Reproducing Research: Causal Effect of Government Transfers and Government Support

1 Empirical Analysis using Data from Manacorda, Miguel, & Vigorito (2011, American Economic Journal: Applied Economics)

This exercise uses data from Manacorda, Miguel, & Vigorito's paper, "Government Transfers and Political Support," published in the *American Economic Journal: Applied Economics* in 2011. This paper studies how receipt of a government anti-poverty cash transfer changes how beneficiary households support and view the government.

The data can be found on Edward Miguel's faculty website. Download and extract the contents from the `Government_Transfers_replication.zip` file.

2 Set up and constructing the data

The original data used in the paper is confidential. The authors instead provide the `reg_panes.dta` data file which is anonymized and created from the original data.

Open the `reg_panes.dta` file. To complete this problem set you will need the following variables from this data file:

Name	Description
aprobado	Ever received PANES 2005-2007
untracked07	Untracked in 2007
h_89	Supports current government 2007 [1 to 3]
hv34	Supports current government 2008 [1 to 3]
ind_reest	Predicted Income
newtreat	PANES eligibility

2.1 Loading the Packages

Load any R packages we will be using:

Code:

```
library(haven)
library(dplyr)
library(ggplot2)
library(lfe)
library(stargazer)
library(kableExtra)
library(statar)
```

Drop all other variables. If needed, give the variables you are keeping more intuitive names.

Code:

```
df<- read_dta('C:\\Users\\roryq\\Downloads\\reg_panes.dta')%>%
  select(aprobado,untracked07,h_89, hv34, ind_reest,newtreat)
```

2.2 The data as downloaded will require that you clean the variables of interest and construct a new dataset to generate the graphs. Start by generating the following cleaned variable:

-An indicator for receiving PANES that is NA if a respondent is untracked in 2007

Code:

```
df$indicator_panes <- ifelse(df$untracked07 == 1, NA, df$aprobado)
```

2.3 We are going to re-scale the variables that indicate support for the current government so that responses range from 0 to 1. To do this, tabulate the current variable to see how it is distributed and then generate a variable that will be NA if it is currently coded as 9, 0 if currently 2, 0.5 if currently 1 and 1 if currently 3. Do this for both the 2007 and 2008 variable.

Note: This is how the authors modify this variable in their code. It seems counter intuitive and does not correspond to the description of how this variable is coded in the survey questionnaire as reported in their appendix though it does correspond to their discussion in footnote 12. My guess is the transcription/translation of the survey question is incorrect.

Code:

```
table(df$h_89) %>% kbl() %>% kable_classic(full_width = F,
 html_font = "Cambria",latex_options = "hold_position")
```

Var1	Freq
1	397
2	122
3	1570
9	130

```
table(df$hv34) %>% kbl() %>% kable_classic(full_width = F,
 html_font = "Cambria",latex_options = "hold_position")
```

Var1	Freq
1	476
2	159
3	1313
9	103

```
df$h_89_scaled <- ifelse(df$h_89 == 9, NA,
                         ifelse(df$h_89 == 2, 0,
                                ifelse(df$h_89 == 1, 0.5,
                                       ifelse(df$h_89 == 3, 1, NA))))

# Rescale support for the current government in 2008 (hv34)
df$hv34_scaled <- ifelse(df$hv34 == 9, NA,
                          ifelse(df$hv34 == 2, 0,
                                 ifelse(df$hv34 == 1, 0.5,
                                        ifelse(df$hv34 == 3, 1, NA))))
```

2.4 Generate a variable that is the square of predicted income.

Code:

```
df$income_sq<- df$ind_reest
```

3 We start by reproducing the main figures (2, 3,and 4) of the paper as good figures are key to any regression discontinuity paper.

3.0.1 The data consists of over 3000 observations. How many points are plotted on these figures? How should we interpret the y axis? How many points are plotted below the threshold? How many points are plotted above the threshold?

Answer:

There are 45 points, because we will split the data into percentile groups (15 below the threshold and 30 above). Each of these points is the average gov support, and predicted income within each of these percentile groups.

3.0.2 Why is the number of points above the threshold different from the number below?

Answer:

Because many more observations fell below the threshold than those above, so to make sure each percentile consisted of the same number of observations, we needed more percentiles below, and these percentiles are what is graphed. Leading to more obserevations plotted below than above.

