HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 19/16

Regression Discontinuity with Integer Running Variable and Non-Integer Cutoff: Dental Care Program Effect on Expenditure

Myoung-jae Lee; Sang-soo Park and Hyae-chong Shim

July 2019

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Regression Discontinuity with Integer Running Variable and Non-Integer Cutoff: Dental Care Program Effect on Expenditure

(June 2019)

Myoung-jae Lee*
Dept. Economics
Korea University
Seoul Korea 02841
myoungjae@korea.ac.kr

Sang-soo Park
Dept. Economics
Korea University
Seoul Korea 02841
starpac@korea.ac.kr

Hyae-chong Shim
KITA, 511 Youngdongdae-ro
Gangnam-gu
Seoul Korea 06164
assashim@gmail.com

In regression discontinuity (RD), the treatment is determined by a continuous running variable G crossing a known cutoff c or not. However, often G is observed only as a rounded-down integer S (e.g., birth year observed instead of birth date), and c is not an integer. In this case, the "cutoff sample" (the observations with the same S value around c) cannot be used, because it is not clear whether their G actually crossed c or not. This paper shows that if the distribution of the measurement error $e \equiv G - S$ is specified, then despite non-integer c, the cutoff sample can be used fruitfully in estimating the treatment effect and in testing for the distributional assumption on e. Particularly, there are good reasons to believe that e is uniform on $[0, 1]$, not least because e is close to a popular way how pseudo uniform random numbers are generated in simulation studies. Also, whereas two-step estimation has been proposed in the RD literature, we show that the treatment effect can be estimated with single-step OLS/IVE as in typical RD with G observed. A simulation study and an empirical analysis for effects of a dental care support program on dental expenditure are provided.

* Corresponding author.

JEL Classification Codes: C21, C24, I18.

Key Words: regression discontinuity, integer running variable, non-integer cutoff.

1 Introduction

Regression discontinuity (RD) is widely used to find treatment effects with observational data; see Imbens and Lemieux (2008), Lee and Lemieux (2010), Choi and Lee (2017), Cattaneo and Escanciano (2017), Cattaneo et al. (2019), and references therein. In typical RD, there are a binary treatment D, an outcome Y, and a continuous running variable G, where D is determined by whether G crosses a known cutoff c or not. If D is fully determined by G so that $E(D|G) = D$, then the RD is a 'sharp RD (SRD)', otherwise, if D is partly determined by an error ε, the RD is a 'fuzzy RD (FRD)'.

The main attraction of RD is that local observations around c are balanced in all covariates, observed or not. Hence, using only such local observations enables estimating the treatment effect while avoiding confounders despite no covariate controlled. Theoretically speaking, this 'local randomization' requires using a small neighborhood of c, but in practice, using a sizable neighborhood is unavoidable if there are not too many observations around c.

Often in RD, the rounded-down integer S of G is observed instead of G:

$$S \equiv \lfloor G \rfloor.$$

There are many examples for this, but the most notable is birth year observed for confidentiality instead of birth date. In this case, if c is not an integer, it is unclear whether individuals around $G = c$ are treated or not. For instance, Y is health behavior and an education law D goes into effect on Sept. 1, 2016 so that a student is subject to the law if born on or after Sept. 1, 2010, i.e., $D = 1[2010.67 \le G]$ with birth date G and $c = 2010.67$, where $1[A] \equiv 1$ if A holds and 0 otherwise. Here, we do not know whether the "cutoff sample" (i.e., those with $S = 2010$) are treated or not.

In the literature, a couple of studies addressed RD with an integer running variable. Lee and Card (2008) examined integer (or discrete) running variables where c is also an integer. They recommended clustered inference with clustering at each integer point, which is problematic, however, as pointed out by Kolesár and Rothe (2018). Dong (2015) proposed two-step estimators, assuming that the measurement error

$$e \equiv G - \lfloor G \rfloor = G - S \text{ is independent of } S \text{ ("} e \amalg S \text{")}$$

and follows a known distribution. Any distribution on $[0, 1]$ can be adopted, but the most basic is the uniform distribution on $[0, 1]$ ("$Uni[0, 1]$"), which was also adopted in Dong's

(2015) applications. Dong (2015) addressed integer c, and suggested dropping the cutoff sample (i.e., the sample with $S = \lfloor c \rfloor$) if c is not an integer.

In this paper, under a known distribution for e, we address integer S and any c that may not be an integer to show that the cutoff sample can be used fruitfully in estimating the treatment effect and in testing for the distributional assumption on e. Although there are many possible distributions to adopt for e, we focus on $e \sim Uni[0,1]$, as there are good reasons to believe $e \sim Uni[0,1]$, not least because $e \equiv G - \lfloor G \rfloor$ is close to a popular way ('linear congruential generator') how pseudo uniform random numbers are generated in simulation studies. That is, '$e \sim Uni[0,1]$' is a highly plausible assumption to make, which will be demonstrated through simulation and empirical studies. When $e \sim Uni[0,1]$, as well known,

$$\mu_k \equiv E(e^k) = \int_0^1 \tau^k d\tau = \frac{1}{k+1}.$$

Other than the above goal with $e \sim Uni[0,1]$, this paper has two other goals. First, we generalize the identification findings of Dong (2015) obtained under integer c to any c, integer or non-integer. Second, whereas Dong (2015) proposed two-stage indirect estimators with bootstrap inference, we show that (one-step direct) ordinary least squares estimator (OLS) or instrumental variable estimator (IVE) is enough for SRD or FRD, respectively, if we modify some regressors used in RD estimation only slightly.

When G is observed, a well-known RD identification finding is (Hahn et al 2001)

$$\beta_d \equiv \frac{E(Y|G=c^+) - E(Y|G=c^-)}{E(D|G=c^+) - E(D|G=c^-)} = E(\text{effect on "compliers" on } S = c^+) \qquad (1.1)$$

where $E(Y|G=c^+) \equiv \lim_{g\downarrow c} E(Y|G=g)$ and $E(Y|G=c^-) \equiv \lim_{g\uparrow c} E(Y|G=g)$; the weakest condition for this complier interpretation appears in Choi and Lee (2018). The first equation of (1.1) is equivalent to (Lee 2016, p.103)

$$E(Y|G) = \beta_d E(D|G) + m(G) \qquad (1.2)$$

for an unknown function $m(G)$ continuous at c. For SRD, the denominator in (1.1) equals one, and $E(D|G)$ in (1.2) equals D.

With the usual normalization $G - c$, implementation of (1.2) differs, depending on how $m(G)$ is specified locally around c: constant, linear, quadratic, etc. For the local linear specification, one would set up

$$Y = \beta_0 + \beta_d D + \beta_1 (G-c) + \beta_{1\delta}\delta(G-c) + U \quad \text{with} \quad E(U|S,e) = 0 \qquad (1.3)$$

where the β's are parameters, U is an error, and

$$\delta \equiv 1[c \leq G].$$

For SRD with $D = \delta$, the OLS of Y on $\{1, D,\ G - c,\ \delta(G - c)\}$ is done using a local sample with $G \simeq c$, and for FRD, IVE is done using δ as an instrument for D.

The local linear model (1.3) has become almost the "industry standard", and the local quadratic version obtains adding $\beta_2(G - c)^2 + \beta_{2\delta}\delta(G - c)^2$ to (1.3):

$$Y = \beta_0 + \beta_d D + \beta_1(G - c) + \beta_{1\delta}\delta(G - c) + \beta_2(G - c)^2 + \beta_{2\delta}\delta(G - c)^2 + error. \quad (1.4)$$

If desired, we can go further than quadratic, but Gelman and Imbens (2019) advise against adding higher-order terms.

The rest of this paper is organized as follows. Section 2 shows why $e \sim Uni[0, 1]$ is plausible, and introduces Dong's (2015) identification finding and two-step estimators under integer c. Section 3 presents our generalized identification finding for any c (integer or non-integer) and the test for $e \sim Uni[0, 1]$ using the cutoff sample, and explains in detail the OLS and IVE that are consistent for any c. Section 4 provides a simulation study, where the proposed estimators perform well even when the distribution of e is extremely far from uniform. Section 5 presents an empirical analysis for effects of a Korean dental care support program on the dental expenditure of the elderly. Finally, Section 6 concludes. The appendix contains lengthy proofs. What is observed is (S_i, Y_i) for SRD and (D_i, S_i, Y_i) for FRD, $i = 1, ..., N$, which are independent and identically distributed; we often omit the subscript i. Since the expression 'running variable' appears frequently, call it just 'score'; S for score and G for "genuine" score.

