Principled Estimation of Regression Discontinuity Designs

L. Jason Anastasopoulos

School of Public and International Affairs (SPIA)
Department of Statistics
Institute for Artificial Intelligence

University of Georgia

Me as a grad student at UC Berkeley in 2013

RDD paper that I was working on...

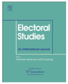
Electoral Studies xxx (2016) 1–8

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

ELSEVIER



Estimating the gender penalty in House of Representative elections using a regression discontinuity design

Lefteris Anastasopoulos

UC Berkeley School of Information, 102 S. Hall Road, Berkeley, CA 94704, USA

A R T I C L E I N F O

Article history:
Received 12 September 2015
Received in revised form
9 November 2015
Accepted 21 April 2016
Available online xxx

A B S T R A C T

While the number of female candidates running for office in U.S. House of Representative elections has increased considerably since the 1980s, women continue to account for about only 20% of House members. Whether this gap in female representation can be explained by a gender penalty female candidates face as the result of discrimination on the part of voters or campaign donors remains uncertain. In this paper, I estimate the gender penalty in U.S. House of Representative general elections using a regression discontinuity design (RDD). Using this RDD, I am able to assess whether chance nomination of female candidates to run in the general election affected the amount of campaign funds raised, general election vote share and probability of victory in House elections between 1982 and 2012. I find no evidence of a gender penalty using these measures. These results suggest that the deficit of female representation in the House is more likely the result of barriers to entering politics as opposed to overt gender discrimination by voters and campaign donors.

Me by the time the paper was published in 2016

“Researcher degrees of freedom” for treatment effect estimation

General (3)	Regression Discontinuity Design (9)
Dependent variable/outcome.	Dependent variable/outcome.
Treatment.	Treatment.
Pre-treatment covariates.	Pre-treatment covariates.
	Forcing variable.
	Kernel.
	Bandwidth.
	Bandwidth method.
	Data subset (based on forcing variable).
	Type of model.

Too many researcher degrees of freedom → more false positive results

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons, Leif D. Nelson, Uri Simonsohn

First Published October 17, 2011 | Research Article | Find in PubMed | Check for updates
https://doi.org/10.1177/0956797611417632

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Also...most RDD treatment effects are underpowered...

Journal (Year),Author(s)	Title	DV	Forcing	Covariate Types	Lowest N
APSR (2009) Eggers and Hainmueller	"MPs for Sale? Returns to Office in Postwar British Politics"	Logged wealth at death	Vote share margin	Candidate/official level traits	165
APSR (2014), Ferwerda and Miller	"Political devolution and resistance to foreign rule: A natural experiment"	attacks	commune distance from demarcation line	mean elevation, train station distance, communications available, farmed area, ruggedness of the landscape, population	15
APSR (2015), Hall	"What happens when extremists win primaries?"	party victory	Vote share margin	Congress fixed effects	35
APSR (2018), Szakonyi	"Businesspeople in elected office: Identifying private benefits from firm-level returns"	Revenue and profit margins	vote share margin	sector, region, year fixed effects, candidate level covariates	136
AJPS (2011), Boas and Hidalgo	"Controlling the airwaves: Incumbency advantage and community radio in Brazil"	radio station coverage	vote share margin	municipal population	33
JOP (2014), Boas, Hidalgo, and Richardson	"The spoils of victory: campaign donations and government contracts in Brazil"	total contracts	vote share margin	firm level fixed effects	45

Table 1 – Covariate types chosen for RDD estimation in top political science journals.

