Rounding errors in standardized tests: Revisiting regression discontinuity applications with ACT composite scores

Background/Context

Empirical research in higher education policy has focused on establishing the causal link between policy-relevant variables and academic outcomes. As a quasi-experimental design, regression discontinuity (RD) design can create a local-randomized sample and provide unbiased estimates of the treatment effects, leading to the growing popularity as a way to evaluate higher education policy. Some of the RD studies in higher education are derived from policy cut-offs in standardized exams. In those designs, test scores are usually treated as the primary running variables (e.g. ACT or SAT score). Sample around the cut-off will be treated as a random experiment if certain assumptions are satisfied. The differences in outcomes between above and below the cutoff become good estimates of the average treatment effect.

However, standardized test scores are not perfect running variables in RD design. They are usually reported as discrete variables (or being rounded). For example, the ACT composite score is reported as the average value of the four subject scores, rounded to the nearest whole number[1]. Hence, when applying RD design, due to the discretization of the reported scores, researchers have to extrapolate from the largest value of the score just below the cut-off value to the cut-off value, resulting in the rounding errors in estimates. The extrapolation may create bias in point estimation (Dong, 2015) and also give misleading inference (McCall & Bielby, 2012).

Purpose/Objective/Research Question

To cover the rounding errors and get more robust inference, Lee and Card (2008) suggest using standard errors clustered by the running variable. Oppositely, Kolesár and Rothe (2018) believe this traditional practice has issues and do not recommend clustered errors.

In this paper, we will discuss the issues of rounding errors from test scores in educational settings. Specifically, we primarily focus on the ACT composite score. First, we will summarize the current RD practices from a set of recent education papers using ACT composite score. Second, empirically, we plan to examine the effect of a state-funded merit financial aid program on public college choice. The RD estimates based on both rounded and unrounded scores under various model settings will be compared. Last, we will recommend better guidance for education researchers when applying RD with test scores.

Setting, Population, and Program

We use administrative datasets provided by the Missouri Department of Higher Education (MDHE). The sample includes all Missouri ACT takers who graduated from high school from 1996 to 2006. The primary program we are looking at is the Missouri Bright Flight Scholarship, a merit-based financial aid, which is eligible based on student's highest ACT composite score with the policy cut-off at 30.

[1] For more information, please see here: https://www.act.org/content/act/en/products-and-services/the-act/scores/understanding-your-scores.html

Research Design and Data Analysis

We compare the RD estimates of the effect on public college choice with both rounded and unrounded the ACT composite score. The empirical strategy will follow Harrington et al. (2016) and Zhang et al. (2016). Two model settings are introduced. The first is under the framework of sharp RD. In sharp RD, the highest ACT score is used as the running variable. To remove the bias from retaking behaviors, the times of ACT attempts are added as control variables. The second is under the framework of fuzzy RD. In fuzzy RD, the first ACT score is used as the running variable. Usually, the first test score is assumed as not manipulated (Goodman et al., 2018; Harrington et al., 2016; Zhang et al., 2016) (our data also indicates similar patterns, see Figure 1 and Figure 2). But it will bring biased estimates due to the noncompliance (Lee & Lemieux, 2010). Fuzzy RD can remove the noncompliance bias through a two-stage procedure. In the first stage, the probability of being Bright Flight eligible is estimated based upon being Bright Flight eligible on first ACT attempt (first ACT>=30). In the second stage, actually being Bright Flight eligible is replaced by the predicted being Bright Flight eligible in the first stage. The outcome variable is being enrolled in Missouri public institutions. Please see appendix for details about the model specification.

Findings/Results

Table 1 summarizes some RD applications in recent evaluations of state-funded merit aids. Those financial aids often have requirements on ACT composite scores and the cut-offs have been utilized to implement RD. However, their RD applications are different. Some studies follow the practice from Lee and Card (2008) and use standard errors clustered by ACT to make the statistical inference (Harrington et al., 2016; Leguizamon & Hammond, 2015; Welch, 2014). But others claim the procedure is not clearly an improvement in her study due to the unbalanced size and small number of clusters and universally reduces the estimated standard errors (Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019). Instead, they report robust unclustered errors.

Table 3 presents the estimates from the local linear model with both robust and cluster errors. The finding implies that the robust standard error performs better in smaller bandwidth with the unrounded first ACT score. A possible explanation is, due to the manipulation of the highest ACT, the model errors may be highly correlated within certain clusters (e.g. retaking behaviors, see Figure 3; similar patterns in SAT (Goodman et al., 2018)). Besides, due to the long distance to the cut-off, the cluster errors usually have better coverage properties with a larger bandwidth.

