

SUPPLEMENT TO “A TEST FOR INSTRUMENT VALIDITY”:
PROOFS AND MONTE CARLO EXPERIMENTS
(*Econometrica*, Vol. 83, No. 5, September 2015, 2043–2063)

BY TORU KITAGAWA¹

APPENDIX A: PROOF OF PROPOSITION 1.1

IN ADDITION TO THE NOTATIONS introduced in the main text, we introduce the individual type indicator T :

$$T = c, \text{ complier, if } D_1 = 1, D_0 = 0,$$

$$T = n, \text{ never-taker, if } D_1 = 0, D_0 = 0,$$

$$T = a, \text{ always-taker, if } D_1 = 1, D_0 = 1,$$

$$T = df, \text{ defier, if } D_1 = 0, D_0 = 1.$$

When instrument exclusion is imposed, we suppress the z subscript in the potential outcome notation, and define $Y_1 \equiv Y_{11} = Y_{10}$ and $Y_0 \equiv Y_{01} = Y_{00}$ as a pair of potential outcomes indexed solely by $D = 1$ and 0 . Note that the joint restriction of instrument exclusion and random assignment is equivalent to $(Y_1, Y_0, T) \perp Z$.

PROOF OF PROPOSITION 1.1: (i) Let P and Q satisfying the inequalities (1.1) be given, and assume instrument exclusion. Our goal is to show that there exists a joint distribution of (Y_1, Y_0, T, Z) that is consistent with the given P and Q , and satisfies $(Y_1, Y_0, T) \perp Z$ and instrument monotonicity. Since the marginal distribution of Z is not important in the following argument, we focus on constructing the conditional distribution of (Y_1, Y_0, T) given Z . Let $p(\cdot, d) = \frac{dP(\cdot, d)}{d\mu}$ and $q(y, d) = \frac{dQ(\cdot, d)}{d\mu}$. Define nonnegative functions

$$h_{Y_1, c}(y) \equiv p(y, 1) - q(y, 1),$$

$$h_{Y_0, c}(y) \equiv q(y, 0) - p(y, 0),$$

$$h_{Y_1, a}(y) = q(y, 1),$$

$$h_{Y_0, n}(y) = p(y, 0),$$

$$h_{Y_1, df}(y) = 0,$$

$$h_{Y_0, df}(y) = 0;$$

¹Financial support from the ESRC through the ESRC Center for Microdata Methods and Practice (CEMMAP) (Grant RES-589-28-0001) and the Merit Dissertation Fellowship from the Graduate School of Economics in Brown University are gratefully acknowledged.

$h_{Y_0,a}(y)$ and $h_{Y_1,n}(y)$ are arbitrary nonnegative functions supported on \mathcal{Y} and satisfy $\int_{\mathcal{Y}} h_{Y_0,a}(y) d\mu = \Pr(D = 1|Z = 1)$ and $\int_{\mathcal{Y}} h_{Y_1,n}(y) d\mu = \Pr(D = 1|Z = 0)$. These nonnegative functions, $h_{Y_d,t}(y)$, $d \in \{1, 0\}$, $t \in \{c, n, a, df\}$, are introduced for the purpose of imputing a probability density of $\frac{\partial}{\partial \mu} \Pr(Y_d \in \cdot, T = t)$ that match the data distribution P and Q . Consider the following probability law of (Y_1, Y_0, T) given Z defined on the product σ -algebra of $\mathcal{Y} \times \mathcal{Y} \times \{c, n, a, df\}$,

$$\begin{aligned}
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = c|Z = 1) \\
&= \Pr(Y_1 \in B_1, Y_0 \in B_0, T = c|Z = 0) \\
&\equiv \begin{cases} \frac{\int_{B_1} h_{Y_1,c}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_1,c}(y) d\mu} \times \frac{\int_{B_0} h_{Y_0,c}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_0,c}(y) d\mu} \times [P(\mathcal{Y}, 1) - Q(\mathcal{Y}, 1)] \\ 0 \text{ if } [P(\mathcal{Y}, 1) - Q(\mathcal{Y}, 1)] = 0, \end{cases} \\
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = n|Z = 1) \\
&= \Pr(Y_1 \in B_1, Y_0 \in B_0, T = n|Z = 0) \\
&\equiv \begin{cases} \frac{\int_{B_1} h_{Y_1,n}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_1,n}(y) d\mu} \times \frac{\int_{B_0} h_{Y_0,n}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_0,n}(y) d\mu} \times P(\mathcal{Y}, 0) & \text{if } P(\mathcal{Y}, 0) > 0, \\ 0 & \text{if } P(\mathcal{Y}, 0) = 0, \end{cases} \\
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = a|Z = 1) \\
&= \Pr(Y_1 \in B_1, Y_0 \in B_0, T = a|Z = 0) \\
&\equiv \begin{cases} \frac{\int_{B_1} h_{Y_1,a}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_1,a}(y) d\mu} \times \frac{\int_{B_0} h_{Y_0,a}(y) d\mu}{\int_{\mathcal{Y}} h_{Y_0,a}(y) d\mu} \times Q(\mathcal{Y}, 1) & \text{if } Q(\mathcal{Y}, 1) > 0, \\ 0 & \text{if } Q(\mathcal{Y}, 1) = 0, \end{cases} \\
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = df|Z = 1) \\
&= \Pr(Y_1 \in B_1, Y_0 \in B_0, T = df|Z = 0) \\
&\equiv 0,
\end{aligned}$$

where $P(\mathcal{Y}, d) = \Pr(D = d|Z = 1)$ and $Q(\mathcal{Y}, d) = \Pr(D = d|Z = 0)$. Note that this is a probability measure on the product σ -algebra of $\mathcal{Y} \times \mathcal{Y} \times \{c, a, n, df\}$,

since it is nonnegative, additive, and sums up to 1:

$$\sum_{t \in \{c, n, a, df\}} \Pr(Y_1 \in \mathcal{Y}, Y_0 \in \mathcal{Y}, T = t | Z = z) = 1, \quad z = 1, 0.$$

The proposed probability distribution of $(Y_1, Y_0, T|Z)$ clearly satisfies joint independence and instrument monotonicity by construction, and it induces the given data generating process, that is, the proposed probability distribution of $(Y_1, Y_0, T|Z)$ satisfies

$$\begin{aligned} \text{(A.1)} \quad P(B, 1) &= \Pr(Y_1 \in B, Y_0 \in \mathcal{Y}, T = a | Z = 1) \\ &\quad + \Pr(Y_1 \in B, Y_0 \in \mathcal{Y}, T = c | Z = 1), \\ Q(B, 1) &= \Pr(Y_1 \in B, Y_0 \in \mathcal{Y}, T = a | Z = 0) \\ &\quad + \Pr(Y_1 \in B, Y_0 \in \mathcal{Y}, T = df | Z = 0), \\ P(B, 0) &= \Pr(Y_1 \in \mathcal{Y}, Y_0 \in B, T = n | Z = 1) \\ &\quad + \Pr(Y_1 \in \mathcal{Y}, Y_0 \in B, T = df | Z = 1), \\ Q(B, 0) &= \Pr(Y_1 \in \mathcal{Y}, Y_0 \in B, T = n | Z = 0) \\ &\quad + \Pr(Y_1 \in \mathcal{Y}, Y_0 \in B, T = c | Z = 0). \end{aligned}$$

This completes the proof of the first claim.

(ii) Let arbitrary P and Q satisfying inequalities (1.1) be given. We maintain instrument exclusion, so, in what follows, we construct a probability law of (Y_1, Y_0, T) given Z that is consistent with P and Q , but violates $(Y_1, Y_0, T) \perp Z$. Consider the probability distribution of (Y_1, Y_0, T) given Z :

$$\begin{aligned} \Pr(Y_1 \in B_1, Y_0 \in B_0, T = c | Z = 1) &= 0, \\ \Pr(Y_1 \in B_1, Y_0 \in B_0, T = c | Z = 0) &= \begin{cases} \frac{Q(B_1, 0)Q(B_0, 0)}{Q(\mathcal{Y}, 0)} & \text{if } Q(\mathcal{Y}, 0) > 0, \\ 0 & \text{if } Q(\mathcal{Y}, 0) = 0, \end{cases} \\ \Pr(Y_1 \in B_1, Y_0 \in B_0, T = n | Z = 1) &= \begin{cases} \frac{P(B_1, 0)P(B_0, 0)}{P(\mathcal{Y}, 0)} & \text{if } P(\mathcal{Y}, 0) > 0, \\ 0 & \text{if } P(\mathcal{Y}, 0) = 0, \end{cases} \\ \Pr(Y_1 \in B_1, Y_0 \in B_0, T = n | Z = 0) &= 0, \\ \Pr(Y_1 \in B_1, Y_0 \in B_0, T = a | Z = 1) &= \begin{cases} \frac{P(B_1, 1)P(B_0, 1)}{P(\mathcal{Y}, 1)} & \text{if } P(\mathcal{Y}, 1) > 0, \\ 0 & \text{if } P(\mathcal{Y}, 1) = 0, \end{cases} \end{aligned}$$

$$\begin{aligned}
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = a | Z = 0) = 0, \\
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = df | Z = 1) = 0, \\
& \Pr(Y_1 \in B_1, Y_0 \in B_0, T = df | Z = 0) \\
& = \begin{cases} \frac{Q(B_1, 1)Q(B_0, 1)}{Q(\mathcal{Y}, 1)} & \text{if } Q(\mathcal{Y}, 1) > 0, \\ 0 & \text{if } Q(\mathcal{Y}, 1) = 0. \end{cases}
\end{aligned}$$

Note that, in this construction, Z and T are dependent, that is, $Z = 1$ is assigned only to never-takers and always-takers, and $Z = 0$ is assigned only to compliers and defiers, so it violates $T \perp Z$ (and the no-defier condition as well if $Q(\mathcal{Y}, 1) > 0$). Furthermore, the proposed distribution of $(Y_1, Y_0, T|Z)$ satisfies (A.1), so it is consistent with P and Q . Since the proposed construction is feasible for any P and Q , we conclude that for any P and Q that meet the testable implications, there exists a distribution of (Y_1, Y_0, T, Z) that violates IV validity. *Q.E.D.*

APPENDIX B: PROOF OF THEOREM 2.1

B.1. Notations

In addition to the notation introduced in the main text, we introduce the following notation that is used throughout this appendix. Let \mathcal{F} be a set of indicator functions defined on $\mathcal{X} \equiv \mathcal{Y} \times \{0, 1\}$,

$$\begin{aligned}
\mathcal{F} = & \{1_{\{[y, y'], 1\}}(Y, D) : -\infty \leq y \leq y' \leq \infty\} \\
& \cup \{1_{\{[y, y'], 0\}}(Y, D) : -\infty \leq y \leq y' \leq \infty\},
\end{aligned}$$

where $1_{\{B, d\}}(Y, D)$ is the indicator function for event $\{Y \in B, D = d\}$. The Borel σ -algebra of \mathcal{X} is denoted by $\mathcal{B}(\mathcal{X})$. Note that \mathcal{F} is a Vapnik–Chervonenkis (VC) class of functions since a class of connected intervals is a VC class of subsets. We denote a generic element of \mathcal{F} by f . For generic $P \in \mathcal{P}$, let P_m be an empirical probability measure constructed by a size m independent and identically distributed (i.i.d.) sample from P . We define shorthand notations $P(f) \equiv P([y, y'], d)$ and $P_m(f) \equiv P_m([y, y'], d)$. Denote empirical processes indexed by \mathcal{F} by

$$G_{m, P}(\cdot) = \sqrt{m}(P_m - P)(\cdot).$$

For a probability measure P on \mathcal{X} , we denote the mean zero P -Brownian bridge processes indexed by \mathcal{F} by $G_P(\cdot)$. Let $\rho_\omega(f, f') = [\omega((f - f')^2)]^{1/2}$ be a seminorm on \mathcal{F} defined in terms of the L_2 metric with respect to a finite measure ω on \mathcal{X} . Given a deterministic sequence of the sizes of two samples, $\{(m(N), n(N)) : N = 1, 2, \dots\}$, let $\{(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{P}^2 : N = 1, 2, \dots\}$ be a

sequence of the two-sample probability measures that drift with the sample sizes $(m(N), n(N))$, where superscripts with brackets index a sequence. We often omit the arguments of $(m(N), n(N))$ unless any confusion arises.

