Econometrica Supplementary Material

SUPPLEMENT TO "A TEST OF EXOGENEITY WITHOUT INSTRUMENTAL VARIABLES IN MODELS WITH BUNCHING"

(*Econometrica*, Vol. 83, No. 4, July 2015, 1581–1600)

BY CAROLINA CAETANO

This supplement provides details of the implementation of the test statistic described in Section 3.1 in "A Test of Exogeneity Without Instrumental Variables in Models With Bunching." It also develops the theorems that describe the test statistic's asymptotic behavior. Finally, it presents a Monte Carlo study of the small sample behavior of the test statistic using real data (the same data set used in the paper's Section 3).

S1. A TEST STATISTIC BASED ON θ

AS DESCRIBED IN THE PAPER in Section 3.2, the empirical application is based on the parameter

$$\theta = \lim_{x\downarrow 0} \mathbb{E}\big[\mathbb{E}[Y|X=0, Z] - Y|X=x\big]. \tag{S1}$$

In order to estimate θ, a simple two-step process is suggested, which consists of first estimating the term $\mathbb{E}[Y|X=0, Z]$ and then estimating the outer limit as $x \downarrow 0$. The requirements of the approach are the following.

ASSUMPTION S1: *Suppose that*
1. $0 < \mathbb{P}(X=0) < 1$.
2. $\mathbb{E}[Y|X=0, Z] = Z'\gamma$.
3. *The sample* $\{(Y_i, X_i, Z_i')\}_{i=1^n}$ *is independent*.
4. $\mathbb{V}\text{ar}[Y|X=0, Z]$ *is finite and uniformly bounded*.

The first requirement permits that the sample be divided between observations such that $X=0$, and $X \neq 0$. This requirement can be relaxed so that the test can still be applied to cases in which there is no bunching (see Remark S1.1). The other requirements are there to guarantee that $\mathbb{E}[Y|X=0, Z]$ is estimable at the $\sqrt{n}$ rate, but they have an effect on the null hypothesis (see Remark S1.5).

If Assumption S1(2) holds, then $\theta = \lim_{x\downarrow 0} \mathbb{E}[Z'\gamma - Y|X=x]$. The estimation of γ is done with an OLS regression of Y onto Z using only observations such that $X=0$. The outer limit in θ is a boundary quantity, and so the limit estimator needs to take this into account. The issues with nonparametric boundary estimation are well known and addressed extensively in the Regresson Discontinuity Design (RDD) literature (e.g., Hahn, Todd, and Van der Klaauw (2001), Porter (2003), and Imbens and Lemieux (2008)). The quantity in (S1) is of the same nature as that in the RDD, with the only difference being that the dependent variable in the regression, $Z'\gamma - Y$, has to be estimated. However, Assumption S1 guarantees that γ is estimated at the rate of $\sqrt{n}$, whereas

 DOI: 10.3982/ECTA11231

the fastest rate possible for the nonparametric estimation of the outer limit is $n^{2/5}$ (see Theorem 5 in Porter (2003)). Therefore, asymptotically it is irrelevant whether the estimate $\hat{\gamma}$ or the true γ was used, and so it is possible to proceed identically to the RDD literature by simply substituting $Z'\hat{\gamma} - Y$ in place of the dependent variable.

The recommended approach is to do a local linear regression of the terms $Z'\hat{\gamma} - Y$ onto X at $X = 0$, using only observations such that $X > 0$. A detailed description of the procedure can be seen in Fan and Gijbels (1992), though practitioners may benefit from the exposition in Imbens and Lemieux (2008). The practitioner must choose a kernel K[1] and a bandwidth h.[2] The procedure is equivalent to the estimation of the constant in a weighted least squares regression:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \mathbf{1}(X_i > 0) K(X_i/h) \big(Z_i'\hat{\gamma} - Y_i - \theta - \beta X_i\big)^2.$$

For convenience, the following assumption reproduces the assumptions of Hahn, Todd, and Van der Klaauw (2001) for the convergence of the boundary local linear estimator in the context of this paper.

