**POLS 602: Advnaced Quantitative Methods**

**Ani Tepnadze**

**Final Project**

In order to understand what socioeconomic and political factors affect scale variable envreg, which contains variables associated with citizens’ favorability of environmental regulations, I decided to incorporate 5 independent variables in the regression model. Before running the regression, I made some changes to envreg and the independent variables. Detailed changed to the variables are in the raw data file.

The variables that I incorporated in this regression are: envreg , pid7 , ideo5, educ, CC18\_332c

Envreg1 -is a dependent variable which I treated as a continuous variable. I converted it into 7 groups.

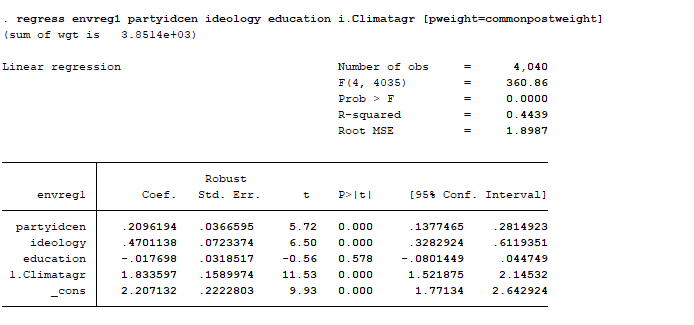
partyidcent- had a likert scale variable and I coded it so that 0-independent, values from 1-3 are in increasing order based on how strong democrat a person is. -3 is equal to strong republican and -1 is equal to lean republican. I expect that higher values of the variable should predict more support for environmental regulations.

Ideology -This variable has 5-point likert scale. I reversed the coding and higher values now indicate liberal view and lower values indicate conservative view. The variable is centered on moderate, which is equal to 0. When I incorporate this variable in the regression, I expect that more 1 unit increase of the value of the variable, should result in an increase of the envreg1 variable.

Education- is a categorical variable. I assume that those who have higher education should support environmental regulations higher. It has 6 categories starting at the lowest value- no high school education to the highest- graduate education.

Climatagr- This is a dummy variable. I assume that those who did not support President Trump’s decision to leave Paris Climate agreement, should be more supportive to environmental regulations than those who supported it. I coded this variable so that 1=those who didn’t support him, 0-those who supported Trump’s decision.

In order to get accurate results, I used weight variable “commonpostweight “, because in the codebook it’s advised to use commonpostweight instead of commonweight if a researcher uses any variables from the post-election survey.



As we can see from the table, some of my expectations appear to be true, while other are not. Coefficient on partyidcen (party id) indicate, that all else equal, one unit increase of the partyid variable, will result in 0.21unit increase in the support for environmental regulations. P value is equal to 0.000 which means that the coefficient is statistically significant. Values of the party identification variable vary from -3 to 3. So, it means that the value of the dependent variable envreg1 might increase maximum by 1.26 units (6\*0.21) on a 7-point scale of envreg1. I would say that the coefficient is substantively significant based on the substantive change in the dependent variable.

The coefficient on variable ideology is positive 0.47 and the p-value is very low, which means that the coefficient is statistically significant. The variable ideology was coded so that the lowest value 1 “Very Conservative” and the highest value is “Very Liberal”. All else equal, 1 unit increase in the ideology is expected to result in 0.47-point increase in supporting environmental regulations. In overall, change of ideology from the lower value to the highest value might increase the dependent variable by 2.82 units. The coefficient is statistically and substantively significant. The magnitude of change is substantively significant. The coefficient is positive and significant, the way I expected before I ran the regression.