3.0.3 Replicating these figures will require restructuring our data and calculating the values that are plotted. Generate a variable that will indicate the percentile group the observation is in. Note the difference in the number of percentile groups above and below the threshold.

Code:

```
df$pct[df$ind_reest<0]<-xtile(df$ind_reest[df$ind_reest<0],n=30)
df$pct1[df$ind_reest>=0]<-xtile(df$ind_reest[df$ind_reest>=0],n=15)
df$pct1<-df$pct1+30
df$pct[df$ind_reest>=0]<-df$pct1[df$ind_reest>=0]
```

3.1 For each of the percentile groups, calculate the mean of each of the variables we will use for plotting: predicted income, receipt of PANES, support for the government in 2007, and support for the government in 2008.

Code:

```
df_summary <- df %>%
  group_by(pct) %>%
  summarize(
    mean_predicted_income = mean(ind_reest, na.rm = TRUE),
    mean_indicator_panes = mean(indicator_panes, na.rm = TRUE),
    mean_support_2007 = mean(h_89_scaled, na.rm = TRUE),
    mean_support_2008 = mean(hv34_scaled, na.rm = TRUE)
  )

df_summary %>% kbl %>% kable_classic(full_width = F, html_font = "Cambria")
```

pct	mean_predicted_income	mean_indicator_panes	mean_support_2007	mean_support_2008
1	-0.0197014	1.0000000	0.9404762	0.9125000
2	-0.0190208	1.0000000	0.9358974	0.9210526
3	-0.0182388	1.0000000	0.8875000	0.7777778
4	-0.0176463	1.0000000	0.8953488	0.8461538
5	-0.0168932	1.0000000	0.9512195	0.9078947
6	-0.0160684	1.0000000	0.9150943	0.8979592
7	-0.0155678	1.0000000	0.9358974	0.8108108
8	-0.0150752	1.0000000	0.9285714	0.8333333
9	-0.0144864	1.0000000	0.9431818	0.8260870
10	-0.0138619	1.0000000	0.9642857	0.8421053
11	-0.0131890	1.0000000	0.8510638	0.9047619
12	-0.0124049	1.0000000	0.8604651	0.8488372
13	-0.0116631	1.0000000	0.8676471	0.8333333
14	-0.0110519	1.0000000	0.9000000	0.7580645
15	-0.0104131	1.0000000	0.8111111	0.8139535
16	-0.0097351	1.0000000	0.9342105	0.8513514
17	-0.0089364	1.0000000	0.8787879	0.8387097
18	-0.0082203	1.0000000	0.8777778	0.8690476
19	-0.0074421	1.0000000	0.9117647	0.7941176
20	-0.0067005	0.9534884	0.8809524	0.7894737
21	-0.0061787	1.0000000	0.9634146	0.8783784
22	-0.0054319	1.0000000	0.8837209	0.8088235
23	-0.0048985	1.0000000	0.9487179	0.8714286
24	-0.0041983	1.0000000	0.8292683	0.8068182
25	-0.0035217	0.9756098	0.9078947	0.8472222
26	-0.0028936	1.0000000	0.8888889	0.7647059
27	-0.0022771	1.0000000	0.8863636	0.9125000
28	-0.0018811	1.0000000	0.9166667	0.8939394
29	-0.0013588	1.0000000	0.8902439	0.8250000
30	-0.0008293	1.0000000	0.8703704	0.8461538
31	0.0009064	0.0000000	0.7698413	0.8189655
32	0.0022822	0.0000000	0.8303571	0.7678571
33	0.0038540	0.0000000	0.7053571	0.6764706
34	0.0052335	0.0294118	0.7923077	0.6545455
35	0.0066122	0.0000000	0.6803279	0.6982759
36	0.0080199	0.0000000	0.7846154	0.6428571
37	0.0091936	0.0158730	0.8416667	0.7641509
38	0.0106284	0.0000000	0.7619048	0.6916667
39	0.0120029	0.0000000	0.8362069	0.8181818
40	0.0132785	0.0163934	0.7777778	0.7549020
41	0.0146722	0.0000000	0.6885246	0.6909091
42	0.0158196	0.0000000	0.8061224	0.7674419
43	0.0171899	0.0000000	0.7651515	0.7419355
44	0.0182551	0.0000000	0.8555556	0.8023256
45	0.0190257	0.0000000	0.7238806	0.6637931

3.2 Replicate figure 2. Make the figure as clear and informative as possible. You may want to create an indicator variable for percentiles above and below the threshold.

