Econometrica, Vol. 69, No. 1 (January, 2001), 201–209

NOTES AND COMMENTS

IDENTIFICATION AND ESTIMATION OF TREATMENT EFFECTS WITH A REGRESSION-DISCONTINUITY DESIGN

BY JINYONG HAHN, PETRA TODD, AND WILBERT VAN DER KLAAUW[1]

1. INTRODUCTION

THE REGRESSION DISCONTINUITY (RD) data design is a quasi-experimental design with the defining characteristic that the probability of receiving treatment changes discontinuously as a function of one or more underlying variables. This data design arises frequently in economic and other applications but is only infrequently exploited as a source of identifying information in evaluating effects of a treatment.

In the first application and discussion of the RD method, Thistlethwaite and Campbell (1960) study the effect of student scholarships on career aspirations, using the fact that awards are only made if a test score exceeds a threshold. More recently, Van der Klaauw (1997) estimates the effect of financial aid offers on students' decisions to attend a particular college, taking into account administrative rules that set the aid amount partly on the basis of a discontinuous function of the students' grade point average and SAT score. Angrist and Lavy (1999) estimate the effect of class size on student test scores, taking advantage of a rule stipulating that another classroom be added when the average class size exceeds a threshold level. Finally, Black (1999) uses an RD approach to estimate parents' willingness to pay for higher quality schools by comparing housing prices near geographic school attendance boundaries. Regression discontinuity methods have potentially broad applicability in economic research, because geographic boundaries or rules governing programs often create discontinuities in the treatment assignment mechanism that can be exploited under the method.

Although there have been several discussions and applications of RD methods in the literature, important questions still remain concerning sources of identification and ways of estimating treatment effects under minimal parametric restrictions. Here, we show that identifying conditions invoked in previous applications of RD methods are often overly strong and that treatment effects can be nonparametrically identified under an RD design by a weak functional form restriction. The restriction is unusual in that it requires imposing continuity assumptions in order to take advantage of the known discontinuity in the treatment assignment mechanism. We also propose a way of nonparametrically estimating treatment effects and offer an interpretation of the Wald estimator as an RD estimator.

[1] We would like to thank Joshua Angrist, James Heckman, Guido Imbens, Alan Krueger, Tony Lancaster, Sendhil Mullainathan, and Ken Wolpin for helpful comments. The paper has also benefited from comments received at the 1997 Midwestern Econometrics Group conference, the 1997 AEA meetings, the 1998 Econometric Society Summer Meetings in Montreal, the University of Michigan, the University of Rochester, the University of Wisconsin, the joint Harvard/MIT econometrics workshop, and the joint Brown-Yale-NYU-Penn-JHU labor conference. Van der Klaauw thanks the C. V. Starr Center of Applied Economics at NYU for research support. Todd is grateful to the NSF for support under #SBR-9730688.

2. REGRESSION-DISCONTINUITY DESIGN AND SOURCES OF IDENTIFICATION

The goal of an evaluation is to determine the effect that some binary treatment variable x_i has on an outcome y_i. The evaluation problem arises because persons either receive or do not receive treatment and no individual is observed in both states at the same time. Let y_{1i} denote the outcome with treatment and y_{0i} that in the absence of treatment, and let $x_i = 1$ if treatment is received and $x_i = 0$ otherwise. The model for the observed outcome can be written as $y_i = \alpha_i + x_i \cdot \beta_i$, where $\alpha_i \equiv y_{0i}$, $\beta_i \equiv y_{1i} - y_{0i}$.

There are two main types of discontinuity designs considered in the literature—the sharp design and the so-called fuzzy design.[2] With a sharp design, treatment x_i is known to depend in a deterministic way on some observable variable z_i, $x_i = f(z_i)$, where z_i takes on a continuum of values and the point z_0, where the function $f(z)$ is discontinuous, is assumed to be known. With a fuzzy design, x_i is a random variable given z_i, but the conditional probability $f(z) \equiv E[x_i \mid z_i = z] = \Pr[x_i = 1 \mid z_i = z]$ is known to be discontinuous at z_0. The fuzzy design differs from the sharp design in that the treatment assignment is not a deterministic function of z_i; there are additional variables unobserved by the econometrician that determine assignment to treatment. The common feature it shares with the sharp design is that the probability of receiving treatment, $\Pr[x_i = 1 \mid z_i]$, viewed as a function of z_i, is discontinuous at z_0:

ASSUMPTION (RD): (i) *The limits* $x^+ \equiv \lim_{z \to z_0^+} E[x_i \mid z_i = z]$ *and* $x^- \equiv \lim_{z \to z_0^-} E[x_i \mid z_i = z]$ *exist*.[3] (ii) $x^+ \neq x^-$.