Let $\sigma_P^2(\cdot, \cdot) : \mathcal{F}^2 \rightarrow \mathbb{R}$ denote the covariance kernel of P -Brownian bridges, $\sigma_P^2(f, g) = P(fg) - P(f)P(g)$. We denote by $\sigma_{P,Q}^2(f, g) : \mathcal{F}^2 \rightarrow \mathbb{R}$ the covariance kernel of the independent two-sample Brownian bridge processes $(1 - \lambda)^{1/2}G_P(\cdot) - \lambda^{1/2}G_Q(\cdot)$,

$$\sigma_{P,Q}^2(f, g) = (1 - \lambda)\sigma_P^2(f, g) + \lambda\sigma_Q^2(f, g),$$

and denote $\sigma_{P_m, Q_n}^2(\cdot, \cdot)$ as its sample analogue,

$$\begin{aligned} \sigma_{P_m, Q_n}^2(f, g) &= (1 - \hat{\lambda})[P_m(fg) - P_m(f)P_m(g)] \\ &\quad + \hat{\lambda}[Q_n(fg) - Q_n(f)Q_n(g)]. \end{aligned}$$

Note that with the current notation, $\sigma_{P_m, Q_m}^2([y, y'], d)$ defined in the main text is equivalent to $\sigma_{P_m, Q_n}^2(f, f)$ for $f = 1_{\{[y, y'], d\}}$. For a sequence of random variables $\{W_N : N = 1, 2, \dots\}$ whose probability law is governed by a sequence of two-sample probability measures $(P^{[m(N)]}, Q^{[n(N)]})$, $W_N \xrightarrow{p^{[m]}, q^{[n]}} c$ denotes convergence in probability in the sense of, for every $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} \Pr_{p^{[m]}, q^{[n]}} (|W_N - c| > \varepsilon) = 0.$$

In particular, if $W_N \xrightarrow{p^{[m]}, q^{[n]}} 0$, we notate it as $W_N = o_{p^{[m]}, q^{[n]}}(1)$.

B.2. Auxiliary Lemmas

We first present a set of lemmas to be used in the proofs of Theorems 2.1 and 2.2.

LEMMA B.1: *Let $\{P^{[m]} \in \mathcal{P} : m = 1, 2, \dots\}$ be a sequence of probability measures on \mathcal{X} . Then*

$$\sup_{f \in \mathcal{F}} |(P_m^{[m]} - P^{[m]})(f)| \xrightarrow{p^{[m]}} 0.$$

PROOF: Variable \mathcal{F} is the class of indicator functions corresponding to the interval VC class of subsets, so an application of the Glivenko–Cantelli theorem uniform in \mathcal{P} (Theorem 2.8.1 of van der Vaart and Wellner (1996)) yields the claim. Q.E.D.

LEMMA B.2: *Suppose Condition RG. Let $\{P^{[m]} \in \mathcal{P} : m = 1, 2, \dots\}$ be a sequence of data generating processes on \mathcal{X} that weakly converges to $P_0 \in \mathcal{P}$ as $m \rightarrow \infty$. Then*

$$\sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)| \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

PROOF: We first consider the case of μ being the Lebesgue measure. Suppose the conclusion is false, that is, there exists $\xi > 0$ and a sequence $\{B_m \in \mathcal{B}(\mathcal{X}) : m = 1, 2, \dots\}$ such that $\limsup_{m \rightarrow \infty} |(P^{[m]} - P_0)(B_m)| > \xi$. By uniform tightness of Condition RG(b), there exists a compact set $K \in \mathcal{B}(\mathcal{X})$ such that

$$\limsup_{m \rightarrow \infty} |(P^{[m]} - P_0)(B_m \cap K)| > \xi/2$$

holds. Let $\{b_m\}$ be a subsequence of $\{m\}$ such that $|(P^{[b_m]} - P_0)(B_{b_m} \cap K)| > \xi/2$ holds for all $b_m \geq b_m^*$. We metricize $\mathcal{B}(\mathcal{X})$ by the L_1 metric $d_{\mathcal{B}(\mathcal{X})}(B, B') = (\mu \times \delta_d)(B \Delta B')$, where μ is the measure defined in Condition RG(a) and δ_d is the mass measure on $d \in \{0, 1\}$. Since $\{B_{b_m} \cap K : m = 1, 2, \dots\}$ is a sequence in a compact subset of $\mathcal{B}(\mathcal{X})$, there exists a subsequence c_{b_m} of b_m , such that $\{B_{c_{b_m}} \cap K\}$ converges to $B^* \in \mathcal{B}(\mathcal{X})$ in terms of metric $d_{\mathcal{B}(\mathcal{X})}(\cdot, \cdot)$, and

$$(B.1) \quad |(P^{[c_{b_m}]} - P_0)(B_{c_{b_m}} \cap K)| > \xi/2$$

holds by the construction of $\{b_m\}$ for all $c_{b_m} \geq c_{b_m}^*$. Under the bounded density assumption of Condition RG(a), it holds that

$$\begin{aligned} & |(P^{[c_{b_m}]} - P_0)(B_{c_{b_m}} \cap K) - (P^{[c_{b_m}]} - P_0)(B^*)| \\ & \leq 2Md_{\mathcal{B}(\mathcal{X})}(B_{c_{b_m}} \cap K, B^*) \rightarrow 0, \quad \text{as } m \rightarrow \infty. \end{aligned}$$

Hence, (B.1) implies

$$(B.2) \quad \limsup_{m \rightarrow \infty} |(P^{[c_{b_m}]} - P_0)(B^*)| > \xi/2.$$

Since μ is the Lebesgue measure and, by Condition RG(a), P_0 as a weak limit of $\{P^{[m]} : m = 1, 2, \dots\}$ is absolutely continuous in $\mu \times \delta_d$, we have $P_0(\delta B^*) = 0$, where δB^* is the boundary of B^* . Accordingly, by applying the Portmanteau theorem (see, e.g., Theorem 1.3.4 of [van der Vaart and Wellner \(1996\)](#)), we obtain $\lim_{m \rightarrow \infty} |(P^{[m]} - P_0)(B^*)| = 0$. This contradicts (B.2). Hence, $\lim_{m \rightarrow \infty} \sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)| = 0$ holds.

When μ is a discrete mass measure with finite support points, then the weak convergence of $P^{[m]}$ to P_0 is equivalent to the pointwise convergence of the probability mass functions, and the $\sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)|$ is equivalent to the supremum over power sets of the finite support points. Hence, the claim follows.

For the case of μ being a mixture of the Lebesgue and a discrete mass measure with finite support points, the claim holds as an immediate corollary of each of the two cases already shown. *Q.E.D.*

The next lemma is a corollary of Lemma B.1 and B.2.

LEMMA B.3: Suppose Condition RG. Let $\{P^{[m]} \in \mathcal{P} : m = 1, 2, \dots\}$ be a sequence of data generating processes on \mathcal{X} that weakly converges to $P_0 \in \mathcal{P}$ as $m \rightarrow \infty$:

$$\sup_{f \in \mathcal{F}} |(P_m^{[m]} - P_0)(f)| \xrightarrow{P^{[m]}} 0.$$

LEMMA B.4: Suppose Condition RG. Let $\{(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{P}^2 : N = 1, 2, \dots\}$ be a sequence of two-sample probability measures with sample size $(m, n) = (m(N), n(N)) \rightarrow (\infty, \infty)$ as $N \rightarrow \infty$. We have

$$\sup_{f, g \in \mathcal{F}} |\sigma_{P_m^{[m]}, Q_n^{[n]}}^2(f, g) - \sigma_{P^{[m]}, Q^{[n]}}^2(f, g)| \xrightarrow{P^{[m]}, Q^{[n]}} 0.$$

PROOF: Consider

$$\begin{aligned} \text{(B.3)} \quad & |\sigma_{P_m^{[m]}, Q_n^{[n]}}^2(f, g) - \sigma_{P^{[m]}, Q^{[n]}}^2(f, g)| \\ & \leq (1 - \lambda) |P_m^{[m]}(fg) - P_m^{[m]}(f)P_m^{[m]}(g) - P^{[m]}(fg) + P^{[m]}(f)P^{[m]}(g)| \\ & \quad + \lambda |Q_n^{[n]}(fg) - Q_n^{[n]}(f)Q_n^{[n]}(g) - Q^{[n]}(fg) + Q^{[n]}(f)Q^{[n]}(g)| \\ & \quad + o(1), \end{aligned}$$

where $o(1)$ is the approximation error of order $|\hat{\lambda} - \lambda|$. Regarding the first term in the right-hand side of this inequality, the following inequalities hold:

$$\begin{aligned} \text{(B.4)} \quad & (1 - \lambda) |P_m^{[m]}(fg) - P_m^{[m]}(f)P_m^{[m]}(g) - P^{[m]}(fg) + P^{[m]}(f)P^{[m]}(g)| \\ & \leq |(P_m^{[m]} - P^{[m]})(fg)| + |P_m^{[m]}(f)P_m^{[m]}(g) - P^{[m]}(f)P^{[m]}(g)| \\ & \leq |(P_m^{[m]} - P^{[m]})(fg)| + |(P_m^{[m]} - P^{[m]})(f)P_m^{[m]}(g)| \\ & \quad + |(P_m^{[m]} - P^{[m]})(g)P^{[m]}(f)| \\ & \leq |(P_m^{[m]} - P^{[m]})(fg)| + |(P_m^{[m]} - P^{[m]})(f)| + |(P_m^{[m]} - P^{[m]})(g)|. \end{aligned}$$

The second and the third terms of (B.4) are $o_{P^{[m]}}(1)$ uniformly in \mathcal{F} by Lemma B.1. Furthermore, since the class of indicator functions $\{fg : f, g \in \mathcal{F}\}$ is also a VC-class,

$$\sup_{f, g \in \mathcal{F}} |(P_m^{[m]} - P^{[m]})(fg)| \xrightarrow{P^{[m]}} 0$$

holds also by Lemma B.1. This proves that the first term in the right-hand side of (B.3) converges to zero uniformly in $f, g \in \mathcal{F}$. The case for the second term of (B.3) is made by the same argument. Hence, the conclusion follows. Q.E.D.

