Nick Hemauer
PLSC 503
4th May 2023

Final Project

Abstract

Previous research (Ojeda 2015) has shown that depression affects various facets of an individual's political behavior. In this data analysis, I build upon my PLSC502 project and explore further into the ANES+GSS2020 data. Previously, my anterior analysis provided evidence to suggest that the most depressed are significantly less likely to vote when compared to the average American. In this analysis, I remodel this hypothesis with logistic regression and include two additional models that explore the further implications of depression. In addition, using R, I test my new additional control variables to show linearity and prove there is no multicollinearity within my models. Finally, I visualize the marginal effects of my original model.

Data

To complete this analysis, I used the joint General Social Survey (GSS) and American National Election Studies (ANES) data. I specifically used the 2020 wave which was conducted during the summer of 2020, during the COVID pandemic, over the phone and online. My final models include 1,084 respondents.

Variables

First I rearranged the reformatted the original data for the needs of this analysis. I merged both datasets by the "YEARID" tag that was given as an identifier for each respondent. Next, I subsetted my data and removed answers that included "don't know" and non-responses. To measure emotional distress, I use the question, "In the past 7 days, how often has R been bothered by emotional problems?" The answers are then scaled from (1) Never, (2) Rarely, (3) Sometimes, (4) Often, and (5) Always. The unfortunate part of this data is that this question is really just a proxy for depression– It is one question that cannot accurately distinguish between those with clinical depression and those with short-term grief (not depression).

Next, I take the question which asks if the respondent voted in the 2020 election. This is a self-reported question and has yet to be matched with voting records. Additionally, I gather the respondent's age, the respondent's highest level of education, self-reported party affiliation (Dem., Rep., Indep.), participation in political discussion

(0-7 days a week), confidence in the political actors within American political institutions (1-100/Bad-Good), and the respondent's feelings towards police (1-100/Bad-Good).

To ensure the linearity of my variables, I test a handful of them using their residuals. Below, in figure 1, shows the plots of four variables. None of these show heteroskedasticity.
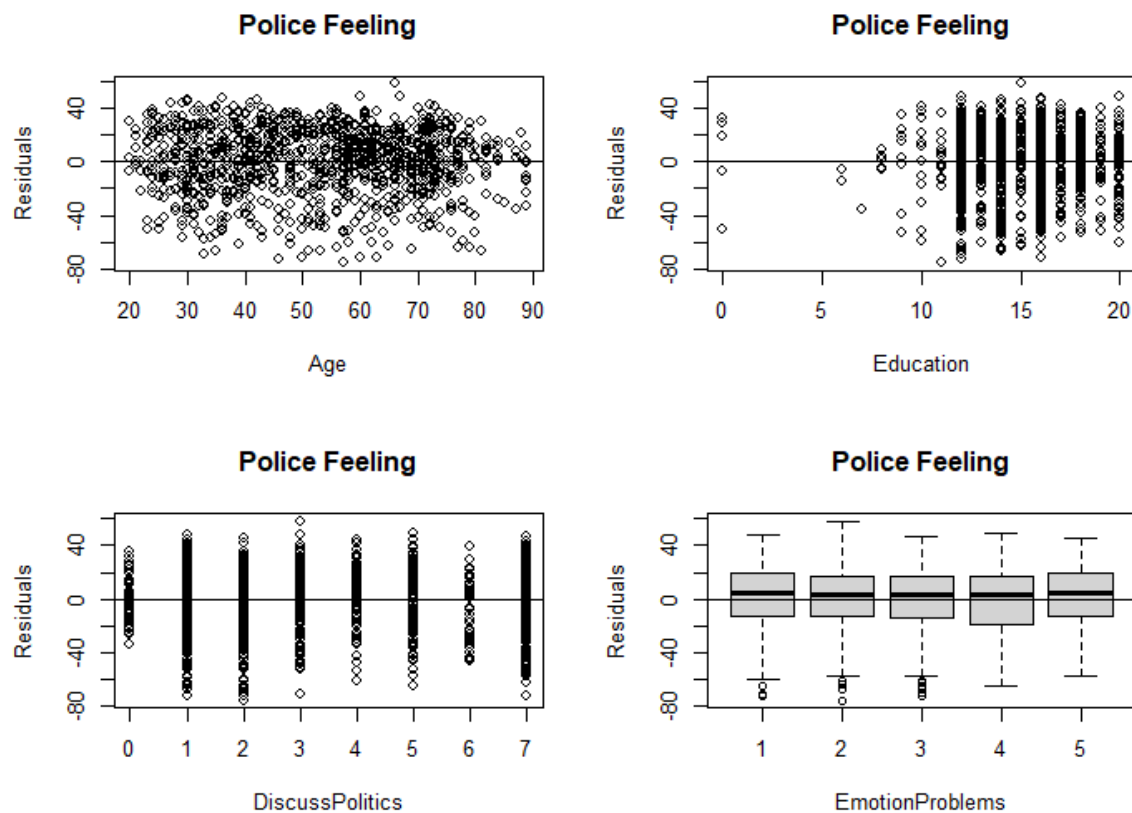

Figure 1. Residual Plots.