...and many different types of pre-treatment covariates are used w/ similar data

Journal (Year),Author(s)	Title	DV	Forcing	Covariate Types	Lowest N
APSR (2009) Eggers and Hainmueller	"MPs for Sale? Returns to Office in Postwar British Politics"	Logged wealth at death	Vote share margin	Candidate/official level traits	165
APSR (2014), Ferwerda and Miller	"Political devolution and resistance to foreign rule: A natural experiment"	attacks	commune distance from demarcation line	mean elevation, train station distance, communications available, farmed area, ruggedness of the landscape, population	15
APSR (2015), Hall	"What happens when extremists win primaries?"	party victory	Vote share margin	Congress fixed effects	35
APSR (2018), Szakonyi	"Businesspeople in elected office: Identifying private benefits from firm-level returns"	Revenue and profit margins	vote share margin	sector, region, year fixed effects, candidate level covariates	136
AJPS (2011), Boas and Hidalgo	"Controlling the airwaves: Incumbency advantage and community radio in Brazil"	radio station coverage	vote share margin	municipal population	33
JOP (2014), Boas, Hidalgo, and Richardson	"The spoils of victory: campaign donations and government contracts in Brazil"	total contracts	vote share margin	firm level fixed effects	45

Table 1 – Covariate types chosen for RDD estimation in top political science journals.

Combination leads to unstable treatment effects

- Bandwidth.
- Covariates included.

A Simulated Example

$N = 100$ simulated election forcing variable $F \sim Unif(-0.5, -0.5)$, two covariates, weakly correlated w/ F, $r = 0.02$.

$$Y = \alpha + \tau D + \beta_1 F + \beta_2 F^2 + \beta_3 X_1 + \beta_4 X_2 + \epsilon$$

Quantity	No Covariates	Two Covariates
$\hat{\tau}_{cct}$	0.036	0.091
h	0.074	0.063
p-value	0.414	0.109

Real example from Szakonyi (2018) APSR

Replication of "Political Connections and Firm Profitability" in Szakonyi (2018) with Adaptive LASSO Adjusted Treatment Effects.

	Original (APSR)	Adaptive	Original 5% (APSR)	Adaptive 5%	Adaptive CCT Robust
District Win	0.146***	0.102*	0.198**	0.097**	0.140***
	(0.065)	(0.060)	(0.090)	(0.038)	(0.052)
Bandwidth	0.113	0.120	0.050	0.050	0.120
Covariates Dropped	*	4	*	2	4
Firm and Cand Covariates	Full	Select	Full	Select	Select
Region,Sector Year FE	Full	Full	No	No	No
Observations	481	520	201	201	520

Omitting all covariates is not a good solution

Covariates increase precision of treatment effect estimates. [2]

Work on covariate adjusted treatment effects w/ regularization provides a way forward

- Bloniarz et. al (2016) "Lasso adjustments of treatment effect estimates in randomized experiments."
- Wager et al. (2016) "High-dimensional regression adjustments in randomized experiments."
- LASSO adjustment of treatment effects as a principled means of including covariates in experimental research.
- Argue for LASSO pre–processing + OLS.

Solution: incorporate LASSO into RDD treatment effect estimation process

- Get benefits of including covariates while minimizing costs:
- Algorithmic selection of *final* covariates $\rightarrow$ reduced "researcher degrees of freedom".
- LASSO covariate selection increases LATE precision via MSE minimization.
-

RDD estimation algorithm w/ adaptive LASSO.

Step 1	Researcher pre-treatment covariate selection	Covariates selected by the researcher on the basis of substantive concerns.
Step 2	Adaptive lasso regularization	Model from **Step 1** estimated using weighted adaptive LASSO (Zou 2006)
Step 3	Covariate adjustment	Covariates, higher-order terms whose coefficients are shrunk to 0 are excluded from the final model.
		Weights designed s.t. **treatment effect, forcing variable & variables in kernel are NOT penalized.**
Step 4	CCT robust estimation of final model	The modified model from Step 3 is estimated via the CCT robust procedure [1].

Remainder of talk

(1) Briefly discuss each stage of the estimation process.

(2) Applied example using replication of close election RDD (Szakonyi 2018, APSR).

(3) Simulation results: bias, % coverage, MSE.

Background: treatment effect estimation for RDDs w/ covariates

Local average treatment effect (LATE) estimate in potential outcomes framework:

$$LATE = \tau = \lim_{F_i \downarrow 0} E[Y(1)_i | F_i = f + \epsilon] - \lim_{F_i \uparrow 0} E[Y(0)_i | F_i = f - \epsilon]$$

- $Y(1)_i$: Outcome of treated unit i.
- $Y(0)_i$: Outcome of control unit i.
- F_i: Forcing variable.

