Are Some Innovation Self-Reports in the Annual Business Survey Biased? A Regression Discontinuity Test

Luyi Han, Zheng Tian, and Stephan J. Goetz
Northeast Regional Center for Rural Development
Penn State University

Timothy R. Wojan
National Center for Science and Engineering Statistics
National Science Foundation

This presentation provides results of exploratory research sponsored in part by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF). This information is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any views expressed are those of the authors and not necessarily those of NCSES or NSF.

Motivation

- An ongoing debate within innovation survey methodology concerns possible underreporting of innovation in joint innovation-R&D surveys relative to innovation-only surveys.

- Design of the 2018 ABS: microbusinesses (fewer than 10 employees) answer both an R&D and innovation module but where businesses with 10 or more employees do not receive an R&D module.

Research Question

- Whether microbusinesses just under the sharp threshold (10 employees) are less likely to report innovation relative to small businesses just over the threshold using a regression discontinuity design (RDD)? Or if we can observe a jump in relevant innovation variables from 9 -10 employees.

Preview of Results

- We find no statistically significant effects at the 10-employee cutoff for both innovation variables (new-to-market and new-to-business innovation).

- The RDD power tests show our RDD results are estimated with a high degree of power against a reasonable range of alternatives, permitting the more definitive conclusion that our nonsignificant results can be interpreted as no bias in self-reported innovation due to the ABS design.

Previous Literature

- Both the U.S. and Norway had self-reported innovation rates considerably below those of their EU peers.

- Tian, et al. (2022) attempted to compare responses from the same firm to an innovation-only and combination innovation-R&D survey by linking the 2014 BRDIS to the 2014 Annual Survey of Entrepreneurs (ASE)—the precursor to the ABS that also included an innovation module but no R&D module.

- Hoskens, et al. (2016), Cirera and Muzi (2016), Tourangeau, et al. (2000), etc.

Data

- 2018 Annual Business Survey (ABS, 2017 reference year) contains innovation self-reports, NAICS industry, and firm size.

- Employment data in the Longitudinal Business Database (LBD) come from the 2016 Business Register (BR) that was used to draw the 2018 ABS sample, with employment in the BR determining which firms received the microbusiness (<10 employees) survey containing the R&D module and all remaining firms receiving the innovation-only survey.

Methods and Models

- $Innovation_i = \alpha + \tau D_i + \beta(Emp_i - 10) + \gamma(Emp_i - 10) \times D_i + \rho X_i + \epsilon_i$
- RDD for the local average treatment effect (LATE)
  - Dependent variable: new-to-market or new-to-business innovation, dummy or fitted value
  - The RDD treatment variable $D_i = 1$ if firms have 10 or more employees
  - The $Emp_i - 10$ is the normalized running variable for number of employees
  - We include an interaction term of the treatment and the normalized running variables to allow the slopes to vary on both sides of the cutoff point
  - X_i are the firm level control variables, including the highest education of the firm owner, and the NAICS 2-digit industry dummies

