




Weak Identification in Fuzzy Regression Discontinuity Designs

Donna Feir, Thomas Lemieux & Vadim Marmer

To cite this article: Donna Feir, Thomas Lemieux & Vadim Marmer (2016) Weak Identification in Fuzzy Regression Discontinuity Designs, Journal of Business & Economic Statistics, 34:2, 185-196, DOI: [10.1080/07350015.2015.1024836](https://doi.org/10.1080/07350015.2015.1024836)

To link to this article: <http://dx.doi.org/10.1080/07350015.2015.1024836>

 View supplementary material [↗](#)

 Accepted author version posted online: 02 Apr 2015.
Published online: 17 Mar 2016.

 Submit your article to this journal [↗](#)

 Article views: 76

 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 2 View citing articles [↗](#)

Weak Identification in Fuzzy Regression Discontinuity Designs

Donna FEIR

Department of Economics, University of Victoria, Victoria, BC, V8W 2Y2, Canada
(dfeir@uvic.ca)

Thomas LEMIEUX and Vadim MARMER

Vancouver School of Economics, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada
(thomas.lemieux@ubc.ca; vadim.marmar@ubc.ca)

In fuzzy regression discontinuity (FRD) designs, the treatment effect is identified through a discontinuity in the conditional probability of treatment assignment. We show that when identification is weak (i.e., when the discontinuity is of a small magnitude), the usual t -test based on the FRD estimator and its standard error suffers from asymptotic size distortions as in a standard instrumental variables setting. This problem can be especially severe in the FRD setting since only observations close to the discontinuity are useful for estimating the treatment effect. To eliminate those size distortions, we propose a modified t -statistic that uses a null-restricted version of the standard error of the FRD estimator. Simple and asymptotically valid confidence sets for the treatment effect can be also constructed using this null-restricted standard error. An extension to testing for constancy of the regression discontinuity effect across covariates is also discussed. Supplementary materials for this article are available online.

KEY WORDS: Nonparametric inference; Regression discontinuity design; Treatment effect; Uniform asymptotic size; Weak identification.

1. INTRODUCTION

Since the late 1990s regression discontinuity (RD) and fuzzy regression discontinuity (FRD) designs have been of growing importance in applied economics. There is extensive theoretical work on RD and FRD designs. A few examples include Hahn, Todd, and Van der Klaauw (1999, 2001); Porter (2003); Buddelmeyer and Skoufias (2004); McCrary (2008); Frölich (2007); Frölich and Melly (2008); Otsu, Xu, and Matsushita (2015); Imbens and Kalyanaraman (2012); Calonico, Cattaneo, and Titiunik (2014); Arai and Ichimura (2013); Papay, Willett, and Murnane (2011); Imbens and Zajonc (2011); Dong and Lewbel (2010); and Fe (2012). See Van der Klaauw (2008) and Lee and Lemieux (2010) for a review of much of this literature. Hundreds of recent applied articles have used RD, and in many cases FRD designs. (For example, as of July 18, 2013, Imbens and Lemieux (2008) review of RD and FRD best practices was cited in 990 articles according to Google Scholar, with 372 of these articles explicitly considering FRD.) Around the same time, the seminal works of Bound, Jaeger, and Baker (1995) and Staiger and Stock (1997) made weak identification in an instrumental variables (IV) context an important consideration in applied work (see Stock, Wright, and Yogo 2002; Andrews and Stock 2007 for surveys of the literature). However, despite the close parallel between an IV setting and the FRD design (see Hahn, Todd, and Van der Klaauw 2001) there has been no theoretical or practical attempt to deal with weak identification in the FRD design more broadly.

To get a sense of the practical importance of weak identification in the FRD design, we have examined a sample of influential applied articles that use the design. We then apply

the F -statistic standards discussed below to see how many of these articles may suffer from a weak identification problem. We find that in about half of the articles where enough information is reported to compute the F -statistic, weak identification appears to be a problem in at least one of the empirical specifications. (For the procedure followed to obtain the sample of articles, see the online supplement, Section 1.) We take this as evidence that weak identification is a serious concern in the applied FRD design literature. Since it is a matter of practical importance, we examine weak identification in the context of the FRD design, demonstrate the problems that arise, and propose uniformly valid testing procedures for treatment (RD) effects.

In this article, we show that the local-to-zero analytical framework common in the weak instruments literature can be adapted to FRD, and when identification is weak, we show that the usual t -test based on the FRD estimator and its standard error suffers from asymptotic size distortions. The usual confidence intervals constructed as estimate \pm constant \times standard error are also invalid because their asymptotic coverage probability can be below the assumed nominal coverage when identification is weak. We rely on novel techniques recently developed in the literature on uniform size properties of tests and confidence sets (Andrews, Cheng, and Guggenberger 2011) to formally justify our local-to-zero framework. Unlike the framework used in the weak IV literature, ours depends not only on the sample size but also on a smoothing parameter (the bandwidth).

We suggest a simple modification to the t -test that eliminates the asymptotic size distortions caused by weak identification. Unlike the usual t -statistic, the modified t -statistic uses a null-restricted version of the standard error of the FRD estimator. The modified statistic can be used with standard normal critical values for two-sided testing. For two-sided testing, the proposed test is equivalent to the Anderson–Rubin test (Anderson and Rubin 1949) adopted in the weak IV literature (Staiger and Stock 1997). For one-sided testing, the modified t -statistic has to be used with nonstandard critical values that must be simulated on a case-by-case basis following the approach of Moreira (2001, 2003).

We discuss how to evaluate the magnitude of potential size distortions in practice following the approach of Stock and Yogo (2005). The strength of identification is measured by the concentration parameter, which in the case of FRD depends on the magnitude of the discontinuity in the treatment variable and on the density of the assignment variable (the variable that determines treatment assignment). The magnitude of potential size distortions can be tested by testing hypotheses about the concentration parameter with noncentral χ_1^2 critical values using the F -statistic, which is an analog of the first-stage F -statistic in IV regression. Surprisingly, we find critical values that are much higher than would be required in a simple IV setting. When the F -statistic is only around 10, which is often used as a threshold value for weak/strong identification in the IV literature, a two-sided test with nominal size of 5% is in fact a 13.6% test, and a 5% one-sided test is in fact a 16.9% test. Nearly zero (under 0.5%) size distortions of a 5% two-sided test correspond to the values of the F -statistic above 93.

Asymptotically valid confidence sets for the treatment effect can be obtained by inverting tests based on the modified t -statistic. Since the FRD is an exactly identified model, these confidence sets are easy to compute, as their construction only involves solving a quadratic equation. Most of the literature on weak instruments deals with the case of over identified models (see, e.g., Andrews and Stock 2007). In exactly identified models, the approach suggested by Anderson and Rubin (1949) results in efficient inference if instruments turn out to be strong and remains valid if instruments are weak. However, in over-identified models, Anderson and Rubin's tests are no longer efficient even when instruments are strong. Several articles (Kleibergen 2002; Moreira 2003; Andrews, Moreira, and Stock 2006) proposed modifications to Anderson and Rubin's basic procedure to gain back efficiency in over identified models. Since the FRD design is an exactly identified model, we can adapt Anderson and Rubin's approach without any loss of power. These confidence sets are expected to be as informative as the standard ones, when identification is strong. However, unlike the usual confidence intervals, the confidence sets we propose can be unbounded with positive probability. This property is expected from valid confidence sets in the situations with local identification failure and an unbounded parameter space (see Dufour 1997). In a recent article, Otsu, Xu, and Matsushita (2015) proposed empirical likelihood-based inference for the RD effect. Using the profile empirical likelihood function, they proposed confidence sets for the RD effect, which are expected to be robust against weak identification. However,

they did not provide a formal analysis of the weak identification. While their method does not involve variances estimation and for that reason can enjoy better higher-order properties than our approach, it requires computation of the empirical likelihood function numerically and is computationally more demanding.

We also discuss testing whether the RD effect is homogenous over differing values of some covariates. The proposed testing approach is designed to remain asymptotically valid when identification is weak. This is achieved by building a robust confidence set for a common RD effect across covariates. The null hypothesis of the common RD effect is rejected when that confidence set is empty.

To illustrate how our proposed confidence sets may differ from the standard ones in practice, we compare the results of applying the standard confidence sets and the proposed confidence sets in two separate applications that use the FRD design to estimate the effect of class size on student achievement. Our main finding is that, as weak identification becomes more likely, the standard confidence sets and the weak identification robust confidence sets become increasingly divergent. Interestingly, in a number of cases the robust confidence sets provide more informative answers than the standard ones. More generally, the empirical applications, along with a Monte Carlo study reported in an online supplement, suggest that our simple and robust procedure for computing confidence sets performs well when identification is either strong or weak.

The rest of the article proceeds as follows. In Section 2, we describe the FRD model, derive the uniform asymptotic size of usual t -tests for FRD, discuss size distortions and testing for potential size distortions, and describe weak-identification-robust inference for FRD. Section 3 discusses robust testing for constancy of the RD effect across covariates. We present our empirical applications in Section 4. The online supplement contains additional materials including the proofs and the Monte Carlo results.

2. THEORETICAL RESULTS

2.1 The Model, Estimation, and Standard Inference Approach

In RD designs, the observed outcome variable y_i is modeled as $y_i = y_{0i} + x_i\beta_i$, where x_i is the treatment indicator variable, y_{0i} is the outcome without treatment, and β_i is the random treatment effect for observation i . If x_i is binary, it takes on value one if the treatment is received and zero otherwise. When there are treatments of different intensity, x_i may be nonbinary. The treatment assignment depends on another observable assignment variable, z_i through $E(x_i|z_i = z)$. The main feature in this framework is that $E(x_i|z_i = z)$ is discontinuous at some known cutoff point z_0 , while $E(y_{0i}|z_i)$ is assumed to be continuous at z_0 .