Background: local linear regression estimation of LATE w/ covariates

$$\hat{Y}_i = \beta_0 + \hat{\tau} T_i + \delta(F_i \cdot T_i) + X\beta$$

- T_i: Treatment dummy s.t. $\mathbb{I}(T_i > 0)$.
- $X\beta$: Pre-treatment covariates.
- F_i: Forcing variable.

* Inclusion of covariates increases precision (Calonico, Cattaneo, Farrell & Titunik, 2019).

Principled RDD estimation algorithm

Step 1: Researcher pre-treatment covariate selection

$$Y_i = \alpha + \tau T_i + \gamma F_i + \delta(F_i \cdot T_i) + X\beta + \epsilon_i$$

Considerations:

- Availability of pre-treatment covariates eg) fixed effects, demographics, etc.
- Are they likely predictive of the outcome?
- Sample size.

Step 2: Adaptive LASSO regularization

$X^s \subseteq X$

Choose an X^{s*} that minimizes the mean squared error (MSE)

$$\arg\min_{\Theta} \sum_{i=1}^{N} (Y_i - [\alpha + \tau T_i + \gamma F_i + \delta(F_i \cdot T_i) + X^{s*}\beta])^2$$

$$\Theta = (\tau, \gamma, \delta, \beta)$$

Ordinary L_1 LASSO regularization

$$\arg\min_{\Theta} \sum_{i=1}^{N} \left[Y_i - (\alpha + \tau T_i + \gamma F_i + \delta(F_i \cdot T_i) + X\beta)\right]^2 + \lambda \left[\sum_{j=3}^{p} |\beta_j|\right]$$

- λ tuned via automated 10-fold cross validation.

Adaptive L_1 LASSO regularization

- Ordinary LASSO inconsistently selects models.
- Adaptive LASSO has **oracle properties** (correct, consistent model selection) (Zou, 2006).
- Easy to incorporate 0 penalty weights for RDD.

Adaptive L_1 LASSO regularization

$$\underset{\Theta}{\arg\min} \sum_{i=1}^{N} [Y_i - (\alpha + \tau T_i + \gamma F_i + \delta(F_i \cdot T_i) + X\beta)]^2 + \lambda \left[\sum_{j=3}^{p} \omega_j |\beta_j| \right]$$

$$\omega_j = \frac{1}{|\beta_j|^\gamma}$$

- λ and γ tuned via automated 10-fold cross validation process.

Step 3: Automated model selection

- Choose "final" model with covariates by excluding those shrunk to zero by the adaptive LASSO.

$X^o \subseteq X$ is the truncated set of covariates selected out by the adaptive lasso described above.

$$\mathbb{E}(Y_i | T_i, F_i, X^o) = \alpha + \tau T_i + \gamma F_i + \delta(F_i \cdot T_i) + X^o \beta$$

Step 3: Automated model selection - bandwidth selection

- **When optimal bandwidth is used** – automated model selection *before* selecting optimal bandwidth.
- Optimal bandwidth algorithms (e.g. Imbens-Kalyanaraman) use model MSE as bandwidth selection criteria.
- **When fixed bandwidth is used (e.g. 5% guideline)** – automated model selection *after* bandwidth selection.

Step 4: Regularized CCT Robust Estimation

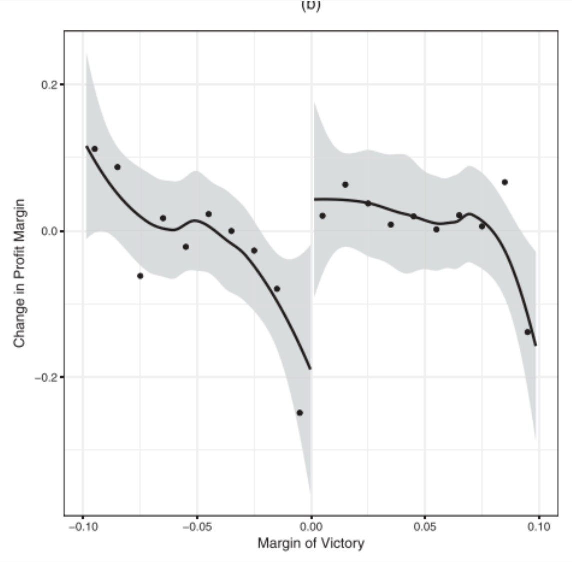
- Doubly robust estimation combining regularized model with Calonico, Cattaneo, Titiunik (CCT) robust estimation.