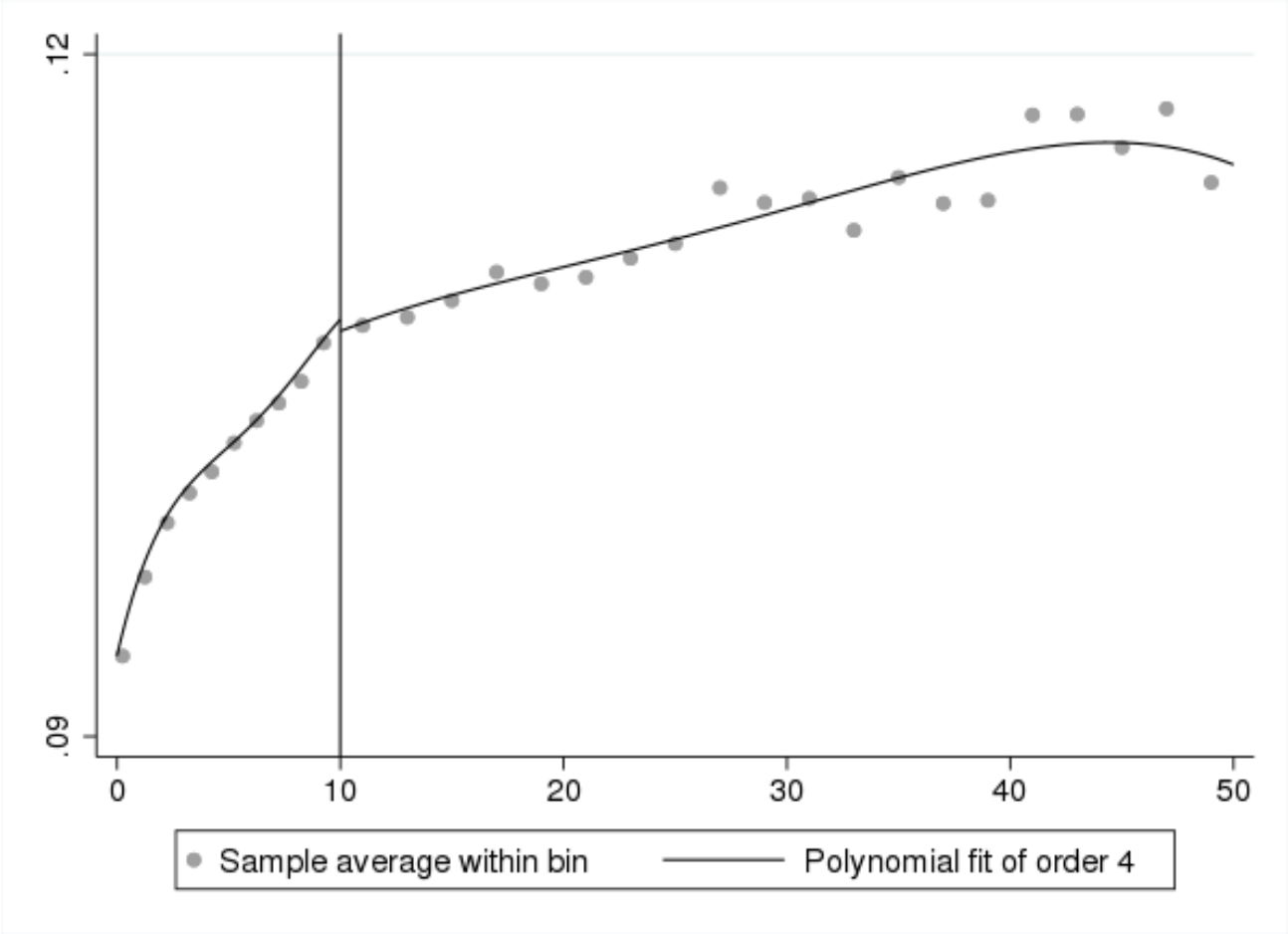
Methods and Models

- Bandwidths in RDD regressions: we use the mean-squared-error (MSE) optimal bandwidth choice method to determine the bandwidth. We also report estimates using half bandwidth and two bandwidths as robustness checks.

- RDD validity: assignment wrt threshold *ex ante*. Respondents denied ability to manipulate assignment on either side of threshold.

- RDD power tests: evaluate both the robustness of our findings along with the validity of inferences from negative findings
  - What effect size are we trying to detect? 2.5% change in innovation rate

Figure 1 RDD Plot for Fitted New-to-market Innovation



Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database.

Notes: Disclosure Avoidance Request 10151 for FSRDC Project 2083.

Table 1 Summary Statistics

	firmsize_emp	New-to-market	New-to-business	Observations
All sample	16.74	0.1031	0.1783	464000
std.err	(0.13)	(0.0004)	(0.0006)	
Below cutoff 10	2.987	0.0955	0.1706	334000
std.err	(0.004)	(0.0005)	(0.0007)	
Above cutoff 10	51.97	0.1227	0.1981	130000
std.err	(0.4390)	(0.0009)	(0.0011)	

Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database.

Notes: Standard errors are shown in the parentheses. Disclosure Avoidance Request 10151 for FSRDC Project 2083

Table 2 RDD for New-to-market Innovation

	(1)	(2)	(3)	(4)
Panel A: LPM Model of Binary Y				
treated10 = 1	0.0074***	-0.0069	0.0012	0
	(0.0018)	(0.0079)	(0.0051)	(0.0030)
firmsize_emp_rd10	0.0011***	0.0043	0	0.0013***
	(0.0002)	(0.0047)	(0.0019)	(0.0004)
1.treated10#c.firmsize_emp_rd10	-0.0011***		0.0026	-0.0004
	(0.0002)		(0.0025)	(0.0007)
Observations	464000	28500	71000	203000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth
Panel B: Logit Model of Binary Y				
treated10 = 1	0.0662***	-0.0706	0.0129	-0.0019
	(0.0197)	(0.0818)	(0.0538)	(0.0324)
firmsize_emp_rd10	0.0140***	0.0442	0	0.0147***
	(0.0024)	(0.0488)	(0.0201)	(0.0044)
1.treated10#c.firmsize_emp_rd10	-0.0138***		0.0266	-0.0058
	(0.0024)		(0.0259)	(0.0077)
Observations	464000	28500	71000	203000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth
Panel A: Probit Fitted Y				
treated10 = 1	0.0023***	-0.0008	-0.0004	0
	(0.0003)	(0.0014)	(0.0009)	(0.0006)
firmsize_emp_rd10	0.0014***	0.0011	0.0013***	0.0010***
	0.0000	(0.0008)	(0.0003)	(0.0001)
1.treated10#c.firmsize_emp_rd10	-0.0014***		-0.0011***	-0.0009***
	0.0000		(0.0004)	(0.0003)
Observations	464000	28500	71000	185000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth

Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database.