LEMMA B.5: *Suppose Condition RG. Let $\{P^{[m]} \in \mathcal{P} : m = 1, 2, \dots\}$ be a sequence of probability measures that converges weakly to $P_0 \in \mathcal{P}$. Then the empirical processes $G_{m, P^{[m]}}(\cdot)$ on index set \mathcal{F} converge weakly to P_0 -Brownian bridges $G_{P_0}(\cdot)$.*

PROOF: To prove this lemma, we apply a combination of Theorem 2.8.3 and Lemma 2.8.8 of van der Vaart and Wellner (1996) restricted to a class of indicator functions. It claims that, given \mathcal{F} is a class of measurable indicator functions and a sequence of probability measure $\{P^{[m]} : m = 1, 2, \dots\}$ in \mathcal{P} , if (i) $\int_0^1 \sup_R \sqrt{\log N(\varepsilon, \mathcal{F}, L_2(R))} d\varepsilon < \infty$, where R ranges over all finitely discrete probability measures and $N(\varepsilon, \mathcal{F}, L_2(R))$ is the covering number of \mathcal{F} with radius ε in terms of $L_2(R)$ metric $[R(|f - f'|^2)]^{1/2}$,² and (ii) there exists $P^* \in \mathcal{P}$ such that $\lim_{m \rightarrow \infty} \sup_{f, g \in \mathcal{F}} \{|\rho_{P^{[m]}}(f, g) - \rho_{P^*}(f, g)|\} = 0$, then $G_{m, P^{[m]}}(\cdot)$ weakly converges to the P^* -Brownian bridge process $G_{P^*}(\cdot)$. Condition (i) is known to hold if \mathcal{F} is a VC class (see Theorem 2.6.4 of van der Vaart and Wellner (1996)).

Therefore, what remains to show is condition (ii). By the construction of seminorm $\rho_P(f, g)$, we have

$$\sup_{f, g \in \mathcal{F}} |\rho_{P^{[m]}}^2(f, g) - \rho_{P_0}^2(f, g)| \leq \sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)|.$$

Hence, to validate condition (ii) with $P^* = P_0$, it suffices to have

$$\lim_{m \rightarrow \infty} \sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)| = 0,$$

which follows from Lemma B.2. Q.E.D.

LEMMA B.6: *Suppose Condition RG. Let $\{(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{P}^2 : N = 1, 2, \dots\}$ be a sequence of probability measures of the two independent samples that converges weakly to (P_0, Q_0) as $N \rightarrow \infty$. Then stochastic processes indexed by the VC class of indicator functions \mathcal{F} ,*

$$(B.5) \quad v_N(\cdot) = \frac{(1 - \hat{\lambda})^{1/2} G_{m, P^{[m]}}(\cdot) - \hat{\lambda}^{1/2} G_{n, Q^{[n]}}(\cdot)}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(\cdot, \cdot)}, \quad \xi > 0,$$

converge weakly to mean zero Gaussian processes

$$v_0(\cdot) = \frac{(1 - \lambda)^{1/2} G_{P_0}(\cdot) - \lambda^{1/2} G_{Q_0}(\cdot)}{\xi \vee \sigma_{P_0, Q_0}(\cdot, \cdot)},$$

where $G_{P_0}(\cdot)$ and $G_{Q_0}(\cdot)$ are independent Brownian bridge processes.

²The covering number $N(\varepsilon, \mathcal{F}, L_2(R))$ is defined as the minimal number of balls of radius ε needed to cover \mathcal{F} .

PROOF: The VC class \mathcal{F} is totally bounded with seminorm ρ_P for any finite measure P . Hence, following Section 2.8.3 of [van der Vaart and Wellner \(1996\)](#), what we want to show for the weak convergence of $v_N(\cdot)$ is that (i) the finite dimensional marginal, $(v_N(f_1), \dots, v_N(f_K))$, converges to that of $v_0(\cdot)$, (ii) $v_N(\cdot)$ is asymptotically uniformly equicontinuous along a sequence of seminorms such as the $L_2(P^{[m]} + Q^{[n]})$ norm, $\rho_{P^{[m]}+Q^{[n]}}(f, g) = [(P^{[m]} + Q^{[n]})(f - g)^2]^{1/2}$, that is, for arbitrary $\varepsilon > 0$,

$$(B.6) \quad \lim_{\delta \searrow 0} \limsup_{N \rightarrow \infty} P_{P^{[m]}, Q^{[n]}}^* \left(\sup_{\rho_{P^{[m]}+Q^{[n]}}(f, g) < \delta} |v_N(f) - v_N(g)| > \varepsilon \right) = 0,$$

where $P_{P^{[m]}, Q^{[n]}}^*$ is the outer probability, and (iii)

$$\sup_{f, g \in \mathcal{F}} |\rho_{P^{[m]}+Q^{[n]}}(f, g) - \rho_{P_0+Q_0}(f, g)| \rightarrow 0$$

as $N \rightarrow \infty$. Note that (i) is implied by [Lemma B.4](#) and [Lemma B.5](#), and (iii) follows as a corollary of [Lemma B.2](#), since

$$\begin{aligned} & \sup_{f, g \in \mathcal{F}} |\rho_{P^{[m]}+Q^{[n]}}^2(f, g) - \rho_{P_0+Q_0}^2(f, g)| \\ & \leq \sup_{B \in \mathcal{B}(\mathcal{X})} |(P^{[m]} - P_0)(B)| + \sup_{B \in \mathcal{B}(\mathcal{X})} |(Q^{[n]} - Q_0)(B)| \\ & \rightarrow 0 \quad \text{as } N \rightarrow \infty. \end{aligned}$$

To verify (ii), consider, for $f, g \in \mathcal{F}$ with $\rho_{P^{[m]}+Q^{[n]}}(f, g) < \delta$,

$$(B.7) \quad \begin{aligned} & |v_N(f) - v_N(g)| \\ & \leq \left| \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f)} - \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g)} \right| \\ & \quad \times |(1 - \lambda)^{1/2} G_{m, P^{[m]}}(g) - \lambda^{1/2} G_{n, Q^{[n]}}(g)| \\ & \quad + ((1 - \lambda)^{1/2} |G_{m, P^{[m]}}(f) - G_{m, P^{[m]}}(g)| \\ & \quad + \lambda^{1/2} |G_{n, Q^{[n]}}(f) - G_{n, Q^{[n]}}(g)|) / (\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g)) \\ & \quad + o(|\hat{\lambda} - \lambda|). \end{aligned}$$

Note that

$$(B.8) \quad \begin{aligned} & \left| \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f)} - \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g)} \right| \\ & = \left| \frac{1}{\xi \vee \sigma_{P^{[m]}, Q^{[n]}}(f, f)} - \frac{1}{\xi \vee \sigma_{P^{[m]}, Q^{[n]}}(g, g)} \right| + o_{P^{[m]}, Q^{[n]}}(1) \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{\xi^2} \left| \xi \vee \sigma_{P^{[m]}, Q^{[n]}}(f, f) - \xi \vee \sigma_{P^{[m]}, Q^{[n]}}(g, g) \right| + o_{P^{[m]}, Q^{[n]}}(1) \\
&\leq \frac{1}{\xi^2} \left| \sigma_{P^{[m]}, Q^{[n]}}(f, f) - \sigma_{P^{[m]}, Q^{[n]}}(g, g) \right| + o_{P^{[m]}, Q^{[n]}}(1),
\end{aligned}$$

where the first line follows from Lemma B.4. By noting the inequalities

$$\begin{aligned}
&\left| \sigma_{P^{[m]}, Q^{[n]}}(f, f) - \sigma_{P^{[m]}, Q^{[n]}}(g, g) \right|^2 \\
&\leq \left| \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f) - \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g) \right| \\
&\quad \times \left| \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f) + \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g) \right| \\
&= \left| \sigma_{P^{[m]}, Q^{[n]}}^2(f, f) - \sigma_{P^{[m]}, Q^{[n]}}^2(g, g) \right|
\end{aligned}$$

and

$$\begin{aligned}
&\left| \sigma_{P^{[m]}, Q^{[n]}}^2(f, f) - \sigma_{P^{[m]}, Q^{[n]}}^2(g, g) \right| \\
&\leq \left| (1 - \lambda)(P^{[m]}(f) - P^{[m]}(g))(1 - P^{[m]}(f) - P^{[m]}(g)) \right| \\
&\quad + \left| \lambda(Q^{[n]}(f) - Q^{[n]}(g))(1 - Q^{[n]}(f) - Q^{[n]}(g)) \right| \\
&\leq \left| (1 - \lambda)(P^{[m]}(f) - P^{[m]}(g)) \right| + \left| \lambda(Q^{[n]}(f) - Q^{[n]}(g)) \right| \\
&\leq (1 - \lambda)\rho_{P^{[m]}}^2(f, g) + \lambda\rho_{Q^{[n]}}^2(f, g) \\
&\leq \rho_{P^{[m]}+Q^{[n]}}^2(f, g),
\end{aligned}$$

we have

$$(B.9) \quad \left| \sigma_{P^{[m]}, Q^{[n]}}(f, f) - \sigma_{P^{[m]}, Q^{[n]}}(g, g) \right| \leq \rho_{P^{[m]}+Q^{[n]}}(f, g).$$

Combining (B.8) and (B.9) then leads to

$$\begin{aligned}
(B.10) \quad &\left| \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f)} - \frac{1}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(g, g)} \right| \\
&\leq \frac{\rho_{P^{[m]}+Q^{[n]}}(f, g)}{\xi^2} + o_{P^{[m]}, Q^{[n]}}(1).
\end{aligned}$$

Hence, (B.7) and (B.10) yield

$$\begin{aligned}
(B.11) \quad &\sup_{\rho_{P^{[m]}+Q^{[n]}}(f, g) < \delta} \left| v_N(f) - v_N(g) \right| \\
&\leq \frac{\delta}{\xi^2} \left| (1 - \lambda)^{1/2} G_{m, P^{[m]}}(g) - \lambda^{1/2} G_{n, Q^{[n]}}(g) \right|
\end{aligned}$$

$$\begin{aligned}
& + \frac{(1-\lambda)^{1/2}}{\xi} \sup_{\rho_{P^{[m]}+Q^{[n]}}(f,g) < \delta} |G_{m,P^{[m]}}(f) - G_{m,P^{[m]}}(g)| \\
& + \frac{\lambda^{1/2}}{\xi} \sup_{\rho_{P^{[m]}+Q^{[n]}}(f,g) < \delta} |G_{n,Q^{[n]}}(f) - G_{n,Q^{[n]}}(g)| + o_{P^{[m]},Q^{[n]}}(1).
\end{aligned}$$

Since $\rho_{P^{[m]}}(f, g) \leq \rho_{P^{[m]}+Q^{[n]}}(f, g)$ for every $f, g \in \mathcal{F}$, we have

$$\begin{aligned}
& \sup_{\rho_{P^{[m]}+Q^{[n]}}(f,g) < \delta} |G_{m,P^{[m]}}(f) - G_{m,P^{[m]}}(g)| \\
& \leq \sup_{\rho_{P^{[m]}}(f,g) < \delta} |G_{m,P^{[m]}}(f) - G_{m,P^{[m]}}(g)| \\
& = o_{P^{[m]}}^*(\delta),
\end{aligned}$$

where $o_{P^{[m]}}^*(\delta)$ denotes the convergence to zero in outer probability along $\{P^{[m]}\}$ as $\delta \searrow 0$, and the equality follows since the uniform convergence of $G_{m,P^{[m]}}(f)$ as established by Lemma B.5 implies

$$\lim_{\delta \searrow 0} \limsup_{m \rightarrow \infty} P_{P^{[m]}}^* \left(\sup_{\rho_{P^{[m]}}(f,g) < \delta} |G_{m,P^{[m]}}(f) - G_{m,P^{[m]}}(g)| > \varepsilon \right) = 0.$$

Similarly, we obtain $\sup_{\rho_{P^{[m]}+Q^{[n]}}(f,g) < \delta} |G_{n,Q^{[n]}}(f) - G_{n,Q^{[n]}}(g)| = o_{Q^{[n]}}^*(\delta)$.

