The Korean Economic Review
Volume 39, Number 1, Winter 2023, 73-101.
DOI : 10.22841/kerdoi.2023.39.1.003

Regression Discontinuity with Integer Score and Non-Integer Cutoff*

Myoung-Jae Lee** · Hyae-Chong Shim*** · Sang Soo Park****

In regression discontinuity (RD), the treatment is determined by a continuous score G crossing a cutoff c or not. However, often G is observed only as the 'rounded-down integer S' (e.g., birth year observed instead of birth time), and c is not an integer. In this case, the "cutoff sample" (i.e., the observations with S equal to the rounded-down integer of c) is discarded due to the ambiguity in G crossing c or not. We show that, first, if the usual RD estimators are used with the integer nature of S ignored, then a bias occurs, but it becomes zero if a slope symmetry condition holds or if c takes a certain "middle" value. Second, the distribution of the measurement error $e = G - S$ *can be specified and tested for, and if the distribution is accepted, then the cutoff sample can be used fruitfully. Third, two-step estimators and bootstrap inference are available in the literature, but a single-step ordinary least squares or instrumental variable estimator is enough. We also provide a simulation study and an empirical analysis for a dental support program based on age in South Korea.*

JEL Classification: C21, C24
Keywords: Regression Discontinuity, Integer Running Variable, Non-integer Cutoff

I. 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, 2021), Cattaneo and Escanciano (2017), Cattaneo et al. (2019), and references therein. Typically, there are a binary treatment D, an outcome Y and

Received: Nov. 11, 2020. Revised: July 6, 2021. Accepted: Sept. 17, 2021.
* Myoung-jae Lee's research has been supported by a Korea University Grant (K2122221) and Sang Soo Park's by another Korea University Grand (K1809281)..
** Corresponding Author, Professor, Department of Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea, Phone: +82-2-3290-2229, E-mail: myoungjae@korea.ac.kr
*** Co-Author, Senior Researcher, Korea International Trade Association, 511 Yeongdong-daero, Gangnam-gu, Seoul 06164, Korea, E-mail: assashim@gmail.com
**** Third Author, Professor, Department of Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea, E-mail: starpac@korea.ac.kr

a continuous score (or running variable) G, where D is determined by G crossing a cutoff c or not. If D is fully determined by G so that $E(D|G) = D$, the RD is a 'sharp RD (SRD)', otherwise, if D is partly determined by an error term, the RD is a 'fuzzy RD (FRD)'.

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

Suppose the rounded-down integer $S \equiv \lfloor G \rfloor$ is observed, instead of the genuine score G. There are many examples for this (Oreopoulos, 2006; Battistin et al., 2009; Li et al., 2015, among others), and the most notable is birth year (or age in years) observed for confidentiality instead of the exact birth time. In this case, if c is not an integer (e.g., January 1), it is unclear whether individuals around $G = c$ are treated or not. For instance, an education law D goes into effect on September 1, 2016, and a student is subject to the law if born on or after September 1, 2010: $D = 1[c \le G]$ with birth time G and $c = 2010.67$, where $1[A] \equiv 1$ if A holds and 0 otherwise. It is not clear whether the "cutoff sample" (i.e., those with $S = 2010$) are treated or not.

A couple of studies addressed RD with an integer score. Lee and Card (2008) examined genuinely integer/discrete score (i.e., $G = S$; e.g., the number of students in a school). They recommended clustered inference with clustering at each support point of S, but their assumptions are untenable (Kolesár and Rothe, 2018). Dong (2015) proposed two-step estimators, assuming that the 'measurement error'

$$e = G - \lfloor G \rfloor = G - S \text{ is independent of } S \quad (`e \amalg S') \tag{1}$$

and follows a known distribution. Imbens and Wager (2019) proposed an optimal estimation and inference for SRD, which allows a discrete or continuous score. Bartalotti et al. (2021) generalized Dong (2015) to group-varying measurement errors. Broadly viewed, the problem of observing $S = \lfloor G \rfloor$ instead of G is a special case of mismeasured scores (Lee, 2017; Davesize and Le Barbanchon, 2017).

Dong (2015) addressed integer c under a known distribution of e, and suggested dropping the cutoff sample if c is not an integer. This amounts to ruling out non-integer c, because the problem of non-integer c is relevant only to the cutoff sample. Thus we write henceforth that the Dong's approach is for integer c. In this paper, under a known distribution for e with $e \amalg S$, we address RD with integer S and any c, to show that the cutoff sample can be used fruitfully in testing for the distribution assumption on e and estimating the treatment effect if the assumption is not rejected.

Any distribution on $[0,1]$ can be adopted, but the most basic is the uniform distribution ('$U[0,1]$') adopted in Dong (2015) as well as in this paper. '$e \sim U[0,1]$' is plausible particularly when G is birth time (Dong, 2015, p. 428). Note that

$$e \sim U[0,1]: \mu_k \equiv E(e^k) = \int_0^1 \tau^k d\tau = (k+1)^{-1};$$

we need mostly only $E(e)$ and $E(e^2)$, not all of $U[0,1]$. Since our approach with integer c becomes that in Dong (2015), assume a non-integer c, unless otherwise mentioned.

We make three contributions for RD with integer score S and non-integer c. First, the usual RD estimators ignoring the integer nature of S are inconsistent, but under $e \sim U[0,1]$ and $e \amalg S$, they are consistent if a slope symmetry around c holds or if c takes a "middle" value. Second, the cutoff sample can be used to test for $e \sim U[0,1]$ and to estimate the treatment effect. Third, two-step estimators with bootstrap inference have been proposed for RD with S by Dong (2015), but single-step ordinary least squares estimator (OLS) or instrumental variable estimator (IVE) is enough.

The requisite assumptions for our proposals are the same as the assumptions 1-6 in Dong (2015, p. 427), but there are differences between our and Dong's (2015) approaches. First, whereas Dong allowed a J-order polynomial to account for a con-tinuous function of the score in Y, we focus on $J=1$ (linear) and $J=2$ (quadratic), following Gelman and Imbens (2019). Second, Dong considered integer c, and dropped the cutoff sample when c is not an integer; as a result, her no-bias condition when the integer nature of S is ignored is only the slope symmetry around c, missing the condition under non-integer c. Third, when c is not an integer, we use the cutoff sample to test for the distribution/moment assumption on e, whereas Dong simply drops the cutoff sample as was already noted. Fourth, when the assumption is not rejected by the test, we use the cutoff sample along with the non-cutoff sample for a higher efficiency.

In RD with G, often the location normalization $G-c$ is done so that the cutoff becomes 0 for the normalized score $G-c$. With $S=\lfloor G \rfloor$, the analogous normalization is, not a non-integer number $S-c$, but an integer

$$S_0 \equiv S - \lfloor c \rfloor \quad (S_0 = 0 \text{ for cutoff sample}) \text{ with } c_0 \equiv c - \lfloor c \rfloor \quad (0 \le c_0 < 1). \quad (2)$$

In the education law example where $c = 2010.67$, we have $\lfloor c \rfloor = 2010$, $S_0 \equiv S - \lfloor c \rfloor = S - 2010$, and $c_0 \equiv c - \lfloor c \rfloor = 0.67$ showing where the cutoff falls within the "cutoff year" 2010.

Before proceeding further, to prevent confusion, we note the difference between the 'classical measurement error' and 'measurement error in RD with integer score'. In the former, a mismeasured continuous S is generated by $S = G + e$ with $e \coprod G$: a continuous error e is added to the genuine continuous score G. In the latter, G is generated by $S + e$ with $e \coprod S$: a continuous error e is added to the integer S ; '$S = G + e$' does not work, as $G + e$ is not an integer. For example, continuous birth time 2010.5 is generated by the integer 2010 plus 0.5; 'e within the year S' is independent of S.

In the remainder of this paper, Sections 2-4 present the aforementioned three con-tributions one by one, with most proofs relegated to the appendix. Section 5 conducts a simulation study, and Section 6 presents an empirical example for a dental support program in South Korea where G is age. Finally, Section 7 concludes. What is observed is (S_i, Y_i) for SRD and (D_i, S_i, Y_i) for FRD, $i = 1, \ldots, N$, which are independent and identically distributed; we often omit the subscript i, as has been done already. In most cases, we address FRD first, and then SRD later as a special/limiting case of FRD.

II. No Bias Condition for Ignored Integer Score

2.1. Local Approaches with Continuous Score

Suppose G is observed for a while to see what is usually done in RD. Define

$$\delta \equiv 1[c \le G] = 1[0 \le G - c];$$

$\delta(\neq D)$ is to be used as an instrument variable (IV) for D in FRD, whereas $\delta = D$ in SRD. Let (D^0, D^1) be the 'potential treatments' corresponding $\delta = 0$, $1 - D$ is an outcome variable when δ is taken as the "deep treatment"—and let (Y^0, Y^1) be the potential outcomes corresponding to $D = 0,1$. Define 'compliers' as those with $(D^0 = 0, D^1 = 1)$—'compliers' because they get treated complying to the 'assignment' δ—and assume $D^0 \le D^1$ to rule out 'defiers' with $(D^0 = 1, D^1 = 0)$.

Define the 'treatment effect on the just-treated compliers' β_d for FRD:

$$\beta_d \equiv E(Y^1 + Y^0 \mid \text{complier}, \ G = c^+) \equiv \lim_{g \downarrow c} E(Y^1 - Y^0 \mid \text{complier}, \ G = g);$$

in SRD, everybody is a complier due to $\delta = D$. For FRD, the most popular approach to find β_d is using a local linear model: for some β parameters and error U_1,

$$(i):\ \ Y = \beta_0 + \beta_d D + \beta_1 (G-c) + \beta_{1\delta}\delta(G-c) + U_1,\ \ E(U_1 \mid G) = 0,$$
$$(ii):\ \ E(Y \mid G) = \beta_0 + \beta_d E(D \mid G) + \beta_1 (G-c) + \beta_{1\delta}\delta(G-c)$$
$$\text{(taking } E(\cdot \mid G) \text{ on } (i)). \tag{3}$$

Here, β_d can be estimated by the IVE of Y on

$$\{1, D, (G-c), \delta(G-c)\}' \ \text{ with IV } \ Z_G \equiv \{1, \delta, (G-c), \delta(G-c)\}'.$$

Locally around $G = c$, this is not a parametric model, because the local linear function $\beta_0 + \beta_1(G-c) + \beta_{1\delta}\delta(G-c)$ in (3) with the left and right slopes β_1 and $\beta_1 + \beta_{1\delta}$ is simply to improve the finite sample behavior of the "local-constant" IVE of Y on $(1, D)$ with $(1, \delta)$ as the IV, where the local linear function is simplified into the constant β_0. See Hahn et al. (2001), Dong (2018), Choi and Lee (2018) and references therein for identification (and estimation) of β_d in FRD.

