A Root-N Consistent Estimation of Regression Discontinuity Models[1]

Chunrong Ai
Department of Economics, University of Florida,
Gainesville, FL 32611, USA

This verion is still under revision; please do not quote.

[1]I have benefited greatly from conversation with David Lee and the comments made by seminar participants at Maryland, Ohio State, NYU-Columbia joint econometric seminar. The Warrington College of Business at the University of Florida provided partial support through 2006 and 2007 Summer research grant. I remain responsible for all errors.

Abstract

This paper proposes a new estimation for the regression discontinuity model studied in Hahn, Todd, and Van der Klaauw (2001) and Porter (2003). Under some parameterizations of the treatment effect, the proposed procedure consistently estimates the treatment effect function as well as the average outcome in the absence of treatment. We show that the the treatment effect estimator is root-n consistent and establish the asymptotic distribution. Moreover, we also show that the proposed procedure still consistently estimates the treatment effect for the threshold individuals even if the parametric specification is false.

KEYWORDS: Discountiuity regression; series approximation; least suares; consistency.

1 Introduction

Regression discontinuity (RD) model has recently been suggested as a framework for evaluating the treatment program in a quasi experimental design. It is formulated by imposing restrictions on the following dummy variable equation:

$$y_i = \alpha_i + \beta_i x_i, \tag{1}$$

where i indexes individual, y_i denotes the observed outcome, x_i denotes the observed treatment status: $x_i = 1$ if treatment is received and $x_i = 0$ if treatment is not received, α_i denotes the outcome if treatment is not received, and $\alpha_i + \beta_i$ denotes the outcome if treatment is received (so that β_i is the treatment effect). The variables (α_i, β_i, x_i) are often related to some underlying variables z_i (which could be more than one variables) through the probability of receiving treatment $p(z) = E\{x_i|z_i = z\}$, the average baseline effect $\alpha(z) = E\{\alpha_i|z_i = z\}$, and the average treatment effect $\beta(z) = E\{\beta_i|z_i = z\}$. The key restriction is that the probability of receiving treatment, $p(z)$, is a discontinuous function. Specifically, for some exclusive decomposition $\mathcal{Z} = \mathcal{Z}_- \cup \mathcal{Z}_+$, there exists some z_0 on the boundary of both subsets such that

$$\begin{aligned} \lim_{z \to z_0^+} p(z) &= \lim_{t \to 0} p(z^+(t)) \text{ for some path } z^+(t) \in \mathcal{Z}_+ \text{ for } t > 0 \text{ and } z^+(0) = z_0, \\ \lim_{z \to z_0^-} p(z) &= \lim_{t \to 0} p(z\_(t)) \text{ for some path } z\_(t) \in \mathcal{Z}_- \text{ for } t > 0 \text{ and } z\_(0) = z_0. \end{aligned}$$

Discontinuity means that

$$\lim_{z \to z_0^+} p(z) \neq \lim_{z \to z_0^-} p(z)$$

holds. In economic applications, the subset $\mathcal{Z}_+$ defines eligibility criterion for the treatment program with z_0 as the minimum criterion and $\mathcal{Z}_-$ consists of individuals who do not meet the minimum criterion. Since ineligible individuals are not allowed to participate in the program, we have $p(z) = 0$ for all $z \in \mathcal{Z}_-$. Thus, the discontinuity restriction is satisfied if there is a nontrivial probability of receiving treatment for any eligible individual: $p(z) > 0$ for any $z \in \mathcal{Z}_+$. It is worth noting that the eligibility criterion $\mathcal{Z}_+$, though determined by a single underlying variable in most economic applications, may depend on several underlying variables. For instance, a perspective student's eligibility for a need-based scholarship requires her SAT score to be above certain threshold and her family income is below certain level. If eligibility is determined by more than one variable, then z_0 is not unqiue. There may exist other points at which $p(z)$ is discontinuous. The proposed approach in this paper applies to any discontinuity point so we should treat z_0 as a generic discontinuity point.