Since $|(1-\lambda)^{1/2}G_{m,P^{[m]}}(g) - \lambda^{1/2}G_{n,Q^{[n]}}(g)|$ converges weakly to the tight Gaussian processes, (B.11) is written as

$$\begin{aligned}
& \sup_{\rho_{P^{[m]}+Q^{[n]}}(f,g) < \delta} |v_N(f) - v_N(g)| \\
& = \delta O_{P^{[m]},Q^{[n]}}(1) + o_{P^{[m]},Q^{[n]}}^*(\delta) + o_{P^{[m]},Q^{[n]}}(1) \\
& = o_{P^{[m]},Q^{[n]}}^*(\delta),
\end{aligned}$$

where $O_{P^{[m]},Q^{[n]}}(1)$ denotes that $\lim_{N \rightarrow \infty} \Pr_{P^{[m]},Q^{[n]}}(|W_N| > a_N) = 0$ for every diverging sequence $a_N \rightarrow \infty$. This establishes the asymptotic uniform equicontinuity (B.6). *Q.E.D.*

The next lemma states that the null hypothesis of our test defined by inequalities (1.1) for every Borel set B can be reduced without loss of information to the hypothesis that inequalities (1.1) hold for all connected intervals. This lemma is a direct corollary of Lemma C1 in Andrews and Shi (2013).

LEMMA B.7: *We have that $P(B, 1) - Q(B, 1) \geq 0$ and $Q(B, 0) - P(B, 0) \geq 0$ hold for every Borel set B if and only if $P(V, 1) - Q(V, 1) \geq 0$ and $Q(V, 0) - P(V, 0) \geq 0$ hold for all $V \in \mathcal{V} \equiv \{[y, y'] : -\infty \leq y \leq y' \leq \infty\}$.*

PROOF: The “only if” statement is obvious. To prove the “if” statement, we apply Lemma C1 of [Andrews and Shi \(2013\)](#). By viewing \mathcal{V} as \mathcal{R} and $P(\cdot, 1) - Q(\cdot, 1)$ as $\mu(\cdot)$ in the notation of Lemma C1 of [Andrews and Shi \(2013\)](#), it follows that $P(B, 1) - Q(B, 1) \geq 0$ for all B in the Borel σ -algebra generated by \mathcal{V} . Since the Borel σ -algebra generated by \mathcal{V} coincides with $\mathcal{B}(\mathcal{Y})$, $P(V, 1) - Q(V, 1) \geq 0$ for every $V \in \mathcal{V}$ implies $P(B, 1) - Q(B, 1) \geq 0$ for every $B \in \mathcal{B}(\mathcal{Y})$. The same results hold for the other inequalities $Q(\cdot, 0) - P(\cdot, 0) \geq 0$. *Q.E.D.*

The next lemma shows that the version of testable implications with conditioning covariates as given in (3.3) can be reduced without any loss of information to the unconditional moment inequalities of (3.4).

LEMMA B.8: *Assume that $\Pr(Z = 1|X)$ is bounded away from 0 and 1, X -a.s. Then*

$$(B.12) \quad \begin{aligned} \Pr(Y \in B, D = 1|Z = 1, X) - \Pr(Y \in B, D = 1|Z = 0, X) &\geq 0, \\ \Pr(Y \in B, D = 0|Z = 0, X) - \Pr(Y \in B, D = 0|Z = 1, X) &\geq 0 \end{aligned}$$

hold for all $B \in \mathcal{B}(\mathcal{Y})$, X -a.s., if and only if

$$\begin{aligned} E[\kappa_1(D, Z, X)g(Y, X)] &\geq 0, \\ E[\kappa_0(D, Z, X)g(Y, X)] &\geq 0, \quad \text{for all } g(\cdot, \cdot) \in \mathcal{G}, \end{aligned}$$

where κ_1 , κ_0 , and \mathcal{G} are as defined in Section 3.2 of the main text.

PROOF: By applying Theorem 3.1 of [Abadie \(2003\)](#) with conditioning of X , the first inequalities of (B.12) can be equivalently written as

$$(B.13) \quad E[1\{Y \in B\}\kappa_1(D, Z, X)|X] \geq 0, \quad X\text{-a.s.}$$

Hence, the “only if” statement immediately follows.

To show the “if” statement, we again invoke Lemma C1 in [Andrews and Shi \(2013\)](#). Let us read \mathcal{R} and $\mu(\cdot)$ of their notation as

$$\begin{aligned} \mathcal{V} \equiv \{ & [y, y'] \times [x_1, x'_1] \times \cdots \times [x_{d_x}, x'_{d_x}] : \\ & -\infty \leq y \leq y' \leq \infty, -\infty \leq x_l \leq x'_l \leq \infty, l = 1, \dots, d_x \} \end{aligned}$$

and $\mu(\cdot) = E[\kappa_1(D, Z, X)1\{(Y, X) \in \cdot\}]$, respectively. By the assumption that $\Pr(Z = 1|X)$ is bounded away from 0 and 1, κ_1 is bounded X -a.s. Hence, the thus-defined $\mu(\cdot)$ satisfies the boundedness condition to apply Lemma C1 in [Andrews and Shi \(2013\)](#). Moreover, \mathcal{V} meets the condition for a semi-ring. Hence, $\mu(V) = E[\kappa_1(D, Z, X)1\{(Y, X) \in V\}] \geq 0$ for all $V \in \mathcal{V}$ implies $\mu(C) = E[\kappa_1(D, Z, X)1\{(Y, X) \in C\}] \geq 0$ for all C in the Borel σ -algebra generated by \mathcal{V} . Since the Borel σ -algebra generated by \mathcal{V} coincides with

$\mathcal{B}(\mathcal{Y} \times \mathbb{X})$, and any product set $B \times V_x$, $B \in \mathcal{B}(\mathcal{Y})$, and $V_x \in \mathcal{B}(\mathbb{X})$ belongs to $\mathcal{B}(\mathcal{Y} \times \mathbb{X})$, it implies $E[1\{Y \in B\} \kappa_1(D, Z, X) 1\{X \in V_x\}] \geq 0$ for all $B \in \mathcal{B}(\mathcal{Y})$ and $V_x \in \mathcal{B}(\mathbb{X})$. Hence, (B.13) follows. A similar line of reasoning yields the equivalence of the second inequalities of (B.12) to $E[\kappa_0(D, Z, X)g(Y, X)] \geq 0$ for all $g(\cdot, \cdot) \in \mathcal{G}$. Q.E.D.

B.3. Proof of Theorem 2.1

Let $\mathcal{F}_1 = \{1_{[[y, y', 1]]}(Y, D) : -\infty \leq y \leq y' \leq \infty\}$ and $\mathcal{F}_0 = \{1_{[[y, y', 0]]}(Y, D) : -\infty \leq y \leq y' \leq \infty\}$. We want to show

$$(B.14) \quad \limsup_{N \rightarrow \infty} \sup_{(P, Q) \in \mathcal{H}_0} \Pr(T_N > c_{N, 1-\alpha}) \leq \alpha,$$

where

$$T_N = \max \left\{ \sup_{f \in \mathcal{F}_1} \left\{ \frac{\hat{\lambda}^{1/2} Q_n(f) - (1 - \hat{\lambda})^{1/2} P_m(f)}{\xi \vee \sigma_{P_m, Q_n}(f, f)} \right\}, \right. \\ \left. \sup_{f \in \mathcal{F}_0} \left\{ \frac{((1 - \hat{\lambda})^{1/2} P_m(f) - \hat{\lambda}^{1/2} Q_n(f))}{\xi \vee \sigma_{P_m, Q_n}(f, f)} \right\} \right\}.$$

Consider a sequence $(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{H}_0$ at which $\Pr_{P^{[m(N)]}, Q^{[n(N)]}}(T_N > c_{N, 1-\alpha})$ differs from its supremum over \mathcal{H}_0 by $\varepsilon_N > 0$ or less with $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. Since $(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{P}^2$ are sequences in the uniformly tight class of probability measures (Condition RG(b)), there exists an a_N subsequence of N such that $(P^{[m(a_N)]}, Q^{[n(a_N)]})$ converges weakly to $(P_0, Q_0) \in \mathcal{P}^2$ as $N \rightarrow \infty$. Note that since $(P^{[m(N)]}, Q^{[n(N)]}) \in \mathcal{H}_0$, Lemma B.2 leads to $(P_0, Q_0) \in \mathcal{H}_0$. With abuse of notation, we read a_N as N and $(m(a_N), n(a_N))$ as (m, n) with $m + n = N$. Along such a sequence, we aim to show that $\limsup_{N \rightarrow \infty} \Pr_{P^{[m]}, Q^{[n]}}(T_N > c_{N, 1-\alpha}) \leq \alpha$ holds.

Using the notation of the weighted empirical processes introduced in Lemma B.6, we can write the test statistic as

$$T_N = \max \left\{ \sup_{f \in \mathcal{F}_1} \{-v_N(f) - h_N(f)\}, \sup_{f \in \mathcal{F}_0} \{v_N(f) + h_N(f)\} \right\},$$

where

$$h_N(f) = \sqrt{\frac{mn}{N}} \frac{P^{[m]}(f) - Q^{[n]}(f)}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}(f, f)}, \quad d = 1, 0.$$

By the almost sure representation theorem (see, e.g., Theorem 9.4 of Pollard (1990)), weak convergence of $(v_N(\cdot), P_m^{[m]}(\cdot), Q_n^{[n]}(\cdot), \sigma_{P_m^{[m]}, Q_n^{[n]}}^2(\cdot, \cdot))$ to $(v_0(\cdot), P_0(\cdot), Q_0(\cdot), \sigma_{P_0, Q_0}^2(\cdot, \cdot))$, as established in Lemmas B.3, B.4, and B.6, implies

the existence of a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ and random objects $\tilde{v}_0(\cdot)$, $\tilde{v}_N(\cdot)$, $\tilde{P}_m^{[m]}(\cdot)$, $\tilde{Q}_n^{[n]}(\cdot)$, and $\tilde{\sigma}_{P_m^{[m]}, Q_n^{[n]}}^2(\cdot, \cdot)$ defined on it, such that (i) $\tilde{v}_0(\cdot)$ has the same probability law as $v_0(\cdot)$, (ii) $(\tilde{v}_N(\cdot), \tilde{P}_m^{[m]}(\cdot), \tilde{Q}_n^{[n]}(\cdot), \tilde{\sigma}_{P_m^{[m]}, Q_n^{[n]}}^2(\cdot, \cdot))$ has the same probability law as $(v_N(\cdot), P_m^{[m]}(\cdot), Q_n^{[n]}(\cdot), \sigma_{P_m^{[m]}, Q_n^{[n]}}^2(\cdot, \cdot))$ for all N , and (iii)

$$(B.15) \quad \sup_{f \in \mathcal{F}} |\tilde{v}_N(f) - \tilde{v}_0(f)| \rightarrow 0,$$

$$(B.16) \quad \sup_{f \in \mathcal{F}} |\tilde{P}_m^{[m]}(f) - P_0(f)| \rightarrow 0,$$

$$(B.17) \quad \sup_{f \in \mathcal{F}} |\tilde{Q}_n^{[n]}(f) - Q_0(f)| \rightarrow 0,$$

$$(B.18) \quad \sup_{f, g \in \mathcal{F}} |\tilde{\sigma}_{P_m^{[m]}, Q_n^{[n]}}^2(f, g) - \sigma_{P_0, Q_0}^2(f, g)| \rightarrow 0 \quad \text{as } N \rightarrow \infty, \mathbb{P}\text{-a.s.}$$

Let \tilde{T}_N be the analogue of T_N defined on probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$,

$$\tilde{T}_N = \max \left\{ \sup_{f \in \mathcal{F}_1} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\}, \sup_{f \in \mathcal{F}_0} \{\tilde{v}_N(f) + \tilde{h}_N(f)\} \right\},$$

where $\tilde{h}_N(f) = \sqrt{\frac{mn}{N}} \frac{P_m^{[m]}(f) - Q_n^{[n]}(f)}{\xi \vee \sigma_{P_m^{[m]}, Q_n^{[n]}}^2(f, f)}$. Let $\tilde{c}_{N, 1-\alpha}$ be the bootstrap critical values,

which we view as a random object defined on the same probability space as $(\tilde{v}_N, \tilde{P}_m^{[m]}, \tilde{Q}_n^{[n]}, \tilde{\sigma}_{P_m^{[m]}, Q_n^{[n]}}^2)$ are defined. Note that the probability law of $\tilde{c}_{N, 1-\alpha}$ under \mathbb{P} is identical to the probability law of bootstrap critical value $c_{N, 1-\alpha}$ under $(P_m^{[m]}, Q_n^{[n]})$ for every N , because the distributions of $\tilde{c}_{N, 1-\alpha}$ and $c_{N, 1-\alpha}$ are determined by the distributions of $(\tilde{P}_m^{[m]}, \tilde{Q}_n^{[n]})$ and $(P_m^{[m]}, Q_n^{[n]})$, respectively, and $(\tilde{P}_m^{[m]}, \tilde{Q}_n^{[n]}) \sim (P_m^{[m]}, Q_n^{[n]})$ for every N , as claimed by the almost sure representation theorem.