Assumption 1.

- $\lim_{z \downarrow z_0} E(x_i|z_i = z) \neq \lim_{z \uparrow z_0} E(x_i|z_i = z)$.
- $\lim_{z \downarrow z_0} E(y_{0i}|z_i = z) = \lim_{z \uparrow z_0} E(y_{0i}|z_i = z)$.

For binary x_i , when $|\lim_{z \uparrow z_0} E(x_i | z_i = z) - \lim_{z \downarrow z_0} E(x_i | z_i = z)| = 1$, we have a sharp RD design, and a fuzzy design otherwise. When x_i is a continuous treatment variable, the design is sharp if x_i is a deterministic function of z_i , and fuzzy otherwise.

The focus of this article is fuzzy designs, and the main object of interest is the RD effect:

$$\beta = (y^+ - y^-)/(x^+ - x^-), \quad (1)$$

where $y^+ = \lim_{z \downarrow z_0} E(y_i | z_i = z)$, $y^- = \lim_{z \uparrow z_0} E(y_i | z_i = z)$, and x^+ and x^- are defined similarly with y_i replaced by x_i . The exact interpretation of β depends on the assumptions that the econometrician is willing to make in addition to Assumption 1. As discussed by Hahn, Todd, and Van der Klaauw (2001), if β_i and x_i are assumed to be independent conditional on z_i , then β captures the average treatment effect (ATE) at $z_i = z_0$: $\beta = E(\beta_i | z_i = z_0)$. When x_i is binary and under an alternative set of conditions, which allow for dependence between x_i and β_i , Hahn, Todd, and Van der Klaauw (2001) showed that the RD effect captures the local ATE (LATE) or ATE for compliers at z_0 , where compliers are observations for which x_i switches its value from zero to one when z_i changes from $z_0 - e$ to $z_0 + e$ for all small $e > 0$. (See the discussion on page 204 of their article.)

Regardless of its interpretation, the RD effect is estimated by replacing the unknown population objects in (1) with their estimates. Following Hahn, Todd, and Van der Klaauw (2001), it is now a standard approach to estimate y^+ , y^- , x^+ , and x^- using local linear kernel regression. Let $K(\cdot)$ and h_n denote the kernel function and bandwidth, respectively. For estimation of y^+ , the local linear regression is

$$(\hat{a}_n, \hat{b}_n) = \arg \min_{a,b} \sum_{i=1}^n (y_i - a - (z_i - z_0)b)^2 \times K\left(\frac{z_i - z_0}{h_n}\right) 1\{z_i \geq z_0\}, \quad (2)$$

and the local linear estimator of y^+ is given by $\hat{y}_n^+ = \hat{a}_n$. The local linear estimator for y^- can be constructed analogously by replacing $1\{z_i \geq z_0\}$ with $1\{z_i < z_0\}$ in (2). Similarly, one can estimate x^+ and x^- by replacing y_i with x_i . Let \hat{y}_n^- , \hat{x}_n^+ , and \hat{x}_n^- denote the local linear estimators of y^- , x^+ , and x^- , respectively. The corresponding estimator of β is given by

$$\hat{\beta}_n = (\hat{y}_n^+ - \hat{y}_n^-)/(\hat{x}_n^+ - \hat{x}_n^-).$$

The asymptotic properties of the local linear estimators and $\hat{\beta}_n$ are discussed in Hahn, Todd, and Van der Klaauw (1999) and Imbens and Lemieux (2008). We assume that the following conditions are satisfied.

Assumption 2.

- $K(\cdot)$ is continuous, symmetric around zero, nonnegative, and compactly supported second-order kernel.
- $\{(y_i, x_i, z_i)\}_{i=1}^n$ are iid; y_i, x_i, z_i have a joint distribution F such that

- $f_z(\cdot)$ (the marginal PDF of z_i) exists and is bounded from above, bounded away from zero, and twice continuously differentiable with bounded derivatives on \mathcal{N}_{z_0} (a small neighborhood of z_0).
- $E(y_i | z_i)$ and $E(x_i | z_i)$ are bounded on \mathcal{N}_{z_0} and twice continuously differentiable with bounded derivatives on $\mathcal{N}_{z_0} \setminus \{z_0\}$; $\lim_{e \downarrow 0} \frac{d^p}{dz^p} E(y_i | z_i = z_0 \pm e)$ and $\lim_{e \downarrow 0} \frac{d^p}{dz^p} E(x_i | z_i = z_0 \pm e)$ exist for $p = 0, 1, 2$.
- $\sigma_y^2(z_i) = \text{var}(y_i | z_i)$ and $\sigma_x^2(z_i) = \text{var}(x_i | z_i)$ are bounded from above and bounded away from zero on \mathcal{N}_{z_0} ; $\lim_{e \downarrow 0} \sigma_y^2(z_0 \pm e)$, $\lim_{e \downarrow 0} \sigma_x^2(z_0 \pm e)$, and $\lim_{e \downarrow 0} \sigma_{xy}(z_0 \pm e)$ exist, where $\sigma_{xy}(z_i) = \text{cov}(x_i, y_i | z_i)$; $|\rho_{xy}| \leq \bar{\rho}$ for some $\bar{\rho} < 1$, where $\rho_{xy} = \sigma_{xy}/(\sigma_x \sigma_y)$, $\sigma_{xy} = \lim_{e \downarrow 0} (\sigma_{xy}(z_0 + e) + \sigma_{xy}(z_0 - e))$, and σ_x^2 and σ_y^2 defined similarly with the conditional covariance replaced by the conditional variances of x_i and y_i , respectively.
- For some $\delta > 0$, $E(|y_i - E(y_i | z_i)|^{2+\delta} | z_i)$ and $E(|x_i - E(x_i | z_i)|^{2+\delta} | z_i)$ are bounded on \mathcal{N}_{z_0} .
- As $n \rightarrow \infty$, $\sqrt{nh_n}h_n^2 \rightarrow 0$ and $nh_n^3 \rightarrow \infty$.

Remark. (1) The smoothness conditions imposed in Assumption 2(b) are standard for kernel estimation except for the left/right limit conditions in parts (ii) and (iii), which are due to the discontinuity design and have been used in Hahn, Todd, and Van der Klaauw (1999). (2) Asymptotic normality of the local linear estimators is established using Lyapounov's central limit theorem (CLT), and part (iv) of Assumption 2(b) can be used to verify Lyapounov's condition (see Davidson 1994, Theorem 23.12, p. 373). (3) With twice differentiable functions, the bias of the local linear estimators is of order h_n^2 even near the boundaries. The condition $\sqrt{nh_n}h_n^2 \rightarrow 0$ in Assumption 2(c) is an under-smoothing condition, which makes the contribution of the bias term to the asymptotic distribution negligible. The condition $nh_n^3 \rightarrow \infty$ ensures that the variance of the local linear estimator tends to zero. Assumption 2(c) is satisfied if the bandwidth is chosen according to the rule $h_n = \text{constant} \times n^{-r}$ with $1/5 < r < 1/3$.

It is convenient for our purposes to present the asymptotic properties of the local linear estimators and the FRD estimator as follows. Define

$$k = \frac{\int_0^\infty (\int_0^\infty s^2 K(s) ds - u \int_0^\infty s K(s) ds)^2 K^2(u) du}{\left(\int_0^\infty u^2 K(u) du \int_0^\infty K(u) du - \left(\int_0^\infty u K(u) du\right)^2\right)^2}.$$

The constant k is known as it depends only on the kernel function. In the case of asymmetric kernels, we will have two different constants for the left and right estimators, with the bounds of integration replaced by $(-\infty, 0]$ for the left estimators. For $\Delta y = y^+ - y^-$, $\Delta y_n = \hat{y}_n^+ - \hat{y}_n^-$, and similarly defined Δx and Δx_n , by Assumption 2 and Lyapounov's CLT we have

$$\sqrt{nh_n} \begin{pmatrix} \Delta y_n - \Delta y \\ \Delta x_n - \Delta x \end{pmatrix} \rightarrow_d \sqrt{\frac{k}{f_z(z_0)}} \begin{pmatrix} \sigma_y \mathcal{Y} \\ \sigma_x \mathcal{X} \end{pmatrix},$$

where \mathcal{Y} and \mathcal{X} are two bivariate normal variables with zero means, unit variances, and correlation coefficient ρ_{xy} (the latter

is defined in Assumption 2(b)iii together with σ_x and σ_y . This in turn implies that under standard asymptotics, $\sqrt{nh_n}(\hat{\beta}_n - \beta) \rightarrow_d N(0, k\sigma^2(\beta)/(f_z(z_0)(\Delta x)^2))$, where $\sigma^2(b) = \sigma_y^2 + b^2\sigma_x^2 - 2b\sigma_{xy}$. The last result holds due to Assumption 1(a), that is, only when $\Delta x \neq 0$ and is fixed.