The local linear model has become almost the "industry standard", but a local quadratic version is sometimes used: for an error U_2,

$$Y = \beta_0 + \beta_d D + \beta_1(G-c) + \beta_{1\delta}\delta(G-c) + \beta_2(G-c)^2 + \beta_{2\delta}\delta(G-c)^2 + U_2 \tag{4}$$

where two additional terms $\beta_2(G-c)^2$ and $\beta_{2\delta}\delta(G-c)^2$ appear. For equation (4), β_d can be estimated by the IVE of Y on

$$\{1, D, G-c, \delta(G-c), (G-c)^2, \delta(G-c)^2\} \ \text{ with IV}$$
$$\{1, \delta, G-c, \delta(G-c), (G-c)^2, \delta(G-c)^2\}.$$

If desired, we can go further than quadratic for more flexibility, but Gelman and Imbens (2019) advised against the higher-order terms, because higher-order terms often make both estimation and inference unstable, and also because they lead to extreme weights in the resulting causal effect as a weighted average of $Y^1 - Y^0$. For this reason, we also restrict our discussion to local linear or quadratic models as was already noted.

Consider two locally linear models for $E(D \mid G)$ and $E(Y \mid G)$ analogous to (3)(ii), but with δ replacing $E(D \mid G)$ on the right side: for some α and γ parameters,

$$(i):\ \ E(D \mid G) = \alpha_0 + \alpha_\delta \delta + \alpha_1(G-c) + \alpha_{1\delta}\delta(G-c),$$
$$(ii):\ \ E(Y \mid G) = \gamma_0 + \gamma_\delta \delta + \gamma_1(G-c) + \gamma_{1\delta}\delta(G-c). \tag{5}$$

(5)(i) would be the model for the first-stage OLS for D in two-stage OLS when

D is an endogenous regressor for Y, and (5)(ii) is the reduced form for Y obtained by substituting (5)(i) into the $E(D|G)$ in (3)(ii) and defining properly the γ's as functions of the α's and β's.

Suppose we do the OLS of D and Y on $Z_G \equiv \{1, \delta, (G-c), \delta(G-c)\}'$ for (5) to get the slope estimators $\hat{\alpha}_\delta$ and $\hat{\gamma}_\delta$ of δ, respectively. Then, denoting the slope estimator of D in the IVE for (3)(i) as $\hat{\beta}_d$, '$\hat{\beta}_d = \hat{\gamma}_\delta / \hat{\alpha}_\delta$' holds. We refer to this as 'IVE=Wald' because $\hat{\gamma}_\delta / \hat{\alpha}_\delta$ is often called 'the Wald estimator'. More generally, given any regressor vector Z, it is well-known that 'IVE=Wald' holds, whenever the OLS of D and Y on (δ, Z) is done for $\hat{\gamma}_\delta / \hat{\alpha}_\delta$ and the IVE of Y on (D, Z) is done for $\hat{\beta}_d$ with the IV (δ, Z). As a special case, 'IVE=Wald' also holds for the local quadratic approach in (4) with $Z = \{Z_G', (G-c)^2, \delta(G-c)^2\}'$.

2.2. Approaches with Integer Score

Turning back to S observed instead of G, one might try to use equation (3) with G replaced by S, or $G-c$ replaced by $S_0 \equiv S - \lfloor c \rfloor$; equivalently, one might try to do the same for the Wald estimator based on (5). However, this does not work, because $E(\cdot|S)$ should be derived from $E(\cdot|G)$ as follows.

Recall $G = S + e$. Using $e \sim U[0,1] \amalg S$, we have

$$E(G|S) = E(S+e|S) = S + E(e|S) = S + E(e) = S + 0.5.$$

'$E(e) = 0.5$' reveals that only $E(e)$ is needed, not the full distribution of e, and that a distribution assumption other than $e \sim U[0,1]$ can be used as long as $E(e)$ is known.

Since $0 \le c_0 < 1$ and $e \sim U[0,1]$, (2) gives

$$\delta \equiv 1[c \le G] = 1[0 \le S + e - c] = 1[-e \le S - \lfloor c \rfloor - (c - \lfloor c \rfloor)] = 1[-e \le S_0 - c_0] = 0$$
$$\text{for } S_0 \le -1, \text{ unclear for } S_0 = 0, \text{ 1 for } S_0 \ge 1.$$

Hence, as long as $S_0 \ne 0$, we have $\delta \equiv 1[c \le G] = 0$ if $S_0 \le -1$, and 1 if $S_0 \ge 1$; i.e., as long as $S_0 \ne 0$, δ is fully determined by S. Observe

$$S + 0.5 - c = S - \lfloor c \rfloor + 0.5 - (c - \lfloor c \rfloor) = S_0 + 0.5 + c_0.$$

For $S \ne \lfloor c \rfloor \Leftrightarrow S_0 \ne 0$, taking $E(\cdot|S)$ on $E(D|G)$ in (5)(i) and substituting $E(G|S) = S + 0.5$ gives

$$(i): \ E(D|S) = \alpha_0 + \alpha_\delta \delta + \alpha_1 (S_0 + 0.5 - c_0) + \alpha_{1\delta} \delta (S_0 + 0.5 - c_0)$$

$$(ii):\ = \alpha_0 + \alpha_1(0.5-c_0) + \{\alpha_\delta + \alpha_{1\delta}(0.5-c_0)\}\delta + \alpha_1 S_0 + \alpha_{1\delta}\delta S_0 . \tag{6}$$

(6)(ii) shows that the slope of δ in the OLS of D on $(1,\delta,S_0,\delta S_0)$ with the observations $S_0 \neq 0$ is $\alpha_\delta + \alpha_{1\delta}(0.5-c_0)$, not α_δ that is the slope of δ in $E(D|G)$ of (5)(i). Doing analogously for $E(Y|S)$ in (5)(ii), we get

$$\begin{aligned} &(i):\ E(Y|S) = \gamma_0 + \gamma_\delta\delta + \gamma_1(S_0+0.5-c_0) + \gamma_{1\delta}\delta(S_0+0.5-c_0) \\ &(ii):\ = \gamma_0 + \gamma_1(0.5-c_0) + \{\gamma_\delta + \lambda_{1\delta}(0.5-c_0)\}\delta + \gamma_1 S_0 + \gamma_{1\delta}\delta S_0 . \end{aligned} \tag{7}$$

(7)(ii) shows that the slope of δ in the OLS of Y on $(1,\delta,S_0,\delta S_0)$ with the observations $S_0 \neq 0$ is $\gamma_\delta + \gamma_{1\delta}(0.5-c_0)$, not γ_δ that is the slope of δ in $E(Y|G)$ of (5)(ii).

From (6)(ii) and (7)(ii), under $E(e|S)=E(e)=0.5$ which holds for any distribution on $[0,1]$ symmetric about 0.5 independently S, it follows that if

$$\alpha_{1\delta}(0.5-c_0) = \gamma_{1\delta}(0.5-c_0) = 0 , \tag{8}$$

i.e., if either the local slope symmetry $\alpha_{1\delta}=\gamma_{1\delta}=0$ holds or $c_0 = 0.5\{=E(e)\}$, then the slopes of δ in (6)(ii) and (7)(ii) become α_δ and γ_δ, respectively, so that the Wald estimator is consistent for β_d in (3) despite S_0, not $G-c$, used in the OLS's. Then, due to IVE=Wald, the slope of D in the IVE to the Y model with the regressors $(1,D,S_0,\delta S_0)$ and IV $(1,\delta,S_0,\delta S_0)$ is also consistent for β_d in (3). Even when (8) is not exactly zero, each product becomes small if either factor is small.

A caution is that, although (5) with G is nonparametric because the true model is polynomially approximated locally around $G=c$, its derived versions (6) and (7) with integer S_0 are not, because a local neighborhood $(S_0=\pm1,\pm2,\ldots)$ of $S_0=0$ is not really local. That is, the nonparametric local linear model in (3) with G becomes a parametric model with S_0 in (6) and (7).

Instead of requiring (8) that is restrictive, we can modify the slopes of δ to identify β_d: from (6)(ii) and (7)(ii), still ruling out $S_0=0$, we get

$$\beta_d = \frac{\{\delta \text{ slope for } E(Y|S) \text{ in (7)(ii)}\} - \{\delta S_0 \text{ slope for } E(Y|S) \text{ in (7)(ii)}\}(0.5-c_0)}{\{\delta \text{ slope for } E(D|S) \text{ in (6)(ii)}\} - \{\delta S_0 \text{ slope for } E(D|S) \text{ in (6)(ii)}\}(0.5-c_0)} . \tag{9}$$

For SRD, replace the denominator with one.

For the local quadratic model (4), using the quadratically extended versions of (5) in between (A.1) and (A.2) of the appendix, the appendix proves that β_d equals

$$\frac{\{\delta \text{ slope for } E(Y|S) \text{ in (A.3)}\} - \{\delta S_0 \text{ slope in (A.3)}\}c_1 + \{\delta S_0^2 \text{ slope in (A.3)}\}c_2}{\{\delta \text{ slope for } E(D|S) \text{ in (A.2)}\} - \{\delta S_0 \text{ slope in (A.2)}\}c_1 + \{\delta S_0^2 \text{ slope in (A.2)}\}c_2}$$

$$\text{where} \quad c_1 \equiv 0.5 - c_0 \quad \text{and} \quad c_2 \equiv (1/6) - c_0 + c_0^2, \tag{10}$$

still ruling out $S_0 = 0$. For SRD, replace the denominator with one.

