UNIVERSITY OF MINNESOTA TWIN CITIES
Driven to Discover®

Effect of Poverty Alleviation Program on Political Support to the Implementing Government

Team 11

Manish C Jain, Priyanka Rajendra Bhosale, Zecong Zhao, Yufan Li

Table of Contents

Executive summary

PANES (Plan de Atención Nacional a la Emergencia Social) was a large anti-poverty program carried out by the Government of Uruguay. The beneficiary households received benefits in cash (USD70 per month) and kind, independent of the household size. We attempt to analyze the effect of the PANES program on political support for the government that implemented it. Our data is a sample of 1,948 families with the income normalized to remove self-selection bias, participation that capture the recipients of benefit and support for government. We use an inferential method called RDD to evaluate the effect in the case for people close to income cut-off, as we assume them to be similar. We found that receiving funds increased families' support for the Uruguay government by 0.093 units. Hence, the implementation of the PANES plan helped to enhance public support for the government.

Introduction

PANES (Plan de Atención Nacional a la Emergencia Social) was a large anti-poverty program carried out by the Government of Uruguay from April 2005 to December 2007. The beneficiary households received benefits in cash (USD70 per month) and kind, independent of the household size. This amount was significant for the target population as it was more than 50% of the average pre-program self-reported income [1]. We attempt to analyze the effect of the PANES program on political support for the government that implemented it.

Program eligibility was determined by a predicted income score based on many socioeconomic variables, and only households with scores below a predetermined threshold were eligible. The income score was designed by researchers at the University of the Republic. It was based on a probit model of the likelihood of being below a critical per capita income level. The eligibility threshold, the predicted income score, and the exact variables and weights attached to them to find the score were not disclosed to avoid any external manipulation leading to a quasi-random assignment. The researchers suggested this discontinuous rule for program assignment, which allows us to evaluate the impact of policy without having any external factors affecting it.

Dataset Overview

The PANES program lasted from April 2005 to December 2007. Households with income scores in the neighborhood of the threshold were surveyed to find their support for the current government about 18 months after the program's start. A second similar follow-up survey took place in 2008 after the program ended.

Our data is a sample of 1,948 families who applied for PANES, with beneficiaries and nonrecipients in the program's qualifying threshold range. Our data contains three features, income score, participation, and support. The variables are:

Outcome Variables	Support (Categorical)	Support: Measure of support for the government 1: support more than the previous government 0.5: support the same as the previous government 0: support worse than the previous government
Input Variables	Income_centered (Numerical) **Forcing Variable**	Predicted income score, centered around program cutoff (0). Families with Income_centered less than 0 will receive funds. And families with Income_centered larger than 0 will not receive funds. Range: -0.02 to 0.02
	Participation (Categorical) **Threshold Dummy**	0: not received funds; 1: received funds

Based on our data quality checks, we found no missing values in the data. A brief univariate summary of the features in the dataset-

1. **Distribution of Support:** Number of households supporting more than the previous government counts for 67.4%; Number of households supporting the same as the

previous government counts for 24.44%; Number of households supporting less than the previous government counts for 8.16%.

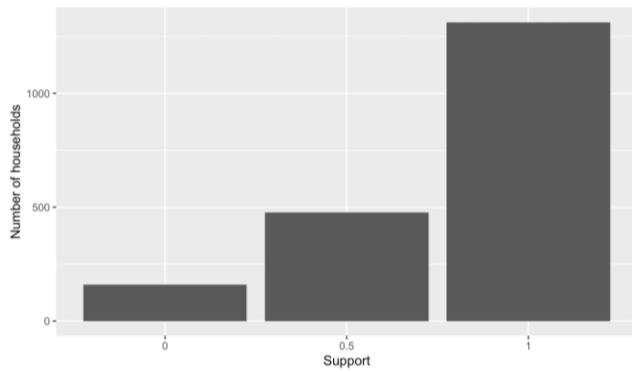


Figure 1: Distribution of Support Variable

2. **Distribution of Income_centered:** The Income_centered graph should be normally distributed to infer no self-selection. However, based on the structure of the experiment, this has been resolved by creating the income variable through a mix of variables.

