Online Appendix for "Testing Treatment Effect Heterogeneity in Regression Discontinuity Designs"

Yu-Chin Hsu[1], Shu Shen[2]

The appendix collects proofs of lemmas, theorems, and some other equivalence results stated in the paper.

A Identification

Proof of Equations (2.1) **and** (2.2)**:** First we show the identification of $CLATE$ under Assumption 2.2.

$$\begin{aligned}
&\lim_{z\searrow c} E[Y_i|X_i=x,Z_i=z]-\lim_{z\nearrow c} E[Y_i|X_i=x,Z_i=z]\\
=&\lim_{z\searrow c} E[Y_i(1)T_i+Y_i(0)(1-T_i)|X_i=x,Z_i=z]-\lim_{z\nearrow c} E[Y_i(1)T_i+Y_i(0)(1-T_i)|X_i=x,Z_i=z]\\
=&\lim_{z\searrow c} E[Y_i(1)T_i(1)+Y_i(0)(1-T_i(1))|X_i=x,Z_i=z]\\
&-\lim_{z\nearrow c} E[Y_i(1)T_i(0)+Y_i(0)(1-T_i(0))|X_i=x,Z_i=z]\\
=&E[Y_i(1)-Y_i(0)|X_i=x,Z_i=c,T_i(1)-T_i(0)=1]P[T_i(1)-T_i(0)=1|X_i=x,Z_i=c]\\
=&E[Y_i(1)-Y_i(0)|X_i=x,Z_i=c,T_i(1)-T_i(0)=1]E[T_i(1)-T_i(0)|X_i=x,Z_i=c].
\end{aligned}$$

The first equality holds by the definition of Y_i. The second holds by the definition of T_i. The third holds by the continuity condition in Assumptions 2.2.(i) and (ii).

Further,

$$\begin{aligned}
&\lim_{z\searrow c} E[T_i|X_i=x,Z_i=z]-\lim_{z\nearrow c} E[T_i|X_i=x,Z_i=z]\\
=&\lim_{z\searrow c} E[T_i(1)|X_i=x,Z_i=z]-\lim_{z\nearrow c} E[T_i(0)|X_i=x,Z_i=z]\\
=&E[T_i(1)-T_i(0)|X_i=x,Z_i=c]
\end{aligned}$$

by the definition of T_i and the continuity condition in Assumption 2.2.(ii). Collecting the two results proves the identification of $CLATE$ stated in Equation (2.2).

[1]E-mail: ychsu@econ.sinica.edu.tw. Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University.

[2]E-mail: shushen@ucdavis.edu. Department of Economics, University of California, Davis.

For the unconditional $LATE$, we notice that since $P[T_i(1) - T_i(0) = 1|Z_i = z] = \int P[T_i(1) - T_i(0) = 1|X_i = x, Z_i = z] f_{X|Z}(x|z)dx$, Assumptions 2.2(ii) and (v) together imply that $P[T_i(1) - T_i(0) = 1|Z_i = z]$ is continuous in z on $\mathcal{N}_{\delta,z}(c)$. Similarly, we can show that $P[T_i(1) - T_i(0) = t|Z_i = z]$ is continuous in z on $\mathcal{N}_{\delta,z}(c)$ for $t \in \{0, 1\}$. For the continuity of the conditional mean of the potential outcome variables, we notice that

$$\begin{aligned} & E[Y_i(t)|T_i(1) - T_i(0) = 1, Z_i = z] \\ = & \int E[Y_i(t)|T_i(1) - T_i(0) = 1, X_i = x, Z_i = z] \frac{P[T_i(1) - T_i(0) = 1|X_i = x, Z_i = z]}{P[T_i(1) - T_i(0) = 1|Z_i = z]} f_{X|Z}(x|z)dx. \end{aligned}$$

So by Assumptions 2.2(i), (ii), (v) and the fact that $P[T_i(1) - T_i(0) = 1|Z_i = z]$ is continuous in z on $\mathcal{N}_{\delta,z}(c)$ as we showed above, we know that $E[Y_i(t)|T_i(1) - T_i(0) = 1, Z_i = z]$ is continuous in z on $\mathcal{N}_{\delta,z}(c)$ for $t \in \{0, 1\}$. Similarly, we can show that $E[Y_i(t)|T_i(1) - T_i(0) = t', Z_i = z]$ is continuous in z on $\mathcal{N}_{\delta,z}(c)$ for $t, t' \in \{0, 1\}$.

