Supplement To "Weak Identification in Fuzzy Regression Discontinuity Designs"

Donna Feir* Thomas Lemieux† Vadim Marmer†

June 29, 2015

Abstract

This supplement contains: i) the description of the procedure for selection and evaluation of the influential empirical RD papers; ii) the proofs of Theorem 1, 3, and 4; iii) the Monte Carlo results for standard and weak-identification-robust confidence sets; and iv) the additional tables from the empirical application.

1 Influential applied papers sample procedure

We start with thirty applied papers that were cited by Lee and Lemieux (2010). Of the thirty papers, sixteen did not report enough information to perform the F-test. Of the remaining papers, more than half had specifications which would be suspect according to the test. We reach similar conclusions when only focusing on the ten most cited paper in the list (Pitt and Khandker (1998); Hoxby (2000); Angrist (1990); (Van der Klaauw, 2002); Thistlethwaite and Campbell (1960); Greenstone and Gallagher (2008); Jacob and Lefgren (2004); (Oreopoulos, 2006); Card et al. (2009); and (Kane, 2003)). These papers had between 203 and 888 Google Scholar citations. Four of the ten papers do not report

*Deparment of Economics, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, V8W 2Y2, Canada. Email: dfeir@uvic.ca.

†Vancouver School of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada. E-mails: thomas.lemieux@ubc.ca (Lemieux) and vadim.marmer@ubc.ca (Marmer).

enough information to compute the test, but four of the remaining six papers presented some specifications that failed the test.

2 Proofs of Theorem 1, 3 and 4

Proof of Theorem 1. In what follows, the population parameters should be viewed as drifting sequences indexed by n. Let $d^*_{n,1} = (f_z(z_0)/k)^{1/2}(nh_n)^{1/2}\Delta x/\sigma_x$, so that $d_{n,1} = |d^*_{n,1}|$, and re-write (6) as

$$\frac{\sigma_x}{\sigma_y}\hat{\beta}_n = \frac{(\Delta Y_n/\sigma_y) + d_{n,3}d^*_{n,1}}{(\Delta X_n/\sigma_x) + d^*_{n,1}}.$$

Since $\Delta y = \beta \Delta x$, we can re-write (5) as

$$T_n(\beta) = \sqrt{\frac{\hat{f}_{z,n}(z_0)}{f_z(z_0)}}\frac{(\Delta Y_n/\sigma_y) - (\sigma_x/\sigma_y)\beta(\Delta X_n/\sigma_x)}{\sqrt{\frac{\hat{\sigma}^2_{y,n}}{\sigma^2_y} + \frac{\hat{\sigma}^2_{x,n}}{\sigma^2_x}\left(\frac{\sigma_x}{\sigma_y}\hat{\beta}_n\right)^2 - 2\frac{\hat{\sigma}_{xy,n}}{\sigma_x\sigma_y}\left(\frac{\sigma_x}{\sigma_y}\hat{\beta}_n\right)}} \times \text{sign}\left((\Delta X_n/\sigma_x) + d^*_{n,1}\right).$$

In addition to $d_{n,1}, d_{n,2}, d_{n,3}$, the finite-sample distribution of $T_n(\beta)$ can also be indexed by $d_{n,4} = f_z(z_0)$, where $d_{n,4}$ takes values in a compact set by Assumption 2(b)(i). However, by usual results for kernel estimators and under Assumption 2(a) and part (i) of Assumption 2(b), $(E\hat{f}_{z,n}(z_0) - f_z(z_0))/f_z(z_0) \to 0$ and $Var(\hat{f}_{z,n}(z_0)/f_z(z_0)) \to 0$ (Li and Racine, 2007, Chapter 1), since f_z is bounded away from zero around z_0. It follows that $\hat{f}_{z,n}(z_0)/d_{n,4} \to_p 1$ as $n \to \infty$, and therefore for any subsequence $\{p_n\}$ of $\{n\}$, $\hat{f}_{z,p_n}(z_0)/d_{p_n,4} \to_p 1$. Hence, $d_{n,4}$ does not affect $AsySz$ of $T_n(\beta)$. Next, let $d_{n,5} = \sigma_y$ and $d_{n,6} = \sigma_x$. By Assumption 2(b)(iii), they take values in compact sets, and $\hat{\sigma}_{x,n}/d_{n,6} \to_p 1$, $\hat{\sigma}_{y,n}/d_{n,5} \to_p 1$, and $\hat{\sigma}_{xy,n}/(d_{n,5}d_{n,6}) - d_{n,2} \to_p 0$. Thus,

$$T_n(\beta) = \frac{((\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6}))\,\text{sign}\left((\Delta X_n/d_{n,6}) + d^*_{n,1}\right)}{\sqrt{1 + \left(\frac{(\Delta Y_n/d_{n,5}) + d_{n,3}d^*_{n,1}}{(\Delta X_n/d_{n,6}) + d^*_{n,1}}\right)^2 - 2d_{n,2}\frac{(\Delta Y_n/d_{n,5}) + d_{n,3}d^*_{n,1}}{(\Delta X_n/d_{n,6}) + d^*_{n,1}}}} + o_p(1).$$

Suppose now that for any subsequence $\{p_n\}$ of $\{n\}$,

$$(\Delta Y_{p_n}/d_{p_n,5}, \Delta X_{p_n}/d_{p_n,6}) \to_d (\mathcal{Y}, \mathcal{X}). \tag{S.1}$$

Note that $d_{n,5}$ and $d_{n,6}$ do not affect $AsySz$. Suppose further that $d^*_{p_n,1} \to d^*_1$ for some $|d^*_1| < \infty$, $d_{p_n,2} \to d_2 \in [-\bar{\rho}, \bar{\rho}]$, and $d_{p_n,3} \to d_3$ for some $|d_3| < \infty$. In this case,

$$T_{p_n}(\beta) \to_d \frac{\mathcal{Y} - d_3\mathcal{X}}{\sqrt{1 + \left(\frac{\mathcal{Y}+d_3 d^*_1}{\mathcal{X}+d^*_1}\right)^2 - 2d_2 \frac{\mathcal{Y}+d_3 d^*_1}{\mathcal{X}+d^*_1}}} \times \text{sign}(\mathcal{X} + d^*_1).$$

If $d^*_1 < 0$, one can multiply $\mathcal{Y} + d_3 d^*_1$, $\mathcal{X} + d^*_1$, and $\mathcal{Y} - d_3\mathcal{X}$ each by -1, and re-define $\mathcal{Y}$ and $\mathcal{X}$ as their negatives without changing the resulting asymptotic distribution. Hence, in this case $T_{p_n}(\beta) \to_d \mathcal{T}_{d_1,d_2,d_3}$. Note also that the distribution of $|\mathcal{T}_{d_1,d_2,d_3}|$ is the same as that of $|\mathcal{T}_{d_1,-d_2,-d_3}|$, and therefore, without loss of generality, one can restrict d_2 to $[0, \bar{\rho}]$ for two-sided testing.

Suppose now that $|d^*_1| < \infty$, $d_2 \in [-\bar{\rho}, \bar{\rho}]$, and $d_3 = \pm\infty$. In this case,

$$T_n(\beta) = \frac{((\Delta Y_n/d_{n,5}d_{n,3}) - (\Delta X_n/d_{n,6}))\,\text{sign}\left((\Delta X_n/d_{n,6}) + d^*_{n,1}\right)}{\sqrt{\frac{1}{d^2_{n,3}} + \left(\frac{(\Delta Y_n/d_{n,5}d_{n,3})+d^*_{n,1}}{(\Delta X_n/d_{n,6})+d^*_{n,1}}\right)^2 - \frac{2d_{n,2}}{d_{n,3}} \frac{(\Delta Y_n/d_{n,5}d_{n,3})+d^*_{n,1}}{(\Delta X_n/d_{n,6})+d^*_{n,1}}}} + o_p(1). \tag{S.2}$$

Therefore, $T_{p_n}(\beta) \to_d -\mathcal{X}(\mathcal{X} + d_1)/d_1 =^d \mathcal{T}_{d_1,d_2,\pm\infty}$ for any $d_2 \in [-\bar{\rho}, \bar{\rho}]$.