2 Uniform Error and Two-Step Procedure

In most RD applications with G, the location normalization $G - c$ is done as in (1.3) so that the cutoff becomes 0 for $G - c$. With $S = \lfloor G \rfloor$, the analogous location normalization is

$$S - \lfloor c \rfloor \qquad (\Longrightarrow\ S = \lfloor c \rfloor \text{ becomes } 0)$$

so that $S = 0$ is the normalized cutoff if c is an integer; if c is not an integer, the fraction c shows how far off the cutoff is from 0. In the education law example, this normalization redefines S as $S - 2010$ and gives $c = 0.67$. Because $G = S + e$, G equals e for the cutoff sample ($S = 0$) after the normalization.

2.1 Measurement Error as Pseudo Uniform Number

To show $e \equiv G - \lfloor G \rfloor$ is close to the popular linear congruential generator for pseudo uniform random numbers, note the general form of linear congruential generator (see, e.g., Toral and Colet 2014):

$$m_i = a \times m_{i-1} + c \quad \text{modulo } M$$

where $0 < a < M$, $0 \leq c < M$, and M is a positive integer.

For instance, with $a = 7$, $c = 0$, $M = 10$ and $m_0 = 3$, the generator becomes '$m_i = 7m_{i-1}$ modulo 10', from which the first few numbers are $(m_1, m_2, m_3, ...) = (1, 7, 9, ...)$, where the first number 1 is the remainder in dividing $7m_0 = 7 \times 3 = 21$ by 10, the second number 7 is the remainder in dividing $7m_1 = 7 \times 1 = 7$ by 10, and the third number 9 is the remainder in dividing $7 \times m_2 = 7 \times 7$ by 10. From $(m_1, m_2, m_3, ...) = (1, 7, 9, ...)$, the pseudo uniform numbers are $(u_1, u_2, u_3, ...) = (1/10,\ 7/10,\ 9/10, ...)$, which can be written succinctly as

$$u_i = \frac{m_i}{M} - \lfloor \frac{m_i}{M} \rfloor; \text{ e.g., } u_1 = \frac{21}{10} - \lfloor \frac{21}{10} \rfloor = \frac{1}{10} \text{ and } u_2 = \frac{7}{10}.$$

Of course, with $M = 10$, we quickly run into the problem of the same numbers generated repeatedly, which is why a large number such as $M = 2^{32}$ with $a = 1812433253$ and $c = 1$ is used to generate millions of different pseudo uniform numbers in practice.

For birth year, analogously to the preceding display, we have

$$e = \frac{\textit{birth date}}{\textit{birth year}} - \lfloor \frac{\textit{birth date}}{\textit{birth year}} \rfloor = \frac{\textit{birth date}}{\textit{birth year}} - 1 = \textit{fractional birth date}.$$

Using individuals born different years amounts to using different M's. In short, our main point is that e *is structured in the same way pseudo uniform random numbers are often generated.* Having a small M such as (year) 1990 and 2010 makes the same e getting repeated, which is natural for birth date though, because many people are born on the same date.

Dong (2105, p.428) noted that, for fractional birth date,

$$\begin{aligned} \text{first 4 moments of } Uni[0,1] \quad &: \quad 0.50,\ 0.33,\ 0.25,\ 0.20; \\ \text{first 4 moments of birth date in NLSY97} \quad &: \quad 0.50,\ 0.34,\ 0.25,\ 0.20; \\ \text{first 4 moments of birth date in Italy 2001-2011} \quad &: \quad 0.51,\ 0.34,\ 0.26,\ 0.20. \end{aligned}$$

That is, fractional birth date is almost $Uni[0, 1]$. This kind of small deviations from $Uni[0, 1]$ should be harmless, because even extreme deviations from $Uni[0, 1]$ will be shown to have only a limited influence on our estimators in the simulation study below.

2.2 Two-Step Identification and Estimation

Consider SRD with integer c, so that $c = 0$ after the normalization $S - \lfloor c \rfloor$. Then $\delta \equiv 1[c \le G]$ is fully determined by S because

$$\text{integer } c \implies \delta \equiv 1[c \le G] = 1[c \le S] \implies E(\delta|S) = \delta. \tag{2.1}$$

Along with this, a key equation to be used under $e \sim Uni[0,1] \amalg S$ is

$$E(G|S) = E(S + e|S) = S + E(e|S) = S + 0.5. \tag{2.2}$$

This reveals that another distribution can be adopted without difficulty: for any distribution with a known $E(e|S) = E(e) = \mu_1$, we can use $E(G|S) = S + \mu_1$.

Since G is not observed, take $E(\cdot|S)$ on (1.3) with $D = \delta$ for SRD to obtain

$$E(Y|S) = \beta_0 + \beta_d\delta + \beta_1(S + 0.5) + \beta_{1\delta}\delta(S + 0.5); \tag{2.3}$$

$E(\delta|S) = \delta$ in (2.1) is used. Putting the terms with δ, S and δS in (2.3) together gives

$$E(Y|S) = (\beta_0 + 0.5\beta_1) \ + (\beta_d + 0.5\beta_{1\delta})\delta \ + \beta_1 S \ + \beta_{1\delta}\delta S. \tag{2.4}$$

This shows an important point: *ignoring the integer nature of S to proceed as if S were continuous fails to identify β_d with the slope of $\delta = D$, unless $\beta_{1\delta} = 0$ ($\Longleftrightarrow$ no change in slope).* Hence, Dong (2015) identified β_d in SRD indirectly with

$$\beta_d = \{\delta \text{ slope of } E(Y|S) \text{ in (2.4)}\} - 0.5\{\delta S \text{ slope of } E(Y|S) \text{ in (2.4)}\} \tag{2.5}$$

and estimated β_d in two steps: do the OLS of Y on $(1, \delta, S, \delta S)$ and plug the slope estimates of δ and δS into (2.5). Dong (2015) then proposed bootstrap for asymptotic inference.

For FRD, considering the OLS of D on $(\delta, 1, S, \delta S)$, the generalization of (2.5) is

$$\beta_d = \frac{\{\delta \text{ slope in } E(Y|S)\} \ - \ 0.5\{\delta S \text{ slope in } E(Y|S)\}}{\{\delta \text{ slope in } E(D|S)\} \ - \ 0.5\{\delta S \text{ slope in } E(D|S)\}}. \tag{2.6}$$

This suggests a multi-step procedure to estimate β_d by plugging in the OLS estimates for (2.4) and the OLS estimates of D on $(\delta, 1\ S, \delta S)$.

For the quadratic model (1.4), considering the OLS of Y and D on $(1, \delta,\ S, \delta S,\ S^2, \delta S^2)$, Dong's (2015) identification result for FRD generalizing (2.6) is

$$\beta_d = \frac{\{\delta \text{ slope in } E(Y|S)\} \ - \{\delta S \text{ slope in } E(Y|S)\}/2 + \{\delta S^2 \text{ slope in } E(Y|S)\}/6}{\{\delta \text{ slope in } E(D|S)\} \ - \{\delta S \text{ slope in } E(D|S)/2\} + \{\delta S^2 \text{ slope in } E(D|S)\}/6}; \tag{2.7}$$

for SRD, replace the denominator of (2.7) with one. Dong (2015) provided further results for higher-order models.

3 Any Cutoff, Main OLS/IVE and Test

This section is our main section allowing for any cutoff c, not just integer c; i.e., $c \neq 0$ is allowed after the normalization. First, we generalize the identification findings with integer c in the preceding section to any c. Second, we consider OLS and IVE for linear $m(\cdot)$ in (1.3) without using the cutoff sample. Third, the cutoff sample is used to test for $e \sim Uni[0,1]$ in the linear model, and in case the uniformity is not rejected, we introduce OLS and IVE using the cutoff sample additionally. Fourth, we deal with quadratic $m(\cdot)$ in (1.4). Often we address SRD first, because FRD can be handled secondly as can be seen in (2.5) to (2.6).

