

The Political Methodologist

NEWSLETTER OF THE POLITICAL METHODOLOGY SECTION
AMERICAN POLITICAL SCIENCE ASSOCIATION
VOLUME 22, NUMBER 2, SPRING 2015

Editor:

JUSTIN ESAREY, RICE UNIVERSITY
justin@justinesarey.com

Editorial Assistant:

AHRA WU, RICE UNIVERSITY
ahra.wu@rice.edu

Associate Editors:

RANDOLPH T. STEVENSON, RICE UNIVERSITY
stevenson@rice.edu

RICK K. WILSON, RICE UNIVERSITY
rkw@rice.edu

Contents

Notes from the Editors	1
Articles	2
Colin Elman and Arthur Lupia: Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors	2
Christopher Gandrud: Corrections and Refinements to the Database of Political Institutions' yrcurnt Election Timing Variable . . .	2
Natália S. Bueno and Guadalupe Tuñón: Graphical Presentation of Regression Discontinuity Results	4
Thomas B. Pepinsky: Why Can't We Just Make Up Instrumental Variables?	8
Announcement	11
Justin Esarey: Call for Papers: <i>The Political Methodologist</i> Special Issue on Peer Review . .	11

Notes From the Editors

The editors are proud to present the Spring 2015 print edition of *The Political Methodologist*!

This issue starts off with the Data Access and Research Transparency (DA-RT) statement that has been endorsed by a large number of leading journals in political science. *The Political Methodologist* is proud to be a part of this movement, which we joined in 2014 with a statement by

Associate Editor Rick Wilson. Colin Elman and Arthur Lupia asked journals that have signed on to the statement to simultaneously print it as a strong signal to the discipline that reproducible and transparent research is a priority for political science and must be a part of the practice of researchers working today. Christopher Gandrud's contribution, correcting and refining aspects of an existing data set, underscores how making data and replication code freely available enables the continuous improvement of research.

We also have two papers about the use of causal inference methods in quantitative political research. Natália Bueno and Guadalupe Tuñón write about why presenting graphical robustness checks for regression discontinuity results should be a part of our practice, and helpfully provide examples with R code to facilitate the wider use of this technique. Tom Pepinsky uses a provocative question—"why can't we just make up instrumental variables?"—to make an instructive point about what makes an instrumental variable valid and how we can be clearer in our research and pedagogy about how we choose these variables.

Finally, an update on *TPM*'s circulation. I am pleased to report that our on-line platform continues to be a very successful supplement to this print format: between January and June of this year, our on-line website has received 18,807 unique visitors. I'm hopeful that our upcoming special issue on peer review (note the Call for Papers noted in this issue) will attract even more interest in *TPM* and in our shared project of improving the practice of political methodology.

more valid indicator for research on how election timing affects policymaking:

Country	Refinements to make to the DPI <code>yrcurnt</code> variable for the EU-28
Austria	Use parliamentary rather than (figurehead) presidential elections.
Lithuania	Use parliamentary rather than presidential elections. Lithuania is a semi-presidential system where the president appoints the PM, the legislature's approval is needed. The PM is more responsible for domestic policy.
Romania	Romania is semi-presidential where the president appoints the PM, but the PM must be approved by the parliament. The PM is both the head of government and sets the legislative agenda. Before 2008, presidential and parliamentary elections occurred in the same year. Since then they have diverged, from which point the parliamentary elections should be used.
Slovenia	Use parliamentary rather than (figurehead) presidential elections.

Graphical Presentation of Regression Discontinuity Results

Natália S. Bueno

Yale University

natalia.bueno@yale.edu

Guadalupe Tuñón

University of California, Berkeley

guadalupe.tunon@berkeley.edu

During the last decade, an increasing number of political scientists have turned to regression-discontinuity (RD) designs to estimate causal effects. Although the growth of RD designs has stimulated a wide discussion about RD assumptions and estimation strategies, there is no single shared approach to guide empirical applications. One of the major issues in RD designs involves selection of the “window” or

A data set that implements these corrections and refinements for the EU-28 countries can be found on GitHub.⁶ The data (including both the original and corrected versions of the variable) can be downloaded directly into R using the `repmis` package (Gandrud 2015) with the following code:

```
yrcurnt_corrected <- repmis::source_data('http://bit.ly/1EM8EVE')
```

This file has also updated election timing data through 2013.