Under the conditional mean independence condition:

$$E\{\beta_i x_i | z_i = z\} = E\{\beta_i | z_i = z\} E\{x_i | z_i = z\} = \beta(z) * p(z)$$

and the condition that $\alpha(z)$ and $\beta(z)$ are both continuous at $z = z_0$, Hahn, Todd, and Van der Klaauw (2001) show that the coefficient $\beta(z_0)$ is identified as

$$\beta(z_0) = \frac{\lim_{z \to z_0^+} E\{y_i | z_i = z\} - \lim_{z \to z_0^-} E\{y_i | z_i = z\}}{\lim_{z \to z_0^+} p(z) - \lim_{z \to z_0^-} p(z)}.$$

They then proceed to propose an estimator for $\beta(z_0)$ by replacing the pathwise limits with consistent nonparametric estimates.[1] Using the same proof as in Hahn, Todd, and Van der Klaauw (2001), if there exist other discontinuity points, we can show that the values of $\beta(z)$ at those discontinuity points are also identified.

The attraction of Hahn's et. al. procedure is that it is quite simple to compute and has desirable asymptotic properties. The limitation of their procedure is that it says nothing about the treatment effect on non-threshold eligible individuals. It is possible that the threshold individuals receive no effect: $\beta(z_0) = 0$ but other eligible individuals receive positive effect: $\beta(z) > 0$ for $z \in \mathcal{Z}_+$. It is also possible that the threshold individuals receive negative effect: $\beta(z_0) < 0$ but well qualified individuals receive positive effect: $\beta(z) > 0$ for large $z \in \mathcal{Z}_+$, or that $\beta(z_0) > 0$ and $\beta(z) < 0$ for large $z \in \mathcal{Z}_+$. Thus, to evaluate the effectiveness of the treatment program, it is important to estimate the whole function $\beta(z)$, not just $\beta(z_0)$. Unfortunately, under Hahn's et. al. conditions, $\beta(z_0)$ is the only parameter that is identified. The identification difficulty arises from the fact that there are no individuals with the same underlying variables z in both the ineligible and eligible group and that the eligible individuals alone cannot distinguish the treatment effect $\beta(z)$ from the baseline effect $\alpha(z)$.

Clearly, to identify the treatment effect function, some restrictions must be imposed on either the baseline effect $\alpha(z)$ or the treatment effect $\beta(z)$ or both. Since the treatment effect function is the focus of most empirical studies, it is natural to impose restrictions on the treatment effect function but leave the baseline effect function $\alpha(z)$ unspecified. It is worth noting that any parameterization of $\beta(z)$ such as $h(z, \theta_0)$ is not necessarily identified without some restrictions on $\alpha(z)$. This follows because we can always write

$$h(z, \theta_0) + \alpha(z) = h(z, \theta) + (\alpha(z) + h(z, \theta_0) - h(z, \theta)) = h(z, \theta) + \widetilde{\alpha}(z), \ z \in \mathcal{Z}_+$$

[1] Porter (2003) derives the asympptotic distribution of this and other similar estimators.

for any θ. With restriction that $\alpha(z)$ is continuous at $z = z_0$, we can identify $h(z_0, \theta_0)$ but not θ_0 if θ_0 is not univariate.

Some other form of restriction must be imposed to identify $\beta(z)$. We now show that the following functional form restriction on $\beta(z)$:

$$\frac{\partial^m \beta(z)}{\partial v^m} = 0 \text{ holds over } z \in \mathcal{Z}_+ \text{ almost everywhere,} \tag{2}$$

where m is some known integer (e.g., $m = 2$) and $z = (v, w')'$ with v a continuous scalar, is enough to identify $\beta(z)$ and $\alpha(z)$. To see this, suppose that $\beta(z)$ and $\alpha(z)$ have up to m^{th} derivatives with respect to v almost everywhere. Consider the sharp design $x = 1\{z \in \mathcal{Z}_+\}$. Condition (2) allows us to eliminate the treatment effect $\beta(z)$ by differentiation:

$$\frac{\partial^m E\{y_i|z_i = z\}}{\partial v^m} = \frac{\partial^m \alpha(z)}{\partial v^m} \text{ over } z \in \mathcal{Z} \text{ almost everywhere.}$$

Since we can compute $E\{y_i|z_i = z\}$ for almost all z from observed data, we can compute the derivative $h(z) = \frac{\partial^m \alpha(z)}{\partial v^m}$ almost everywhere. Notice that condition (2) implies

$$\beta(z) = b(w)' t^J(v),$$

for some $J \times 1$ vector of known functions $t^J(v)$ satisfying

$$\frac{d^m t^J(v)}{dv^m} = 0 \text{ almost everywhere,}$$

and $b(w)$ is a vector of unknown functions of the other underlying variables w. Examples of the known functions $t^J(v)$ include power functions:

$$t^J(v) = p^J(v) = (1, v, ..., v^{J-1})'$$

with v_0 as the known threshold, the step functions (e.g., $m = 1$):