ASSUMPTION S2: *Suppose that there exists an interval* $I := (0, \delta]$ *for some* $\delta > 0$ *such that*

1. X *is continuously distributed in* I. *Its density* f *is twice continuously differentiable and bounded away from zero. The limit* $f_{Xr} = \lim_{x \downarrow 0} f(x)$ *exists and is positive*.
2. $\mathbb{E}[\mathbb{V}\mathrm{ar}[Y|X, Z]|X = x]$ *and* $\mathbb{V}\mathrm{ar}[\Delta(x, Z)|X = x]$ *are continuous in* I, *and* $\sigma^2_{Yr} := \lim_{x\downarrow 0} \mathbb{E}[\mathbb{V}\mathrm{ar}[Y|X, Z]|X = x]$ *and* $\sigma^2_{\Delta r} := \lim_{x\downarrow 0} \mathbb{V}\mathrm{ar}[\Delta(x, Z)|X = x]$ *are finite*.
3. *For some* $\zeta > 0$, $\mathbb{E}[|Y - \mathbb{E}[Y|X = 0, Z]|^3|X = x]$ *is uniformly bounded in* I, *with bounded right-limit at zero*.
4. *The kernel* $K(\cdot)$ *is continuous, symmetric, nonnegative-valued, and with compact support*.
5. *The bandwidth* $h \propto n^{-s}$, *with* $1/5 < s < 2/5$.

The assumptions of the local linear estimator are well known. In this paper, item (1) deserves special notice. It indirectly requires that there cannot be any bunching points in I. This can be an excessive imposition in certain examples (such as in the case of labor supply discussed in Section 2.2 in the paper).

[1]Imbens and Lemieux (2008) recommended the use of the rectangular kernel $K(u) = \frac{1}{2}\mathbf{1}(|u| < 1)$, although the Epanechnikov kernel $K = \frac{3}{4}(1 - u^2)\mathbf{1}(|u| < 1)$ possesses a few optimality advantages. In any case, the choice of the kernel is not very influential, except perhaps when the curve is excessively convex or concave near the boundary.

[2]See Remark S1.3.

Remark S1.2 discusses how this condition can be relaxed. The faster bandwidth requirement (5) allows the bias term to vanish. The following result establishes the asymptotic behavior of $\hat{\theta}$.

THEOREM S1: *Suppose that Assumptions* S1 *and* S2 *hold*. *Then*

$$\sqrt{nh}(\hat{\theta}-\theta) \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \int K(u)^2\,du \cdot (\sigma^2_{Yr} + \sigma^2_{\Delta r})/f_{Xr}$.

PROOF: $\hat{\theta}$ can be written as a linear combination $\sum_{i=1}^n w_i(Z_i'\hat{\gamma} - Y_i)$ (see Heckman, Ichimura, and Todd (1998, p. 284)). Thus,

$$\hat{\theta} = \left[\sum_{i=1}^n w_i\right](\hat{\gamma}-\gamma) + \sum_{i=1}^n w_i \Delta(X_i, Z_i) + \sum_{i=1}^n w_i\big(Y_i - \mathbb{E}[Y_i|X_i, Z_i]\big).$$

$\sum_{i=1}^n w_i$ is bounded in probability (see Porter (2003, p. 45)). Assumption S1 guarantees that $\hat{\gamma} - \gamma = o_p(\sqrt{nh})$. The second and third terms are typical local linear estimators and are also uncorrelated. The result is then a straightforward application of Theorem 4 in Hahn, Todd, and Van der Klaauw (2001). *Q.E.D.*

The variance term Ω can be estimated with a plug-in method. Under $\mathbb{H}_0$, $\sigma^2_{\Delta r} = 0$. Both f_{Xr} and σ^2_{Yr} can be estimated directly (see Imbens and Lemieux (2008, p. 630)). The test statistic that uses this variance estimator can be easily implemented using a well-known software package.[3] The convergence in probability of this estimator is a simple application of Theorem 4 in Porter (2003).

The test statistic is thus $T_n := \hat{\theta}/\sqrt{\hat{\Omega}/nh}$, which should be compared to the critical values of the standard normal distribution.

