and } \widehat{F}_{W|XZ}(w|x,z) = \frac{\mathbf{E}_n[\mathbb{1}^*_{\widehat{W}XZ}(w,x,z)]}{\mathbf{E}_n[\mathbb{1}_{XZ}(x,z)]}.$$

In the expectation $\mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)]$ discussed below, we treat $\hat{\delta}$ as an index rather than a random object. Note that

$$\mathbf{E}_n[\mathbb{1}^*_{\widehat{W}XZ}(\cdot,x,z)] = \mathbf{E}_n[\mathbb{1}^*_{WXZ}(\cdot,x,z)] - \mathbf{E}[\mathbb{1}^*_{WXZ}(\cdot,x,z)] + \mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)]$$
$$+ \left\{\mathbf{E}_n[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)] - \mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)] - \mathbf{E}_n[\mathbb{1}^*_{W(\delta)XZ}(\cdot,x,z)] + \mathbf{E}[\mathbb{1}^*_{W(\delta)XZ}(\cdot,x,z)]\right\}$$
$$= \mathbf{E}_n[\mathbb{1}^*_{WXZ}(\cdot,x,z)] - \mathbf{E}[\mathbb{1}^*_{WXZ}(\cdot,x,z)] + \mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)] + o_p(n^{-1/2}),$$

where the last step follows from the fact that $\sqrt{n}\big(\mathbf{E}_n[\mathbb{1}^*_{W(\delta)XZ}(\cdot,x,z)] + \mathbf{E}[\mathbb{1}^*_{W(\delta)XZ}(\cdot,x,z)]\big)$ is stochastically equicontinuous by the empirical process theory (see, e.g., van der Vaart and Wellner, 2007). By the Taylor expansion,

$$\sqrt{n}\ \left\{\mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)] - F_{W|XZ}(w|x,z)\right\} = \frac{\partial \mathbf{E}[\mathbb{1}^*_{W(\delta)XZ}(w,x,z)]}{\partial \delta} \times \sqrt{n}(\hat{\delta} - \delta) + o_p(1).$$

Note that $\frac{\partial \mathbf{E}[\mathbb{1}^*_{W(\delta)XZ}(w,x,z)]}{\partial \delta(x')} = 0$, for all $x' \neq x$, and $\frac{\partial \mathbf{E}[\mathbb{1}^*_{W(\delta)XZ}(w,x,z)]}{\partial \delta(x)} = -f_{W|DXZ}(w|0,x,z) \times \mathbf{P}(D=0, X=x, Z=z)$. Therefore, we have

$$\sqrt{n}\ \left\{\mathbf{E}[\mathbb{1}^*_{W(\hat{\delta})XZ}(\cdot,x,z)] - F_{W|XZ}(w|x,z)\right\}$$
$$+ \sqrt{n}\left\{\mathbf{E}_n[\mathbb{1}^*_{WXZ}(\cdot,x,z)] - \mathbf{E}[\mathbb{1}^*_{WXZ}(\cdot,x,z)]\right\} - f_{WDXZ}(w,0,x,z) \times \sqrt{n}[\hat{\delta}(x) - \delta(x)] + o_p(1).$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

Moreover, $\mathbf{E}_n[\mathbb{1}_{XZ}(x,z)] = \mathbf{P}(X=x, Z=z) + O_p(n^{-1/2})$ under the central limit theorem. Thus, by Slutsky's theorem, we have

$$\begin{aligned}
&\sqrt{n}\left[\widehat{F}_{W|XZ}(w|x,1) - \widehat{F}_{W|XZ}(w|x,0)\right] - \sqrt{n}\left[F_{W|XZ}(w|x,1) - F_{W|XZ}(w|x,0)\right] \\
&= \frac{\sqrt{n}\left\{\mathbf{E}_n[\mathbb{1}^*_{WXZ}(w,x,1)] - \mathbf{E}[\mathbb{1}^*_{WXZ}(w,x,1)]\right\} - f_{WDXZ}(w,0,x,1) \times \sqrt{n}[\hat{\delta}(x) - \delta(x)]}{\mathrm{P}(X=x, Z=1)} \\
&\quad - \frac{\sqrt{n}\left\{\mathbf{E}_n[\mathbb{1}^*_{WXZ}(w,x,0)] - \mathbf{E}[\mathbb{1}^*_{WXZ}(w,x,0)]\right\} - f_{WDXZ}(w,0,x,0) \times \sqrt{n}[\hat{\delta}(x) - \delta(x)]}{\mathbf{P}(X=x, Z=0)} \\
&\quad + \frac{\sqrt{n}\mathbf{P}(W \le w, X=x, Z=1)}{\mathbf{E}_n \mathbb{1}_{XZ}(x,1)} - \frac{\sqrt{n}\mathbf{P}(W \le w, X=x, Z=0)}{\mathbf{E}_n \mathbb{1}_{XZ}(x,0)} + o_p(1).
\end{aligned}$$

By applying the Taylor expansion, we have

$$\begin{aligned}
&\frac{\sqrt{n}\mathbf{P}(W \le w, X=x, Z=z)}{\mathbf{E}_n \mathbb{1}_{XZ}(x,z)} - \sqrt{n}\, F_{W|XZ}(w|x,z) \\
&\qquad = -F_{W|XZ}(w|x,z) \times \frac{\sqrt{n}\left[\mathbf{E}_n \mathbb{1}_{XZ}(x,z) - \mathbf{P}(X=x, Z=z)\right]}{\mathbf{P}(X=x, Z=z)} + o_p(1).
\end{aligned}$$

Moreover, applying Lemma 3, we have

$$\begin{aligned}
&\sqrt{n}\left[\widehat{F}_{W|XZ}(w|x,1) - \widehat{F}_{W|XZ}(w|x,0)\right] - \sqrt{n}\left[F_{W|XZ}(w|x,1) - F_{W|XZ}(w|x,0)\right] \\
&= \sqrt{n}\mathbf{E}_n\left\{[\mathbb{1}(W \le w) - F_{W|XZ}(w|x,1)] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)}\right\} \\
&\quad - \sqrt{n}\mathbf{E}_n\left\{[\mathbb{1}(W \le w) - F_{W|XZ}(w|x,0)] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x, Z=0)}\right\} \\
&\quad + \kappa(w,x) \times \sqrt{n}\mathbf{E}_n\left\{\left[W - \mathbf{E}(W|X=x, Z=0)\right] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)}\right\} \\
&\quad - \kappa(w,x) \times \sqrt{n}\mathbf{E}_n\left\{\left[W - \mathbf{E}(W|X=x, Z=1)\right] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x, Z=0)}\right\} + o_p(1).
\end{aligned}$$

Under the null hypothesis, there is

$$\begin{aligned}
&\sqrt{n}\left[\widehat{F}_{W|XZ}(w|x,1) - \widehat{F}_{W|XZ}(w|x,0)\right] \\
&= \sqrt{n}\mathbf{E}_n\left\{[\mathbb{1}(W \le w) - F_{W|X}(w|x)] \times \left[\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x, Z=0)}\right]\right\} \\
&\quad + \kappa(w,x) \times \sqrt{n}\mathbf{E}_n\left\{\left[W - \mathbf{E}(W|X=x)\right] \times \left[\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x, Z=0)}\right]\right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\psi_{wx,i} + \phi_{wx,i}\right) + o_p(1),
\end{aligned}$$

where $\psi_{wx,i}$ and $\phi_{wx,i}$ are defined by (7) and (8). Note that, for each $x \in \mathcal{X}$, $\{\mathbb{1}(W \le w) - F_{W|X}(w|x) : w \in \mathcal{W}\}$ is a type I class of functions according to Andrews (1994), and this implies that $\{\mathbb{1}(W \le w) - F_{W|X}(w|x) : w \in \mathcal{W}, x \in \mathcal{X}\}$ satisfies Pollard's entropy condition by Theorem 3 of Andrews (1994). Note that $\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x,Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x,Z=0)}$ is a measurable