References

- Alt, James, David Dreyer Lassen, and Joachim Wehner. 2014. “It Isn’t Just About Greece: Domestic Politics, Transparency and Fiscal Gimmickry in Europe.” *British Journal of Political Science* 44 (04): 707-16.
- Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh. 2001. “New Tools in Comparative Political Economy: The Database of Political Institutions.” *World Bank Economic Review* 15 (1): 165-76.
- Gandrud, Christopher. 2015. *repmis: Miscellaneous Tools for Reproducible Research with R*. <http://cran.r-project.org/web/packages/repmis/>.

“bandwidth” – the values of the running variable that define the set of units included in the RD study group.¹

This choice is key for RD designs, as results are often sensitive to bandwidth size. Indeed, even those who propose particular methods to choose a given window agree that “irrespective of the manner in which the bandwidth is chosen, one should always investigate the sensitivity of the inferences to this choice. [...] [I]f the results are critically dependent on a particular bandwidth choice, they are clearly less credible than if they are robust to such variation in bandwidths” (Imbens and Lemieux, 2008, p. 633). Moreover, the existence of multiple methods to justify a given choice opens the door to “fishing” – the intentional or unintentional selection of models that yield positive findings (Humphreys, Schanchez de la Sierra, and van der Windt, 2013).

In this note, we propose a simple graphical way of report-

⁶See https://github.com/christophergandrud/yrcurnt_corrected/tree/master/data.

¹Formally, the window is an interval on the running variable, $W_0 = [r, \bar{r}]$, containing the cutoff value r_0 . The analysis then focuses on all observations with R_i within this interval. Scholars have developed numerous tools to determine the right window for a given application and estimator (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2011; Calonico, Cattaneo and Titiunik, 2015).

²This choice is posterior to defining the estimand and choosing an estimator for the causal effect of treatment. While the plots we propose focus

ing RD results that shows the sensitivity of estimates to a wide range of possible bandwidths.² By forcing researchers to present results for an extensive set of possible choices, the use of these plots reduces the opportunities for fishing, complementing existing good practices in the way in which RD results are presented. Some empirical applications of RD designs have presented their results using plots that are in some ways similar to the ones we propose. However, this note explores the virtues of the plots more systematically (e.g., in connection with balance tests and the discussion of the RD estimator) and provides code so that scholars can adapt them to their own applications. The following paragraphs describe how RD results are usually reported in two top political science journals and the ways in which the graphs we propose can improve on current practices. An R function to construct these plots for a wide set of applications is publicly available on the online Appendix.

Reporting RD Analyses: Current Practice

How do political scientists report the results of regression-discontinuity designs? We reviewed all papers using RDDs published in the *American Political Science Review* and *American Journal of Political Science*. We surveyed these papers and coded (1) their choice of estimators, (2) whether they present any type of balance test and, (3) if they do, the window(s) chosen for this.

Out of a total of twelve RD papers published in these journals, five report results using a single estimator.³ Four articles present results for a single window – usually the full sample. The remaining papers present results using multiple windows, but the number and selection of windows are neither systematic nor extensive.⁴ Seven papers present some type of balance test, but while researchers often report their main results using a handful of windows, they do not report balance tests for different windows to the same extent.⁵

A Graphical Alternative: An Example

We use electoral data from the U.S. House of Representatives from 1942 to 2008 collected by Caughey and Sekhon (2011) and a regression discontinuity design examining incumbency advantage to illustrate how a researcher can use plots to present RD results in a transparent and systematic

way. This application has two advantages for illustrating the use of these plots. First, close-race electoral regression-discontinuity designs are one of the main applications of this type of design in political science – we are thus presenting the plot in one of the most used RDD setups, although researchers can use this type of graph in all sorts of RD designs. Second, the use of close-race RDDs to measure the effect of incumbency advantage has sparked a vigorous debate about the assumptions and validity of these designs in particular settings. Using this application allows us to show an additional advantage of the plots we propose: tests of balance and other types of placebo tests.