Conclusions

The traditional RD practice recommended by Lee and Card (2008) is using clustered standard errors to cover the rounding errors and to make more robust statistical inference. However, Kolesár and Rothe (2018) point out the poor coverage properties of this clustered error theoretically and empirically. But their empirical evidence is based on a different running variable, age. We extend their application to test scores and find similar but not identical conclusions. In summary, if applicable, we recommend researchers to use unrounded ACT composite score with robust standard errors and report clustered errors to check whether the statistical significance is sensitive to different types of standard errors.

Reference

Bruce, D. J., & Carruthers, C. K. (2014). Jackpot? The impact of lottery scholarships on enrollment in Tennessee. *Journal of Urban Economics*, *81*, 30–44. https://doi.org/10.1016/j.jue.2014.01.006

Dong, Y. (2015). Regression discontinuity applications with rounding errors in the running variable. *Journal of Applied Econometrics*, *30*(3), 422–446. https://doi.org/10.1002/jae.2369

Goodman, J., Gurantz, O., & Smith, J. (2018). *Take Two! SAT retaking and college enrollment gaps* (Vol. No. 24945). Cambridge, MA. https://doi.org/10.3386/w24945

Harrington, J. R., Muñoz, J., Curs, B. R., & Ehlert, M. (2016). Examining the impact of a highly targeted state administered merit aid program on brain drain: Evidence from a regression discontinuity analysis of Missouri's Bright Flight program. *Research in Higher Education*, *57*, 423–447. https://doi.org/10.1007/s11162-015-9392-9

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304. https://doi.org/10.1257/aer.20160945

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674. https://doi.org/10.1016/j.jeconom.2007.05.003

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Leeds, D. M., & DesJardins, S. L. (2015). The effect of merit aid on enrollment: A regression discontinuity analysis of Iowa's National Scholars Award. *Research in Higher Education*, *56*(5), 471–495. https://doi.org/10.1007/s11162-014-9359-2

Leguizamon, J. S., & Hammond, G. W. (2015). Merit-based college tuition assistance and the conditional probability of in-state work. *Papers in Regional Science*, *94*(1), 197–218. https://doi.org/10.1111/pirs.12053

McCall, B. P., & Bielby, R. M. (2012). Regression discontinuity design: Recent developments and a guide to practice for researchers in higher education. In *Higher education: Handbook of theory and research* (pp. 249–290). Springer.

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college Achievement. *Journal of Human Resources*, *46*(3), 614–646. https://doi.org/10.3368/jhr.46.3.614

Scott-Clayton, J., & Zafar, B. (2019). Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid. *Journal of Public Economics*, *170*, 68–82. https://doi.org/10.1016/J.JPUBECO.2019.01.006

Welch, J. G. (2014). HOPE for community college students: The impact of merit aid on persistence, graduation, and earnings. *Economics of Education Review*, *43*, 1–20. https://doi.org/10.1016/j.econedurev.2014.08.001

Zhang, L., Hu, S., Sun, L., & Pu, S. (2016). The effect of Florida's Bright Futures program on college choice: A regression discontinuity approach. *The Journal of Higher Education*, *87*(1), 115–146. https://doi.org/10.1353/jhe.2016.0003

Appendix, Tables, and Figures

Empirical strategy of regression discontinuity design

<u>*Sharp RDD (intend to treat)*</u>

In this setting, public college choice is the dependent variable and the highest ACT score is the running variable. Besides a quadratic functional form of highest ACT score (ACT_HT_i) and a set of control variables (X_i) (gender, race, family income level), the ACT attempts are also included to remove the bias from manipulation.

$$Pub\_College_i = \beta BF\_eligible_i + \gamma' f(ACT\_HT_i) + \theta' X_i + Attempt + \varepsilon_i$$

<u>*Fuzzy RDD (intend to treat)*</u>

In the first stage, being Bright Flight eligible ($BF\_eligible_i$) is estimated based on whether the subject is Bright Flight eligible on the first attempt ($BF\_eligible\_FT_i$), a quadratic functional form of first ACT score (ACT_FT_i), and a set of control variables (X_i) (gender, race, family income level).

$$BF\_eligible_i = \beta BF\_eligible\_FT_i + \gamma' f(ACT\_FT_i) + \theta' X_i + \varepsilon_i$$

In the second stage, the predicted probability of being Bright Flight eligible ($BF\widehat{\_eligi}ble_i$) is substituted for being Bright Flight eligible to estimate the effect of being Bright Flight eligible on public college choice.