The asymptotic variance σ_y^2 can be consistently estimated by

$$\hat{\sigma}_{y,n}^2 = \frac{1}{\hat{f}_{z,n}(z_0)} \frac{1}{nh_n} \sum_{i=1}^n (y_i - \hat{y}_n^+ 1\{z_i \geq z_0\} - \hat{y}_n^- 1\{z_i < z_0\})^2 \times K\left(\frac{z_i - z_0}{h}\right),$$

where $\hat{f}_{z,n}(z_0)$ is the kernel estimator of $f_z(z_0)$: $\hat{f}_{z,n}(z_0) = (nh_n)^{-1} \sum_{i=1}^n K((z_i - z_0)/h_n)$. Consistent estimators of σ_x^2 and σ_{xy} can be constructed similarly by replacing $(y_i - \hat{y}_n^+ 1\{z_i \geq z_0\} - \hat{y}_n^- 1\{z_i < z_0\})^2$ with $(x_i - \hat{x}_n^+ 1\{z_i \geq z_0\} - \hat{x}_n^- 1\{z_i < z_0\})^2$ and $(x_i - \hat{x}_n^+ 1\{z_i \geq z_0\} - \hat{x}_n^- 1\{z_i < z_0\})(y_i - \hat{y}_n^+ 1\{z_i \geq z_0\} - \hat{y}_n^- 1\{z_i < z_0\})$, respectively. Hence, a consistent estimator of $\sigma^2(b)$ can be constructed as

$$\hat{\sigma}_n^2(b) = \hat{\sigma}_{y,n}^2 + b\hat{\sigma}_{x,n}^2 - 2b\hat{\sigma}_{xy,n}. \quad (3)$$

A common inference approach for the FRD effect is based on the usual t -statistic. Thus, when testing $H_0 : \beta = \beta_0$ one typically computes

$$T_n(\beta_0) = \sqrt{nh_n} (\hat{\beta}_n - \beta_0) / \sqrt{k\hat{\sigma}_n^2(\hat{\beta}_n) / (\hat{f}_{z,n}(z_0)(\widehat{\Delta x}_n)^2)}$$

and compares it with standard normal critical values, as $T_n(\beta) \rightarrow_d N(0, 1)$, when $\Delta x \neq 0$ and is fixed. Confidence intervals for β are constructed by collecting all values β_0 for which $H_0 : \beta = \beta_0$ cannot be rejected using a test based on $T_n(\beta_0)$.

2.2 Weak Identification in FRD

Weak identification is a finite-sample problem, which occurs when the noise due to sampling errors is of the same magnitude or even dominates the signal in estimation of a model's parameters. In such cases, the asymptotic normality result $T_n(\beta) \rightarrow_d N(0, 1)$ provides a poor approximation to the actual distribution of the t -statistic, and as a result inference may be distorted.

Assuming that $H_0 : \beta = \beta_0$, we can rewrite the t -statistic as

$$T_n(\beta) = \frac{\sqrt{nh_n} (\widehat{\Delta y}_n - \beta \widehat{\Delta x}_n)}{\sqrt{k\hat{\sigma}_n^2(\hat{\beta}_n) / \hat{f}_{z,n}(z_0)}} \times \text{sign}(\widehat{\Delta x}_n). \quad (4)$$

When testing H_0 against two-sided alternatives, one uses the absolute value of $T_n(\beta)$, which eliminates the sign term. Since under standard (fixed distribution) asymptotics $\sqrt{nh_n} (\widehat{\Delta y}_n - \beta \widehat{\Delta x}_n) \rightarrow_d N(0, k\sigma^2(\beta)/f_z(z_0))$, the usual t -test has no size distortions as long as $\hat{\beta}_n$ is consistent and $\hat{\sigma}_n^2(\hat{\beta}_n)$ approximates $\sigma^2(\beta)$ very well. Define $\Delta Y_n = (f_z(z_0)/k)^{1/2}(nh_n)^{1/2}(\widehat{\Delta y}_n - \Delta y)$ and $\Delta X_n = (f_z(z_0)/k)^{1/2}(nh_n)^{1/2}(\widehat{\Delta x}_n - \Delta x)$. We can now write

$$\hat{\beta}_n - \beta = \frac{\Delta Y_n - \beta \Delta X_n}{\Delta X_n + (f_z(z_0)/k)^{1/2}(nh_n)^{1/2} \Delta x}. \quad (5)$$

Note that in the above expression, the estimation errors ΔY_n and ΔX_n represent the noise components, while the signal

component is given by $(nh_n)^{1/2} \Delta x$. Since the noise terms have bounded variances, the signal dominates the noise as long as $(nh_n)^{1/2} \Delta x \rightarrow \infty$. In this case, $\hat{\beta}_n \rightarrow_p \beta$. If, however, $\lim_{n \rightarrow \infty} |(nh_n)^{1/2} \Delta x| < \infty$, the signal and noise are of the same magnitude, which results in inconsistency of the FRD estimator and weak identification.

Thus, similarly to the weak IV literature (Staiger and Stock 1997), it is appropriate to model weak identification by assuming that Δx is inversely related to the square root of the sample size. However, the kernel estimation framework and presence of the bandwidth, which is chosen by the econometrician, require some adjustments. Suppose one models weak identification as $\Delta x \sim 1/(ng_n)^{1/2}$, for some sequence $g_n \rightarrow 0$ as $n \rightarrow \infty$. In this case, the econometrician can obtain consistency of $\hat{\beta}_n$ and resolve weak identification simply by choosing h_n so that $h_n/g_n \rightarrow \infty$. This situation resembles so-called nearly weak or semistrong identification, see Hahn and Kuersteiner (2002), Caner (2009), Antoine and Renault (2009, 2012), and Antoine and Lavergne (2014). Hence, the worst-case scenario, in which the econometrician cannot resolve weak identification by tweaking the bandwidth, occurs when $g_n = h_n$, that is, $\Delta x \sim 1/(nh_n)^{1/2}$.

This idea can be formalized using the results obtained in the recent literature on uniform size properties of tests and confidence sets: Andrews and Guggenberger (2010), Andrews and Cheng (2012), and Andrews, Cheng, and Guggenberger (2011). The latter article provides a general framework of establishing uniform size properties of tests and confidence sets. To describe this framework, let S_n be a test statistic with exact finite-sample distribution (in a sample of size n) determined by $\lambda \in \Lambda$. Note that λ may include infinite-dimensional components such as distribution functions. Let $cr_n(\alpha)$ denote a possibly data-dependent critical region for nominal significance level α . The test rejects a null hypothesis when $S_n \in cr_n(\alpha)$, and the rejection probability is given by $RP_n(\lambda) = P_\lambda(S_n \in cr_n(\alpha))$, where subscript λ in P_λ indicates that the probability is computed for a given value of $\lambda \in \Lambda$. The exact size is defined as $ExSz_n = \sup_{\lambda \in \Lambda} RP_n(\lambda)$. Note that $ExSz_n$ captures the maximum rejection probability for any combination of parameters λ (the worst case scenario). In large samples, the exact size is approximated by asymptotic size $AsySz = \limsup_{n \rightarrow \infty} \sup_{\lambda \in \Lambda} RP_n(\lambda)$. Contrary to the usual point-wise asymptotic approach, $AsySz$ is determined by taking supremum over the parameter space before taking limit with respect to n . It has been argued in many articles that controlling $AsySz$ is crucial for ensuring reliable inference when test statistics have discontinuous asymptotic distribution, that is, when point-wise asymptotic distribution is discontinuous in a parameter. On the importance of uniform size, see, for example, Imbens and Manski (2004, p. 1848), Mikusheva (2007), and references in Andrews, Cheng, and Guggenberger (2011). In what follows, we rely on the following result of Andrews, Cheng, and Guggenberger (2011): (Lemma 1 combines Assumption B and Theorems 2.1 and 2.2 in Andrews, Cheng, and Guggenberger (2011).

Lemma 1 (Andrews, Cheng, and Guggenberger 2011). Let $\{d_n(\lambda) : n \geq 1\}$ be a sequence of functions, where $d_n : \Lambda \rightarrow \mathbb{R}^J$. Define $D = \{d \in \{\mathbb{R} \cup \{\pm\infty\}\}^J : d_{p_n}(\lambda_{p_n}) \rightarrow d \text{ for some subsequence } \{p_n\} \text{ of } \{n\} \text{ and some sequence } \{\lambda_{p_n} \in \Lambda\}\}$. Sup-

pose that for any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{\lambda_{p_n} \in \Lambda\}$ for which $d_{p_n}(\lambda_{p_n}) \rightarrow d \in D$, we have that $\text{RP}_{p_n}(\lambda_{p_n}) \rightarrow \text{RP}(d)$ for some function $\text{RP}(d) \in [0, 1]$. Then, $\text{AsySz} = \sup_{d \in D} \text{RP}(d)$.

To apply Lemma 1, we define

$$\lambda_1 = \left(\frac{f_z(z_0)}{k} \right)^{1/2} \frac{|\Delta x|}{\sigma_x}, \quad \lambda_2 = \rho_{xy}, \quad \lambda_3 = \beta \sigma_x / \sigma_y. \quad (6)$$

We define $\lambda_4 = F$, where F is the joint distribution of x_i, y_i, z_i and is such that, given $\lambda_1 \in \mathbb{R}_+$, $\lambda_2 \in [-\bar{\rho}, \bar{\rho}]$, and $\lambda_3 \in \mathbb{R}$, the three equations in (6) hold. Note that λ_4 is an infinite-dimensional parameter that depends on λ_1, λ_2 , and λ_3 . As explained by Andrews, Cheng, and Guggenberger (2011, pp. 8–9), $d_n(\lambda)$ is chosen so that when $d_n(\lambda_n)$ converges to $d \in D$ for some sequence of parameters $\{\lambda_n \in \lambda : n \geq 1\}$, the test statistic converges to some limiting distribution, which might depend on d . In view of (4) and (5), we therefore define

$$d_{n,1}(\lambda) = \sqrt{nh_n} \lambda_1, \quad d_{n,2}(\lambda) = \lambda_2, \quad d_{n,3}(\lambda) = \lambda_3. \quad (7)$$

While $\lambda_4 = F$ affects the finite-sample distribution of the test statistic, it does not enter its asymptotic distribution, and therefore can be dropped from $d_n(\lambda)$ as discussed by Andrews, Cheng, and Guggenberger (2011, p. 8). Hence, $D = \{\mathbb{R}_+ \cup \{+\infty\}\} \times [-\bar{\rho}, \bar{\rho}] \times \{\mathbb{R} \cup \{\pm\infty\}\}$.

Next, we describe the asymptotic size of tests for FRD based on the usual t -statistic and standard normal critical value. Let z_ν denote the ν th quantile of the standard normal distribution.