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

function, so it follows that

$$\left\{ \begin{array}{ll} \psi_{wx} & = (\mathbb{1}(W \le w) - F_{W|X}(w|x)) \\ & \quad \times \left(\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x,Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x,Z=0)} \right) : w \in \mathcal{W}, x \in \mathcal{X} \end{array} \right\}$$

satisfies Pollard's entropy condition. Furthermore, by Assumption 5, we have that, for all x,

$$\left\{ \phi_{wx} = \kappa(w,x)(W - E[W|X=x]) \times \left(\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x,Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x,Z=0)} \right) : w \in \mathcal{W} \right\}$$

is a Vapnik-Chervonenkis (VC) class of function according to Kosorok (2008), and it follows that

$$\left\{ \begin{array}{ll} \phi_{wx} & = \kappa(w,x)(W - E[W|X=x]) \\ & \quad \times \left(\frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x,Z=1)} - \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x,Z=0)} \right) : w \in \mathcal{W}, x \in \mathcal{X} \end{array} \right\}$$

is a VC class of function that satisfies Pollard's entropy condition. It follows that $\{\psi_{wx} + \phi_{wx} : w \in \mathcal{W}, x \in \mathcal{X}\}$ satisfies Pollard's entropy condition. Then, combined with the previous results and by Theorem 1 of Andrews (1994), we can show that

$$\sqrt{n}(\widehat{F}_{\widehat{W}|XZ}(w|x,0) - \widehat{F}_{\widehat{W}|XZ}(w|x,1) - (F_{W|XZ}(w|x,0) - F_{W|XZ}(w|x,1))) \Rightarrow \mathcal{Z}(w,x). \tag{14}$$

Then, by the continuous mapping theorem (see, e.g., van der Vaart and Wellner, 2007), and under the null, we have $\widehat{\mathcal{T}}_n \xrightarrow{d} \sup_{w \in \mathrm{R};\ x \in \mathcal{S}_X} |\mathcal{Z}(w,x)|$. In addition, (14) implies that

$$\sup_{(w,x)\in S_{WX}} \left| (\widehat{F}_{\widehat{W}|XZ}(w|x,0) - \widehat{F}_{\widehat{W}|XZ}(w|x,1) - (F_{W|XZ}(w|x,0) - F_{W|XZ}(w|x,1))) \right| \xrightarrow{p} 0.$$

It follows that under H_1, we have $n^{-1/2}\widehat{\mathcal{T}}_n \xrightarrow{p} \sup_{(w,x)\in S_{WX}} |F_{W|XZ}(w|x,0) - F_{W|XZ}(w|x,1)| > 0$.

A.5. *Proof of Theorem 3.* The proof of Theorem 3 would follow standard arguments once we establish the validity of the multiplier bootstrapped processes. Note that under Assumption 7, we have that

$$\left| \frac{1}{n} \sum_{i=1}^{n} (\hat{\psi}_{wx,i} + \hat{\phi}_{wx,i}) \cdot (\hat{\psi}_{w'x',i} + \hat{\phi}_{w'x',i}) - Cov[\mathcal{Z}(w,x), \mathcal{Z}(w',x')] \right| \xrightarrow{p} 0$$

uniformly over $((w,x),(w',x')) \in \mathcal{S}_{WX}^2$ by the uniform law of large numbers and the uniform consistency of the various estimators in the estimated influence functions. That is, the covariance kernel of the simulated processes converges to the covariance kernel of $\mathcal{Z}(w,x)$ uniformly. Then, by similar arguments as in Hsu (2017), we can show that $\mathcal{Z}^u(\cdot,\cdot) \overset{p}{\Rightarrow} \mathcal{Z}(\cdot,\cdot)$. Given this result, the size and power properties of our test follow standard arguments such as Andrews (1997).

Discussion on Assumption 7: Here, we provide estimators and low-level conditions so that Assumption 7 would be satisfied. It is straightforward to see that $\hat{\delta}(x)$ satisfies

Assumption 7(iv). For $x, z \in \mathcal{S}_{XZ}$, let

$$\widehat{\mathbf{P}}(X = x, Z = z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{XZ}(x,z), \tag{15}$$

$$\hat{p}(x,z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{DXZ}(1,x,z) \Big/ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{XZ}(x,z).$$

Because X and Z take on a finite number of values, it is straightforward to see that $\widehat{\mathbf{P}}(X = x, Z = z)$ and $\hat{p}(x,z)$ given in (15) satisfy Assumption 7(i) and (ii), respectively.

We propose an estimator for $f_{WD|XZ}(w,0|x,z)$. For $w \in \mathcal{W} = [w_\ell, w_u]$, let

$$\tilde{f}_{WD|XZ}(w,0|x,z) = \frac{\sum_{i=1}^{n} K_{W_i,h}(w)\mathbb{1}(D_i = 0)\mathbb{1}_{X_iZ_i}(x,z)}{\sum_{i=1}^{n} \mathbb{1}_{X_iZ_i}(x,z)}$$

and

$$\hat{f}_{WD|XZ}(w,0|x,z) = \begin{cases} \tilde{f}_{WD|XZ}(w_\ell + h, 0|x,z), & \text{if } w \in [w_\ell, w_\ell + h), \\ \tilde{f}_{WD|XZ}(w, 0|x,z), & \text{if } w \in (w_\ell + h, w_u - h), \\ \tilde{f}_{WD|XZ}(w_u - h, 0|x,z), & \text{if } w \in (w_u - h, w_u]. \end{cases} \tag{16}$$

Assumption 13. Assume that

(i) $K(u)$ is nonnegative and has support $[-1,1]$. $K(u)$ is symmetric around 0 and is continuously differentiable of order 1.
(ii) The bandwidth h satisfies $h \to 0$, $nh^4 \to \infty$, and $nh/\log(n) \to \infty$, as $n \to \infty$.

Then, under Assumption 13 and by the argument of the proof of Lemma 3.2 of Donald, Hsu, and Barrett (2012), we show that the $\hat{f}_{WD|XZ}(w,0|x,z)$ given in (16) satisfies Assumption 7. Finally, for kernel functions, we can pick an Epanechnikov kernel that would satisfy Assumption 13(i), and for h, we can set $h = \hat{\sigma}_w \times n^{-1/4.5}$.

A.6. *Proof of Lemma 2.* **Proof:** Fix the value of X as x, and let $z = 1$ without loss of generality. Note that

$$\begin{aligned}
&\widehat{G}(w,x,1) - \widetilde{G}(w,x,1) \\
&= \mathbf{E}_n\left\{\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(w - \widehat{W})\left[\mathbb{1}(\widehat{W} \le w) - \mathbb{1}(W \le w)\right]\right\} \\
&= \mathbf{E}_n\left\{\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(w - \widehat{W})\left[\mathbb{1}(\widehat{W} \le w) - \mathbb{1}(W \le w)\right] \times \mathbb{1}(|W - w| \le n^{-r})\right\} \\
&\quad + \mathbf{E}_n\left\{\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(w - \widehat{W})\left[\mathbb{1}(\widehat{W} \le w) - \mathbb{1}(W \le w)\right] \times \mathbb{1}(|W - w| > n^{-r})\right\} \\
&\equiv \mathrm{T}_1 + \mathrm{T}_2,
\end{aligned}$$

where $r \in (\frac{1}{4}, \iota)$. It suffices to show both T_1 and T_2 are $o_p(n^{-\frac{1}{2}})$.