THEOREM S2: *Suppose that Assumptions* 2.1 *and* 2.2 *in the paper hold*. *Additionally suppose that Assumptions* S1 *and* S2 *hold*, *and that* $\hat{V} \xrightarrow{p} V$. *Then under* $\mathbb{H}_0$ (*i.e.*, X *is exogenous*),

$$\lim_{n\to\infty} \mathbb{P}\big(|T_n| > c_{1-\alpha/2}\big) = \alpha.$$

Suppose additionally that Assumption 2.3 *in the paper holds with positive probability*, *so that under* $\mathbb{H}_1$ (*i.e.*, X *is endogenous*), $\theta \neq 0$. *Then*

$$\lim_{n\to\infty} \mathbb{P}\big(|T_n| > c_{1-\alpha/2}\big) = 1.$$

[3]Command "lpoly" in Stata. Alternatively, the author can provide the ready-to-use Stata code of the test upon request.

And if there exists a constant δ *such that* $\sqrt{nh}\theta \to \delta$,[4] *then*

$$\lim_{n\to\infty} \mathbb{P}\big(|T_n| > c_{1-\alpha/2}\big) = 1 - \Phi(c_{1-\alpha/2} - \delta/\sqrt{\Omega}),$$

where Φ *denotes the standard normal cumulative distribution function and* $c_{1-\alpha/2}$ *its* $(1-\alpha/2)\cdot 100$*th critical value*.

PROOF: Follows directly from Theorem S1. *Q.E.D.*

REMARK S1.1—No Bunching Point: In cases where there is no bunching but the discontinuity happens at an interior point (see the working hours example in Section 2.2 in the paper), the test can still be performed with minimum modifications. In such cases, $\mathbb{E}[Y|X=0, Z]$ can be substituted by the limit $\lim_{x\uparrow 0}\mathbb{E}[Y|X=x, Z]$. Unfortunately, without further assumptions about the shape of $\mathbb{E}[Y|X=x, Z]$ when $x<0$, this estimation step may affect the asymptotic variance. If it is possible to assume, for example, that $\mathbb{E}[Y|X, Z] = X\alpha + Z'\gamma$ for $X<0$, then γ can be estimated with an OLS regression of Y onto X and Z using only observations such that $X<0$, and all the results hold the same.

REMARK S1.2—Bunching Inside I: If this is the case, the bunching points can simply be eliminated from the estimation. For example, in the working hours example in the paper Section 2.2, if the bandwidth is $h=16$, one could eliminate the data such that $X = 25$, 30, and 35 on the left side. Accounting for this generalization in the theory does burden the notation, but the results hold exactly the same provided the distribution of X is differentiable outside the bunching points.

REMARK S1.3—Bandwidth Selection: In practice, if $\mathbb{E}[Z'\gamma - Y|X=x]$ is very concave or convex near the boundary, the choice of the bandwidth is of paramount importance. In order to optimize this procedure, it is advisable to apply a method of bandwidth selection that is appropriate for boundary estimation. The method developed in Imbens and Kalyanaraman (2012) for the RDD can be transplanted to this test, but it yields the optimal bandwidth. Assumption S2(5) requires undersmoothing, and thus one should choose a bandwidth which is at least slightly smaller than the one given by that method. Nevertheless, it is still advisable to report results for several bandwidths.

REMARK S1.4—A Simplified Test: If K is the rectangular kernel, $\hat{\theta}$ has a very simple interpretation. Restricting the sample to observations such that

[4] A data generating process which generates this is, for example, if $\mathbb{E}[U|X=x, Z=z] = \lambda_1(x)a_n + \lambda_2(z)$, where there exists a constant C such that $\sqrt{nh}a_n \to C$. Then $\theta = a_n(\lambda_1(0) - \lim_{x\downarrow 0}\lambda_1(x))$, and $\delta = C(\lambda_1(0) - \lim_{x\downarrow 0}\lambda_1(x))$.

$0 < X_i < h$, $\hat{\theta}$ can be understood as the intercept of an OLS regression of the $Z_i'\hat{\gamma} - Y_i$ onto X_i. Therefore, T_n is the t-statistic of the intercept of the OLS regression described above. If the assumptions are relaxed to allow for heteroskedasticity, the robust variance estimator should be used instead.