Theorem 1. Suppose that Assumption 2 holds. Let \mathcal{X}, \mathcal{Y} be two bivariate normal variables with zero means, unit variances, and correlation d_2 . Define

$$\mathcal{T}_{d_1, d_2, d_3} = \frac{\mathcal{Y} - d_3 \mathcal{X}}{\sqrt{1 + \left(\frac{\mathcal{Y} + d_3 d_1}{\mathcal{X} + d_1} \right)^2 - 2d_2 \frac{\mathcal{Y} + d_3 d_1}{\mathcal{X} + d_1}}} \times \text{sign}(\mathcal{X} + d_1).$$

- For tests that reject $H_0 : \beta = \beta_0$ in favor of $H_1 : \beta \neq \beta_0$ when $|T_n(\beta_0)| > z_{1-\alpha/2}$, $\text{AsySz} = \sup_{d_1 \in \mathbb{R}_+ \cup \{+\infty\}, d_2 \in [0, \bar{\rho}], d_3 \in \mathbb{R} \cup \{\pm\infty\}} P(|\mathcal{T}_{d_1, d_2, d_3}| > z_{1-\alpha/2})$.
- For tests that reject $H_0 : \beta \leq \beta_0$ in favor of $H_1 : \beta > \beta_0$ when $T_n(\beta_0) > z_{1-\alpha}$, $\text{AsySz} = \sup_{d_1 \in \mathbb{R}_+ \cup \{+\infty\}, d_2 \in [-\bar{\rho}, \bar{\rho}], d_3 \in \mathbb{R} \cup \{\pm\infty\}} P(\mathcal{T}_{d_1, d_2, d_3} > z_{1-\alpha})$.

Remark. A commonly used measure of identification strength is the so-called *concentration parameter*. On the importance of the concentration parameter in IV estimation, see, for example, Stock and Yogo (2005). In our framework, the concentration parameter is given by $d_{n,1}^2$, where $d_{n,1}^2 \rightarrow \infty$ corresponds to strong (or semistrong) identification, and identification is weak when the limit of $d_{n,1}^2$ is finite. As it is apparent from the expressions for λ_1 and $d_{n,1}$ in (6) and (7), the concentration parameter and, therefore, the strength of identification depend not only on the size of discontinuity in treatment assignment Δx , but also on $f_z(z_0)$, the PDF of the assignment variable at z_0 . Hence, smaller values of $f_z(z_0)$ would correspond to a more severe weak identification problem.

For any permitted values of d_2 and d_3 , when $d_1 = \infty$, we have $\mathcal{T}_{\infty, d_2, d_3} \sim N(0, 1)$. Thus, the asymptotic size of tests based on $T_n(\beta_0)$ is equal to nominal size α under strong or semistrong identification. When $d_1 < \infty$, it is straightforward to compute AsySz numerically. To compute asymptotic rejection probabilities given d_1, d_2, d_3 , first using bivariate normal PDFs, one integrates numerically $1(|\mathcal{T}_{d_1, d_2, d_3}| > z_{1-\alpha/2})$ or $1(\mathcal{T}_{d_1, d_2, d_3} > z_{1-\alpha})$ over the support of the joint distribution of \mathcal{Y}, \mathcal{X} . Rejection probabilities then can be numerically maximized over d 's.

Table 1 reports maximal rejection probabilities of one- and two-sided tests based on the usual t -statistic. The rejection probabilities reported in Table 1 were computed by numerical integration using `quad2d` function in Matlab. Integration bounds for normal variables were set to $[-7, 7]$, and the rejection probabilities were maximized over the following grids of values: from -0.99 to 0.99 at 0.01 intervals for d_2 , and from -1000 to 1000 at 0.5 intervals for d_3 . It shows that AsySz approaches one as the concentration parameter approaches zero. Size distortions decrease monotonically as the concentration parameter increases. In the case of two-sided testing, nearly zero size distortions (under 0.5%) correspond to the concentration parameter of order $d_1^2 \geq 64$ for asymptotic 5% tests, and $d_1^2 \geq 50^2$ for asymptotic 1% tests. The table also shows that one-sided tests suffer from more substantial size distortions than two-sided tests, which is due to asymmetries in the distribution of $\mathcal{T}_{d_1, d_2, d_3}$.

2.3 Testing for Potential Size Distortions

Following the approach of Stock and Yogo (2005), Table 1 can be used for testing a null hypothesis about the largest potential size distortion against an alternative hypothesis under which the largest potential size distortion does not exceed a certain prespecified level. Suppose that the econometrician decides that identification is strong enough if, in the case of 1% two-sided testing, the maximal rejection probability does not exceed 5%. Thus, the econometrician effectively adopts tests with 5% significance level, however uses the 1% standard normal critical value. According to the results in Table 1, the corresponding null hypothesis and its alternative in this case can be stated in terms of the concentration parameter d_1^2 as $H_0^W : d_1^2 \leq 9$ and $H_1^S : d_1^2 > 9$, respectively. A test of H_0^W can be based on the estimator of discontinuity Δx . Define

$$F_n = \frac{nh_n(\widehat{\Delta x}_n)^2}{\hat{\sigma}_{x,n}^2 k / \hat{f}_{z,n}(z_0)} = ((\Delta X_n / \sigma_x) + d_{n,1})^2 + o_p(1). \quad (8)$$

As long as the concentration parameter is finite, $F_n \rightarrow_d \chi_1^2(d_1^2)$, a noncentral χ_1^2 distribution with noncentrality parameter d_1^2 . Let $\chi_{1,1-\tau}^2(d_1^2)$ denote the $(1 - \tau)$ th quantile of the $\chi_1^2(d_1^2)$ distribution. Since size distortions are monotonically decreasing when the concentration parameter increases, an asymptotic size τ test of H_0^W should reject it when $F_n > \chi_{1,1-\tau}^2(d_1^2)$.

Noncentral χ_1^2 critical values are reported in the last two columns of Table 1 for selected values of the concentration parameter and $\tau = 0.05, 0.01$. For example, $H_0^W : d_1^2 \leq 9$ should be rejected in favor of $H_1^S : d_1^2 > 9$ by a 5% test when $F_n > 21.57$. In the case of 5% two-sided testing of β , one needs the concentration parameter of at least 64 to en-

Table 1. Maximal asymptotic rejection probabilities for different values of the concentration parameter (d_1^2) of one- and two-sided t -tests for FRD with significance level α , and noncentral χ_1^2 critical values for testing hypotheses about the concentration parameter at significance level τ

d_1^2	Maximal rejection prob. for FRD				Noncentral $\chi_1^2(d_1^2)$ critical values	
	One-sided		Two-sided		$\tau = 0.05$	$\tau = 0.01$
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$		
10^{-4}	0.906	0.885	0.893	0.877	3.84	6.64
0.01	0.691	0.636	0.664	0.622	3.88	6.70
0.25	0.363	0.294	0.322	0.261	4.76	8.08
1.0	0.221	0.153	0.187	0.134	7.00	11.06
4.0	0.144	0.086	0.113	0.070	13.28	18.72
9.0	0.119	0.062	0.099	0.050	21.57	28.37
16.0	0.106	0.051	0.076	0.038	31.87	40.03
25.0	0.097	0.045	0.067	0.031	44.15	53.67
36.0	0.091	0.037	0.060	0.029	58.45	69.34
49.0	0.086	0.033	0.056	0.023	74.73	86.98
64.0	0.081	0.032	0.053	0.022	93.03	106.63
81.0	0.078	0.029	0.052	0.022	113.31	128.28
10^2	0.076	0.029	0.052	0.020	135.60	151.94
25^2	0.061	0.020	0.051	0.015	709.96	746.72
50^2	0.056	0.014	0.051	0.012	2667.17	2738.06

sure that size distortions are under 0.5%. In that case, a 5% test should reject the null hypothesis of weak identification if $F_n > 93.03$.

Note that the critical values in Table 1 substantially exceed the rule-of-thumb of 10, which is often used in the literature as a threshold value for weak IVs. According to our calculations, with an F -statistic of only 10, one cannot reject $H_0^W : d_1^2 \leq 1.51^2$ at 5% significance level. However, a concentration parameter of 1.51^2 corresponds to maximal rejection probabilities of 16.9% and 13.6% for 5% one-sided and two-sided tests, respectively.

The results from Table 1 can also be used for designing valid tests (for the FRD effect β) based on usual t -statistics in combination with somewhat larger than usual critical values. For example, suppose one is interested in a 5% two-sided test about β , and rejects the null hypothesis when $F_n > 21.57$ and $|T_n(\beta_0)|$ exceeds the 1% standard normal critical value. According to Table 1, if the concentration parameter $d_1^2 \geq 9$, the asymptotic size does not exceed 5%. On the other hand, if $d_1^2 \leq 9$, $\lim_{n \rightarrow \infty} P(F_n > 21.75) \leq 0.05$. Hence, overall this test has an asymptotic 5% significance level. Intuitively, such a test is valid because the null-hypothesis for the F -pretest assumes size distortions, and one proceeds using the t -statistic only if it is rejected, that is, if the concentration parameter is found to be large enough. Note, however, that the procedure is conservative. Furthermore, passing the F -test does not completely safeguard against size distortions, and the usual t -statistic must be used with somewhat larger critical values.

Although the F -test provides useful guidance on the potential magnitude of size distortions, practitioners should not solely rely on this test to decide whether it is worth proceeding with the estimation. With this in mind, we present a robust inference approach in the next section that always yields valid confidence intervals regardless of the strength of identification and does not rely on any pretests.

2.4 Weak-Identification-Robust Inference for FRD

A common approach adopted in the weak IV literature is to use weak-identification-robust statistics to test hypotheses about structural parameters directly, instead of using their estimates and standard errors. The Anderson–Rubin (AR) statistic (Anderson and Rubin 1949; Staiger and Stock 1997) is often used for that purpose. In the context of IV regression, the AR statistic can be used to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ by testing whether the null-restricted residuals computed for $\beta = \beta_0$ are uncorrelated with the instruments.