Below, we focus on identification under the fuzzy design treating the sharp design as a special case.

2.1. *Constant Treatment Effects*

Suppose that the treatment effect β is constant across different individuals. Let $e > 0$ denote an arbitrary small number. Suppose that we have a reason to believe that in the absence of treatment, persons close to the threshold z_0 are similar. We would then expect $E[\alpha_i \mid z_i = z_0 + e] \cong E[\alpha_i \mid z_i = z_0 - e]$, which motivates the assumption:

ASSUMPTION (A1): $E[\alpha_i \mid z_i = z]$ *is continuous in* z *at* z_0.

Below, we establish that β is nonparametrically identified solely under this continuity restriction:

THEOREM 1: *Suppose that* β_i *is fixed at* β. *Further suppose that Assumptions* (*RD*) *and* (*A*1) *hold. We then have*

$$\beta = \frac{y^+ - y^-}{x^+ - x^-}, \tag{1}$$

where $y^+ \equiv \lim_{z \to z_0^+} E[y_i \mid z_i = z]$ *and* $y^- \equiv \lim_{z \to z_0^-} E[y_i \mid z_i = z]$.

[2] See Trochim (1984).

[3] Throughout this paper, we also assume that the density of z_i is positive in the neighborhood containing z_0.

PROOF: The mean difference in outcomes for persons above and below the discontinuity point is

$$\begin{aligned} &E[y_i \mid z_i = z_0 + e] - E[y_i \mid z_i = z_0 - e] \\ &\quad = \beta \cdot \{E[x_i \mid z_i = z_0 + e] - E[x_i \mid z_i = z_0 - e]\} \\ &\qquad + \{E[\alpha_i \mid z_i = z_0 + e] - E[\alpha_i \mid z_i = z_0 - e]\}. \end{aligned}$$

Under (A1), we have

$$\begin{aligned} &\lim_{z \to z_0^+} E[y_i \mid z_i = z] - \lim_{z \to z_0^-} E[y_i \mid z_i = z] \\ &\quad = \beta \cdot \left\{ \lim_{z \to z_0^+} E[x_i \mid z_i = z] - \lim_{z \to z_0^-} E[x_i \mid z_i = z] \right\}, \end{aligned}$$

from which the conclusion follows. The denominator in (1) is nonzero by Assumption (RD). *Q.E.D.*

With the sharp design, $x^+ = 1$ and $x^- = 0$ by definition. Therefore, β is identified by

$$\beta = y^+ - y^-. \tag{2}$$

2.2. *Variable Treatment Effects*

Now we consider the question of identification when treatment effects are heterogeneous. To generalize the identification strategy in the constant treatment effect case, we make the following assumption:

ASSUMPTION (A2): $E[\beta_i \mid z_i = z]$, *regarded as a function of* z, *is continuous at* z_0.

We establish that the average treatment effect at z_0, $E[\beta_i \mid z_i = z_0]$, is nonparametrically identified under the functional form restriction and a weak form of conditional independence:

THEOREM 2: *Suppose that* x_i *is independent of* β_i *conditional on* z_i *near* z_0. *Further suppose that Assumptions* (*RD*), (*A*1), *and* (*A*2) *hold. We then have*

$$E[\beta_i \mid z_i = z_0] = \frac{y^+ - y^-}{x^+ - x^-}. \tag{3}$$

PROOF: The mean difference in outcomes for persons above and below the discontinuity point is

$$\begin{aligned} &E[y_i \mid z_i = z_0 + e] - E[y_i \mid z_i = z_0 - e] \\ &\quad = \{E[x_i \cdot \beta_i \mid z_i = z_0 + e] - E[x_i \cdot \beta_i \mid z_i = z_0 - e]\} \\ &\qquad + \{E[\alpha_i \mid z_i = z_0 + e] - E[\alpha_i \mid z_i = z_0 - e]\}. \end{aligned}$$

By conditional independence, we have

$$E[x_i \cdot \beta_i \mid z_i = z \pm e] = E[\beta_i \mid z_i = z \pm e] \cdot E[x_i \mid z_i = z \pm e].$$