Given all these continuity conditions, by the same derivations used to show the identification result for $LATE$ except for conditional only on Z_i this time, we obtain the identification of $LATE$ stated in equation 2.2.

Under the sharp RD design, we know that $T_i = 1(Z_i \geq c)$ so $E[T_i(1) - T_i(0)|X_i = x, Z_i = c] = 1$. The above identification results for $CLATE$ and $LATE$ could then be applied to obtain the identification of $CATE$ and ATE. Also, under sharp RD, Assumption 2.2.(ii)-(iv) hold by construction. So Assumption 2.1 is sufficient for the identificaion results stated in equation (2.1). □

Proof of Lemma 3.1: Without loss of generality, we assume that X is a scalar so $d_x = 1$. To show the equivalence of the hypotheses in (3.1) and (3.3), we just need to show the equivalence of the null hypotheses. The direction from (3.1) to (3.3) is straightforward, so we omit the details. To show the other direction, suppose that there is an $x^* \in \mathcal{X}_c$ such that $CATE(x^*) > 0$. Since $CATE(x)$ is continuous in $x \in \mathcal{X}_c$ as is implied by Assumption 2.1, there exists $x_\ell < x_u$ such that $CATE(x) > 0$ for all $x \in [x_\ell, x_u]$. Then there exists $\ell^* = (x^*, r^*) \in \mathcal{L}$ with $[x^*, x^* + r^*] \subseteq [x_\ell, x_u]$ such that $E[g_{\ell^*}(X_i)CATE(X_i)|Z_i = c] > 0$. Therefore, the null in (3.3) implies the null in (3.1). This then completes the proof of the lemma. □

Proof of Lemma 3.2: If we could show that

$$\sqrt{nh}\,(\hat{m}_+(.) - m_+(.)) \Rightarrow \Phi_{h_{2,m+}}(.) \text{ and } \sqrt{nh}\,(\hat{m}_-(.) - m_-(.)) \Rightarrow \Phi_{h_{2,m-}}(.),$$

where $\Phi_{h_{2,m+}}$ and $\Phi_{h_{2,m-}}$ are mean zero Gaussian processes with covariance kernels $h_{2,m+}(\ell_1,\ell_2) = \frac{\int_0^\infty (\vartheta_2 - u\vartheta_1)^2 K^2(u)du}{(\vartheta_2\vartheta_0 - \vartheta_1^2)^2} \frac{\sigma_+^2(\ell_1,\ell_2)}{f_z(c)}$ and $h_{2,m-}(\ell_1,\ell_2) = \frac{\int_0^\infty (\vartheta_2 - u\vartheta_1)^2 K^2(u)du}{(\vartheta_2\vartheta_0 - \vartheta_1^2)^2} \frac{\sigma_-^2(\ell_1,\ell_2)}{f_z(c)}$, respectively, then by the independence of the data, we know that

$$\sqrt{nh}\,(\hat{\nu}(.) - \hat{\nu}(.))\,\sqrt{nh}(\hat{m}_+(.) - m_+(.)) - \sqrt{nh}(\hat{m}_-(.) - m_-(.)) \Rightarrow \Phi_{h_{2,\nu}}(.),$$

uniformly for all $\ell \in \mathcal{L}$. The lemma is then proven.