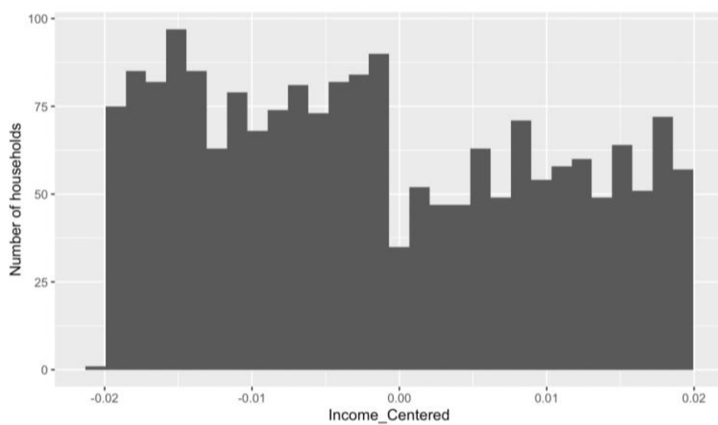


Figure 2: Distribution of Income centered

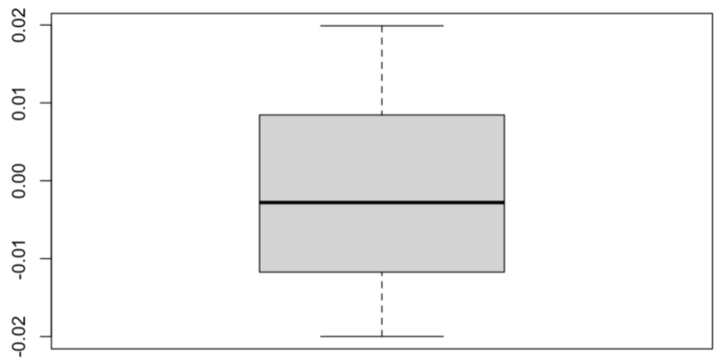


Figure 3: Box Plot of Income centered

3. **Distribution of Participation:** Number of participating households accounts for 57.85%, while the number of non-participating households accounts for 42.15%.

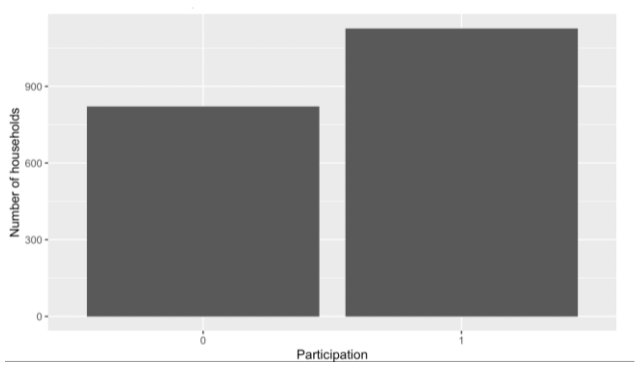


Figure 4: Distribution of Participation Flag

Threats to causal Inference

1. Self-Selection: The households could have reduced their income to qualify for the PANES program threshold.
2. Reverse Causality: The households that supported the government were favored getting the cash transfers from the government.
3. Omitted variables: Inherent political preferences and biases of the people will influence the political support for the government. For example, we might have poor people that strongly support the left-centered government no matter what.

4. Bitterness among non-beneficiaries: The households that barely missed the threshold could be bitter at their exclusion from the program. They can see that other people like them are getting the benefits. In this case, the estimates of political support are a combination of two different effects and might cause issues in interpreting the results.
5. Interference: The cash transfer in PANES was a substantial sum for the target demographic, accounting for more than half of the average self-reported pre-program household income among program participants. There is a chance that households receiving the benefit might share it with a related household that did not qualify.

In order to solve the above issues, we used the Regression Discontinuity Design (RDD) method.

Method Introduction and Assumptions

Regression Discontinuity Design (RDD) is a quasi-experimental pretest-posttest design that aims to determine the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it is possible to estimate the local average treatment effect in environments in which randomization is unfeasible. Below are the required assumptions of the RDD method:

1. **No Self-Selection:** People should not select which group they belong to by themselves. The government did not use income as a running variable; that might have been too easy to manipulate. Instead, the government used a bunch of factors - housing, work, reported income, schooling - and predicted what the income would be from that. To avoid potential manipulation, neither the enumerators nor households were ever even informed that a score was used to determine eligibility, the exact variables entered into the score, the weights attached to them, or the program eligibility threshold. Then, the predicted income

was the running variable, and treatment was assigned based on being below a cutoff. About 14% of the population ended up getting payments.

2. **No Model Miss-specification:** We are required to use the best model to fit the data. We used AIC and BIC to select the model that best fits the untreated group data.
3. Individuals close to the cutoff point should be similar, on average, in observed and unobserved characteristics. Our analysis assumes that there is no systematic difference between the households close to the threshold that qualified and did not qualify for the program.

Besides, as the eligibility threshold, the predicted income score, the exact variables, and weights attached to them to find the score were not disclosed to anyone, and this implementation was strict, we can be sure that there was political isolation, no external manipulation, and assumed quasi-random assignment around the threshold. Because of this, we can safely say that there are no chances of self-selection and reverse causality. The political preferences will not be a problem in RDD analysis as there will be similar types of people on both sides of the threshold. Lastly, to check if the bitterness among non-beneficiaries affects the results, we can potentially look at the change in the government's political support before and after the implementation of the program, specifically for the group that did not get cash transfers. Regarding the possibility of interference, there is no way to know if this happened.

Analysis

Visualize the data

From the graphical regression discontinuity model check, we can see a break at the cut-off point. Overall, higher support is detected at the left cut-off for the treatment group. Besides, the graph shows that the polynomial degree of the model should be larger than 1.