Differently from (9) and (10) needing only $E(e|S) = E(e) = 0.5$, (10) needs $E(e^2|S) = E(e^2) = 1/3$ extra. Whereas any symmetric distribution on [0,1] satisfies $E(e) = 0.5$, it is hard to think of distributions other than $U[0,1]$ satisfying $E(e^2) = 1/3$ exactly. Of course, it is still possible that both moment conditions are closely met, if not exactly met, by the distribution of e.

In (10), if $c_0 = 0.5$, then $c_1 = 0$, but $c_2 = 0$ only when $c_0 = 0.2115$ or 0.7885: c_1 and c_2 cannot be both zero together. Nevertheless, both terms can be small in reality; e.g., if $c_0 = 0.5$, then $c_1 = 0$ and $c_2 = -0.083$. We can test whether the two terms with c_1 and c_2 in (10) are needed or not in the quadratic models: for $E(D|S)$, test for

$$H_0: \quad -\{\delta S_0 \ \text{ slope in (A.2}\}c_1 + \{\delta S_0^2 \ \text{ slope in (A.2)}\}c_2 = 0. \tag{11}$$

The findings (8) to (11) include the findings in Dong (2015, p. 432, (8)) as a special case, when $c_0 = 0$ in our notation and $c_4 = 0$ in the Dong's notation. Specifically, presenting (10) and 'τ_f' in Dong's equation (8) together—c_0, c_1 and c_2 are defined as the slopes of δ, δS_0 and δS_0^2 for $E(Y|S)$ in Dong (2015)—gives two same entities:

$$\frac{\{\delta \text{ slope in (A.3)}\} - (\delta S_0 \text{ slope})/2 + \{\delta S_0^2 \text{ slope})/6}{\{\delta \text{ slope in (A.2)}\} - (\delta S_0 \text{ slope})/2 + \{\delta S_0^2 \text{ slope})/6} \quad \text{and}$$

$$\tau_f = \frac{c_0 - (c_1/2) + (c_2/6)}{s_0 - (s_1/2) + (s_2/6)}.$$

2.3. OLS and IVE instead of Two-Step Estimators

Dong's (2015) two-stage estimators—(9) for the linear model and (10) for the quadratic model (and their generalizations for higher-order models)—are cumbersome to implement, and Dong (2015) proposed bootstrap inference for the two-stage estimators. The source for the complication is using (6)(ii) and (7)(ii), instead of (6)(i) and (7)(i). Here, we show that single-step OLS for SRD and IVE for FRD are enough to estimate β_d in (3), using (6)(i) and (7)(i).

For SRD with $D = \delta$ and $\gamma = \beta$, still without the cutoff sample, (7)(i) becomes

$$E(Y \mid S) = \beta_0 + \beta_\delta \delta + \beta_1 S_{0.5c} + \beta_{1\delta} \delta S_{0.5c}, \quad S_{0.5c} \equiv S_0 + 0.5 - c_0 = S + 0.5 - c. \quad (12)$$

For this, we can just do the OLS of Y on $(1, \delta, S_{0.5c}, \delta S_{0.5c})$ without the cutoff sample, where the slope of δ is consistent for β_δ. Instead of insisting on using the "natural regressors" $(S_0, \delta S_0)$ as in (6)(ii) and (7)(ii), using the "transformed regressors" $(S_{0.5c}, \delta S_{0.5c})$ as in (6)(i) and (7)(i) leads to a big simplification.

For FRD, we can obtain the Wald estimator using (6)(i) and (7)(i) with the regressors $(1, \delta, S_{0.5c}, \delta S_{0.5c})$, but using 'IVE=Wald' is simpler because we can do the IVE of Y on $(1, D, S_{0.5c}, \delta S_{0.5c})$ with $(1, \delta, S_{0.5c}, \delta S_{0.5c})$ as the IV without the cutoff sample. For both OLS for SRD and IVE for FRD, inference can be done with the usual OLS and IVE asymptotic variance estimators—no need for bootstrap. The OLS and IVE for the quadratic model (4) will appear shortly.

III. Measurement Error Distribution Tests

In (12) for SRD without the cutoff sample, the left and right slopes are β_1 and $\beta_1 + \beta_{1\delta}$, which can be rewritten as

$$E(Y \mid S, S_0 \neq 0) = \beta_0 + \beta_d \delta_+ + \beta_- \delta_- S_{0.5c} + \beta_+ \delta_+ S_{0.5c} \quad (13)$$

where $\delta_- \equiv 1[S_0 \leq -1]$, $\delta_+ \equiv 1[1 \leq S_0]$ and $\delta_0 \equiv 1[S_0 = 0]$. (14)

In (13), the left and right slopes are β_- and β_+; i.e., $\beta_- = \beta_1$ and $\beta_+ = \beta_1 + \beta_{1\delta} = \beta_- + \beta_{1\delta}$, so that $\beta_{1\delta} = \beta_+ - \beta_-$.

Let $\hat{\beta}_{ols,1}$ be the OLS for (13) of Y on

$$W_1 \equiv (1, \delta_+, \delta_- S_{0.5c}, \delta_+ S_{0.5c})' \quad \text{for} \quad \beta \equiv (\beta_0, \beta_d, \beta_-, \beta_+)' \quad (15)$$

without the cutoff sample. Our test for $e \sim U[0,1]$ in SRD is based on the next equation that is proven in the appendix:

$$E(Y \mid S_0 = 0) = \beta_0 + \beta_d (1 - c_0) + \beta_- (-0.5 c_0^2) + \beta_+ 0.5 (1 - c_0)^2 \quad \text{under}$$
$$e \sim U[0,1] \amalg S. \quad (16)$$

Differently from (8) and (9) requiring only $E(e) = 0.5$ in the linear model and (10) and (11) requiring $E(e) = 0.5$ and $E(e) = 1/3$ in the quadratic model, (16) makes use of the assumption $e \sim U[0,1]$ fully in the linear model.

The moment in (16)gives a method-of-moment test statistic: with $\delta_0 \equiv 1[S_0 = 0]$,

$$\frac{1}{\sqrt{N}}\sum_i m_1(Y_i,\hat{\beta}_{ols,1}),$$
$$m_1(Y,\beta)\equiv\delta_0\{Y-\beta_0-\beta_d(1-c_0)-\beta_-(-0.5c_0^2)-\beta_+0.5(1-c_0)^2\}. \tag{17}$$

Since the cutoff sample is not used in obtaining $\hat{\beta}_{ols,1}$, we can test for $e\sim U[0,1]\coprod S$ with $E\{m_1(Y,\beta)\}=0$ by plugging in c_0 and $\hat{\beta}_{ols,1}$.

Taking into account the first-stage error as in Lee (2010, p. 109), the appendix shows that the test statistic is asymptotically normal with the variance estimable by

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

where $C_1\equiv\{1,1-c_0,-0.5c_0^2,0.5(1-c_0)^2\}'$, $\hat{p}\equiv\frac{1}{N}\sum_i 1[S_{0i}=0]$,

$$\eta_{1i}(Y_i,\hat{\beta}_{ols,1})\equiv\left(\frac{1}{N}\sum_i 1[S_{0i}\neq 0]W_{1i}W_{1i}'\right)^{-1}\cdot 1[S_{0i}\neq 0]W_{1i}(Y_i-W_{1i}'\hat{\beta}_{ols,1}). \tag{18}$$

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 FRD test statistic is $N^{-1/2}\sum_i m_1(D_i,\hat{\alpha}_{ols,1})$ with the variance estimator

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

Differentiating $m_1(Y,\beta)$ with respect to c_0 gives

$$\frac{\partial m_1(Y,\beta)}{\partial c_0}=\delta_0\{\beta_d+\beta_-c_0+\beta_+(1-c_0)=\delta_0\{\beta_d+\beta_++(\beta_--\beta_+)c_0\}. \tag{19}$$

If $\beta_d+\beta_++(\beta_--\beta_+)c_0<0$ and if the test statistic value is negative, then $\tilde{c}_0$ that would make the test statistic zero is smaller than c_0 as if the treatment eligibility had started earlier. This situation seems to have occurred in our empirical analysis below where $c_0=0.5$ and Y is the dental expenditure of South Korean elders, because people can wait for many dental treatments. For example, if one needs an implant but is not qualified for a dental care support until age 70, then he/she would wait until age 70. The ability to wait makes late eligibility with c_0 as good as early eligibility with $\tilde{c}_0$.

Consider now a local quadratic version equivalent to (4) for SRD with $D=\delta$: $Y=\beta_0+\beta_d\delta+\beta_-(1-\delta)(G-c)+\beta_+\delta(G-c)+\beta_{--}(1-\delta)(G-c)^2+\beta_{++}\delta(G-c)^2+U_2$;

β_{--} is the left slope of the second order term, and β_{++} is the right slope. The appendix shows that, for SRD without the cutoff sample, do

$$\text{OLS } \hat{\beta}_{ols,2} \text{ of } Y \text{ on } W_2 \equiv (1, \delta_+, \delta_- S_{0.5c}, \delta_+ S_{0.5c}, \delta_- S_0^*, \delta_+ S_0^*)'$$
$$\text{where } S_0^* \equiv S_0^2 + (1-2c_0)S_0 + (1/3) - c_0 + c_0^2. \quad (20)$$

The moment condition to test for $e \sim U[0,1]$ uses (see the appendix again)

$$m_2(Y,\beta) \equiv \delta_0 \left\{ Y - \beta_0 - \beta_d(1-c_0) - \beta_-(-0.5c_0^2) - \beta_+ 0.5(1-c_0)^2 - \beta_{--}\frac{c_0^3}{3} - \beta_{++}\frac{(1-c_0)^3}{3} \right\}. \quad (21)$$

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

$$\frac{1}{N}\sum_i m_2(Y_i, \hat{\beta}_{ols,2})^2 + \hat{p}^2 \frac{1}{N}\sum_i C_2' \eta_{2i}(Y_i, \hat{\beta}_{ols,2}) \eta_{2i}(Y_i, \hat{\beta}_{ols,2})' C_2$$
$$\text{where } C_2 \equiv \left\{ 1, 1-c_0, -\frac{c_0^2}{2}, \frac{(1-c_0)^2}{2}, \frac{c_0^3}{3}, \frac{(1-c_0)^3}{3} \right\}',$$
$$\eta_{2i}(Y_i, \hat{\beta}_{ols,2}) \equiv \left(\frac{1}{N}\sum_i 1[S_{0i} \neq 0] W_{2i} W_{2i}' \right)^{-1} \cdot 1[S_{0i} \neq 0] W_{2i}(Y_i - W_{2i}' \hat{\beta}_{ols,2}).$$

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

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

IV. One-Step OLS and IVE Using Cutoff Sample

If '$e \sim U[0,1]$' is not rejected, we can use all observations including the cutoff sample. Combining (14) and (16) gives (recall $\delta_0 \equiv 1[S_0 = 0]$), for local linear SRD,

$$E(Y \mid S) = \beta_0 + \beta_d\{(1-c_0)\delta_0 + \delta_+\} + \beta_-(-0.5c_0^2\delta_0 + \delta_- S_{0.5c}) + \beta_+\{0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}\}. \quad (22)$$

Compared with (13) excluding δ_0, all regressors in (22) have a δ_0 component. '$(1-c_0)\delta_0$' for β_d makes sense, as the proportion of the treated in the cutoff sample is $1-c_0$.