Recall that

$$\hat{m}_+(\ell) = \frac{\sum_{i=1}^n 1(Z_i \geq c)\cdot K(\frac{Z_i - c}{h})[S_{n,2}^+ - S_{n,1}^+(Z_i - c)]g_\ell(X_i)Y_i}{\sum_{i=1}^n 1(Z_i \geq c)\cdot K(\frac{Z_i - c}{h})[S_{n,2}^+ - S_{n,1}^+(Z_i - c)]} = \sum_{i=1}^n w_{ni}^+ g_\ell(X_i)Y_i,$$

and it is true that

$$\sqrt{nh}(\hat{m}_+(\ell) - m_+(\ell)) = \sqrt{nh}\,(\hat{m}_+(\ell) - E_Z[\hat{m}_+(\ell)]) + \sqrt{nh}\,(E_Z[\hat{m}_+(\ell)] - m_+(\ell))$$

in which E_Z denotes the conditional expectation conditional on sample path $\{Z_1, Z_2, \ldots\}$. By Theorem 4 of Fan and Gijbels (1992), we know that

$$\sqrt{nh}E_Z[\hat{m}_+(\ell) - m_+(\ell)] = O_p(\sqrt{nh^5}) = o_p(1).$$

The first equality holds because the magnitude is proportional to $m_+''(\ell)$ which is equal to $E_Z[g_\ell(X)\cdot(\partial^2\mu_1(x,z)/\partial z\partial z)]$ and $|\partial^2\mu_1(x,z)/\partial z\partial z|$ is assumed to be uniformly bounded on $x \in \mathcal{X}_c$ and $z \in \mathcal{N}_{\delta,z}(c)$. Therefore,

$$\begin{aligned}&\sqrt{nh}(\hat{m}_+(\ell) - m_+(\ell)) \equiv \sqrt{nh}\,(\hat{m}_+(\ell) - E_Z[\hat{m}_+(\ell)]) + o_p(1),\\ &= \sqrt{nh}\sum_{i=1}^n w_{ni}^+\big(g_\ell(X_i)Y_i - E_Z[g_\ell(X_i)Y_i]\big) + o_p(1).\end{aligned}$$

We use the functional central limit theorem (FCLT), Theorem 10.6 of Pollard (1990), to show that

$$\sqrt{nh}\sum_{i=1}^n w_{ni}^+\big(g_\ell(X_i)Y_i - E_Z[g_\ell(X_i)Y_i]\big) \Rightarrow \Phi_{h_{2,m+}}(\ell).$$

Our arguments condition on the sample path of Z_i's and, in other words, w_{ni}^+ can be treated as constants. Define our triangular array as $\{f_{ni}(\ell) : \ \ell \in \mathcal{L},\ i \leq n, n \geq$

1} and $f_{ni}(\ell) = \sqrt{nh}w_{ni}^{+}\big(g_\ell(X_i)Y_i - E_Z[g_\ell(X_i)\cdot Y_i]\big)$. Let the envelope functions be $\{F_{ni}:\ i \le n, n \ge 1\}$ with $F_{ni} = \sqrt{nh}|w_{ni}^{+}|\cdot(|Y_i| + E_Z[|Y_i|])$. Define our empirical process as $\widehat{\Phi}_n^{+}(\ell) = \sum_{i=1}^{n} f_{ni}(\ell)$. First, $\{g_\ell(X):\ \ell \in \mathcal{L}\}$ is a Type I class of functions in Andrews (1994) and by Lemma E1 of Andrews and Shi (2013), $\{f_{ni}(\ell):\ \ell \in \mathcal{L},\ i \le n, n \ge 1\}$ satisfies condition (i) of Theorem 10.6 in Pollard (1990). To show condition (ii), note that

$$
\begin{aligned}
\hat{h}_{2,m+}(\ell_1,\ell_2) &= E_Z[\widehat{\Phi}_n^{+}(\ell_1)\widehat{\Phi}_n^{+}(\ell_2)] = E[f_{ni}(\ell_1)f_{ni}(\ell_2)]\\
&= nh\sum_{i=1}^{n}(w_{ni}^{+})^2\Big(E_Z[g_{\ell_1}(X_i)g_{\ell_2}(X_i)Y_i^2] - E_Z[g_{\ell_1}(X_i)\cdot Y_i]E_Z[g_{\ell_2}(X_i)\cdot Y_i]\Big)\\
&\to h_{2,m+}(\ell_1,\ell_2),
\end{aligned}
$$