Figure 1 plots the estimates of local average treatment effects as a function of the running variable, here vote margin. For example, the first solid black circle represents the average difference in vote share between an incumbent party that won by 0.45% or less and an incumbent party that lost by 0.45% or less is about 11 percentage points. It also reports the average difference in vote share between incumbent parties that won and lost by sequential increases of 0.2% in vote margin between 0.45% to 9.85%. Figure 1 has an additional feature: the solid gray line represents the results using an additional estimator that enables us to compare the effects estimated from two different estimators, across different windows, in the same plot. In this case, we present the estimated effects of party incumbency on vote share using a local linear regression with a triangular kernel, across different bandwidths. Researchers could use different estimators, such as a polynomial regression model.⁶ Note, however, that the black circles and the solid gray line represent different quantities. The black circles represent estimates of the average causal effect for the RD study group N . The difference of means estimator ($Y^T - Y^C$) – the difference between average vote share for an incumbent party minus the average vote share for a non-incumbent party – is unbiased for the average causal effect (τ_{ACE}), represented in equation (1).⁷ The gray line presents estimates of the average causal effect precisely at the point of discontinuity (c). We fit a local linear regression with a triangular kernel, within the RDD bandwidth, to estimate this limit parameter (τ_{lim}), represented in equation (2).⁸

$$\tau_{ACE} = \frac{1}{N} \sum_{i=1}^N [Y_i(1) - Y_i(0)]. \quad (1)$$

on presenting the sensitivity of results to bandwidth choice, we also show how they can be used to explore the sensitivity of results to these other choices. For discussions of estimands in RD design see Dunning (2012) and Calonico et al. (2015).

³Polynomial regression is the most popular model: nine papers use a type of polynomial regression, five employ a local linear regression, and three use a difference of means (via OLS). Only one presents results using all three estimators.

⁴Gerber and Hopkins (2011), Ferwerda and Miller (2014), and Eggers et al. (2015) are exceptions—they show the robustness of the main result using plots similar to the one we suggest here.

⁵All of the papers that use local linear regressions also use a type of standard procedure to choose the “optimal” bandwidth – either Imbens and Lemieux (2008) or Imbens and Kalyanaraman (2011).

⁶We present an example of this plot using a fourth-degree polynomial regression in Figure A.1 available at the online Appendix.

⁷See Dunning (2012) and Bueno, Dunning, and Tuñón (2014) for a discussion and proofs.

⁸We use the terms “window” and “bandwidth” interchangeably, since both denote the values of the running variable (r) that define the set of units included in the RD study group. However, in local linear regression with kernel smoothers, bandwidth refers to the width of the kernel.

$$\tau_{lim} = \lim_{r \downarrow c} [\bar{Y}_i(1)|R_i = r] - \lim_{r \uparrow c} [\bar{Y}_i(0)|R_i = r] \quad (2)$$

The key part of the plotting function is the section that produces the estimate for each window. For this, we first pre-specify functions for each estimator. For example, for the difference of means we have:⁹

```
#Difference of mean (using OLS with robust standard errors)
dom <- function(rescaled, treat, outcome){
  model <- lm(outcome~treat)
  est <- NA
  est[1] <- model$coefficients[2]
  est[2] <- sqrt(diag(vcovHC(model,type="HC3"))[2])
  return(est)
}
```

We then include a loop which takes a window value and subsets the data to keep only the observations within that window.

```
ests <- matrix(NA,length(windows),3) #object to store loop output
for (i in 1:length(windows)){
  # select data
  temp <- as.data.frame(data[abs_running<=windows[i],])
```

We take this subset of the data to calculate the estimates of interest for that window, here the difference of means. The plot requires that we calculate both the point estimates and the confidence intervals. In these figures, confidence intervals are calculated using a normal approximation and unequal variances are assumed for standard errors in the treatment and control groups. If the researcher wanted to include an additional estimator, the calculation of the estimate for a particular window would also be included in the loop.¹⁰

```
ests[i,1:2] <- with(temp, dom(rescaled=rescaled, treat=treat,
  outcome=outcome))

if (ci=="95%") CI <- cbind(ests[,1]+1.96*ests[,2],ests[,1]-1.96*
  ests[,2])
if (ci=="90%") CI <- cbind(ests[,1]+1.64*ests[,2],ests[,1]-1.64*
  ests[,2])
```