$$t^J(v) = (1, 1\{e_1 \leq v < e_2\}, ..., 1\{e_J \leq v\})'$$

for some known $e_1, e_2, ..., e_J$, and the piecewise linear functions (e.g., $m = 2$):

$$t^J(v) = (1, 1\{e_1 \leq v < e_2\}, ..., 1\{e_J \leq v\})' \otimes (1, v)'.$$

In most applications, it is preferrable to specify $t^J(v)$ as low order spline basis functions. The component $b(w)$ is either parameterized (such as linear form) or left unspecified. We will discuss both cases. But for demonstration purpose, we will use the simplest case: $z = v$, $t^J(v) = 1$ and $b(w) = \theta_0$ as example.

Compute the indefinite integration of $h(z)$ with respect to v for m times, we obtain

$$\begin{aligned}\alpha(z) &= \int \left(\int \cdots \int h(v, w) dv \cdots \right) dv + c(w)' p^J(v) \\ &= g(z) + c(w)' p^J(v)\end{aligned}$$

for almost all z, where $c(w)$ is a vector of unknown functions and $g(z)$ is the known part of the indefinite integration. This means that $\alpha(z)$ is identified if $c(w)$ is identified. Let $\mathcal{W}$ denote the support of w and let $\mathcal{V}$ denote the support of v. Suppose that $\mathcal{Z}_+ = \{v \in \mathcal{V}, v \geq v_0\} \times \mathcal{W}$ and $\mathcal{Z}_- = \{v \in \mathcal{V}, v < v_0\} \times \mathcal{W}$. Then, $c(w)$ is identified by ineligible individuals as $c(w)$ solves

$$\min_{c(\cdot)} E\left\{(y_i - g(z_i) - c(w_i)' p^J(v_i))^2 | w_i = w, x_i = 0\right\}.$$

After computing $\alpha(z) = g(z) + c(w)' p^J(v)$, $b(w)$ is identified by eligible individuals:

$$\min_{b(\cdot)} E\left\{y_i - \alpha(z_i) - b(w_i)' t^J(v_i))^2 | w_i = w, x_i = 1\right\},$$

provided that

$$E\{t^J(v_i) t^J(v_i)' | w_i = w, x_i = 1\} \text{ is nonsingular for almost all } w.$$

We summarize these results in the following assumption and lemma.

Assumption 1. *(i) The observations* $\{(y_i, x_i, z_i), i = 1, 2, ..., N\}$ *are identically distributed; (ii) the following conditional independence condition*

$$E\{\beta_i x_i | z_i = z\} = E\{\beta_i | z_i = z\} E\{x_i | z_i = z\} = \beta(z) * p(z)$$

holds for all z*; (iii) for some known basis functions* $t^J(v)$, $\beta(z) = b(w)' t^J(v)$*; (iv) for some known integer* m, $t^J(v)$ *and* $\alpha(z)$ *have up to* m^{th} *derivatives with respect to* v *almost everywhere and*

$$\frac{d^m t^J(v)}{dv^m} = 0 \textit{ holds almost everywhere;}$$

(v) $\mathcal{Z}_+ = \{v \in \mathcal{V}, v \geq v_0\} \times \mathcal{W}$ *and* $\mathcal{Z}_- = \{v \in \mathcal{V}, v < v_0\} \times \mathcal{W}$*; (vi)*

$$E\{t^J(v_i)t^J(v_i)'|w_i = w, x_i = 1\}$$

is nonsingular for every w*; (vii)* $x = 1\{v \geq v_0\}$ and $p(z) = x$*.*

Lemma 1. *For the sharp design, Assumption 1 identifies* $\alpha(z)$ *for* $z \in \mathcal{Z}$ *and* $\beta(z) = b(w)'t^J(v)$ *for* $z \in \mathcal{Z}_+$*.*

Before we turn to the case of fuzzy design, we illustrate the identification through the simplest example: $z = v$, $t^J(v) = 1$ and $b(w) = \theta_0$. In this example, we have

$$\begin{aligned} h(v) &= \frac{dE\{y|v, x=0\}}{dv} \text{ for } v < v_0 \text{ and } h(v) = \frac{dE\{y|v, x=1\}}{dv} \text{ for } v \geq v_0, \\ \alpha(v) &= \int h(v)dv + c = g(v) + c. \end{aligned}$$