where the third equality holds because $f_{ni}(\ell_1)$ and $f_{nj}(\ell_2)$ are mutually independent for $i \ne j$. Then by the arguments of the second part of Theorem 4 of Fan and Gijbels (1992), we can show that $E_Z[\widehat{\Phi}_n^{+}(\ell_1)\widehat{\Phi^{+}}_n(\ell_2)]$ converges to $h_{2,m+}(\ell_1,\ell_2)$. Furthermore, it is true that the convergence result holds uniformly over $\ell_1,\ell_2 \in \mathcal{L}$. Condition (iii) can be shown by the same arguments for condition (ii). To show condition (iv) of Theorem 10.6 in Pollard (1990), note that for any $\epsilon > 0$,

$$
\sum_{i=1}^{n} E_Z[F_{ni}^2\cdot 1(F_{ni} > \epsilon)] \le \sum_{i=1}^{n} E_Z\Big[\frac{F_{ni}^4}{\epsilon^2}\Big] = \epsilon^{-2}(nh)^2\sum_{i=1}^{n}(w_{ni}^{+})^4 E_Z\big[(|Y_i| + E_Z[|Y_i|])^4\big].
$$

The first inequality holds because $1(F_{ni} > \epsilon) \le (F_{ni}/\epsilon)^{\delta}$ for any $\delta > 0$ and we take $\delta = 2$ here. By the same arguments from the second part of Theorem 4 of Fan and Gijbels (1992), we can show that

$$
\epsilon^{-2}(nh)^2\sum_{i=1}^{n}(w_{ni}^{+})^4 E_Z\big[(|Y_i| + E_Z[|Y_i|])^4\big] = \epsilon^{-2}(nh)^2 O_p((nh)^{-3}) = O_p((nh)^{-1}) = o_p(1),
$$

and this implies that condition (iv) holds.

To show condition (v) of Theorem 10.6 in Pollard (1990), note that

$$
\begin{aligned}
\hat{\rho}_{n,m+}^2(\ell_1,\ell_2) =& \sum_{i=1}^{n}(f_{ni}(\ell_1) - f_{ni}(\ell_2))^2 = \sum_{i=1}^{n} f_{ni}^2(\ell_1) - 2\sum_{i=1}^{n} f_{ni}(\ell_1)f_{ni}(\ell_2) + \sum_{i=1}^{n} f_{ni}^2(\ell_2)\\
=& H_{1n}(\ell_1,\ell_1) - 2H_{1n}(\ell_1,\ell_2) + H_{1n}(\ell_2,\ell_2)\\
\to& h_{2,m+}(\ell_1,\ell_1) - 2h_{2,m+}(\ell_1,\ell_2) + h_{2,m+}(\ell_2,\ell_2) \equiv \rho_{m+}^2(\ell_1,\ell_2).
\end{aligned}
$$

Note that similar to condition (ii), the convergence holds uniformly over $\ell_1, \ell_2 \in \mathcal{L}$. This is sufficient for condition (v). Then, by the FCLT of Pollard (1990), we can show that $\sqrt{nh}(\hat{m}_+(\ell) - m_+(\ell)) \Rightarrow \Phi_{h_{2,m+}}(\ell)$. By the same arguments, we can show that $\sqrt{nh}(\hat{m}_-(\ell) - m_-(\ell)) \Rightarrow \Phi_{h_{2,m-}}(\ell)$ and it follows that $\sqrt{nh}(\hat{\nu}(\ell) - \nu(\ell)) = \sqrt{nh}(\hat{m}_+(\ell) - m_+(\ell)) - \sqrt{nh}(\hat{m}_-(\ell) - m_-(\ell)) \Rightarrow \Phi_{h_{2,\nu}}(\ell)$. $\square$