REMARK S1.5—Effect of Assumption S1 on $\mathbb{H}_0$: One should keep in mind that Assumption S1(2) becomes part of $\mathbb{H}_0$. Hence, the test is now a test both of exogeneity and of the validity of this assumption (both with respect to the linearity and to the correct covariate specification). If the researcher intends to estimate a model which assumes linearity in covariates, this is in fact desirable. For example, if the model is $Y = m(X) + U$, with $\mathbb{E}[U|X, Z] = Z'\lambda$ (which could be estimated using a method such as in Robinson (1988)), then in this case $\mathbb{H}_0 : \mathbb{E}[U|X, Z] = Z'\lambda$.

S2. MONTE CARLO STUDY

This section studies the test statistic presented in the previous section in the context of Monte Carlo simulations using real data. Consider the following model:

$$Y = \beta X + Z'\gamma + U,$$

where Y is the baby's birth weight, X is the average daily number of cigarettes smoked during pregnancy, Z is the covariate specification used in the application Section 3, and U is unobservable. Suppose that

$$U = \delta Q + \varepsilon,$$

where

$$\begin{pmatrix} Q \\ \varepsilon \end{pmatrix} \Big| Z \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_Q^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}\right). \tag{S2}$$

Assume that $\mathbb{E}[\varepsilon|X, Z] = 0$. The variable Q is the source of the endogeneity of X, because it determines the optimal smoking choice X^* in the equation

$$X^* = Z'\pi + Q. \tag{S3}$$

Though the optimal choice X^* could be negative, the actual smoking amount X cannot. Therefore, the actual smoking choice is subject to a corner restriction

$$X = \max\{X^*, 0\}. \tag{S4}$$

Notice that this is not a censoring model in the strictest sense. The structural equation is not a function of the latent intended smoking amount, X^*, but of the observed actual smoking amount X. If there is no endogeneity ($\delta = 0$), β and γ can be estimated even if the censoring is ignored.

The expected birth weight conditional on observables is

$$\mathbb{E}[Y|X,Z] = (\beta+\delta)X + Z'(\gamma - \pi\delta) + \delta\big(Z'\pi - \sigma_Q\lambda(Z)\big)\mathbf{1}(X=0), \tag{S5}$$

where $\lambda(Z) = \phi(-Z'\pi/\sigma_Q)/\Phi(-Z'\pi/\sigma_Q)$ is the inverse Mills ratio. When $\delta \neq 0$, the endogeneity generates a discontinuity in birth weight as a function of smoking.

This model completely specifies the endogeneity. Although it does not explain the essence of Q, it establishes how this "black box" interacts with X. If the objective is the identification of β, this model may be arguably sensible.

The simulations use the same data set as in the application Section 3 in the paper. The parameters π and σ_Q^2 are estimated using a Tobit regression on equation (S3). The estimates are then plugged into equation (S5), where a simple OLS regression yields estimates for β, γ, δ, and σ_ε^2.[5] It is interesting to observe that, under this model, the estimate of β is 0.0073, with a 95% confidence interval of $[-0.57, 0.71]$. The average smoker mother smokes 12.9 cigarettes. If the model is true, then smoking causes an average loss of at most 7.3 grams (-0.57×12.9), which is not an important amount under any circumstances.

Disregarding endogeneity has severe consequences. Conditional on covariates, the difference in birth weight between mothers that smoke $X = x$ and those that do not smoke is $(\beta+\delta)x - \delta\mathbb{E}[Z'\pi - \sigma_Q\lambda(Z)|X = x]$. If the entire difference is erroneously attributed to the causal effect of smoking, the bias of endogeneity is $\delta(x - \mathbb{E}[Z'\pi - \sigma_Q\lambda(Z)|X = x])$. In the original sample, the average smoker has a bias of 191.7 grams, which accounts for almost all the effect found in Almond, Chay, and Lee (2005).

The experiments study the test behavior for three sample sizes: 1000, 5000, and 20,000. For this, three subsamples of corresponding sizes are randomly drawn from the original population. The samples are representative, as can be verified in Table S-I. This table shows the summary statistics of a group of the most important covariates for each of the subsamples, which can be compared to the values in the original data set.