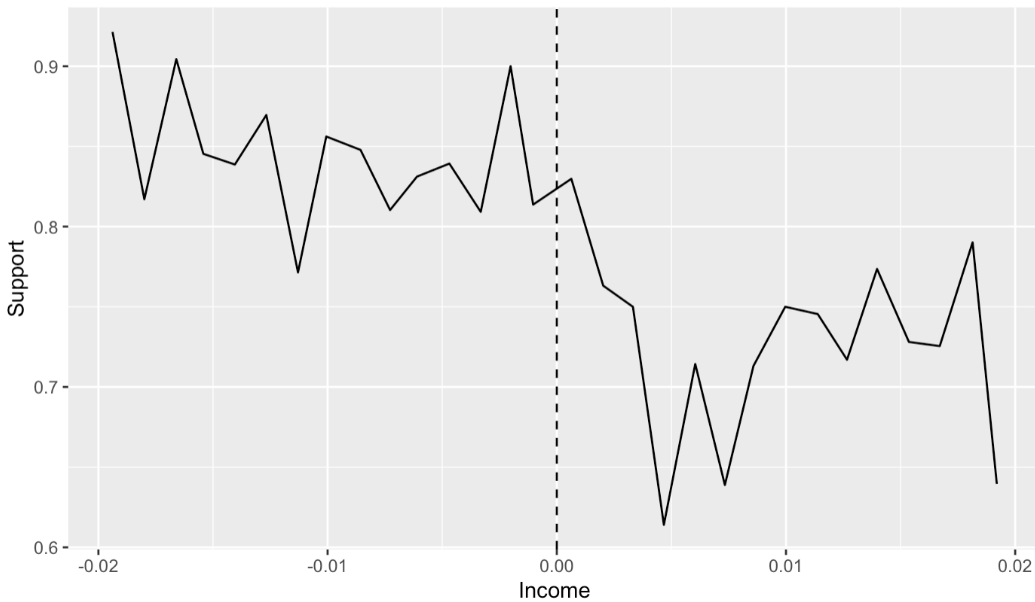


Figure 5: Government support by Income with Policy Cut-off

Fitting models on untreated data

In order to find the best model, we tried different polynomial degrees to fit the untreated group. The graph shows that when the polynomial degree equals 3, AIC and BIC are the lowest, and the model is the best, but the results are pretty close. Therefore, in the following RDD analysis, we will first use a linear regression model with polynomial degree 3, and based on the result, we will make some adjustments.

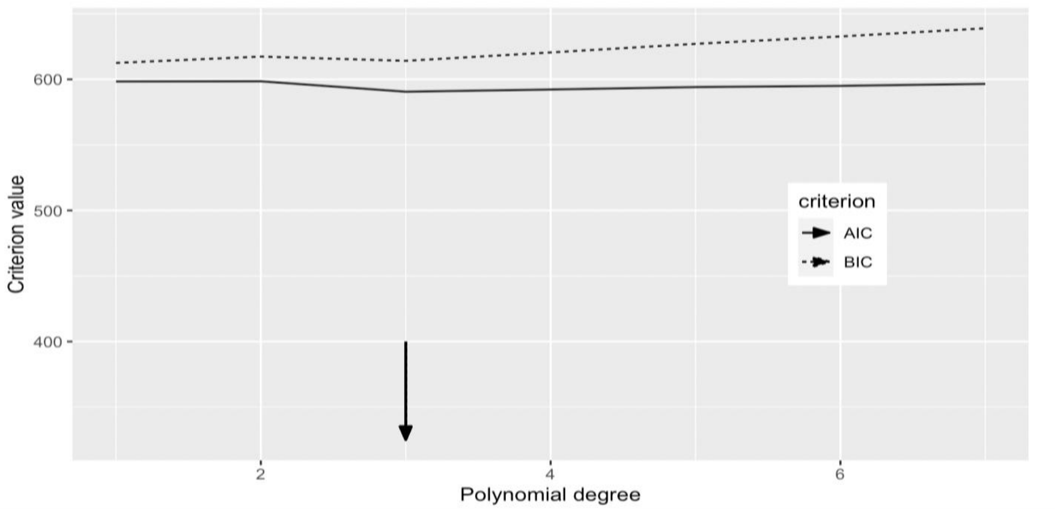


Figure 6: AIC and BIC

RDROBUST Results

We used RDROBUST to find out the best bandwidth, which is 0.005. And from the rdplot, we can again see the drop above and below the cutoff point.

```
[1] "Mass points detected in the running variable."
Call: rdrobust

Number of Obs.                 1948
BW type                       mserd
Kernel                   Triangular
VCE method                       NN

Number of Obs.                 1127          821
Eff. Number of Obs.             291          194
Order est. (p)                    1            1
Order bias  (q)                   2            2
BW est. (h)                   0.005        0.005
BW bias (b)                   0.010        0.010
rho (h/b)                     0.509        0.509
Unique Obs.                     841          639

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     0.025     0.062     0.396     0.692    [-0.098 , 0.147]
        Robust         -         -     0.624     0.533    [-0.097 , 0.188]
=============================================================================
[1] "Mass points detected in the running variable."
```

