DiD, IV, RDD implementation

Mauricio Romero
(Based on the examples in "The effect" by Nick C. Huntington-Klein)

Difference-in-Difference

Don't Take 'No' For An Answer: An Experiment With Actual Organ Donor Registrations
Judd B. Kessler and Alvin E. Roth
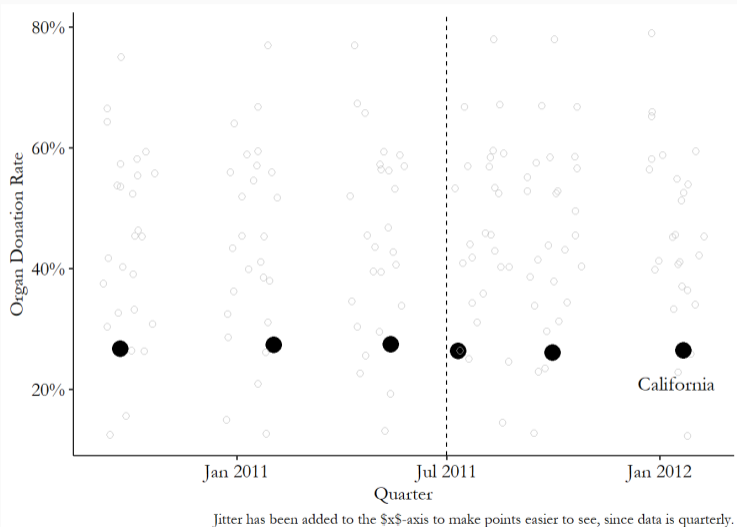NBER Working Paper No. 20378
August 2014
JEL No. D02,D04,D47

ABSTRACT

Over 10,000 people in the U.S. die each year while waiting for an organ. Attempts to increase organ transplantation have focused on changing the registration question from an opt-in frame to an active choice frame. We analyze this change in California and show it decreased registration rates. Similarly, a "field in the lab" experiment run on actual organ donor registration decisions finds no increase in registrations resulting from an active choice frame. In addition, individuals are more likely to support donating the organs of a deceased who did not opt-in than one who said "no" in an active choice frame.
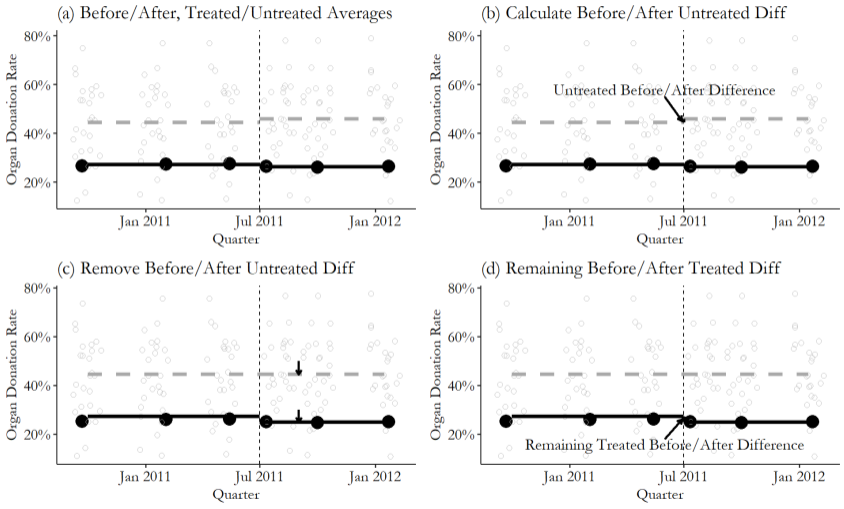
Active choice

- In the US, people not signed up to be organ donors by default

- When you get a driver's license, you can choose to "opt in" to be an organ donor

- In July 2011, the state of California switched from "opt-in" to "active choice"

  - Need to activly choose "no" to "opt out"

- What is the effect of this policy?

Raw data

Jitter has been added to the x-axis to make points easier to see, since data is quarterly.