The importance of the cutoff sample diminishes as the number of the integer support points in use increases. But since we do not want to compare individuals too far apart, it is better to keep the number of integer support points small, which may result in the proportion of the cutoff sample being substantial for the local sample in use. With h denoting the localizing bandwidth, the number of support points is 3 with $h = 1$ (for $S = 0, \pm 1$), 5 with $h = 2$ (for $S = 0, \pm 1, \pm 2$), and so on.

3.1 Identification with Any Cutoff for Linear/Quadratic Model

Suppose c is any cutoff. With the linear model (1.3) for FRD, the appendix proves

$$\beta_d = \frac{\{\delta \text{ slope in } E(Y|S)\} - (0.5 - c)\{\delta S \text{ slope in } E(Y|S)\}}{\{\delta \text{ slope in } E(D|S)\} - (0.5 - c)\{\delta S \text{ slope in } E(D|S)\}} \tag{3.1}$$

which becomes (2.6) when $c = 0$ for integer c; for SRD, the denominator in (3.1) should be replaced with one. This reveals that the adjustment using the slopes of δS is unnecessary if either $c = 0.5$ (as in our empirical analysis below) or $\beta_{1\delta} = 0$ ($\Longleftrightarrow$ no change in slope). That is, *if either $c = 0.5$ or no change in slope in the linear model, then we can proceed with S as we do with G in the usual RD to ignore the integer nature of S*. However, the following shows that this finding is specific to the linear model.

For the quadratic model (1.4) in FRD, the appendix proves that β_d equals

$$\frac{\{\delta \text{ slope in } E(Y|S)\} - (0.5 - c)\{\delta S \text{ slope in } E(Y|S)\} + c^*\{\delta S^2 \text{ slope in } E(Y|S)\}}{\{\delta \text{ slope in } E(D|S)\} - (0.5 - c)\{\delta S \text{ slope in } E(D|S)\} + c^*\{\delta S^2 \text{ slope in } E(D|S)\}}$$
$$\text{where} \quad c^* \equiv (1/6) - c + c^2; \tag{3.2}$$

for SRD, the denominator should be replaced with one. Clearly, (3.2) becomes (2.7) for integer c (i.e., $c = 0$). If $c = 0.5$, then the adjustment term with δS drops out due to $0.5 - c$, but c^* becomes zero only when $c = 0.2115$ or 0.7885. That is, the two adjustment terms in (3.2) cannot drop out together.

3.2 Estimation for Any Cutoff without Cutoff Sample

For SRD, consider a linear model that is equivalent to (1.3):

$$Y = \beta_0 \ + \beta_d \delta \ + \beta_-(1-\delta)(G-c) + \beta_+ \delta(G-c) \ + U. \tag{3.3}$$

Using $e \sim Uni[0,1] \amalg S$, as in (2.2), we have

$$E(G-c|S) = E(S+e-c|S) = S + E(e|S) - c = S_{0.5c}, \quad S_{0.5c} \equiv S + 0.5 - c; \tag{3.4}$$

any distribution can be used for e as long as $E(e|S)$ is known. To ease presentation, define

$$\delta_0 \equiv 1[S=0], \quad \delta_- \equiv 1[S \le -1] \quad \text{and} \quad \delta_+ \equiv 1[1 \le S].$$

Since '$S \le -1 \Longrightarrow \delta = 0$' and '$S \ge 1 \Longrightarrow \delta = 1$', (3.3) gives

$$\begin{aligned} S &\le -1 : E(Y|S) = \beta_0 + \beta_- E(G-c|S) = \beta_0 + \beta_- S_{0.5c}; \\ S &\ge 1 : E(Y|S) = \beta_0 + \beta_d + \beta_+ E(G-c|S) = \beta_d + \beta_0 + \beta_+ S_{0.5c}. \end{aligned} \tag{3.5}$$

Hence, ignoring the cutoff sample $S = 0$ gives

$$E(Y|S) = \beta_0 + \beta_d \delta_+ + \beta_- \delta_- S_{0.5c} + \beta_+ \delta_+ S_{0.5c}. \tag{3.6}$$

Then $\beta \equiv (\beta_0, \beta_d, \beta_-, \beta_+)'$ can be estimated by the OLS, say $\hat{\beta}_{ols,1}$, of Y on

$$W_1 \equiv (1,\ \delta_+,\ \delta_- S_{0.5c},\ \delta_+ S_{0.5c})';$$

β_d is estimated by the OLS slope of δ_+. This OLS as well as the IVE below require $h \ge 2$.

Whereas (3.1) suggests a two-stage estimation, β_d can be estimated simply by the OLS of Y on W_1 that has modified regressors based on $S_{0.5c}$. For FRD, apply IVE to

$$Y = \beta_0 + \beta_d D + \beta_- \delta_- S_{0.5c} + \beta_+ \delta_+ S_{0.5c} + error \qquad \text{where}$$
$$\text{the regressor is} \quad X_1 \equiv (1,\ D,\ \delta_- S_{0.5c},\ \delta_+ S_{0.5c})' \quad \text{and} \quad \text{the instrument is } W_1.$$

The above OLS and IVE are the same as the corresponding estimators in Dong (2015), because the treatment status ambiguity occurs only for the cutoff sample when c is non-integer. The equality can be seen by rewriting (3.6) as, with $\beta_{1\delta} = \beta_+ - \beta_-$,

$$\begin{aligned} &\beta_0 + \beta_d \delta_+ + \beta_- S_{0.5c} + \beta_{1\delta} \delta_+ S_{0.5c} = \beta_0 + \beta_d \delta_+ + \beta_-(S + 0.5 - c) + \beta_{1\delta} \delta_+ (S + 0.5 - c) \\ &= \beta_0 + \beta_-(0.5-c) \ + \{\beta_d + \beta_{1\delta}(0.5-c)\}\delta_+ \ + \beta_- S \ + \beta_{1\delta} \delta_+ S. \end{aligned}$$

This gives the numerator of (3.1), because of $\delta = \delta_+$ when the cutoff sample is unused.

3.3 Test and Estimation with Cutoff Sample

Our test is based on the next equation proven in the appendix for SRD:

$$E(Y|S=0) = \beta_0 + \beta_d(1-c) + \beta_-(-0.5c^2) + \beta_+ 0.5(1-c)^2 \quad \text{under } e \sim Uni[0,1] \amalg S. \quad (3.7)$$

This gives a method-of-moment test statistic:

$$\frac{1}{\sqrt{N}} \sum_i m_1(Y_i, \hat{\beta}_{ols,1}), \quad m_1(Y,\beta) \equiv \delta_0\{Y - \beta_0 - \beta_d(1-c) - \beta_-(-0.5c^2) - \beta_+ 0.5(1-c)^2\}. \quad (3.8)$$

Since the cutoff sample is not used in obtaining $\hat{\beta}_{ols,1}$, we can test for $e \sim Uni[0,1] \amalg S$ using $E\{m_1(Y,\beta)\} = 0$ by plugging in c and $\hat{\beta}_{ols,1}$.