As expected, the confidence intervals in Figure 1 become increasingly smaller for results associated with larger vote margins because the number of observations is larger. This increase in the number of observations associated with larger windows can also be reported in the plot, which we do in the upper axis of Figure 1. To include the number of observations as a function of window size, we order the observed values of the outcome of interest according to their value for the running variable and allow the user to set the different number of observations that she would want to show in the plot. We calculate the number of observations at the specified values of the running variable – then, we add the number of observations to an upper axis in the plot.

```
# as an argument in the function, the user defines nr_obs_lab,
# the labels for the number of observations she would like to
```

⁹Note that we compute the difference of means by regressing the outcome variable on a dummy for treatment assignment, with robust standard errors allowing for unequal variances, which is algebraically equivalent to the *t*-test with unequal variances.

¹⁰Our plot, and the accompanying documented R code, is flexible to incorporating different ways of estimating standard errors and constructing confidence intervals. For a detailed discussion of standard errors and confidence intervals in RD designs, see Calonico et al. (2015).

¹¹In these plots, we chose to omit the axis with the number of observations because even though there are different rates of missing observations for covariates, the number of observations for the windows we were mostly interested in did not vary substantially from those in Figure 1.

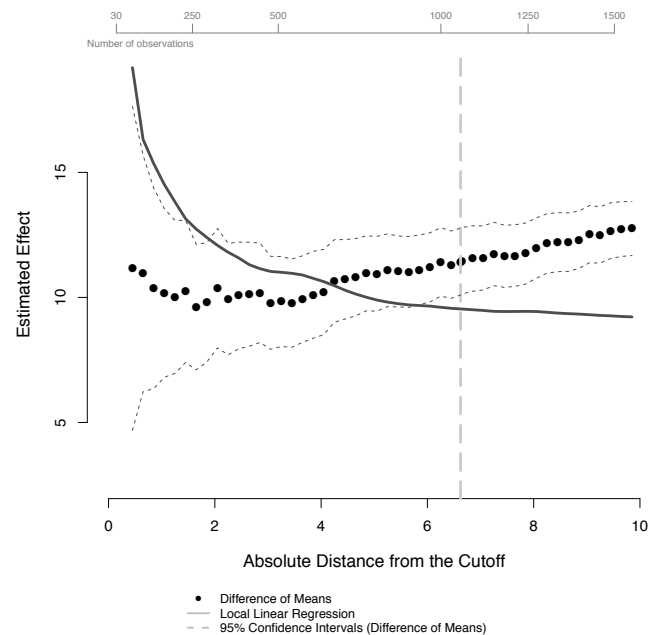
```
# include in the plot

# ordering the data by the values for the running variable
data <- as.data.frame(data[order(abs_running),])

if (nr_obs==T) {
  # binding the labels with the corresponding value for the running
  variable
  nr_obs_lab <- cbind(nr_obs_lab, data$abs_running[nr_obs_lab])
}

# Finally, we include an additional axis in the plot
axis(3, at=c(nr_obs_lab[,2]), labels=c(nr_obs_lab[,1]), cex=.6,
  col="grey50", lwd = 0.5, padj=1, line=1, cex.axis=.7,
  col.axis="grey50")
mtext("Number_of_observations", side=3, col="grey50", cex=.7, adj
  =0)
```

Figure 1: Mean vote share difference between winners and losers by Democratic margin of victory in previous election, U.S. House of Representatives from 1942 to 2008.



Note: Dashed gray line at the optimal bandwidth estimated by the method of Imbens and Kalyanaraman (2011). Difference of means is the average difference in vote share for an incumbent party minus the average vote share for a non-incumbent party. The local linear regression regression uses a triangular kernel.

Figure 2 follows the same logic described for Figure 1 but reports balance tests: each plot shows the effects of incumbency on a pre-treatment covariate as a function of the run-

ning variable.¹¹ These plots allow for an extensive examination of the sensitivity of balance to different windows, reporting the magnitude of the difference between treatment and control and its confidence interval. For a given identifying assumption (such as continuity of potential outcomes or as-if random assignment near the threshold), Figures 1 and 2 help the reader to evaluate whether or not – or for which window – these assumptions are plausible.