Proof of Lemma 3.3: Recall that $\widehat{\Phi}^u_n(\ell) = \sum_{i=1}^n U_i \cdot \hat{\phi}_{\nu,ni}(\ell)$, where

$$\hat{\phi}_{\nu,ni}(\ell) = \sqrt{nh}\Big(w^+_{ni} \cdot (g_\ell(X_i)Y_i - \hat{m}_+(\ell)) - w^-_{ni} \cdot (g_\ell(X_i)Y_i - \hat{m}_-(\ell))\Big).$$

So it is sufficient for us to show that

$$\widehat{\Phi}^{+,u}_n(.) = \sum_{i=1}^n U_i \cdot \hat{\phi}_{m+,ni}(.) \overset{p}{\Rightarrow} \Phi_{h_{2,m+}}(.), \text{ and}$$
$$\widehat{\Phi}^{-,u}_n(.) = \sum_{i=1}^n U_i \cdot \hat{\phi}_{m-,ni}(.) \overset{p}{\Rightarrow} \Phi_{h_{2,m-}}(.),$$

where $\hat{\phi}_{m+,ni}(\ell) = \sqrt{nh}\Big(w^+_{ni}(g_\ell(X_i)Y_i - \hat{m}_+(\ell))\Big)$ and $\hat{\phi}_{m-,ni}(\ell) = \sqrt{nh}\Big(w^-_{ni}(g_\ell(X_i)Y_i - \hat{m}_-(\ell))\Big)$.

First, it is straightforward to see that the triangular array $\{\hat{f}_{ni}(\ell) = U_i \cdot \hat{\phi}_{m+,ni}(\ell) : \ell \in \mathcal{L},\ i \le n,\ n \ge 1\}$ is manageable with respect to envelope functions $\{\widehat{F}_{ni} = \sqrt{nh}|U_i| \cdot \big(|w^+_{ni}| \cdot (|Y_i| + \overline{|Y|}^+_n)\big) :\ i \le n,\ n \ge 1\}$ in which $\overline{|Y|}^+_n \equiv \sum_{i=1}^n |w^+_{ni}| \cdot |Y_i|$. Define $\hat{h}_{2,m+}(\ell_1, \ell_2) = \sum_{i=1}^n \hat{\phi}_{m+,ni}(\ell_1)\hat{\phi}_{m+,ni}(\ell_2)$. First, by the same argument in (12.24)-(12.26) of Andrews and Shi (2014) and the same argument from the second part of Theorem 4 of Fan and Gijbels (1992), we can show that

$$\sup_{\ell_1,\ell_2 \in \mathcal{L}} |\hat{h}_{2,m+}(\ell_1, \ell_2) - h_{2,m+}(\ell_1, \ell_2)| \overset{p}{\to} 0.$$

Note that this result implies that $\sup_{\ell \in \mathcal{L}} |\hat{\sigma}^2_{\mu,n}(\ell) - \sigma^2_\mu(\ell)| \overset{p}{\to} 0$.

Also, we can show that

$$nh \sum_{i=1}^n \big(|w^+_{ni}| \cdot (|Y_i| + \overline{|Y|}^+_n)\big)^2 \overset{p}{\to} M_1 < \infty,$$
$$n^3h^3 \sum_{i=1}^n \big(|w^+_{ni}| \cdot (|Y_i| + \overline{|Y|}^+_n)\big)^4 \overset{p}{\to} M_2 < \infty,$$

for some positive M_1 and M_2.