Following Lee (2010, p. 109), the test statistic can be shown to be asymptotically normal with the variance estimable by recalling W_1,

$$\frac{1}{N}\sum_i m_1(Y_i, \hat{\beta}_{ols,1})^2 \ + \ \hat{p}^2 \frac{1}{N}\sum_i C' \eta_{1i}(Y_i, \hat{\beta}_{ols,1}) \eta_{1i}(Y_i, \hat{\beta}_{ols,1})' C_1$$

$$\text{where} \quad C_1 \equiv \{1,\ 1-c,\ -0.5c^2,\ 0.5(1-c)^2\}', \quad \hat{p} \equiv \frac{1}{N}\sum_i 1[S_i = 0],$$

$$\eta_{1i}(Y_i, \hat{\beta}_{ols,1}) \equiv (\frac{1}{N}\sum_i 1[S_i \neq 0] W_{1i} W_{1i}')^{-1} \cdot 1[S_i \neq 0] W_{1i}(Y_i - W_{1i}\hat{\beta}_{ols,1}).$$

For FRD, do the test with Y replaced by D, as the relationship between δ and Y in SRD is analogous to that between δ and D in FRD. Denoting the OLS of D on W_1 as $\hat{\alpha}_{ols,1}$, the test statistic is $N^{-1/2}\sum_i m_1(D_i, \hat{\alpha}_{ols,1})$ with the variance estimable by

$$\frac{1}{N}\sum_i m_1(D_i, \hat{\alpha}_{ols,1})^2 \ + \ \hat{p}^2 \frac{1}{N}\sum_i C_1' \eta_{1i}(D_i, \hat{\alpha}_{ols,1}) \eta_{1i}(D_i, \hat{\alpha}_{ols,1})' C_1.$$

Differentiating $m_1(Y,\beta)$ with respect to c under $\beta_- = \beta_+ = 0$ gives β_d. Hence, if $\beta_d < 0$, then the test statistic decreases as c increases. If the test statistic value is negative, then the value of c that would make the test statistic zero is lower than the nominal value of c, as if the treatment had started earlier. This situation seems to have occurred in our empirical analysis where $c = 0.5$ and Y is dental expenditure of South Korean elders, because people can wait for many dental treatments. For example, if one needs a denture but is not qualified for a dental care support until age 70, then he/she would wait until age 70, which is likely when the person is poor as many elders in South Korea are.

If the test does not reject, we can do the OLS/IVE using the cutoff sample additionally. For this, combine (3.6) and (3.7) to get, for any integer S,

$$E(Y|S) = \beta_0 + \beta_d\{(1-c)\delta_0 + \delta_+\} + \beta_-(-0.5c^2\delta_0 + \delta_- S_{0.5c}) + \beta_+\{0.5(1-c)^2\delta_0 + \delta_+ S_{0.5c}\}. \quad (3.9)$$

Compared with (3.6) excluding $\delta_0 \equiv 1[S=0]$, the regressor for β_d has $(1-c)\delta_0$ extra; also, the regressors for β_- and β_+ have extra components involving δ_0. The appearance of $(1-c)\delta_0$ for β_d makes a good sense, because the proportion of the treated in the cutoff sample is $1-c$. For estimation, do the OLS of Y on (with $h \geq 2$)

$$W_{1c} \equiv \{1,\ (1-c)\delta_0 + \delta_+,\ -0.5c^2\delta_0 + \delta_- S_{0.5c},\ 0.5(1-c)^2\delta_0 + \delta_+ S_{0.5c}\}'. \quad (3.10)$$

For FRD, do, using W_{1c} as an instrument for X_{1c},

$$\text{IVE of } Y \text{ on } X_{1c} \equiv \{1,\ D,\ -0.5c^2\delta_0 + \delta_- S_{0.5c},\ 0.5(1-c)^2\delta_0 + \delta_+ S_{0.5c}\}'.$$

3.4 Estimation and Test for Quadratic Model

Consider a quadratic model equivalent to (1.4) for SRD:

$$Y = \beta_0 + \beta_d\delta + \beta_-(1-\delta)(G-c) + \beta_+\delta(G-c) + \beta_{--}(1-\delta)(G-c)^2 + \beta_{++}\delta(G-c)^2 + error.$$

The appendix shows that, for SRD without the cutoff sample, do

$$\text{OLS } \hat\beta_{ols,2} \text{ of } Y \text{ on } W_2 \equiv (1,\ \delta_+,\ \delta_- S_{0.5c},\ \delta_+ S_{0.5c},\ \delta_- S^*,\ \delta_+ S^*)'$$
$$\text{where} \quad S^* \equiv S^2 + (1-2c)S + (1/3) - c + c^2.$$

For FRD, using W_2 as instruments for X_2, do

$$\text{IVE of } Y \text{ on } X_2 \equiv (1,\ D,\ \delta_- S_{0.5c},\ \delta_+ S_{0.5c},\ \delta_- S^*,\ \delta_+ S^*)'.$$

Regarding the test for SRD, the appendix shows that the moment to be used is

$$m_2(Y,\beta) \equiv \delta_0\{Y - \beta_0 - \beta_d(1-c) - \beta_-(-0.5c^2) - \beta_+ 0.5(1-c)^2 - \beta_{--}\frac{c^3}{3} - \beta_{++}\frac{(1-c)^3}{3}\}.$$

The asymptotic variance of $N^{-1/2}\sum_i m_2(Y_i, \hat\beta_{ols,2})$ can be estimated with

$$\frac{1}{N}\sum_i m_2(Y_i, \hat\beta_{ols,2})^2 \ + \ \hat p^2 \frac{1}{N}\sum_i C_2' \eta_{2i}(Y_i, \hat\beta_{ols,2})\eta_{2i}(Y_i, \hat\beta_{ols,2})' C_2$$
$$\text{where} \quad C_2 \equiv \{1,\ 1-c,\ -\frac{c^2}{2},\ \frac{(1-c)^2}{2},\ \frac{c^3}{3},\ \frac{(1-c)^3}{3}\}',$$
$$\eta_{2i}(Y_i, \hat\beta_{ols,2}) \equiv (\frac{1}{N}\sum_i 1[S_i \neq 0] W_{2i} W_{2i}')^{-1} \cdot 1[S_i \neq 0] W_{2i}(Y_i - W_{2i}\hat\beta_{ols,2}).$$

For FRD, the test statistics is $N^{-1/2}\sum_i m_2(D_i,\hat{\alpha}_{ols,2})$ where $\hat{\alpha}_{ols,2}$ is the OLS of D on W_2. The asymptotic variance can be estimated with

$$\frac{1}{N}\sum_i m_2(D_i,\hat{\alpha}_{ols,2})^2 \ + \ \hat{p}^2\frac{1}{N}\sum_i C_2'\eta_{2i}(D_i,\hat{\alpha}_{ols,2})\eta_{2i}(D_i,\hat{\alpha}_{ols,2})'C_2.$$

As for the estimation using the cutoff sample additionally, do

$$\text{OLS of } Y \text{ on } W_{2c} \equiv \{1,\ (1-c)\delta_0+\delta_+,\ -0.5c^2\delta_0+\delta_-S_{0.5c},$$
$$0.5(1-c)^2\delta_0+\delta_+S_{0.5c},\ \frac{c^3}{3}\delta_0+\delta_-S^*,\ \frac{(1-c)^3}{3}\delta_0+\delta_+S^*\}'.$$

For FRD, using W_{2c} as instruments for X_{2c}, do

$$\text{IVE of } Y \text{ on } X_{2c} \equiv \{1,\ D,\ -0.5c^2\delta_0+\delta_-S_{0.5c},$$
$$0.5(1-c)^2\delta_0+\delta_+S_{0.5c},\ \frac{c^3}{3}\delta_0+\delta_-S^*,\ \frac{(1-c)^3}{3}\delta_0+\delta_+S^*\}'.$$

4 Simulation Study

Our base simulation design is

Each S value has 500 or 1000 observations, 5000 repetitions, $V \sim N(0,1)$,

$D = (1-\delta)1[V < -0.5] + \delta 1[V > -0.5]$, $G = S + e$, $c = 0.2,\ 0.5,\ 0.9$,

$Y = \beta_0 + \beta_d D + \beta_-(G-c) + \beta_+(G-c) + V + N(0,1)$, $\beta_0 = \beta_d = 1$, $\beta_- = 0.5$, $\beta_+ = 1$.

This is a FRD with D and δ the same for about 70% of the observations.

To see how robust the estimators are to violations of $e \sim Uni[0.1]$, we try three non-uniform distributions: (i) e has an asymmetric density linearly increasing from 0.5 to 1.5, (ii) $e \sim Beta(2,2)$ whose density is symmetric about 0.5 and proportional to $e(1-e)$, going up and down with a peak of 1.5 at 0.5, and (iii) $e \sim Beta(0.5,0.5)$ whose density is symmetric about 0.5 and proportional to $e^{-0.5}(1-e)^{-0.5}$, going down and up. All three distributions are highly non-uniform; particularly $e \sim Beta(0.5,0.5)$ is extremely so, because its density is unbounded at both ends. Note that both Beta distributions have mean 0.5. The three non-uniform distributions are dubbed, respectively, "Inc 0.5-1.5", "QuadPeak", and "DownUp" in the simulation tables below.

