Economics Bulletin

Volume 35, Issue 3

Exploiting Regression-Discontinuity Design to Estimate Peer Effects in College – The Case of Class Attendance

Qihui Chen
China Agricultural University

Tade Okediji
University of Minnesota

Tian Guoqiang
China Agricultural University

Abstract

We exploit a regression-discontinuity design to estimate peer effects on college students' attendance, using data from a classroom experiment, which required students who scored below a cut-off on the first midterm exam to attend subsequent classes. Since within a small interval around the cut-off, which side of the cut-off a student falls is randomly determined, so is the proportion of a student's classmates falling on one side of the cut-off in the same interval. Using this proportion to instrument peer attendance, we find that a one-point (out of ten) increase in classmates' average attendance score raises a student's attendance score by 0.7 points.

Citation: Qihui Chen and Tade Okediji and Tian Guoqiang, (2015) "Exploiting Regression-Discontinuity Design to Estimate Peer Effects in College – The Case of Class Attendance", *Economics Bulletin*, Volume 35, Issue 3, pages 1563-1571
Contact: Qihui Chen - Chen1006@umn.edu, Tade Okediji - tokediji@umn.edu, Tian Guoqiang - tiangq@cau.edu.cn.
Submitted: October 03, 2014. **Published:** July 11, 2015.

Economics Bulletin

Submission Number: EB-14-00808

Exploiting Regression-Discontinuity Design to Estimate Peer Effects in College – The Case of Class Attendance

Qihui Chen
Assistant Professor

Tade Okediji
Associate Professor

Tian Guoqiang
Associate Professor

Abstract

We exploit a regression-discontinuity design to estimate peer effects on college students' attendance, using data from a classroom experiment, which required students who scored below a cut-off on the first midterm exam to attend subsequent classes. Since within a small interval around the cut-off, which side of the cut-off a student falls is randomly determined, so is the proportion of a student's classmates falling on one side of the cut-off in the same interval. Using this proportion to instrument peer attendance, we find that a one-point (out of ten) increase in classmates' average attendance score raises a student's attendance score by 0.7 points.

Submitted: October 03, 2014.

1. Introduction

Peer effect on student performance has important implications for designing college courses. In the absence of peer effects, grouping students by levels of ability is desirable, since administrators and instructors can then modify resources to fit each group's ability. However, in the presence of positive peer effects, mixing poor-performing students with better-performing ones is more efficient provided that the arrangement does not harm the latter. Nevertheless, peer effects are difficult to identify due to problems such as self-selection (students joining a peer group with similar ability) and reverse-causality (students affecting their peers). Recent quasi-experimental studies (Sacerdote 2001; Winston and Zimmerman 2003; Zimmerman 2003; Chen *et al*. 2014) employed random roommate assignment to resolve these problems. However, roommates may not be a peer group of potential academic influence in college especially if they do not take the same courses together (Stinebrickner and Stinebrickner 2006).

This paper examines the impact of another kind of peer group, potentially of stronger academic influence – classmates – on a student's educational outcomes. It also illustrates how a regression- discontinuity (RD) design can help achieve identification using data from a policy experiment conducted in a large economics course. The experiment required students who scored below a cut-off on the first midterm exam to attend all subsequent classes. The RD design is helpful in that within a small interval around the cut-off, students cannot perfectly manipulate their scores in order to score just above the cutoff; thus, which side of the cut-off a student falls is randomly determined (Lee and Lemieux 2010). The *proportion* of students falling on either side of the cut-off within the same interval is also randomly determined. Using this proportion as an instrumental variable (IV) for peer attendance, we find that a one-point increase (out of 10) in classmates' average attendance score raises a student's attendance score by 0.7 points.

2. Background

The experiment was conducted in the course *Principles of Microeconomics* at the University of Minnesota in Fall 2010 with an enrolment of 232 students (Table 1). Each week, all students attended the same lectures but were assigned to seven different discussion sections that met once a week. Class attendance, which accounted for 10% of the final course grade (the total = 100 points), was recorded in the discussion sections on a weekly basis. In this setting, students assigned to the same discussion section naturally form the peer group of interest in our paper.

Our policy, which required students who scored below 70% on the first midterm exam (which indicates risk of failing the course) should attend all subsequent discussion sessions, was announced at the beginning of the class. Given this policy, a student's midterm score (z) switches on their treatment status D when it passes the cut-off (c =70) smoothly:

$$D = 1(z < c), \qquad (1)$$

Chen and Okediji (2014) show that this RD policy had a significantly positive impact on students' attendance. The current paper further illustrates how it can also help identify peer effect on attendance.