Code:

```
df <- df %>%
  mutate(
    above_threshold = ifelse(ind_reest >= 0, 1, 0)
  )

# Create the plot
ggplot(df, aes(x = ind_reest, y = aprobado, color = factor(above_threshold))) +
  geom_point(alpha = 0.6, size = 2) +
  scale_color_manual(values = c("cornflowerblue", "firebrick"),
                     labels = c("Below Threshold", "Above Threshold")) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
  labs(
    title = "PANES Program Eligibility and Participation",
    x = "Predicted Income",
    y = "",
    color = "Eligibility Group",
     caption = "Figure 2") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 14, face = "bold",hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    plot.caption = element_text(size = 10, face = "italic"),
    legend.position = "top"
  )
```

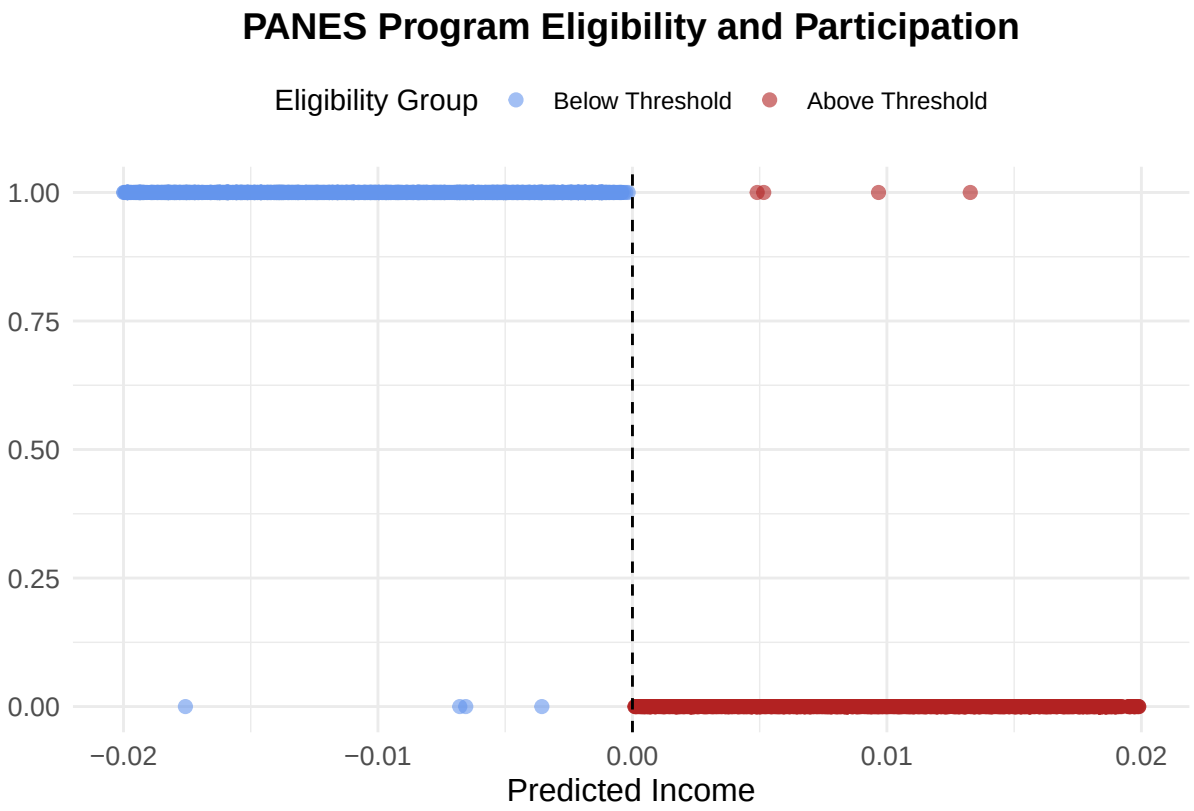
PANES Program Eligibility and Participation
Eligibility Group
Below Threshold
Above Threshold
1.00
0.75
0.50
0.25
0.00
−0.02
−0.01
0.00
0.01
0.02
Predicted Income

Figure 2

3.3 What is the purpose of this figure and what should we take away from it?

Answer: The purpose of this graph is to determine if it is a fuzzy or sharp regression continuity, by evaluating the discontinuity at the threshold between x and treatment.