Figure 7: RDROBUST results

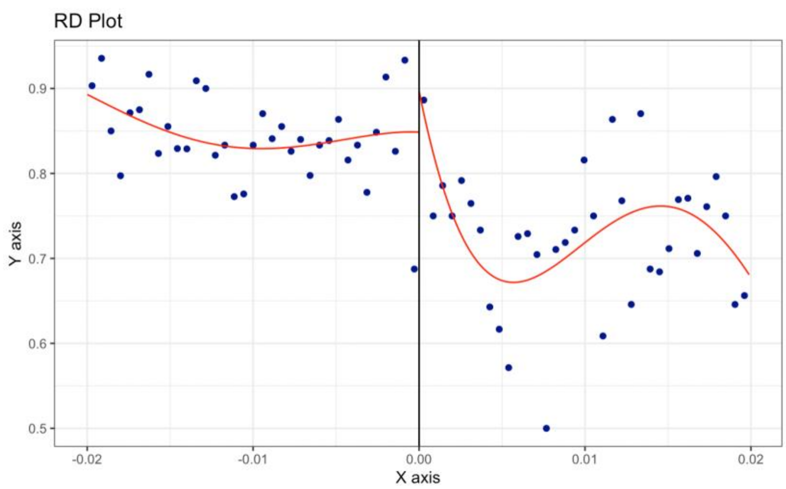


Figure 8: Plotting a polynomial degree 3 model on data

Result with bandwidth 0.005 and polynomial degree 3

Equation:

Support = β_0 + β_1*(Income_Centered-X_c) + β_2* Participation + β_3*$(Income\_Centered - X_c)^2$+ β_4*$(Income\_Centered - X_c)^3$ + β_5 *(Income_Centered-X_c)* Participation + β_6 *$(Income\_Centered - X_c)^2$* Participation + β_7 *$(Income\_Centered - X_c)^3$* Participation

Notes: *X_c is cut-off, X_c=0; Participation=1 if Income_Centered>0 (If Income_Centered is greater than the cut off)*

The coefficient of Participation is -0.364, which indicates that receiving funds reduces support for the government by 0.364. This is a little bit against our expectations. As per Figure 5, we can clearly observe that families who got those financial funds with low income should be more supportive of the government since they got benefits. However, the opposite result shows the opposite result that they support less.

This may be because we used a smaller bandwidth -- 0.005, which sharply narrows down the sample size from 1948 to 462 and further decreases statistical power. Also, we noticed that the p-value of $(Income\_Centered)^2$, $(Income\_Centered)^3$ is high. This could be the signal that the model and bandwidth are not suitable.

```
Call:
lm(formula = Support ~ Income_Centered * Participation + I(Income_Centered^2) *
    Participation + I(Income_Centered^3) * Participation, data = gt2)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8903 -0.2419  0.1281  0.1942  0.3253

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          8.914e-01  8.991e-02   9.914   <2e-16 ***
Income_Centered                     -9.686e+01  1.570e+02  -0.617   0.5375
Participation                       -3.640e-01  1.740e-01  -2.092   0.0370 *
I(Income_Centered^2)                 1.838e+04  7.349e+04   0.250   0.8026
I(Income_Centered^3)                -1.541e+06  9.592e+06  -0.161   0.8724
Income_Centered:Participation       -3.693e+02  2.527e+02  -1.461   0.1446
Participation:I(Income_Centered^2) -2.043e+05  1.078e+05  -1.895   0.0587 .
Participation:I(Income_Centered^3) -2.009e+07  1.349e+07  -1.489   0.1372
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2974 on 454 degrees of freedom
Multiple R-squared:  0.05034,	Adjusted R-squared:  0.0357
F-statistic: 3.438 on 7 and 454 DF,  p-value: 0.001351
```

Figure 9: Model summary of polynomial degree 3 on bandwidth of -0.005 to 0.005

Result with bandwidth 0.02 and polynomial degree 2

Because the range of X(Income_centered) of original data is only 0.02, which is significantly small, we decided to use 0.02 as our new bandwidth. On the other hand, because of the high p-value of some coefficients of the above model, we changed the polynomial degree to 2 to see the result.