First, note that

$$\begin{aligned}
\mathrm{T}_1 = &\mathbf{E}_n\left\{\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(w - W)\left[\mathbb{1}(\widehat{W} \le w) - \mathbb{1}(W \le w)\right] \times \mathbb{1}(|W - w| \le n^{-r})\right\} \\
&+ \mathbf{E}_n\left\{\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(W - \widehat{W})\left[\mathbb{1}(\widehat{W} \le w) - \mathbb{1}(W \le w)\right] \times \mathbb{1}(|W - w| \le n^{-r})\right\}.
\end{aligned}$$

This is because

$$\mathbf{E}\left\{\left|\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(w-W)\left[\mathbb{1}(\widehat{W}\le w)-\mathbb{1}(W\le w)\right]\times\mathbb{1}(|W-w|\le n^{-r})\right|\right\}$$
$$\le\mathbf{E}\left\{\left|\hat{f}_{XZ}(X_1,0)\times(w-W)\times\mathbb{1}(|W-w|\le n^{-r})\right|\right\}=O(1)\times O(n^{-2r})=o(n^{-\frac{1}{2}}),$$

where the last step holds when $r>\frac{1}{4}$. Moreover,

$$\mathbf{E}\left\{\left|\mathbb{1}^*_{XZ}(x,1)\hat{f}_{XZ}(X,0)(W-\widehat{W})\left[\mathbb{1}(\widehat{W}\le w)-\mathbb{1}(W\le w)\right]\times\mathbb{1}(|W-w|\le n^{-r})\right|\right\}$$
$$\le\mathbf{E}\left\{\left|\hat{f}_{XZ}(X_1,0)\times(W-\widehat{W})\times\mathbb{1}(|W-w|\le n^{-r})\right|\right\}=O(1)\times O(n^{-\iota})\times O(n^{-r})=o(n^{-\frac{1}{2}}).$$

Then, we have $\mathrm{T}_1=o_p(n^{-\frac{1}{2}})$.

For term T_2, note that

$$\mathbf{E}[|\mathrm{T}_2|]\le\frac{\overline{K}}{h}\times\sqrt{\mathbf{E}[(w-\widehat{W})^2]}\times\sqrt{\mathbf{P}\left(|\widehat{W}-W|>n^{-r}\right)}$$
$$\le\frac{\overline{K}}{h}\times\sqrt{\mathbf{E}[\widehat{W}^2]-2w\times\mathbf{E}[\widehat{W}]+w^2}\times\sqrt{\mathbf{P}\left[|\hat{\delta}(X)-\delta(X)|>n^{-r}\right]},$$

where $\overline{K}$ is the upper bound of $K(\cdot)$. Because W is a bounded random variable and w belongs to a compact set, then $\sqrt{\mathbf{E}[\widehat{W}^2]-2w\times\mathbf{E}[\widehat{W}]+w^2}=O(1)$. Moreover, by Lemma 4, $\mathbf{E}|\mathrm{T}_2|\le o(n^{-k})$, for any $k>0$. Hence, $\mathrm{T}_2=o_p(n^{-\frac{1}{2}})$.

A.7. *Proof of Theorem 4.* **Proof:** By Lemma 2, we have

$$\widehat{\mathcal{T}}^c_n=\sqrt{n}\left|\widetilde{G}(w,x,1)-\widetilde{G}(w,x,0)\right|+o_p(1).$$

Note that

$$\widetilde{G}(w,x,z)=\mathrm{U}_1(w,x,z)+\mathrm{U}_2(w,x,z)+o_p(n^{-1/2}),$$

where

$$\mathrm{U}_1(w,x,z)\equiv\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}^*_{W_i}(w)\times\mathbb{1}^*_{X_iZ_i}(x,z)\times\hat{f}_{XZ}(X_i,z')\times(W_i-\widehat{W}_i),$$

$$\mathrm{U}_2(w,x,z)\equiv\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}^*_{W_i}(w)\times\mathbb{1}^*_{X_iZ_i}(x,z)\times\hat{f}_{XZ}(X_i,z')\times(w-W_i).$$

Therefore,

$$\begin{aligned}&\sqrt{n}\left[\widetilde{G}(w,x,1)-\widetilde{G}(w,x,0)\right]\\&\quad=\sqrt{n}\{\mathrm{U}_1(w,x,1)-\mathrm{U}_1(w,x,0)-[\mathbf{E}\mathrm{U}_1(w,x,1)-\mathbf{E}\mathrm{U}_1(w,x,0)]\}\\&\qquad+\sqrt{n}\{\mathrm{U}_2(w,x,1)-\mathrm{U}_2(w,x,0)-[\mathbf{E}\mathrm{U}_2(w,x,1)-\mathbf{E}\mathrm{U}_2(w,x,0)]\}\\&\qquad+\sqrt{n}[\mathbf{E}\mathrm{U}_1(w,x,1)-\mathbf{E}\mathrm{U}_1(w,x,0)]+\sqrt{n}[\mathbf{E}\mathrm{U}_2(w,x,1)-\mathbf{E}\mathrm{U}_2(w,x,0)].\end{aligned}$$

We first discuss U_2 terms. By definition,

$$\mathrm{U}_2(w,x,z)=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\{\mathbb{1}^*_{X_iZ_i}(x,z)\lambda(W_i-w)\times K_{X_j,h}(X_i)\mathbb{1}(Z_j=z')\}$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \zeta_{n,ij}(w,x,z),$$

where $\zeta_{n,ij}(w,x,z) = \mathbb{1}^*_{X_i Z_i}(x,z) \times \lambda(W_i - w) \times K_{X_j,h}(X_i) \times \mathbb{1}(Z_j = z')$.

Let $\zeta^*_{n,ij}(w,x,z) = \frac{1}{2}\left[\zeta_{n,ij}(w,x,z) + \zeta_{n,ji}(w,x,z)\right]$. Then, $\zeta^*_{n,ij}$ is symmetric in indices i and j. Therefore,

$$\mathrm{U}_2(w,x,z) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \zeta^*_{n,ij}(w,x,z),$$

which is a $\mathcal{U}$-process indexed by (w,x,z_ℓ). By Nolan and Pollard (1988, Thm. 5) and Powell et al. (1989, Lem. 3.1),

$$\begin{aligned} &\mathrm{U}_2(w,x,z) - \mathbf{E}[\mathrm{U}_2(w,x,z)] \\ &= \frac{2}{n} \sum_{i=1}^{n} \left\{ \mathbf{E}[\zeta^*_{n,ij}(w,x,z) | Y_i, D_i, X_i, Z_i] - \mathbf{E}[\zeta^*_{n,ij}(w,x,z)] \right\} + o_p(n^{-1/2}), \end{aligned}$$

where the $o_p(n^{-1/2})$ applies uniformly over (w,x). Note that

$$\begin{aligned} &\mathbf{E}[\zeta^*_{n,ij}(w,x,z) | Y_i, D_i, X_i, Z_i] \\ &\quad = \frac{1}{2} \left\{ \mathbb{1}^*_{XZ}(x,z) f_{XZ}(X,z') \lambda(W-w) + \mathbb{1}^*_{XZ}(x,z') f_{XZ}(X,z) \Pi(w|X,z) \right\} + o_p(1). \end{aligned}$$