The sharp discontinuity of those that received panes falling below the threshold, show a strict enforcement of the program eligibility threshold

3.4 Replicate figures 3 and 4. Make these figures as clear and informative as possible.

```
ggplot(df_summary, aes(x = mean_predicted_income, y = mean_support_2007)) +
  geom_point(aes(fill = mean_predicted_income < 0), size = 2, alpha = 0.6, stroke = 0.5,
             shape = 21, color = "black") +
  geom_smooth(data = subset(df_summary, mean_predicted_income < 0),
              aes(x = mean_predicted_income, y = mean_support_2007),
              method = "lm", se = FALSE, color = "cornflowerblue", alpha = 0.6) +
  geom_smooth(data = subset(df_summary, mean_predicted_income >= 0),
              aes(x = mean_predicted_income, y = mean_support_2007),
              method = "lm", se = FALSE, color = "firebrick", alpha = 0.6) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
  labs(
    title = "PANES Program Eligibility and Political Support (2007)",
    x = "Predicted Income",
    y = "Support for Current Government",
    caption = "Figure 3."
  ) +
  scale_fill_manual(values = c("TRUE" = "cornflowerblue", "FALSE" = "firebrick"),
                    labels = c("Below Threshold", "Above Threshold"),
                    breaks = c("TRUE", "FALSE")) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    plot.caption = element_text(size = 10, face = "italic"),
    legend.position = "top"
  ) +
  guides(fill = guide_legend(title = ""))
```

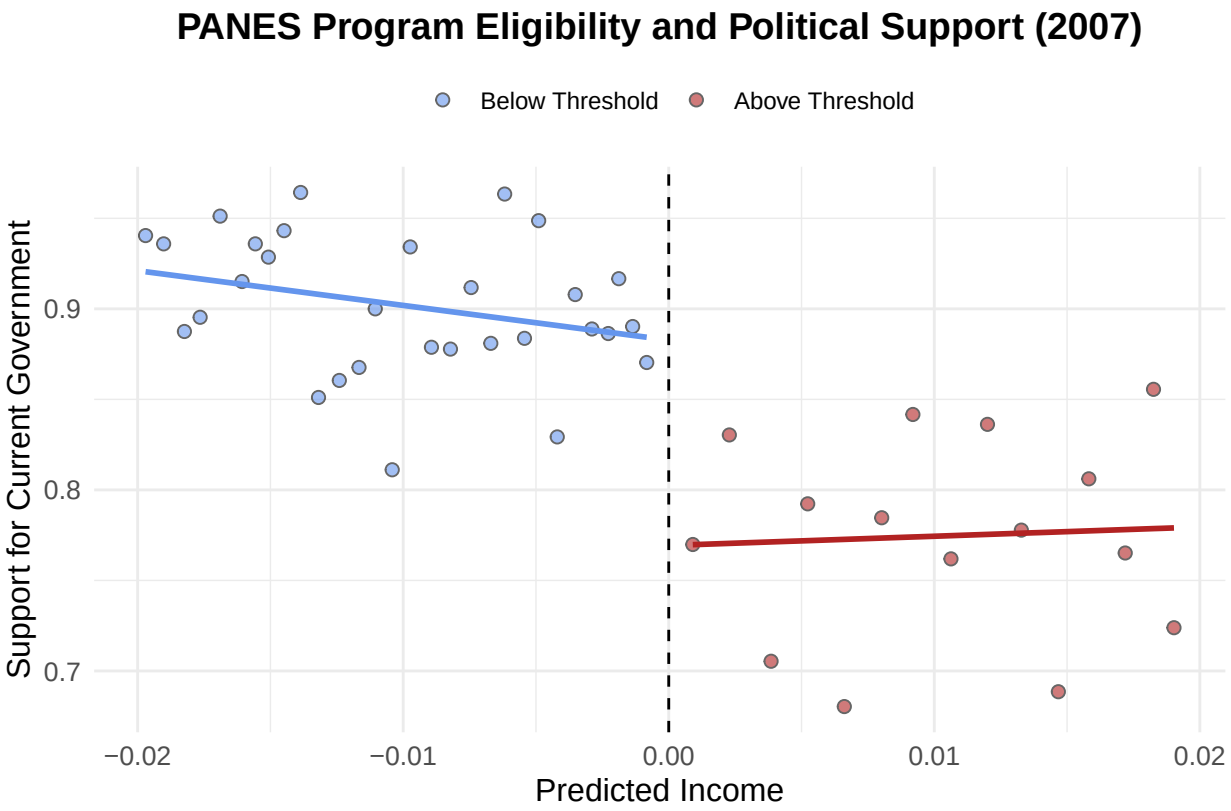
PANES Program Eligibility and Political Support (2007)
Below Threshold
Above Threshold
Support for Current Government
0.9
0.8
0.7
−0.02
−0.01
0.00
0.01
0.02
Predicted Income