Next, suppose that $|d^*_1| = \infty$, $d_2 \in [-\bar{\rho}, \bar{\rho}]$, and $|d_3| < \infty$. We have

$$T_n(\beta) = \frac{((\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6}))\,\text{sign}\left((\Delta X_n/d_{n,6}) + d^*_{n,1}\right)}{\sqrt{1 + \left(\frac{(\Delta Y_n/d_{n,5}d^*_{n,1})+d_{n,3}}{(\Delta X_n/d_{n,6}d^*_{n,1})+1}\right)^2 - 2d_{n,2} \frac{(\Delta Y_n/d_{n,5}d^*_{n,1})+d_{n,3}}{(\Delta X_n/d_{n,6}d^*_{n,1})+1}}} + o_p(1),$$

and, therefore, $T_{p_n}(\beta)$ converges in distribution to $(\mathcal{Y} - d_3\mathcal{X})/(1 + d_3^2 - 2d_2 d_3)^{1/2} \times \text{sign}(d^*_1) =^d \mathcal{T}_{\infty,d_2,d_3} \sim N(0,1)$ for any $d_2 \in [-\bar{\rho}, \bar{\rho}]$. The case of $|d^*_1| = \infty$ and $|d_3| = \infty$ can be handled similarly to the previous to cases with $T_{p_n}(\beta) \to_d \mathcal{T}_{\infty,d_2,\pm\infty} \sim N(0,1)$ for any $d_2 \in [-\bar{\rho}, \bar{\rho}]$.

The results of the theorem now follow from Lemma 1 provided that (S.1) holds. To show (S.1), consider $\hat{y}^+_n$ first. As in Hahn et al. (1999, Lemma 2), write

$$\sqrt{nh_n}\begin{pmatrix} \hat{y}^+_n - y^+ \\ h_n(\hat{y}^{(1),+}_n - y^{(1),+}) \end{pmatrix} = \left(\frac{1}{nh_n}\sum_{i=1}^n Z_i Z_i' K_i\right)^{-1}$$

$$\times \left(\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} \xi_{ni} + \sqrt{nh_n} E y_i^* Z_i K_i \right), \quad \text{(S.3)}$$

where $y^{(1),+} = \lim_{e \downarrow 0} dE(y_i|z_i = z_0 + e)/dz_i$, $\hat{y}_n^{(1),+}$ denotes the estimator of $y^{(1),+}$, $Z_i' = (1, (z_i - z_0)/h_n)$, $K_i = K((z_i - z_0)/h_n) 1\{z_i \geq z_0\}$, and $y_i^* = y_i - y^+ - y^{(1),+}(z_i - z_0)/h_n$, and

$$\xi_{ni} = y_i^* Z_i K_i - E y_i^* Z_i K_i.$$

Hahn et al. (1999) show that $E y_i^* Z_i K_i = h_n^2 f_z(z_0)(\lim_{e \downarrow 0} d^2 E(y_i|z_i = z_0 + e)/dz_i^2) \times (k_1 + o(1))$, where k_1 is a vector of constants depending only on $K(\cdot)$. Since $f_z(z)$ and the second derivative of $E(y_i|z_i = z)$ are bounded in the neighborhood of z_0 by Assumption 2(b)(i)-(ii), it follows from Assumption 2(c) that $\sqrt{p_n h_{p_n}} E y_i^* Z_i K_i \to 0$ for all subsequences $\{p_n\}$ of $\{n\}$. Similarly, since the variances are bounded from below by Assumption 2(b)(iii),

$$\left(Var \left(\frac{1}{\sqrt{p_n h_{p_n}}} \sum_{i=1}^{n} \xi_{p_n i} \right) \right)^{-1/2} \sqrt{p_n h_{p_n}} E y_i^* Z_i K_i \to 0. \quad \text{(S.4)}$$

By Lyapounov's CLT (see for example, Lehmann and Romano, 2005, Corollary 11.2.1, p. 427) and the Cramér-Wold device (Davidson, 1994, Theorem 25.5, p. 405),

$$\left(Var \left(\frac{1}{\sqrt{p_n h_{p_n}}} \sum_{i=1}^{n} \xi_{p_n i} \right) \right)^{-1/2} \frac{1}{\sqrt{p_n h_n}} \sum_{i=1}^{n} \xi_{p_n i} \to_d N(0,1), \quad \text{(S.5)}$$

where Lyapounov's condition can be verified by Theorem 23.12 on p. 373 in Davidson (1994) using Assumption 2(b)(iv). Uniform positive definiteness of the variance-covariance matrix, which is needed to apply the Cramér-Wold device, holds because $\sigma_y^2(z_i)$ is bounded away from zero around z_0 by Assumption 2(b)(iii), and by Lemma 4 in Hahn et al. (1999). Let

$$\Omega_n = \left(\frac{1}{nh_n} \sum_{i=1}^{n} Z_i Z_i' K_i \right)^{-1} Var \left(\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} \xi_{ni} \right) \left(\frac{1}{nh_n} \sum_{i=1}^{n} Z_i Z_i' K_i \right)^{-1}.$$

By (S.3)-(S.5),

$$\Omega_{p_n}^{-1/2}\sqrt{p_n h_{p_n}}\begin{pmatrix} \hat{y}_{p_n}^{+} - y^{+} \\ h_{p_n}(\hat{y}_{p_n}^{(1),+} - y^{(1),+}) \end{pmatrix} \to_d N(0,1).$$

Now, by Lemmas 1 and 4 in Hahn et al. (1999) and in view of Assumption 2, we conclude that $\sqrt{p_n h_{p_n}}(\hat{y}_{p_n}^{+} - y^{+})/(\sigma_y^{+}\sqrt{k/f_z(z_0)}) \to_d N(0,1)$, where $\sigma_y^{+} = \lim_{e\downarrow 0}\sigma_y(z_0+e)$.

Let $d_{n,7} = \sigma_y^{+}$, $d_{n,8} = \sigma_y^{-} = \lim_{e\downarrow 0}\sigma_y(z_0 - e)$, $\Delta Y_n^{+} = \sqrt{nh_n/(k/f_z(z_0))}(\hat{y}_n^{+} - y^{+})$, and let ΔY_n^{-} be defined similarly with the plus-terms replaced with the minus-terms. Using the same arguments as above and applying the Cramér-Wold device, we can show that

$$(\Delta Y_{p_n}^{+}/d_{p_n,7}, \Delta Y_{p_n}^{-}/d_{p_n,8}) \to_d \left(\mathcal{Y}^{+}, \mathcal{Y}^{-}\right). \tag{S.6}$$

where $\mathcal{Y}^{+}, \mathcal{Y}^{-}$ are independent standard normal random variables. Next,

$$\frac{\Delta Y_{p_n}}{d_{p_n,5}} = \frac{\Delta Y_{p_n}^{+}}{d_{p_n,7}}\frac{d_{p_n,7}}{d_{p_n,5}} + \frac{\Delta Y_{p_n}^{-}}{d_{p_n,8}}\frac{d_{p_n,8}}{d_{p_n,5}}.$$

Now, $\Delta Y_{p_n}/d_{p_n,5} \to_d \mathcal{Y}$ in (S.1) can be argued using (S.6), Assumption 2(b)(iii), and Lemma 1.