In our case, the structural parameter is defined by (1). Hence, to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, following the AR approach, we can test instead $H_0 : \Delta y - \beta_0 \Delta x = 0$ against $H_1 : \Delta y - \beta_0 \Delta x \neq 0$. A test, therefore, can be based on

$$\frac{nh_n (\widehat{\Delta y}_n - \beta_0 \widehat{\Delta x}_n)^2}{k \hat{\sigma}_n^2(\beta_0) / \hat{f}_{z,n}(z_0)} = |T_n^R(\beta_0)|^2,$$

where $T_n^R(\beta_0)$ denotes a modified or null-restricted version of the usual t -statistic:

$$T_n^R(\beta_0) = \sqrt{nh_n} (\hat{\beta}_n - \beta_0) / \sqrt{k \hat{\sigma}_n^2(\beta_0) / (\hat{f}_{z,n}(z_0) (\widehat{\Delta x}_n)^2)},$$

and the equality holds by (4). Unlike the usual t -statistic, $T_n^R(\beta_0)$ uses the null-restricted value β_0 instead of $\hat{\beta}_n$ when computing the standard error. In view of the discussion at the beginning of Section 2.2 and since the asymptotic distribution of $|T_n^R(\beta_0)|$ does not depend on the concentration parameter, replacing $\hat{\sigma}_n^2(\hat{\beta}_n)$ by $\hat{\sigma}_n^2(\beta_0)$ eliminates size distortions.

Theorem 2. Suppose that Assumption 2 holds. Tests that reject $H_0 : \beta = \beta_0$ in favor of $H_1 : \beta \neq \beta_0$ when $|T_n^R(\beta_0)| > z_{1-\alpha/2}$ have AsySz equal to α .

Consider now a one-sided testing problem $H_0 : \beta \leq \beta_0$ versus $H_1 : \beta > \beta_0$. Again, one can base a test on the null-

restricted statistic. In this case under H_0 when $\beta = \beta_0$, we have $T_n^R(\beta) = (\Delta Y_n - \beta \Delta X_n) \times \text{sign}(\Delta X_n \pm d_{n,1}) / \sigma(\beta) + o_p(1)$. When identification is strong or semistrong, $d_{n,1} \rightarrow \infty$, and the sign term is constant with probability one. Since the first term is asymptotically $N(0, 1)$, $T_n^R(\beta)$ is also asymptotically $N(0, 1)$, and one could use standard normal critical values. On the other hand, when identification is weak and the concentration parameter is small, the sign term is random, and therefore, the null asymptotic distribution of the product differs from standard normal. To obtain an asymptotically uniformly valid test, one can use data-dependent critical values that automatically adjust to the strength of identification. Such critical values can be generated using the approach of Moreira (2001, 2003) by conditioning on a statistic that is (i) asymptotically independent of $\Delta Y_n - \beta \Delta X_n$, and (ii) summarizes the information on the strength of identification (see also Andrews, Moreira, and Stock 2006; Mills, Moreira, and Vilela 2014).

Define $S_n = (\Delta Y_n - \beta \Delta X_n) / \sigma(\beta)$ and $Q = \Delta X_n / \sigma_x - (\sigma_{xy} - \beta \sigma_x^2) S_n / (\sigma_x \sigma(\beta))$, so that, when $\beta = \beta_0$, $T_n^R(\beta) = S_n \times \text{sign}[Q_n \pm d_{n,1} + (\sigma_{xy} - \beta \sigma_x^2) S_n / (\sigma_x \sigma(\beta))] + o_p(1)$.

When identification is weak, S_n and Q_n are asymptotically independent by construction, while $S_n \rightarrow_d N(0, 1)$. Therefore, one can construct data-dependent critical values as follows. First, compute

$$\hat{Q}_n(\beta_0) = \frac{\sqrt{nh_n} \widehat{\Delta x}_n}{\sqrt{k \hat{\sigma}_{x,n}^2 / \hat{f}_{z,n}(z_0)}} - \frac{\hat{\sigma}_{xy,n} - \beta_0 \hat{\sigma}_{x,n}^2}{\hat{\sigma}_{x,n} \hat{\sigma}_n(\beta_0)} \times \left(\frac{\sqrt{nh_n} (\widehat{\Delta y}_n - \beta_0 \widehat{\Delta x}_n)}{\sqrt{k \hat{\sigma}_n^2(\beta_0) / \hat{f}_{z,n}(z_0)}} \right).$$

Second, simulate M -independent $N(0, 1)$ random variables $\{S_1, \dots, S_M\}$ for some large M . Third, for $m = 1, \dots, M$ compute

$$\hat{T}_{n,m}^R(\beta_0, \hat{Q}_n(\beta_0)) = S_m \times \text{sign} \left(\hat{Q}_n(\beta_0) + \frac{\hat{\sigma}_{xy,n} - \beta_0 \hat{\sigma}_{x,n}^2}{\hat{\sigma}_{x,n} \hat{\sigma}_n(\beta_0)} S_m \right).$$

Let $\hat{c}_{v_{n,1-\alpha}}(\beta_0, \hat{Q}_n(\beta_0))$ denote the $(1 - \alpha)$ th quantile of the sample distribution of $\{\hat{T}_{n,m}^R(\beta_0, \hat{Q}_n(\beta_0)) : m = 1, \dots, M\}$. To obtain an asymptotically uniformly valid one-sided test, one can use $\hat{c}_{v_{n,1-\alpha}}(\beta_0, \hat{Q}_n(\beta_0))$ as the critical value.

Theorem 3. Suppose that Assumption 2 holds. Tests that reject $H_0 : \beta \leq \beta_0$ in favor of $H_1 : \beta > \beta_0$ when $T_n^R(\beta_0) > \hat{c}_{v_{n,1-\alpha}}(\beta_0, \hat{Q}_n(\beta_0))$ have AsySz equal to α .

Weak-identification-robust confidence sets for β can be constructed by inversion of the robust tests. For example, a confidence set for β with asymptotic coverage probability $1 - \alpha$ can be constructed by collecting all values β_0 that cannot be rejected by the two-sided robust test:

$$CS_{1-\alpha,n} = \{\beta_0 \in \mathbb{R} : |T_n^R(\beta_0)| \leq z_{1-\alpha/2}\}. \quad (9)$$

This confidence set can be easily computed analytically by solving for values of β_0 that satisfy the inequality

$$(\hat{\beta}_n - \beta_0)^2 \hat{\sigma}_{x,n}^2 F_n - z_{1-\alpha/2}^2 (\hat{\sigma}_{y,n}^2 + \beta_0^2 \hat{\sigma}_{x,n}^2 - 2\hat{\sigma}_{xy,n} \beta_0) \leq 0, \quad (10)$$

where F_n is defined in (8).

Depending on the coefficients of the second-order polynomial (in β_0) in Equation (10), $CS_{1-\alpha,n}$ can take one of the following

forms: (i) an interval, (ii) a union of two disconnected half-lines $(-\infty, a_1] \cup [a_2, \infty)$, where $a_1 < a_2$, or (iii) the entire real line. One will see cases (ii) or (iii) if the coefficient on β_0^2 in (10) is negative, which occurs when

$$F_n - z_{1-\alpha/2}^2 < 0. \quad (11)$$

Thus, in practice one will see nonstandard confidence sets if the null hypothesis $\Delta x = 0$ cannot be rejected using the F -statistic and central $\chi_{1,1-\alpha}^2$ critical values. Case (iii) arises when the discriminant of the quadratic polynomial in (10) is negative, which occurs if

$$F_n \hat{\sigma}_n^2(\hat{\beta}_n) - z_{1-\alpha/2}^2 (\hat{\sigma}_{y,n}^2 - \hat{\sigma}_{xy,n}^2 / \hat{\sigma}_{x,n}^2) < 0. \quad (12)$$

Positive definiteness of the variance-covariance matrix composed of $\hat{\sigma}_{x,n}^2$, $\hat{\sigma}_{y,n}^2$, and $\hat{\sigma}_{xy,n}$ implies that (11) holds whenever (12) holds. Thus, negative discriminants implied by (12) are inconsistent with $F_n > z_{1-\alpha/2}^2$ or positive coefficients on β_0^2 in (10). This in turn implies that $CS_{1-\alpha,n}$ cannot be empty.

When identification is strong or semistrong, the concentration parameter and, therefore, F_n diverge to infinity. In such cases, both the discriminant and the coefficient on β_0^2 tend to be positive, and consequently, $CS_{1-\alpha,n}$ will be an interval with probability approaching one.

Furthermore, one can show that when identification is strong and under local alternatives of the form $\beta = \beta_0 + \mu / (nh_n)^{1/2}$, tests based on $T_n(\beta_0)$ and $T_n^R(\beta_0)$ have the same asymptotic power. Thus, in practice there is no loss of asymptotic power from adopting the robust inference approach if identification is strong.

3. TESTING FOR CONSTANCY OF THE RD EFFECT ACROSS COVARIATES

In this section, we develop a test of constancy of the RD effect across covariates, which is robust to weak identification issues. Such a test can be useful in practice when the econometrician wants to argue that the treatment effect is different for different population subgroups. For example, in Section 4, we use this test to argue that the effect of class sizes on educational achievements is different for secular and religious schools, and therefore it might be optimal to implement different rules concerning class sizes in those two categories of schools. The problem is related to the classical analysis of variance (ANOVA) hypothesis of homogenous populations (see, e.g., Casella and Berger 2002, chap. 11).