Next, we derive $\mathbf{E}[\zeta^*_{n,ij}(w,x,z)]$. Let $u_1(w,x,z) = \mathbf{E}[\mathbb{1}^*_{XZ}(x,z) f_{XZ}(X,z') \lambda(W-w)]$ and $u_2(w,x,z) = \mathbf{E}[\mathbb{1}^*_{XZ}(x,z') f_{XZ}(X,z) \Pi(w|X,z)]$. Note that under $\mathcal{H}_0$

$$u_1(w,x,z) = u_2(w,x,z) = \int \mathbb{1}(X \leq x) \Pi(w|X) f_{X|Z}(X|1) f_{X|Z}(X|0) dX \times \mathbf{P}(Z=1) \mathbf{P}(Z=0),$$

invariant with z. Therefore, $\mathbf{E}[\zeta^*_{n,ij}(w,x,z)] = \frac{1}{2}[u_1(w,x,z) + u_2(w,x,z)]$ is also invariant with z. Let $u^e(w,x) = \mathbf{E}[\zeta^*_{n,ij}(w,x,z)]$. Moreover, by Powell et al. (1989, Thm. 3.1),

$$\begin{aligned} &\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \mathbf{E}[\zeta^*_{n,ij}(w,x,z) | Y_i, D_i, X_i] - \mathbf{E}[\zeta^*_{n,ij}(w,x,z)] \right\} \\ &\quad = \mathbf{E}_n \left\{ \mathbb{1}^*_{XZ}(x,z) f_{XZ}(X,z') \lambda(W-w) - u^e(w,x) \right\} \\ &\qquad + \mathbf{E}_n \left\{ \mathbb{1}^*_{XZ}(x,z') f_{XZ}(X,z) \Pi(w|X,z) - u^e(w,x) \right\} + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

where the $o_p(n^{-1/2})$ holds uniformly over (w,x). Moreover, under $\mathcal{H}_0$, there is $\Pi(w|X,z) = \mathbf{E}(\lambda(W-w)|X)$. Thus,

$$\begin{aligned}
&\mathrm{U}_2(w,x,1) - \mathrm{U}_2(w,x,0) - [\mathbf{E}\mathrm{U}_2(w,x,1) - \mathbf{E}\mathrm{U}_2(w,x,0)] \\
&\quad = \mathbf{E}_n\left\{\left[\frac{\mathbb{1}^*_{XZ}(x,1)}{f_{XZ}(X,1)} - \frac{\mathbb{1}^*_{XZ}(x,0)}{f_{XZ}(X,0)}\right] f_{XZ}(X,0) f_{XZ}(X,1)\big[\lambda(W-w) - \mathbf{E}(\lambda(W-w)|X)\big]\right\} \\
&\qquad + o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

We now turn to $\mathrm{U}_1(w,x,z)$. Note that

$$\mathrm{U}_1(w,x,z) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{1}^*_{W_iX_iZ_i}(w,x,z) f_{XZ}(X_i,z')(1-D_i)[\hat{\delta}(X_i) - \delta(X_i)]\right\} + o_p(n^{-\frac{1}{2}}),$$

provided that $\mathbf{E}\left|\left[\hat{f}_{XZ}(X_i,z') - f_{XZ}(X_i,z')\right] \times \left[\hat{\delta}(X_i) - \delta(X_i)\right]\right| = o_p(n^{-\frac{1}{2}})$ holds. By a similar decomposition argument on $\hat{\delta}(X) - \delta(X)$ in Lemma 4, we have

$$\mathrm{U}_1(w,x,z) = -\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\xi_{n,ij}(w,x,z) + o_p(n^{-1/2}),$$

where $\xi_{n,ij}(w,x,z) = \mathbb{1}^*_{W_iX_iZ_i}(w,x,z) f_{XZ}(X_i,z')(1-D_i)\frac{[W_j - \mathbf{E}(W_j|X_i)]K_{X_j,h}(X_i)}{p(X_i,1) - p(X_i,0)}\left[\frac{\mathbb{1}(Z_j=1)}{f_{XZ}(X_i,1)} - \frac{\mathbb{1}(Z_j=0)}{f_{XZ}(X_i,0)}\right]$. Moreover, let $\xi^*_{n,ij}(w,x,z) = \frac{1}{2}[\xi_{n,ij}(w,x,z) + \xi_{n,ji}(w,x,z)]$.

By a similar argument for U_2,

$$\begin{aligned}
&\mathrm{U}_1(w,x,z) - \mathbf{E}[\mathrm{U}_1(w,x,z)] \\
&\quad = -\frac{2}{n}\sum_{i=1}^{n}\left\{\mathbf{E}[\xi^*_{n,ij}(w,x,z)|Y_i,D_i,X_i,Z_i] - \mathbf{E}[\xi^*_{n,ij}(w,x,z)]\right\} + o_p(n^{-1/2}).
\end{aligned}$$

Note that $\mathbf{E}[\xi_{n,ij}(w,x,z)|Y_i,D_i,X_i,Z_i] = 0$ and

$$\begin{aligned}
&\mathbf{E}[\xi_{n,ji}(w,x,z)|Y_i,D_i,X_i,Z_i] = \mathbf{E}\left\{\mathbf{E}[\xi_{n,ji}(w,x,z)|X_j,Z_j,Y_i,D_i,X_i,Z_i]\big|Y_i,D_i,X_i,Z_i\right\} \\
&= \mathbf{E}\Bigg\{\mathbb{1}^*_{X_jZ_j}(x,z) f_{XZ}(X_j,z')\mathbf{P}(W \le w; D=0|X_j,Z_j)[W_i - \mathbf{E}(W|X_j)] \\
&\quad \times \frac{K_{X_j,h}(X_i)}{p(X_j,1) - p(X_j,0)}\left[\frac{\mathbb{1}(Z_i=1)}{f_{XZ}(X_j,1)} - \frac{\mathbb{1}(Z_i=0)}{f_{XZ}(X_j,0)}\right]\Bigg|Y_i,D_i,X_i,Z_i\Bigg\} \\
&= F^*_{WD|XZ}(w,0|X_i,z)[W_i - \mathbf{E}(W|X_i)]\frac{f_{XZ}(X_i,0)f_{XZ}(X_i,1)}{p(X_i,1) - p(X_i,0)}\left[\frac{\mathbb{1}^*_{X_i,Z_i}(x,1)}{f_{XZ}(X_i,1)} - \frac{\mathbb{1}^*_{X_i,Z_i}(x,0)}{f_{XZ}(X_i,0)}\right] \\
&\quad + o_p(1),
\end{aligned}$$

where the last step comes from Bochner's lemma (see, e.g., Rudin, 1962) and uses the fact the integrant equals zero if $Z_j = z'$.

Thus, we have

$$\begin{aligned}
&\mathrm{U}_1(w,x,z) - \mathbf{E}[\mathrm{U}_1(w,x,z)] \\
&\quad = -\mathbf{E}_n\left\{[W - \mathbf{E}(W|X)]\frac{F^*_{WD|XZ}(w,0|X,z)}{p(X,1)-p(X,0)}\left[\frac{\mathbb{1}^*_{XZ}(x,1)}{f_{XZ}(X,1)} - \frac{\mathbb{1}^*_{XZ}(x,0)}{f_{XZ}(X,0)}\right] f_{XZ}(X,1) f_{XZ}(X,0)\right\} \\
&\qquad + o_p(n^{-\frac{1}{2}}),
\end{aligned}$$

where the $o_p(n^{-1/2})$ holds uniformly over (w,x). It follows that

$$\mathrm{U}_1(w,x,1) - \mathbf{E}\mathrm{U}_1(w,x,1) - [\mathrm{U}_1(w,x,0) - \mathbf{E}\mathrm{U}_1(w,x,0)] = \mathbf{E}_n\phi^c_{wx} + o_p(n^{-\frac{1}{2}}).$$

By Assumption 9, we have $\mathbf{E}[\mathrm{U}_1(w,x;z)] = o_p(n^{-\frac{1}{2}})$. Therefore, under $\mathcal{H}_0$,

$$\begin{aligned}
&\sqrt{n}\left[\widetilde{G}(w,x,1) - \widetilde{G}(w,x,0)\right] \\
&= \sqrt{n}\{\mathrm{U}_1(w,x,1) - \mathrm{U}_1(w,x,0) - \{\mathbf{E}[\mathrm{U}_1(w,x,1)] - \mathbf{E}[\mathrm{U}_1(w,x,0)]\}\} \\
&\quad + \sqrt{n}\{\mathrm{U}_2(w,x,1) - \mathrm{U}_2(w,x,0) - \{\mathbf{E}[\mathrm{U}_2(w,x,1)] - \mathbf{E}[\mathrm{U}_2(w,x,0)]\}\} + o_p(1) \\
&= \sqrt{n} \times \mathbf{E}_n(\psi^c_{wx} + \phi^c_{wx}) + o_p(1),
\end{aligned}$$

which converges to a zero-mean Gaussian process with the given covariance kernel.