Combined with (A1) and (A2), we obtain

$$\lim_{z \to z_0^+} E[y_i \mid z_i = z] - \lim_{z \to z_0^-} E[y_i \mid z_i = z]$$
$$= E[\beta_i \mid z_i = z_0] \cdot \left\{ \lim_{z \to z_0^+} E[x_i \mid z_i = z] - \lim_{z \to z_0^-} E[x_i \mid z_i = z] \right\},$$

from which the conclusion follows. *Q.E.D.*

With a sharp design, $E[\beta_i \mid z_i = z_0]$ is identified by

$$E[\beta_i \mid z_i = z_0] = y^+ - y^-. \tag{4}$$

The conditional independence assumption maintains that individuals do not select into treatment on the basis of anticipated gains from treatment. Although such assumptions are routinely invoked in the literature on matching estimators, this type of assumption may be considered unrealistic in a world in which individuals self-select into treatment.[4] To examine the consequence of dropping the assumption, we consider an alternative set of conditions that allows selection into the program on the basis of prospective gains. Suppose, as in Imbens and Angrist (1994), that for each observation i, treatment assignment is a deterministic function of z, but the function is different for different persons or groups of persons. Consider the following set of assumptions on impacts and treatment assignment:

ASSUMPTION (A3): (i) $(\beta_i, x_i(z))$ *is jointly independent of* z_i *near* z_0. (ii) *There exists* $\varepsilon > 0$ *such that* $x_i(z_0 + e) \geq x_i(z_0 - e)$ *for all* $0 < e < \varepsilon$.

THEOREM 3: *Suppose that Assumptions* (*RD*), (*A*1), *and* (*A*3) *hold. We then have*

$$\lim_{e \to 0^+} E[\beta_i \mid x_i(z_0 + e) - x_i(z_0 - e) = 1] = \frac{y^+ - y^-}{x^+ - x^-}. \tag{5}$$

PROOF: Invoking the reasoning in Imbens and Angrist (1994), we obtain

$$E[x_i \cdot \beta_i \mid z_i = z_0 + e] - E[x_i \cdot \beta_i \mid z_i = z_0 - e]$$
$$= E[\beta_i \mid x_i(z_0 + e) - x_i(z_0 - e) = 1]$$
$$\cdot \{E[x_i \mid z_i = z_0 + e] - E[x_i \mid z_i = z_0 - e]\},$$

from which the conclusion follows. *Q.E.D.*

For $e > 0$ sufficiently small, the conditioning event $\{x_i(z_0 + e) - x_i(z_0 - e) = 1\}$ in (5) corresponds to the subgroup of persons for whom treatment changes discontinuously at z_0. Therefore, (5) identifies the local average treatment effect (LATE) at z_0.

2.3. *Discussion*

In each of the cases considered, identification was made possible by comparing persons arbitrarily close to the point z_0 who did and did not receive treatment. Without further assumptions such as the common effect assumption, treatment effects can only be identified at $z = z_0$. This notion of identification is similar to the notion of identification at infinity.[5]

[4] See Heckman, Lalonde, and Smith (1999) for related discussion.
[5] See Chamberlain (1986).

For identification of treatment effects, we relied heavily on a local continuity restriction on $E[\alpha_i \mid z_i]$ and a known discontinuity in $E[x_i \mid z_i]$. We now show, in the context of a common effects model, that such functional form restrictions are necessary, and that without them the model is nonparametrically unidentified. We can put the model for outcomes in more familiar econometric notation by writing

$$y_i = \alpha(z_i) + \beta \cdot x_i + v_i$$

where $\alpha(z_i) \equiv E[\alpha_i \mid z_i]$ and $v_i \equiv \alpha_i - \alpha(z_i)$. We argue that the usual conditional mean independence restriction, $E[v_i \mid z_i] = 0$, is not sufficient for identification of the treatment effect, even for the common treatment effect case. For this purpose consider another DGP, where we have

$$y_i = \alpha^*(z_i) + 0 \cdot x_i + v_i^*,$$

and where

$$\alpha^*(z_i) \equiv \alpha(z_i) - \beta \cdot E[x_i \mid z_i], \qquad v_i^* \equiv v_i + \beta \cdot \{x_i - E[x_i \mid z_i]\}.$$

These two models are equivalent except that the treatment effect in the former case is β whereas in the latter case it is equal to 0. We cannot distinguish the models in the population if $E[v_i \mid z_i] = 0$ is the only restriction available.