Table 1: Student Profiles in Principles of Microeconomics, Fall 2010

	Observations/[Mean]	%/[SD]
Gender		
Female	167	72.0
Male	65	28.0
Year in college		
High school	17	7.3
Freshman	82	35.3
Sophomore	59	25.4
Junior	44	19.0
Senior	30	12.9
Major		
Agriculture	10	4.3
Animal science	89	38.4
Design	14	6.0
Economics/Business	45	19.4
Engineering	8	3.5
Natural science	14	6.0
Social science	30	12.9
Undeclared	14	6.0
Missing	8	3.5
Attendance score	[9.05]	[1.08]
Mean peer attendance	[9.05]	[0.39]
Midterm I score	[79.73]	[12.78]
Observations	232	100

3. Identification

Consider the commonly adopted linear-in-mean peer-effect specification:

$$y_i = \alpha + \beta \overline{y}_{-i} + \gamma x_i + \varepsilon_i, \qquad (2)$$

where y_i is an educational outcome of interest of student i in a group of size N; $\overline{y}_{-i} = \sum_{j \neq i} y_j /(N-1)$ is the leave-me-out mean of y for student i's peers; x_i is student i's own characteristics; ε is the error term. If equation "(2)" is correctly specified, β is the parameter that represents the peer effect of interest. Yet standard ordinary least-squares (OLS) regression might lead to a biased estimate of β if $\overline{y}_{-i}$ is correlated with ε (e.g. due to reverse causality). To illustrate the problems, we expand $\overline{y}_{-i}$ one step further, which reveals three sources of variation in the peer outcome $\overline{y}_{-i}$:

$$\begin{aligned} \overline{y}_{-i} &= \sum_{j \neq i} y_j /(N-1) = \sum_{j \neq i} (\alpha + \beta \overline{y}_{-j} + \gamma x_j + \varepsilon_j)/(N-1) \\ &= \alpha + \beta_1 \overline{y} + \beta_2 y_i + \gamma \overline{x}_{-i} + \overline{\varepsilon}_j, \qquad (3) \end{aligned}$$

where $\beta_1 = \beta[N(N-2)/(N-1)^2]$, $\beta_2 = \beta/(N-1)^2$, $\overline{y} = (\sum_i y_i)/N$, and $\overline{x}_{-i} = \sum_{j \neq i} x_j/(N-1)$. The first source of variation is the mean outcome of the entire group $\overline{y}$, which is hardly a concern since it is the same for each *i*. The second source is student *i*'s own outcome y_i, which is the source of reverse causality. The third source is the mean peer characteristics $\overline{x}_{-i}$, the appearance of which in equation "(3)" suggests an identification strategy – a sufficiently large exogenous variation in $\overline{x}_{-i}$ (which is not affected by y_i) can be exploited to create an IV for $\overline{y}_{-i}$.

The RD policy under discussion generates such an exogenous variation. To the extent that students within a small interval around the policy cut-off cannot perfectly manipulate their scores to be just above the cutoff, which student eventually passes the cut-off is randomly determined. This implies that the *proportion* of students falling on one side of the cut-off within the same given interval is also randomly determined. Thus for student *i*, the proportion of treated peers $\overline{D}_{-i} = \sum_{j \neq i} D_j/(N-1)$ (which is part of $\overline{x}_{-i}$) defined over a small interval around the cut-off serves as an IV for $\overline{y}_{-i}$ (in practice, different interval widths should be used for robustness checks). Intuitively, suppose student *i* decides to attend classes more frequently after observing an increase in the overall attendance of the class. Part of this observed increase in peer attendance is exogenously induced by the policy, and this part of the policy-induced exogenous variation is exploited to create the candidate IV.

4. Results

Before presenting the main results of this paper, we verify three conditions required for the proposed identification method to work. First, a formal density test (McCrary 2008) verifies continuity of the forcing variable (i.e. midterm I scores z) at the cut-off (t =1.26), the fundamental condition for applying the RD approach (Hahn *et al*. 2001). Second, the policy significantly raised the attendance of poor-performing students' within small (i.e. 15-, 10-, and 5-point) intervals around the cut-off *c*, thereby creating an exogenous variation in students' attendance. Columns 1-3 of Table 2 adopt the standard parametric RD specification, controlling for a cubic function of midterm I scores *z* on either side of the cut-off, and they indicate that the policy raised the attendance score of poor-performing students' by 2.6-3.3 points (out of 10), relative to that of better-performing students near the cut-off. Third, peer attendance is balanced across the cut-off. Columns 4-6 of Table 2 adopt the same RD parametric specification as in columns 1-3 but replace the outcome variable y_i with the mean peer attendance $\overline{y}_{-i}$, and the results indicate a small and insignificant impact of the policy on $\overline{y}_{-i}$. This suggests that the discontinuity in y_i at the cut-off (Fig. 1) is indeed due to the policy, rather than unbalanced peer attendance across the cut-off.[1]