*Notes: Standard errors are shown in the parentheses. *** p < 0.01. Disclosure Avoidance Request 10151 for FSRDC Project 2083.*

Table 3 RDD for New-to-business Innovation

	(1)	(2)	(3)	(4)
Panel A: LPM Model of Binary Y				
treated10 = 1	0.0126***	-0.0034	0.0057	-0.0018
	(0.0023)	(0.0098)	(0.0064)	(0.0038)
firmsize_emp_rd10	0.0008***	0.0062	-0.001	0.0016***
	(0.0003)	(0.0059)	(0.0024)	(0.0005)
1.treated10#c.firmsize_emp_rd10	-0.0007***		0.0019	0.0006
	(0.0003)		(0.0031)	(0.0009)
Observations	464000	28500	71000	203000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth
Panel B: Logit Model of Binary Y				
treated10 = 1	0.0836***	-0.0222	0.0392	-0.0125
	(0.0158)	(0.0655)	(0.0432)	(0.0261)
firmsize_emp_rd10	0.0057***	0.0415	-0.0068	0.0113***
	(0.0019)	(0.0391)	(0.0161)	(0.0035)
1.treated10#c.firmsize_emp_rd10	-0.0056***		0.0132	0.0028
	(0.0019)		(0.0208)	(0.0062)
Observations	464000	28500	71000	203000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth
Panel A: Probit Fitted Y				
treated10 = 1	0.0004	-0.0005	-0.0008	-0.0004
	(0.0003)	(0.0014)	(0.0009)	(0.0006)
firmsize_emp_rd10	0.0013***	0.0005	0.0012***	0.0009***
	0.0000	(0.0008)	(0.0003)	(0.0001)
1.treated10#c.firmsize_emp_rd10	-0.0013***		-0.0013***	-0.0010***
	0.0000		(0.0004)	(0.0003)
Observations	464000	28500	71000	185000
Sample	all	0.5 bandwidth	bandwidth	2 bandwidth

Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database.

*Notes: Standard errors are shown in the parentheses. *** $p < 0.01$. Disclosure Avoidance Request 10151 for FSRDC Project 2083.*

Table 4 RDD Power Test for New-to-market Innovation

Cutoff c = 0	Left of c	Right of c
Number of obs	334000	130000
Eff. Number of obs	40000	31000
BW loc. poly. (h)	3.844	3.844
Order loc. poly. (p)	4	4
Sampling BW	3.844	3.844
New sample	40000	31000

Number of obs =	464000
BW type =	mserd
Kernel =	Triangular
VCE method =	NN
Derivative =	0
HA: tau =	0.05

Power against:	H0: tau=0	0.2*tau =0.01	**0.5*tau =0.025**	0.8*tau =0.04	Tau =0.05
Robust bias-corrected	0.05	0.492	**0.998**	1	1

Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database. .

Notes: Disclosure Avoidance Request 10151 for FSRDC Project 2083.

Table 5 RDD Power Test for New-to-business Innovation

Cutoff c = 0	Left of c	Right of c
Number of obs	334000	130000
Eff. Number of obs	40000	31000
BW loc. poly. (h)	3.975	3.975
Order loc. poly. (p)	4	4
Sampling BW	3.975	3.975
New sample	40000	31000

Number of obs =	464000
BW type =	mserd
Kernel =	Triangular
VCE method =	NN
Derivative =	0
HA: tau =	0.05

Power against:	H0: tau=0	0.2*tau =0.01	**0.5*tau =0.025**	0.8*tau =0.04	Tau =0.05
Robust bias-corrected	0.05	0.345	**0.974**	1	1

Sources: 2018 Annual Business Survey and 2016 Longitudinal Business Database. .

Notes: Disclosure Avoidance Request 10151 for FSRDC Project 2083.

Summary of Results

- Our results show no statistical difference in self-reported innovation below or above the 10-employee cutoff.

- The power tests show our RDD results are estimated with a high degree of power, allowing a statistical inference from our nonsignificant results that can be interpreted as no bias in self-reported innovation

Implications

- For innovation survey methodology, our results suggest that combination R&D/innovation surveys need not cause respondents to underreport innovation if they are insulated from potentially irrelevant questions (e.g., those with no R&D activity should not be asked about hiring of science and engineering PhDs).
- This extends to the broader interest in a future modular, single enterprise platform where firms receive a single, customized Census survey each year
  - Sequence modules from more general to more specific topic areas
  - Pay special attention to how responses in one module may influence responses in other modules
  - Web survey collection essential for insulating respondents from irrelevant Qs