3. ESTIMATION

For both the sharp design and fuzzy design, the ratio

$$\frac{y^+ - y^-}{x^+ - x^-} \tag{6}$$

identifies the treatment effect at $z = z_0$. Thus, given consistent estimators $\hat{y}^+$, $\hat{y}^-$, $\hat{x}^+$, and $\hat{x}^-$ of the four one-sided limits in (6), the treatment effect can be consistently estimated by

$$\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-}.$$

In principle, we can use any nonparametric estimator to estimate the limits. We first consider one-sided kernel estimation and observe that under certain conditions an estimate based on kernel regression will be numerically equivalent to a standard Wald estimator. We then argue that such an estimator may have a poor finite sample property due to the boundary problem and propose to avoid the boundary problem by using local linear nonparametric regression (LLR) methods.

Consider the special case where we use kernel regression estimators based on one-sided uniform kernels. For the uniform kernel, it is not difficult to show that

$$\hat{y}^+ = \frac{\sum_{i \in \mathscr{P}} y_i \cdot w_i}{\sum_{i \in \mathscr{P}} w_i}, \qquad \hat{y}^- = \frac{\sum_{i \in \mathscr{P}} y_i \cdot (1 - w_i)}{\sum_{i \in \mathscr{P}} (1 - w_i)},$$

$$\hat{x}^+ = \frac{\sum_{i \in \mathscr{P}} x_i \cdot w_i}{\sum_{i \in \mathscr{P}} w_i}, \qquad \hat{x}^- = \frac{\sum_{i \in \mathscr{P}} x_i \cdot (1 - w_i)}{\sum_{i \in \mathscr{P}} (1 - w_i)},$$

where $\mathscr{S}$ denotes the subsample such that $z_0 - h < z_i < z_0 + h$, $w_i \equiv 1$ $(z_0 < z_i < z_0 + h)$, and $h > 0$ denotes the bandwidth. The estimator is *numerically* equivalent to an IV estimator for the regression of y_i on x_i which uses w_i as an instrument, applied to the subsample $\mathscr{S}$. Denote this estimator by $\hat{\beta}_w$.

It is interesting to note that the regression discontinuity can 'justify' a Wald estimator even when the standard IV assumption is violated. To see this, put the model in more familiar econometric notation by writing $\alpha_i = E[\alpha_i] + v_i = \alpha + v_i$. Under the common treatment assumption, this yields a model

$$y_i = \alpha + x_i \cdot \beta + v_i.$$

Identification of β does not require that the error term v_i be uncorrelated with z_i. All that is required is continuity assumption (A1). As long as the researcher is willing to change h appropriately as a function of the sample size, $\hat{\beta}_w$ is consistent. Thus $\hat{\beta}_w$ is motivated by a different principle than is the usual Wald estimator, but for a particular choice of kernel and subsample they are numerically equivalent.

Although $\hat{\beta}_w$ is numerically equivalent to a local Wald estimator, inference based on $\hat{\beta}_w$ will be different from that based on a Wald estimator. $\hat{\beta}_w$ will be asymptotically biased, as are many other nonparametric-regression-based estimators, whereas the Wald estimator is asymptotically unbiased by assumption. The bias problem is exacerbated in the regression-discontinuity case due to the bad boundary behavior of the kernel regression estimator: at boundary points, the bias of the kernel regression estimator converges to zero at a slower rate than at interior points. Under conventional assumptions on the kernel function, the order of the bias of the standard kernel estimator is $O(h)$ at boundary points and $O(h^2)$ at interior points.[6] For our problem, all the points of estimation are at boundaries, so the bias could be substantial in finite samples.[7] It would be misleading to use the conventional confidence interval based on the asymptotic distribution of the (asymptotically unbiased) Wald estimator as the true coverage probability would be very different from the nominal coverage probability.

Because of the poor boundary performance of standard kernel estimators, we propose instead to estimate the limits by local linear regression (LLR), shown by Fan (1992) to have better boundary properties than the traditional kernel regression estimator. The local linear estimator for y^+, for example, is given by $\hat{a}$, where

$$(\hat{a}, \hat{b}) \equiv \underset{a,b}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - a - b(z_i - z_0))^2 K\left(\frac{z_i - z_0}{h}\right) 1(z_i > z_0).$$

Here, $K(\cdot)$ is a kernel function and $h > 0$ is a suitable bandwidth. The smaller bias associated with the LLR estimator implies that it is more rate-efficient than the kernel-based estimator. Another advantage of LLR is that the bias does not depend on the design density of the data. Because of these advantages, local linear methods are

[6] See Härdle (1990) or Härdle and Linton (1994) for further discussion of the boundary bias problem. Under slightly stronger assumptions than ours, Porter (1998) recently proposed an alternative estimator for the sharp discontinuity design, constant effect model for which the boundary bias problem does not exist.