The unknown constant c solves

$$\min_c E\left\{(y_i - g(v_i) - c)^2 | x_i = 0\right\},$$

which yields

$$c = E\left\{y_i - g(v_i) | x_i = 0\right\}.$$

The treatment effect solves

$$\min_\theta E\left\{y_i - g(v_i) - c - \theta)^2 | x_i = 1\right\},$$

which yields

$$\theta_0 = E\left\{y_i - g(v_i) - c) | x_i = 1\right\}.$$

It is clear from the above expression that we use the whole treatment group, not just the threshold individuals, to estimate the average treatment effect. Thus, the proposed procedure yields a root-N consistent estimator for the average treatment effect. The proposed procedure is criticized, however, for being multistep (e.g., step I eliminates the treatment effect through differentiation; step II recovers the baseline effect from the ineligible group; step III estimates the treatment effect from eligible group) and not better than Porter's (2003) dummy variable approach. Again, in our simplest example, Porter's dummy variable approach is the model

$$y_i = \alpha(v_i) + x_i\theta_0 + u_i, i = 1, 2, ..., N.$$

Applying Robinson's partialling out procedure, we obtain

$$y_i - E\{y_i|v_i\} = (x_i - E\{x_i|v_i\})\,\theta_0 + u_i.$$

The problem here is that $x_i - E\{x_i|v_i\}$ is identically zero for all v_i. Porter suggests using the approximation $x_i - \widetilde{E}\{x_i|v_i\}$ and estimate the treatment effect by

$$\widehat{\theta}_{Porter} = \frac{\sum_{i=1}^{N} \left(x_i - \widetilde{E}\{x_i|v_i\}\right) y_i}{\sum_{i=1}^{N} \left(x_i - \widetilde{E}\{x_i|v_i\}\right)^2}.$$

This estimator appears using all observations, not just those near threshold. However, looking deeper, we find this estimator is no better than Hahn's estimator which use observations near threshold. The reason is that the approximation error $x_i - \widetilde{E}\{x_i|v_i\}$, though not identically zero, converges to zero at differential rate. The approximation error converges to zero for all v outside a neighborhood of v_0 faster than for v in the neighborhood of v_0. Since the slower rate terms dominate the faster rate terms in both the numerator and denominator of Porter's estimator, his estimator essentially uses individuals near threshold. Moreover, because of the approximation error converges to zero at rate slower than root-N, Porter's estimator is not root-N consistent. We are able to achieve root-N consistentcy because we exploit information on derivatives, not just the discontinuity, wheras Porter (2003) does not.

We now turn to the fuzzy design. Under our assumption, we have $E\{x|z\} = E\{x|v\}$. So the probability of receiving treatment is $p(v) = E\{x|v\}$. The difference between fuzzy and sharp design is that $p(v)$ is no longer an indicator function. Our approach is to transform the model so that the transformed model is identical to the sharp design. Let $\widetilde{p}(v)$ denote any positive and continuously differentiable probability function for all v and satisfying $\widetilde{p}(v) = p(v)$ for all $v \geq v_0$. Define

$$\widetilde{y}_i = \frac{y_i}{\widetilde{p}(v_i)} \text{ and } \widetilde{\alpha}(z_i) = \frac{\alpha(z_i)}{\widetilde{p}(v_i)}.$$

It is easy to show that

$$E\{\widetilde{y}_i|z_i = z\} = \widetilde{\alpha}(z) + \beta(z)\widetilde{x}$$

where $\widetilde{x} = 1\{v \geq v_0\}$ is the eligibility status variable. If we treat any eligible individuals as "treated" regardless of whether they actually receive the treatment or not, the above equation is the RD with sharp design with $\widetilde{x}$ as the "treatment status" variable, $\widetilde{\alpha}(z)$ as the baseline effect, and $\widetilde{y}$ as the observed effect. Applying Lemma 1 to this model, we identify $\widetilde{\alpha}(z)$ and $\beta(z) = b(w)'t^J(v)$. Since $p(v)$ is identified by regressing x on v, we obtain:

Lemma 2. *For the fuzzy design, suppose that $\widetilde{p}(z)$ has up to m^{th} derivatives with respect to v almost everywhere. Assumption 1 identifies $\alpha(z)$ for $z \in \mathcal{Z}$ and $\beta(z)$ for $z \in \mathcal{Z}_+$.*