In each entry of all tables below, Bias, Sd, and Root mean squared error (Rmse) are shown. In each table, three panels for $c = 0.2$, 0.5 and 0.9 are provided. Also, 'IVE_{nc}' is the

IVE without the cutoff sample, 'IVE_c' is the IVE using the cutoff sample, and 'P(Reject)' is the rejection proportion of the test with ± 1.96 as the critical values.

Table 1 is for $S = 0, \pm 1, \pm 2$ so that $N = 2500$ because each integer point has 500 observations. In the first panel for $c = 0.2$, IVE_c using the cutoff sample has lower Rmse's than IVE_{nc}. Both IVE_{nc} and IVE_c are much biased for the asymmetric distribution 'Inc 0.5-1.5', but almost unbiased for the other distributions. The test has the correct size 0.05, and rejects the asymmetric non-uniform distribution far more easily than the two Beta distributions.

In the second panel for $c = 0.5$, IVE_{nc}=IVE_c as if the cutoff sample is uninformative when $c = 0.5$. This result must be specific to our simulation design, as $\text{IVE}_{nc} \neq \text{IVE}_c$ in our empirical analysis below with $c = 0.5$. The test fails to reject the Beta distributions, which is not necessarily bad because the estimates are almost unbiased for the Beta distributions. In the third panel for $c = 0.9$, the findings are similar to those in the first panel, but the test power is much lower for the asymmetric distribution with $c = 0.9$.

Table 1. Bias, Sd & Rmse and Test for Linear Model ($N = 2500$)

	Uni[0,1]	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.2$		
IVE_{nc}	-0.02 0.56 0.56	-0.17 0.58 0.61	0.00 0.57 0.57	-0.02 0.58 0.58
IVE_c	0.02 0.44 0.44	-0.10 0.57 0.58	0.01 0.40 0.40	0.05 0.50 0.51
P(Reject)	0.05	0.49	0.14	0.13
		$c = 0.5$		
IVE_{nc}	0.00 0.56 0.56	-0.15 0.57 0.59	-0.01 0.56 0.56	0.00 0.57 0.57
IVE_c	0.00 0.56 0.56	-0.15 0.57 0.59	-0.01 0.56 0.56	0.00 0.57 0.57
P(Reject)	0.06	0.16	0.05	0.05
		$c = 0.9$		
IVE_{nc}	0.03 0.58 0.58	-0.12 0.57 0.58	0.03 0.59 0.59	0.02 0.57 0.57
IVE_c	0.00 0.40 0.40	-0.17 0.40 0.44	0.00 0.37 0.37	-0.04 0.45 0.45
P(Reject)	0.05	0.06	0.09	0.14

$S = 0, \pm 1, \pm 2$ (500 obs. each); Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak, $f(e) \propto e(1-e)$; DownUp, $f(e) \propto e^{-0.5}(1-e)^{-0.5}$; IVE_{nc}, IVE not using cutoff sample; IVE_c, IVE using cutoff sample; P(Reject), test rejection proportion.

In Table 2, the sample size doubles as each integer support point gets 1000 observations.

The findings in Table 2 are almost the same as those in Table, except that the Bias's, Sd's and Rmse's are mostly smaller than in Table 1 due to the larger sample size. Also, the rejection proportions are higher for the non-uniform distributions when the estimates are biased so that the uniformity should be rejected.

Table 2. Bias, Sd & Rmse and Test for Linear Model ($N = 5000$)

	Uni[0,1]	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.2$		
IVE$_{nc}$	0.00 0.39 0.39	-0.16 0.40 0.43	0.00 0.38 0.38	-0.01 0.40 0.40
IVE$_{c}$	0.01 0.31 0.31	-0.12 0.39 0.41	0.00 0.28 0.28	0.04 0.35 0.35
P(Reject)	0.05	0.79	0.24	0.25
		$c = 0.5$		
IVE$_{nc}$	0.01 0.38 0.38	-0.15 0.39 0.42	0.00 0.37 0.37	0.00 0.38 0.38
IVE$_{c}$	0.01 0.38 0.38	-0.15 0.39 0.42	0.00 0.37 0.37	0.00 0.38 0.38
P(Reject)	0.05	0.26	0.05	0.05
		$c = 0.9$		
IVE$_{nc}$	0.00 0.40 0.40	-0.13 0.39 0.42	0.02 0.40 0.40	0.01 0.40 0.40
IVE$_{c}$	-0.02 0.28 0.28	-0.17 0.28 0.32	0.00 0.25 0.25	-0.02 0.31 0.31
P(Reject)	0.05	0.06	0.13	0.23

$S = 0, \pm 1, \pm 2$ (1000 obs. each); Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak, $f(e) \propto e(1-e)$; DownUp, $f(e) \propto e^{-0.5}(1-e)^{-0.5}$; IVE$_{nc}$, IVE not using cutoff sample; IVE$_{c}$, IVE using cutoff sample; P(Reject), test rejection proportion.

Table 3 uses 7 integer points $S = 0, \pm 1, \pm 2, \pm 3$, each with 500 observations so that $N = 3500$, and the quadratic model with $\beta_{--} = 0.05$ and $\beta_{++} = 0.2$ is employed; otherwise, the simulation designs are the same as in the above linear model design. There are two more estimators compared: IVE$_{ncq}$ and IVE$_{cq}$, which are the quadratic-model versions of IVE$_{nc}$ and IVE$_{c}$, respectively. Also 'P(Reject)' presents two rejection proportions of the linear-model and quadratic-model tests.

The first panel for $c = 0.2$ shows that IVE$_{nc}$ and IVE$_{c}$ are heavily biased, compared with IVE$_{ncq}$ and IVE$_{cq}$, although their Sd's are much smaller. Both tests have almost correct sizes, but the power is lower in the quadratic model test, which is not necessarily bad because the biases are much lower in IVE$_{ncq}$ and IVE$_{cq}$. In the second panel for $c = 0.5$, we have

IVE$_{nc}$=IVE$_c$ and IVE$_{ncq}$=IVE$_{cq}$ as in the second panel of Table 2. IVE$_{nc}$ and IVE$_c$ are much biased with smaller Sd's as in the first panel. Also, the power is lower for the quadratic model test, and both tests essentially fail to reject the Beta distributions.

Table 3. Bias, Sd & Rmse and Test for Quadratic Model ($N = 3500$)

	Uni[0,1]	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.2$		
IVE$_{nc}$	-1.18 0.23 1.20	-1.25 0.23 1.27	-1.19 0.22 1.21	-1.17 0.23 1.19
IVE$_c$	-0.68 0.18 0.71	-0.86 0.21 0.89	-0.66 0.17 0.68	-0.71 0.20 0.74
IVE$_{ncq}$	-0.04 0.67 0.68	-0.10 0.67 0.68	-0.04 0.65 0.65	-0.03 0.67 0.67
IVE$_{cq}$	0.01 0.40 0.40	-0.03 0.61 0.61	-0.01 0.35 0.35	0.04 0.48 0.49
P(Reject)	0.05, 0.04	1.00, 0.73	0.67, 0.23	0.60, 0.21
		$c = 0.5$		
IVE$_{nc}$	-0.72 0.21 0.75	-0.79 0.21 0.82	-0.72 0.21 0.75	-0.71 0.21 0.74
IVE$_c$	-0.72 0.21 0.75	-0.79 0.21 0.82	-0.72 0.21 0.75	-0.71 0.21 0.74
IVE$_{ncq}$	0.00 0.59 0.59	-0.07 0.60 0.60	0.00 0.58 0.58	0.00 0.60 0.60
IVE$_{cq}$	0.00 0.59 0.59	-0.07 0.60 0.60	0.00 0.58 0.58	0.00 0.60 0.60
P(Reject)	0.06, 0.06	0.54, 0.25	0.07, 0.05	0.06, 0.05
		$c = 0.9$		
IVE$_{nc}$	-0.16 0.21 0.26	-0.23 0.21 0.31	-0.17 0.20 0.26	-0.15 0.21 0.26
IVE$_c$	-0.36 0.17 0.40	-0.44 0.17 0.47	-0.35 0.16 0.39	-0.40 0.19 0.44
IVE$_{ncq}$	0.03 0.69 0.69	-0.04 0.67 0.67	0.04 0.67 0.68	0.04 0.68 0.68
IVE$_{cq}$	-0.01 0.33 0.33	-0.09 0.33 0.34	-0.01 0.29 0.29	-0.02 0.40 0.40
P(Reject)	0.05, 0.05	0.05, 0.05	0.42, 0.12	0.66. 0.21

$S = 0, \pm1, \pm2, \pm3$ (500 obs. each); Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak $f(e) \propto e(1 - e)$; DownUp, $f(e) \propto e^{-0.5}(1 - e)^{-0.5}$; IVE$_{nc}$, IVE not using cutoff sample; IVE$_c$, IVE using cutoff sample; IVE$_{ncq}$ & IVE$_{cq}$, quadratic model versions of IVE$_{nc}$ & IVE$_c$; P(Reject), rejection proportion of test (linear, quad.).