By Lemma C.1 shown below, $\tilde{c}_{N, 1-\alpha} \rightarrow c_{1-\alpha}$ as $N \rightarrow \infty$, \mathbb{P} -a.s., where $c_{1-\alpha}$ is the $(1 - \alpha)$ th quantile of statistic

$$(B.19) \quad T_H \equiv \max \left\{ \sup_{f \in \mathcal{F}_1} \{-G_{H_0}(f) / (\xi \vee \sigma_{H_0}(f, f))\}, \right. \\ \left. \sup_{f \in \mathcal{F}_0} \{G_{H_0}(f) / (\xi \vee \sigma_{H_0}(f, f))\} \right\},$$

where $H_0 = \lambda P_0 + (1 - \lambda) Q_0$.

Since $\Pr_{P_m^{[m]}, Q_n^{[n]}}(T_N > c_{N, 1-\alpha}) = P(\tilde{T}_N > \tilde{c}_{N, 1-\alpha})$ for all N and $\tilde{c}_{N, 1-\alpha} \rightarrow c_{1-\alpha}$ as $N \rightarrow \infty$, \mathbb{P} -a.s., if there exists a random variable \tilde{T}^* defined on $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$,

such that

$$(A) \quad \limsup_{N \rightarrow \infty} \tilde{T}_N \leq \tilde{T}^*, \quad \mathbb{P}\text{-a.s.}, \quad \text{and}$$

$$(B) \quad \text{the cumulative distribution function of } \tilde{T}^* \text{ is continuous at } c_{1-\alpha} \text{ and} \\ \mathbb{P}(\tilde{T}^* > c_{1-\alpha}) \leq \alpha,$$

then the claim of the proposition follows from

$$\begin{aligned} \limsup_{N \rightarrow \infty} \Pr_{P^{[m]}, Q^{[m]}}(T_N > c_{N,1-\alpha}) &= \limsup_{N \rightarrow \infty} \mathbb{P}(\tilde{T}_N > \tilde{c}_{N,1-\alpha}) \\ &\leq \mathbb{P}(\tilde{T}^* > c_{1-\alpha}) \\ &\leq \alpha, \end{aligned}$$

where the second line follows from Fatou's lemma. Hence, in what follows, we aim to find a random variable \tilde{T}^* that satisfies (A) and (B).

Let η_N be a deterministic sequence that satisfies $\eta_N \rightarrow \infty$ and $\eta_N/\sqrt{N} \rightarrow 0$. Fix $\omega \in \Omega$ and define a sequence of subclass of \mathcal{F}_1 ,

$$\begin{aligned} \mathcal{F}_{1,\eta_N} &= \{f \in \mathcal{F}_1 : \tilde{h}_N(f) \leq \eta_N\} \\ &= \left\{ f \in \mathcal{F}_1 : \sqrt{\hat{\lambda}(1-\hat{\lambda})} \frac{P^{[m]}(f) - Q^{[m]}(f)}{\xi \vee \tilde{\sigma}_{P^{[m]}, Q^{[m]}}^2(f, f)} \leq \frac{\eta_N}{\sqrt{N}} \right\}. \end{aligned}$$

The first term in the maximum operator of \tilde{T}_N satisfies

$$\begin{aligned} (B.20) \quad &\sup_{f \in \mathcal{F}_1} \{-\tilde{v}_N(f) - \tilde{h}_{1,N}(f)\} \\ &= \max \left\{ \sup_{f \in \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\}, \sup_{f \in \mathcal{F}_1 \setminus \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\} \right\} \\ &\leq \max \left\{ \sup_{f \in \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f)\}, \sup_{f \in \mathcal{F}_1 \setminus \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\} \right\} \\ &\leq \max \left\{ \sup_{f \in \bigcup_{N' \geq N} \mathcal{F}_{1,\eta_{N'}}} \{-\tilde{v}_N(f)\}, \sup_{f \in \mathcal{F}_1 \setminus \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f)\} - \eta_N \right\} \end{aligned}$$

for every N , where the second line follows since $\tilde{h}_{1,N}(f) \geq 0$ for all $f \in \mathcal{F}_1$ under the assumption that $(P^{[m]}, Q^{[m]}) \in H_0$, and the third line follows because $\tilde{h}_N(f) > \eta_N$ for all $f \in \mathcal{F}_1 \setminus \mathcal{F}_{1,\eta_N}$. Since $\tilde{v}_N(\cdot)$ is \mathbb{P} -a.s. bounded and $\eta_N \rightarrow \infty$, it holds that

$$(B.21) \quad \sup_{f \in \mathcal{F}_1 \setminus \mathcal{F}_{1,\eta_N}} \{-\tilde{v}_N(f)\} - \eta_N \rightarrow -\infty, \quad \text{as } N \rightarrow \infty, \mathbb{P}\text{-a.s.}$$

On the other hand, since $\tilde{v}_N(\cdot)$, \mathbb{P} -a.s., converges to $\tilde{v}_0(\cdot)$ uniformly in \mathcal{F} , we have

$$(B.22) \quad \sup_{f \in \bigcup_{N' \geq N} \mathcal{F}_{1, \eta_{N'}}} \{-\tilde{v}_N(f)\} \rightarrow \sup_{f \in \mathcal{F}_{1, \infty}} \{-\tilde{v}_0(f)\}, \quad \text{as } N \rightarrow \infty, \mathbb{P}\text{-a.s.},$$

where $\mathcal{F}_{1, \infty} = \lim_{N \rightarrow \infty} \bigcup_{N' \geq N} \mathcal{F}_{1, \eta_{N'}}$. Let $\mathcal{F}_1^* = \{f \in \mathcal{F}_1 : P_0(f) = Q_0(f)\}$. By the construction of \mathcal{F}_{1, η_N} , every $f \in \mathcal{F}_{1, \infty}$ satisfies

$$(B.23) \quad \liminf_{N \rightarrow \infty} \left\{ \sqrt{\hat{\lambda}(1 - \hat{\lambda})} \frac{P^{[m]}(f) - Q^{[m]}(f)}{\xi \vee \tilde{\sigma}_{P^{[m]}, Q^{[m]}}^2(f, f)} \right\} = 0.$$

Since $P^{[m]}(f) - Q^{[m]}(f)$ converges to $P_0(f) - Q_0(f)$ by Lemma B.2, any f satisfying (B.23) belongs to \mathcal{F}_1^* . Hence, we have

$$(B.24) \quad \sup_{f \in \mathcal{F}_{1, \infty}} \{-\tilde{v}_0(f)\} \leq \sup_{f \in \mathcal{F}_1^*} \{-\tilde{v}_0(f)\}, \quad \mathbb{P}\text{-a.s.}$$

By combining (B.20), (B.21), (B.22), and (B.24), we obtain

$$\limsup_{N \rightarrow \infty} \sup_{f \in \mathcal{F}_1} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\} \leq \sup_{f \in \mathcal{F}_1^*} \{-\tilde{v}_0(f)\}, \quad \mathbb{P}\text{-a.s.}$$

In a similar manner, it can be shown that

$$\limsup_{N \rightarrow \infty} \sup_{f \in \mathcal{F}_0} \{\tilde{v}_N(f) + \tilde{h}_N(f)\} \leq \sup_{f \in \mathcal{F}_0^*} \{\tilde{v}_0(f)\}, \quad \mathbb{P}\text{-a.s.},$$

where $\mathcal{F}_0^* = \{f \in \mathcal{F}_0 : P_0(f) = Q_0(f)\}$. Hence, \tilde{T}^* defined by

$$\tilde{T}^* = \max \left\{ \sup_{f \in \mathcal{F}_1^*} \{-\tilde{v}_0(f)\}, \sup_{f \in \mathcal{F}_0^*} \{\tilde{v}_0(f)\} \right\}$$

satisfies condition (A).

Next, we show that the thus-defined \tilde{T}^* satisfies (B). First, we show that \tilde{T}^* is stochastically dominated by T_H . Note that statistic T_H defined in (B.19) can be written as

$$T_H = \max \left\{ T_H^*, \sup_{f \in \mathcal{F}_1 \setminus \mathcal{F}_1^*} \left\{ -\frac{G_{H_0}(f)}{\xi \vee \sigma_{H_0}(f, f)} \right\}, \sup_{f \in \mathcal{F}_0 \setminus \mathcal{F}_0^*} \left\{ \frac{G_{H_0}(f)}{\xi \vee \sigma_{H_0}(f, f)} \right\} \right\},$$

where

$$T_H^* = \max \left\{ \sup_{f \in \mathcal{F}_1^*} \{-G_{H_0}(f)/(\xi \vee \sigma_{H_0}(f, f))\}, \right. \\ \left. \sup_{f \in \mathcal{F}_0^*} \{G_{H_0}(f)/(\xi \vee \sigma_{H_0}(f, f))\} \right\}.$$

If the distribution of T_H^* is identical to \tilde{T}^* , then the distribution of T_H stochastically dominates \tilde{T}^* so that we can ascertain the second part of (B). Hence, in what follows we show that T_H^* and \tilde{T}^* follow the same probability law. Stochastic processes defined on the subdomain of \mathcal{F} , $\mathcal{F}^* = \mathcal{F}_1^* \cup \mathcal{F}_0^*$ are

$$\begin{aligned} u(f) &= -v_0(f)1\{f \in \mathcal{F}_1^*\} + v_0(f)1\{f \in \mathcal{F}_0^*\}, \\ u_H(f) &= -\frac{G_{H_0}(f)}{\xi \vee \sigma_{H_0}(f, f)}1\{f \in \mathcal{F}_1^*\} + \frac{G_{H_0}(f)}{\xi \vee \sigma_{H_0}(f, f)}1\{f \in \mathcal{F}_0^*\}. \end{aligned}$$