To illustrate, we again use the simplest case. The probability of receiving treatment can be estimated parametrically by probit or logit regression of x on v and v^2 or nonparametrically. Denote

$$\begin{aligned} \widetilde{h}(v) &= \frac{dE\{\widetilde{y}|v, \widetilde{x}=0\}}{dv} \text{ for } v < v_0 \text{ and } \widetilde{h}(v) = \frac{dE\{\widetilde{y}|v, \widetilde{x}=1\}}{dv} \text{ for } v \geq v_0, \\ \widetilde{\alpha}(v) &= \int \widetilde{h}(v)dv + \widetilde{c} = \widetilde{g}(v) + \widetilde{c}, \end{aligned}$$

$$\widetilde{c} = E\left\{\widetilde{y}_i - \widetilde{g}(v_i)|\widetilde{x}_i = 0\right\}.$$

The treatment effect is given by

$$\theta_0 = E\left\{\widetilde{y}_i - \widetilde{g}(v_i) - \widetilde{c})|\widetilde{x}_i = 1\right\}.$$

The paper is organized as follows: Section 2 details an estimation strategy, Section 3 discusses the consequence if the treatment effect function is misspecified, and Section 4 concludes the paper.

2 Estimation

As we discussed above, the key step for estimating the treatment effect is to estimate the baseline effect. To estimate the baseline effect, we need to estimate the conditional mean function. We will use the sieve regression which is also applied by Gallant and Nychka (1987), Andrews (1994), Newey (1997), Ai and Chen (2003), Ai (2005a,b) and others. Specifically, let $s(w) = (s_1(w), s_2(w), \cdots)$ denote a sequence of known basis functions that can approximate any measurable function $c(w)$ arbitrarily well in the sense that there exist coefficients $(\pi_1, \pi_2 \cdots)$ such that

$$\left\| c(w) - \sum_{j=1}^{K_2} s_j(w)\pi_j \right\|_\infty = \sup_{z \in \mathcal{Z}} \left| c(w) - \sum_{j=1}^{K_2} s_j(w)\pi_j \right| \to 0 \text{ as } K_2 \to +\infty.$$

Examples of the basis functions include power series and B-splines. Let $t(v) = (t^J(v)', t_{J+1}(v), t_{J+2}(v), \cdots)$ denote the known basis functions that can approximate any measurable function of v arbitrarily well under the norm $\|\cdot\|_\infty$. Here we include $t^J(v)$ in the basis functions for a technical convenience. For some integers $K_1(> J)$ and K_2, denote $t^{K_1}(v) = (t^J(v)', t_{J+1}(v), ..., t_{K_1}(v))'$ and $s^{K_2}(w) = (s_1(w), s_2(w), \cdots, s_{K_2}(w))'$. With $K = K_1K_2$, denote

$$q^K(z) = t^{K_1}(v) \otimes s^{K_2}(w),$$

where $\otimes$ is the Kronecker product. Then, for any measurable function $f(z)$, there exist coefficients π_K such that

$$\left\|f(z) - q^K(z)'\pi_K\right\|_\infty = \sup_{z\in\mathcal{Z}} \left|f(z) - q^K(z)'\pi_K\right| \to 0 \text{ as } K_1, K_2 \to +\infty.$$

Moreover, if $f(z)$ is continuously differentiable with respect to v, we will assume

$$\left\|\frac{\partial^s f(z)}{\partial v^s} - \frac{\partial^s q^K(z)'\pi_K}{\partial v^s}\right\|_\infty = O(K^{-r}) \text{ uniformly in } K_1, K_2 \text{ and for } s = 0, 1, ..., m.$$

These approximation results permit us to use the truncated series $q^K(z)$ and $s^{K_2}(w)$ to approximate the unknown functions, and use $\frac{\partial^m q^K(z)}{\partial v^m}$ to approximate the derivative.

Let $\{(y_i, z_i, x_i), i = 1, 2, \cdots, N\}$ denote a sample of observations. To simplify exposition and without loss of generality, we assume that the first N_c individuals are ineligible for treatment while the last $N - N_c$ individuals are eligible for treatment.