[7] The boundary bias formula of the kernel estimator suggests that the bias is the smallest when the conditional expectations $E[y_i | z_i]$ and/or $E[x_i | z_i]$ have one-sided derivatives around z_0 equal to zero. We thus find that $\hat{\beta}_w$ has a small bias only for the case where α_i has no correlation with z_i, i.e., the case where z_i is a proper instrument and the Wald assumption is exactly satisfied near the discontinuity.

deemed to be a better choice than standard kernel methods. The asymptotic distribution of the treatment effect estimator based on local linear regression is derived in the Appendix.

4. SUMMARY

The RD method provides a way of identifying mean treatment impacts for a subgroup of the population under minimal assumptions. An advantage of the method is that it bypasses many of the questions concerning model specification: both the question of which variables to include in the model for outcomes and of their functional forms. A limitation of the approach is that it only identifies treatment effects locally at the point at which the probability of receiving treatment changes discontinuously. However, in some cases this localized parameter is precisely the parameter of interest. It would be of interest, for example, if the policy change being considered is a small change in the program rules, such as lowering or raising the threshold for program entry, in which case we would want to know the effect of treatment for the subpopulation affected by the change.

In this paper, we considered the question of identification and estimation under two RD designs, the sharp and the fuzzy design. The estimator we propose uses recently developed local linear nonparametric regression techniques that avoid the poor boundary behavior of the kernel regression estimator. We also discussed why the regression-discontinuity design sometimes provides a possible justification for the Wald estimator, even when the zero correlation condition is violated.

Department of Economics, Brown University, Box B, Providence, RI 02912, U.S.A.,
Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.,
and
Department of Economics, UNC-Chapel Hill, CB#3305, Gardner Hall, Chapel Hill, NC 27599, U.S.A.

Manuscript received February, 1998; final revision received December, 1999.

APPENDIX

We next present the distribution theory for the estimator $\hat{\beta}$ of the ratio (6), where the limits are estimated by local linear regression. Define $m(z) \equiv E[y_i \mid z_i = z]$ and $p(z) \equiv [x_i \mid z_i = z]$, and define the limits $\lim_{z \to z_0^+} E[y_i \mid z_i = z]$, $\lim_{z \to z_0^-} E[y_i \mid z_i = z]$, $\lim_{z \to z_0^+} E[x_i \mid z_i = z]$, and $\lim_{z \to z_0^-} E[x_i \mid z_i = z]$ by $m^+(z_0)$, $m^-(z_0)$, $p^+(z_0)$, and $p^-(z_0)$, respectively. Additionally, define

$$\sigma^{2^+}(z_0) = \lim_{z \to z_0^+} \operatorname{var}[y_i \mid z_i = z], \qquad \sigma^{2^-}(z_0) = \lim_{z \to z_0^-} \operatorname{var}[y_i \mid z_i = z],$$

$$\eta^+(z_0) = \lim_{z \to z_0^+} \operatorname{cov}[y_i, x_i \mid z_i = z], \qquad \text{and} \qquad \eta^-(z_0) = \lim_{z \to z_0^-} \operatorname{cov}[y_i, x_i \mid z_i = z].$$

THEOREM 4 (Asymptotic Distribution): *Suppose that*:

(i) *For $z > z_0$, $m(z)$ and $p(z)$ are twice continuously differentiable. There exists some $M > 0$ such that $|m^+(z)|$, $|m'^+(z)|$, $|m''^+(z)|$ and $|p^+(z)|$, $|p'^+(z)|$, $|p''^+(z)|$ are uniformly bounded on $(z_0, z_0 + M]$. Similarly, $|m^-(z)|$, $|m'^-(z)|$, $|m''^-(z)|$ and $|p^-(z)|$, $|p'^-(z)|$, $|p''^-(z)|$ are uniformly bounded on $[z_0 - M, z_0)$.*

(ii) *The limits* $m^+(z_0)$, $m^-(z_0)$, $m'^+(z_0)$, $m'^-(z_0)$, $m''^+(z_0)$, $m''^-(z_0)$, $p^+(z_0)$, $p^-(z_0)$, $p'^+(z_0)$, $p'^-(z_0)$, $p''^+(z_0)$, *and* $p''^-(z_0)$ *exist and are finite.*

(iii) *The density of* z_i, $f(z)$, *is continuous and bounded near* z_0. *It is also bounded away from zero near* z_0.