Raw data

(a) Before/After, Treated/Untreated Averages

(b) Calculate Before/After Untreated Diff
Untreated Before/After Difference

(c) Remove Before/After Untreated Diff

(d) Remaining Before/After Treated Diff
Remaining Treated Before/After Difference

Implementation

```
library(tidyverse)
library(modelsummary)
library(fixest)
library(causaldata)
od <- causaldata::organ_donations

# Treatment variable
od <- od %>%
  mutate(Treated = State == 'California' &
           Quarter %in% c('Q32011','Q42011','Q12012'))

# feols clusters by the first
# fixed effect by default, no adjustment necessary
clfe <- feols(Rate ~ Treated | State + Quarter,
              data = od)
msummary(clfe, stars = c('*' = .1, '**' = .05, '***' = .01),output="latex")
```

	(1)
Treated	-0.022***
	(0.006)
Num.Obs.	162
R2	0.979
Std.Errors	by: State
FE: State	X
FE: Quarter	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Placebo

```
od <- causaldata::organ_donations %>%
  # Use only pre-treatment data
  filter(Quarter_Num <= 3)
# Create our fake treatment variables
od <- od %>%
  mutate(FakeTreat1 = State == 'California' &
           Quarter %in% c('Q12011','Q22011'),
         FakeTreat2 = State == 'California' &
           Quarter == 'Q22011')

# Run the same model we did before but with our fake treatment
clfe1 <- feols(Rate ~ FakeTreat1 | State + Quarter,
               data = od)
clfe2 <- feols(Rate ~ FakeTreat2 | State + Quarter,
               data = od)
```

|	(1)	(2)
FakeTreat1	0.006	
	(0.005)	
FakeTreat2		-0.002
		(0.003)
Num.Obs.	81	81
R2	0.994	0.994
Std.Errors	by: State	by: State
FE: State	X	X
FE: Quarter	X	X

Placebo

```r
# Treatment variable
od <- od %>% mutate(California = State == 'California')

# Interact quarter with being in the treated group using
# the fixest i() function, which also lets us specify
# a reference period (using the numeric version of Quarter)
clfe <- feols(Rate ~ i(Quarter_Num, California, ref = 3) |
                State + Quarter_Num, data = od)

# And use iplot() for a graph of effects
iplot(clfe, bty="L", las=2, cex=1.5, cex.lab=1.5, lwd=2,
        grid=F, zero=T, xlab="Time to treatment",
     x=as.character(c(-3,-2,-1,0,1,2)))
```

Effect on Rate
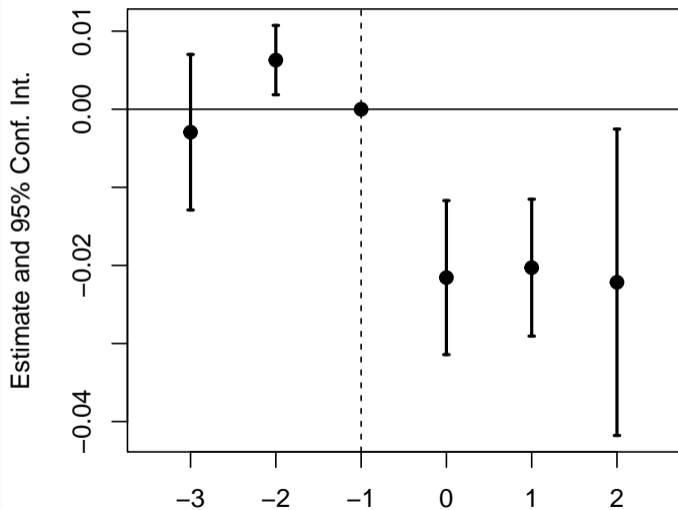
Difference-in-Difference: Staggered roll out

JOURNAL ARTICLE

Bargaining in the Shadow of the Law: Divorce Laws and Family Distress *

Betsey Stevenson, Justin Wolfers

📄 PDF   ▌▌ Split View   66 Cite   🔑 Permissions   ⋘ Share ▾

Abstract

This paper exploits the variation occurring from the different timing of divorce law reforms across the United States to evaluate how unilateral divorce changed family violence and whether the option provided by unilateral divorce reduced suicide and spousal homicide. Unilateral divorce both potentially increases the likelihood that a domestic violence relationship ends and acts to transfer bargaining power toward the abused, thereby potentially stopping the abuse in extant relationships. In states that introduced unilateral divorce we find a 8−16 percent decline in female suicide, roughly a 30 percent decline in domestic