2.1 Sharp design

Regressing y_i on $q^K(z_i)$ for the ineligible group, we obtain the regression coefficients

$$\widehat{\pi}_{Kc} = (Q_c'Q_c)^{-1}Q_c'Y_c,$$

where $Q_c = (q^K(z_1), q^K(z_2), \cdots, q^K(z_{N_c}))'$ and $Y_c = (y_1, \cdots, y_{N_c})'$. The conditional mean function for ineligible individuals is estimated by

$$\widehat{E}\{y_i|z_i = z\} = q^K(z)'\widehat{\pi}_{Kc} = q_1^K(z)'\widehat{\pi}_{1Kc} + q_2^K(z)'\widehat{\pi}_{2Kc} \text{ for } v < v_0,$$

where $q_1^K(z) = t^J(v) \otimes s^{K_2}(w)$, and the derivative is estimated by

$$\frac{\partial^m \widehat{E}\{y_i|z_i = z\}}{\partial v^m} = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2Kc}}{\partial v^m} \text{ for } v < v_0. \tag{3}$$

Regressing y_i on $q^K(z_i)$ for the eligible group, we obtain the regression coefficients

$$\widehat{\pi}_{Kt} = (Q_t'Q_t)^{-1}Q_t'Y_t,$$

where $Q_t = (q^K(z_{N_c+1}), q^K(z_{N_c+2}), \cdots, q^K(z_N))'$ and $Y_t = (y_{N_c+1}, \cdots, y_N)'$. The conditional mean function for eligible individuals is estimated by

$$\widehat{E}\{y_i|z_i = z\} = q^K(z)'\widehat{\pi}_{Kt} = q_1^K(z)'\widehat{\pi}_{1Kt} + q_2^K(z)'\widehat{\pi}_{2Kt} \text{ for } v \geq v_0,$$

and the derivative is estimated by

$$\frac{\partial^m \widehat{E}\{y_i|z_i = z\}}{\partial v^m} = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2Kt}}{\partial v^m} \text{ for } v \geq v_0. \tag{4}$$

Notice that the derivative estimators in (3) and (4) should estimate the same function $\frac{\partial^m \alpha(z)}{\partial^m v}$ but the estimates $\widehat{\pi}_{2Kc}$ and $\widehat{\pi}_{2Kt}$ are not necessarily identical. Our proposal is to fit a smoothed curve

$$\widehat{h}(z) = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2K}}{\partial v^m},$$

where $\widehat{\pi}_{2K}$ solves

$$\widehat{\pi}_{2K} : \min_{\pi_2} \sum_{i=1}^{N_c} (\frac{\partial^m q_2^K(z_i)'\widehat{\pi}_{2Kc}}{\partial v^m} - \frac{\partial^m q_2^K(z_i)'\pi_2}{\partial v^m})^2 + \sum_{i=N_c+1}^{N} (\frac{\partial^m q_2^K(z_i)'\widehat{\pi}_{2Kt}}{\partial v^m} - \frac{\partial^m q_2^K(z_i)'\pi_2}{\partial v^m})^2.$$

The baseline effect is estimated by

$$\widehat{\alpha}(z) = q_2^K(z)'\widehat{\pi}_{2K} + (s^{K_2}(w) \otimes p^J(v))'\widehat{c}$$

where $\widehat{c}$ is obtained by

$$\text{regressing } y_i - q_2^K(z_i)'\widehat{\pi}_{2K} \text{ on } s^{K_2}(w_i) \otimes p^J(v_i), i = 1, 2, ..., N_c.$$

After computing the baseline effect, we estimate the treatment effect function by

$$\text{regressing } y_i - q_2^K(z_i)'\widehat{\pi}_{2K} - (s^{K_2}(w_i) \otimes p^J(v_i))'\widehat{c} \text{ on } t^J(v_i)' \otimes s^{K_2}(w_i)', i = N_c + 1, ..., N.$$

we obtain the OLS estimates $\widehat{\gamma}$ and

$$\widehat{\beta}(z) = \left(t^J(v)' \otimes s^{K_2}(w)' \right) \widehat{\gamma}.$$

The consistency of $\widehat{\alpha}(z)$ and $\widehat{\beta}(z)$ can be easily established using arguments similar to those of Newey (1997) and Ai and Chen (2003).

To illusrate this step, we consider the case $z = v$ and $\beta(z) = \beta_0 + \beta_1 v$. So $J = 2$ and $t^J(v) = (1, v)$. For some $K_1 > 2$, denote $t^{K_1}(v) = (1, v, v^2, ..., v^{K_1})'$. Obtain the least squares estimates $\widehat{\pi}_{Kc} = (\widehat{\pi}_{Kc0}, \widehat{\pi}_{Kc1}, ..., \widehat{\pi}_{KcK_1})$ from

$$\text{regressing } y_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } i = 1, 2, ..., N_c.$$

Then

$$\frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} = 2\widehat{\pi}_{Kc2} + 6v\widehat{\pi}_{Kc3} + ... + K_1(K_1 - 1)v^{K_1 - 2}\widehat{\pi}_{KcK_1} \text{ for } v < v_0.$$