Figure 3.

```r
ggplot(df_summary, aes(x = mean_predicted_income, y = mean_support_2008)) +
  geom_point(aes(fill = mean_predicted_income < 0), size = 2, alpha = 0.6, stroke = 0.5,
             shape = 21, color = "black") +
  geom_smooth(data = subset(df_summary, mean_predicted_income < 0),
              aes(x = mean_predicted_income, y = mean_support_2008),
              method = "lm", se = FALSE, color = "cornflowerblue", alpha = 0.6) +
  geom_smooth(data = subset(df_summary, mean_predicted_income >= 0),
              aes(x = mean_predicted_income, y = mean_support_2008),
              method = "lm", se = FALSE, color = "firebrick", alpha = 0.6) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
  labs(
    title = "PANES Program Eligibility and Political Support (2008)",
    x = "Predicted Income",
    y = "Support for Current Government",
    caption = "Figure 4."
  ) +
  scale_fill_manual(values = c("TRUE" = "cornflowerblue", "FALSE" = "firebrick"),
                    labels = c("Below Threshold", "Above Threshold"),
                    breaks = c("TRUE", "FALSE")) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),
    plot.caption = element_text(size = 10, face = "italic"),
    legend.position = "top"
  ) +
  guides(fill = guide_legend(title = ""))
```