A.8. *Proof of Theorem 5.* The proof is similar to that of Theorem 3, and we skip it for brevity.

Discussion on Assumptions 10–12: By Masry (1996) and Lemma A.1 of Abrevaya et al. (2015), we have the bias terms satisfying

$$\begin{aligned}
&\sup_{x\in\mathcal{S}^\xi_X}\left|\mathbf{E}\left[\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{Z_j}(z)K_{X_j,h}(x)\right] - f_{XZ}(x,z)\right| = O(h^k), \\
&\sup_{x\in\mathcal{S}^\xi_X}\left|\mathbf{E}\left[\frac{1}{n}\sum_{j=1}^n D_j\mathbb{1}_{Z_j}(z)K_{X_j,h}(x)\right] - p(x,z)f_{XZ}(x,z)\right| = O(h^k), \\
&\sup_{x\in\mathcal{S}^\xi_X}\left|\mathbf{E}\left[\frac{1}{n}\sum_{j=1}^n Y_j\mathbb{1}_{Z_j}(z)K_{X_j,h}(x)\right] - \mathbf{E}(Y|X=x,Z=z)f_{XZ}(x,z)\right| = O(h^k), \\
&\sup_{x\in\mathcal{S}^\xi_X}\left|\mathbf{E}[\hat{\delta}(x)] - \delta(x)\right| = O(h^k),
\end{aligned}$$

where k is the order of kernel K. According to Assumptions 10 and 11, we require that $h^k = o(n^{-1/2})$ and $\sqrt{nh^{d_X}}$ diverges at a rate faster than n^ι with $\iota > 1/4$. Therefore, if we pick a kth order Epanechnikov kernel with $k = 2d_X + 2$, $h = O(n^{1/(2d_X+1)})$, then Assumptions 10 and 11 will be satisfied. Note that under the same conditions, Assumption 12(i), (iii), and (iv) is also satisfied.

We next consider Assumption 12(ii). To estimate $E[W|X=x]$, we will replace W_i's with $\widehat{W}_i$'s in the kernel regressions. Because $|W_i - \widehat{W}_i| = o_p(1)$ uniformly, we can show that Assumption 12(ii) holds.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

We finally consider Assumption 12(iv). Define the estimator for $F_{W|DXZ}(w|0,x,z)$ as

$$\widehat{F}_{W|DXZ}(w|0,x,z) = \frac{\sum_{i=1}^{n} \mathbb{1}(\widehat{W}_i \leq w)\mathbb{1}(D_i = 0, Z_i = z)K^F_{X_i,h_F}(X_i)}{\sum_{i=1}^{n} \mathbb{1}(D_i = 0, Z_i = z)K^F_{X_i,h_F}(X_i)},$$

where K^F is a regular Epanechnikov kernel and h_F is the bandwidth. It is straightforward to see that if $h_F \to 0$ and $nh^{d_X} \to \infty$, then we can show that $\sup_{x\in\mathcal{S}_X^{\xi}, z\in\mathcal{S}_Z, w\in\mathcal{S}_W} |\widehat{F}_{W|DXZ}(w|0, x,z) - F_{W|DXZ}(w|0,x,z)| \xrightarrow{p} 0$. Given that the regular Epanechnikov kernel is always nonnegative, $\widehat{F}_{W|DXZ}(w|0,x,z)$ is monotonically increasing as well. Hence, Assumption 12(iv) also holds.

B. Technical Lemmas

Let $\Delta p(x) \equiv p(x,1) - p(x,0)$, which is strictly positive by Assumption 5.

LEMMA 3. *Suppose Assumptions 1 and 5 hold. Then, we have*

$$\sqrt{n}[\hat{\delta}(x) - \delta(x)] = \frac{1}{\Delta p(x)} \times \sqrt{n}\mathbf{E}_n\left\{\left[W - \mathbf{E}(W|X = x, Z = 0)\right] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X = x, Z = 1)}\right\} - \frac{1}{\Delta p(x)} \times \sqrt{n}\mathbf{E}_n\left\{\left[W - \mathbf{E}(W|X = x, Z = 1)\right] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X = x, Z = 0)}\right\} + o_p(1). \quad \mathbf{(17)}$$

Proof of Lemma 3: Fix $X = x$. For expositional simplicity, we suppress x in the following proof. Moreover, let $\mathrm{A}_n(z) = \mathbf{E}_n[Y\mathbb{1}_{XZ}(x,z)]$, $\mathrm{B}_n(z) = \mathbf{E}_n[D\mathbb{1}_{XZ}(x,z)]$, $\mathrm{C}_n(z) = \mathbf{E}_n\mathbb{1}_{XZ}(x,z)$, $\mathrm{A}(z) = \mathbf{E}[Y\mathbb{1}_{XZ}(x,z)]$, $\mathrm{B}(z) = \mathbf{E}[D1_{XZ}(x,z)]$, and $\mathrm{C}(z) = \mathbf{E}\mathbb{1}_{XZ}(x,z) = \mathbf{P}(X = x, Z = z)$. By definition, note that

$$\hat{\delta}(x) = \frac{\mathrm{A}_n(1)\mathrm{C}_n(0) - \mathrm{A}_n(0)\mathrm{C}_n(1)}{\mathrm{B}_n(1)\mathrm{C}_n(0) - \mathrm{B}_n(0)\mathrm{C}_n(1)} \quad \text{and} \quad \delta(x) = \frac{\mathrm{A}(1)\mathrm{C}(0) - \mathrm{A}(0)\mathrm{C}(1)}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)}.$$

It follows that

$$\hat{\delta}(x) - \delta(x) = \frac{\mathrm{A}_n(1)\mathrm{C}_n(0) - \mathrm{A}_n(0)\mathrm{C}_n(1) - [\mathrm{A}(1)\mathrm{C}(0) - \mathrm{A}(0)\mathrm{C}(1)]}{\mathrm{B}_n(1)\mathrm{C}_n(0) - \mathrm{B}_n(0)\mathrm{C}_n(1)} + \left\{\frac{\mathrm{A}(1)\mathrm{C}(0) - \mathrm{A}(0)\mathrm{C}(1)}{\mathrm{B}_n(1)\mathrm{C}_n(0) - \mathrm{B}_n(0)\mathrm{C}_n(1)} - \frac{\mathrm{A}(1)\mathrm{C}(0) - \mathrm{A}(0)\mathrm{C}(1)}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)}\right\} \equiv \mathrm{I} + \mathrm{II}.$$