Lastly, the joint convergence in (S.1) can be shown using the same arguments as above in combination with the Cramér-Wold device applied to y- and x-terms. Note that since $|\rho_{xy}|$ is bounded away from one by Assumption 2(b)(iii), the variance-covariance matrices will be positive definite, which ensures that the Cramér-Wold device can be applied. □

Proof of Theorem 3. Again, the population parameters should be viewed as drifting sequences indexed by n. First, note that under H_0, the rejection probability is largest when $\beta = \beta_0$. Next, as in the proof of Theorem 1, we can write

$$T_n^R(\beta) = \frac{((\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6}))\,\mathrm{sign}\left((\Delta X_n/d_{n,6}) + d_{n,1}^{*}\right)}{\sqrt{1 + d_{n,3}^2 - 2d_{n,2}d_{n,3}}} + o_p(1). \tag{S.7}$$

Suppose that $d_{p_n,1}^{*} \to \pm\infty$, and $d_{p_n,3} \to d_3$, where $|d_3| < \infty$. By (S.1) we have that

$T^R_{p_n}(\beta) \to_d N(0,1)$. Next, similarly to (S.7),

$$\hat{Q}_n(\beta) = \frac{\Delta X_n}{d_{n,6}} + d^*_{n,1} - \frac{(d_{n,2} - d_{n,3})\left((\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6})\right)}{1 + d^2_{n,3} - 2d_{n,2}d_{n,3}} + o_p(1),$$

$$\hat{\mathcal{T}}^R_{n,m}(\beta, \hat{Q}_n(\beta)) = \mathcal{S}_m \times \text{sign}\left(\hat{Q}_n(\beta) + \frac{d_{n,2} - d_{n,3}}{\sqrt{1 + d^2_{n,3} - 2d_{n,2}d_{n,3}}}\mathcal{S}_m\right) + o_p(1). \quad \text{(S.8)}$$

We have that $\hat{Q}_{p_n}(\beta)$ diverges to $\pm\infty$, and $\hat{\mathcal{T}}^R_{p_n,m}(\beta, \hat{Q}_{p_n}(\beta)) \to_d N(0,1)$. Hence, it follows that for all subsequences $\{p_n\}$ of $\{n\}$, $\hat{cv}_{p_n,1-\alpha}(\beta_0, \hat{Q}_{p_n}(\beta)) \to_p z_{1-\alpha}$, and

$$P(T^R_{p_n}(\beta) > \hat{cv}_{p_n,1-\alpha}(\beta_0, \hat{Q}_{p_n}(\beta))) \to \alpha. \quad \text{(S.9)}$$

Suppose now that $d^*_{p_n,1} \to d^*_1$, where $|d^*_1| < \infty$, and $d_{p_n,3} \to d_3$, where $|d_3| < \infty$. We can re-write (S.7) as

$$T^R_n(\beta) = \frac{(\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6})}{\sqrt{1 + d^2_{n,3} - 2d_{n,2}d_{n,3}}} \times \text{sign}\left(\hat{Q}_n(\beta) + \frac{(d_{n,2} - d_{n,3})\left((\Delta Y_n/d_{n,5}) - d_{n,3}(\Delta X_n/d_{n,6})\right)}{1 + d^2_{n,3} - 2d_{n,2}d_{n,3}}\right) + o_p(1).$$

By (S.1),

$$\frac{(\Delta Y_{p_n}/d_{p_n,5}) - d_{p_n,3}(\Delta X_{p_n}/d_{p_n,6})}{\sqrt{1 + d^2_{p_n,3} - 2d_{p_n,2}d_{p_n,3}}} \to_d \mathcal{S} = \frac{\mathcal{Y} - d_3\mathcal{X}}{\sqrt{1 + d^2_3 - 2d_2}},$$

$$\hat{Q}_{p_n}(\beta) \to_d \mathcal{Q} + d^*_1, \text{ where}$$

$$\mathcal{Q} = \mathcal{X} - \frac{d_2 - d_3}{\sqrt{1 + d^2_3 - 2d_2}}\mathcal{S}.$$

Note that the two limiting distributions represented by $\mathcal{S}$ and $\mathcal{Q}$ are independent, and $\mathcal{S} \sim N(0,1)$. Hence

$$T^R_{p_n}(\beta) \to_d \mathcal{S} \times \text{sign}\left(\mathcal{Q} + d^*_1 + \frac{d_2 - d_3}{\sqrt{1 + d^2_3 - 2d_2}}\mathcal{S}\right). \quad \text{(S.10)}$$

By (S.8), we have

$$\hat{\mathcal{T}}^R_{p_n,m}(\beta,\hat{Q}_{p_n}(\beta)) \to_d \mathcal{S}_m \times \text{sign}\left(\mathcal{Q} + d_1^* + \frac{d_2 - d_3}{\sqrt{1+d_3^2 - 2d_2d_3}}\mathcal{S}_m\right), \tag{S.11}$$

where $\mathcal{S}_m \sim N(0,1)$ and is independent from $\mathcal{Q}$ by construction. The results in (S.10) and (S.11) imply that (S.9) holds also for $|d_1^*| < \infty$.

Equation (S.9) remains true in the case of $d_{p_n,3} \to \pm\infty$, which can be handled as in the proof of Theorem 1, equation (S.2). The result of the theorem now follows by Lemma 1. □

Proof of Theorem 4. The result in part (a) holds since the concentration parameter does not affect the asymptotic distribution of $G_n(\beta)$. To prove part (b), let

$$G_n^*(\beta_0) = \sum_{j=1}^{J^*} \frac{n_j h_{n_j}\left(\hat{\beta}_n(\bar{w}^j) - \beta_0\right)^2}{k\hat{\sigma}_n^2(\beta_0,\bar{w}^j)/(\hat{f}_{z,n}(z_0|\bar{w}^j)(\widehat{\Delta x}_n(\bar{w}^j))^2)} \le G_n(\beta_0).$$

In the proof below, we allow for $G_n^*(b)$ to be minimized at a set of points or infinity. Let $\Delta x(\bar{w}^j) = x^+(\bar{w}^j) - x^-(\bar{w}^j)$, $\sigma^2(b,\bar{w}^j) = \sigma_y^2(\bar{w}^j) + b^2\sigma_x^2(\bar{w}^j) - 2\sigma_{xy}(\bar{w}^j)$, and

$$G^*(b) = \sum_{j=1}^{J^*} p_j \frac{(\beta(\bar{w}^j) - b)^2(\Delta x(\bar{w}^j))^2}{\sigma^2(b,\bar{w}^j)}\frac{f_z(z_0|\bar{w}^j)}{k}.$$

Since under the theorem's assumptions $\inf_{b\in\mathbb{R}} G^*(b) > 0$, it suffices to show that

$$|\inf_{b\in\mathbb{R}} G_n^*(b)/(nh_n) - \inf_{b\in\mathbb{R}} G^*(b)| \to_p 0, \tag{S.12}$$

as (S.12) implies that $P(\inf_{b\in\mathbb{R}} G_n^*(b) > a) \to 1$ for all $a \in \mathbb{R}$ as $n \to \infty$. However, the last equation can be shown to establishing that

$$\sup_{b\in\mathbb{R}} |G_n^*(b)/(nh_n) - G^*(b)| \to_p 0. \tag{S.13}$$

Since $(\beta(\bar{w}^j) - b)^2$ and $\sigma^2(b,(\bar{w}^j))$ are continuous for all $b \in \mathbb{R}$, and the asymptotic variance-covariance matrix composed of $\sigma_y^2(\bar{w}^j)$, $\sigma_x^2(\bar{w}^j)$, and $\sigma_{xy}(\bar{w}^j)$ is positive definite, it follows that the function $(\beta(\bar{w}^j)-b)^2/\sigma^2(b,\bar{w}^j)$ is continuous for all $b\in\mathbb{R}$ and bounded. By the same arguments,

$$\sup_{b\in\mathbb{R}}\frac{\left(\hat{\beta}_n(\bar{w}^j)-b\right)^2}{\hat{\sigma}_n^2(b,\bar{w}^j)}=O_p(1). \tag{S.14}$$