Equation:

Support = β_0 + β_1*(Income_Centered-X_c) + β_2* Participation + β_3*$(Income\_Centered - X_c)^2$ + β_4 *(Income_Centered-X_c)* Participation + β_5 *$(Income\_Centered - X_c)^2$* Participation

Notes: *X_c is cut-off, X_c=0; Participation=1 if Income_centered>0 (If Income_centered is greater than the cut off)*

The coefficient of Participation here is 0.093, meaning that receiving funds made support for the government increase by 0.093 units. This is consistent with our expectations and visualization results in Figure 5.

```
Call:
lm(formula = Support ~ Income_Centered * Participation + I(Income_Centered^2) *
    Participation, data = gt)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8887 -0.2276  0.1528  0.1706  0.2905

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                           0.76900    0.03445  22.321   <2e-16 ***
Income_Centered                     -11.56663    7.77729  -1.487   0.1371
Participation                         0.09286    0.04585   2.025   0.0430 *
I(Income_Centered^2)                562.24726  372.18239   1.511   0.1310
Income_Centered:Participation        19.29999   10.44517   1.848   0.0648 .
Participation:I(Income_Centered^2) -101.10250  500.19558  -0.202   0.8398
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3127 on 1942 degrees of freedom
Multiple R-squared:  0.03629,   Adjusted R-squared:  0.03381
F-statistic: 14.63 on 5 and 1942 DF,  p-value: 4.245e-14
```

Figure 10: Model summary of polynomial degree 2 on bandwidth of -0.02 to 0.02

Findings and Conclusions

We conclude from the above analysis that under the optimal bandwidth suggested by RDROBUST, families who received funds are 36.4% less likely to support the government. This result comes from a very narrow range – 0.5% around the cutoff line where there is a small jump. However, when we look at a slightly larger sample, 2% around the cutoff line, we found that receiving funds increased families’ support for the Uruguay government by 0.093 units. The implementation of the PANES plan helped to enhance public support for the government.

Limitations

1. RDD analysis only looks at a local sub-group of the overall sample as it estimates local average treatment effects (LATE) around the cutoff point, where treatment and comparison units are most similar
2. RDD requires optimal bandwidth. In order to improve similarity in groups, we can reduce the bandwidth even further. But we see an opposite trend. It is due to low statistical power as the number of records drop. If we utilize a larger bandwidth to increase the statistical power, the units might not be comparable enough for the assumptions of RDD analysis to be true.
3. RDD requires no model misspecification. During our exploration, we found that even though we can use AIC, BIC method, it is still not easy to find the best model to fit the data, and the result and statistical power could be affected by this.

References

[1]. *Manacorda, Marco, Edward Miguel, and Andrea Vigorito. 2011. “Government Transfers and Political Support.” American Economic Journal: Applied Economics 3 (3): 1–28.*

Appendix

Import data

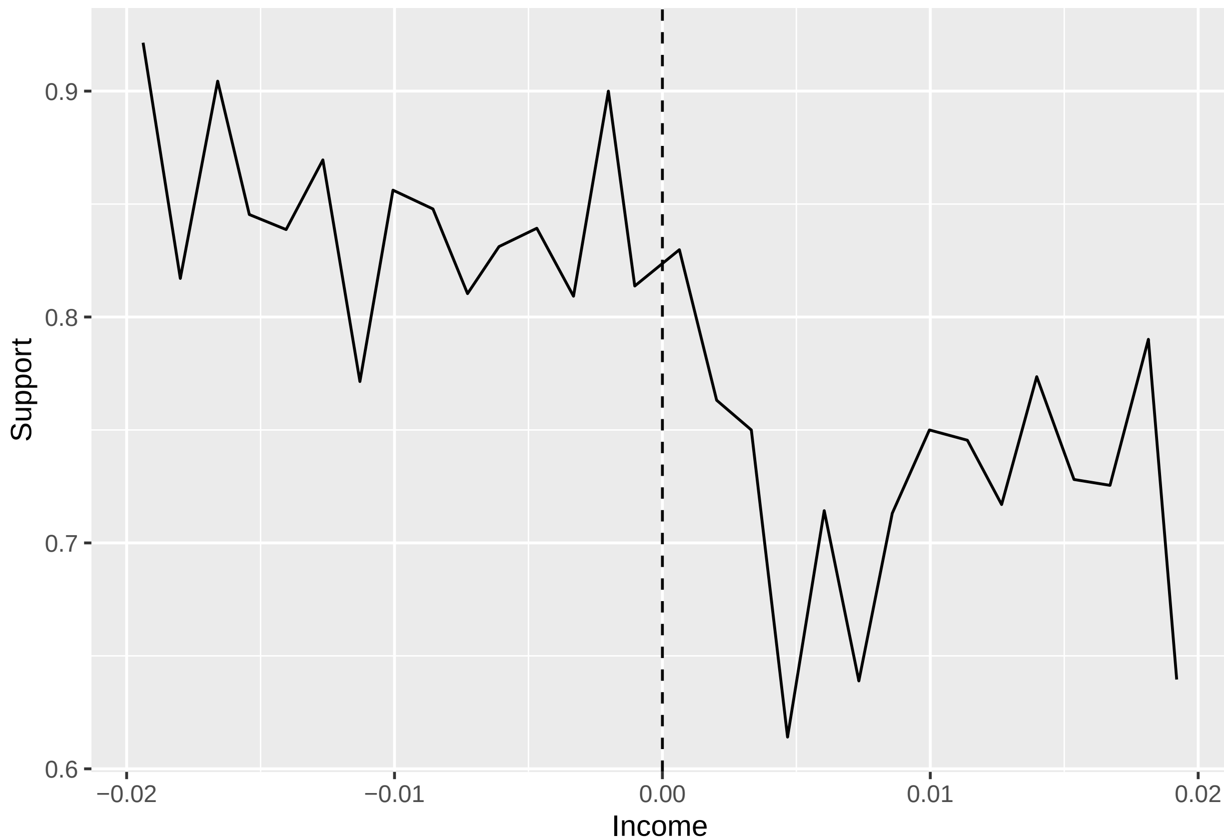
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Basic linear regression