Based on (22), for the linear model in SRD, do the OLS of Y on

$$W_{1c} \equiv \{1, (1-c_0)\delta_0 + \delta_+, -0.5c_0^2\delta_0 + \delta_- S_{0.5c}, 0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}\}'. \tag{23}$$

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

$$\text{IVE of } Y \text{ on } X_{1c} \equiv \{1, D, -0.5c_0^2\delta_0 + \delta_- S_{0.5c}, 0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}\}'. \tag{24}$$

Because four parameters are estimated in the OLS and IVE, at least four or five support points ($S = 0, \pm 1, \pm 2$) should be used.

For the quadratic model in SRD, recalling S_0^* in (20), do

$$\text{OLS of } Y \text{ on } W_{2c} \equiv \{1, (1-c_0)\delta_0 + \delta_+, -0.5c_0^2\delta_0 + \delta_- S_{0.5c},$$
$$0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}, \frac{c_0^3}{3}\delta_0 + \delta_- S_0^*, \frac{(1-c_0)^3}{3}\delta_0 + \delta_+ S_0^*\}'. \tag{25}$$

This OLS is based on an equation analogous to (22), which is derived in (A.9) of the appendix. For FRD, using W_{2c} as instruments for X_{2c} next, do

$$\text{IVE of } Y \text{ on } X_{2c} \equiv \{1, D, -0.5c_0^2\delta_0 + \delta_- S_{0.5c},$$
$$0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}, \frac{c_0^3}{3}\delta_0 + \delta_- S_0^*, \frac{(1-c)^3}{3}\delta_0 + \delta_+ S_0^*\}'. \tag{26}$$

The above OLS and IVE with non-integer .are improvements on Dong's (2015) two-step procedures using (9) and (10) in two aspects. First, Dong (2015) proposed bootstrap inference after obtaining the estimates using (9) and (10), but the OLS and IVE inferences are much easier. Second, the cutoff sample is used to enhance the efficiency. If 5 integers ($S = 0, \pm 1, \pm 2$) are used with the sample sizes similar across the integers, then we get to use about $100 \times 1/5 = 20\%$ more observations with the cutoff sample added, which translates into a standard deviation (SD) reduction by

$$100\frac{(1.2N)^{1/2} - (N)^{1/2}}{(N)^{1/2}} = 100 \times \{(1.2)^{1/2} - 1\} = 9.5\%.$$

as the SD declines at the $N^{-1/2}$ rate. With 7 integers ($S = 0, \pm 1, \pm 2, \pm 3$), we use $100 \times 1/7 = 14\%$ more observations, which translates into a 6.8% reduction in SD.

V. Simulation Study

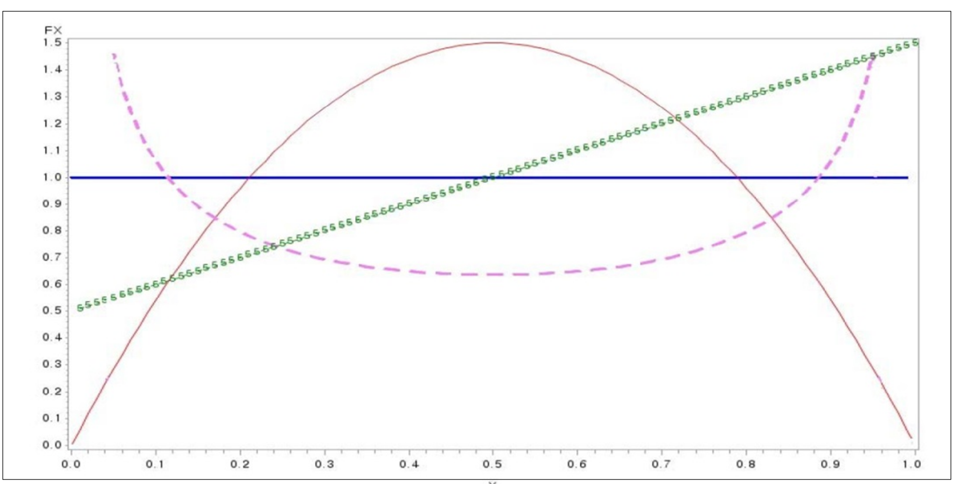
Our base simulation design is a FRD: with 5000 repetitions and $e \sim U(0,1)$,

$c = 0.5,\ 0.2,\ 0.5,\ 0.7$ (implying $c_0 = c$ and $S_0 = S$, due to $\lfloor c \rfloor = 0$)
$N = 2500,\ 5000$ for $S = 0, \pm 1, \pm 2$ and $N = 3500,\ 7000$ for $S = 0, \pm 1, \pm 2, \pm 3$,
$\sum_i 1[S_i = s] = 500 - 50s$ ($\times 2$ for $N = 5000,\ 7000$) with $s = 0, \pm 1, \pm 2, \pm 3$,
$D = (1-\delta)1[V < -0.5] + \delta 1[V > -0.5]$, $V \sim N(0,1) \amalg (e, S)$, $\delta = 1[c \le G]$,
$Y = \beta_0 + \beta_d D + \beta_-(1-\delta)(G-c) + \beta_+\delta(G-c) + V + N(0,1)$,
$\beta_- = 0.5$, $\beta_0 = \beta_d = \beta_+ = 1$.

In the design with birth year in mind for S, we set the number of observations for each support point s to decline as s increases, as the number of births has been declining in developed countries; e.g., they are 600, 550, ..., 400 for $S = -2, -1, \ldots, 2$ for $N = 2500$. We set $c = 0.05$, 0.2, 0.5 and 0.7,as similar cutoffsoccur in reality. For instance, school starts on March 1 in South Korea ($c = 1/6 = 0.17$), on April 1 in Japan ($c = 0.25$), and often on September 1 in the U.S.A. ($c = 0.67$). The reason for $c = 0.05$ is to see a large efficiency gain in using the cutoff sample as follows.

It is likely that the efficiency gain in using the cutoff sample becomes greater as c approaches 0 or 1 because $c \simeq 0,1$ means that the cutoff sample could have been used without ambiguity even when only S is observed. The opposite is likely to hold; i.e., the efficiency gain is likely to be minimal when $c = 0.5$ so that the ambiguity of the cutoff sample treatment status is at its maximum.

[Figure 1] Four Density Functions with Flat Line for $U[0,1]$



To see how robust the estimators are to violations of $e \sim U[0,1]$, 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 either end. 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. Figure 1 plots the four density functions.

In each entry of all tables below, Bias, SD, and Root mean squared error (Rmse) are presented. In each table, four panels for $c = 0.05$, 0.2, 0.5 and 0.7 are provided. 'NAIVE' is the IVE ignoring the integer nature of S; i.e., NAIVE is the IVE of Y on $\{1, D, (1-\delta)S_0, \delta S_0\}$ with $\{1, \delta, (1-\delta)S_0, \delta S_0\}$ as the IV. 'IVE_c' is the IVE in (24) which uses the cutoff sample, 'IVE_{nc}' is the same as IVE_c except that the cutoff sample is dropped, and 'P(Reject)' is the rejection proportion of the uniform

[Table 1] Bias, SD & Rmse and Test for Linear Model ($N = 2500$)

	$U[0,1]$	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.05$		
NAIVE	0.60 0.56 0.82	0.45 0.56 0.72	0.59 0.57 0.82	0.60 0.58 0.83
IVE_{nc}	-0.03 0.61 0.61	-0.17 0.62 0.64	-0.06 0.63 0.63	-0.04 0.63 0.64
IVE_c	0.01 0.36 0.36	-0.17 0.54 0.57	0.00 0.34 0.34	0.02 0.41 0.41
P(Reject)	0.05	0.61	0.06	0.11
		$c = 0.2$		
NAIVE	0.40 0.56 0.69	0.24 0.56 0.61	0.41 0.56 0.69	0.39 0.57 0.69
IVE_{nc}	-0.02 0.59 0.59	-0.17 0.60 0.62	-0.03 0.60 0.60	0.00 0.59 0.59
IVE_c	0.01 0.43 0.43	-0.09 0.58 0.59	-0.01 0.39 0.39	0.06 0.49 0.49
P(Reject)	0.05	0.47	0.13	0.14
		$c = 0.5$		
NAIVE	-0.01 0.57 0.58	-0.14 0.55 0.57	0.00 0.56 0.56	0.00 0.56 0.56
IVE_{nc}	-0.01 0.57 0.58	-0.14 0.55 0.57	0.00 0.56 0.56	0.00 0.56 0.56
IVE_c	0.00 0.57 0.57	-0.12 0.56 0.57	0.00 0.55 0.55	0.02 0.56 0.56
P(Reject)	0.05	0.14	0.05	0.05
		$c = 0.7$		
NAIVE	-0.29 0.56 0.63	-0.41 0.57 0.70	-0.28 0.57 0.63	-0.27 0.57 0.63
IVE_{nc}	0.02 0.55 0.55	-0.13 0.56 0.57	0.00 0.55 0.55	0.01 0.57 0.57
IVE_c	-0.01 0.51 0.51	-0.18 0.52 0.55	0.00 0.49 0.49	-0.04 0.56 0.56
P(Reject)	0.05	0.06	0.14	0.11

Note: $\lfloor c \rfloor = 0$, $S = 0, \pm 1, \pm 2, \pm 3$; Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak $f(e) \propto e(1-e)$; DownUp, $f(e) \propto e^{-0.5}(1-e)^{-0.5}$; NAIVE, ignoring integer; IVE_{nc}, IVE without cutoff sample; IVE_c, IVE with cutoff sample; P(Reject), test rejection proportion.

distribution test for e in (20) 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 about 500 observations. In the first and second panels for $c = 0.05$ and $c = 0.2$, NAIVE is higher biased in all cases. IVE_c using the cutoff sample has much lower Rmse's than IVE_{nc}, which was expected because c is close to 0.Both IVE_{nc} and IVE_c are much biased for the asymmetric distribution Inc 0.5-1.5, but hardly biased for the other distributions. The test has the correct size 0.05 for $U[0,1]$, and rejects the asymmetric non-uniform distribution far more easily than the two Beta distributions.