In the third panel for $c = 0.9$, IVE$_c$ is far more biased than IVE$_{nc}$ whereas the Sd of IVE$_c$ is only a little smaller than the Sd of IVE$_{nc}$: using the cutoff sample does not always result in improvements, if the local neighborhood is not local enough so that the model misspecification matters. Differently from IVE$_c$ relative to IVE$_{nc}$, however, IVE$_{cq}$ does much better than IVE$_{ncq}$. Again, the quadratic-model test rejects less than the linear-model test, which is not necessarily bad because the quadratic model estimators are little biased under the Beta distributions. We omit the simulation results for the quadratic model with each integer point having 1000 observations, because what can be learned from this relative to Table 3 is similar to what was learned from Table 2 relative to Table 1.

5 Empirical Analysis

The National Health Insurance Service (NHIS) in South Korea expanded its dental support program coverage for denture and dental implant on July 1, 2015, to the elderly of age 70 or above whereas the existing age cutoff used to be 75. NHIS expected that about 100,000 to 120,000 elders would benefit by paying about \$1000 less per year for the covered dental treatments that would cost about \$1600 on average if not for the program. We use the 2016 wave from 'the Korea Longitudinal Study of Aging' for the 2015 information with $N = 7089$ to assess the effect of the dental support program extension D on the 2015 dental expenditure Y in $10,000$ Korean Won (a little less than \$10). This is a SRD, as there is no exception for the qualification condition based on age.

Table 4. Descriptive Statistics of Variables ($N = 7089$)

Variable	Mean (Sd)	Min, Max	Variable	Mean (Sd)
dental expenditure Y	24.0 (118)	0, 2200	smoking	0.11 (0.31)
age in years S	66.8 (10.3)	52, 106	drinking	0.34 (0.48)
male	0.43 (0.49)		elementary school	0.38 (0.49)
married	0.77 (0.42)		middle school	0.16 (0.37)
household income	3101 (2692)	0, 35000	high school	0.34 (0.47)
			university	0.11 (0.32)

Y, income in $\simeq$ \$10; Min, Max for dummy omitted; education dummy for completion

Table 4 provides descriptive statistics for Y and S, along with those for covariates on gender, marital status, household income, smoking, drinking, and education. For each value of S, there are about 200 persons; 182 persons in our cutoff sample with $S = 0$. We provide effect estimates with the covariates uncontrolled first and controlled later, but as it turns out, controlling the covariates makes little difference.

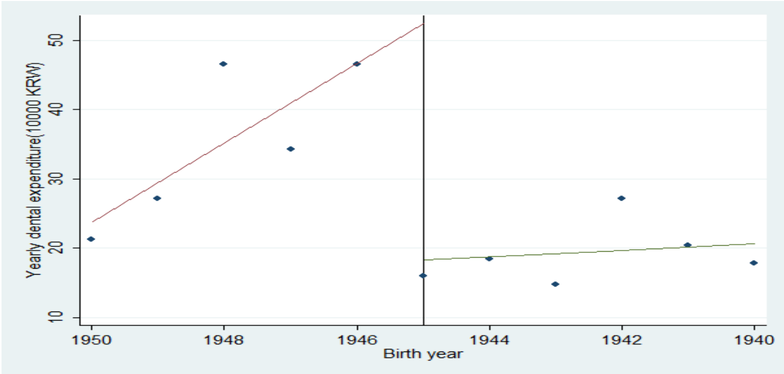

Figure 1: 2015 Mean Dental Expenditure for Each Birth Year Cohort

The cutoff c varies across individuals in the cutoff sample, because persons reaching age 70 on a date after July 1, 2015, become treated on the date and onwards. However, since Y is the dental expenditure in 2015, those cutoff-sample individuals who became eligible later than July 1, 2015, are only partly treated, compared with the fully treated persons who became eligible on July 1, 2015. This case differs from the usual RD where individuals are fully treated with different cutoffs. That is, our RD case with an individually varying cutoff is unusual, because individuals in the cutoff sample are partially treated to different degrees.

Another unusual aspect in our RD case is that dentures and implants can wait for several months, which means that those who are partly treated may be almost as good as fully treated if they wait until they become eligible within 2015. Indeed, Figure 1 plotting $E(Y|S)$ for birth years $S = 1940 \sim 1950$ reveals this feature. Figure 1 presents two linearly fitted lines to the right and left of the cutoff sample dot for birth year 1945, where a clear break of magnitude greater than 30 is seen at birth year 1945; the 1945 dot was not used for the linear line estimation. Although we expected the 1945 dot to be somewhere vertically in the middle, it actually is as low as the dots for earlier birth years 1944-1940. This also raises a possibility that despite $c = 0.5$, the "de facto" c could have been earlier, as individuals could have waited until their eligibility date for their dental problems that occurred much earlier.

Table 5 presents the RD estimation results without controlling any covariates. Analogously to 'IVE$_{nc}$' and 'IVE$_c$' in the simulation tables of the preceding section, 'OLS$_{nc}$' is the linear-model OLS without the cutoff sample, whereas 'OLS$_c$' is the linear model OLS with the cutoff sample; β_- and β_+ are the slope estimates from OLS$_c$. 'OLS$_{c0}$' is the OLS$_c$ under the assumption $c = 0$ as if the treatment had started on January 1, 2015. 'ts' is the test statistic value for the linear model. 'OLS_{ncq}', 'OLS_{cq}', 'OLS_{cq0}' and 'ts_q' are analogously defined for the quadratic model; β_-, β_+, β_{--} and β_{++} are the slope estimates from OLS_{cq}.

Table 5. Effect Estimate (t-value) & Test Statistic (Covariate Uncontrolled)

h (N_h)	$h = 2$ (959)	$h = 3$ (1386)	$h = 4$ (1812)	$h = 5$ (2217)
OLS_{nc}	-36.7 (-1.2)	-31.6 (-1.4)	-34.6 (-2.2)	-34.2 (-2.6)
OLS_c	-39.3 (-1.3)	-32.4 (-1.5)	-35.4 (-2.3)	-35.0 (-2.6)
β_-	3.19 (0.28)	-2.35 (-0.29)	2.55 (0.65)	4.26 (1.6)
β_+	6.76 (0.86)	7.20 (1.5)	4.13 (1.6)	2.28 (1.2)
OLS_{c0}	-35.4 (-1.8)	-29.6 (-1.9)	-33.0 (-2.7)	-33.0 (-3.0)
ts	-1.65	-1.06	-1.99	-2.50
OLS_{ncq}		-44.8 (-0.83)	-30.5 (-0.79)	-33.8 (-1.2)
OLS_{cq}		-49.7 (-0.90)	-31.6 (-0.80)	-34.9 (-1.2)
β_-		8.78 (0.24)	-12.5 (-0.56)	-7.80 (-0.52)
β_+		17.26 (0.60)	15.5 (0.98)	14.3 (1.4)
β_{--}		2.80 (0.29)	-3.29 (-0.77)	-2.17 (-0.90)
β_{++}		-2.48 (-0.35)	-2.56 (-0.85)	-2.20 (-1.3)
OLS_{cq0}		-40.9 (-1.6)	-29.0 (-1.3)	-31.5 (-1.7)
ts_q		-1.69	-0.69	-0.89

N_h, sample size with h; OLS_{nc}, linear-model OLS without cutoff sample; OLS_c, linear-model OLS with cutoff sample; β_- & β_+ from OLS_c; OLS_{c0}, linear-model OLS with $c = 0$; ts, test statistic; OLS_{ncq}, OLS_{cq}, OLS $_{qc0}$, ts_q for quadratic model; β_-, β_+, β_{--} , β_{++} from OLS_{cq}

The effect estimates from the linear model are insignificant for $h = 2$ and 3, but become significant for $h = 4$ and 5; the effect magnitude stays fairly stable around $-32 \sim -39$ as Figure 1 indicates. The slope estimates for β_- and β_+ look insignificant for all h values. Also, OLS_{c0} with $c = 0$ are similar to OLS_{nc} and OLS_c. Somewhat disappointingly, OLS_c hardly differs from OLS_{nc} with rather small efficiency gains, which might be due to $c = 0.5$.