Unilateral divorce

- Panel from 1964 through 1996 of the US

- No-fault divorce reforms and female suicide rates

- Treatment (i.e. when the law went into effect in each state) is staggered

- Some states never received treatment

Implementation

```r
library(data.table) ## For some minor data wrangling
library(fixest)     ## NB: Requires version >=0.9.0
# Load and prepare data
dat = fread("https://raw.githubusercontent.com/LOST-STATS/LOST-STATS.github
# Let's create a more user-friendly indicator of which states received treat
dat[, treat := ifelse(is.na(`_nfd`), 0, 1)]
```

Implementation

```
# Create a "time_to_treatment" variable for each state
#so that treatment is  relative for all treated units.
#For the never-treated (i.e. control) units,
# we'll arbitrarily set the "time_to_treatment" value at 0.
#This value  doesn't really matter
# it will be canceled by the treat==0 interaction anyway
#But we want to make sure they aren't NA
#otherwise feols would drop these never-treated observations

dat[, time_to_treat := ifelse(treat==1, year - `_nfd`, 0)]
```
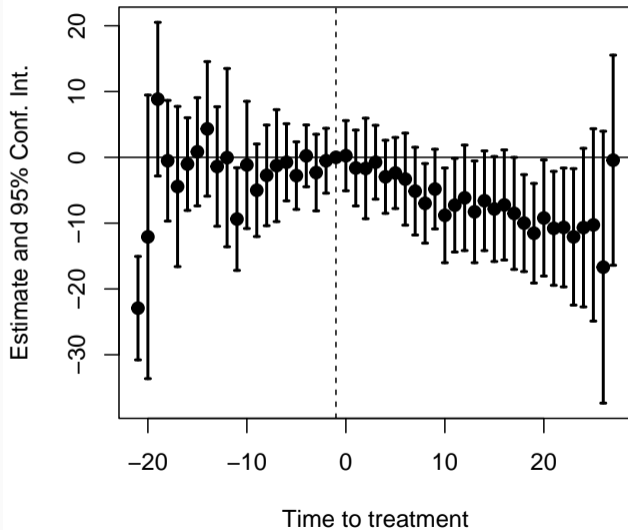
Implementation

```
mod_twfe = feols(asmrs ~ i(time_to_treat, treat, ref = -1) +
                    pcinc + asmrh + cases |
                    stfips + year,
                 cluster = ~stfips,
                 data = dat)


iplot(mod_twfe,
      xlab = 'Time to treatment',
      main = 'Event study: Staggered treatment (TWFE)',
      las=2,cex=1.5,cex.lab=1.5,lwd=2,
      grid=F,zero=T)
```

Event study: Staggered treatment (TWFE)

Instrumental variables

Social Networks and the Decision to Insure[†]

By Jing Cai, Alain de Janvry, and Elisabeth Sadoulet*

Using data from a randomized experiment in rural China, we study the influence of social networks on weather insurance adoption and the mechanisms through which they operate. To quantify network effects, the experiment provides intensive information sessions about the product to a random subset of farmers. For untreated farmers, the effect of having an additional treated friend on take-up is equivalent to granting a 13 percent reduction in the insurance premium. By varying the information available about peers' decisions and randomizing default options, we show that the network effect is driven by the diffusion of insurance knowledge rather than purchase decisions. (JEL G22, O12, O16, P36, Q12, Q54, Z13)

18

- Decision to buy insurance against weather events

- In particular, whether information about insurance travels through social networks

- How much does what your friends learn about insurance affect your own takeup?

  - Does your friends actually buying insurance make you more likely to buy?