Similarly to Otsu, Xu, and Matsushita (2015), we consider the RD effect conditional on some covariate w_i . (See also Frölich 2007.) Let \mathcal{W} denote the support of the distribution of w_i . Next, for $w \in \mathcal{W}$ we define $y^+(w)$ using the conditional expectation given z_i and $w_i = w$: $y^+(w) = \lim_{z \downarrow z_0} E(y_i | z_i = z, w_i = w)$. Let $y^-(w)$, $x^+(w)$, and $x^-(w)$ be defined similarly. The conditional RD effect given $w_i = w$ is defined as $\beta(w) = (y^+(w) - y^-(w)) / (x^+(w) - x^-(w))$. Similarly to the case without covariates, under an appropriate set of assumptions, $\beta(w)$ captures the (local) ATE at z_0 conditional on $w_i = w$. We are interested in testing the null hypothesis of constancy of the RD effect

$$H_0 : \beta(w) = \beta \text{ for some } \beta \in \mathbb{R} \text{ and all } w \in \mathcal{W}, \quad (13)$$

against a general alternative $H_1 : \beta(w) \neq \beta(v)$ for some $v, w \in \mathcal{W}$. When identification is strong, the econometrician can esti-

mate the conditional RD effect function consistently and then use it for testing of H_0 . (Such a test can be constructed similarly to the ANOVA F -test as in Casella and Berger (2002, chap. 11) and is discussed in the supplement.) However, this approach can be unreliable if identification is weak. We therefore take an alternative approach.

Suppose that $\mathcal{W} = \{\bar{w}^1, \dots, \bar{w}^J\}$, that is, the covariate is categorical and divides the population into J groups. The assumption of a categorical covariate is plausible in many practical applications where the econometrician may be interested in the effect of gender, school type, etc. However, even when the covariate is continuous, in a nonparametric framework it might be sensible to categorize it to have sufficient power (as is often done in practice). For $j = 1, \dots, J$, let $\hat{y}_n^+(\bar{w}^j)$, $\hat{y}_n^-(\bar{w}^j)$, $\hat{x}_n^+(\bar{w}^j)$, and $\hat{x}_n^-(\bar{w}^j)$ denote the local linear estimators of the corresponding population terms computed using only the observations with $w_i = \bar{w}^j$. Let n_j be the number of such observations. Define $\sigma_y^2(\bar{w}^j)$, $\sigma_x^2(\bar{w}^j)$, and $\sigma_{xy}(\bar{w}^j)$ as the conditional versions of the corresponding population terms, and let $\hat{\sigma}_{y,n}^2(\bar{w}^j)$, $\hat{\sigma}_{x,n}^2(\bar{w}^j)$, and $\hat{\sigma}_{xy,n}(\bar{w}^j)$ denote the corresponding estimators.

Suppose that Assumption 2 holds for each of the J categories, and none of the categories is redundant asymptotically: $n_j h_{n_j} / (nh_n) \rightarrow p_j > 0$ for $j = 1, \dots, J$, where $n = \sum_{j=1}^J n_j$. If H_0 is true and the FRD effect is independent of w , one can construct a robust confidence set for the common effect: $CS_{1-\alpha,n}^J = \{\beta_0 \in \mathbb{R} : G_n(\beta_0) \leq \chi_{J,1-\alpha}^2\}$, where

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

$\hat{\beta}_n(\bar{w}^j) = \widehat{\Delta y}_n(\bar{w}^j) / \widehat{\Delta x}_n(\bar{w}^j)$, $\widehat{\Delta x}_n(\bar{w}^j) = \hat{x}_n^+(\bar{w}^j) - \hat{x}_n^-(\bar{w}^j)$; $\hat{\sigma}_n^2(\beta_0, \bar{w}^j)$ is defined similarly to $\hat{\sigma}_n^2(\beta_0)$ in (3) using the estimators conditional on $w_i = \bar{w}^j$; and $\hat{f}_{z,n}(z_0 | \bar{w}^j) = (n_j h_{n_j})^{-1} \sum_{i=1}^n K((z_i - z_0) / h_{n_j}) 1\{w_i = \bar{w}^j\}$ is the estimator for $f_z(z_0 | \bar{w}^j)$, which denotes the conditional density of z_i at z_0 conditional on $w_i = \bar{w}^j$.

Under $H_0 : \beta(w) = \beta$ for some $\beta \in \mathbb{R}$, $CS_{1-\alpha,n}^J$ is an asymptotically valid confidence set since $G_n(\beta) \rightarrow_d \chi_{J,1-\alpha}^2$ under weak or strong identification. We consider the following size α asymptotic test: Reject H_0 if $CS_{1-\alpha,n}^J$ is empty. The test is asymptotically valid because under H_0 , $P(CS_{1-\alpha,n}^J = \emptyset) \leq P(\beta \notin CS_{1-\alpha,n}^J) = P(G_n(\beta) > \chi_{J,1-\alpha}^2) \rightarrow \alpha$, which again holds under weak or strong identification. Under the alternative, there is no common value β that will provide a proper recentering for all J categories, and therefore, one can expect deviations from the asymptotic χ_J^2 distribution.

We show below that the test is consistent if there is strong (or semistrong) identification for at least two values \bar{w}^{j_1} and \bar{w}^{j_2} that satisfy $\beta(\bar{w}^{j_1}) \neq \beta(\bar{w}^{j_2})$. Let $d_{n,1}^2(\bar{w}^j) = n_j h_{n_j} |x^+(\bar{w}^j) - x^-(\bar{w}^j)|^2 f_z(z_0 | \bar{w}^j) / (k \sigma_x^2(\bar{w}^j))$ be the conditional version of the concentration parameter.

Theorem 4. Suppose that $n_j h_{n_j} / (nh_n) \rightarrow p_j > 0$ and Assumption 2 holds for each $j = 1, \dots, J$.

- a. Tests that reject H_0 of constancy in (13) when $CS_{1-\alpha,n}^J = \emptyset$ have $AsySz$ less or equal to α .

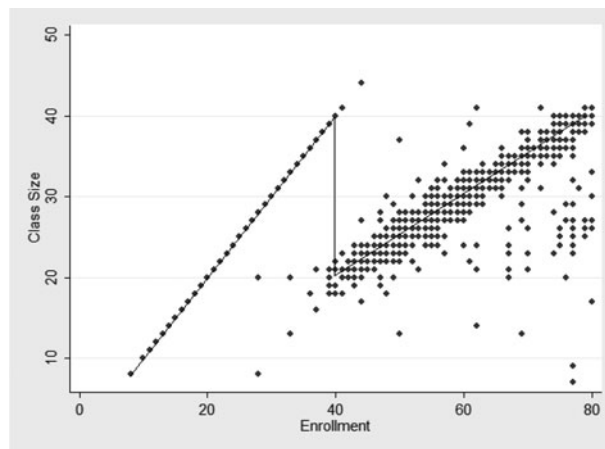


Figure 1. Angrist and Lavy (1999): Empirical relationship between class size and school enrollment.

Note: The solid line shows the relationship when Maimonides' rule (cap of 40 students) is strictly enforced.

- b. Let $\mathcal{W}^* = \{\bar{w}^1, \dots, \bar{w}^{J^*}\} \subset \mathcal{W}$ be such that $d_{n,1}^2(\bar{w}^j) \rightarrow \infty$ for $\bar{w}^j \in \mathcal{W}^*$ and $\beta(\bar{w}^{j_1}) \neq \beta(\bar{w}^{j_2})$ for some $\bar{w}^{j_1}, \bar{w}^{j_2} \in \mathcal{W}^*$. Then, $P(CS_{1-\alpha,n}^J = \emptyset) \rightarrow 1$ as $n \rightarrow \infty$.

4. EMPIRICAL APPLICATIONS

In this section, we compare the results of standard and weak identification robust inference in two separate, but related, applications. We show that the standard method and our proposed method yield significantly different conclusions when weak identification is a problem, but similar results when it is not. We also show that the robust confidence sets can provide more informative answers than the standard confidence intervals in cases when the usual assumptions are violated. We also apply our weak identification robust constancy test.

We begin with a case where weak identification is not a serious issue. In an influential article, Angrist and Lavy (1999) studied the effect of class size on academic success in Israel using the fact class size in Israeli public schools was capped at 40 students during their sample period. As demonstrated in Figure 1, this cap results in discontinuities in the relationship between class size and total school enrollment for a given grade. In practice, school enrollment does not perfectly predict class size and thus the appropriate design is fuzzy rather than sharp. We use the same sample selection rules as Angrist and Lavy (1999) and focus on language scores among 4th graders. The data can be found at <http://econ-www.mit.edu/faculty/angrist/data1/data/anglavy99>. There is a total of 2049 classes in 1013 schools with valid test results. Here, we only look at the first discontinuity at the 40-student cutoff. The number of observations used in the estimation depends on the bandwidth. It ranges from 471 classes in 118 schools for the smallest bandwidth (6), to 722 observations in 484 schools for the widest bandwidth (20). We use the uniform kernel in all cases.