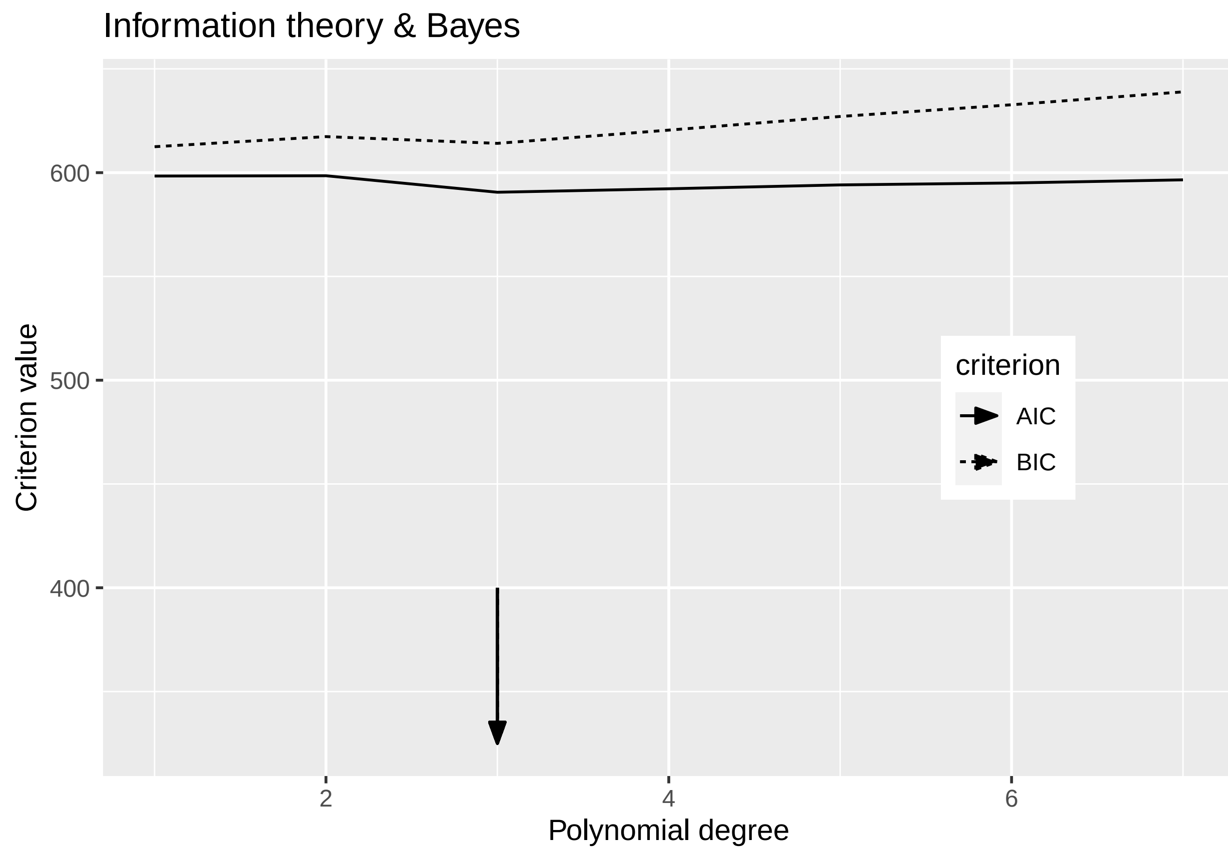
```
##
## Call:
## lm(formula = Support ~ Participation, data = gt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8460 -0.2278  0.1540  0.1540  0.2722
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.72777    0.01092   66.66  < 2e-16 ***
## Participation  0.11828    0.01435    8.24 3.12e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3128 on 1946 degrees of freedom
## Multiple R-squared:  0.03372,    Adjusted R-squared:  0.03322
## F-statistic:  67.9 on 1 and 1946 DF,  p-value: 3.12e-16
```