The test is not-rejecting with $h = 2$ and 3, but rejecting with $h = 4$ and 5. Even when non-rejecting, ts is negative with a fairly large magnitude, which seems to be due to the aforementioned reason that dental procedures can wait several months. That is, the cutoff sample elders seem to have waited until they became eligible for the dental support program, which is equivalent to having c lower than its nominal value 0.5.

For the quadratic model, the effect estimates are insignificant for all h values under consideration; judging from the β_-, β_+, β_{--} and β_{++} estimates, it seems that the quadratic model is over-specified. The test statistic values are all insignificantly negative for the quadratic model.

Table 6 presents the effect estimates (t-values) with the eight covariates in Table 4 controlled. Despite controlling the covariates, the results are remarkably similar to those in Table 5, which demonstrates the robustness of our findings in Table 5.

Table 6. Effect Estimate (t-value) & Test Statistic (Covariates Controlled)

h (N_h)	$h = 2$ (959)	$h = 3$ (1386)	$h = 4$ (1812)	$h = 5$ (2217)
OLS$_{nc}$	-40.2 (-1.3)	-33.1 (-1.5)	-35.7 (-2.3)	-35.3 (-2.6)
OLS$_c$	-41.3 (-1.3)	-33.4 (-1.5)	-36.5 (-2.3)	-36.2 (-2.7)
OLS$_{c0}$	-35.8 (-1.8)	-30.5 (-1.9)	-33.9 (-2.7)	-34.1 (-3.1)
OLS$_{ncq}$		-43.1 (-0.82)	-30.5 (-0.78)	-33.5 (-1.1)
OLS$_{cq}$		-47.1 (-0.86)	-31.1 (-0.78)	-34.5 (-1.1)
OLS$_{cq0}$		-40.5 (-1.5)	-29.3 (-1.3)	-32.0 (-1.8)

N_h, sample size with h; OLS$_{nc}$, linear-model OLS without cutoff sample; OLS$_c$, linear-model OLS with cutoff sample; OLS$_{c0}$, linear-model OLS with $c = 0$; ts, test statistic; OLS$_{ncq}$, OLS$_{cq}$ and OLS$_{cq0}$ are for quadratic model.

6 Conclusions

Often in regression discontinuity (RD) with a cutoff c, the running variable G is observed only as its rounded-down integer S. If one proceeds as usual ignoring the integer nature of S, then there occurs a bias in general. In the RD literature, indirect estimators for the treatment effect have been proposed under the assumption that the measurement error $e \equiv G - S$ follows a known distribution, which may be found with an auxiliary sample or a census. When c is not an integer, the existing indirect estimators do not use the "cutoff sample" (the sample with S equalling the integer version of c), because their treatment status is unclear.

In this paper, we made a number of contributions. First, we explained why the uniform distribution "$Uni[0, 1]$" on $[0, 1]$ for e is likely to hold in reality: $e \equiv G - S$ is close to a popular way pseudo uniform random numbers are generated in simulation studies. Second, we generalized the existing RD identification findings with integer c to those with any c to show that the value of c may play an important role; e.g., the RD bias due to observing S instead of G may disappear if $c = 0.5$ as in our empirical analysis. Third, when c is not an integer, we showed how to make use of the cutoff sample in estimation and in test for the $Uni[0,1]$ assumption.

We conducted a simulation study to demonstrate that our proposed estimators and tests work as they are supposed to, and that the estimators are highly robust to deviations from the $Uni[0,1]$ assumption on e. We also provided an empirical analysis, where the $Uni[0,1]$ assumption is rejected for large bandwidth values, but not for small bandwidth values. Using the cutoff sample in estimation made only a minute difference in our empirical analysis, which seems to be due to $c = 0.5$. Despite this, however, there is no reason not to make use of the cutoff sample, as the requisite extra work is almost zero while there are things to gain: testability of $Uni[0,1]$, and the potential efficiency gain which could be substantial in other applications.

APPENDIX

Identification with Any Cutoff for Linear/Quadratic $m(\cdot)$

Take $E(\cdot|S)$ on the linear model (1.3) with $D = \delta$ for SRD to obtain

$$\begin{aligned} E(Y|S) &= \beta_0 + \beta_d\delta + \beta_1(S + 0.5 - c) + \beta_{1\delta}\delta(S + 0.5 - c) \\ &= \beta_0 + (0.5 - c)\beta_1 \ + \{\beta_d + (0.5 - c)\beta_{1\delta}\}\delta \ + \beta_1 S \ + \beta_{1\delta}\delta S \\ \Longrightarrow \quad & \beta_d = \{\delta \text{ slope in } E(Y|S)\} - (0.5 - c)\{\delta S \text{ slope in } E(Y|S)\}. \end{aligned}$$

Generalizing this for FRD gives (3.1). As for the quadratic (1.4) with $D = \delta$, with $\mu_2 = 1/3$,

$$\begin{aligned} & E\{(G - c)^2|S\} = E\{(S + e)^2 - 2c(S + e) + c^2|S\} = S^2 + 2S\mu_1 + \mu_2 - 2cS - 2c\mu_1 + c^2 \\ = \quad & S^2 + (2\mu_1 - 2c)S + \frac{1}{3} - 2c\mu_1 + c^2 = S^2 + (1 - 2c)S + \frac{1}{3} - c + c^2. \end{aligned} \tag{A.1}$$

Using this renders

$$\begin{aligned} & E(Y|S) = \beta_0 + \beta_d\delta + \beta_1(S + 0.5 - c) + \beta_{1\delta}\delta(S + 0.5 - c) \\ & +\beta_2\{S^2 + (1 - 2c)S + \frac{1}{3} - c + c^2\} + \beta_{2\delta}\delta\{S^2 + (1 - 2c)S + \frac{1}{3} - c + c^2\} \\ = \quad & \beta_0 + \beta_1(0.5 - c) + \beta_2(\frac{1}{3} - c + c^2) \ + \{\beta_d + \beta_{1\delta}(0.5 - c) + \beta_{2\delta}(\frac{1}{3} - c + c^2)\}\delta \\ & +\{\beta_1 + \beta_2(1 - 2c)\}S \ + \ \{\beta_{1\delta} + \beta_{2\delta}(1 - 2c)\}\delta S \ + \beta_2 S^2 + \beta_{2\delta}\delta S^2. \end{aligned}$$

Hence β_d in SRD equals

$$
\begin{aligned}
&(\text{slope of } \delta) - (\text{slope of } \delta S)(\frac{1}{2} - c) + (\text{slope of } \delta S^2)\{(1-2c)(\frac{1}{2} - c) - (\frac{1}{3} - c + c^2)\} \\
&= (\text{slope of } \delta) - (\text{slope of } \delta S)(\frac{1}{2} - c) + (\text{slope of } \delta S^2)(\frac{1}{2} - 2c + 2c^2 - \frac{1}{3} + c - c^2) \\
&= (\text{slope of } \delta) - (\text{slope of } \delta S)(\frac{1}{2} - c) + (\text{slope of } \delta S^2)(\frac{1}{6} - c + c^2). \qquad \text{(A.2)}
\end{aligned}
$$

Generalizing this for FRD gives (3.2).