$$Pub\_College_i = \beta BF\widehat{\_eligi}ble_i + \gamma' f(ACT\_FT_i) + \theta' X_i + \varepsilon_i$$

Table 1
RDD applications in recent evaluations of merit-based financial aid programs

Program	Eligibility	Paper	RD	Running Variable	Error	Outcome Variable
Tennessee Education Lottery Scholarship program	ACT>=21 or HS GPA>=3.0	Bruce and Carruthers (2014)	Fuzzy RD	First ACT (unrounded)	Standard error clustered by first ACT	College choice
Tennessee Education Lottery Scholarship program	ACT>=21 or HS GPA>=3.0	Welch (2014)	Fuzzy RD	First ACT (unrounded)	Standard error clustered by first ACT	College persistence, graduation, and earnings
Missouri Bright Flight Program	ACT>=30	Harrington et al. (2016)	Sharp RD; Fuzzy RD	Highest ACT; first ACT (rounded)	Standard error clustered by highest/first ACT	Being employed in-state after graduation
Iowa National Scholars Award	Index mixed with ACT and HS GPA	Leeds and DesJardins (2015)	Fuzzy RD	Admissions Index Score/ Regent Admission Index	Standard error	College enrollment
West Virginia PROMISE scholarship	ACT>=21 or HS GPA>=3.0	Scott-Clayton and Zafar (2019)	Fuzzy RD	ACT (rounded, probably highest)	Robust unclustered errors	post-graduation debt, socioeconomic outcomes
West Virginia PROMISE scholarship	ACT>=21 or HS GPA>=3.0	Scott-Clayton (2011)	Fuzzy RD	ACT (rounded, probably highest)	Robust unclustered errors	College persistence, performance and completion
West Virginia PROMISE scholarship	ACT>=21 or HS GPA>=3.0	Leguizamon and Hammond (2015)	Sharp RD	ACT (rounded, probably highest)	Standard error clustered by highest ACT	Being employed in-state after graduation
Florida Bright Futures Program	ACT>=20 or ACT>=28	Zhang et al. (2016)	Sharp RD	Highest ACT (rounded)	Standard errors	College choice

Table 2
Summary statistics

Summary	All Missouri ACT takers		Highest ACT27-31.75	
	Mean	SD	Mean	SD
Gender				
Female	0.547	0.498	0.506	0.500
male	0.446	0.497	0.486	0.500
Gender_missing	0.008	0.086	0.008	0.089
Race				
Race_missing	0.050	0.218	0.058	0.234
Black	0.082	0.275	0.012	0.107
American Indian	0.006	0.076	0.004	0.059
White	0.783	0.412	0.842	0.365
Hispanic	0.009	0.097	0.006	0.076
Asian	0.016	0.126	0.023	0.149
Pacific Islander	0.005	0.069	0.004	0.065
Other	0.011	0.103	0.006	0.077
Two or more races	0.011	0.106	0.010	0.101
Prefer not to respond	0.026	0.160	0.036	0.186
Family income				
Income_missing	0.174	0.379	0.201	0.401
Income zero	0.071	0.257	0.027	0.162
Less than $24,000	0.063	0.243	0.031	0.173
$24,000 to $36,000	0.065	0.246	0.038	0.192
$36,000 to $50,000	0.067	0.250	0.048	0.215
$50,000 to $60,000	0.077	0.266	0.061	0.238
$60,000 to $80,000	0.091	0.288	0.083	0.275
$80,000 to $100,000	0.098	0.297	0.101	0.301
$100,000 to $120,000	0.125	0.331	0.150	0.357
$120,000 to $150,000	0.077	0.266	0.106	0.308
Greater than $150,000	0.093	0.290	0.155	0.362
Observation	446,346		59,319	

Table 3
Statistical inference in different model settings

MO_Pub_HEI	SRD(Robust)	SRD(Cluster)	FRD(Robust)	FRD(Cluster)
	Highest ACT		First ACT	
Unrounded ACT				
24.5-34.25	0.132***	0.132***	0.072***	0.072**
	(0.006)	(0.011)	(0.023)	(0.031)
25.75-33	0.128***	0.128***	0.065**	0.065*
	(0.007)	(0.012)	(0.028)	(0.035)
27-31.75	0.134***	0.134***	0.024	0.024
	(0.008)	(0.013)	(0.035)	(0.029)
28.25-30.25	0.136***	0.136***	-0.028	-0.028
	(0.011)	(0.009)	(0.051)	(0.026)
Rounded ACT				
24.5-34.25	0.124***	0.124***	0.053**	0.053*
	(0.006)	(0.012)	(0.022)	(0.029)
25.75-33	0.119***	0.119***	0.045*	0.045
	(0.006)	(0.012)	(0.024)	(0.031)
27-31.75	0.122***	0.122***	0.014	0.014
	(0.007)	(0.012)	(0.030)	(0.025)
28.25-30.25	0.115***	0.115***	-0.016	-0.016**
	(0.009)	(0.006)	(0.034)	(0.006)

Standard errors in parentheses are clustered by ACT score or robust SE.