RDD Analysis

visualization the data



Fitting models on untreated data and AIC BIC score



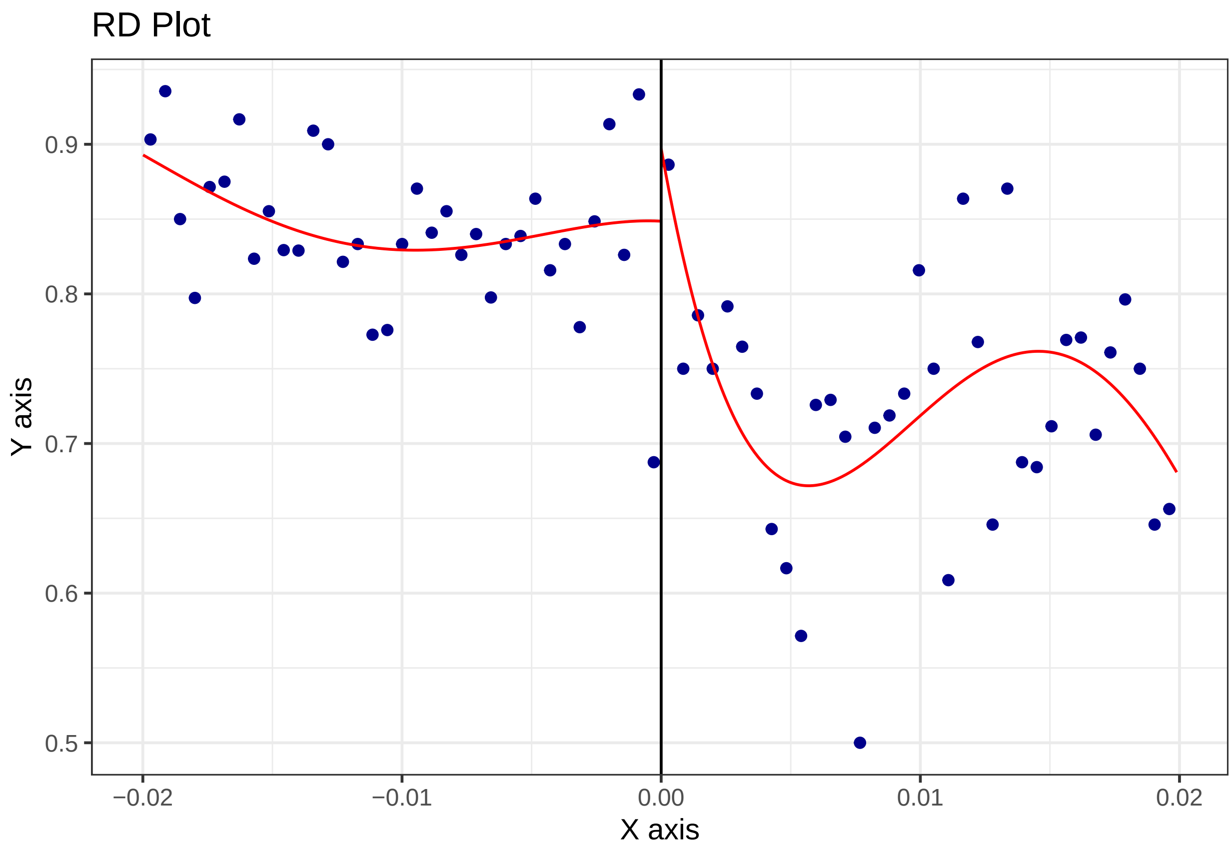
rdrobust result

We use rdrobust to find out the best band width, which is 0.005. And from the rdplot we can know the best model should be high polynomial(larger than 1).

```
## Warning: package 'rdrobust' was built under R version 4.1.3

## [1] "Mass points detected in the running variable."

## Call: rdrobust
##
## Number of Obs.                 1948
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 1127          821
## Eff. Number of Obs.             291          194
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.005        0.005
## BW bias (b)                   0.010        0.010
## rho (h/b)                     0.509        0.509
## Unique Obs.                     841          639
##
## =============================================================================
```

```
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional      0.025     0.062     0.396     0.692    [-0.098 , 0.147]
##         Robust          -         -     0.624     0.533    [-0.097 , 0.188]
## =============================================================================
```

```
## [1] "Mass points detected in the running variable."
```