We first look at term I. By the central limit theorem and Assumption 5, we have $\mathrm{A}_n(z) = \mathrm{A}(z) + O_p(n^{-1/2})$, $\mathrm{B}_n(z) = \mathrm{B}(z) + O_p(n^{-1/2})$, and $\mathrm{C}_n(z) = \mathrm{C}(z) + O_p(n^{-1/2})$. Therefore,

$$\mathrm{I} = \frac{[\mathrm{A}_n(1) - \mathrm{A}(1)]\mathrm{C}(0) + \mathrm{A}(1)[\mathrm{C}_n(0) - \mathrm{C}(0)]}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} - \frac{[\mathrm{A}_n(0) - \mathrm{A}(0)]\mathrm{C}(1) + \mathrm{A}(0)[\mathrm{C}_n(1) - \mathrm{C}(1)]}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} + o_p(n^{-1/2})$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

$$= \frac{\mathrm{A}_n(1)\mathrm{C}(0) - \mathrm{A}(0)\mathrm{C}_n(1) - \mathrm{A}_n(0)\mathrm{C}(1) + \mathrm{A}(1)\mathrm{C}_n(0)}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} + \frac{2[\mathrm{A}(0)\mathrm{C}(1) - \mathrm{A}(1)\mathrm{C}(0)]}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} + o_p(n^{-1/2}).$$

Specifically, we have

$$\begin{aligned}
\mathrm{I} &= \mathbf{E}_n\left\{\left[Y - \mathbf{E}(Y|X=x, Z=0)\right] \times \mathbb{1}_{XZ}(x,1)\right\} \times \frac{\mathbf{P}(X=x, Z=0)}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} \\
&\quad - \mathbf{E}_n\left\{\left[Y - \mathbf{E}(Y|X=x, Z=1)\right] \times \mathbb{1}_{XZ}(x,0)\right\} \times \frac{\mathbf{P}(X=x, Z=1)}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} \\
&\quad + \frac{2[\mathrm{A}(0)\mathrm{C}(1) - \mathrm{A}(1)\mathrm{C}(0)]}{\mathrm{B}(1)\mathrm{C}(0) - \mathrm{B}(0)\mathrm{C}(1)} + o_p(n^{-1/2}) \\
&= \frac{1}{\Delta p(x)} \times \mathbf{E}_n\left\{\left[Y - \mathbf{E}(Y|X=x, Z=0)\right] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)}\right\} \\
&\quad - \frac{1}{\Delta p(x)} \times \mathbf{E}_n\left\{\left[Y - \mathbf{E}(Y|X=x, Z=1)\right] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathbf{P}(X=x, Z=0)}\right\} - 2\delta(x) + o_p(n^{-1/2}).
\end{aligned}$$

For term II, by a similar argument, we have

$$\begin{aligned}
\mathrm{II} &= \frac{-\delta(x)}{\Delta p(x)} \times \mathbf{E}_n\left\{\left[D - p(x,0)\right] \times \frac{\mathbb{1}_{XZ}(x,1)}{\mathbf{P}(X=x, Z=1)}\right\} \\
&\quad + \frac{\delta(x)}{\Delta p(x)} \times \mathbf{E}_n\left\{\left[D - p(x,1)\right] \times \frac{\mathbb{1}_{XZ}(x,0)}{\mathrm{P}(X=x, Z=0)}\right\} + 2\delta(x) + o_p(n^{-1/2}).
\end{aligned}$$

By definition of W, we have $W - \mathbf{E}[W|X=x, Z=z] = Y - \mathbf{E}[Y|X=x, Z=z] - [D - p(x,z)] \times \delta(x)$. Summing up I and II, we obtain (17).

LEMMA 4. *Suppose Assumptions 8–10 hold. Then, for any* $k > 0$ *and* $r \in (\frac{1}{4}, \iota)$*,*

$$\sup_{x \in \mathcal{S}_X} n^k \times \mathbf{P}\left[|\hat{\delta}(x) - \delta(x)| > n^{-r}\right] \to 0.$$

Proof of Lemma 4: First, by a similar decomposition of $\hat{\delta}(x) - \delta(x)$ as that in the proof of Lemma 3, it suffices to show

$$\sup_x n^k \times \mathbf{P}\left\{|a_n(x,z) - a(x,z)| > \lambda_a \times n^{-r}\right\} \to 0,$$
$$\sup_x n^k \times \mathbf{P}\left\{|b_n(x,z) - b(x,z)| > \lambda_b \times n^{-r}\right\} \to 0,$$
$$\sup_x n^k \times \mathbf{P}\left\{|q_n(x,z) - q(x,z)| > \lambda_q \times n^{-r}\right\} \to 0,$$

where λ_a, λ_b, and λ_q are strictly positive constants, and

$$a_n(x,z) = \frac{1}{n}\sum_{j=1}^n Y_j K_{X_j,h}(x)\mathbb{1}(Z_j = z), \quad a(x,z) = \mathbf{E}(Y|X=x, Z=z) \times q(x,z),$$
$$b_n(x,z) = \frac{1}{n}\sum_{j=1}^n D_j K_{X_j,h}(x)\mathbb{1}(Z_j = z), \quad b(x,z) = \mathbf{E}(D|X=x, Z=z) \times q(x,z),$$

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

$$q_n(x,z) = \frac{1}{n}\sum_{j=1}^{n} K_{X_j,h}(x)\mathbb{1}(Z_j = z).$$

For expositional simplicity, we only show the first result. It is straightforward that the rest follow a similar argument.

Let $T_{nxzj} = Y_j K(\frac{X_j - x}{h})1(Z_j = z)$ and $\tau_{nxz} = h \times \left[\lambda_a n^{-r} - |\mathbf{E}[a_n(x,z)] - a(x,z)|\right]$. Note that

$$\begin{aligned}
&\mathbf{P}\left[|a_n(x,z) - a(x,z)| > \lambda_a \times n^{-r}\right] \\
&\leq \mathbf{P}\left[|a_n(x,z) - \mathbf{E}[a_n(x,z)]| + |\mathbf{E}[a_n(x,z)] - a(x,z)| > \lambda_a \times n^{-r}\right] \\
&\quad = \mathbf{P}\left\{\frac{1}{n}\left|\sum_{j=1}^{n}\left(T_{nxzj} - \mathbf{E}[T_{nxzj}]\right)\right| > \tau_{nxz}\right\}.
\end{aligned}$$

Moreover, by Bernstein's tail inequality,

$$\mathbf{P}\left\{\frac{1}{n}\left|\sum_{i=1}^{n}\left(T_{xzj} - \mathbf{E}[T_{xzj}]\right)\right| > \tau_{nxz}\right\} \leq 2\exp\left(-\frac{n \times \tau_{nxz}^2}{2\mathrm{Var}\left(T_{nxzj}\right) + \frac{2}{3}\overline{K} \times \tau_{nxz}}\right),$$

where $\overline{K}$ is the upper bound of kernel K.