In the third panel for $c = 0.5$, $\text{NAIVE} = \text{IVE}_{nc} \simeq \text{IVE}_c$ as if the cutoff sample is uninformative when $c = 0.5$, which also agrees with our intuition mentioned above. The near zero biases for $c = 0.5$ in the three symmetric distributions con.rm the no-bias condition (8). 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 fourth panel for $c = 0.7$, the findings are similar to those in the second panel

[Table 2] Bias, SD & Rmse and Test for Linear Model ($N = 5000$)

	$U[0,1]$	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.05$		
NAIVE	0.59 0.39 0.71	0.44 0.39 0.59	0.60 0.38 0.71	0.59 0.38 0.70
IVE_{nc}	-0.01 0.42 0.42	-0.18 0.43 0.47	-0.02 0.42 0.42	-0.01 0.42 0.42
IVE_c	0.00 0.25 0.25	-0.20 0.37 0.42	0.00 0.23 0.23	0.01 0.28 0.28
P(Reject)	0.05	0.89	0.07	0.16
		$c = 0.2$		
NAIVE	0.39 0.38 0.55	0.26 0.38 0.46	0.39 0.39 0.55	0.40 0.39 0.56
IVE_{nc}	-0.01 0.40 0.40	-0.16 0.42 0.45	-0.02 0.41 0.41	-0.02 0.41 0.41
IVE_c	0.01 0.30 0.30	-0.12 0.39 0.41	-0.01 0.28 0.28	0.03 0.34 0.34
P(Reject)	0.05	0.75	0.21	0.21
		$c = 0.5$		
NAIVE	0.01 0.39 0.39	-0.15 0.39 0.42	-0.01 0.39 0.39	0.00 0.39 0.39
IVE_{nc}	0.01 0.39 0.39	-0.15 0.39 0.42	-0.01 0.39 0.39	0.00 0.39 0.39
IVE_c	0.00 0.38 0.38	-0.14 0.39 0.42	-0.01 0.39 0.39	0.01 0.38 0.38
P(Reject)	0.05	0.24	0.05	0.05
		$c = 0.7$		
NAIVE	-0.27 0.39 0.47	-0.41 0.40 0.57	-0.27 0.39 0.48	-0.27 0.40 0.48
IVE_{nc}	0.01 0.39 0.39	-0.14 0.39 0.41	0.00 0.38 0.38	0.00 0.38 0.38
IVE_c	0.00 0.36 0.36	-0.17 0.36 0.40	0.00 0.33 0.33	-0.04 0.38 0.38
P(Reject)	0.05	0.07	0.23	0.17

Note: $\lfloor c \rfloor = 0$, $S = 0, \pm 1, \pm 2, \pm 3$; Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak $f(e) \propto e(1-e)$; DownUp, $f(e) \propto e^{-0.5}(1-e)^{-0.5}$; NAIVE, ignoring integer; IVE_{nc}, IVE without cutoff sample; IVE_c, IVE with cutoff sample; P(Reject), test rejection proportion.

for $c = 0.2$, except for the test not rejecting Inc 0.5-1.5 despite the large bias magnitudes.

In Table 2, the sample size doubles at each support point. The findings in Table 2 are almost the same as those in Table 1, 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 either the same or higher than in Table 1, except for $e \sim U[0,1]$.

Table 3 uses 7 support points $S = 0, \pm 1, \pm 2, \pm 3$, each with about 500 observations so that $N = 3500$, and the quadratic model with $\beta_{--} = 0.05$ and $\beta_{++} = 0.2$ is estimated as well; to keep the table from becoming too long, we omit the case $c = 0.05$. Other than these, the simulation designs are the same as those for Tables 1 and 2. There are two more estimators compared in Table 3: IVE_{ncq} and IVE_{cq}, which are the quadratic-model versions of IVE_{nc} and IVE_{c}. Also P(Reject) presents two rejection proportions of the linear- and quadratic-model tests

[Table 3] Bias, SD & Rmse and Test for Quadratic Model ($N = 3500$)

	$U[0,1]$	Inc 0.5-1.5	QuadPeak	DownUp
		$c = 0.2$		
NAIVE	-0.45 0.21 0.49	-0.53 0.22 0.57	-0.45 0.21 0.50	-0.44 0.22 0.49
IVE_{nc}	-1.15 0.24 1.17	-1.22 0.25 1.24	-1.15 0.24 1.17	-1.14 0.24 1.17
IVE_{c}	-0.60 0.18 0.63	-0.79 0.22 0.82	-0.57 0.16 0.59	-0.63 0.20 0.66
IVE_{ncq}	-0.04 0.72 0.72	-0.09 0.74 0.74	-0.04 0.71 0.71	-0.02 0.73 0.73
IVE_{cq}	0.01 0.38 0.38	-0.02 0.60 0.60	-0.01 0.33 0.33	0.06 0.46 0.46
P(Reject)	0.05, 0.05	1.00, 0.69	0.65, 0.20	0.58, 0.19
		$c = 0.5$		
NAIVE	-0.69 0.21 0.72	-0.76 0.22 0.79	-0.70 0.21 0.73	-0.68 0.22 0.72
IVE_{nc}	-0.69 0.21 0.72	-0.76 0.22 0.79	-0.70 0.21 0.73	-0.68 0.22 0.72
IVE_{c}	-0.62 0.21 0.66	-0.70 0.22 0.73	-0.64 0.21 0.67	-0.61 0.21 0.65
IVE_{ncq}	0.01 0.60 0.60	-0.07 0.62 0.62	0.01 0.60 0.60	0.00 0.62 0.62
IVE_{cq}	0.00 0.59 0.59	-0.05 0.62 0.62	-0.01 0.59 0.59	0.02 0.61 0.61
P(Reject)	0.05, 0.05	0.54, 0.26	0.05, 0.04	0.04, 0.05
		$c = 0.7$		
NAIVE	-0.83 0.22 0.86	-0.88 0.22 0.91	-0.83 0.21 0.86	-0.82 0.22 0.85
IVE_{nc}	-0.40 0.21 0.45	-0.48 0.21 0.52	-0.42 0.20 0.46	-0.39 0.21 0.45
IVE_{c}	-0.53 0.21 0.57	-0.61 0.21 0.64	-0.52 0.19 0.55	-0.54 0.21 0.58
IVE_{ncq}	0.01 0.58 0.58	-0.08 0.61 0.62	0.01 0.59 0.59	0.03 0.61 0.61
IVE_{cq}	-0.02 0.52 0.52	-0.12 0.53 0.54	0.00 0.46 0.46	-0.02 0.58 0.58
P(Reject)	0.05, 0.05	0.12, 0.07	0.57, 0.23	0.39. 0.19

Note: $S = 0, \pm 1, \pm 2, \pm 3$; Inc 0.5-1.5, $f(e) = 0.5 + e$; QuadPeak $f(e) \propto e(1-e)$; DownUp, $f(e) \propto e^{-0.5}(1-e)^{-0.5}$; NAIVE, ignoring integer; IVE_{nc}, IVE without cutoff sample; IVE_{c}, IVE with cutoff sample; IVE_{ncq} & IVE_{cq}, quadratic model versions of IVE_{nc} & IVE_{c}; P(Reject), rejection proportion of test (linear, quad.).

in (17) and (21). We omit the simulation for the quadratic model with each integer point having about 1000observations, because what can be learned from this relative to Table 3 is similar to what was learned from Table 2 relative to Table 1.

The first panel for $c = 0.2$ shows that, compared with IVE_{ncq} and IVE_{cq}, IVE_{nc} and IVE_{c} are heavily biased although their SD's are much smaller. NAIVE is also much biased, although to a lesser degree than IVE_{nc} and IVE_{c}. 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 $\text{NAIVE} = \text{IVE}_{nc} \simeq \text{IVE}_{c}$ and $\text{IVE}_{ncq} \simeq \text{IVE}_{cq}$ as in the panel for $c = 0.5$ in Tables 1 and 2. NAIVE, IVE_{nc} and IVE_{c} are all much biased with smaller SD's as in the first panel for $c = 0.2$. Also, the power is lower for the quadratic model test as in the first panel. The tests fail to reject the two Beta distributions.

In the third panel for $c = 0.7$, differently from $c = 0.2$ and $c = 0.5$, NAIVE is far more biased than IVE_{c}, which is in turn more biased than IVE_{nc} whereas the SD's of NAIVE, IVE_{nc} and IVE_{c} are similar. This shows that using the cutoff sample does not always result in improvements, if the model is misspecified by omitting the quadratic terms, which makes $E(Y \mid S_0 = 0)$ also misspecified as (A.6) and (A.8) in the appendix show for the linear and quadratic models. Differently from IVE_{c} relative to IVE_{nc}, however, IVE_{cq} does 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 hardly biased under the Beta distributions. Comparing IVE_{ncq} and IVE_{cq} with $c = 0.5$, there is some gain in using the cutoff sample when $c = 0.7$: our conjecture on efficiency gain in using the cutoff sample is borne out also in Table 3.