- Randomized experiment in rural China

  - Households randomized into two rounds of informational sessions about insurance

  - Outcome: Second round of sessions

  - Treatment: what friends did and saw in the first round of sessions

  - Round one: randomly assigned to "default buy" or "default no buy"

IV

```
# Include just the outcome and controls first
#then endogenous ~ instrument
# in the second part, cluster variable
m <- feols(takeup_survey ~ male + age + agpop + ricearea_2010 +
             literacy + intensive + risk_averse + disaster_prob +
             factor(village) | pre_takeup_rate ~ default,
           cluster = ~address, data = d)

#reduced form
m_reduced <- feols(takeup_survey ~ default+male + age +
agpop + ricearea_2010 + literacy +
intensive + risk_averse +disaster_prob +
             factor(village),
           cluster = ~address, data = d)
```

IV

```
#OLS
m_OLS<- feols(takeup_survey ~ pre_takeup_rate+male + age + agpop +
 ricearea_2010 +literacy + intensive + risk_averse
 + disaster_prob +factor(village),cluster = ~address, data = d)

# Show the first and second stage, omitting all
# the controls for ease of visibility
msummary(list('OLS'=m_OLS,
              'First Stage' = m$iv_first_stage[[1]],
              'Reduced Form' = m_reduced,
              'Second Stage' = m),
         coef_map = c(default = 'First Round Default',
                      pre_takeup_rate = 'Friends Purchase Behavior',
                      fit_pre_takeup_rate = 'Friends Purchase Behavior'),
         stars = c('*' = .1, '**' = .05, '***' = .01),
         output="latex_tabular")
```

	OLS	First Stage	Reduced Form	Second Stage
First Round Default		0.118***	0.093***	
		(0.034)	(0.030)	
Friends Purchase Behavior	0.452***			0.791***
	(0.071)			(0.273)
Num.Obs.	1378	1378	1378	1378
R2	0.142	0.469	0.121	0.127
Std.Errors	by: address	by: address	by: address	by: address

Regression Discontinuity: Sharp

Government Transfers and Political Support[†]

*By* Marco Manacorda, Edward Miguel, and Andrea Vigorito[*]

This paper estimates the impact of a large anti-poverty cash transfer program, the Uruguayan PANES, on political support for the government that implemented it. Using the discontinuity in program assignment based on a pretreatment eligibility score, we find that beneficiary households are 11 to 13 percentage points more likely to favor the current government relative to the previous government. Political support effects persist after the program ends. Our results are consistent with theories of rational but poorly informed voters who use policy to infer politicians' redistributive preferences or competence, as well as with behavioral economics explanations grounded in reciprocity. (JEL D72, H23, H53, I38, O15, O17)

- Large poverty alleviation program in Uruguay: sizeable check to a large portion of the population.

- Interested in whether receiving those funds made people more likely to support the newly-installed center-left government that sent them

- Who got the payments?

- The government used a bunch of factors - housing, work, reported income, schooling - and predicted what your income would be from that