*** p<0.01, ** p<0.05, * p<0.10

Note: The bandwidth is under the format of unrounded quarter points. For rounded ACT, 24.5-34.25 equals to 25-34, with 5 data points in each side. Some small bandwidths with rounded ACT do not have enough data points in treatment and control groups. The result is just used for comparison and not recommended for empirical studies. The

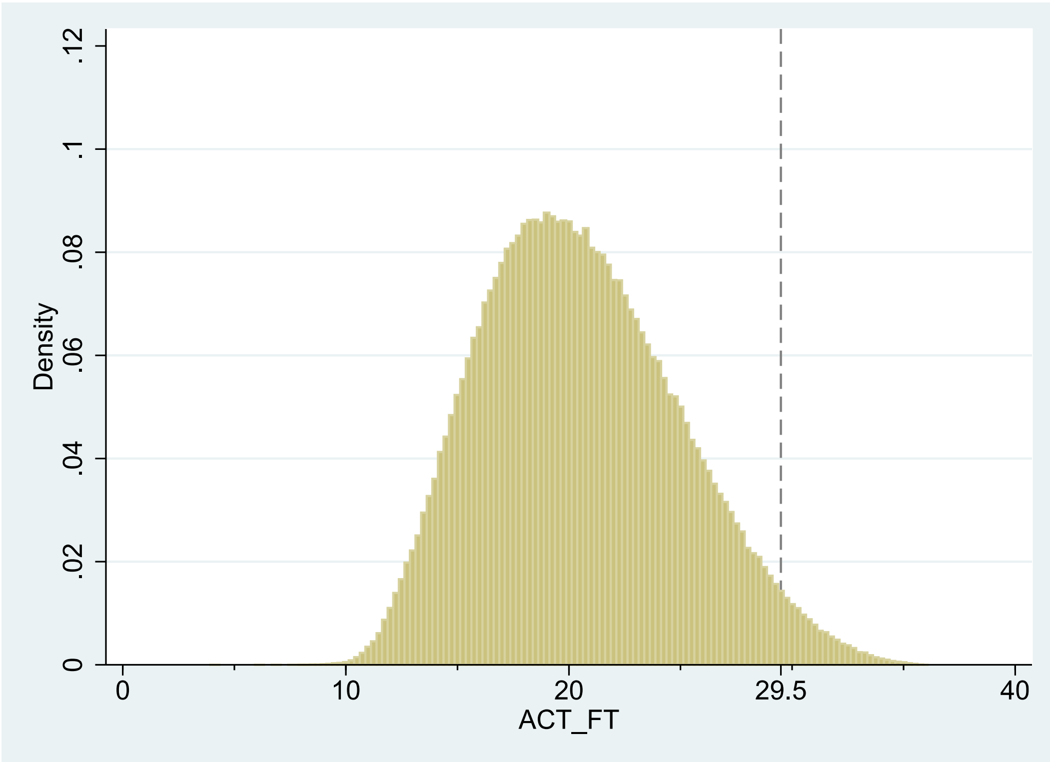

Figure 1 Histogram of first ACT composite score

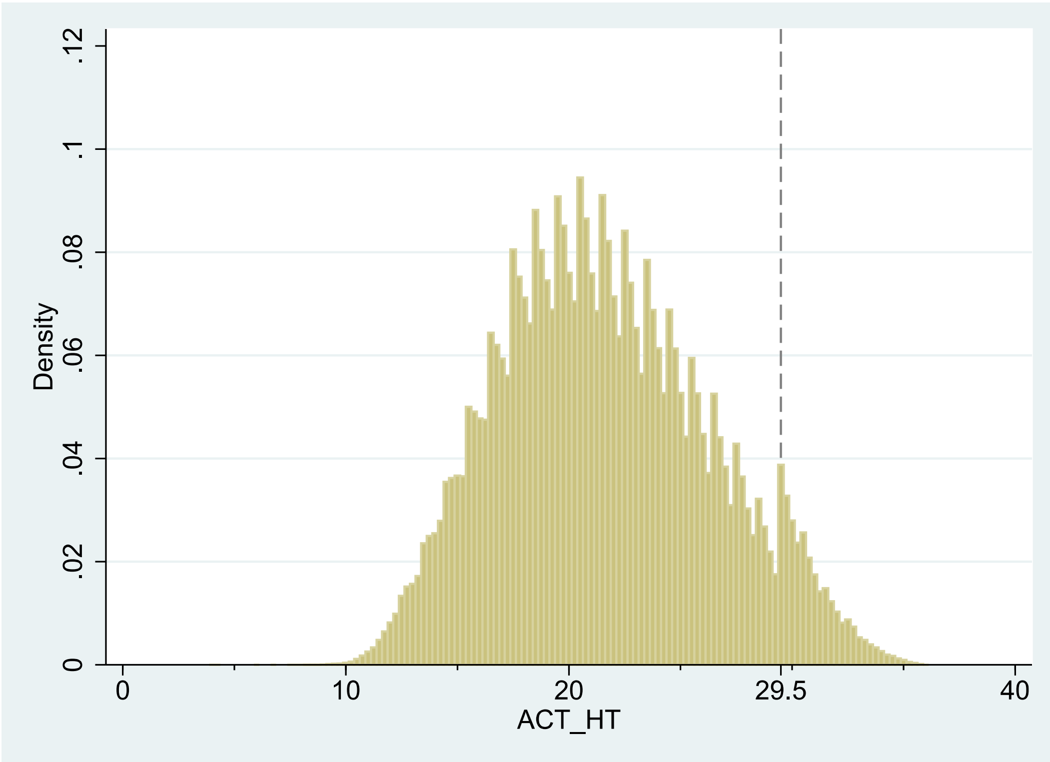

Figure 2 Histogram of highest ACT composite score

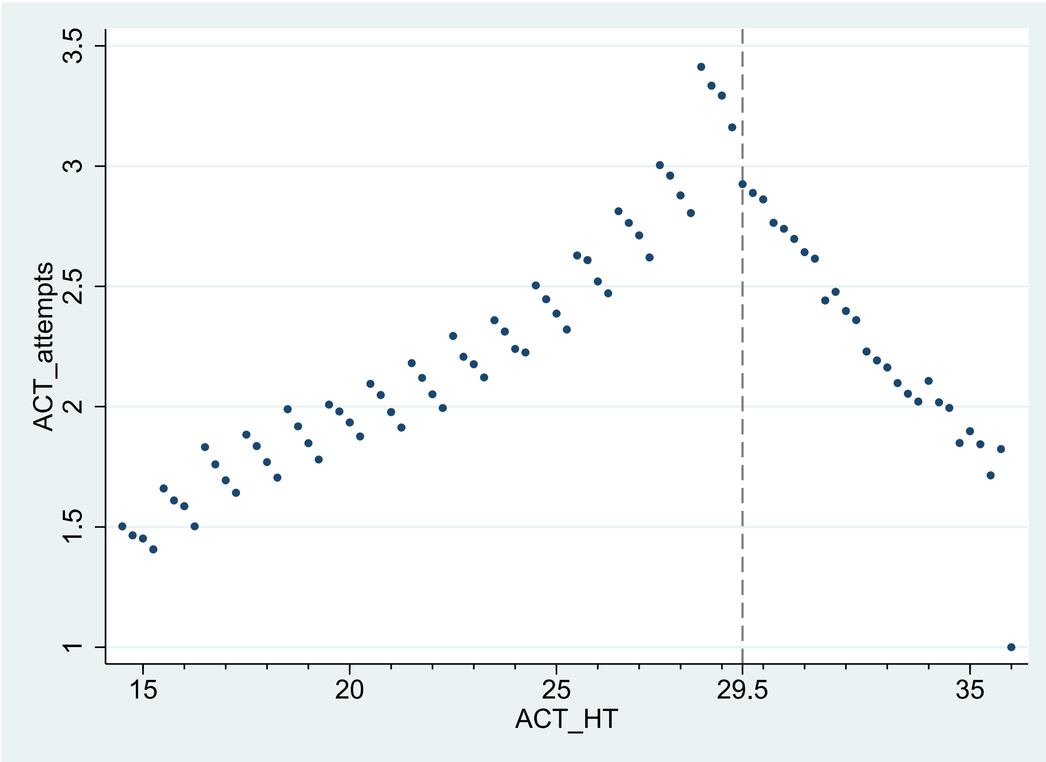

Figure 3 Average ACT attempts by Highest ACT