By triangle inequality,

$$\begin{aligned}|G_n^*(b)/(nh_n)-G^*(b)|\leq&\sum_{j=1}^{J^*}p_j(\Delta x(\bar{w}^j))^2\\ &\times\left|\frac{\left(\hat{\beta}_n(\bar{w}^j)-b\right)^2}{\hat{\sigma}_n^2(b,\bar{w}^j)k/\hat{f}_z(z_0|\bar{w}^j)}-\frac{(\beta(\bar{w}^j)-b)^2}{\sigma^2(b,\bar{w}^j)k/f_z(z_0|\bar{w}^j)}\right|+\sum_{j=1}^{J^*}R_{j,n}(b),\end{aligned} \tag{S.15}$$

where $|R_{j,n}(b)|$ is bounded by

$$\left(\left|\frac{n_jh_{n_j}}{nh_n}-p_j\right|(\Delta x(\bar{w}^j))^2+\left|(\widehat{\Delta x}_n(\bar{w}^j))^2-(\Delta x(\bar{w}^j))^2\right|p_j\right)\left|\frac{\left(\hat{\beta}_n(\bar{w}^j)-b\right)^2}{\hat{\sigma}_n^2(b,\bar{w}^j)k/\hat{f}_z(z_0|\bar{w}^j)}\right|.$$

Since $n_jh_{n_j}/nh_n\to p_j$, it follows from (S.14) that for $j=1,\ldots,J^*$, $\sup_{b\in\mathbb{R}}|R_{j,n}(b)|=o_p(1)$. Similarly, one can show that, for all $j=1,\ldots,J^*$,

$$\sup_{b\in\mathbb{R}}\left|\frac{\left(\hat{\beta}_n(\bar{w}^j)-b\right)^2}{\hat{\sigma}_n^2(b,\bar{w}^j)k/\hat{f}_z(z_0|\bar{w}^j)}-\frac{(\beta(\bar{w}^j)-b)^2}{\sigma^2(b,\bar{w}^j)k/f_z(z_0|\bar{w}^j)}\right|=o_p(1). \tag{S.16}$$

The last result holds since it is assumed that there is strong or semi-strong identification for $j=1,\ldots,J^*$. It also establishes (S.15), which now implies (S.13). □

3 Monte Carlo results for standard and weak-identification-robust confidence sets for FRD designs

In this section, we discuss the performance of standard and robust confidence sets for FRD in Monte Carlo experiments. The model is as in Section 2.1 of the main paper, and specific parametrizations that we use for our simulations are described below. Standard confidence intervals are based on the usual t-statistic for FRD ($T_n(\beta_0)$) and standard normal critical values. Thus, standard two-sided symmetric confidence intervals with asymptotic coverage probability of $1-\alpha$ are constructed as estimator$\pm z_{1-\alpha/2} \times$ std.err, where z_τ denotes the τ-th quantile of $N(0,1)$ distribution, and they correspond to testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. Standard one-sided confidence intervals are constructed as $(-\infty, \text{estimator} - z_\alpha \times \text{std.err}] = (-\infty, \text{estimator} + z_{1-\alpha} \times \text{std.err}]$, and they correspond to testing $H_0 : \beta \geq \beta_0$ against $H_0 : \beta < \beta_0$. Robust confidence sets are constructed as discussed in Section 2.4 of the main paper using robust t-statistic $T_n^R(\beta_0)$. Robust two-sided confidence sets consist of all values β_0 such that $|T_n^R(\beta_0)| < z_{1-\alpha/2}$, and are computed analytically by solving (11) in the main paper.

Robust one-sided confidence sets consist of all values β_0 that satisfy the inequality $T_n^R(\beta_0) > \hat{cv}_{n,1-\alpha}(\beta_0, \hat{Q}_n(\beta_0))$, where $\hat{cv}_{n,1-\alpha}(\beta_0, \hat{Q}_n(\beta_0))$ denotes data-dependent critical values discussed in Section 2.4 of the main paper. In practice one-sided robust confidence sets can be computed numerically by checking the above inequality over a grid of values for β_0. However, to evaluate their coverage probabilities in a Monte Carlo experiment, one can simply compute the relative frequency of occurrence of the event $T_n^R(\beta) > \hat{cv}_{n,1-\alpha}(\beta, \hat{Q}_n(\beta))$, where β is the true value used to generate data.

3.1 Data generating process (DGP)

The outcome variable y_i is generated according to the following model with a constant RD effect:

$$\begin{aligned} y_i &= y_{0i} + x_i\beta, \\ y_{0i} &= g_y(u_{yi}, v_i), \\ x_i &= g_x(z_i, u_{xi}, c), \end{aligned}$$

where u_{yi} and u_{xi} are bivariate normal with correlation parameter κ:

$$\begin{pmatrix} u_{yi} \\ u_{xi} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \kappa \\ \kappa & 1 \end{pmatrix} \right),$$

the assignment variable x_i follows a normal distribution,

$$z_i \sim N(0, \sigma_z^2),$$

and v_i follows a χ^2-distribution with 3 degrees of freedom,

$$v_i \sim \chi_3^2.$$

Note also that $(u_{yi}, u_{xi})'$, z_i, and v_i are jointly independent. In our setup, y_{0i} captures the outcome in the absence of treatment, and the function $g_y(\cdot,\cdot)$ takes one of the following three forms:

$$g_y(u_{yi}, v_i) \quad = \quad \begin{cases} u_{yi}, & y_{0i} \sim N(0,1), \\ \exp(u_{yi}), & y_{0i} \sim \text{Log-normal}, \\ u_{yi}/v_i, & y_{0i} \sim t_3. \end{cases}$$

Thus, the marginal distribution of the outcome without treatment is either standard normal, log-normal, or a t-distribution with 3 degrees of freedom. Log-normal and t- distributions are used to evaluate the effect of deviations from normality: asymmetries in the first case and heavier tails in the second case.

The function $g_x(\cdot,\cdot,\cdot)$ controls when and what kind of treatment is received, and it can take one of the two following forms. In the first case,

$$g_x(z_i, u_{xi}, c) = u_i \times 1\{z_i \leq 0\} + (u_i + c) \times 1\{z_i > 0\}.$$

This is a DGP with continuous treatment variables (thus, there are treatments of different intensity). In this case, $x^+ - x^- = c$, and the concentration parameter (equations (7) and (8) in the main paper) is given by

$$\frac{nh_n c^2}{2k\sqrt{2\pi\sigma_z^2}},$$

since $z_0 = 0$, $f_z(0) = 1/\sqrt{2\pi\sigma_z^2}$, $Var(x_i|z_i) = 1$, so $\sigma_x^2 = 2$. Thus, weaker designs can be generated by reducing the value of $|c|$ or by increasing the value of σ_z^2. Alternatively, treatment assignment can be generated using

$$g_x(z_i, u_{xi}, c) = 1\{u_i \leq 0\} \times 1\{z_i \leq 0\} + 1\{u_i \leq c\} \times 1\{z_i > 0\}.$$

With this DGP, the treatment variable is binary. Let $\Phi(\cdot)$ denote the standard normal CDF. Then, $x^+ - x^- = \Phi(c) - \Phi(0)$, and $\sigma_x^2 = \Phi(0)(1 - \Phi(0)) + \Phi(c)(1 - \Phi(c))$. Hence, the concentration parameter is given by

$$\frac{nh_n(\Phi(c) - \Phi(0))^2}{2k\sqrt{2\pi\sigma_z^2}\,(\Phi(0)(1 - \Phi(0)) + \Phi(c)(1 - \Phi(c)))}.$$

Similarly to the continuous case DGP, the concentration parameter is increasing in c, however, it is now bounded from above for fixed nh_n and σ_z^2, since $\lim_{c\to\infty} \Phi(c) = 1$.

In our DGP, u_{xi} determines whether the treatment is received, and therefore the parameter κ captures the degree of endogeneity of treatment.

Observations are simulated to be independent across i's. The number of Monte Carlo replications is set to 10,000. Our sample size is set to $n = 1,000$. Our base bandwidth value has been chosen as $h_n = n^{-1/4} \approx 0.1778$. We also explore sensitivity of the results to bandwidth choices by also using $h_n = 0.0889$ and 0.8891. We use the uniform kernel function $K(z) = 1/2 \times 1\{-1 \leq z \leq 1\}$, which corresponds to $k = 4$.