Proof for $E(Y|S=0)$ in SRD with Linear $m(G)$

For (1.3) with $D = \delta$, we have

$$
\begin{aligned}
E(Y|\delta &= 0, S=0) = \beta_0 + \beta_-\{E(G|\delta=0, S=0) - c\}; \qquad \text{(A.3)} \\
E(Y|\delta &= 1, S=0) = \beta_0 + \beta_d + \beta_+\{E(G|\delta=1, S=0) - c\}.
\end{aligned}
$$

Because e is uniform on any subinterval of $(0,1)$, we have

$$
\begin{aligned}
E(G|\delta &= 0, S=0) = E\{e|e \in (0,c)\} = 0.5c, \\
E(G|\delta &= 1, S=0) = E\{e|e \in (c,1)\} = 0.5(c+1). \qquad \text{(A.4)}
\end{aligned}
$$

Substituting these into (A.3) renders

$$
\begin{aligned}
E(Y|\delta &= 0, S=0) = \beta_0 + \beta_-(0.5c - c) = \beta_0 + \beta_-(-0.5c), \qquad \text{(A.5)} \\
E(Y|\delta &= 1, S=0) = \beta_0 + \beta_d + \beta_+(0.5 - 0.5c).
\end{aligned}
$$

Because $P(\delta=0|S=0) = P(e \in [0,c)) = c$ as $e \sim Uni[0,1] \amalg S$,

$$
\begin{aligned}
&E(Y|S=0) = E(Y|\delta=0,S=0)P(\delta=0|S=0) + E(Y|\delta=1,S=0)P(\delta=1|S=0) \\
= \;& \{\beta_0 + \beta_-(-0.5c)\}c + \{\beta_0 + \beta_d + \beta_+(0.5 - 0.5c)\}(1-c) \\
= \;& \beta_0 \; + \beta_d(1-c) \; + \beta_-(-0.5c^2) \; + \beta_+ 0.5(1-c)^2. \qquad \text{(A.6)}
\end{aligned}
$$

Proofs for Quadratic $m(\cdot)$

For the SRD quadratic model, we get, recalling (A.1),

$$
\begin{aligned}
S \le -1 : E(Y|S=s) &= \beta_0 + \beta_-\{E(G|S) - c\} + \beta_{--}E\{(G-c)^2|S\} \\
&= \beta_0 + \beta_- S_{0.5c} + \beta_{--}S^*; \qquad \text{(A.7)} \\
S \ge 1 : E(Y|S=s) &= \beta_0 + \beta_d + \beta_+ S_{0.5c} + \beta_{++}S^*.
\end{aligned}
$$

Combining these two cases gives, for all S values other than 0,

$$E(Y|S) = \beta_0 + \beta_d\delta_+ + \beta_-\delta_- S_{0.5c} + \beta_+\delta_+ S_{0.5c} + \beta_{--}\delta_- S^* + \beta_{++}\delta_+ S^*. \tag{A.8}$$

Observe

$$\begin{aligned} E(e^2|\delta &= 0) = E(e^2|0 < e < c) = \int_0^c e^2 de/c = \frac{1}{3c}e^3|_0^c = \frac{c^2}{3}; \\ E(e^2|\delta &= 1) = E(e^2|c \le e < 1) = \int_c^1 e^2 de/(1-c) = \frac{1}{3(1-c)}e^3|_c^1 = \frac{1+c+c^2}{3}. \end{aligned} \tag{A.9}$$

This gives, as $G - c = e - c$ on $S = 0$,

$$\begin{aligned} &E\{(G-c)^2|\delta = 0, S = 0)\} = E(e^2 - 2ec + c^2|\delta = 0) \\ &= E(e^2|\delta = 0) - 2cE(e|\delta = 0) + c^2 = \frac{c^2}{3} - 2c \times \frac{c}{2} + c^2 = \frac{c^2}{3}; \\ &E\{(G-c)^2|\delta = 1, S = 0)\} = E(e^2|\delta = 1) - 2cE(e|\delta = 1) + c^2 \\ &= \frac{1+c+c^2}{3} - 2c \times \frac{1+c}{2} + c^2 = \frac{1+c+c^2-3c-3c^2+3c^2}{3} = \frac{(1-c)^2}{3}. \end{aligned}$$

Using these, we have

$$\begin{aligned} &E(Y|\delta = 0, S = 0) = \beta_0 + \beta_-\{E(G|\delta = 0, S = 0) - c\} \\ &+\beta_{--}E\{(G-c)^2|\delta = 0, S = 0)\} = \beta_0 + \beta_-(0.5c - c) + \beta_{--}\frac{c^2}{3}; \\ &E(Y|\delta = 1, S = 0) = \beta_0 + \beta_d + \beta_+\{0.5(c+1) - c\} + \beta_{++}\frac{(1-c)^2}{3}. \end{aligned}$$

Hence, analogously to (A.6), $E(Y|S = 0)$ becomes

$$\begin{aligned} &\{\beta_0 + \beta_-(-0.5c) + \beta_{--}\frac{c^2}{3}\}c + \{\beta_0 + \beta_d + \beta_+(0.5 - 0.5c) + \beta_{++}\frac{(1-c)^2}{3}\}(1-c) \\ = \; &\beta_0 + \beta_d(1-c) + \beta_-(-0.5c^2) + \beta_+ 0.5(1-c)^2 + \beta_{--}\frac{c^3}{3} + \beta_{++}\frac{(1-c)^3}{3}. \end{aligned} \tag{A.10}$$

This gives the uniformity test in the main text.

Combine (A.8) and (A.10) to obtain

$$\begin{aligned} &E(Y|S) = \beta_0 + \beta_d\{(1-c)\delta_0 + \delta_+\} + \beta_-\{-0.5c^2\delta_0 + \delta_- S_{0.5c}\} \\ &+\beta_+\{0.5(1-c)^2\delta_0 + \delta_+ S_{0.5c}\} + \beta_{--}(\frac{c^3}{3}\delta_0 + \delta_- S^*) + \beta_{++}\{\frac{(1-c)^3}{3}\delta_0 + \delta_+ S^*\}]. \end{aligned}$$

This gives the OLS and IVE with quadratic $m(\cdot)$ using the cutoff sample.

REFERENCES

Cattaneo, M.D. and J.C. Escanciano, 2017, Introduction, in Advances in Econometrics 38 (entitled ‘Regression discontinuity designs: theory and applications), edited by M.D. Cattaneo and J.C. Escanciano, Emerald Publishing.

Cattaneo, M.D., N. Idrobo and R. Titiunik, 2019, A practical introduction to regression discontinuity designs, Cambridge University Press, forthcoming.

Choi, J.Y. and M.J. Lee, 2017, Regression discontinuity: review with extensions, Statistical Papers 58, 1217-1246.

Choi, J.Y. and M.J. Lee, 2018, Relaxing conditions for local average treatment effect in fuzzy regression discontinuity, Economics Letters 173, 47-50.

Dong, Y., 2015, Regression discontinuity applications with rounding errors in the running variable, Journal of Applied Econometrics 30, 422-446.

Gelman, A. and G. Imbens, 2019, Why high-order polynomials should not be used in regression discontinuity designs, Journal of Business and Economic Statistics, forthcoming.

Hahn, J., P. Todd and W. Van der Klaauw, 2001, Identification and estimation of treatment effects with a regression-discontinuity design, Econometrica 69, 201-209.

Imbens, G.W. and T. Lemieux, 2008, Regression discontinuity designs: a guide to practice, Journal of Econometrics 142, 615-635.

Kolesár, M. and C. Rothe, 2018, Inference in regression discontinuity designs with a discrete running variable, American Economic Review 108, 2277-2304.

Lee, D.S. and D. Card, 2008, Regression discontinuity inference with specification error, Journal of Econometrics 142, 655-674.

Lee, D.S. and T. Lemieux, 2010, Regression discontinuity designs in economics, Journal of Economic Literature 48, 281-355.

Lee, M.J., 2010, Micro-econometrics: methods of moments and limited dependent variables, Springer.

Lee, M.J., 2016, Matching, regression discontinuity, difference in differences, and beyond, Oxford University Press.

Toral, R. and C. Colet, 2014, Stochastic numerical methods: an Introduction for students and scientists, Wiley.