Note first that, for $f \in \mathcal{F}^*$, $P_0(f) = Q_0(f) = H_0(f)$ implies that

$$\sigma_{P_0, Q_0}^2(f, f) = P_0(f)(1 - P_0(f)) = \sigma_{H_0}^2(f, f).$$

Hence, $\text{Var}(u(f)) = \text{Var}(u_H(f))$ holds for every $f \in \mathcal{F}^*$. To also show equivalence of the covariance kernels of $u(\cdot)$ and $u_H(\cdot)$, consider, for $f, g \in \mathcal{F}^*$,

$$\begin{aligned} &\text{Cov}(u(f), u(g)) \\ &= \frac{(1 - \lambda)[P_0(fg) - P_0(f)P_0(g)] + \lambda[Q_0(fg) - Q_0(f)Q_0(g)]}{(\xi \vee \sigma_{P_0, Q_0}(f, f))(\xi \vee \sigma_{P_0, Q_0}(g, g))} \\ &= \frac{[(1 - \lambda)P_0 + \lambda Q_0](fg) - H_0(f)H_0(g)}{(\xi \vee \sigma_{H_0}(f, f))(\xi \vee \sigma_{H_0}(g, g))}. \end{aligned}$$

If $f \in \mathcal{F}_1^*$ and $g \in \mathcal{F}_0^*$, then $P_0(fg) = Q_0(fg) = H_0(fg) = 0$. If $f, g \in \mathcal{F}_1^*$, then $(P_0, Q_0) \in \mathcal{H}_0$ implies $0 \geq (P_0 - Q_0)(fg) \geq (P_0 - Q_0)(f) = 0$, so $P_0(fg) = Q_0(fg) = H_0(fg)$. Similarly, if $f, g \in \mathcal{F}_0^*$, then $(P_0, Q_0) \in \mathcal{H}_0$ implies $0 \leq (P_0 - Q_0)(fg) \leq (P_0 - Q_0)(f) = 0$, so $P_0(fg) = Q_0(fg) = H_0(fg)$ holds as well. Thus, we obtain

$$\begin{aligned} \text{Cov}(u(f), u(g)) &= \frac{H_0(fg) - H_0(f)H_0(g)}{(\xi \vee \sigma_{H_0}(f, f))(\xi \vee \sigma_{H_0}(g, g))} \\ &= \text{Cov}(u_H(f), u_H(g)) \end{aligned}$$

for every $f, g \in \mathcal{F}^*$. Equivalence of the covariance kernels implies equivalence of the probability laws of the mean zero Gaussian processes, so we conclude $T_H^* \sim \tilde{T}^*$. Hence, $P(\tilde{T}^* > c_{1-\alpha}) \leq \Pr(T_H > c_{1-\alpha}) = \alpha$.

To check the first requirement of (B), we show continuity of the cumulative distribution function (c.d.f.) of \tilde{T}^* at $c_{1-\alpha}$ by applying the absolute continuity theorem for the supremum of Gaussian processes (Tsirelson (1975)), which says the supremum of Gaussian processes has a continuous c.d.f. except at the left limit of its support. By the definition of $u_H(\cdot)$, T_H can be equivalently written as $T_H = \sup_{f \in \mathcal{F}} \{u_H(f)\}$. Note first that the support of T_H contains 0 since

\mathcal{F} contains an indicator function for a singleton set in \mathcal{X} at which $u_{H_0}(f) = 0$ holds with probability 1. Following the symmetry argument of the mean zero Gaussian process, which we borrowed from the proof of Proposition 2.2 in [Abadie \(2002\)](#), we have

$$\Pr(T_H \leq 0) = \Pr((\nexists f \in \mathcal{F}, u_H(f) > 0)) = \Pr((\nexists f \in \mathcal{F}, u_H(f) < 0)).$$

By Condition RG(a), $u_H(\cdot)$ is not a degenerate process, so

$$\Pr((\nexists f \in \mathcal{F}, u_H(f) < 0) \cap (\nexists f \in \mathcal{F}, u_H(f) < 0)) = 0.$$

Hence,

$$\begin{aligned} 1 &\geq \Pr((\nexists f \in \mathcal{F}, u_H(f) < 0) \cup (\nexists f \in \mathcal{F}, u_H(f) < 0)) \\ &= 2\Pr(T_H \leq 0), \end{aligned}$$

implying that the probability mass that T_H can have at the left limit of its support is less than or equal to $1/2$. As a result, $c_{1-\alpha}$ for $\alpha \in (0, 1/2)$ lies in the region where the c.d.f. of T_H is continuous. Since \tilde{T}^* is also a supremum of the mean zero Gaussian process and, as already shown, it is stochastically dominated by T_H , the c.d.f. of \tilde{T}^* is also continuous at $c_{1-\alpha}$. This completes the proof of Theorem 2.1(i).

To prove claim (ii), assume that the first inequality of (1.1) is violated for some Borel set $B \subset \mathcal{Y}$. By Lemma B.7, there exists some $f^* \in \mathcal{F}_1$ such that $0 \leq P(f^*) < Q(f^*)$ holds. Then we have

$$\begin{aligned} \text{(B.25)} \quad T_N &= \max \left\{ \sup_{f \in \mathcal{F}_1} \left\{ \frac{\hat{\lambda}^{1/2} Q_n(f) - (1 - \hat{\lambda})^{1/2} P_m(f)}{\xi \vee \sigma_{P_m, Q_n}(f, f)} \right\}, \right. \\ &\quad \left. \sup_{f \in \mathcal{F}_0} \left\{ \frac{(1 - \hat{\lambda})^{1/2} P_m(f) - \hat{\lambda}^{1/2} Q_n(f)}{\xi \vee \sigma_{P_m, Q_n}(f, f)} \right\} \right\} \\ &\geq \frac{(\hat{\lambda}^{1/2} G_{n, Q}(f^*) - (1 - \hat{\lambda})^{1/2} G_{m, P}(f^*))}{\xi \vee \sigma_{P_m, Q_n}(f^*, f^*)} \\ &\quad + \sqrt{\frac{mn}{N}} \frac{Q(f^*) - P(f^*)}{\xi \vee \sigma_{P_m, Q_n}(f^*, f^*)}, \end{aligned}$$

where the second term of (B.25) diverges to positive infinity, while the first term is stochastically bounded asymptotically. Since the bootstrap critical values $c_{N, 1-\alpha}$ converge to $c_{1-\alpha} < \infty$ irrespective of whether the null holds true, the rejection probability converges to 1.

APPENDIX C: CONVERGENCE OF THE BOOTSTRAP CRITICAL VALUES AND
PROOF OF THEOREM 2.2

C.1. *Lemma on Convergence of the Bootstrap Critical Values*

The proof of Theorem 2.1 given in the previous section assumes \mathbb{P} -almost sure convergence of the bootstrap critical value $\tilde{c}_{N,1-\alpha}$ to $c_{1-\alpha}$. This convergence claim is proven by the next lemma. The probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ and the random objects with tildes used in the following proof are defined in the proof of Theorem 2.1(i) by the almost sure representation theorem.

LEMMA C.1: *Suppose Condition RG. Let $\tilde{c}_{N,1-\alpha}$ be the bootstrap critical value of Algorithm 2.1 constructed from $\tilde{H}_N^{[N]} = \hat{\lambda} \tilde{P}_m^{[m]} + (1 - \hat{\lambda}) \tilde{Q}_n^{[n]}$, which is viewed as a sequence of random variables $\{\tilde{c}_{N,1-\alpha} : N = 1, 2, \dots\}$ defined on probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$. It holds that $\tilde{c}_{N,1-\alpha}$ converges to $c_{1-\alpha}$ as $N \rightarrow \infty$, \mathbb{P} -a.s., where $c_{1-\alpha}$ is the $(1 - \alpha)$ th quantile of statistic*

$$T_H = \max \left\{ \sup_{f \in \mathcal{F}_1} \left\{ -G_{H_0}(f) / (\xi \vee \sigma_{H_0}(f, f)) \right\}, \right. \\ \left. \sup_{f \in \mathcal{F}_0} \left\{ G_{H_0}(f) / (\xi \vee \sigma_{H_0}(f, f)) \right\} \right\},$$

where $H_0 = \lambda P_0 + (1 - \lambda) Q_0$.

PROOF: Let sequence $\{\tilde{H}_N^{[N]} : N = 1, 2, \dots\}$ be given, and let P_m^* and Q_n^* be the bootstrap empirical probability measures with size m and size n , respectively, drawn i.i.d. from $\tilde{H}_N^{[N]}$. Define bootstrap weighted empirical processes indexed by $f \in \mathcal{F}$ as

$$v_N^*(\cdot) = \sqrt{\frac{mn}{N}} \frac{P_m^*(\cdot) - Q_n^*(\cdot)}{\xi \vee \sigma_{P_m^*, Q_n^*}(\cdot, \cdot)} \\ = \frac{(1 - \hat{\lambda})^{1/2} G_{m, \tilde{H}_N^{[N]}}^*(\cdot) - \hat{\lambda}^{1/2} G_{n, \tilde{H}_N^{[N]}}^{*'}(\cdot)(f)}{\xi \vee \sigma_{P_m^*, Q_n^*}(\cdot, \cdot)},$$

where $G_{m, \tilde{H}_N^{[N]}}^*(\cdot) = \sqrt{m}(P_m^* - \tilde{H}_N^{[N]})(\cdot)$ and $G_{n, \tilde{H}_N^{[N]}}^{*'}(\cdot) = \sqrt{n}(Q_n^* - \tilde{H}_N^{[N]})(\cdot)$ are two independent bootstrap empirical processes given $\{\tilde{H}_N^{[N]} : N = 1, 2, \dots\}$. Let (X_1, \dots, X_N) be the N support points of $\tilde{H}_N^{[N]}$ and let δ_X be the point-mass measure at X . To apply the uniform central limit theorem with exchangeable multipliers (Theorem 3.6.13 of [van der Vaart and Wellner \(1996\)](#)), we introduce multinomial random vector $(M_{m,1}, \dots, M_{m,N})$ that is independent of (X_1, \dots, X_N) and has parameters $(m, \frac{1}{N}, \dots, \frac{1}{N})$. We express

$G_{m, \tilde{H}_N^{[N]}}^*(\cdot)$ as

$$\begin{aligned} G_{m, \tilde{H}_N^{[N]}}^*(\cdot) &= \frac{1}{\sqrt{m}} \sum_{i=1}^N \left(M_{m,i} - \frac{m}{N} \right) \delta_{X_i}(\cdot) \\ &= \frac{1}{\sqrt{m}} \sum_{i=1}^N \left(M_{m,i} - \frac{m}{N} \right) (\delta_{X_i} - H^{[N]})(\cdot) \\ &= \hat{\lambda}^{-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_{m,i} (\delta_{X_i} - H^{[N]})(\cdot), \end{aligned}$$

where $\xi_{m,i} = M_{m,i} - \frac{m}{N}$, $i = 1, \dots, N$. Note that $(\xi_{m,1}, \dots, \xi_{m,N})$ are exchangeable random variables by construction and $E(\frac{1}{N} \sum_{i=1}^N \xi_i^2) = \frac{m}{N}(1 - \frac{1}{N}) \rightarrow \lambda$, as $N \rightarrow \infty$. On the other hand, since $H^{[N]}$ converges weakly to H_0 , an application of Lemma B.5 yields $\frac{1}{\sqrt{N}} \sum_{i=1}^N (\delta_{X_i} - H^{[N]})(\cdot) \rightsquigarrow G_{H_0}(\cdot)$. Hence, the uniform central limit theorem with exchangeable multipliers (Theorem 3.6.13 of van der Vaart and Wellner (1996)) leads to $G_{m, \tilde{H}_N^{[N]}}^*(\cdot) \rightsquigarrow G_{H_0}(\cdot)$ for \mathbb{P} -almost every sequence $\{\tilde{H}_N^{[N]} : N = 1, 2, \dots\}$. By the same reasoning, we have $G_{n, \tilde{H}_N^{[N]}}^{*'}(\cdot) \rightsquigarrow G_{H_0}'(\cdot)$ for \mathbb{P} -almost every sequence $\{\tilde{H}_N^{[N]} : N = 1, 2, \dots\}$, where $G_{H_0}'(\cdot)$ is an H_0 -Brownian bridge process independent of $G_{H_0}(\cdot)$.