Table 2 shows that the estimated discontinuity in the treatment variable ranges from 8 to 14 students depending on the

Table 2. Angrist and Lavy (1999): Estimated discontinuity in the treatment variable for the first cutoff and their standard errors, estimated effect of class size on class average verbal score, and standard and robust 95% confidence sets (CSs) for the class size effect for different values of the bandwidth

Bandwidth	Discont.	Std. errors	F-stat	Effect	Standard CS	Robust CS
6	-8.40	1.60	27.5	-0.07	[-0.145, 0.007]	[-0.170, -0.000]
8	-9.90	1.26	61.9	-0.07	[-0.129, -0.015]	[-0.138, -0.019]
10	-10.83	1.03	110.2	-0.06	[-0.103, -0.015]	[-0.103, -0.015]
12	-12.00	0.92	172.0	-0.02	[-0.056, 0.011]	[-0.058, 0.010]
14	-12.62	0.78	258.8	-0.03	[-0.061, 0.000]	[-0.062, -0.000]
16	-13.21	0.69	370.1	-0.02	[-0.048, 0.008]	[-0.049, 0.007]
18	-13.87	0.61	525.8	-0.02	[-0.046, 0.003]	[-0.047, 0.003]
20	-14.35	0.56	667.7	-0.02	[-0.042, 0.005]	[-0.043, 0.004]

NOTES: Silverman's normal rule-of-thumb bandwidth is 7.84 and the optimal bandwidth suggested by Imbens and Kalyanaraman (2012) is 7.90. The scores are given in terms of standard deviations from the mean.

bandwidth chosen. The table also shows that, as expected, the F -statistic becomes smaller as the bandwidth gets smaller. Silverman's normal rule-of-thumb and the optimal bandwidth procedure of Imbens and Kalyanaraman (2012) both suggest a bandwidth value of approximately 8, which corresponds to a relatively large value of the F -statistic (approximately 62). Applying the standards of Table 1, we then conclude that weak identification is not a serious concern in this application. Using the 5% noncentral χ^2 critical value, we reject the null hypothesis that the concentration parameter is below 36, and therefore, the maximal size distortions of the 5% two-sided tests are expected to be under 1%. Note that even at the smallest bandwidth, the F -statistic is relatively large. This is consistent with Figure 2 that shows that the 95% standard and robust confidence sets for the class size effect are very similar. The figure shows that the two sets of confidence intervals are essentially indistinguishable for larger bandwidths, and only differ slightly for smaller bandwidths.

In this application, we also compare the results of the standard constancy test of the treatment effect across subgroups to the results of our robust constancy test. The first set of results reported in Section 5 of the online supplement compares the

treatment effect for secular and religious schools. The null hypothesis (the treatment effect is the same across subgroups) can never be rejected using a standard test. By contrast, the robust constancy test rejects the null hypothesis for the largest values of the bandwidth (18 and 20). We reach similar conclusions when comparing the treatment effect for schools with above and below median proportions of disadvantaged students. The null hypothesis is rejected by the robust test under the largest bandwidth (20). This suggests that our proposed test may have greater power against alternatives than the standard test in some contexts.

The second application considers a similar policy in Chile originally studied by Urquiola and Verhoogen (2009). It should be noted that Urquiola and Verhoogen (2009) are not attempting to provide causal estimates of the effect of class size on tests score. They instead showed how the RD design can be invalid when there is manipulation around the cutoff, which results in a violation of Assumption 1(b) (exogeneity of z_i). So while this particular application is useful for illustrating some pitfalls linked to weak identification in an FRD design, the results should be interpreted with caution. In this application, the class sizes are capped at 45 students. Figure 3 shows the fuzzy discontinuity in the empirical relationship between class size

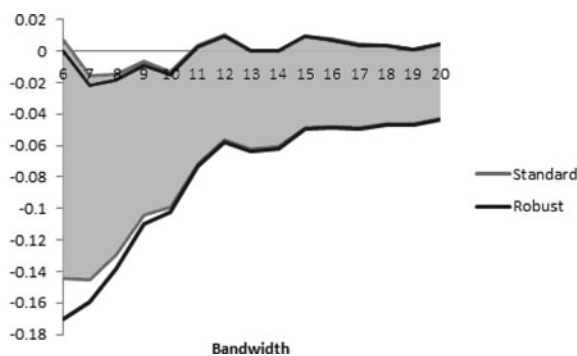


Figure 2. Angrist and Lavy (1999): 95% confidence intervals for the effect of class size on verbal test scores for different values of the bandwidth.

Note: This figure is for the enrollment cutoff of 40. The bandwidth according to Silverman's normal rule-of-thumb is 7.94. The optimal bandwidth selected according to Imbens and Kalyanaraman (2012) is 7.90. The scores are given in terms of standard deviations from the mean.

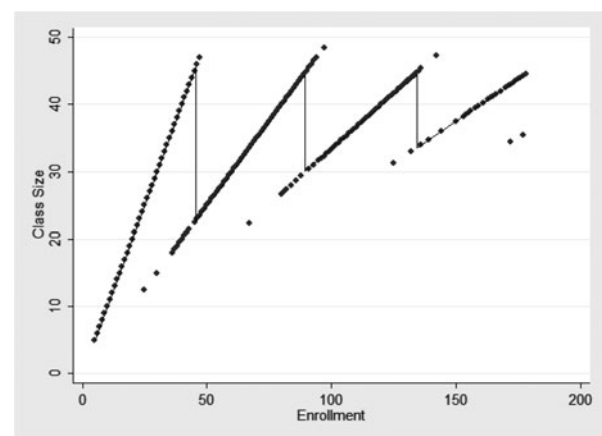


Figure 3. Urquiola and Verhoogen (2009): Empirical relationship between class size and enrollment.

Note: The solid line shows the relationship when the rule (cap of 45 students) is strictly enforced.

Table 3. Urquiola and Verhoogen (2009): The estimated effect of class size on the class average math score and its 95% standard and robust confidence sets (CSs) for different values of the bandwidth

Bandwidth	Estimated effect	Standard CS	Robust CS
First cutoff (45)			
6	0.146	[-0.061, 0.353]	$(-\infty, -0.433] \cup [0.043, \infty)$
8	3.378	[-74.820, 81.576]	$(-\infty, -0.120] \cup [0.129, \infty)$
10	-0.437	[-1.867, 0.993]	$(-\infty, -0.078] \cup [0.181, \infty)$
12	-0.173	[-0.360, 0.014]	[-1.720, -0.065]
14	-0.136	[-0.246, -0.026]	[-0.376, -0.060]
16	-0.091	[-0.153, -0.029]	[-0.186, -0.042]
18	-0.073	[-0.115, -0.031]	[-0.127, -0.037]
20	-0.063	[-0.099, -0.027]	[-0.107, -0.032]
Second cutoff (90)			
6	0.128	[-0.025, 0.281]	[0.004, 3.093]
8	0.261	[-0.061, 0.582]	$(-\infty, -0.587] \cup [0.085, \infty)$
10	0.227	[-0.111, 0.566]	$(-\infty, -0.241] \cup [0.046, \infty)$
12	0.306	[-0.296, 0.908]	$(-\infty, -0.118] \cup [0.053, \infty)$
14	0.486	[-1.092, 2.063]	$(-\infty, -0.056] \cup [0.068, \infty)$
16	1.636	[-18.745, 22.017]	$(-\infty, 0.002] \cup [0.065, \infty)$
18	-1.056	[-10.968, 8.856]	$(-\infty, \infty)$
20	-0.425	[-2.041, 1.190]	$(-\infty, 0.005] \cup [0.162, \infty)$
Third cutoff (135)			
6	-2.145	[-15.627, 11.336]	$(-\infty, -0.076] \cup [0.584, \infty)$
8	-0.298	[-0.692, 0.097]	[-21.482, 0.007]
10	-0.307	[-0.850, 0.236]	$(-\infty, 0.027] \cup [1.414, \infty)$
12	-0.309	[-0.861, 0.243]	$(-\infty, 0.027] \cup [1.550, \infty)$
14	-0.328	[-0.885, 0.228]	$(-\infty, -0.001] \cup [1.838, \infty)$
16	-0.231	[-0.652, 0.190]	$(-\infty, 0.034] \cup [1.604, \infty)$
18	-0.181	[-0.500, 0.138]	$(-\infty, 0.041] \cup [21.933, \infty)$
20	-0.136	[-0.389, 0.117]	[-1.642, 0.063]
Fourth cutoff (180)			
10	0.048	[-0.119, 0.216]	$(-\infty, \infty)$
12	0.035	[-0.130, 0.200]	$(-\infty, \infty)$
14	-0.047	[-0.371, 0.278]	$(-\infty, \infty)$
16	-0.045	[-0.343, 0.254]	$(-\infty, \infty)$
18	-0.039	[-0.316, 0.238]	$(-\infty, \infty)$
20	-0.029	[-0.299, 0.242]	$(-\infty, \infty)$

NOTES: Silverman's rule-of-thumb bandwidth is 8.59. The optimal bandwidth suggested by Imbens and Kalyanaraman (2012) for the cutoff of 45 is 9.67 and for the cutoff of 90, the suggested bandwidth is 11.60. The optimal bandwidth suggested by Imbens and Kalyanaraman (2012) for the cutoff of 135 is 14.12 and for the cutoff of 180, the suggested bandwidth is 17.81. The scores are given in terms of standard deviations from the mean.

and enrollment at the various multiples of 45. The figure also shows that the discontinuity becomes smaller as enrollment increases. In this example, the outcome variable is average class scores on state standardized math exams and we restrict attention to 4th graders. We also strictly adhere to the sample selection rules used by Urquiola and Verhoogen (2009). The total number of observations is 1636. The effective number of observations varies with the bandwidth and the enrollment cutoff. The range of the number of observations is 201 to 402 at the 90 student enrollment cut off; 45 to 95 at the 135 student enrollment cutoff, and 17 to 34 at the 180 student enrollment cutoff. The uniform kernel is used to compute all the results below. Table 3 reports the FRD estimates and the confidence sets for the different values of the bandwidth and cutoff points. As before, we set the size of the test at 5%. Starting with the first cutoff point, Table 3 shows that the robust and conventional confidence sets diverge dramatically as the bandwidth gets smaller. Interestingly, while the robust confidence interval is much wider than the conventional one, it nevertheless rejects the null hypothesis that the

effect of class size is equal to zero while the conventional fails to reject the null.