We use the following parameter values:

$$\beta = 0,$$

$$\sigma_z = 1,\ 5,$$
$$\kappa = 0.5,\ 0.99,$$
$$c = 2,\ 0.5,\ 0.1$$

where with $c = 2$ identification is considered to be relatively strong, and it becomes weak as c decreases. The values for our endogeneity parameter (κ) are the same as those used in the weak IV literature (Staiger and Stock, 1997). However, since our DGP is non-linear, κ is different from the asymptotic correlation between estimation errors ρ_{xy} (see Section 2.1 of the main paper), where ρ_{xy} is typically smaller in absolute value than κ. Note that ρ_{xy} directly affects asymptotic rejection probabilities.

3.2 Results

First, we consider the effect of weak identification on the distribution of the usual t-statistic $T_n(\beta)$. Figure S.1 shows the densities of $T_n(\beta)$ estimated by kernel smoothing for binary x_i, normal y_{0i}, and $\sigma_z = 1$. As a comparison, we also plot the standard normal density. From panels (a) and (b) constructed using $\kappa = 0.50$ or $\kappa = 0.99$ and $c = 2$ (strong identification), it is apparent that the standard normal distribution is a very good approximation to the distribution of $T_n(\beta)$. When $\kappa = 0.99$, the distribution of $T_n(\beta)$ is slightly skewed to the left. However, the normal approximation should still work reasonably well (as we show below), because there are no substantial deviations of the extreme values of the distribution of $T_n(\beta)$ from those of the standard normal distribution.

Figures S.1 (c) and (d) show the density of $T_n(\beta)$ under very weak identification ($c = 0.1$). In this case, the distribution of $T_n(\beta)$ is very different from normal. It is strongly skewed to the left, although when $\kappa = 0.50$ it is also more concentrated around zero. A consequence of concentration is that there will be no size distortions when identification is weak but the degree of endogeneity is small. The picture changes drastically when $\kappa = 0.99$. The distribution of $T_n(\beta)$ is strongly skewed to the left and no longer concentrated as much around zero. One can expect substantial size distortions in this case for two-sided tests or confidence intervals. Even more severe distortions can be expected for one-sided tests of $H_0 : \beta \geq \beta_0$ against $H_1 : \beta < \beta_0$. However, there will be no size distortions for tests of $H_0 : \beta \leq \beta_0$ against $H_1 : \beta > \beta_0$, since the probability mass is shifted to the left. Similarly, one can expect that the coverage probability of one-sided confidence intervals of the form $(-\infty, \text{estimator} + z_{1-\alpha} \times \text{std.err}]$ will be below the nominal coverage of $1-\alpha$. The discrepancy between the actual and nominal coverage for such intervals would be even larger than for two-sided confidence intervals. At the same time, one can expect that the actual coverage of one-sided confidence intervals of the form $[\text{estimator} - z_{1-\alpha} \times \text{std.err}, \infty)$ will exceed the nominal coverage.

Simulated coverage probabilities for different combinations of the model's parameters are reported in Table S.1. With moderate degree of endogeneity ($\kappa = 0.5$) and when identification is relatively strong (the concentration parameter is around 35 or 7), the usual confidence intervals, two-sided and one-sided, have coverage probabilities very close to the nominal ones. Their coverage probabilities remain very close to nominal even when the concentration parameter drops to very small values (0.09 and 0.02), as long as endogeneity remains moderate.

When the degree of endogeneity is very high ($\kappa = 0.99$), the coverage probabilities of the standard confidence intervals deviate from the nominal levels. Even with a large value of the concentration parameter of 35, the simulated coverage of one-sided intervals can be below the nominal level by 5% (while two-sided intervals remain quite accurate). For a concentration parameter around 7, distortions can be up to 10% for one-sided intervals and 5% for two-sided intervals. The situation becomes substantially worse when identification is very weak ($c = 0.1$ and the corresponding values of the concentration parameter below 0.1). In this case we observe severe size distortions for one-sided and two-sided interval. For example, the actual coverage probabilities of the 90%, 95% and 99% two-sided confidence intervals are approximately 51%, 55%, and 62%, respectively when the concentration parameter is around 0.09. More substantial size distortions were observed with the concentration parameter equal to 0.009: the actual coverages of the 90%, 95% and 99% two-sided confidence intervals were

32%, 35%, and 41%, respectively.

As expected, the performance of one-sided intervals was even worse due to the skewness of the distribution of $T_n(\beta)$ in the case of weak identification and strong endogeneity (Figure S.1(d)). For example, when $c = 0.1$ and $\kappa = 0.99$, the actual coverage of the 90% one-sided confidence intervals does not exceed 46%, and it goes as low as 28% in the case of the concentration parameter equal to 0.009.

The last rows of Table S.1 show coverage when identification is still weak, but less so ($c = 0.5$), which results in the concentration parameter values of 2.22, 0.44, or 0.22 depending on σ_z and the bandwidth. Still, distortions remain serious when κ is large. For example, the actual coverage for two-sided intervals, when nominal coverage is set to 90%, ranges from 83% when the concentration parameter is 2.22, to 55% when it is 0.22.

Table S.1 makes clear that, as long as the concentration parameter is similar, coverage will be similar even if some of the primitive parameters such as the bandwidth and σ_z differ. For example, compare the seventh row of Table S.1 with the tenth row. In the seventh row, the bandwidth is 0.1778, and σ_z is equal to one, which corresponds to the concentration parameter of approximately 0.09. In the tenth row, σ_z is five, but the bandwidth has been increased five times leaving the concentration parameter unchanged. The actual coverage probabilities of the standard confidence intervals are equal in both cases.[1]

Table S.1 also clearly demonstrates how the magnitude of the concentration parameter relates to the degree of distortions observed. When the concentration parameter is relatively large as in the first and fifth rows of Table S.1, the coverage probabilities are close to the nominal ones. On the other hand, the closer to zero the concentration parameter is, the more severe the distortions. It is important to note, however, that the degree of endogeneity is kept the same. Even with equivalent concentration parameters, if there is a higher degree of endogeneity then distortions will be more severe. We can see this by comparing the third and seventh rows.

We have also computed the simulated coverage probabilities of the weak identification

[1] Note that in each experiment (i.e. for each combination of parameters), we used the same sequence of primitive random variables by controlling the random numbers generator.

robust confidence sets. We find that regardless of the strength of identification and degree of endogeneity, the simulated coverage probabilities of two-sided and one-sided robust confidence sets are uniformly very close to the nominal coverage probabilities. This supports our claim that the inference based on the null-restricted statistic $T_n^R(\beta_0)$ does not suffer from size distortions.

Table S.2 repeats the exercises presented in Table S.1 when the treatment variable is binary rather than continuous. The distortions observed in this case are less severe under all specifications. The reason is that κ does not map exactly to the asymptotic correlation between the estimation errors ρ_{xy}, which controls directly asymptotic rejection probabilities (see Section 2.2 of the main paper). This is due to the non-linear nature of the DGP. For example, when $\kappa = 0.99$ and $\sigma_z = 5$, the implied correlation between $\widehat{\Delta y}_n$ and $\widehat{\Delta x}_n$ is approximately 0.80. Under our standard choice of bandwidth, the coverage probabilities of the 90%, 95%, and 99% confidence sets are 78.7%, 84.9%, and 92.2%, respectively.

Table S.3 presents the results for non-normal DGPs. Again, the distortions appear less severe than those in Table S.1 due to the non-linear mapping of κ to the degree of endogeneity. We report the implied value of ρ_{xy} in Table S.3. Nevertheless, the table demonstrates that even when the endogeneity is not as severe and the system is non-normal, size distortions are still present.