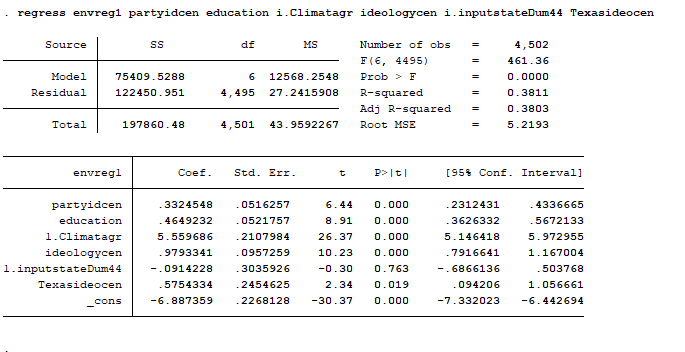
The coefficient on education is -.02 and it’s not statistically significant, because p value is not statistically significant neither at 99% nor at 95%. The magnitude of change is not substantively significant either. Even if the coefficient was statistically significant, changing education from the minimum (no high school) to maximum (graduate) would result 0.1 unit decrease in the dependent variable.

Dummy variable Climatagr (Climate agreement) is a dummy variable. It’s equal to 0 when respondents supported President Trump’s decision to withdraw from the Paris agreement, and 1 when they didn’t. As we can see from the regression, dummy 1 group is expected to support environmental regulation policies 1.83 units higher than dummy 0 group. It’s probably because, Paris climate agreement is very important for the environmental stability and those who do not support it, are expected to not support other environmental regulations as well.

2) The Bureau is interested in if the magnitude of the change independent variables bring are the same in Texas and California. I decided to see if the ideology is equally important in determining the environmental policy support in Texas and California. So, I crated state dummy variables and then incorporated dummy variables for California and Texas in two separate models as interactions with ideology. This way I will see in which state has ideology higher magnitude when it interacts with the state dummy variable.

inputstateDum44- Texas Dummy variable

ideologycen- centered ideology at 0- Moderate



**B4+B6= 0.98(coefficient on ideology) +0.58(Coefficient on the interaction term) =1.56. coefficient on the interaction term is statistically significant as p value (0.02) < 0.05.**

margins inputstateDum44, at(ideologycen=(-3(1)3))

Predictive margins Number of obs = 4,502

Model VCE : OLS

Expression : Linear prediction, predict()

1.\_at : ideologycen = -3

2.\_at : ideologycen = -2

3.\_at : ideologycen = -1

4.\_at : ideologycen = 0

5.\_at : ideologycen = 1

6.\_at : ideologycen = 2

7.\_at : ideologycen = 3

-------------------------------------------------------------------------------------

| Delta-method

| Margin Std. Err. t P>|t| [95% Conf. Interval]

--------------------+----------------------------------------------------------------

\_at#inputstateDum44 |

1 0 | -4.761696 .2985755 -15.95 0.000 -5.347051 -4.176342

1 1 | -4.853119 .4117898 -11.79 0.000 -5.66043 -4.045809

2 0 | -3.782362 .2080356 -18.18 0.000 -4.190214 -3.37451

2 1 | -3.873785 .3510317 -11.04 0.000 -4.56198 -3.18559

3 0 | -2.803028 .1254481 -22.34 0.000 -3.048968 -2.557088

3 1 | -2.894451 .3085491 -9.38 0.000 -3.499359 -2.289543

4 0 | -1.823694 .0807622 -22.58 0.000 -1.982028 -1.66536

4 1 | -1.915117 .2924188 -6.55 0.000 -2.488402 -1.341832

5 0 | -.8443601 .1250391 -6.75 0.000 -1.089498 -.599222

5 1 | -.9357829 .3068255 -3.05 0.002 -1.537312 -.334254

6 0 | .134974 .2075426 0.65 0.516 -.2719116 .5418595

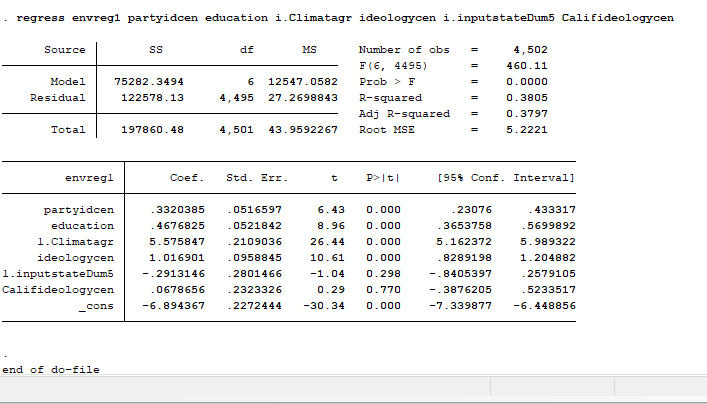
6 1 | .0435511 .3479971 0.13 0.900 -.6386943 .7257966

7 0 | 1.114308 .2980604 3.74 0.000 .5299631 1.698653

7 1 | 1.022885 .407908 2.51 0.012 .2231849 1.822586

As we can see from Margins command, highest value on the Texas dummy variable can get is 1.022885, when ideology is equal to 3 - Strong Liberal.