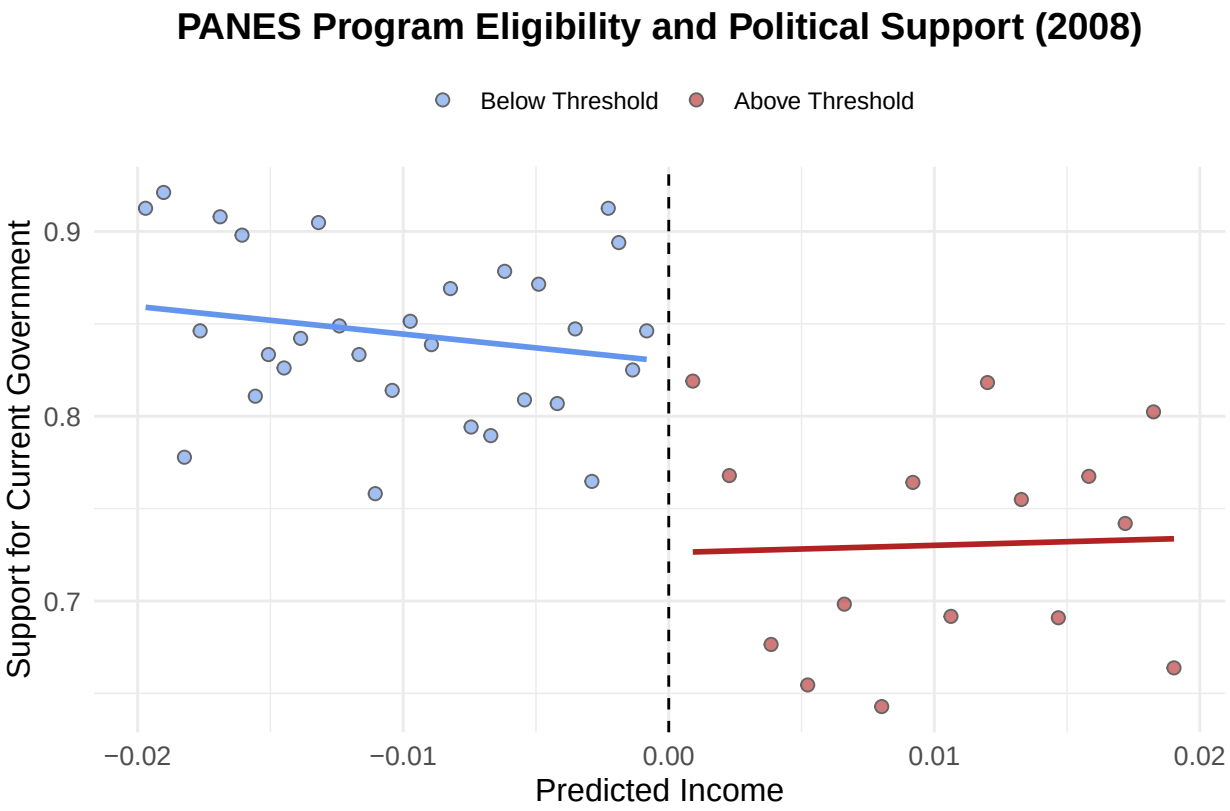
PANES Program Eligibility and Political Support (2008)
Below Threshold
Above Threshold
Support for Current Government
0.9
0.8
0.7
−0.02
−0.01
0.00
0.01
0.02
Predicted Income

Figure 4.

3.5 Interpret these figures. What should we take away from them?

Answer:

Figure 3:

The estimated discontinuity implies that program eligibility was associated with a 13 percentage point increase in support for the government over the opposition coalition. This provides evidence that households' political views are responsive to government transfers

Figure 4:

similar though somewhat smaller gains in FA support—of between 8 and 12 percentage points

Overall: suggests that past transfers also factor meaningfully into voters' decision making.

3.6 Replicate the results of the three regressions estimated in the first column of table 1. Present your results in a table. Interpret the coefficients.

Code:

```
# Regression 1: Ever received PANES (2005-2007)
model1 <- lm(indicator_panes ~ newtreat, data = df)

# Regression 2: Government support (2007)
model2 <- lm(h_89_scaled ~ newtreat, data = df)

# Regression 3: Government support (2008)
model3 <- lm(hv34_scaled ~ newtreat, data = df)


stargazer(model1, model2, model3,
          type = "latex",
          dep.var.labels = c("Ever Received PANES", "Gov. Support 2007", "Gov. Support 2008"),
          covariate.labels = c("Coefficient (Standard Error)"),omit.stat = c("ser"),
          column.labels = c("(1)", "(2)", "(3)"),digits = 3)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Sat, Dec 07, 2024 - 1:54:53 PM

Table 2:

	Dependent variable:		
	Ever Received PANES	Gov. Support 2007	Gov. Support 2008
	(1)	(2)	(3)
	(1)	(2)	(3)
Coefficient (Standard Error)	0.993***	0.129***	0.118***
	(0.002)	(0.012)	(0.014)
Constant	0.004**	0.772***	0.728***
	(0.002)	(0.009)	(0.011)
Observations	2,232	2,089	1,948
R^2	0.987	0.050	0.034
Adjusted R^2	0.987	0.049	0.033
F Statistic	172,740.500*** (df = 1; 2230)	109.220*** (df = 1; 2087)	67.901*** (df = 1; 1946)
Note:			*p<0.1; **p<0.05; ***p<0.01

Answer:

The coefficient of 0.993 suggests that being eligible for the PANES program almost 99.3 percentage point increase in the likelihood of receiving PANES for eligible households).

The coefficient of 0.129 suggests households eligible for PANES had an average 12.9 percentage point higher support for the government in 2007, during the program.

The coefficient of 0.118 suggests that households eligible for PANES had 11.8 percentage points more support for the government in 2008, even after the program had ended.

3.7 Question: Write down the specifications used in row 2 of columns 1,2 and 3 of table 1.

Answer:

Column 1

$$Gov.Support_{07} = \beta_0 + \beta_1 Eligibility + \epsilon$$

Column 2

$$Gov.Support_{07} = \beta_0 + \beta_1 Eligibility + \beta_2 Score + \beta_3 (Eligibility : Score) + \epsilon$$

Column 3

$$Gov.Support_{07} = \beta_0 + \beta_1 Eligibility + \beta_2 Score + \beta_3 Score^2 + \beta_4 (Eligibility : Score) + \beta_5 (Eligibility : Score)^2 + \epsilon$$

3.8 Replicate the results reported in row 2 of Table 1 columns 1, 2 and 3. Explain the difference between these specifications and interpret their coefficients.

Hint: the variables listed in the table above after newtreat are the controls you will want to include.

Code:

```
model1 <- lm(h_89_scaled ~ newtreat, data = df)

# Column 2: Linear score control (Eligibility + Score + Interaction)
model2 <- lm(h_89_scaled ~ newtreat + ind_reest + newtreat:ind_reest, data = df)

# Column 3: Quadratic score control (Eligibility + Score + Score^2 + Interaction)
model3 <- lm(h_89_scaled ~ newtreat + ind_reest + I(ind_reest^2) + newtreat:ind_reest
+ newtreat:I(ind_reest^2), data = df)

# Display the results using stargazer
stargazer(model1, model2, model3,type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Sat, Dec 07, 2024 - 1:54:53 PM

Table 3:

	Dependent variable:		
	h_89_scaled		
	(1)	(2)	(3)
newtreat	0.129***	0.110***	0.130***
	(0.012)	(0.026)	(0.040)
ind_reest		−0.011	0.812
		(1.646)	(6.736)
I(ind_reest^2)			−40.457
			(321.113)
newtreat:ind_reest		−1.916	2.377
		(2.170)	(9.072)
newtreat:I(ind_reest^2)			292.215
			(433.228)
Constant	0.772***	0.772***	0.769***
	(0.009)	(0.019)	(0.030)
Observations	2,089	2,089	2,089
R^2	0.050	0.051	0.051
Adjusted R^2	0.049	0.049	0.049
Residual Std. Error	0.280 (df = 2087)	0.280 (df = 2085)	0.280 (df = 2083)
F Statistic	109.220*** (df = 1; 2087)	37.023*** (df = 3; 2085)	22.353*** (df = 5; 2083)
Note:			*p<0.1; **p<0.05; ***p<0.01

3.9 What is the point of including all of these specifications?

Answer:

No controls gives a simple estimate of how eligibility affects government support in 2007, without accounting for the potential impact of income or other covariates.

A linear control for ind_reest (predicted income), along with an interaction term between newtreat and ind_reest. This helps to assess whether the effect of eligibility on government support varies across different levels of predicted income (linearly).

A quadratic term for ind_reest (ind_reest^2) and its interaction with newtreat. This allows the model to capture any non-linear effects of income on government support and investigate whether the relationship between eligibility and government support changes at different levels of income in a non-linear way.

3.10 Using the coefficients estimated above, write out the function you would use to predict the probability a household supports the current government based on their predicted income score:

a) If they are eligible for the transfer using the results from column 1.

$$Gov.Support_{07} = 0.772 + 0.129(1) = 0.901$$

b) If they are not eligible for the transfer using the results from column 1.

$$Gov.Support_{07} = 0.772 + 0.129(0) = 0.772$$

c) If they are eligible for the transfer using the results from column 2.

$$Gov.Support_{07} = 0.772 + 0.129(1) - 0.011(Score) - 1.916(Score)$$

$$Gov.Support_{07} = 0.901 - 1.927(Score)$$

d) If they are not eligible for the transfer using the results from column 2.

$$Gov.Support_{07} = 0.772 + 0.129(0) - 0.011(Score) - 1.916(Score)$$
$$Gov.Support_{07} = 0.772 - 0.011(Score)$$

e) If they are eligible for the transfer using the results from column 3.

$$Gov.Support_{07} = 0.901 - 0.011(Score) - 40.457(Score^2)$$

f) If they are not eligible for the transfer using the results from column 3.

$$Gov.Support_{07} = 0.901 - 0.011(Score) - 40.457(Score^2)$$

3.11 How narrow is the "bandwidth" used by the authors. Why does this matter? Check that the results are robust to a narrower bandwidth.