(iv) $K(\cdot)$ *is continuous, symmetric, and nonnegative-valued with compact support.*

(v) $\sigma^2(z_i) = \mathrm{var}(y_i \mid z_i)$ *is uniformly bounded near* z_0. *Similarly,* $\eta(z_i) = \mathrm{cov}(y_i, x_i \mid z_i)$ *is uniformly bounded near* z_0. *Furthermore, the limits* $\sigma^{2^+}(z_0)$, $\sigma^{2^-}(z_0)$, $\eta^+(z_0)$, *and* $\eta^-(z_0)$ *exist and are finite.*

(vi) $\lim_{z \to z_0^+} E[|y_i - m(z_i)|^3 \mid z_i = z]$ *and* $\lim_{z \to z_0^-} E[|y_i - m(z_i)|^3 \,\big|\, z_i = z]$ *exist and are finite.*

(vii) *The bandwidth sequence satisfies* $h = h_n = \varrho \cdot n^{-1/5}$ *for some* ϱ. *Then,*

$$n^{\frac{2}{5}}\left(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-} - \frac{y^+ - y^-}{x^+ - x^-}\right) \to \mathcal{N}(\mu_f, \Omega_f),$$

where

$$\mu_f \equiv \frac{1}{x^+ - x^-}(\rho^+ m''^+(z_0) - \rho^- m''^-(z_0)) - \frac{y^+ - y^-}{(x^+ - x^-)^2}(\rho^+ p''^+(z_0) - \rho^- p''^-(z_0)),$$

and

$$\begin{aligned}\Omega_f \equiv{}& \frac{1}{(x^+ - x^-)^2}(\omega^+ \sigma^{2+}(z_0) + \omega^- \sigma^{2-}(z_0)) \\ &- 2\frac{y^+ - y^-}{(x^+ - x^-)^3}(\omega^+ \eta^+(z_0) + \omega^- \eta^-(z_0)) \\ &+ \frac{(y^+ - y^-)^2}{(x^+ - x^-)^4}(\omega^+ p^+(z_0)(1 - p^+(z_0)) + \omega^- p^-(z_0)(1 - p^-(z_0))),\end{aligned}$$

and where

$$\rho^+ \equiv \frac{\left(\int_0^\infty u^2 K(u)\,du\right)^2 - \left(\int_0^\infty u^3 K(u)\,du\right)\left(\int_0^\infty u K(u)\,du\right)}{\left(\int_0^\infty u^2 K(u)\,du\right)\left(\int_0^\infty K(u)\,du\right) - \left(\int_0^\infty u K(u)\,du\right)^2}\,\frac{\varrho^2}{2},$$

$$\omega^+ \equiv \frac{\int_0^\infty \left(\left(\int_0^\infty s^2 K(s)\,ds\right) - \left(\int_0^\infty s K(s)\,ds\right)\cdot u\right)^2 K(u)^2\,du}{f(z_0)\varrho \cdot \left[\left(\int_0^\infty u^2 K(u)\,du\right)\left(\int_0^\infty K(u)\,du\right) - \left(\int_0^\infty u K(u)\,du\right)^2\right]^2}.$$

with ρ^- *and* ω^- *similarly defined but now with the integral in the limits of integration over* $(-\infty, 0)$.

PROOF: A derivation of the distribution of the estimator is available in Hahn, Todd, and Van der Klaauw (1999), or upon request from the authors.

REFERENCES

ANGRIST, J., AND V. LAVY (1999): "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

BLACK, S. (1999): "Do 'Better' Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114, 577–599.

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models with Censoring," *Journal of Econometrics*, 32, 189–218.

FAN, J. (1992): "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (1999): "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design," NBER Working Paper 7131.

HECKMAN, J., R. LALONDE, AND J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs," forthcoming in *Handbook of Labor Economics*, *Volume III*, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*. New York: Cambridge University Press.

HÄRDLE, W., AND O. LINTON (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics*, *4*, ed. by D. F. McFadden and R. F. Engle. Amsterdam: North Holland, 2295–2339.

IMBENS, G., AND J. ANGRIST (1994): "Identification of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

PORTER, J. (1998): "Estimation of Regression Discontinuities," Seminar Notes.

THISTLETHWAITE, D., AND D. CAMPBELL (1960): "Regression-discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51, 309–317.

TROCHIM, W. (1984): *Research Design for Program Evaluation*: *the Regression-Discontinuity Approach*. Beverly Hills: Sage Publications.

VAN DER KLAAUW, W. (1997): "A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment," C. V. Starr Center Research Report 97-10, New York University.