In the main body of the paper, we have discussed the possibility that the robust confidence sets may cover the entire real line or be the union of two half-lines. Table S.4 demonstrates the likelihood of this possibility under different scenarios. When identification is relatively strong ($c = 2$), the form of the robust confidence sets complies to the standard one. By contrast, the shape of weak-identification-robust confidence sets is non-standard when identification becomes weaker. As reported in Table S.4, in the case of a relatively strong FRD ($c = 2$), the probabilities that the robust confidence sets are unbounded are very small regardless of the value of κ, and even negligible in the case of continuous treatment. For binary treatment, the probability of seeing unbounded confidence sets when $c = 2$ varies between 0.0007 and 0.11 depending on the nominal coverage and the value of κ. On the other hand, in the case of a weak FRD ($c = 0.1$ or $c = 0.5$), unbounded robust confidence sets are obtained with very high probabilities. For example, in the case of continuous treatment and when $c = 0.1$ and $\kappa = 0.99$, the entire real line is obtained with probabilities 23%, 35% and 60% and the union of two half lines with probabilities 65%, 60% and 39% for the confidence sets with nominal coverage of 90%, 95% and 99% respectively.

4 Monte Carlo results for the test of constancy of the RD effect

In this section, we present the simulated size and power of the standard and weak-identification-robust constancy tests. As discussed in Section 3 of the main paper, it is assumed that in addition to y_i, x_i, z_i, the econometrician observes the covariate variable w_i that takes values in $\mathcal{W} = \{\bar{w}^1, \ldots, \bar{w}^J\}$. The econometrician is interested in testing the null hypothesis that the RD effect is independent of w_i. We maintain the same basic design as in Section 3 with normally distributed outcomes in the absence of treatment and continuous treatment variables. However, we now consider three population sub-groups ($J = 3$) with $n_j = 1,000$ for all $j = 1, 2, 3$. We use the uniform kernel for all sub-groups, select the bandwidth according to $h_{n_j} = n_j^{-1/4}$, and use $\sigma_z = 1$ and $\kappa = 0.99$ in all simulations for all categories of w_i. Under H_0 of constancy, we generate data with $\beta_j = 0$ for all $j = 1, 2, 3$. Under the alternative of heterogenous treatment effects, we generate data with $\beta_1 = 0$, $\beta_2 = -1$ and $\beta_3 = 1$, or $\beta_2 = -3$ and $\beta_3 = 3$.

The standard constancy test can be constructed along the lines of the ANOVA F-test. See for example, Casella and Berger (2002, Chapter 11). Using the notation of Section 3 of the main paper, let

$$CB_n = \sum_{j=1}^{J} \left(\hat{\beta}_n(\bar{w}^j) - \bar{\beta}_n\right)^2 / \hat{V}_n(\bar{w}^j), \text{ where}$$

$$\bar{\beta}_n = \frac{\sum_{j=1}^{J} \hat{\beta}_n(\bar{w}^j)/\hat{V}_n(\bar{w}^j)}{\sum_{j=1}^{J} 1/\hat{V}_n(\bar{w}^j)}, \text{ and}$$

$$\hat{V}_n(\bar{w}^j) \quad = \quad \frac{1}{n_j h_{n_j}} \frac{k\hat{\sigma}_n^2(\beta_0, \bar{w}^j)}{\hat{f}_{z,n}(z_0|\bar{w}^j)(\widehat{\Delta x}_n(\bar{w}^j))^2}.$$

Under H_0 of constancy of the RD effect across the covariate's values and under strong identification for all J categories, $CB_n \to_d \chi^2_{J-1}$. Thus, the standard test of constancy will reject H_0 when$CB_n > \chi^2_{J-1,1-\alpha}$. Since the standard test relies on the asymptotic normal approximation for the FRD estimator, one can expect that it will be distorted when identification is weak.

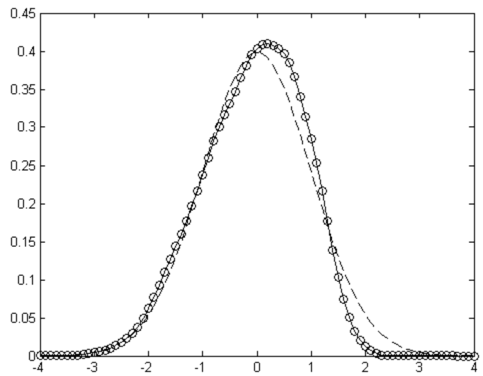
Weak-identification-robust constancy test is proposed in Section 3 of the main paper. The robust statistic $G_n(\beta_0)$ is evaluated over a grid of values for β_0 that covers the interval $[-10, 10]$. The null hypothesis of constancy is rejected by the robust test if the smallest value of $G_n(\beta_0)$ obtained on the grid exceeds $\chi^2_{J,1-\alpha}$. Our theoretical results predict that the robust test has accurate size regardless of the strength of identification, and good power if at least two categories with different RD effects have sufficiently strong identification.

Table S.5 reports simulated rejection probabilities for the standard and robust tests when the treatment effect is strongly identified for all sub-groups and when it is weakly identified for some sub-groups. As one can see from the first row of Table S.5, which reports the results under H_0, the standard test is under-sized when identification is relatively strong for all three categories, while the robust test has rejection probabilities very close to the desired significance level.

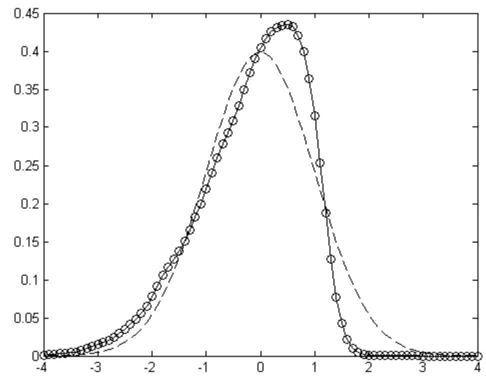
When the treatment effect for one of the groups is only weakly identified as shown in the second row of Table S.5, the standard test over-rejects the null hypothesis of equality: the simulated rejection probabilities of the standard test are equal to 43%, 38%, and 31% for the significance levels of 10%, 5%, and 1% respectively. The rejection probabilities for the robust test remain very close to the corresponding significance levels.

When the treatment effect differs between the groups, rows three and four of Table S.5 demonstrate that the robust test has reasonable power to reject the null hypothesis of constancy even when the treatment effect is weakly identified for one or more groups. For example, when the treatment effect is 0, -1, and 1 from groups 1, 2 and 3 respectively, the robust test rejects the null hypothesis 80%, 70% and 47% of the time for a 10%, 5% and 1% test. When the difference in the treatment effect between the groups is greater, 0, -3 and 3 for example, the robust test rejects the null hypothesis of equality nearly 100% of the time. Rows five and six of Table S.5 demonstrate that similar results hold when weak identification is a problem for more than one of the sub-groups. The last row of the table shows that, when identification is strong for all groups, the two tests have comparable power.

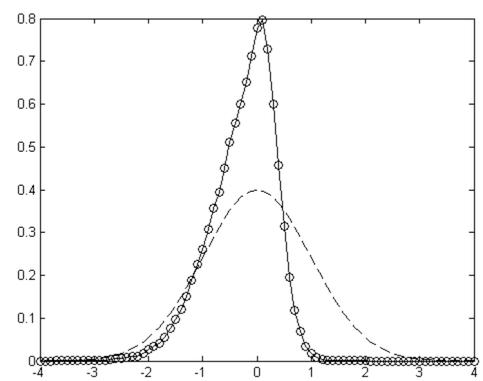
Figure S.1: Kernel estimated density of the usual T statistic (solid line) under strong (c = 2) and weak (c = 0.1) identification for different values of the endogeneity parameter against the standard normal PDF (dashed line)



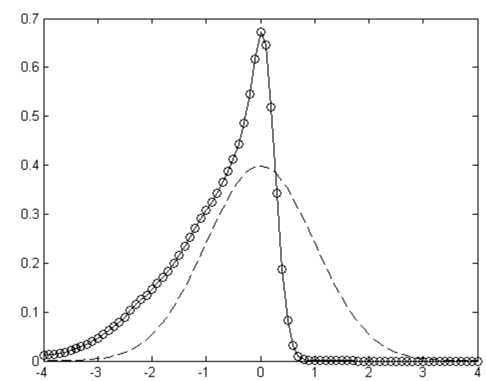
(a) Strong identification, $\kappa = 0.50$



(b) Strong identification, $\kappa = 0.99$



(c) Weak identification, $\kappa = 0.50$



(d) Weak identification, $\kappa = 0.99$

Table S.1: (Continuous x_i & normal y_{0i}) Simulated coverage probabilities of standard and weak-identification-robust confidence sets for different values of the standard deviation of the assignment variable (σ_z), size of discontinuity in treatment assignment (c), degree of endogeneity (κ), and bandwidth (h_n).