In the experiments, the endogeneity levels are characterized as the correlation between U and Q, and denoted ρ.[6] However, in order to give the correlation level a tangible interpretation, Table S-II also reports the value of the

[5]The coefficient of the discontinuity term estimates δ. Then, β equals the coefficient of X, minus δ, and γ equals the coefficient of Z, plus $\pi\delta$. Obtaining the variance σ_ε^2 is more delicate, since when $X > 0$, $Y - \mathbb{E}[Y|X,Z] = \varepsilon$, but the same is not true when $X = 0$. Thus, $\hat\sigma_\varepsilon^2 = [\sum_{i=1}^n [Y_i - (\widehat{\beta+\delta})X_i - Z'(\widehat{\gamma-\pi\delta})]^2 \mathbf{1}(X_i > 0)]/[\sum_{i=1}^n \mathbf{1}(X_i > 0) - d - 2]$ is the average of the squared residuals of the smoker observations, where d is the dimension of Z.

[6]The endogeneity is defined as $\rho = \mathbb{C}\mathrm{orr}(U,Q) = \delta\sigma_Q/\sqrt{\delta^2\sigma_Q^2 + \sigma_\varepsilon^2}$. For each endogeneity level ρ, the equation above determines the corresponding value of δ.

TABLE S-I

DESCRIPTIVE STATISTICS OF KEY VARIABLES ACROSS SAMPLES

	Original Data Set	Subsamples		
	$n = 488{,}152$	$n = 1000$	$n = 5000$	$n = 20{,}000$
Outcome Variables				
Birth weight (in grams)	3366.17	3380.96	3371.79	3368.65
	(590.25)	(580.04)	(603.68)	(589.67)
Gestation (in weeks)	39.22	39.25	39.26	39.24
	(2.63)	(2.52)	(2.62)	(2.63)
APGAR (5 minutes)	9.03	9.03	9.02	9.03
	(0.79)	(0.75)	(0.80)	(0.80)
Mother's Variables				
Proportion of smokers	19.69	19.84	18.89	19.74
Avg. cigarettes per smoker	12.93	12.10	12.94	13.06
	(8.00)	(6.89)	(8.18)	(8.09)
Age	26.79	26.50	26.77	26.82
	(5.65)	(5.70)	(5.63)	(5.63)
Proportion of blacks	0.152	0.154	0.153	0.152
	(0.359)	(0.361)	(0.360)	(0.359)
Years of education	12.79	12.84	12.81	12.79
	(2.18)	(2.24)	(2.20)	(2.19)
Proportion of unmarried	0.293	0.292	0.288	0.293
	(0.455)	(0.455)	(0.453)	(0.455)
Prenatal visits	2.64	2.60	2.62	2.62
	(1.52)	(1.46)	(1.52)	(1.50)
Father's Variables				
Age	29.42	29.12	29.34	29.46
	(6.38)	(6.36)	(6.33)	(6.35)
Proportion of blacks	0.154	0.153	0.154	0.152
	(0.361)	(0.360)	(0.361)	(0.359)
Years of education	12.98	13.08	12.99	12.96
	(2.28)	(2.39)	(2.27)	(2.29)

average bias in the identified effect which corresponds to each endogeneity level ρ: $\delta\mathbb{E}[X - Z'\pi - \sigma_Q\lambda(Z)]$. The values can be seen in the second row of Table S-II. This is a best case scenario bias, because it assumes that the researcher has the correct specification of the functional form of equation (S5). If the researcher misspecifies the functional form in any way, for example by running a simple OLS regression of Y on X and Z, then the resulting effects will be biased not only because of the endogeneity, but also because of the misspecification.

TABLE S-II

PERCENTAGE OF REJECTIONS (5% SIGNIFICANCE, 10,000 REPETITIONS)

		Size	Power				
	Endogeneity (ρ):	0	−0.1	−0.25	−0.5	−0.75	−0.90
	Bias (in grams):	0	−73	−187	−418	−822	−1,497
n	h						
1000	3	5.6	5.8	7.1	13.1	25.4	39.0
($\approx$ 200 with $X > 0$)	7	4.6	6.0	8.9	21.3	41.9	55.5
	13	9.3	10.4	15.4	34.5	54.4	62.9
5000	3	4.7	9.4	31.7	82.6	93.3	95.9
($\approx$ 1000 with $X > 0$)	7	4.6	13.5	58.6	92.3	96.2	97.7
	13	6.6	22.2	76.4	94.4	97.2	98.5
20,000	3	5.1	26.2	89.0	98.7	99.7	99.9
($\approx$ 4000 with $X > 0$)	7	5.5	50.9	96.8	99.2	99.8	100.0
	13	5.5	71.6	98.3	99.5	100.0	100.0