- Treatment was assigned based on being below a cutoff

Implementation

```
gt <- causaldata :: gov_transfers
#Check if its a fuzzy or a sharp RD
# Use cut () to create bins , using breaks to make sure it breaks at 0
# (-15:15)*.02/15 gives 15 breaks from -.02 to .02
binned <- gt %>%
  mutate ( Inc_Bins = cut ( Income_Centered ,
                          breaks = (-15:15)*(.02/15))) %>%
  group_by ( Inc_Bins ) %>%
  summarize ( Participation = mean ( Participation ),
             Income = mean ( Income_Centered ))
# Taking the mean of Income lets us plot data roughly at the bin midpoints

ggplot ( binned , aes (x = Income , y = Participation )) +
  geom_line () +
  # Add a cutoff line
  geom_vline ( aes ( xintercept = 0) , linetype = 'dashed ')
```
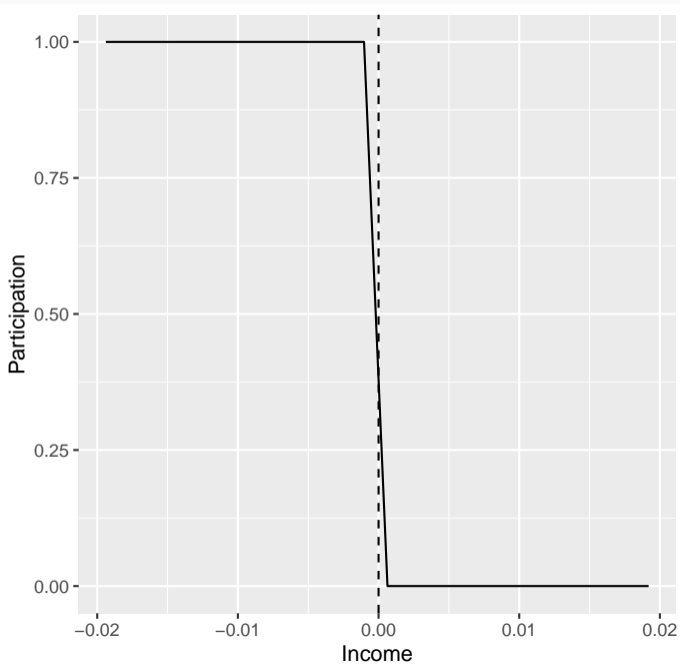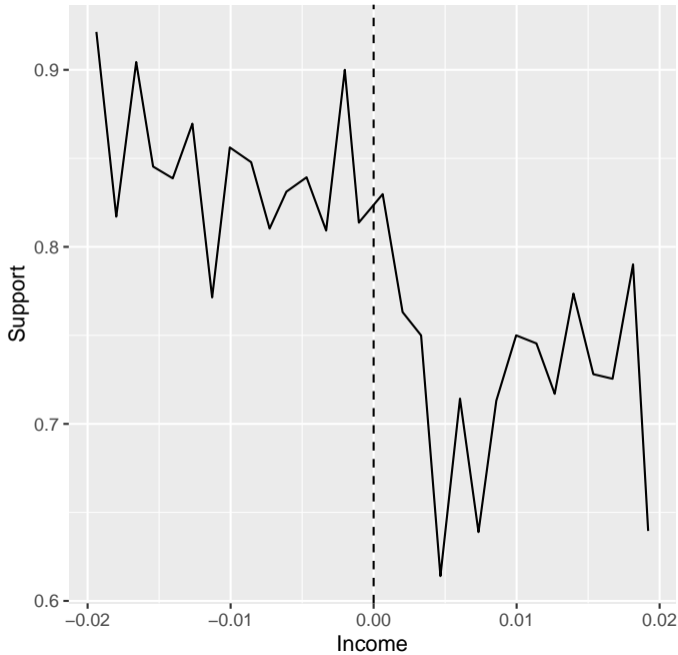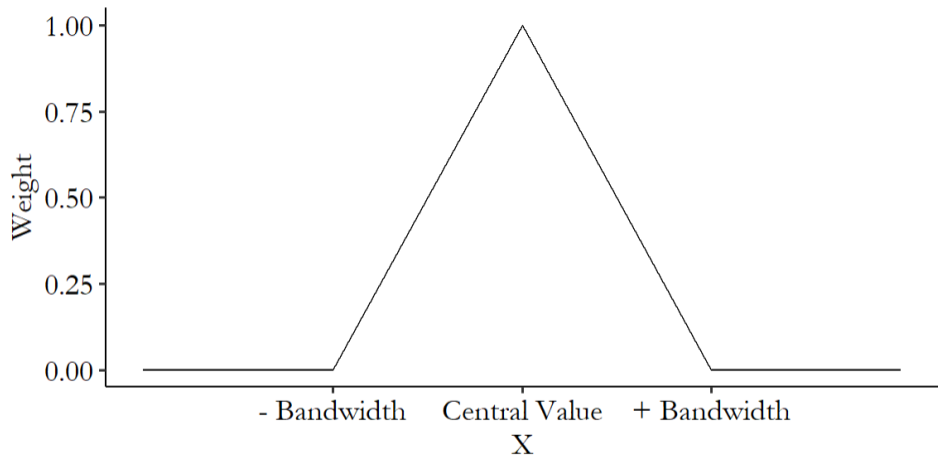
Implementation

```
# Use cut() to create bins, using breaks to make sure it breaks at 0
# (-15:15)*.02/15 gives 15 breaks from -.02 to .02
binned <- gt %>%
  mutate(Inc_Bins = cut(Income_Centered,
                        breaks = (-15:15)*(.02/15))) %>%
  group_by(Inc_Bins) %>%
  summarize(Support = mean(Support),
            Income = mean(Income_Centered))
# Taking the mean of Income lets us plot data roughly at the bin midpoints

ggplot(binned, aes(x = Income, y = Support)) +
  geom_line() +
  # Add a cutoff line
  geom_vline(aes(xintercept = 0), linetype = 'dashed')
```