Obtain the least squares estimates $\widehat{\pi}_{Kt} = (\widehat{\pi}_{Kt0}, \widehat{\pi}_{Kt1}, ..., \widehat{\pi}_{KtK_1})$ from

$$\text{regressing } y_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } i = N_c + 1, ..., N.$$

Then

$$\frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} = 2\widehat{\pi}_{Kt2} + 6v\widehat{\pi}_{Kt3} + ... + K_1(K_1 - 1)v^{K_1 - 2}\widehat{\pi}_{KtK_1} \text{ for } v \geq v_0.$$

Obtain $\widehat{\pi}_{2K}$ from

$$\text{regressing } \frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} \text{ on } 2, 6v_i, ..., K_1(K_1 - 1)v_i^{K_1 - 2} \text{ for } i = 1, ..., N.$$

Then we estimate the baseline effect by

$$\widehat{\alpha}(v) = \widehat{c}_0 + \widehat{c}_1 v + (v^2, ..., v^{K_1})\widehat{\pi}_{2K},$$

where $\widehat{c}_0, \widehat{c}_1$ are obtained by

$$\text{regressing } y_i - (v_i^2, ..., v_i^{K_1})\widehat{\pi}_{2K} \text{ on } 1, v_i \text{ for } i = 1, 2, ..., N_c.$$

After computing the baseline effect, we estimate the treatment effect function by

$$\text{regressing } y_i - (v_i^2, ..., v_i^{K_1})\widehat{\pi}_{2K} - (1, v_i)\widehat{c} \text{ on } 1, v_i, i = N_c + 1, ..., N.$$

we obtain the OLS estimates $\widehat{\gamma}$ and

$$\widehat{\beta}(v) = (1, v)\,\widehat{\gamma}.$$

2.2 Fuzzy design

For the fuzzy design, we need to estimate the probability of receiving treatment. A simplest approach is to fit a probit model by

$$\text{probit regression of } x_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } x_i = 1.$$

Denote the fitted values by

$$\widehat{P}_i = \Phi((1, v_i, ..., v_i^{K_1})\widehat{\delta}), i = 1, 2, ..., N$$

where $\widehat{\delta}$ is probit estimate. Denote

$$\widetilde{y}_i = \frac{y_i}{\widehat{P}_i}, i = 1, 2, ..., N.$$

The treatment effect function can be estimated by applying the procedure of sharp design with y_i replaced by $\widetilde{y}_i$.

3 Misspecification Bias

The proposed procedure produces consistent estimates under the functional form restriction (2). Since the true treatment effect function is unknown, it is a legitimate concern that the proposed procedure may produce biased estimates if (2) is not satisfied. In the case that the functional form restriction (2) is false, what does the proposed procedure actually estimate? To investigate this question, consider the sharp design. Notice that differentiation now yields

$$\begin{aligned}
\frac{\partial^m E\{y_i|z_i = z\}}{\partial v^m} &= \frac{\partial^m \alpha(z)}{\partial v^m} \text{ over } z \in \mathcal{Z}_- \text{ almost everywhere,} \\
\frac{\partial^m E\{y_i|z_i = z\}}{\partial v^m} &= \frac{\partial^m [\alpha(z) + \beta(z)]}{\partial v^m} \text{ over } z \in \mathcal{Z}_+ \text{ almost everywhere.}
\end{aligned}$$

The smoothing step in the proposed procedure, in this case, finds an approximation $h_K(z) = q^K(z)'\pi_K$ to $\frac{\partial^m E\{y_i|z_i=z\}}{\partial v^m}$ in the sense that

$$\sup_{z\in\mathcal{Z}_-}\left|\frac{\partial^m \alpha(z)}{\partial v^m} - h_K(z)\right| \to 0 \text{ and } \sup_{z\in\mathcal{Z}_+}\left|\frac{\partial^m [\alpha(z)+\beta(z)]}{\partial v^m} - h_K(z)\right| \to 0.$$

Suppose that $\frac{\partial^m E\{y_i|z_i=z\}}{\partial v^m}$ is bounded everywhere except for the threshold (i.e. $v = v_0$). Then $h_K(z)$ is bounded for all z and all K. Let $b^*(w)'t^J(v)$ denote the least squares projection of $\beta(z)$ and satisfy

$$b^*(w)'t^J(v_0) = \beta(v_0, w) \text{ for all } w, \tag{5}$$

in the sense that

$$b^*(w) : \arg\min E\{[\beta(z) - b(w)'t^J(v)]^2|w\}$$

subject to the constraint (5). Write

$$\beta(z) = b^*(w)'t^J(v) + r_J(z).$$

Let $g_K(z)$ denote the integration of $h_K(z)$ and let $c_K(w)$ solve

$$\min_{c(\cdot)} E\left\{(y_i - g_K(z_i) - c(w_i)'t^J(v_i))^2|w_i = w, x_i = 0\right\}.$$