To understand why the test power varies across $U[0,1]$ and the two Beta distributions as c changes, examine the following from the appendix proof for (16): with $\beta_{-} = 0.5$ and $\beta_0 = \beta_d = \beta_{+} = 1$ in the simulation design,

$$\begin{aligned} E(Y \mid S_0 = 0) &= E(Y \mid \delta = 0, S_0 = 0)P(e < c_0) + E(Y \mid \delta = 1, S_0 = 0)P(e > c_0) \\ &= \beta_0 + \beta_{-}\{E(e \mid e < c_0) - c_0\}P(e < c_0) + [\beta_d + \beta_{+}\{E(e \mid e > c_0) - c_0\}]P(e > c_0) \\ &= 1 + 0.5\{E(e \mid e < c_0) - c_0\}P(e < c_0) + [1 + E(e \mid e > c_0) - c_0\}P(e > c_0)\,. \end{aligned} \tag{27}$$

(27) reveals how the differences across the distributions for the test came about: as the test used (27) with the moments and probabilities obtained under $U[0,1]$, the test power depends on how different the moments and probabilities are under the Beta distributions. For this, Table 4 presents $P(e < c_0)$, $E(e \mid e < c_0)$ and $E(e \mid e > c_0)$ in the left column with $c = 0.05$, 0.2 and 0.5 for the three distributions, and then shows (27) minus 1in the right column.

[Table 4] Differences for Test Moment (3.4) across Three Distributions

	$P(e<c_0), E(e\|e<c_0), E(e\|c_0<e)$	(5.1) minus 1
	$c=0.05$	
$U[0,1]$	0.05 0.03 0.53	0.5(0.03-0.05)0.05+(1+0.53-0.05)0.95=1.41
QuadPeak	0.01 0.03 0.51	0.5(0.03-0.05)0.01+(1+0.51-0.05)0.99=1.45
DownUp	0.15 0.02 0.58	0.5(0.02-0.05)0.15+(1+0.58-0.05)0.85=1.30
	$c=0.2$	
$U[0,1]$	0.20 0.10 0.60	0.5(0.10-0.20)0.20+(1+0.60-0.20)0.80=1.11
QuadPeak	0.10 0.13 0.54	0.5(0.13-0.20)0.10+(1+0.54-0.20)0.90=1.20
DownUp	0.30 0.07 0.68	0.5(0.07-0.20)0.30+(1+0.68-0.20)0.70=1.02
	$c=0.5$	
$U[0,1]$	0.50 0.25 0.75	0.5(0.25-0.50)0.50+(1+0.75-0.20)0.50=0.71
QuadPeak	0.50 0.31 0.69	0.5(0.31-0.50)0.50+(1+0.69-0.20)0.50=0.70
DownUp	0.50 0.18 0.82	0.5(0.18-0.50)0.50+(1+0.82-0.20)0.50=0.73

Note: $S=0,\pm1,\pm2$; QuadPeak $f(e)\propto e(1-e)$; DownUp, $f(e)\propto e^{-0.5}(1-e)^{-0.5}$.

First, for $c=0.05$, the values of (27)-1 in Table 4 explain why the test power under DownUp was about twice as high as that under QuadPeak in Tables 1 and 2: 1.41 for $U[0,1]$ is 0.11 apart from 1.30 for DownUp, but only 0.04 apart from 1.45 for QuadPeak. Second, for $c=0.2$, the value 1.11 for $U[0,1]$ is equally apart (by 0.09) from 1.20 and 1.02 for QuadPeak and DownUp, explaining why the test power was almost the same for QuadPeak and DownUp in Tables 1 and 2. Third, for $c=0.5$, the value 0.71 for $U[0,1]$ little differs from 0.70 and 0.73 for QuadPeak and DownUp, explaining why the power was so low in Tables 1 and 2 when $c=$ 0.5. Analogous analyses can be done for $c=0.7$and for Table 3, which are omitted.

VI. Empirical Analysis

The National Health Insurance Service 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. It was expected that about 100,000 to 120,000 elders would bene.t by paying about $1000 less per year for the covered dental treatments that would cost about $1600 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 effects 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. Since $c_0=0.5$ for July 1, if the linear model (3) holds, then there should be no bias even if we use .0 with its integer nature ignored due to the no-bias condition (8).

Table 5 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_0, there are about 200 persons:

S_0 (born 19*xx*) :	−3(48)	−2(47)	−1 (46)	0 (45)	1 (44)	2 (43)	3 (42)
# individuals :	212	205	208	182	176	188	215

For example, there are 182 persons born in 1945 with $S_0 = 0$.

We provide effect estimates with the covariates uncontrolled first and controlled later, but as it turns out, controlling the covariates makes little difference. In our data with birth time G, we have $D = 1[G \le 1945.5]$. To make this fit our framework with $D = 1[c \le G]$, we transform the birth year S into $S_0 = -(S - 1945)$; e.g., $S_0 = -1$ if $S = 1946$, and $S_0 = 2$ if $S = 1943$. Note that, differently from $S_0 \equiv S - \lfloor c \rfloor$ in the previous sections, now $S_0 \equiv -(S - \lfloor c \rfloor)$ to reverse the inequality direction in $G \le c$.

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 become eligible later than July 1 are only partly treated, compared with the fully treated persons who become eligible exactly on July 1. This case differs from the usual RD where individually 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.

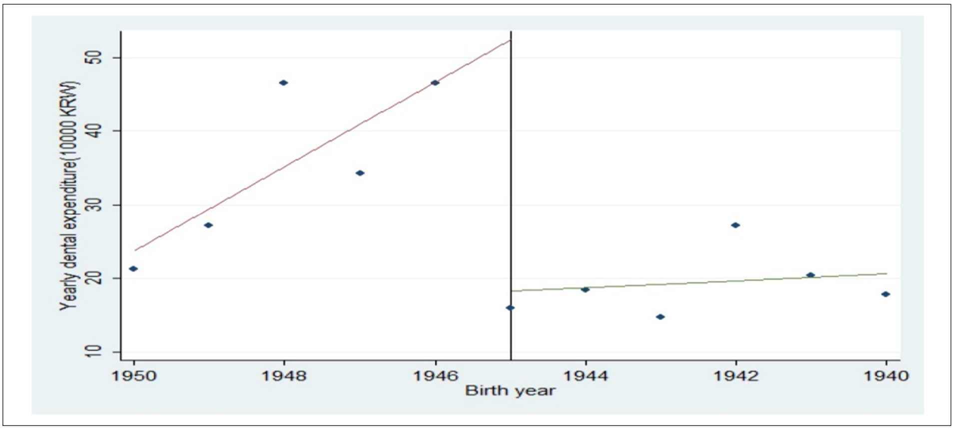
Another unusual aspect in our RD case is that dentures and implants can wait for several months, which means that those who are partly treated may be almost as good as fully treated if they wait until they become eligible in 2015. Indeed, Figure 2 plotting $E(Y|S)$ for birth years $S = 1940 \sim 1950$ reveals this feature. Figure 2 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 data were not used for the linear line estimation. Although we

[Table 5] 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 (for 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)
(median income)	2400		college or higher	0.11 (0.32)

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

[Figure 2] 2015 Mean Dental Expenditure for Each Birth Year Cohort



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 as if the cutoff sample were fully treated. That is, despite $c_0 = 0.5$, the "de facto" c_0 could have been much earlier, because individuals could have waited within 2015 until their eligibility date, postponing the requisite dental treatment to get fully treated in essence.

Viewed more broadly, the real treatment variable may be the continuous waiting time until D is taken as in Angelov et al. (2019), but we do not entertain this general yet complicating view in this section, as it goes beyond the scope of the current paper.

The possibility for waiting until becoming eligible raises an intriguing possibility: even if G were available, using G might lead to a worse finding, because the actual c is ambiguous. That is, using S might be preferred in this kind of unusual RD cases with an ambiguous actual cutoff. To explore this possibility, in our empirical result tables below, we try OLS with $c_0 = 0$, because $c_0 = 0$ corresponds to everybody born in 1945 getting fully treated. These unusual aspects in our empirical example—partial treatment and ambiguity of the de facto c due to waiting—deserve to be fully addressed, but we eschew them, as they also go beyond the scope of the current paper.

Table 6 presents the estimation results without controlling covariates. '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 $c_0 = 0$ as if the treatment had started on January 1, 2015. 'ts' is the moment test statistic value for the linear model. 'ts(pv)-sym' is the χ_1^2. Wald test statistic value (and its p-value) for $\beta_- = \beta_+$ in the linear model. 'OLS_{ncq}', 'OLS_{cq}', 'OLS_{cq0}', 'ts_q' and 'ts_q (pv)-sym' are analogously defined for the quadratic model; β_-, β_+, β_{--} and β_{++} are the slope estimates from OLS_{cq}.

[Table 6] 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)
moment ts	-1.65	-1.06	-1.99	-2.50
ts(pv)-sym	0.17 (0.68)	1.67 (0.20)	0.17 (0.68)	0.54 (0.46)
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)
moment ts_q		-1.69	-0.69	-0.89
ts_q (pv)-sym		0.12 (0.73)	0.01 (0.91)	0.00 (0.99)

Note: N_h, sample size for bandwidth h; OLS_{nc}, linear-model OLS w/o cutoff sample; OLS_c, linear-model OLS with cutoff sample; β_-, β_+ from OLS_c; OLS_{c0}, linear-model OLS with $c_0=0$; moment ts, moment test statistic; ts(pv)-sym, test statistic (p-value) for $\beta_-=\beta_+$; OLS_{ncq}, OLS_{cq}, OLS_{cq0}, ts_q, ts_q (pv)-sym for quadratic model; β_-, β_+, β_{--}, β_{++} from OLS_{cq}.

The effect estimates from the linear model are insignificant for the localizing band-width $h=2$. 3, but become significant for $h=4$. 5; the effect magnitude is fairly stable around $-32\sim-39$ (i.e., $-\$320\sim-\$390$) as Figure 2 indicates. The slope estimates for β_- and β_+ look insignificant for all h values. Also, OLS_{c0} with $c_0=0$ is simi-lar to OLS_{nc} and OLS_c Somewhat disappointingly, OLS_c hardly differs from OLS_{nc} with rather small efficiency gains, which is likely due to $c_0=0.5$. None of the slope symmetry test rejects.

The linear model moment test is not-rejecting with $h=2$. 3, but rejecting with $h=4$. 5. Even when non-rejecting, the moment test statistic 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 D, which is equivalent to having c lower than its nominal value 0.5, as was noted in relation to (19).