Then by the same proof of Theorem 2.1 of Hsu (2016), we can show that $\widehat{\Phi}_n^{+,u}(.) \overset{p}{\Rightarrow} \Phi_{h_{2,m+}}(.)$. These results complete the proof for Lemma 3.3. $\square$

Proof of Lemma 3.4: Without loss of generality we assume that X is a scalar as in Lemma 3.1. It is straightforward to show that the null hypothesis in (3.7) implies the null hypothesis in (3.8), so we omit the details. To show the other direction, suppose that there is $x^* \in \mathcal{X}$ such that $CATE(x^*) \neq 0$ and, without loss of generality, we assume that $CATE(x^*) > 0$. Since $CATE(x)$ is continuous in $x \in \mathcal{X}_c$ as is implied by Assumption 2.1, there exists $x_\ell < x_u$ such that $CATE(x) > 0$ for all $x \in [x_\ell, x_u]$. Then there exists $\ell^* = (x^*, r^*) \in \mathcal{L}$ with $[x^*, x^* + r^*] \subseteq [x_\ell, x_u]$ such that $\nu(\ell^*) = E[g_{\ell^*}(X_i)CATE(X_i)|Z_i = c] > 0$. That is, $\nu(\ell^*) \neq 0$. This completes the proof. $\square$

Proof of Lemma 3.5: Without loss of generality we assume that X is a scalar as in Lemmas 3.1 and 3.4, . It is straightforward to show the null hypothesis in (3.9) implies the null hypothesis in (3.10), so we omit the details. To show the other direction, suppose that there is $x^* \in \mathcal{X}_c$ such that $CATE(x^*) \neq ATE = \nu((0,1))$ and without loss of generality, we assume that $CATE(x^*) > \nu((0,1))$. Since $CATE(x)$ is continuous in $x \in \mathcal{X}_c$, as is implied by Assumption 2.1, there exists $x_\ell < x_u$ such that $CATE(x) > \nu((0,1))$ for all $x \in [x_\ell, x_u]$. Then there exists $\ell^* = (x^*, r^*) \in \mathcal{L}$ with $[x^*, x^* + r^*] \subseteq [x_\ell, x_u]$ such that $E[g_{\ell^*}(X_i)CATE(X_i)|Z_i = c] > E[g_{\ell^*}(X_i)\nu((0,1))|Z_i = c] = \nu((0,1)) \cdot p(\ell)$. That is, $\nu_{hetero,ate}(\ell^*) \neq 0$. This completes the proof. $\square$

B Theorems

Proof of Theorem 3.1: Theorem 3.1 follows from the results of Lemma 3.2 and Lemma 3.3 by the same proof arguments for Proposition 3 of Barrett and Donald (2003). We omit the details for brevity. $\square$

Proofs of Theorems 3.3 and 3.4: Note that the process results and simulated process results for Theorems 3.3 and 3.4 are similar to Lemma 3.2 and Lemma 3.3, so we omit the details for brevity. The proofs for Theorem 3.3 and 3.4 are similar to that for Theorem 3.1. $\square$

C Other Results

Proof of Equation (3.4):

For any $\ell \in \mathcal{L}$, we have

$$\begin{aligned}
\nu(\ell) =& E[g_\ell(X_i) CATE(X_i) | Z_i = c] \\
=& E\big[g_\ell(X_i)\left(E[Y_i(1)|X_i, Z_i = c] - E[Y_i(0)|X_i, Z_i = c]\right)|Z_i = c\big] \\
=& \lim_{z \searrow c} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = z\big] - \lim_{z \nearrow c} E\big[g_\ell(X_i) E[Y_i(0)|X_i, Z_i = c]|Z_i = z\big] \\
=& \lim_{z \searrow c} E\big[g_\ell(X_i) E[Y_i|X_i, Z_i = z]|Z_i = z\big] - \lim_{z \nearrow c} E\big[g_\ell(X_i) E[Y_i|X_i, Z_i = z]|Z_i = z\big] \\
=& \lim_{z \searrow c} E[g_\ell(X_i) Y_i | Z_i = z] - \lim_{z \nearrow c} E[g_\ell(X_i) Y_i | Z_i = z].
\end{aligned}$$