Each experiment is characterized by a sample size n, an endogeneity level ρ, and a bandwidth h, and consists of 10,000 repetitions of the same procedure. For instance, take the case of $n = 1000$. For each repetition, 1000 vectors (Q_i, ε_i) are drawn from the distribution in (S2). Then the corresponding triad $(Z_i, Q_i, \varepsilon_i)$ generates X_i from equations (S3) and (S4). Y_i is generated using the structural equation with the δ which corresponds to the endogeneity level ρ. After the entire sample $\{(Y_i, X_i, Z_i); i = 1, \dots, 1000\}$ is generated, the test is performed at the 5% significance level, using the Epanechnikov kernel, bandwidth h, and polynomial degree $p = 1$.[7] The reported numbers in Table S-II are the percentages of rejections out of the 10,000 repetitions.

It is important to notice that only around 20% of the observations are smokers. Hence, when the sample has 1000 observations, only somewhere around 200 of those can be counted for the estimation of the nonparametric component of the test. Therefore, the first set of results reflects a very small sample even for the standards of parametric estimators. The test has no severe size distortion at the smaller bandwidths, but the bias due to the use of the larger bandwidth causes the test to be oversized when $h = 13$. The 5000 observation experiments count with about 1000 smoker observations to be used in the nonparametric component of estimation, which is a quantity much better aligned

[7]The experiments were also run for $p = 3$, with qualitatively similar results. For the combination of $n = 1000$ and bandwidth $h = 13$, $p = 3$ helped decrease the bias. In other respects, $p = 3$ proved inferior to $p = 1$. The cubic polynomial is more unstable at the boundary, since it can react rather harshly to outliers at the end of the support of the kernel. This is a consequence of Runge's phenomenon that higher order polynomial fits have very large variability at the boundary. Although in essence a local polynomial estimator is not a polynomial fit, it behaves like that in finite samples. Monte Carlo results for $p = 3$ are available by request from the author.

with the types of data requirements in nonparametric estimation. As expected, there are still size distortions for the higher bandwidths, though the distortion decreases. The 20,000 observation samples have around 4000 smoker observations, and perform significantly better. The smallest bandwidth has almost no size distortion at all.

For the smallest samples, the test has power, although it is not large unless the endogeneity is very high. The 5000 observation experiments show that the test performs excellently when the endogeneity generates a bias of −418 grams. The 20,000 observation sample has excellent power for a much smaller endogeneity bias of −187 grams. To put these numbers in perspective, observe that the average birth weight among nonsmoker mothers is 3428 grams, and the consensus in the medical literature is that the baby should not weigh less than 2500 grams. Another way to understand the power of the test is to consider that Almond, Chay, and Lee (2005) found that smoking causes a loss of around 200 grams. However, if smoking has no effect, and the number they found is due entirely to endogeneity (as is indeed the case if the simulation model is correct), then with 20,000 observations, this test would detect this problem over 90% of the times.

REFERENCES

ALMOND, D., K. Y. CHAY, AND D. S. LEE (2005): "The Costs of Low Birth Weight," *Quarterly Journal of Economics*, 120 (3), 1031–1083. [6,9]

FAN, J., AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20 (4), 2008–2036. [2]

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69 (1), 201–209. [1-3]

HECKMAN, J. J., H. ICHIMURA, AND P. TODD (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65 (2), 261–294. [3]

IMBENS, G. W., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79 (3), 933–959. [4]

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142 (2), 615–635. [1-3]

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," Report, Department of Economics, University of Wisconsin. [1-3]

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56 (4), 931–954. [5]

Dept. of Economics, University of Rochester, 280 Hutchison Road, Box 270156, Rochester, NY 14627, U.S.A.; carol.caetano@rochester.edu.

Manuscript received November, 2012; final revision received February, 2015.