						Simulated Coverage			
						Two-sided		One-sided	
σ_z	c	κ	Concentration Parameter	h_n	Nominal Coverage	Standard	Robust	Standard	Robust
					0.90	0.9180	0.9051	0.8784	0.8970
1	2	0.50	35.48	0.1778	0.95	0.9583	0.9522	0.9320	0.9505
					0.99	0.9885	0.9931	0.9793	0.9907
					0.90	0.9262	0.9157	0.8713	0.9025
5	2	0.50	7.10	0.1778	0.95	0.9604	0.9693	0.9273	0.9585
					0.99	0.9896	0.9984	0.9820	0.9969
					0.90	0.9600	0.9051	0.9073	0.8970
1	0.1	0.50	0.09	0.1778	0.95	0.9815	0.9522	0.9600	0.9505
					0.99	0.9969	0.9931	0.9938	0.9907
					0.90	0.9403	0.9157	0.8824	0.9025
5	0.1	0.50	0.02	0.1778	0.95	0.9659	0.9693	0.9404	0.9585
					0.99	0.9892	0.9984	0.9831	0.9969
					0.90	0.9072	0.9051	0.8580	0.8970
1	2	0.99	35.48	0.1778	0.95	0.9393	0.9522	0.9072	0.9505
					0.99	0.9722	0.9931	0.9629	0.9907
					0.90	0.8676	0.9157	0.8189	0.9025
5	2	0.99	7.10	0.1778	0.95	0.9011	0.9693	0.8676	0.9585
					0.99	0.9443	0.9984	0.9299	0.9969
					0.90	0.5165	0.9051	0.4625	0.8970
1	0.1	0.99	0.09	0.1778	0.95	0.5573	0.9522	0.5165	0.9505
					0.99	0.6203	0.9931	0.5949	0.9907
					0.90	0.3741	0.9157	0.3264	0.9025
5	0.1	0.99	0.02	0.1778	0.95	0.4113	0.9693	0.3741	0.9585
					0.99	0.4718	0.9984	0.4475	0.9969
					0.90	0.3203	0.9289	0.2829	0.8999
5	0.1	0.99	0.009	0.0889	0.95	0.3521	0.9670	0.3203	0.9569
					0.99	0.4139	0.9847	0.3902	0.9845
					0.90	0.5165	0.9051	0.4625	0.8970
5	0.1	0.99	0.09	0.8891	0.95	0.5573	0.9522	0.5165	0.9505
					0.99	0.6203	0.9931	0.5949	0.9907

Table S.1: (Continued)

						Simulated Coverage			
						Two-sided		One-sided	
σ_z	c	κ	Concentration Parameter	h_n	Nominal Coverage	Standard	Robust	Standard	Robust
					0.90	0.9367	0.9051	0.8828	0.8970
1	0.5	0.50	2.22	0.1778	0.95	0.9656	0.9522	0.9367	0.9505
					0.99	0.9917	0.9931	0.9845	0.9907
					0.90	0.9362	0.9157	0.8780	0.9025
5	0.5	0.50	0.44	0.1778	0.95	0.9646	0.9693	0.9365	0.9585
					0.99	0.9895	0.9984	0.9823	0.9969
					0.90	0.8287	0.9051	0.7843	0.8970
1	0.5	0.99	2.22	0.1778	0.95	0.8617	0.9522	0.8287	0.9505
					0.99	0.9066	0.9931	0.8908	0.9907
					0.90	0.6783	0.9157	0.6183	0.9025
5	0.5	0.99	0.44	0.1778	0.95	0.7195	0.9693	0.6783	0.9585
					0.99	0.7812	0.9984	0.7585	0.9969
					0.90	0.5535	0.9289	0.4983	0.8999
5	0.5	0.99	0.22	0.0889	0.95	0.5946	0.9670	0.5535	0.9569
					0.99	0.6579	0.9847	0.6347	0.9845
					0.90	0.8287	0.9051	0.7843	0.8970
5	0.5	0.99	2.22	0.8891	0.95	0.8617	0.9522	0.8287	0.9505
					0.99	0.9066	0.9931	0.8908	0.9907

Table S.2: (Binary x_i & normal y_{0i}) Simulated coverage probabilities of standard and weak-identification-robust confidence sets for different values of the standard deviation of the assignment variable (σ_z), size of discontinuity in treatment assignment (c), degree of endogeneity (κ), and bandwidth (h_n).

						Simulated Coverage			
						Two-sided		One-sided	
σ_z	c	κ	Concentration Parameter	h_n	Nominal Coverage	Standard	Robust	Standard	Robust
1	2	0.50	7.42	0.1778	0.90	0.9376	0.9051	0.8925	0.8970
					0.95	0.9745	0.9522	0.9472	0.9505
					0.99	0.9961	0.9931	0.9912	0.9907
5	2	0.50	1.48	0.1778	0.90	0.9606	0.9157	0.9120	0.9025
					0.95	0.9838	0.9693	0.9625	0.9585
					0.99	0.9966	0.9984	0.9932	0.9969
1	0.1	0.50	0.03	0.1778	0.90	0.9783	0.9051	0.9394	0.8970
					0.95	0.9913	0.9522	0.9783	0.9505
					0.99	0.9987	0.9931	0.9972	0.9907
5	0.1	0.50	0.006	0.1778	0.90	0.9627	0.9157	0.9184	0.9025
					0.95	0.9816	0.9693	0.9628	0.9585
					0.99	0.9942	0.9984	0.9913	0.9969
1	2	0.99	7.42	0.1778	0.90	0.9206	0.9051	0.8704	0.8970
					0.95	0.9546	0.9522	0.9211	0.9505
					0.99	0.9857	0.9931	0.9758	0.9907
5	2	0.99	1.48	0.1778	0.90	0.9141	0.9157	0.8514	0.9025
					0.95	0.9493	0.9693	0.9141	0.9585
					0.99	0.9839	0.9984	0.9750	0.9969
1	0.1	0.99	0.03	0.1778	0.90	0.8201	0.9051	0.7382	0.8970
					0.95	0.8732	0.9522	0.8201	0.9505
					0.99	0.9448	0.9931	0.9204	0.9907
5	0.1	0.99	0.006	0.1778	0.90	0.7872	0.9157	0.7057	0.9025
					0.95	0.8495	0.9693	0.7872	0.9585
					0.99	0.9226	0.9984	0.8976	0.9969
5	0.1	0.99	0.003	0.0889	0.90	0.7325	0.9287	0.6487	0.8997
					0.95	0.7868	0.9668	0.7327	0.9567
					0.99	0.8654	0.9845	0.8385	0.9843
5	0.1	0.99	0.03	0.8891	0.90	0.8201	0.9051	0.7382	0.8970
					0.95	0.8732	0.9522	0.8201	0.9505
					0.99	0.9448	0.9931	0.9204	0.9907

Table S.3: (Non-normal y_{0i}) Simulated coverage probabilities of standard and weak-identification-robust confidence sets for different values of the standard deviation of the assignment variable (σ_z), size of discontinuity in treatment assignment (c), degree of endogeneity (κ), correlation between estimation errors for y and x (ρ_{xy}), and different distributions of the outcome without treatment (y_{0i}). Bandwidth is equal to 0.1778.