The second equality holds by the definition of $CATE$. The third equality holds given Assumption 2.1.(ii). To be specific, Assumption 2.1.(ii) implies that for all $\epsilon > 0$, there exists $\delta > 0$ such that $\sup_{x \in \mathcal{X}_c} |f_{x|z}(x|z) - f_{x|z}(x|c)| < \epsilon$ for all $|z - c| < \delta$. Then given that $g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]$ is uniformly bounded above, we have for all $|z - c| < \delta$,

$$\Big| E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = z\big] - E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = c\big]\Big| \leq C \cdot \epsilon$$

for some $C > 0$. Same result applies to $Y_i(0)$ case. Therefore, this implies that third equality holds. The fourth equality holds by the continuity of $E[Y_i(1)|X_i, Z_i = c]$ at c. To be specific, let $a_\ell = \lim_{z \searrow c} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = z\big]$. It is true that for any $z_n \searrow c$, $\lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = z_n\big] = a_\ell$. Also, by assumption, we have $\sup_{x \in \mathcal{X}} |E[Y_i(1)|X_i = x, Z_i = c] - E[Y_i(1)|X_i = x, Z_i = z_n]| \leq C|z_n - c|$ for some $C > 0$. Then we have

$$\begin{aligned}
& \lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = c]|Z_i = z_n\big] \\
=& \lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = z_n]|Z_i = z_n\big] \\
& + \lim_{n \to \infty} E\big[g_\ell(X_i)(E[Y_i(1)|X_i, Z_i = c] - E[Y_i(1)|X_i, Z_i = z_n])|Z_i = z_n\big] \\
=& \lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = z_n]|Z_i = z_n\big] + \lim_{n \to \infty} C|z_n - c| \\
=& \lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = z_n]|Z_i = z_n\big] + 0 \\
=& \lim_{n \to \infty} E\big[g_\ell(X_i) E[Y_i(1)|X_i, Z_i = z_n]|Z_i = z_n\big].
\end{aligned}$$

Similar argument applies to $Y_i(0)$ case, so fourth equality holds. The last equality is from the law of iterated expectation. □

Proof of Equation (3.12): Note that

$$
\begin{aligned}
&\sqrt{nh}\left(\hat{\nu}_{hetero,ate}(\ell) - \nu_{hetero,ate}(\ell)\right) \\
=&\sqrt{nh}(\hat{\nu}(\ell) - \hat{\nu}((\mathbf{0},1)) \cdot \hat{p}(\ell) - \nu(\ell) + \nu((\mathbf{0},1)) \cdot p(\ell)) \\
=&\sqrt{nh}(\hat{\nu}(\ell) - \nu(\ell)) - \sqrt{nh}(\hat{\nu}((\mathbf{0},1)) \cdot \hat{p}(\ell) - \nu((\mathbf{0},1)) \cdot p(\ell)) \\
=&\sqrt{nh}(\hat{\nu}(\ell) - \nu(\ell)) - \hat{p}(\ell) \cdot \sqrt{nh}(\hat{\nu}((\mathbf{0},1)) - \nu((\mathbf{0},1))) - \nu((\mathbf{0},1))\sqrt{nh}(\hat{p}(\ell) - p(\ell)) \\
=&\sqrt{nh}(\hat{\nu}(\ell) - \nu(\ell)) - p(\ell) \cdot \sqrt{nh}(\hat{\nu}((\mathbf{0},1)) - \nu((\mathbf{0},1))) + o_p(1) \\
&- \nu((\mathbf{0},1))\sqrt{nh}(\hat{p}(\ell) - p(\ell)) \\
=&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} \phi_{\nu,ni}(\ell) - p(\ell)\phi_{\nu,ni}((\mathbf{0},1)) - \nu((\mathbf{0},1)) \cdot \phi_{p,ni}(\ell) + o_p(1),
\end{aligned}
$$

where the $o_p(1)$ result holds uniformly over all $\ell \in \mathcal{L}$. Note that the second to last equality follows from the uniform consistency of $\hat{p}(\cdot)$, and the last equality is due to the inference function representations in Equations (3.5) and (3.11). This completes the proof. □