Applied example using replication of close election RDD (Szakonyi 2018, APSR)

Empirical Illustration: Do Firms Profit from Having Elected Board Members?

- Szakonyi (2018) uses close election RDD to explore whether office-holding affects profits of firms whose board members held political office in Russia.
- Results replicated using adaptive LASSO process.

Finds evidence of big returns to office...



General form of RDD LLR estimated by Szakonyi (2018)

$$
\begin{aligned}
\text{Firm Profits} &= \alpha + \hat{\tau}(\text{District Win}) + \gamma\text{Vote Margin} \\
&+ \delta(\text{District Win} \times \text{Vote Margin}) + X\beta \\
&+ \mathsf{Y}_j + \mathsf{S}_j + \mathsf{R}_j
\end{aligned}
$$

- **Treatment**: District win.
- **Forcing variable**: Vote margin.
- **Covariates**: X, Y, S, R: candidate covariates, state, region and district fixed effects.

Replication of “Political Connections and Firm Profitability” in Szakonyi (2018) with Adaptive LASSO Adjusted Treatment Effects.

	Original (APSR)	Adaptive	Original 5% (APSR)	Adaptive 5%	Adaptive CCT Robust
District Win	0.146***	0.102*	0.198**	0.097**	0.140***
	(0.065)	(0.060)	(0.090)	(0.038)	(0.052)
Bandwidth	0.113	0.120	0.050	0.050	0.120
Covariates Dropped	*	4	*	2	4
Firm and Cand Covariates	Full	Select	Full	Select	Select
Region,Sector Year FE	Full	Full	No	No	No
Observations	481	520	201	201	520

Simulation results

Simulations

- Realistic simulations using parameters from election and profit data from Szakonyi (2018).
- True simulated treatment effect set to a known value τ_{RDD}.
- 2,000 simulated data sets w/ the same covariance structure and mean of the original dataset.

Simulations: models estimated

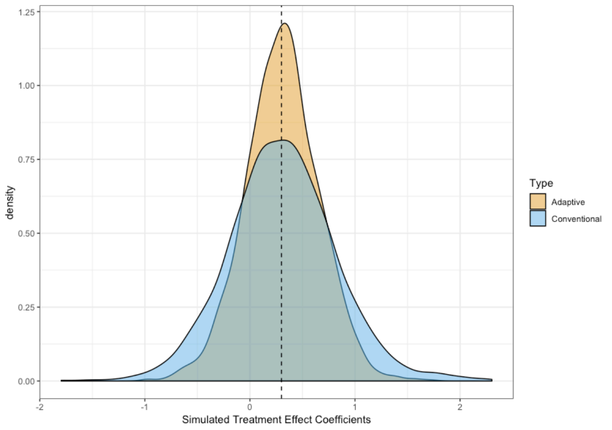
For each simulation $s = 1, \ldots, 2000$. The true model is:

$$Y_s = 0.3(District\ Win_s) + \gamma(Margin_s) + \delta(District\ Win_s \times Margin_s) + \eta_s$$

The estimated model with covariates is:

$$\mathrm{Y}_s = \alpha^s + \hat{\tau}^s_{RDD}(\text{District Win}_s) + \gamma^s \text{Margin}_s + \\ \delta^s(\text{District Win}_s \times \text{Margin}_s) + X^s\beta^s + \epsilon^s$$