Table 3 presents the main results of estimating Equation "(2)". Column (1) uses the OLS regression, controlling for gender, major, year in college, midterm I scores *z* and the mean peer midterm I score $\overline{z}_{-i}$. It indicates that one additional point in the mean peer

[1] Table 2 in Chen and Okediji (2014) also indicates the balance of students' own characteristics.

attendance $\overline{y}_{-i}$ is associated with a 0.57-point increase in one's own attendance score y_i. To account for possible endogeneity in $\overline{y}_{-i}$, column (2) instruments $\overline{y}_{-i}$ using the proportion of treated peer $\overline{D}_{-i}$ computed based on all students as the IV, which yields a larger and marginally significant peer effect (β = 1.18). However, Shea's partial R^2 for this IV is very low (= 0.059) in the first-stage regression (column 3), signifying a weak-IV problem (Bound *et al*. 1995). As alternatives, $\overline{D}_{-i,c\pm 15}$ (column 5) and $\overline{D}_{-i,c\pm 10}$ (column 7) both have strong predictive power for $\overline{y}_{-i}$ in the first-stage regressions (Shea's partial $R^2 = 0.21$ and 0.43, respectively), indicating no sign of weak-IV problems, while $\overline{D}_{-i,c\pm 5}$ (column 9) has a very small predictive power for $\overline{y}_{-i}$ (Shea's partial $R^2 = 0.062$), probably due to lack of variation in $\overline{D}_{-i,c\pm 5}$. Thus, the relative performance of these IVs suggests the model in column (6) as the preferred model. The result indicates a peer effect of 0.71 points per one-point increase in the average peer attendance score. Given the standard deviation (SD) of the mean peer attendance of 0.39 points (Table 1), the above result implies that a one-SD increase in the mean peer attendance raises a student's attendance score by 0.27 SDs [= 0.71×(1/0.39)], quite a sizable impact.

To further assess how this 0.71-point increase in attendance score translates into a student's final course grade, we quantify the impact of attendance on the final course grade using a fuzzy-RD regression − a regression of students' final course grade on their attendance scores using the treatment status *D* as the IV for the latter. We find that a 1-point increase in students' attendance is associated with a 5.9-point (out of 100) increase in their final course grade, an impact that is statistically significant at the 1% level.[2] Thus, a 0.7-point increase in attendance can lead to a 4.1-point (0.7×5.9) improvement in a student's final course grade. This is close enough to raise a student's final letter grade category by one level, that is, from a C+ (65) to a B-(70).

Table 2: Impacts of RD Policy on Attendance Scores

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	*Student i's* own *attendance score* (y_i)			Mean peer attendance score ($\overline{y}_{-i}$)		
Sample	c ± 15	c ± 10	c ± 5	c ± 15	c ± 10	c ± 5
Treatment status	2.588**	2.647**	3.348**	0.214	0.339	0.317
(D_i)	(1.036)	(0.984)	(1.283)	(0.323)	(0.298)	(0.227)
Observations	116	73	27	116	73	27
R^2	0.156	0.144	0.580	0.251	0.334	0.646

Notes:
All regressions specify a cubic function of midterm I scores on either side of the cutoff, controlling for gender, major, and year in college.
Robust SEs in parentheses, clustered at bins with 3 midterm points.
$^{**}p < .05$.

[2] Detailed results are not reported but available upon request.