Proof of the Equivalence of (4.3) and (4.4): Recall that $H_{0,late}^{hetero}$ in (4.3) is equivalent to: $H_{0,late}^{hetero}:\ CLATE(x) = LATE$ for all $x \in \mathcal{X}_c$ in which

$$
\begin{aligned}
CLATE(x) &= \frac{\lim_{z\searrow c} E[Y_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[Y_i|X_i = x, Z_i = z]}{\lim_{z\searrow c} E[T_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[T_i|X_i = x, Z_i = z]} \\
LATE &= \nu((\mathbf{0},1))/\mu((\mathbf{0},1)).
\end{aligned}
$$

Therefore, $H_{0,late}^{hetero}$ is equivalent to

$$
\begin{aligned}
H_{0,late}^{hetero}:&(\lim_{z\searrow c} E[Y_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[Y_i|X_i = x, Z_i = z]) \cdot \mu((\mathbf{0},1)) \\
&- (\lim_{z\searrow c} E[T_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[T_i|X_i = x, Z_i = z]) \cdot \nu((\mathbf{0},1)) \\
&= 0 \text{ for all } x \in \mathcal{X}_c.
\end{aligned}
$$

It is straightforward to show that the above equality implies the null hypothesis in (4.4), so we omit the details. Next, we show the equality in the other direction.

Denote $\lim_{z\searrow c} E[Y_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[Y_i|X_i = x, Z_i = z]$ as $CARE(x)$, or conditional average reduced-from effect, and $\lim_{z\searrow c} E[T_i|X_i = x, Z_i = z] - \lim_{z\nearrow c} E[T_i|X_i =$

$x, Z_i = z]$) as $CAFE(x)$, or conditional average first-stage effect. Suppose there is an $x^* \in \mathcal{X}_c$ such that $CARE(x^*) \cdot \mu((\mathbf{0}, 1)) - CAFE(x^*) \cdot \nu((\mathbf{0}, 1)) \neq 0$ and, without loss of generality, we assume that $CARE(x^*) \cdot \mu((\mathbf{0}, 1)) - CAFE(x^*) \cdot \nu((\mathbf{0}, 1)) > 0$. Assumptions 2.2.(i) and (ii) imply that both $CARE(x)$ and $CAFE(x)$ are continuous in x for all $x \in \mathcal{X}_c$. Therefore, $CARE(x) \cdot \mu((\mathbf{0}, 1)) - CAFE(x) \cdot \nu((\mathbf{0}, 1))$ is continuous in x and there exists $x_l < x_u$ such that for all $x \in [x_l, x_u]$, $CARE(x) \cdot \mu((\mathbf{0}, 1)) - CAFE(x) \cdot \nu((\mathbf{0}, 1)) > 0$. Then there exists $\ell^* = (x^*, r^*) \in \mathcal{L}$ with $[x^*, x^* + r^*] \subseteq [x_\ell, x_u]$ such that $E[g_{\ell^*}(X_i)\,(CARE(X_i) \cdot \mu((\mathbf{0}, 1)) - CAFE(X_i) \cdot \nu((\mathbf{0}, 1)))\,|Z_i = c] = \nu(\ell) \cdot \mu((\mathbf{0}, 1)) - \mu(\ell) \cdot \nu((\mathbf{0}, 1)) > 0$. This completes the proof. $\square$

References

ANDREWS, D. W. K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, North Holland, vol. 4, 2111–3155.

ANDREWS, D. W. K. AND X. SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

——— (2014): "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, 179, 31–45.

BARRETT, G. F. AND S. G. DONALD (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

FAN, J. AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20(4), 2008–2036.

HSU, Y.-C. (2016): "Multiplier Bootstrap," Tech. rep., Academia Sinica.

POLLARD, D. (1990): "Empirical Processes: Theory and Applications," in *NSF-CBMS regional conference series in probability and statistics*.