Then $\alpha_K(z) = g_K(z) + c_K(w)'t^J(v)$ is an approximation to $f_J(z) = \alpha(z)$ over $z \in \mathcal{Z}_-$ and $f_J(z) = \alpha(z) + r_J(z)$ over $z \in \mathcal{Z}_+$. It is worth pointing out that constraint (5) is critical for this approximation result. If this constraint is not satisfied, $f_J(z)$ is discontinuous at the threshold. From the approximation theory, a discontinuous function cannot be approximated by a sequence of functions with bounded derivatives. However, the derivative of $\alpha_K(z)$ is $h_K(z)$ which is bounded for all z and K.

Let $b_K(w)$ solve:

$$\min_{b(\cdot)} E\left\{y_i - \alpha_K(z_i) - b_K(w_i)'t^J(v_i))^2|w_i = w, x_i = 1\right\}.$$

Then, $b_K(w)$ is an approximation to $b^*(w)$. Hence the estimator $\widehat{\beta}(z)$ is a consistent estimator of $\beta^*(z) = b^*(w)'t^J(v)$.

There are two implications of this approximation result. First, it implies that the average treatment effect for threshold individuals can be estimated consistently by

$$\widehat{\mu} = \frac{1}{N - N_c} \sum_{i=N_c+1}^{N} \widehat{\beta}(v_0, w_i)$$

regardless of whether (2) holds. Moreover, because the average treatment effect estimator uses all eligible individuals, it is root-N consistent. The root-N consistency is achieved by exploiting the derivative information which is not utilized by the existing estimators. The second implication is that the approximation error can be small for large J if $t^J(v)$ is part of some approximating basis functions. Thus, for the non-threshold individuals, $\widehat{\beta}(z)$ is an estimate for $\beta(z)$ with small bias if we choose J large enough.

4 Conclusion

In this paper, we propose a new estimation for regression discontinuity model under the additional assumption that the treatment effect function satisfies some functional form restriction, and the average outcome in the absence of treatment has higher order of continuous derivatives. These additional assumptions allow us to remove the treatment effect through differentiation and to use all observations, not just the few near the threshold, to estimate the model. Thus, our procedure yields better estimates than the existing procedures. We also discuss the consequence of misspecifying the treatment effect function, and find that our estimators still provide useful information about the treatment effect even if our parameterization is false.

5 Reference

Ai, C. (2005a), "A Nonparametric Maximum Likelihood Estimation of Conditional Moment Restrictions Models," *Manuscript.*

Ai, C. (2005b), "An Alternative Root-N Consistent Estimator for Panel Data Binary Choice Models," *Manuscript*

Ai, C, and X. Chen (2004), "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditional Variables", forthcoming in *Journal of Econometrics*

Ai, C, and X. Chen (2003), "Efficient Estimation of Conditional Moment Restrictions Models Containing Unknown Functions", *Econometrica* 71, pp. 1795-1843.

Angrist, J. and V. Lavy (1999), "Using Maimonides Rule to Estimate the Effects of Class Size on Scholastic Achievement", *Quarterly Journal of Economics* 114, pp. 533 - 575.

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

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola (2005): "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools," *American Economic Review* 95, pp. 1237-1258.

Chay, Kenneth Y. and Michael Greenstone (2005): "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy* 113, pp. 376-424.

DiNardo, John and David S. Lee (2004): "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001," *Quarterly Journal of Economics*

Donald, Imbens, and Newey (2004), "Empirical likelihood estimation and consistent tests with conditional moment restrictions", *Journal of Econometrics* 117, pp. 55-93.

Gallant, A.R. and D.W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica* 55, pp. 363-390.

Hahn, J., P. Todd, and W. Van der Klaauw (2000), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica* 69, pp. 201-209

Newey, Whitney K. (1997) "Convergence rates and Asymptotic Normality for Series Estimators", *Journal of Econometrics* 79, pp. 147 - 168.

Porter, J. (1998), "Estimation of Regression Discontinuity", Manuscript.