28

Implementation

```
# Add a triangular kernel weight
kweight <- function(x) {
  # To start at a weight of 0 at x = 0, and impose a bandwidth of .01,
  # we need a "slope" of -1/.01 = 100,
  # and to go in either direction use the absolute value
  w <- 1 - 100*abs(x)
  # if further away than .01, the weight is 0, not negative
  w <- ifelse(w < 0, 0, w)
  return(w)
}
```

Implementation

```r
# Run linear
m <- lm(Support ~ Income_Centered*Participation, data = gt,
        subset=abs(Income_Centered)<=0.01)

# Run the same model but with the weight
mw <- lm(Support ~ Income_Centered*Participation, data = gt,
         weights = kweight(Income_Centered))

# Linear term and a squared term with "treated" interactions
m2 <- lm(Support ~ Income_Centered*Participation +
         I(Income_Centered^2)*Participation, data = gt,
         subset=abs(Income_Centered)<=0.01)

# Run the same model but with the weight
m2w <- lm(Support ~ Income_Centered*Participation +
          I(Income_Centered^2)*Participation, data = gt,
          weights = kweight(Income_Centered))
```

	Quadratic	Quadratic Kernel	Linear	Linear Kernel
(Intercept)	0.891***	0.883***	0.773***	0.819***
	(0.044)	(0.021)	(0.032)	(0.015)
Income_Centered	-80.455***	-73.751***	-10.338*	-23.697***
	(22.417)	(11.830)	(5.658)	(3.219)
Participation	-0.032	-0.042	0.077*	0.033
	(0.067)	(0.036)	(0.041)	(0.021)
	6944.814***	6124.497***		
	(2204.994)	(1281.007)		
Income_Centered × Participation	87.748***	69.095***	12.272*	26.594***
	(31.584)	(18.784)	(7.398)	(4.433)
Participation × I(Income_Centered^2)	-6428.250**	-7012.181***		
	(3014.945)	(1969.275)		
Num.Obs.	937	937	937	937
R2	0.048	0.047	0.037	0.041
Std.Errors	HC3	HC3	HC3	HC3

32

- In general, you don't do things by hand

- Instead use packages (rdrobust) that do things in the best way possible

- https://rdpackages.github.io/

  - Do not have to choose the bandwidth. It will do an optimal bandwidth selection procedure for us

  - Implement local regression with your polynomial order of choice (linear by default) and a kernel (triangular by default)

  - Heteroskedasticity-robust standard errors

  - Bias correction for the standard errors

```
library(rdrobust)
# Estimate regression discontinuity and plot it
m <- rdrobust(gt$Support, gt$Income_Centered, c = 0)
summary(m)
# Note, by default, rdrobust and rdplot use different numbers
# of polynomial terms. You can set the p option to standardize them.
rdplot(gt$Support, gt$Income_Centered)
```
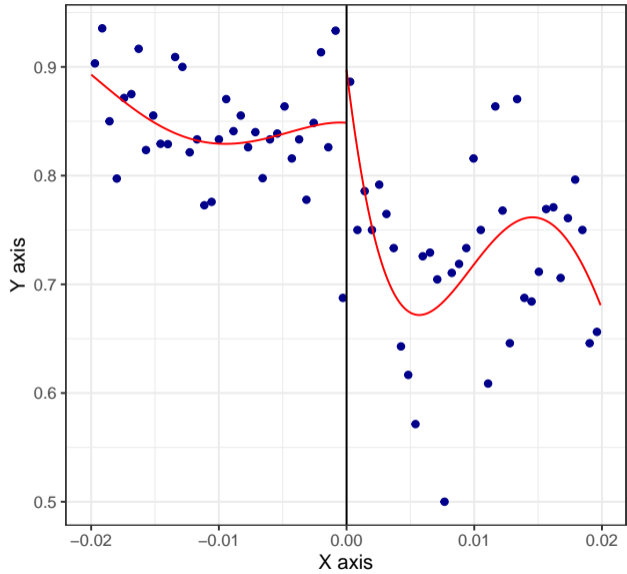
RD Plot