Code:

```
cutoff <- 0

# Example of different bandwidths (in terms of score)
bandwidths <- c(0.01, 0.05, 0.15)  # 1%, 5%, 10% bandwidths around the cutoff

# Subset data for each bandwidth
narrow_bandwidth_data <- lapply(bandwidths, function(bw) {
  subset(df, abs(df$ind_reest - cutoff) <= bw)
})
```

```
models <- lapply(narrow_bandwidth_data, function(sub_data) {
  lm(h_89_scaled ~ newtreat * ind_reest + I(ind_reest^2), data = sub_data)
})
```

```
stargazer(models[[1]], models[[2]], models[[3]],
          type = "latex",
          title = "Regression Results for Different Bandwidths",
          dep.var.labels = "Government Support 2007",
          digits = 3,
          model.names = FALSE,
          omit.stat = c("f", "ser"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Sat, Dec 07, 2024 - 1:54:53 PM

Answer:

The author used 2% bandwidth. All of the coefficients are very similar indicating that the results are robust, because the smaller bandwidth estimates are very similar to the larger bandwidth results.

Table 4: Regression Results for Different Bandwidths

	Dependent variable:		
	Government Support 2007		
	(1)	(2)	(3)
newtreat	0.116***	0.110***	0.110***
	(0.037)	(0.026)	(0.026)
ind_reest	−10.985	−2.454	−2.454
	(13.509)	(4.683)	(4.683)
I(ind_reestˆ2)	1,042.423	120.084	120.084
	(1,255.887)	(215.527)	(215.527)
newtreat:ind_reest	19.180	2.967	2.967
	(26.476)	(9.028)	(9.028)
Constant	0.789***	0.781***	0.781***
	(0.035)	(0.025)	(0.025)
Observations	1,007	2,089	2,089
R^2	0.052	0.051	0.051
Adjusted R^2	0.049	0.049	0.049
Note:			*p<0.1; **p<0.05; ***p<0.01

3.12 The authors attribute these effects to the causal effect of receiving the government transfers. What is the implied assumption behind this interpretation?

Answer:

In a regression discontinuity design is that the potential outcome for a unit (household, in this case) that receives the treatment (eligibility for PANES) is continuous around the threshold for eligibility. This means that, had the threshold been slightly different, the outcome for eligible individuals would have been similar to the outcome for ineligible individuals, except for the effect of the treatment.

The eligibility rule creates a "cutoff" where households just above or below the threshold are assumed to be similar in all respects except for their eligibility status. This allows the comparison between households just below and just above the threshold to be interpreted as a randomized experiment near the cutoff.

3.13 What evidence do they provide to support this assumption?

Answer:

"Because the targeting rule was thus insulated from political considerations, and its implementation was remarkably strict, assignment to the pro gram near the threshold is "as good as random." Around 18 months following the start of the program, households with income scores in the neighborhood of the threshold were surveyed and asked a series of questions, including their support for the current government. A second similar fol low-up survey took place in 2008 after the program had already ended. The quasi random assignment of applicants in the neighborhood of the threshold allows us to circumvent the problems of reverse causality, endogenous political targeting, and omitted variables highlighted above, and thus credibly estimate the impact of trans fers on support for the incumbent. "

3.14 Was this threshold eligibility score specifically designed for this particular program? Why does this matter?

Answer:

"The eligibility thresholds varied across the country's five main administrative regions in order to entitle the same proportion of poor households in each area to the program. The regions are: Montevideo, North (Artigas, Salto, Rivera); Center-North (Paysandú, Río Negro, Tacuarembó, Durazno, Treinta y Tres, Cerro Largo); Center-South (Soriano, Florida, Flores, Lavalleja, Rocha); and South (Colonia, San José, Canelones, Maldonado). Only households with predicted income scores below a predetermined threshold were assigned to program treatment."