							Simulated Coverage			
							Two-sided		One-sided	
σ_z	c	Distribution of y_{0i}	κ	ρ_{xy}	Concentration Parameter	Nominal Coverage	Standard	Robust	Standard	Robust
continuous treatment x_i										
5	0.1	Log-normal	0.99	0.85	0.02	0.90	0.7189	0.9207	0.6359	0.9025
						0.95	0.7788	0.9769	0.7189	0.9609
						0.99	0.8643	0.9996	0.8321	0.9989
5	0.1	t_3	0.99	0.64	0.02	0.90	0.8774	0.9267	0.8067	0.9015
						0.95	0.9194	0.9785	0.8774	0.9638
						0.99	0.9663	0.9991	0.9521	0.9987
5	0.5	Log-normal	0.99	0.85	0.44	0.90	0.7701	0.9207	0.6949	0.9025
						0.95	0.8193	0.9769	0.7701	0.9609
						0.99	0.8995	0.9996	0.8725	0.9989
5	0.5	t_3	0.99	0.64	0.44	0.90	0.8837	0.9290	0.8166	0.9057
						0.95	0.9245	0.9814	0.8840	0.9636
						0.99	0.9686	0.9992	0.9556	0.9985
binary treatment x_i										
5	0.1	Log-normal	0.99	-0.6435	0.006	0.90	0.8861	0.9207	0.8106	0.9025
						0.95	0.9255	0.9769	0.8861	0.9609
						0.99	0.9716	0.9996	0.9575	0.9989
5	0.1	t_3	0.99	-0.5327	0.006	0.90	0.9327	0.9320	0.8735	0.9051
						0.95	0.9627	0.9812	0.9327	0.9652
						0.99	0.9877	0.9992	0.9801	0.9978

Table S.4: Simulated probabilities for weak-identification-robust confidence sets to be unbounded for different values of the discontinuity parameter (c), different degrees of endogeneity (κ), and different types of treatment variable x_i. The bandwidth is set to 0.1778, assignment variable is standard normal, and the outcome without treatment is normal.

c	κ	nominal coverage	entire real line	two half-lines
continuous treatment x_i				
2	0.50	0.90	0	0.0002
		0.95	0	0.0002
		0.99	0	0.0020
2	0.99	0.90	0	0.0001
		0.95	0	0.0001
		0.99	0	0.0008
0.1	0.50	0.90	0.7282	0.1614
		0.95	0.8494	0.0973
		0.99	0.9681	0.0221
0.1	0.99	0.90	0.2354	0.6544
		0.95	0.3534	0.5909
		0.99	0.5972	0.3908
0.5	0.50	0.90	0.3572	0.2148
		0.95	0.4894	0.1980
		0.99	0.7348	0.1324
0.5	0.99	0.90	0	0.5637
		0.95	0	0.6847
		0.99	0	0.8626
binary treatment x_i				
2	0.50	0.90	0.0061	0.0139
		0.95	0.0183	0.0222
		0.99	0.0630	0.0551
2	0.99	0.90	0.0007	0.0196
		0.95	0.0021	0.0402
		0.99	0.0134	0.1103
0.1	0.50	0.90	0.7348	0.1592
		0.95	0.8532	0.0945
		0.99	0.9689	0.0217
0.1	0.99	0.90	0.7218	0.1745
		0.95	0.8385	0.1077
		0.99	0.9619	0.0282

Table S.5: Simulated size and power of the standard and weak-identification-robust tests for constancy of the RD effect across covariates. There are three groups with RD effects β_j and discontinuities in treatment assignments c_j.

Size of Discontinuity			Treatment Effect			Nominal	Rejection probabilities	
c_1	c_2	c_3	β_1	β_2	β_3	Size	Standard	Robust
						0.10	0.0309	0.0858
2	2	2	0	0	0	0.05	0.0070	0.0372
						0.01	0	0.0060
						0.10	0.4310	0.0745
2	2	0.1	0	0	0	0.05	0.3864	0.0335
						0.01	0.3171	0.0047
						0.10	0.9739	0.7961
2	2	0.1	0	-1	1	0.05	0.9615	0.6978
						0.01	0.9238	0.4660
						0.10	1	0.9840
2	2	0.1	0	-3	3	0.05	1	0.9684
						0.01	1	0.8946
						0.10	0.6581	0.0676
2	0.1	0.1	0	0	0	0.05	0.6104	0.0312
						0.01	0.5009	0.0043
						0.10	0.6068	0.4059
2	0.1	0.1	0	-1	1	0.05	0.5794	0.2709
						0.01	0.5188	0.0932
						0.10	0.9531	0.9958
2	2	2	0	-1	1	0.05	0.9239	0.9921
						0.01	0.8251	0.9656

5 Empirical Applications: Additional Tables

Table S.6: Angrist and Lavy (1999): Test of equality of RD effect across groups at 5% significance level for different values of the bandwidth

bandwidth	estimated effect		reject H_0 of equality? robust	reject H_0 of equality? standard
	religious	secular		
6	−0.0524	−0.1131	no	no
8	−0.0540	−0.0985	no	no
10	−0.0381	−0.0756	no	no
12	−0.0170	−0.0364	no	no
14	−0.0274	−0.0363	no	no
16	−0.0035	−0.0382	no	no
18	0.0052	−0.0505	yes	no
20	0.0107	−0.0523	yes	no
	$<=$ 10% disadvantaged	$>$ 10% disadvantaged		
6	−0.0390	−0.0909	no	no
8	−0.0626	−0.0469	no	no
10	−0.0387	−0.0488	no	no
12	−0.0259	−0.0192	no	no
14	−0.0343	−0.0226	no	no
16	−0.0290	−0.0079	no	no
18	−0.0368	−0.0037	no	no
20	−0.0360	−0.0008	yes	no

Table S.7: Urquiola and Verhoogen (2009): Estimated discontinuity in the treatment variable for the first cutoff and F-statistics for testing for potential size distortions for various values of the bandwidth

bandwidth	discontinuity estimates	F-statistic
6	1.388	0.8226
8	−0.387	0.0812
10	−3.107	6.8069
12	−4.779	20.684
14	−6.092	41.037
16	−7.870	84.236
18	−8.934	129.80
20	−9.968	188.43

Note: Silverman's normal rule-of-thumb is only 8.59 and the optimal bandwidth suggested by Imbens and Kalyanaraman (2012) is 9.67. The scores are given in terms of standard deviations from the mean.

References

Angrist, J. D. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 313–336.

Card, D., Dobkin, C., and Maestas, N. (2009), "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124, 597–636.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, Duxbury Press, Pacific Grove, CA, 2nd ed.

Davidson, J. (1994), *Stochastic Limit Theory*, New York: Oxford University Press.

Greenstone, M. and Gallagher, J. (2008), "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *Quarterly Journal of Economics*, 123, 951–1003.

Hahn, J., Todd, P., and Van der Klaauw, W. (1999), "Evaluating the Effect of an Antidiscrimination Law Using a Regresion-Discontinuity Design," NBER Working Paper 7131.

Hoxby, C. M. (2000), "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285.

Imbens, G. W. and Kalyanaraman, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959.

Jacob, B. and Lefgren, L. (2004), "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *Review of Economics and Statistics*, 86, 226–244.

Kane, T. J. (2003), "A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going," NBER Working Paper 9703.

Lee, D. S. and Lemieux, T. (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Lehmann, E. L. and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer, 3rd ed.

Li, Q. and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, New Jersey: Princeton University Press.

Oreopoulos, P. (2006), "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter," *American Economic Review*, 152–175.

Pitt, M. M. and Khandker, S. R. (1998), "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy*, 106, 958–996.

Staiger, D. and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586.

Thistlethwaite, D. L. and Campbell, D. T. (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51, 309–317.

Van der Klaauw, W. (2002), "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach," *International Economic Review*, 43, 1249–1287.