Figure 1: RD Policy Impacts on Individual and Peer Attendance

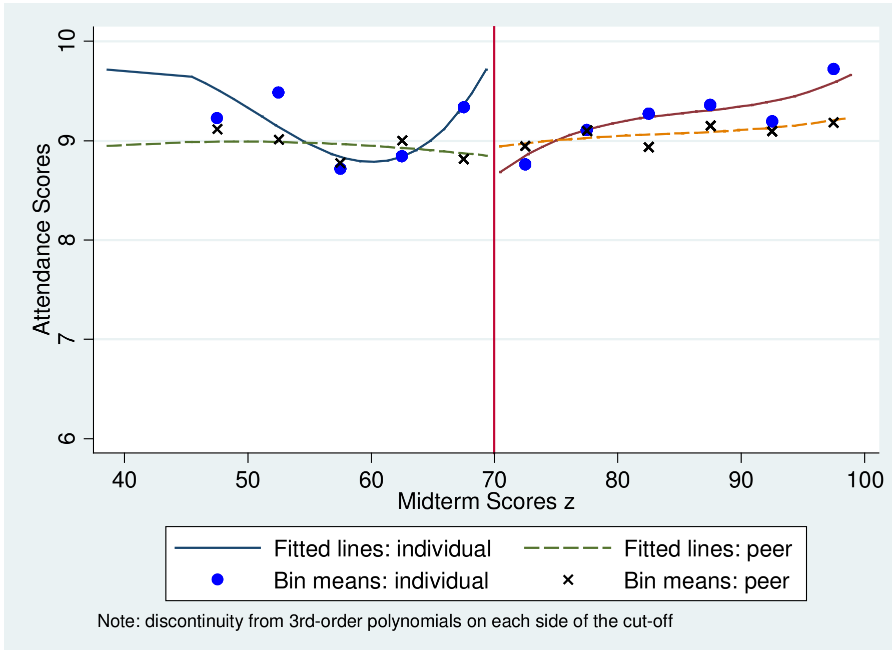

Table 3: Peer Effects on Attendance

	(1) Own attendance (y_i)	(2) Own attendance (y_i)	(3) Mean peer attendance ($\bar{y}_{-i}$)	(4) Own attendance (y_i)	(5) Mean peer attendance ($\bar{y}_{-i}$)	(6) Own attendance (y_i)	(7) Mean peer attendance ($\bar{y}_{-i}$)	(8) Own attendance (y_i)	(9) Mean peer attendance ($\bar{y}_{-i}$)
Estimator	OLS	IV	1st-stage	IV	1st-stage	IV	1st-stage	IV	1st-stage
IV = proportion of treated peer within interval of:	–	full range	–	$c \pm 15$	–	$c \pm 10$	–	$c \pm 5$	–
Mean peer attendance score ($\bar{y}_{-i}$)	0.567***	1.175*		0.982*		0.712**		1.529	
	(0.099)	(0.695)		(0.548)		(0.358)		(0.985)	
Proportion of "treated" peer ($\bar{D}_{-i,\, c\pm15/10/5}$)			0.867***		1.432***		0.672***		0.383***
			(0.262)		(0.227)		(0.128)		(0.114)
Weak-IV tests									
F-statistic			9.74***		46.87***		170.56***		38.54***
P-value			0.0066		0.0000		0.0000		0.0000
Shea's partial R^2			0.059		0.208		0.432		0.062
Observations	227	224	224	224	224	224	224	223	223
R^2	0.223	–	0.626	–	0.685	–	0.774	–	0.624

Notes:
Robust SEs in parentheses, clustered at bins with 3 midterm points.
All regressions control for gender, major, year in college, midterm I score and classmates' mean midterm I score.
$*p < .10$; $**p < .05$; $***p < .01$.

5. Conclusion

We propose a regression discontinuity approach to estimate peer effects on class attendance at the collegiate level by exploiting the notion of a regression discontinuity design as a locally randomized experiment. Using data from a classroom policy experiment in which students who scored below a cut-off on the first midterm exam were required to attend subsequent classes, we used the proportion of a student’s classmates falling on one side of the cut-off within a small interval around the cut-off as an instrument for peer attendance. We found that a one-point increase in classmates' average attendance score increases a student’s attendance score by 0.7 points. Such an impact can translate into a 4.1-point increase in a student's final course grade.

References

Bound, J., D. A. Jaeger, and R. Baker (1995) "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak" *Journal of the American Statistical Association* **90**, 443-450.

Chen, Q and T. O. Okediji (2014) "Incentive Matters!—The Benefit of Reminding Students about Their Academic Standing in Introductory Economics Courses" *Journal of Economic Education* **45**, 1-24.

Chen, Q., G. Tian and T. O. Okediji. (2014). "Quasi-Experimental Evidence of Peer Effects in First-Year Economics Courses at a Chinese University". *Journal of Economic Education* 45(4): 304-319.

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.

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

McCrary, J. (2008) "Manipulation of the running variable in the regression discontinuity design: A density test" *Journal of Econometrics* **142**, 698–714.

Sacerdote, B. (2001) "Peer Effects with Random Assignment: Results for Dartmouth Roommates"*Quarterly Journal of Economics* **116**, 681-704.

Stinebrickner, T., and R. Stinebrickner (2006) "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds" *Journal of Public Economics* **90,** 1435-1454.

Winston, G and D. Zimmerman (2003) "Peer Effects in Higher Education" in *College Decisions: How Students Actually Make Them and How They Could* by C. Hoxby, Ed., University of Chicago Press, 395-424.

Zimmerman, D. (2003) "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment" *Review of Economics and Statistics* **85,** 9–23.