For instance, panel (a) in Figure 2, reports the difference in previous Democratic victories between the incumbent and non-incumbent party and shows a large imbalance, strikingly larger for smaller windows – a point made by Caughey and Sekhon (2011). Panel (b) in Figure 2 shows the difference in voter turnout between treatment and control districts. For the entire set of windows covered by the plot the difference is never statistically different from zero, suggesting that the groups are balanced in terms of this covariate. Note that the size of the difference between treatment and control is much smaller in panel (b) than in panel (a) – also, relatively to the size of the effect of incumbency, the imbalance in panel (a) is substantial.¹²

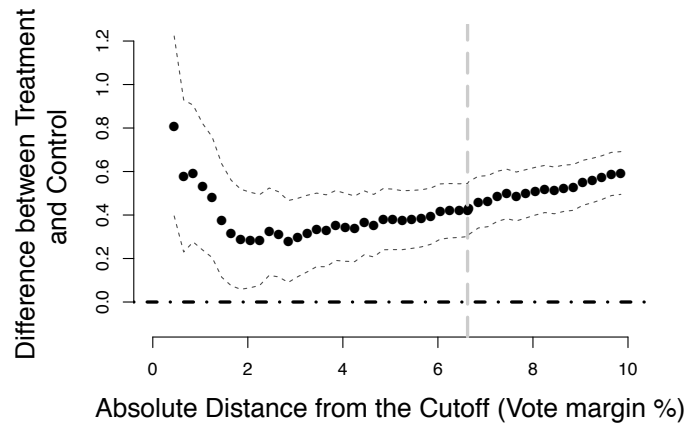
For ease of presentation, here we present plots for only two pre-treatment covariates. However, analysts should be encouraged to present plots for all pre-treatment covariates at their disposal.¹³ Some plots may be more informative than others, for instance, because some pre-treatment covariates are expected to have a stronger prognostic relationship to the outcome. However, presentation of balance plots for the full set of available pre-treatment covariates may reduce opportunities for intentional or unintentional fishing.

Concluding Remarks

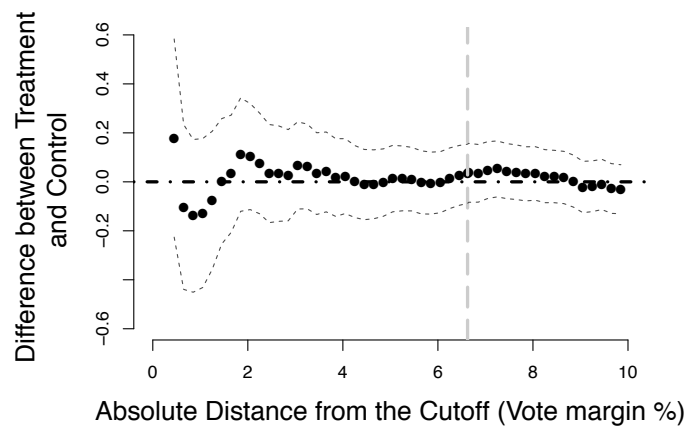
We believe that these simple plots are a useful complement to the standard way in which scholars report results and balance tests from regression-discontinuity designs. They provide a simple visual comparison of how different estimators perform as a function of different windows, communicating the robustness of the results. Moreover, using these plots both for analysis of effects and placebo tests enables an informal visual inspection of how important confounders may be, relative to the size of the effect – this is particularly informative when researchers use pre-treatment values of the outcome variable as a placebo test. However, researchers may also compare the size of treatment effects relative to other placebo tests by using standardized effect sizes across different windows, so that the scale of all plots is comparable. In summary, these plots improve the communication of findings from regression-discontinuity designs by showing readers the results from an extensive set of bandwidths, thus reducing researchers’ latitude in presentation of their

main findings and increasing the transparency of RD design applications.

Figure 2: Tests for balance: Standardized difference of means of pre-treatment covariates by Democratic margin of victory (95% confidence intervals), U.S. House of Representatives, from 1942 to 2008.