By Assumption 10, $|\mathbf{E}[a_n(x,z)] - a(x,z)| = O(n^{-\iota}) = o(n^{-r})$. Then, for sufficiently large n, there is $0.5\lambda_a n^{-r} h \leq \tau_n(x,z) \leq \lambda_a n^{-r} h$. Moreover,

$$\mathrm{Var}\left(T_{nxzj}\right) \leq \mathbf{E}[T_{nxzj}^2] \leq \mathbf{E}\left[\mathbf{E}(Y^2|X)K^2(\frac{X-x}{h})\right] \leq Ch,$$

where $C = \sup_x \mathbf{E}[Y^2|X=x] \times \sup_x f_X(x) \times \overline{K} \times \int |K(u)|du < \infty$. It follows that

$$\mathbf{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}\left(T_{xzj} - \mathbf{E}[T_{xzj}]\right)\right| > \tau_{nxz}\right\} \leq 2\exp\left(-\frac{\frac{\lambda_a}{4} nhn^{-2r}}{2C + \frac{2}{3}\overline{K}\lambda_a n^{-r}}\right).$$

For sufficiently large n, we have $\frac{2}{3}\overline{K}\lambda_a n^{-r} \leq 1$. Therefore, for sufficiently large n,

$$\mathbf{P}\left\{\frac{1}{n}\left|\sum_{\ell=1}^{n}\left(T_{xzj} - \mathbf{E}[T_{xzj}]\right)\right| > \tau_{nxz}\right\} \leq 2\exp\left(-\frac{n^{2\iota - 2r}}{2C+1}\right) = o(n^{-k}),$$

where the inequality comes from Assumption 10. Note that the upper bound does not depend on x or z. Therefore,

$$\sup_{x,z}\mathbf{P}\left[|a_n(x,z) - a(x,z)| > \lambda_a \times n^{-r}\right] = o(n^{-k}).$$

C. Testing with Both Discrete and Continuous Covariates

We briefly discuss how to implement our test when the covariates contain both discrete and continuous variables. Let $X = (X_d', X_c')$, where X_d is a d_{X_d}-dimensional vector of discrete variables taking a finite number of values in $\mathcal{X}_d$, and X_c is a d_{X_c}-dimensional vector of continuous covariates with support $\mathcal{X}_c = \prod_{j=1}^{d_{X_c}} [x_{c,\ell j}, x_{c,uj}]$. Define $\mathcal{X}_c^{\xi} = \prod_{j=1}^{d_{X_c}} [x_{c,\ell j} + \xi, x_{c,uj} - \xi]$.

Note that in this case, $W \perp\!\!\!\perp Z|X$ is equivalent to

$$\Pi(w|x_c, x_d, 0) = \Pi(w|x_c, x_d, 1),\ \forall (w, x_c, x_d) \in \mathcal{S}_{WX_cX_d}. \tag{18}$$

With a similar argument, we can further show that (18) is equivalent to

$$G(w, x_c, x_d, 0) = G(w, x_c, x_d, 1),\ \forall (w, x_c, x_d) \in \mathcal{S}_{WX_cX_d}, \tag{19}$$

where, for $z = 0, 1$,

$$G(w, x_c, x_d, z) = \mathbf{E}[\lambda(W - w)\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d, z) f_{X_c|X_dZ}(x_c|x_d, z')\Pr(X_d = x_d, Z = z')].$$

To see this, by the same arguments as in the previous subsection, we have

$$\begin{aligned} G(w, x_c, x_d, z) =& \mathbf{E}\left[\lambda(W - w)\mathbb{1}^*_{X_c}(x_c) f_{X_c|X_dZ}(X_c|x_d, z')|X_d = x_d, Z = z\right] \\ &\times \mathbf{P}(X_d = x_d, Z = 0)\mathbf{P}(X_d = x_d, Z = 1). \end{aligned}$$

We estimate $\delta(X_{di}, X_{ci})$ by $\hat{\delta}(X_{di}, X_{ci}) = \hat{\delta}_1(X_{di}, X_{ci})/\hat{\delta}_2(X_{di}, X_{ci})$, where

$$\begin{aligned} \hat{\delta}_1(X_{di}, X_{ci}) =& \sum_{j\neq i, X_{dj}=X_{di}} Y_j Z_j K_{X_{cj},h}(X_{ci}) \times \sum_{j\neq i, X_{dj}=X_{di}} K_{X_{cj},h}(X_{ci}) \\ &- \sum_{j\neq i, X_{dj}=X_{di}} Y_j K_{X_{cj},h}(X_{ci}) \times \sum_{j\neq i, X_{dj}=X_{di}} Z_j K_{X_{cj},h}(X_{ci}), \\ \hat{\delta}_2(X_{di}, X_{ci}) =& \sum_{j\neq i, X_{dj}=X_{di}} D_j Z_j K_{X_{cj},h}(X_{ci}) \times \sum_{j\neq i, X_{dj}=X_{di}} K_{X_{cj},h}(X_{ci}) \\ &- \sum_{j\neq i, X_{dj}=X_{di}} D_j K_{X_{cj},h}(X_{ci}) \times \sum_{j\neq i, X_{dj}=X_{di}} Z_j K_{X_{cj},h}(X_{ci}). \end{aligned}$$

Moreover, let $f_{X_dX_cZ}(x_c, x_d, z') = f_{X_c|X_dZ}(x|x_d, z)\mathbf{P}(X_d = x_d, Z = z)$, which can be estimated by

$$\hat{f}_{X_cX_dZ}(X_{ci}, X_{di}, z) = \frac{1}{n} \sum_{j\neq i, X_{dj}=X_{di}} K_{X_{cj},h}(X_{ci})\mathbb{1}_{Z_j}(z).$$

In turn, we let $\widehat{W}_i = Y_i + (1 - D_i) \times \hat{\delta}(X_{ci}, X_{di})$ and can estimate $G(w, x_c, x_d, z)$ as

$$\widehat{G}(w, x_c, x_d, z) = \frac{1}{n} \sum_{\{i: X_{ci}\in\mathcal{X}_c^{\xi}\}} \lambda(\widehat{W}_i - w)\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d, z)\hat{f}_{X_cX_dZ}(X_{ci}, X_{di}, z'),$$

and define our test statistic as follows:

$$\widehat{\mathcal{T}}_n^m = \sup_{w\in\mathcal{S}_W, x_c\in\mathcal{X}_c^{\xi}, x_d\in\mathcal{X}_d} \sqrt{n}\left|\widehat{G}(w, x_c, x_d, 0) - \widehat{G}(w, x_c, x_d, 1)\right|.$$

Here, we provide the influence functions for $\sqrt{n}(\widehat{G}(w, x_c, x_d, 0) - \widehat{G}(w, x_c, x_d, 1) - G(w, x_c, x_d, 0) - G(w, x_c, x_d, 1))$.Let $F^*_{WD|X_cX_dZ}(w, d|x_c, x_d, z) \equiv F_{W|DX_CX_dZ}(w|d, x_c, x_d, z) \times$

$\mathbf{P}(D=d|X_c=x_c,X_d=x_d,Z=z)$ and

$$\kappa^m(w,x_c,x_d) = -\frac{F^*_{WD|X_cX_dZ}(w,0|x_c,x_d,1) - F^*_{WD|X_cX_dZ}(w,0|x_c,x_d,0)}{p(x_c,x_d,1)-p(x_c,x_d,0)}.$$

Moreover, we define

$$\begin{aligned}
\psi^m_{wx} &= \mathbb{1}(X \in \mathcal{S}^{\xi}_{X_c})\Bigg\{\Big[\lambda(W-w) - \mathbf{E}[\lambda(W-w)|X_c,X_d=x_d,Z=1]\Big] \times \frac{\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d,0)}{f_{X_cX_dZ}(X_c,x_d,0)} \\
&\quad - \Big[\lambda(W-w) - \mathbf{E}[\lambda(W-w)|X_c,X_d=x_d,Z=0]\Big] \times \frac{\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d,1)}{f_{X_cX_dZ}(X_c,x_d,1)}\Bigg\} \\
&\quad \times f_{X_cX_dZ}(X_c,x_d,0) f_{X_cX_dZ}(X_c,x_d,1); \\
\phi^m_{wx} &= \mathbb{1}(X \in \mathcal{S}^{\xi}_{X_c})\kappa^c(w,X_c,x_d)\Bigg\{\Big[W - \mathbf{E}[W|X_c,X_d=x_d,Z=0]\Big] \times \frac{\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d,1)}{f_{X_cX_dZ}(X_c,x_d,1)} \\
&\quad - \Big[W - \mathbf{E}[W|X_c,X_d=x_d,Z=1]\Big] \times \frac{\mathbb{1}^*_{X_c}(x_c)\mathbb{1}_{X_dZ}(x_d,0)}{f_{X_cX_dZ}(X_c,x_d,0)}\Bigg\} \\
&\quad \times f_{X_cX_dZ}(X_c,x_d,0) f_{X_cX_dZ}(X_c,x_d,1).
\end{aligned}$$

Given such influence function representations, we can implement our test as before, so we omit the details.