For the quadratic model, the effect estimates are insignificant for all h values; judging from the "erratic" β_-, β_+, β_{--} and β_{++} estimates, the quadratic model seems over-specified. The moment test statistic values are all insignificantly negative, and none of the slope symmetry test rejects for the quadratic model.

Table 7 presents the effect estimates (t-values) with the eight covariates in Table 5 controlled. Despite controlling the covariates, the results are remarkably similar to those in Table 6, which demonstrates the robustness of our findings in Table 6.

[Table 7] 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)

VII. Conclusions

Often in regression discontinuity (RD) with a cutoff c, the score (i.e., running variable) G is observed only as its rounded-down integer S. If one proceeds as usual to use S while ignoring the integer nature of S, then there occurs a bias in general. In the RD literature, two-step estimators for the treatment effect have been proposed under the assumption that the measurement error $e = G - S$ follows a known distribution (e.g., $U[0,1]$). When c is not an integer, the existing two-step indirect estimators do not use the "cutoff sample" (the sample with S equalling the integer part of c), because their treatment status is unclear.

In this paper, we made a number of contributions. First, we generalized the existing RD identification findings with integer c to those with any c to show that c may play an important role; e.g., the RD estimator bias due to using S instead of G disappears if the fractional part c_0 of c is 0.5 and the popular linear model is used. Second, when c is not an integer, we use the cutoff sample to test for the distribution assumption on e, and if not rejected, the cutoff sample gets to be used in treatment effect estimation for an efficiency gain. Third, we showed that one-step OLS/IVE whose inference is straightforward is consistent for any c, integer or not, which is easier to implement than the two-step procedure and bootstrap inference suggested for integer c in the literature.

Appendix

Proof for (10)

With $S = S_0 + \lfloor c \rfloor$, $c = \lfloor c \rfloor + c_0$, $E(e) = \mu_1 = 0.5$ and $E(e^2) = \mu_2 = 1/3$,

$$
\begin{aligned}
&E\{(G-c)^2 \mid S\} = E\{(S+e)^2 - 2c(S+e) + c^2 \mid S\} \\
&= S^2 + 2S\mu_1 + \mu_2 - 2cS - 2c\mu_1 + c^2 = S^2 + (1-2c)S + \frac{1}{3} - c + c^2 \\
&= (S_0 + \lfloor c \rfloor)^2 + (1 - 2\lfloor c \rfloor - 2c_0)(S_0 + \lfloor c \rfloor) + \frac{1}{3} - (\lfloor c \rfloor + c_0) + (\lfloor c \rfloor + c_0)^2 \\
&= S_0^2 + 2\lfloor c \rfloor S_0 + (\lfloor c \rfloor)^2 + (1 - 2\lfloor c \rfloor - 2c_0)(S_0 + \lfloor c \rfloor) + \frac{1}{3} - \lfloor c \rfloor - c_0 \\
&\quad + (\lfloor c \rfloor)^2 + 2c_0 \lfloor c \rfloor + c_0^2 = S_0^2 + (1-2c_0)S_0 + \frac{1}{3} - c_0 + c_0^2 .
\end{aligned} \tag{A-1}
$$

Consider the quadratic versions of (5):

$$
\begin{aligned}
E(D \mid G) &= \alpha_0 + \alpha_d \delta + \alpha_1 (G-c) + \alpha_{1\delta}\delta(G-c) + \alpha_2 (G-c)^2 + \alpha_{2\delta}\delta(G-c)^2 , \\
E(Y \mid G) &= \gamma_0 + \gamma_d D + \gamma_1 (G-c) + \gamma_{1\delta}\delta(G-c) + \gamma_2 (G-c)^2 + \gamma_{2\delta}\delta(G-c)^2 .
\end{aligned}
$$

For $E(D \mid G)$, when $S \neq \lfloor c \rfloor \Leftrightarrow S_0 \neq 0$, (A-1) gives

$$
\begin{aligned}
&E(D \mid S) = \alpha_0 + \alpha_d \delta + \alpha_1 (S_0 + 0.5 - c_0) + \alpha_{1\delta}\delta(S_0 + 0.5 - c_0) \\
&+ \alpha_2 \left\{ S_0^2 + (1-2c_0)S_0 + \frac{1}{3} - c_0 + c_0^2 \right\} + \alpha_{2\delta}\delta \left\{ S_0^2 + (1-2c_0)S_0 + \frac{1}{3} - c_0 + c_0^2 \right\} \\
&= \alpha_0 + \alpha_1 \left(\frac{1}{2} - c_0 \right) + \alpha_2 \left(\frac{1}{3} - c_0 + c_0^2 \right) + \left\{ \alpha_d + \alpha_{1\delta} \left(\frac{1}{2} - c_0 \right) + \alpha_{2\delta} \left(\frac{1}{3} - c_0 + c_0^2 \right) \right\} \delta \\
&+ \{\alpha_1 + \alpha_2 (1-2c_0)\} S_0 + \{\alpha_{1\delta} + \alpha_{2\delta}(1-2c_0)\} \delta S_0 + \alpha_2 S_0^2 + \alpha_{2\delta} \delta S_0^2 .
\end{aligned} \tag{A-2}
$$

This makes α_d equal to

$$
\begin{aligned}
&(\delta \text{ slope}) - (\delta S_0 \text{ slope}) \left(\frac{1}{2} - c_0 \right) + (\delta S_0^2 \text{ slope}) \left\{ (1-2c_0) \left(\frac{1}{2} - c_0 \right) - \left(\frac{1}{3} - c_0 + c_0^2 \right) \right\} \\
&= (\delta \text{ slope}) - (\delta S_0 \text{ slope}) \left(\frac{1}{2} - c_0 \right) + (\delta S_0^2 \text{ slope}) \left(\frac{1}{2} - 2c_0 + 2c_0^2 - \frac{1}{3} + c_0 - c_0^2 \right) \\
&= (\delta \text{ slope}) - (\delta S_0 \text{ slope}) \left(\frac{1}{2} - c_0 \right) + (\delta S_0^2 \text{ slope}) \left(\frac{1}{6} - c_0 + c_0^2 \right) .
\end{aligned}
$$

The result analogous to (A-2) holds for γ_d in the next $E(Y|S)$, which then gives (10):

$$E(Y|S) = \gamma_0 + \gamma_1\left(\frac{1}{2} - c_0\right) + \gamma_2\left(\frac{1}{3} - c_0 + c_0^2\right) + \left\{\gamma_d + \gamma_{1\delta}\left(\frac{1}{2} - c_0\right) + \gamma_{2\delta}\left(\frac{1}{3} - c_0 + c_0^2\right)\right\}\delta$$
$$+\{\gamma_1 + \gamma_2(1-2c_0)\}S_0 + \{\gamma_{1\delta} + \gamma_{2\delta}(1-2c_0)\}\delta S_0 + \gamma_2 S_0^2 + \gamma_{2\delta}\delta S_0^2 . \quad \text{(A-3)}$$

Proof for (16)

(3)(i) for SRD with $D = \delta$ can be written as

$$Y = \beta_0 + \beta_d \delta + \beta_-(1-\delta)(G-c) + \beta_+\delta(G-c) + U_1 .$$

This gives

$$E(Y|\delta=0, S=\lfloor c \rfloor) = \beta_0 + \beta_-\{E(G|\delta=0, S=\lfloor c \rfloor) - c\} ;$$
$$E(Y|\delta=1, S=\lfloor c \rfloor) = \beta_0 + \beta_d + \beta_+\{E(G|\delta=1, S=\lfloor c \rfloor) - c\} . \quad \text{(A-4)}$$

Because e is uniform on any subinterval of (0.1), we have, as $\lfloor c \rfloor = c - c_0$,

$$E(G|\delta=0, S=\lfloor c \rfloor) = E(S+e|e<c_0, S=\lfloor c \rfloor) = \lfloor c \rfloor + \frac{c_0}{2} = c - \frac{c_0}{2}$$
$$E(G|\delta=1, S=\lfloor c \rfloor) = E(e|e>c_0, S=\lfloor c \rfloor) = \lfloor c \rfloor + \frac{c_0+1}{2} = c + \frac{1-c_0}{2} .$$

Substituting these into (A-4) renders

$$E(Y|\delta=0, S=\lfloor c \rfloor) = \beta_0 + \beta_-\left(c - \frac{c_0}{2} - c\right) = \beta_0 + \beta_-\left(-\frac{c_0}{2}\right),$$
$$E(Y|\delta=1, S=\lfloor c \rfloor) = \beta_0 + \beta_d + \beta_+\left(c + \frac{1-c_0}{2} - c\right) = \beta_0 + \beta_d + \beta_+\frac{1-c_0}{2} . \quad \text{(A-5)}$$

Because $P(\delta=0|S=\lfloor c \rfloor) = P(e<c_0) = c_0$ as $e \sim U[0,1] \amalg S$,

$$E(Y|S_0=0)$$
$$= E(Y|\delta=0, S_0=0)P(\delta=0|S_0=0) + E(Y|\delta=1, S_0=0)P(\delta=1|S_0=0)$$
$$= \{\beta_0 + \beta_-(-0.5c_0)\}c_0 + \{\beta_0 + \beta_d + \beta_+(0.5-0.5c_0)\}(1-c_0)$$
$$= \beta_0 + \beta_d(1-c_0) + \beta_-(-0.5c_0^2) + \beta_+ 0.5(1-c_0)^2 . \quad \text{(A-6)}$$

Proof for the asymptotic variance of test statistic (17)

Following Lee (2010, p. 109), the asymptotic variance can be estimated with

$$\frac{1}{N}\sum_i A_i^2 \quad \text{where} \quad A_i \equiv m_1(Y_i, \hat{\beta}_{ols,1}) + \left\{\frac{1}{N}\sum_i \frac{\partial m_1(Y_i, \hat{\beta}_{ols,1})}{\partial \beta'}\right\} \eta_{1i}(\hat{\beta}_{ols,1}).$$

Observe

$$\frac{1}{N}\sum_i \frac{\partial m_1(Y_i, \hat{\beta}_{ols,1})}{\partial \beta'} = -\frac{1}{N}\sum_i 1[S_{0i}=0]\{1, 1-c_0, -0.5c_0^2, 0.5(1-c_0)^2\} = -\hat{p}C_1'.$$

The first term of A has $\delta_0 \equiv 1[S_0 = 0]$, and the second term has $1[S_0 \neq 0]$ in $\eta_{1i}(\hat{\beta}_{ols,1})$. Hence, the covariance of the two terms is zero, and the above variance estimator becomes the sum of two individual variances in (18).