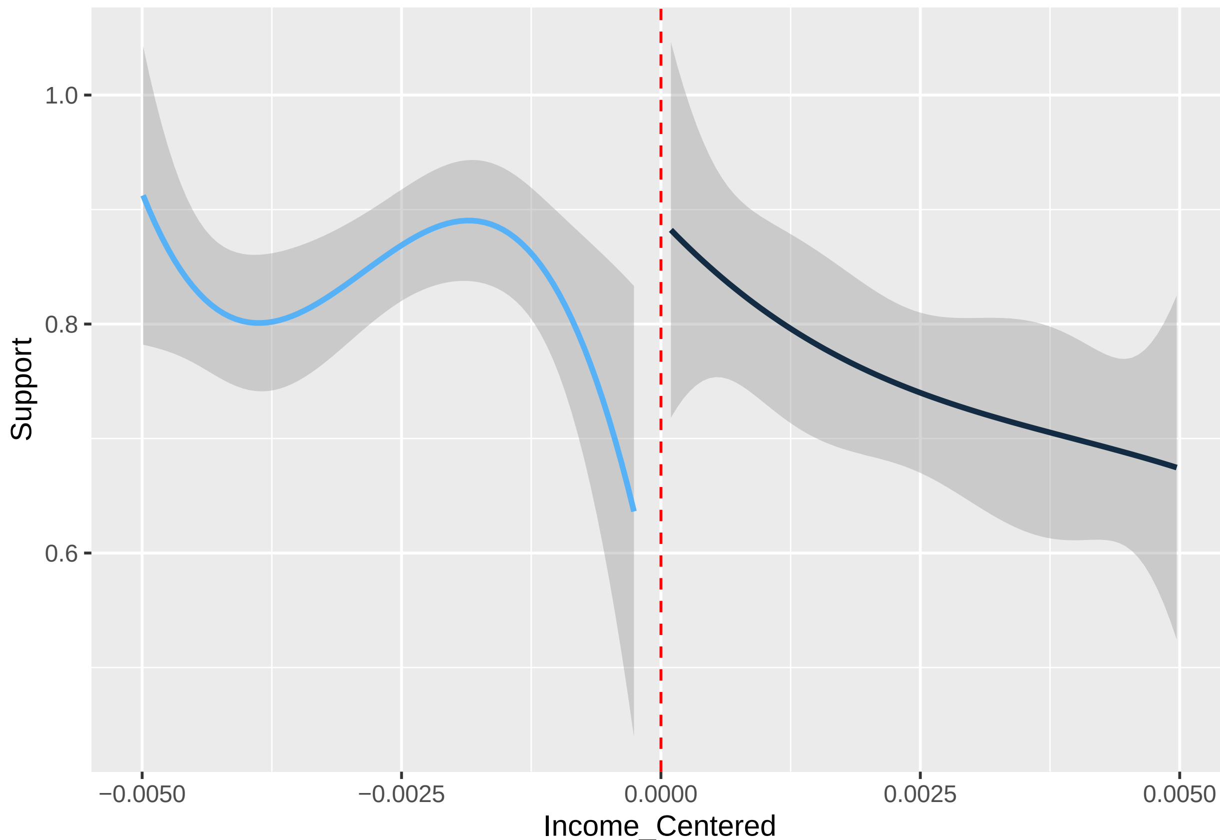
Run RDD

The coefficient of Participation is -0.364. This is a little bit against our expectation that we assume families with treatment, that is, families got those financial fund with low income, should be more supportive for the government since they got benefits. But here the result shows the opposite result that they support less. From the graph, we can also see this gap. This may be because we used smaller band width – 0.005, which sharply narrow down the sample size from 1948 into 462 and further decreases statistical power.

Run RDD with bandwidth 0.005 and polynomial degree 3

```
##
## Call:
## lm(formula = Support ~ Income_Centered * Participation + I(Income_Centered^2) *
##     Participation + I(Income_Centered^3) * Participation, data = gt2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8903 -0.2419  0.1281  0.1942  0.3253
##
```

```
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         8.914e-01  8.991e-02   9.914   <2e-16 ***
## Income_Centered                    -9.686e+01  1.570e+02  -0.617   0.5375
## Participation                      -3.640e-01  1.740e-01  -2.092   0.0370 *
## I(Income_Centered^2)                1.838e+04  7.349e+04   0.250   0.8026
## I(Income_Centered^3)               -1.541e+06  9.592e+06  -0.161   0.8724
## Income_Centered:Participation      -3.693e+02  2.527e+02  -1.461   0.1446
## Participation:I(Income_Centered^2) -2.043e+05  1.078e+05  -1.895   0.0587 .
## Participation:I(Income_Centered^3) -2.009e+07  1.349e+07  -1.489   0.1372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2974 on 454 degrees of freedom
## Multiple R-squared:  0.05034,    Adjusted R-squared:  0.0357
## F-statistic: 3.438 on 7 and 454 DF,  p-value: 0.001351
```



Run RDD with bandwidth 0.02 and polynomial degree 2

```
##
## Call:
## lm(formula = Support ~ Income_Centered * Participation + I(Income_Centered^2) *
##     Participation, data = gt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.8887 -0.2276  0.1528  0.1706  0.2905
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.76900    0.03445  22.321   <2e-16 ***
## Income_Centered                  -11.56663    7.77729  -1.487   0.1371
## Participation                      0.09286    0.04585   2.025   0.0430 *
## I(Income_Centered^2)             562.24726  372.18239   1.511   0.1310
## Income_Centered:Participation     19.29999   10.44517   1.848   0.0648 .
## Participation:I(Income_Centered^2) -101.10250  500.19558  -0.202   0.8398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3127 on 1942 degrees of freedom
## Multiple R-squared:  0.03629,    Adjusted R-squared:  0.03381
## F-statistic: 14.63 on 5 and 1942 DF,  p-value: 4.245e-14
```