Hence, the numerator of $v_N^*(\cdot)$ converges weakly to $(1 - \lambda)^{1/2} G_{H_0}(\cdot) - \lambda^{1/2} G_{H_0}'(\cdot)$, \mathbb{P} -a.s., sequences of $\{\tilde{H}_N^{[N]}\}$. Note that the covariance kernel of $(1 - \lambda)^{1/2} G_{H_0}(\cdot) - \lambda^{1/2} G_{H_0}'(\cdot)$ coincides with that of the H_0 -Brownian bridge, so we conclude that

$$(C.1) \quad (1 - \hat{\lambda})^{1/2} G_{m, \tilde{H}_N^{[N]}}^*(\cdot) - \hat{\lambda}^{1/2} G_{n, \tilde{H}_N^{[N]}}^{*'}(\cdot)(f) \rightsquigarrow G_{H_0}(\cdot),$$

\mathbb{P} -a.s. sequences of $\{\tilde{H}_N^{[N]}\}$.

Regarding the bootstrap covariance kernel, we have convergence of $\sup_{f \in \mathcal{F}} |\sigma_{P_m^*, Q_n^*}(f, f) - \sigma_{H_0}(f, f)|$ to zero (in probability in terms of the probability law of bootstrap resampling given $\tilde{H}_N^{[N]}$) for \mathbb{P} -a.s. sequences of $\{\tilde{H}_N^{[N]}\}$, since

$$(C.2) \quad \sup_{f \in \mathcal{F}} |\sigma_{P_m^*, Q_n^*}^2(f, f) - \sigma_{H_0}^2(f, f)| \leq \sup_{f \in \mathcal{F}} |\sigma_{P_m^*, Q_n^*}^2(f, f) - \sigma_{\tilde{H}_N^{[N]}}^2(f, f)| \\ + \sup_{f \in \mathcal{F}} |\sigma_{\tilde{H}_N^{[N]}}^2(f, f) - \sigma_{H_0}^2(f, f)|,$$

where the first term in the right-hand side converges to zero (in probability in terms of the probability law of bootstrap resampling) by applying the

Glivenko–Cantelli theorem for the triangular arrays as given in Lemma B.1, and the convergence to zero \mathbb{P} -a.s. for the second term follows from the almost sure representation theorem, (B.16) and (B.17).

By putting together (C.1) and (C.2), and repeating the proof of the asymptotic uniform equicontinuity as given in (B.11) above, we obtain

$$\begin{aligned} v_N^*(\cdot) &\rightsquigarrow \frac{(1-\lambda)^{1/2}G_{H_0}(\cdot) - \hat{\lambda}^{1/2}G'_{H_0}(\cdot)}{\xi \vee \sigma_{H_0}(\cdot, \cdot)} \\ &\sim \frac{G_{H_0}(\cdot)}{\xi \vee \sigma_{H_0}(\cdot, \cdot)} \quad \text{as } N \rightarrow \infty, \end{aligned}$$

for \mathbb{P} -almost every sequence of $\{\tilde{H}_N^{[N]}\}$. The bootstrap test statistics T_N^* is a continuous functional of $v_N^*(\cdot)$, so the continuous mapping theorem leads to

$$\begin{aligned} T_N^* &\rightsquigarrow T_H = \max \left\{ \sup_{f \in \mathcal{F}_1} \{-G_{H_0}(f)/(\xi \vee \sigma_{H_0}(f, f))\}, \right. \\ &\quad \left. \sup_{f \in \mathcal{F}_0} \{G_{H_0}(f)/(\xi \vee \sigma_{H_0}(f, f))\} \right\} \quad \text{as } N \rightarrow \infty, \end{aligned}$$

for \mathbb{P} -almost every sequence of $\{\tilde{H}_N^{[N]}\}$. We already showed in the proof of Theorem 2.1(i) that the c.d.f. of T_H is continuous at $c_{1-\alpha}$ for $\alpha \in (0, 1/2)$. Hence, the bootstrap critical values $\tilde{c}_{N,1-\alpha}$ converge to $c_{1-\alpha}$, \mathbb{P} -a.s. *Q.E.D.*

C.2. Proof of Theorem 2.2

PROOF: By Assumption LA(c) and the Portmanteau theorem, $(P^{[N]}, Q^{[N]} \in \mathcal{P}^2 : N = 1, 2, \dots)$ converges weakly to $(P_0, Q_0) \in \mathcal{H}_0$. We can therefore apply all the lemmas established in Appendixes B and C.1, and, as done in the proof of Theorem 2.1(i), we can define via the almost sure representation theorem a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ and random objects with tildes that copy the ones defined in a sequence of probability spaces in terms of $(P^{[N]}, Q^{[N]} : N = 1, 2, \dots)$. By Lemma C.1, the bootstrap critical values $\tilde{c}_{N,1-\alpha}$ converge to $c_{1-\alpha}$, the $(1-\alpha)$ th quantile of T_H , \mathbb{P} -a.s., which depends only on $(\alpha, \xi, \lambda, P_0, Q_0)$. Suppose that $([y, y'], d = 1)$ satisfies Assumption LA(a) and (d). Let $\tilde{v}_N(\cdot) = \frac{(1-\lambda)^{1/2}\tilde{G}_{m,P^{[N]}}(\cdot) - \hat{\lambda}^{1/2}\tilde{G}_{n,Q^{[N]}}(\cdot)}{\xi \vee \tilde{\sigma}_{P_m^{[N]}, Q_n^{[N]}}(\cdot, \cdot)}$ be the weighted empirical process defined on $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$, where $\tilde{G}_{m,P^{[N]}}(\cdot) = \sqrt{m}(\tilde{P}_m^{[N]} - P^{[N]})(\cdot)$ and $\tilde{G}_{n,Q^{[N]}}(\cdot) = \sqrt{m}(\tilde{Q}_n^{[N]} - Q^{[N]})(\cdot)$. Note the probability law of the test statistic is that of

$$\tilde{T}_N = \max \left\{ \sup_{f \in \mathcal{F}_1} \{-\tilde{v}_N(f) - \tilde{h}_N(f)\}, \sup_{f \in \mathcal{F}_0} \{\tilde{v}_N(f) + \tilde{h}_N(f)\} \right\},$$

induced by \mathbb{P} , where

$$\tilde{h}_N(f) = \sqrt{\frac{mn}{N}} \frac{P^{[N]}(f) - Q^{[N]}(f)}{\xi \vee \tilde{\sigma}_{P_m^{[N]}, Q_n^{[N]}}(f, f)}.$$

Since \tilde{T}_N is bounded from below by

$$-\tilde{v}_N([y, y'], 1) - \tilde{h}_N([y, y'], 1),$$

the rejection probability is also bounded from below by

$$\mathbb{P}(-\tilde{v}_N([y, y'], 1) - \tilde{h}_N([y, y'], 1) \geq \tilde{c}_{N, 1-\alpha}).$$

By Assumption LA(c), and by applying Lemmas B.4 and B.6, $\tilde{v}_N([y, y'], 1) - \tilde{h}_N([y, y'], 1)$ converges \mathbb{P} -a.s. to

$$-\tilde{v}_0([y, y'], 1) - \frac{[\lambda(1-\lambda)]^{1/2} \Delta\beta([y, y'], 1)}{\xi \vee \sigma_{P_0, Q_0}([y, y'], 1)},$$

which follows the Gaussian with mean

$$-\frac{[\lambda(1-\lambda)]^{1/2} \Delta\beta([y, y'], 1)}{\xi \vee \sigma_{P_0, Q_0}([y, y'], 1)}$$

and variance

$$\min\left\{\frac{\sigma_{P_0, Q_0}^2([y, y'], 1)}{\xi^2}, 1\right\}.$$

Hence, we obtain

$$\begin{aligned} & \mathbb{P}(-\tilde{v}_N([y, y'], 1) - \tilde{h}_N([y, y'], 1) \geq \tilde{c}_{N, 1-\alpha}) \\ & \rightarrow \mathbb{P}\left(-\tilde{v}_0([y, y'], 1) - \frac{[\lambda(1-\lambda)]^{1/2} \Delta\beta([y, y'], 1)}{\xi \vee \sigma_{P_0, Q_0}([y, y'], 1)} \geq c_{1-\alpha}\right) \\ & = 1 - \Phi\left(\left(\frac{\sigma_{P_0, Q_0}^2([y, y'], 1)}{\xi^2} \wedge 1\right)^{-1}\right. \\ & \quad \left. \times \left(c_{1-\alpha} - \frac{[\lambda(1-\lambda)]^{1/2} |\Delta\beta([y, y'], 1)|}{\xi \vee \sigma_{P_0, Q_0}([y, y'], 1)}\right)\right). \end{aligned}$$

In case $([y, y'], d = 0)$ satisfies Assumption LA(i) and (iv), a similar argument yields the same lower bound. *Q.E.D.*

APPENDIX D: MONTE CARLO STUDIES

This section examines the finite sample performance of the Monte Carlo test. In assessing finite sample type I errors of the test, we consider a data generating process on a boundary of \mathcal{H}_0 , so that the theoretical type I error of the test equals a nominal size asymptotically,

$$p(y, D = 1) = q(y, D = 1) = 0.5 \times \mathcal{N}(1, 1),$$

$$p(y, D = 0) = q(y, D = 0) = 0.5 \times \mathcal{N}(0, 1),$$

where $N(\mu, \sigma^2)$ is the probability density of a normal random variable with mean μ and σ^2 .

In computing the first (second) supremum of the test statistic, the boundary points of intervals are chosen by every pair of Y -values observed in the subsample of $\{D = 1, Z = 0\}$ ($\{D = 0, Z = 1\}$). To assess how the test performance depends on the choice of trimming constant, we run simulations for each of the four specification of the trimming constant:

$$\xi_1 = \sqrt{0.005(1 - 0.005)} \approx 0.07,$$

$$\xi_2 = \sqrt{0.05(1 - 0.05)} \approx 0.22,$$

$$\xi_3 = \sqrt{0.1(1 - 0.1)} = 0.3,$$

$$\xi_4 = 1.$$

Note that ξ_k , $k = 1, 2, 3$, has the form of $\sqrt{\pi_k(1 - \pi_k)}$, and π_k can be interpreted as that if both $P_m([y, y'], d)$ and $Q_n([y, y'], d)$ are less than π_k , we weigh the difference of the empirical distribution by the inverse of ξ instead of the inverse of its standard deviation estimate. Accordingly, as π_k becomes larger, we put relatively less weight on the differences of the empirical probabilities for thinner probability events. The fourth choice of ξ , $\xi_4 = 1$, makes the test statistic identical to the nonweighted KS statistic.

Table II shows the simulated test size. The rejection probabilities are slightly upwardly biased relative to the nominal sizes, while they overall show good size performance even in the cases with the sample sizes being as small as $(m, n) = (100, 100)$ and being unbalanced as much as $(m, n) = (100, 1000)$. It is also worth noting that these test sizes are not sensitive to the choice of trimming constant.