(a) Democratic Win in $t - 1$



(b) Voter Turnout %

Note: Dashed gray line at the optimal bandwidth estimated by the method of Imbens and Kalyanaraman (2011). Difference of means is the average difference in vote share for an incumbent party minus the average vote share for a non-incumbent party.

References

Bueno, Natália S., Thad Dunning, and Guadalupe Tuñón. 2014. “Design-Based Analysis of Regression Discontinuities: Evidence From Experimental Benchmarks. *Paper Presented at the 2014 APSA Annual Meeting*.

¹²See Figure A.2 in the online Appendix for the standardized effect of incumbency on vote share.

¹³An extensive set of balance plots for pre-treatment covariates in Caughey and Sekhon (2011) can be found in Figure A.3 of the online Appendix.

- Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2015. "Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs." *Econometrica* 86 (2): 2295-326.
- Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942-2008." *Political Analysis* 19 (4): 385-408.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: a Design-Based Approach*. New York: Cambridge University Press.
- Eggers, Andrew C., Olle Folke, Anthony Fowler, Jens Hainmueller, Andrew B. Hall and James M. Snyder. 2015. "On the Validity of the Regression Discontinuity Design For Estimating Electoral Effects: New Evidence From Over 40,000 Close Races." *American Journal of Political Science* 59 (1): 259-74.
- Ferwerda, Jeremy and Nicholas L. Miller. 2014. "Political Devolution and Resistance to Foreign Rule: A Natural Experiment." *American Political Science Review* 108 (3): 642-60.
- Gerber, Elisabeth R. and Daniel J. Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55 (2): 326-39.
- Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Non-binding Research Registration." *Political Analysis* 21 (1): 1-20.
- Imbens, Guido and Karthik Kalyanaraman. 2011. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3):933-59.
- Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615-35.

Why Can't We Just Make Up Instrumental Variables?

Thomas B. Pepinsky
Cornell University
pepinsky@cornell.edu

Introduction

Valid instrumental variables are hard to find. The key identification assumption that an instrumental variable Z be correlated with an endogenous variable X , but not causally related to Y , is difficult to justify in most empirical applications. Even worse, the latter of these requirements must be defended theoretically, it is untestable empirically. Recent concerns about "overfishing" of instrumental variables (Morck and Yeung 2011) underscore the severity of the problem. Every new argument that an instrumental variable is valid in one research context undermines the validity of that instrumental variable in other research contexts.

Given such challenges, it might be useful to consider more creative sources of instrumental variables that do not depend on any theoretical assumptions about how Z affects X or Y . Given an endogenous variable X , in a regression $Y = \beta X + u$, why not simply generate random variables until we discover one that is, by happenstance, correlated with X , and then use that variable Z_{random} as an instrument? Because Z_{random} is generated by the researcher, we can be absolutely certain that there is no theoretical channel through which it causes Y . And because we have deliber-

ately selected Z_{random} so that it is highly correlated with X , weak instrument problems should not exist. The result should be an instrument that is both valid and relevant. Right?

If it were only so easy. In Figure 1 below I show what would happen if we used random instruments, generated in the manner just described, for identification purposes. The data generating process, with notation following Sovey and Green (2011), is

$$Y = \beta X + \lambda Q + u \quad (1)$$

$$X = \gamma Z + \delta Q + e \quad (2)$$

X is the endogenous variable, Q is an exogenous control variable, and Z is an instrumental variable. Endogeneity arises from correlation between u and e , which are distributed multivariate standard normal with correlation $\rho = .5$. For the simulations, $\beta = 5$, with $\lambda = \gamma = \delta = 1$.

Figure 1 shows the distribution of estimates \hat{b} from 100 simulations of the data-generating process above. In each simulation, I generate each variable in Equations 1 and 2, and then repeatedly draw random variables until I happen upon one that is correlated with the randomly generated X at the 99.9% confidence level. I use that variable Z_{random} as an instrument to estimate $\hat{b}_{IV(Fake)}$, which I compare to the naive OLS estimates \hat{b}_{OLS} and estimates using the real instrument Z from Equation 2, \hat{b}_{IV} .

Clearly, OLS estimates are biased. But more importantly, the "fake" IV estimates are biased and inefficient, while "true" IV estimates using Z are accurate. What has happened to the fake IVs?