Implementation

```
library(rddensity)
mcr=rddensity(gt$Income_Centered)
rdplotdensity(mcr,X=gt$Income_Centered)
```
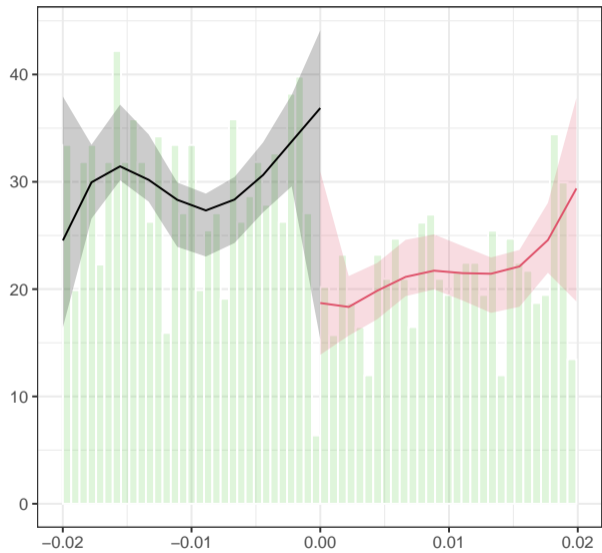
Regression Discontinuity: Fuzzy

How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills

Daniel K. Fetter

Download Full Text PDF

Article Information | Comments (0)

Abstract

The largest twentieth-century increase in US home ownership occurred between 1940 and 1960, associated largely with declining age at first ownership. I shed light on the contribution of coincident government mortgage market interventions by examining home loan benefits granted under the World War II and Korean War GI Bills. Veterans' benefits increased home ownership rates primarily by shifting purchase earlier in life, explaining 7.4 percent of the overall 1940-1960 increase, and 25 percent of the increase for affected cohorts. A rough extrapolation suggests that broader changes in mortgage terms can explain 40 percent of the 1940-1960 increase. (JEL G21, N22, N92, R21, R31)

- How much of the increase in the home ownership rate in the midcentury US was due to mortgage subsidies given out by the government

- Veterans of major wars like World War II or the Korean War received special mortgage subsidies

- What does this have to do with regression discontinuity?There's an age requirement to join the military.

- If you were born one year too late to join the military to fight in a War, then you won't get these mortgage subsidies (or at least far fewer veterans were eligible)

- Of course, not everybody born in a given year joins the military.

- Treatment rates jump from 0% to... above 0% but not 100%

Implementation

```
vet <- causaldata::mortgages
# Create an "above-cutoff" variable as the instrument
vet <- vet %>% mutate(above = qob_minus_kw > 0)
# Impose a bandwidth of 12 quarters on either side
vet <- vet %>%  filter(abs(qob_minus_kw) < 12)
m <- feols(home_ownership ~
    nonwhite   | # Control for race
    bpl + qob | # fixed effect controls
    qob_minus_kw*vet_wwko ~ # Instrument our standard RDD
    qob_minus_kw*above, # with being above the cutoff
    se = 'hetero', # heteroskedasticity-robust SEs
    data = vet)
# And look at the results
msummary(m, stars = c('*' = .1, '**' = .05, '***' = .01))
```

|	(1)
fit_qob_minus_kw	-0.007***
	(0.002)
fit_vet_wwko	0.170***
	(0.046)
fit_qob_minus_kw × vet_wwko	-0.003
	(0.003)
nonwhite	-0.190***
	(0.007)
Num.Obs.	56901
R2	0.053
Std.Errors	Heteroskedasticity-robust
FE: bpl	X
FE: qob	X