To help interpret the results, we also graphically illustrate the difference between standard and robust confidence sets in Figure 4. The first panel plots the standard confidence sets as a function of the bandwidth. The second panel does the same for the weak identification robust method. The shaded area is the region covered by the confidence sets. As the bandwidth increases, the robust confidence sets evolve from two disjoint sections of the real line to a well-defined interval. Note that class size is a discrete rather than a strictly continuous variable, hence the break between bandwidths 11 and 12 when the robust confidence set switches from two disjoint half lines to a single interval. This is consistent with the size of the discontinuity in class size as a function of enrollment estimated at different bandwidths and the corresponding F -statistic. At bandwidths below 10, the estimated discontinuity is small and the F -statistic is below 7. However at bandwidths higher than 12, the estimated discontinuity is progressively closer to 10 students and the

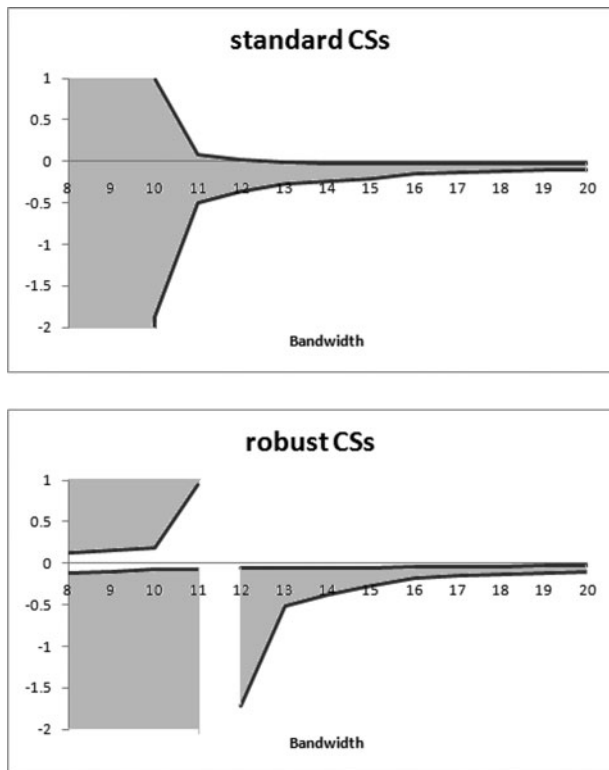


Figure 4. Urquiola and Verhoogen (2009): 95% standard and robust confidence sets (CSs) for the effect of class size on class average math score for different values of the bandwidth.

Note: This figure is for the first enrollment cutoff of 45. The bandwidth according to Silverman's normal rule-of-thumb is 8.59. The optimal bandwidth selected according to Imbens and Kalyanaraman (2012) is 9.67. The scores are given in terms of standard deviations from the mean.

F -statistic ranges from just over 40 to just over 188. This is important since the bandwidth suggested by Silverman's normal rule-of-thumb is only 8.59 and the optimal bandwidth suggested by Imbens and Kalyanaraman (2012) is 9.67. See Section 5 in the online supplement for a complete listing of the F -statistic and discontinuity estimates at different bandwidths.

Identification is considerably weaker for the second cutoff point. At all bandwidths, the standard confidence intervals fail to reject the null that the effect of class size is zero. However, for most bandwidths, the robust confidence sets do not include a zero effect. For example, for a bandwidth of 8, we cannot reject the null that class size is not related to grades when using the standard method, while the robust method suggests rejecting the null.

Identification is even weaker at the third cutoff and, for most bandwidths, the robust confidence sets consist of two disjoint intervals. Finally, results get very imprecise at the fourth cutoff and the robust confidence sets now map the entire real line. This suggests that identification is very weak at these levels and the standard confidence intervals are overly narrow.

In summary, our results suggest that when weak identification is not a problem, the robust and standard confidence sets are similar. But when the discontinuity in the treatment variable is not large enough, the robust confidence sets are very different from those obtained using the standard method. We also

demonstrate that our robust inference method provides more informative results than the standard method.

SUPPLEMENTARY MATERIALS

The supplementary materials contain: (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.

ACKNOWLEDGMENTS

The authors thank their editor, Shakeeb Khan, an associate editor, and two anonymous referees for very helpful comments. The authors also thank Chris Muris, Moshe Buchinsky, Karim Chalak, Jae-Young Kim, Sokbae Lee, Arthur Lewbel, Taisuke Otsu, Eric Renault, and Yoon-Jae Whang for their comments on early drafts of the article. Vadim Marmer gratefully acknowledges the financial support of the SSHRC under grants 410-2010-1394 and 435-2013-0331.

[Received November 2012. Revised February 2015.]

REFERENCES

- Anderson, T. W., and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63. [186,190]
- Andrews, D. W. K., and Cheng, X. (2012), "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211. [188]
- Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2011), "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper 1813. Cowles Foundation for Research in Economics, Yale University. Available online at <http://cowles.econ.yale.edu/P/cd/d18a/d1813.pdf> [185,188,189]
- Andrews, D. W. K., and Guggenberger, P. (2010), "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory*, 26, 426–468. [188]
- Andrews, D. W. K., Moreira, M. J., and Stock, J. H. (2006), "Optimal Invariant Similar Tests For Instrumental Variables Regression," *Econometrica*, 74, 715–752. [186,191]
- Andrews, D. W. K., and Stock, J. H. (2007), "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society* (Vol. III), eds. R. Blundell, W. K. Newey and T. Persson, Cambridge, UK: Cambridge University Press, pp. 122–173. [185,186]
- Angrist, J. D., and Lavy, V. (1999), "Using Maimonides' Rule to Estimate The Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575. [192]
- Antoine, B., and Lavergne, P. (2014), "Conditional Moment Models Under Semi-Strong Identification," *Journal of Econometrics*, 182, 59–69. [188]
- Antoine, B., and Renault, E. (2009), "Efficient GMM with Nearly-Weak Instruments," *Econometrics Journal*, 12, S135–S171. [188]
- (2012), "Efficient Minimum Distance Estimation with Multiple Rates of Convergence," *Journal of Econometrics*, 170, 350–367. [188]
- Arai, Y., and Ichimura, H. (2013), "Optimal Bandwidth Selection for Differences of Nonparametric Estimators with an Application to the Sharp Regression Discontinuity Design," Working Paper No. 13-09, National Graduate Institute for Policy Studies. [185]
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450. [185]
- Buddelmeyer, H., and Skoufias, E. (2004), "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," World Bank Policy Research Working Paper 3386, World Bank Publications. [185]
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Bias-Corrected Inference in the Regression Discontinuity Design," *Econometrica*, 82, 2295–2326. [185]

- Caner, M. (2009), "Testing, Estimation in GMM and CUE with Nearly-Weak Identification," *Econometric Reviews*, 29, 330–363. [188]
- Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Duxbury Press. [191]
- Davidson, J. (1994), *Stochastic Limit Theory*, New York: Oxford University Press. [187]
- Dong, Y., and Lewbel, A. (2010), "Regression Discontinuity Marginal Threshold Treatment Effects," Working Paper No. 759, Boston College. [185]
- Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387. [186]
- Fe, E. (2012), "Efficient Estimation in Regression Discontinuity Designs via Asymmetric Kernels," MPRA Working Paper No. 38164, Munich Personal RePEc Archive. Available at <http://mpra.ub.uni-muenchen.de/38164/>. [185]
- Frölich, M. (2007), "Regression Discontinuity Design with Covariates," Working Paper 2007-32, University of St. Gallen. [185,191]
- Frölich, M., and Melly, B. (2008), "Quantile Treatment Effects in the Regression Discontinuity Design," IZA Discussion Paper 3638, Institute for the Study of Labor. [185]
- Hahn, J., and Kuersteiner, G. (2002), "Discontinuities of Weak Instrument Limiting Distributions," *Economics Letters*, 75, 325–331. [188]
- Hahn, J., Todd, P., and Van der Klaauw, W. (1999), "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design," NBER Working Paper 7131, National Bureau of Economic Research. [185,187]
- (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209. [185,187]
- Imbens, G. W., and Kalyanaraman, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959. [185,193,194,195]
- Imbens, G. W., and Lemieux, T. (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635. [185,187]
- Imbens, G. W., and Manski, C. F. (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [188]
- Kleibergen, F. (2002), "Pivotal Statistics For Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803. [186]
- Lee, D. S., and Lemieux, T. (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355. [185]
- McCrary, J. (2008), "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714. [185]
- Mikusheva, A. (2007), "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [188]
- Mills, B., Moreira, M. J., and Vilela, L. P. (2014), "Tests Based on t -Statistics for IV Regression with Weak Instruments," *Journal of Econometrics*, 182, 351–363. [191]
- Moreira, M. J. (2001), "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," unpublished manuscript, Department of Economics, University of California, Berkeley. [186,191]
- (2003), "A Conditional Likelihood Ratio Test For Structural Models," *Econometrica*, 71, 1027–1048. [186,191]
- Otsu, T., Xu, K.-L., and Matsushita, Y. (2015), "Empirical Likelihood for Regression Discontinuity Design," *Journal of Econometrics*, 186, 94–112. [185,186,191]
- Papay, J. P., Willett, J. B., and Murnane, R. J. (2011), "Extending the Regression-Discontinuity Approach to Multiple Assignment Variables," *Journal of Econometrics*, 161, 203–207. [185]
- Porter, J. (2003), "Estimation in the Regression Discontinuity Model," Working Paper, University of Wisconsin-Madison. [185]
- Staiger, D., and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [185,188,190]
- Stock, J. H., Wright, J. H., and Yogo, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, [185]
- Stock, J. H., and Yogo, M. (2005), "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews and J. H. Stock, New York: Cambridge University Press, pp. 80–108. [186,189]
- Urquiola, M., and Verhoogen, E. (2009), "Class-Size Caps, Sorting, and the Regression-Discontinuity Design," *American Economic Review*, 99, 179–215. [193,195]
- Van der Klaauw, W. (2008), "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics," *Labour*, 22, 219–245. [185]
- Zajonc, T. (2012), "Regression Discontinuity Design with Multiple Forcing Variables," in *Essays on Causal Inference for Public Policy*, Chapter 2, Doctoral Dissertation, Harvard University, pp. 45–81. Available at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:9368030> [185]