To see the finite sample power performance of our test, we simulate the rejection probabilities of the bootstrap test against four different specifications of fixed alternatives. These four data generating processes (DGPs) share

$$\Pr(Z = 1) = \frac{1}{2}, \quad \Pr(D = 1|Z = 1) = 0.55,$$

$$\Pr(D = 1|Z = 0) = 0.45,$$

TABLE II
MONTE CARLO TEST SIZE: 1000 MONTE CARLO ITERATIONS; 300 BOOTSTRAP ITERATIONS^a

Trimming Constant: Nominal Size:	$\xi_1 \approx 0.07$			$\xi_2 \approx 0.21$			$\xi_3 = 0.3$			$\xi_4 = 1$		
	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
$(m, n): (100, 100)$	0.13	0.07	0.01	0.13	0.07	0.01	0.14	0.06	0.01	0.13	0.06	0.01
$(100, 500)$	0.11	0.06	0.01	0.10	0.06	0.01	0.11	0.05	0.01	0.10	0.05	0.01
$(500, 500)$	0.13	0.06	0.02	0.12	0.07	0.02	0.11	0.06	0.02	0.12	0.05	0.01
$(100, 1000)$	0.12	0.06	0.02	0.12	0.06	0.01	0.13	0.06	0.02	0.12	0.06	0.02
$(1000, 1000)$	0.14	0.07	0.02	0.13	0.08	0.02	0.13	0.06	0.02	0.12	0.06	0.01

^aThe statistic is equivalent to the nonweighted KS statistics when $\xi_4 = 1$.

$$p(y, D = 1) = 0.55 \times \mathcal{N}(0, 1),$$

$$p(y, D = 0) = 0.45 \times \mathcal{N}(0, 1), \quad q(y, D = 0) = 0.55 \times \mathcal{N}(0, 1),$$

while they differ in terms of specifications of the treated outcome distribution conditional on $Z = 0$:

$$\text{DGP1: } q(y, D = 1) = 0.45 \times \mathcal{N}(-0.7, 1),$$

$$\text{DGP2: } q(y, D = 1) = 0.45 \times \mathcal{N}(0, 1.675^2),$$

$$\text{DGP3: } q(y, D = 1) = 0.45 \times \mathcal{N}(0, 0.515^2),$$

$$\text{DGP4: } q(y, D = 1) = 0.45 \times \sum_{l=1}^5 w_l \mathcal{N}(\mu_l, 0.125^2),$$

$$(w_1, \dots, w_5) = (0.15, 0.2, 0.3, 0.2, 0.15),$$

$$(\mu_1, \dots, \mu_5) = (-1, -0.5, 0, 0.5, 1).$$

In all these specifications, violations of the testable implication occur only for the treatment outcome densities. As plotted in Figure 5, the ways that the densities $p(y, 1)$ and $q(y, 1)$ intersect differ across the DGPs. In DGP1, $p(y, 1)$ and $q(y, 1)$ are differentiated horizontally, and they intersect only once. In DGP2, the violations occur at the tail parts of $p(y, 1)$ and $q(y, 1)$, whereas, in DGP3, the violation occurs around the modes of $p(y, 1)$ and $q(y, 1)$. In DGP4, $q(y, 1)$ is specified to be oscillating sharply around $p(y, 1)$ and they intersect many times. In all these specifications, $p(y, 1)$ and $q(y, 1)$ are designed to be equally distant in terms of the one-sided total variation distance, that is, $\int_{-\infty}^{\infty} \max\{(q(y, 1) - p(y, 1)), 0\} dy \approx 0.092$ for all the DGPs.

Table III shows the simulated rejection probabilities, based on which several remarks follow. First, we observe that the rejection probabilities vary depending on the DGPs and the choice of trimming constant. When the violations

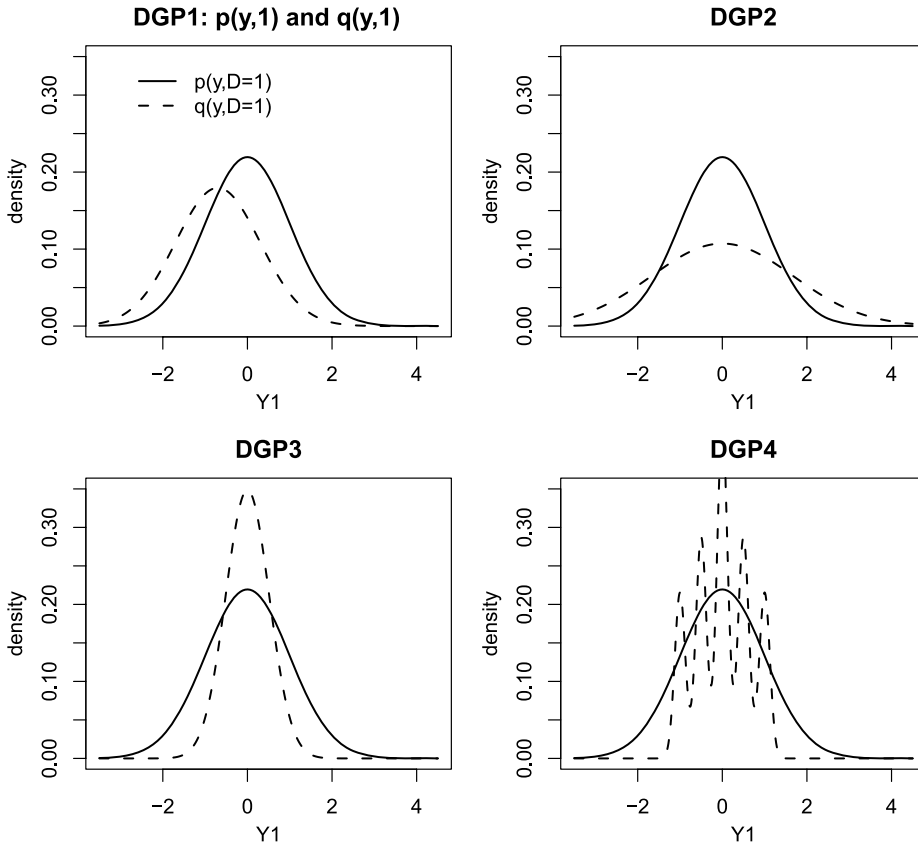


FIGURE 5.—Specification of densities in Monte Carlo experiments of test power.

occur for the tail parts of the densities (DGP2), smaller ξ yields a significantly higher power. In contrast, if violations occur on a fatter part of the densities (DGP1, DGP3, and DGP4), middle range ξ 's and $\xi = 1$ tend to exhibit a slightly higher power than the smallest choice of ξ . This suggests that if a likely violation of the testable implications is expected at the tail parts of the distributions, it is important to use a variance-weighted statistic with a sufficiently small ξ such as $\xi = 0.07$. Given these simulation findings that a power loss by choosing $\xi = 0.07$ instead of the medium size ξ or $\xi = 1$ is not so severe in the other cases, we can argue that in case there is no prior knowledge available about a likely alternative, one default choice of ξ is as small as 0.07. At the same time, it is also worth reporting the test results with several other choices of $\xi \in (0, 0.5]$. Second, the rows of unbalanced sample sizes indicate that the magnitude of the rejection probabilities tends to depend on a smaller sample size of (m, n) , rather than the total sample size N , so a lack of power

TABLE III
REJECTION PROBABILITIES AGAINST FIXED ALTERNATIVES: 1000 MONTE CARLO ITERATIONS;
300 BOOTSTRAP ITERATIONS

Trimming Constant:	$\xi_1 \approx 0.07$			$\xi_2 = 0.22$			$\xi_3 = 0.3$			$\xi_4 = 1$			
	Nominal Size:	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
DGP1 (m, n):	(100, 100)	0.31	0.22	0.10	0.31	0.21	0.10	0.30	0.21	0.10	0.23	0.15	0.05
	(100, 500)	0.42	0.31	0.14	0.56	0.43	0.22	0.57	0.44	0.21	0.38	0.24	0.07
	(500, 500)	0.93	0.88	0.77	0.95	0.91	0.78	0.96	0.92	0.79	0.89	0.80	0.52
	(100, 1000)	0.38	0.28	0.12	0.58	0.46	0.24	0.59	0.46	0.23	0.39	0.26	0.09
	(1000, 1000)	0.99	0.98	0.94	1.00	0.99	0.97	0.99	0.98	0.94	0.99	0.98	0.93
DGP2 (m, n):	(100, 100)	0.16	0.09	0.02	0.15	0.09	0.02	0.08	0.04	0.00	0.01	0.00	0.00
	(100, 500)	0.35	0.23	0.07	0.17	0.10	0.02	0.07	0.02	0.00	0.01	0.00	0.00
	(500, 500)	0.95	0.91	0.73	0.86	0.77	0.53	0.56	0.40	0.16	0.10	0.03	0.01
	(100, 1000)	0.40	0.26	0.08	0.20	0.09	0.02	0.06	0.03	0.00	0.01	0.00	0.00
	(1000, 1000)	1.00	1.00	1.00	1.00	0.99	0.97	0.96	0.90	0.67	0.52	0.27	0.05
DGP3 (m, n):	(100, 100)	0.30	0.20	0.09	0.30	0.20	0.09	0.33	0.22	0.09	0.34	0.22	0.09
	(100, 500)	0.32	0.21	0.08	0.51	0.38	0.19	0.59	0.46	0.23	0.55	0.40	0.15
	(500, 500)	0.77	0.69	0.54	0.83	0.76	0.57	0.87	0.79	0.62	0.89	0.82	0.61
	(100, 1000)	0.30	0.18	0.06	0.53	0.40	0.18	0.61	0.47	0.25	0.54	0.41	0.18
	(1000, 1000)	0.98	0.96	0.89	0.99	0.98	0.93	1.00	0.99	0.96	1.00	0.99	0.95
DGP4 (m, n):	(100, 100)	0.12	0.07	0.02	0.11	0.07	0.02	0.09	0.05	0.02	0.09	0.05	0.01
	(100, 500)	0.23	0.13	0.04	0.20	0.11	0.03	0.15	0.09	0.02	0.12	0.05	0.01
	(500, 500)	0.46	0.33	0.17	0.45	0.33	0.16	0.33	0.22	0.10	0.23	0.13	0.03
	(100, 1000)	0.26	0.15	0.05	0.22	0.13	0.05	0.15	0.10	0.03	0.11	0.06	0.01
	(1000, 1000)	0.78	0.67	0.48	0.80	0.69	0.50	0.51	0.38	0.19	0.45	0.30	0.11

should be acknowledged when one of the sample sizes is small. Third, for the magnitudes of violations considered in these simulations, the rejection probabilities are sufficiently close to 1 (for some smaller choices ξ only for DGP2) if the sample sizes are as large as $(m, n) = (1000, 1000)$.

REFERENCES

- ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. [18]
- (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. [12]
- ANDREWS, D. W. K., AND X. SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666. [11,12]
- POLLARD, D. (1990): "Empirical Processes: Theory and Applications," in *NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. 2. Alexandria, VA: American Statistical Association. [13]
- TSIRELSON, V. S. (1975): "The density of the maximum of a Gaussian Process," *Theory of Probability and Its Applications*, 20, 817–856. [17]
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [5,6,8,9,19,20]

*Dept. of Economics, University College London, Gower Street, London, WC1E
6BT, U.K.; t.kitagawa@ucl.ac.uk.*

Manuscript received October, 2013; final revision received December, 2014.