REFERENCES

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

Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113(2), 231–263.

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

Abrevaya, J., Y.-C. Hsu, & R.P. Lieli (2015) Estimating conditional average treatment effects. *Journal of Business & Economic Statistics* 33(4), 485–505.

Andrews, D.W. (1994) Empirical process methods in econometrics. In F. Engle, & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2247–2294. Elsevier.

Andrews, D.W. (1997) A conditional Kolmogorov test. *Econometrica* 65(5), 1097–1128.

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

Angrist, J.D. & W.N. Evans (1998) Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review* 88(3), 450–477.

Angrist, J.D. & A.B. Krueger (1991) Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106(4), 979–1014.

Athey, S. & G. Imbens (2016) Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America* 113(27), 7353–7360.

Barrett, G.F. & S.G. Donald (2003) Consistent tests for stochastic dominance. *Econometrica* 71(1), 71–104.

Bloom, H.S., L.L. Orr, S.H. Bell, G. Cave, F. Doolittle, W. Lin, & J.M. Bos (1997) The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study. *The Journal of Human Resources* 32(3), 549–576.

Blundell, R. & J.L. Powell (2003) Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L. Hansen, & S. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Econometric Society Monographs, vol. 36, pp. 312–357. Cambridge University Press.

Bouezmarni, T. & A. Taamouti (2014) Nonparametric tests for conditional independence using conditional distributions. *Journal of Nonparametric Statistics* 26(4), 697–719.

Bronars, S.G. & J. Grogger (1994) The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review* 32(5), 1141–1156.

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

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

Chesher, A. (2003) Identification in nonseparable models. *Econometrica* 71(5), 1405–1441.

Chesher, A. (2005) Nonparametric identification under discrete variation. *Econometrica* 73(5), 1525–1550.

Crump, R.K., V.J. Hotz, G.W. Imbens, & O.A. Mitnik (2008) Nonparametric tests for treatment effect heterogeneity. *The Review of Economics and Statistics* 90(3), 389–405.

D'Haultfœuille, X. & P. Février (2015) Identification of nonseparable triangular models with discrete instruments. *Econometrica* 83(3), 1199–1210.

Dauxois, J. & G.M. Nkiet (1998) Nonlinear canonical analysis and independence tests. *Annals of Statistics* 26(4), 1254–1278.

Delgado, M.A. & W.G. Manteiga (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29(5), 1469–1507.

Donald, S.G. & Y.-C. Hsu (2014) Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics* 178(3), 383–397.

Donald, S.G., Y.-C. Hsu, & G.F. Barrett (2012) Incorporating covariates in the measurement of welfare and inequality: Methods and applications. *The Econometrics Journal* 15(1), C1–C30.

Florens, J.-P., J.J. Heckman, C. Meghir, & E. Vytlacil (2008) Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica* 76(5), 1191–1206.

Frölich, M. & B. Melly (2013) Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics* 31(3), 346–357.

Heckman, J.J., D. Schmierer, & S. Urzua (2010) Testing the correlated random coefficient model. *Journal of Econometrics* 158(2), 177–203.

Heckman, J.J., J. Smith, & N. Clements (1997) Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64(4), 487–535.

Heckman, J.J. & E. Vytlacil (2001) Policy-relevant treatment effects. *American Economic Review* 91(2), 107–111.

Heckman, J.J. & E. Vytlacil (2005) Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73(3), 669–738.

Hoderlein, S. & E. Mammen (2009) Identification and estimation of marginal effects in nonseparable, nonmonotonic models. *The Econometrics Journal* 12(1), 1–25.

Hoderlein, S. & H. White (2012) Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics* 168(2), 300–314.

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

Hsu, Y.-C., C.-A. Liu, & X. Shi (2019) Testing generalized regression monotonocity. *Econometric Theory*, 1146–1200.

Huang, M., Y. Sun, & H. White (2016) A flexible nonparametric test for conditional independence. *Econometric Theory* 32(6), 1434–1482.

Huang, T.-M. (2010) Testing conditional independence using maximal nonlinear conditional correlation. *Annals of Statistics* 38(4), 2047–2091.

Imbens, G.W. (2010) Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature* 48(2), 399–423.

Imbens, G.W. & J.D. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.

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

Imbens, G.W. & D.B. Rubin (1997) Estimating distributions for outcome compliers models in instrumental variables. *The Review of Economic Studies* 64(4), 555–574.

Jacobsen, J.P., J.W. Pearce, & J.L. Rosenbloom (1999) The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. *The Journal of Human Resources* 34(3), 449–474.

Jun, S.J., Y. Lee, & Y. Shin (2016) Treatment effects with unobserved heterogeneity: A set identification approach. *Journal of Business & Economic Statistics* 34(2), 302–311.

Kosorok, M. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lee, S., R. Okui, & Y.-J. Whang (2017) Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics* 32(7), 1207–1225.

Lee, S. & Y.-J. Whang (2009) Nonparametric Tests of Conditional Treatment Effects. Manuscript.

Linton, O. & P. Gozalo (2014) Testing conditional independence restrictions. *Econometric Reviews* 33(5–6), 523–552.

Lu, X. & H. White (2014) Testing for separability in structural equations. *Journal of Econometrics* 182(1), 14–26.

Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17(6), 571–599.

Matzkin, R.L. (2003) Nonparametric estimation of nonadditive random functions. *Econometrica* 71(5), 1339–1375.

Nolan, D. & D. Pollard (1988) Functional limit theorems for U-processes. *Annals of Probability* 16(3), 1291–1298.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.

Rosenzweig, M.R. & K.I. Wolpin (1980) Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica* 48(1), 227–240.

Rudin, W. (1962) *Fourier Analysis on Groups*. Interscience.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14(3), 295–325.

Su, L., Y. Tu, & A. Ullah (2015) Testing additive separability of error term in nonparametric structural models. *Econometric Reviews* 34(6–10), 1057–1088.

Su, L. & H. White (2007) A consistent characteristic function-based test for conditional independence. *Journal of Econometrics* 141(2), 807–834.

Su, L. & H. White (2008) A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24(4), 829–864.

Su, L. & H. White (2014) Testing conditional independence via empirical likelihood. *Journal of Econometrics* 182(1), 27–44.

Torgovitsky, A. (2015) Identification of nonseparable models using instruments with small support. *Econometrica* 83(3), 1185–1197.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*, pp. 16–28. Springer.

https://doi.org/10.1017/S0266466622000147 Published online by Cambridge University Press

van der Vaart, A.W. & J.A. Wellner (2007) Empirical processes indexed by estimated functions. In E.A. Cator, G. Jongbloed, C. Kraaikamp, H.P. Lopuhaä, & J.A. Wellner (eds.), *Asymptotics: Particles, Processes and Inverse Problems*. Lecture Notes–Monograph Series, vol. 55, pp. 234–252. Institute of Mathematical Statistics.

Vuong, Q. & H. Xu (2017) Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics* 8(2), 589–610.

Vytlacil, E. (2002) Independence, monotonicity, and latent index models: An equivalence result. *Econometrica* 70(1), 331–341.