Regression with Dummy variable for California (inputstateDum5) and its interaction with centered ideology variable.



As we can see, the estimate effect of the interaction between ideology and California state is equal to B4(coefficient on ideologycen) +B6(coefficient on interaction term) = 1.02+0.07 =1.09. Coefficient is not statistically significant; p value is equal to 0.77. As we can see, ideology has higher and statistically more significant magnitude of effect in Texas than in California.

Margins inputstateDum5, at(ideologycen=(-3(1)3))

Predictive margins Number of obs = 4,502

Model VCE : OLS

Expression : Linear prediction, predict()

1.\_at : ideologycen = -3

2.\_at : ideologycen = -2

3.\_at : ideologycen = -1

4.\_at : ideologycen = 0

5.\_at : ideologycen = 1

6.\_at : ideologycen = 2

7.\_at : ideologycen = 3

------------------------------------------------------------------------------------

| Delta-method

| Margin Std. Err. t P>|t| [95% Conf. Interval]

-------------------+----------------------------------------------------------------

\_at#inputstateDum5 |

1 0 | -4.856133 .3000294 -16.19 0.000 -5.444338 -4.267928

1 1 | -5.147448 .3880656 -13.26 0.000 -5.908247 -4.386648

2 0 | -3.839232 .2093698 -18.34 0.000 -4.2497 -3.428765

2 1 | -4.130547 .325412 -12.69 0.000 -4.768515 -3.49258

3 0 | -2.822332 .1266551 -22.28 0.000 -3.070638 -2.574025

3 1 | -3.113646 .282097 -11.04 0.000 -3.666695 -2.560598

4 0 | -1.805431 .0814556 -22.16 0.000 -1.965124 -1.645738

4 1 | -2.096746 .2676792 -7.83 0.000 -2.621529 -1.571963

5 0 | -.7885304 .1249648 -6.31 0.000 -1.033523 -.5435379

5 1 | -1.079845 .2865542 -3.77 0.000 -1.641632 -.5180579

6 0 | .2283703 .2073284 1.10 0.271 -.1780954 .634836

6 1 | -.0629443 .3331098 -0.19 0.850 -.7160033 .5901146

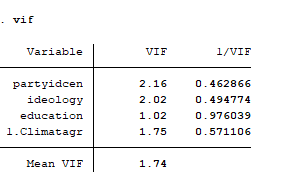
7 0 | 1.245271 .2978954 4.18 0.000 .6612495 1.829292

7 1 | .9539563 .3977419 2.40 0.017 .1741866 1.733726

------------------------------------------------------------------------------------

As we can see from the Margins command, as ideology increases from -3 to 3, from Strong Conservative to Strong Liberal, magnitude of the coefficient on the dummy variable also increases.

In the model, I incorporated both ideology and part affiliation. In order to check if there is multicollinearity, I ran vif command right after the original regression.



As we can see from the table, all vif values are below 10, which means that there is no multicollinearity in the model.

To check if the residuals are normaly distributed with Kernel Density plot



As we can see, residuals are not completely normally distributed, however, they’re approximately normally distributed.

In order to check homoscedasticity of the residuals, I used command rvplot command.



This graph shows that errors are not homoscedastic. We can further run heteroscedasticity test to see if the errors are homoscedastic:

Hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of envreg1

chi2(1) = 387.63

Prob > chi2 = 0.0000

As we can see, p-value is very small, we should reject the hypothesis and accept the alternative hypothesis that the variance is not homogenous. In order to make up for the heteroscedasticity, we can run the robust regression in state, so that outliers cannot contaminate the results.