Implementation

```
# Create our matrix of controls
controls <- vet %>%
    select(nonwhite, bpl, qob) %>%
    mutate(qob = factor(qob))
# and make it a matrix with dummies
conmatrix <- model.matrix(~., data = controls)

# This is fairly slow due to the controls, beware!
m <- rdrobust(vet$home_ownership,
              vet$qob_minus_kw,
              fuzzy = vet$vet_wwko,
              c = 0,
              covs = conmatrix)

summary(m)
rdplot(vet$vet_wwko, vet$qob_minus_kw)
rdplot(vet$home_ownership, vet$qob_minus_kw)
```
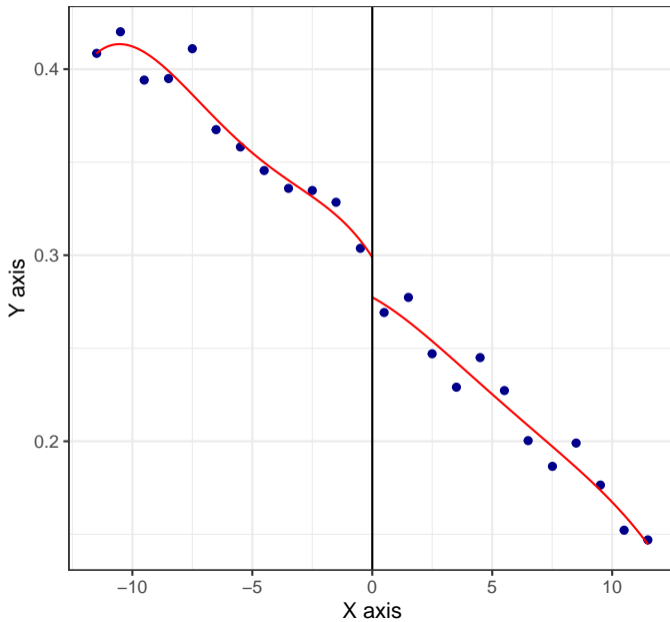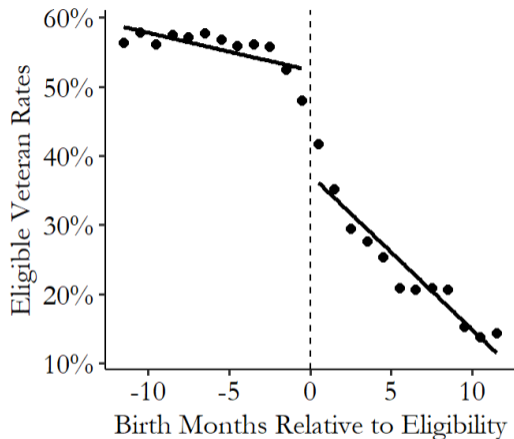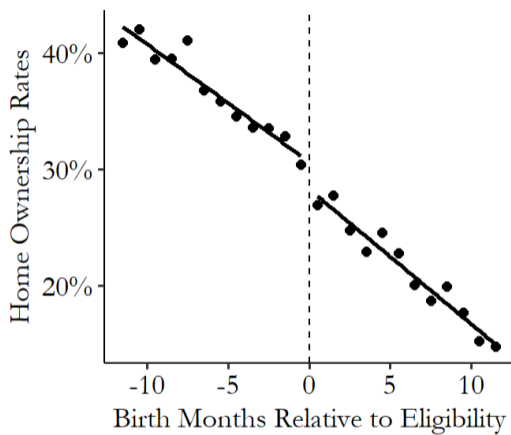
RD Plot

43
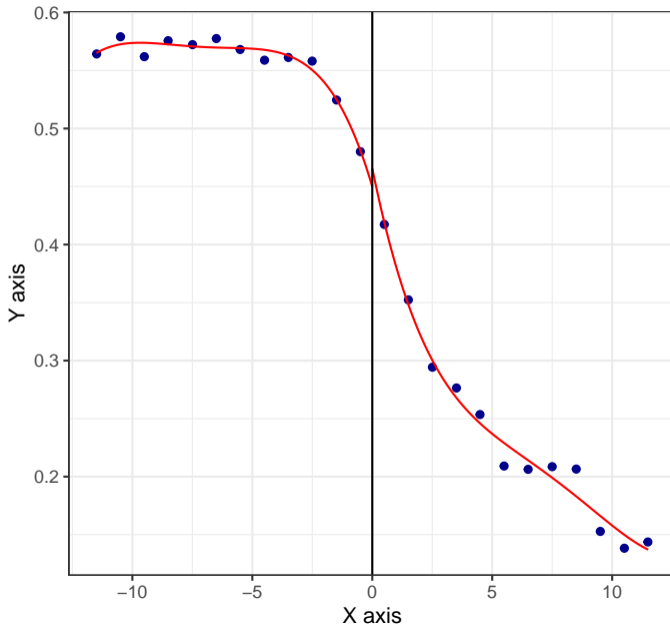
RD Plot

(a) RD for Eligible Vet Status

(b) RD for Home Ownership