Results: treatment effect point estimates

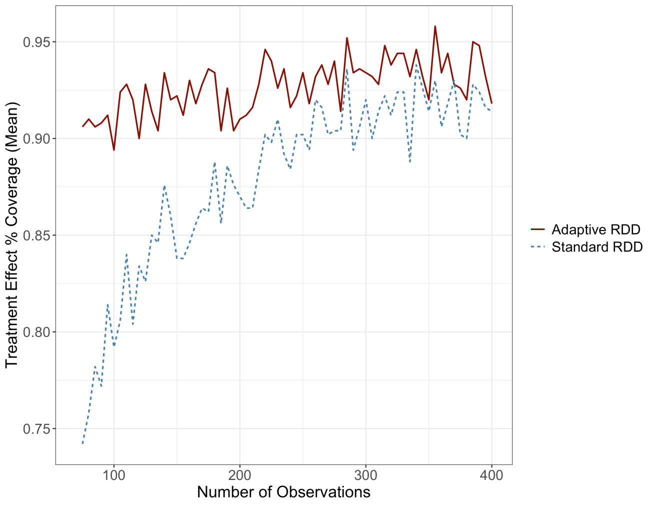


Distribution of simulated treatment effects $\hat{\tau}^{s}_{RDD}$, for adaptive lasso adjusted treatment effects and conventional treatment effects across 2,000 simulated data sets with variable bandwidth select. The true $\tau_{RDD} = 0.30$ is denoted by the black dotted line.

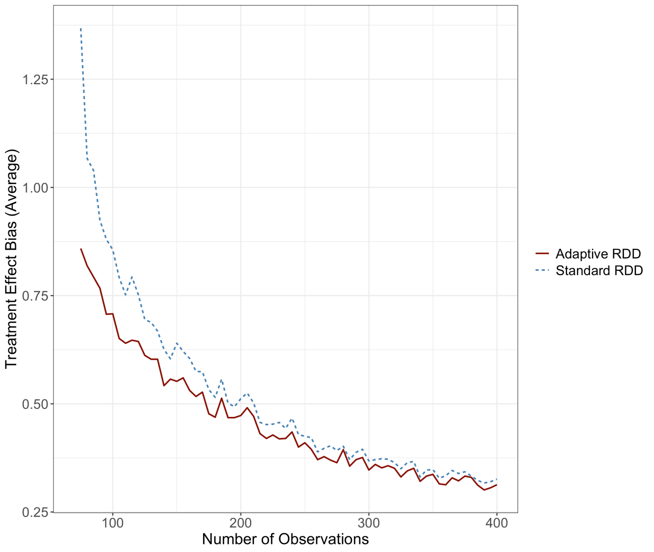
Results: bias and % coverage ($N = 500$)

	Variable Bandwidth*		
	Adaptive	*Conventional*	*Difference (Adaptive - Conventional)*
τ_{RDD} Bias	0.274	0.397	- 0.123***
% Coverage	0.944	0.699	+ 0.245***
τ_{RDD} Estimate	0.308	0.308	-
Bandwidth	0.38	0.292	+ 0.088***
	Fixed Bandwidth∓		
	Adaptive	*Conventional*	*Difference (Adaptive - Conventional)*
τ_{RDD} Bias	0.375	0.375	- 0.001
% Coverage	0.931	0.796	+ 0.135***
τ_{RDD} Estimate	0.300	0.300	- 0.001
Bandwidth	0.200	0.200	-

Treatment Effect % Coverage by Sample Size



Treatment Effect Bias by Sample Size



Conclusions

- Including covariates increases precision of LATE for RDDs but can be problematic for under-powered/low N estimation.
- Adaptive LASSO regularization with CCT robust estimation, provides a doubly robust means of gaining precision from covariates while reducing researcher degrees of freedom.

References

[1] Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

[2] Sebastian Calonico, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik. Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451, 2019.

Extra Slides

Simulations: details

$$?^{S} \sim \mathcal{N}(\mu, \Sigma)$$

- Ξ: matrix which contains the set of covariates plus the vote margin in Szakonyi (2018).
- DGP of Ξ is MVN distribution w/ $\mu = (\mu_1, \mu_2, \cdots, \mu_p)$ & covariance matrix Σ estimated from data.
- 2000 datasets estimated from this data generating process.