Proofs for (20), (21), (25) and (26)

Recall (A-1). For the SRD quadratic model, we get, for all $S_0 \neq 0$,

$$S_0 \leq -1: E(Y \mid S) = \beta_0 + \beta_-\{E(G \mid S) - c\} + \beta_{--}E\{(G-c)^2 \mid S\}$$

$$= \beta_0 + \beta_- S_{0.5c} + \beta_{--}S_0^*, \quad S_0^* = S_0^2 + (1-2c_0)S_0 + \frac{1}{3} - c_0 + c_0^2 \quad \text{from (A-1);}$$

$$S_0 \geq 1: E(Y \mid S = s) = \beta_0 + \beta_d + \beta_+ S_{0.5c} + \beta_{++}S_0^*.$$

Hence,

$$E(Y \mid S, S_0 \neq 0) = \beta_0 + \beta_d\delta_+ + \beta_-\delta_- S_{0.5c} + \beta_+\delta_+ S_{0.5c} + \beta_{--}\delta_- S_0^* + \beta_{++}\delta_+ S_0^*. \quad \text{(A-7)}$$

Observe

$$E(e^2 \mid \delta = 0) = E(e^2 \mid e < c_0) = \int_0^{c_0} e^2 de / c_0 = \frac{1}{3c_0} e^3 \Big|_0^{c_0} = \frac{c_0^2}{3};$$

$$E(e^2 \mid \delta = 1) = E(e^2 \mid c_0 \leq e) = \int_{c_0}^1 e^2 de / (1-c_0) = \frac{1}{3(1-c_0)} e^3 \Big|_{c_0}^1 = \frac{1 + c_0 + c_0^2}{3}.$$

This gives, as $G - c = S + e - c = S_0 + e - c_0 = e - c_0$ on $S_0 = 0$,

$$E\{(G-c)^2 \mid \delta=0, S_0=0)\} = E(e^2 - 2ec_0 + c_0^2 \mid \delta=0)$$
$$= E(e^2 \mid \delta=0) - 2c_0 E(e \mid \delta=0) + c_0^2 = \frac{c_0^2}{3} - 2c_0 \times \frac{c_0}{2} + c_0^2 = \frac{c_0^2}{3};$$
$$E\{(G-c)^2 \mid \delta=0, S_0=0)\} = E(e^2 \mid \delta=1) - 2c_0 E(e \mid \delta=1) + c_0^2$$
$$= \frac{1+c_0+c_0^2}{3} - 2c_0 \times \frac{1+c_0}{2} + c_0^2 = \frac{1+c_0+c_0^2-3c_0-3c_0^2+3c_0^2}{3} = \frac{(1-c_0)^2}{3}.$$

Using these, we have

$$E(Y \mid \delta=0, S_0=0) = \beta_0 + \beta_-\{E(G \mid \delta=0, S_0=0) - c\}$$
$$+\beta_{--}E\{(G-c)^2 \mid \delta=0, S_0=0)\} = \beta_0 + \beta_-(-0.5c_0) + \beta_{--}\frac{c_0^2}{3};$$
$$E(Y \mid \delta=1, S_0=0) = \beta_0 + \beta_d + \beta_+(0.5-0.5c_0) + \beta_{++}\frac{(1-c_0)^2}{3}.$$

Analogously to (A-6), we have ((A-8) below gives the $U[0,1]$ test with (21))

$$E(Y \mid S_0=0) = \left\{\beta_0 + \beta_-(-0.5c_0) + \beta_{--}\frac{c_0^2}{3}\right\}c_0$$
$$+\left\{\beta_0 + \beta_d + \beta_+(0.5-0.5c_0) + \beta_{++}\frac{(1-c_0)^2}{3}\right\}(1-c_0)$$
$$= \beta_0 + \beta_d(1-c_0) + \beta_-(-0.5c_0^2) + \beta_+ 0.5(1-c_0)^2 + \beta_{--}\frac{c_0^3}{3} + \beta_{++}\frac{(1-c_0)^3}{3}. \qquad \text{(A-8)}$$

Combine (A-7) and (A-8) to obtain, for any S including $\lfloor c \rfloor$,

$$E(Y \mid S) = \beta_0 + \beta_d\{(1-c_0)\delta_0 + \delta_+\} + \beta_-(-0.5c_0^2\delta_0 + \delta_- S_{0.5c})$$
$$+\beta_+\{0.5(1-c_0)^2\delta_0 + \delta_+ S_{0.5c}\} + \beta_{--}\left(\frac{c_0^3}{3}\delta_0 + \delta_- S_0^*\right) + \beta_{++}\left\{\frac{(1-c_0)^3}{3}\delta_0 + \delta_+ S_0^*\right\}. \text{(A-9)}$$

Removing δ_0 gives the OLS in (20) for the quadratic model with no cutoff sample.

References

Angelov, N., P. Johansson and M. J. Lee (2019), "Practical Causal Analysis for Treatment-timing Effect on Doubly Censored Duration: Effect of Fertility on Work Span," *Journal of the Royal Statistical Society* (Series A), 182(4), 1561–1585.

Bartalotti, O., Q. Brummet and S. Dieterle (2021), "A Correction for Regression Discontinuity Designs with Group-specific Mismeasurement of the Running Variable," *Journal of Business and Economic Statistics*, 39(3), 833–848.

Battistin, E., A. Brugiavini, E. Rettore and G. Weber (2009), "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Review*, 99(5), 2209-2226.

Cattaneo, M. D. and J. C. Escanciano (2017), *Regression Discontinuity Designs: Theory and Applications*, Emerald Group Publishing.

Cattaneo, M. D., N. Idrobo and R. Titiunik (2019), *A Practical Introduction to Regression Discontinuity Designs*, Cambridge University Press.

Choi, J. Y. and M. J. Lee (2017), "Regression Discontinuity: Review with Extensions," *Statistical Papers*, 58(4), 1217–1246.

______________________ (2018), "Relaxing Conditions for Local Average Treatment Effect in Fuzzy Regression Discontinuity," *Economics Letters*, 173, 47–50.

______________________ (2021), "Basics and Recent Advances in Regression Discontinuity: Difference Versus Regression Forms," *Journal of Economic Theory and Econometrics*, 32, 1–68.

Davesize, L and T. Le Barbanchon (2017), "Regression Discontinuity Design with Continuous Measurement Error in the Running Variable," *Journal of Econometrics*, 200(2), 260–281.

Dong, Y. (2015), "Regression Discontinuity Applications with Rounding Errors in the Running Variable," *Journal of Applied Econometrics*, 30(3), 422–446.

________ (2018), "Alternative Assumptions to Identify LATE in Regression Discontinuity Designs," *Oxford Bulletin of Economics and Statistics*, 80(5), 1020–1027.

Gelman, A. and G. Imbens (2019), "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs," *Journal of Business and Economic Statistics*, 37(3), 447–456.

Hahn, J., P. Todd and W. Van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression-discontinuity Design," *Econometrica*, 69(1), 201–209.

Imbens, G. W. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–476.

Imbens, G. W. and T. Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142(2), 615–635.

Imbens, G. W. and S. Wager (2019), "Optimized Regression Discontinuity Designs," *Review of Economics and Statistics*, 101(2), 264–278.

Kolesar, M. and C. Rothe (2018), "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108(8), 2277–2304.

Lee, D. S. and D. Card (2008), "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142(2), 655–674.

Lee, D. S. and T. Lemieux (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Lee, M. J. (2010), *Micro-econometrics: Methods of Moments and Limited Dependent Variables*, Springer.

_______ (2016), *Matching, Regression Discontinuity, Difference in Differences, and Beyond*, Oxford University Press.

_______ (2017), "Regression Discontinuity with Errors in the Running Variable: Effect on Truthful Margin," *Journal of Econometric Methods*, 6(1), 1–8.

Li, H., X. Shi and B. Wu (2015), "The Retirement Consumption Puzzle in China," *American Economic Review*, 105(5), 437–441.

Oreopoulos, P. (2006), "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *American Economic Review*, 96(1), 152–175.

정수형 점수변수와 비정수 역치를 가진 불연속 회귀모형*

이 명 재** · 심 혜 정*** · 박 상 수****

초 록 불연속회귀 모형에서 점수변수(running variable)가 연속변수인 G대신 내림정수(rounded-down integer) S로 관찰되고, 역치(cutoff) c가 정수가 아닌 경우가 있다. 이 경우, 기존 분석에서는 통제집단인지 처치집단인지 가려내기 힘든 '역치 근처 표본'(cut-off sample)을 배제하고 추정하였다. 이 논문에서는, 먼저 S가 정수임을 간과하고 추정한 경우 추정치 오차가 있음을 보였다 (단, 기울기가 대칭이거나 역치가 어떤 중간 값을 가질 경우, 오차는 0이 됨). 또한, 측정오차 e=G−S의 분포를 가정하고 이를 테스트할 수 있는데, 만약 테스트가 통과된다면 역치 근처 표본을 사용하여 추정의 효율성을 높일 수 있음을 밝혔다. 마지막으로, 기존 2단계 추정이나 부트스트랩 추정보다 간편한 최소자승추정법(OLS) 및 도구변수추정(IVE) 방법도 함께 제시하였다. 실증분석에서는 연령을 기준으로 지급되는 한국의 치과치료 보조비 사례를 분석하였다.

핵심 주제어: 회귀불연속모형, 정수형 점수변수, 비정수 역치

경제학문헌목록 주제분류: C21, C24

투고 일자: 2020. 11. 11. 심사 및 수정 일자: 2021. 7. 6. 게재 확정 일자: 2021. 9. 17.

* 논문에 대하여 도움이 된 지적과 제안을 해 주신 편집위원과 두 명의 익명 심사자께 감사를 표한다.

** 제1저자 및 교신저자, 고려대학교 경제학과 교수, E−mail: myoungjae@korea.ac.kr

*** 공동저자, 한국무역협회 부연구위원, E−mail: assashim@gmail.com.

**** 공동저자, 고려대학교 경제학과 교수, E−mail: starpac@korea.ac.kr