Models

As stated previously, I test three models in this analysis. One model uses logistic regression to explain voter turnout in the 2020 election. The second model is an ordinal regression that seeks to show which variables are most likely to impact the dependent variable emotional problems. Finally, the third model shows how respondents' feeling thermometers toward the police are affected by different variables. Below are the three models in full (non-mathematical) form. Some additional variables that would be ideal to include in these models would be a region or a Southernness measure and gender (my data was not being nice). I also do not have access to the income variable, which would be very important to control for.

Vote2020 ~ Emotion + Age + Education + DiscPol + PartyAffiliation + InstitutionConfidence + PoliceFeeling

Emotion ~ Age + Education + DiscPol + PartyAffiliation + InstitutionConfidence + PoliceFeeling

PoliceFeeling ~ Age + Education + DiscPol + PartyAffiliation + InstitutionConfidence + Emotion

Behavior Models

	Dependent variable:	
	Vote2020	EmotionProblems
	logistic	ordered logistic
	(1)	(2)
EmotionProblems2	-0.395	
	(0.326)	
EmotionProblems3	-0.360	
	(0.318)	
EmotionProblems4	-0.293	
	(0.367)	
EmotionProblems5	-0.765*	
	(0.422)	
Age	0.028***	-0.030***
	(0.007)	(0.004)
Education	0.196***	-0.016
	(0.037)	(0.020)
DiscussPolitics	0.219***	0.093***
	(0.048)	(0.024)
PartyAffiliation	0.071	0.039
	(0.099)	(0.053)
InstitutionConfidence	-0.208	0.177*
	(0.164)	(0.093)
PoliceFeeling	-0.001	-0.006***
	(0.004)	(0.002)
Constant	-2.250***	
	(0.824)	
Observations	1,084	1,084
Log Likelihood	-344.221	
Akaike Inf. Crit.	710.442	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 2 shows both the first and second models. First, the logistic regression will be analyzed. As can be seen, the factored emotional problems variable shows there to only be significance amongst those with the highest levels of depression. According to the model, those with severe depression are -0.765 points less likely to have voted in the 2020 presidential election when compared to the average American.

Next, the model provides evidence to suggest that both voter age and education predict voter turnout. With every one-year increase in age, the probability that an American votes increases by 0.028 points. Likewise, every one-level increase in education (i.e. high school diploma -> some college) increases the likelihood of voting by 0.196 units. Finally, for every level increase in the DiscussPolitics variable (i.e. 1 day to 2 days a week), the model predicts that the likelihood of voting increases by 0.219. This makes sense, as political discussion likely predicts political interest and political knowledge, which are highly positively correlated with voting probability.

Figure 2. First two models.

In the second model, my dependent variable is depression which is on an ordinal scale from 0 (no emotional problems) to 5 (constant emotional problems). As seen by the second column, every one-year increase in age is associated with a decrease in 0.03 units of depression. Next, and more interestingly, those who discuss politics the most are significantly more likely to be depressed. The model predicts that with every one-unit increase in political discussion, the average respondent will have an increase in emotional problems by .093 points. This has implications that I have not seen discussed within any current political science literature and is something I will likely explore in further detail.

Next, my model predicts that those with greater confidence in the political actors that run political institutions will be more depressed. For every one-unit increase in the full 100-scale feeling thermometer, emotional problems increase by 0.177. I think that this finding is odd, but I'm unsure. I would expect that those who have the highest feeling thermometer score would likely be the most politically illiterate, and therefore, have a lower amount of emotional problems. My data, however, predicts that those with the highest trust in political actors are the most depressed. I also think it is interesting that education shows no significance in predicting levels of emotional problems, but maybe that is the problem with not including income as a control variable.

Finally, my model predicts that those with greater levels of positivity toward the police will be the least emotionally distressed. In other words, with every unit increase in police positivity, emotional problems decrease by 0.006. This is another really interesting finding. If you hold the belief that the penal system is medieval, as I do, you probably are a more emotional or socially cognizant person. As a result, you are likely to be more emotionally stressed.

To visualize the results of the 2020 vote model. The analysis calculates the margins. Below is table 1 with the estimates, and a plot that visualizes the significant variables.

| factor	AME	SE	z	p	lower	upper
Age	0.0026	0.0006	4.1367	0.0000	0.0014	0.0038
DiscussPolitics	0.0204	0.0045	4.5370	0.0000	0.0116	0.0292
Education	0.0183	0.0034	5.4305	0.0000	0.0117	0.0249
EmotionProblems2	-0.0340	0.0275	-1.2381	0.2157	-0.0879	0.0198
EmotionProblems3	-0.0307	0.0262	-1.1709	0.2416	-0.0821	0.0207
EmotionProblems4	-0.0244	0.0307	-0.7958	0.4262	-0.0845	0.0357
EmotionProblems5	-0.0747	0.0446	-1.6752	0.0939	-0.1620	0.0127
InstitutionConfidence	-0.0194	0.0152	-1.2735	0.2028	-0.0493	0.0105
PartyAffiliation	0.0066	0.0092	0.7182	0.4726	-0.0115	0.0247
PoliceFeeling	-0.0001	0.0004	-0.1951	0.8453	-0.0008	0.0006

Table 1. Marginal effects.
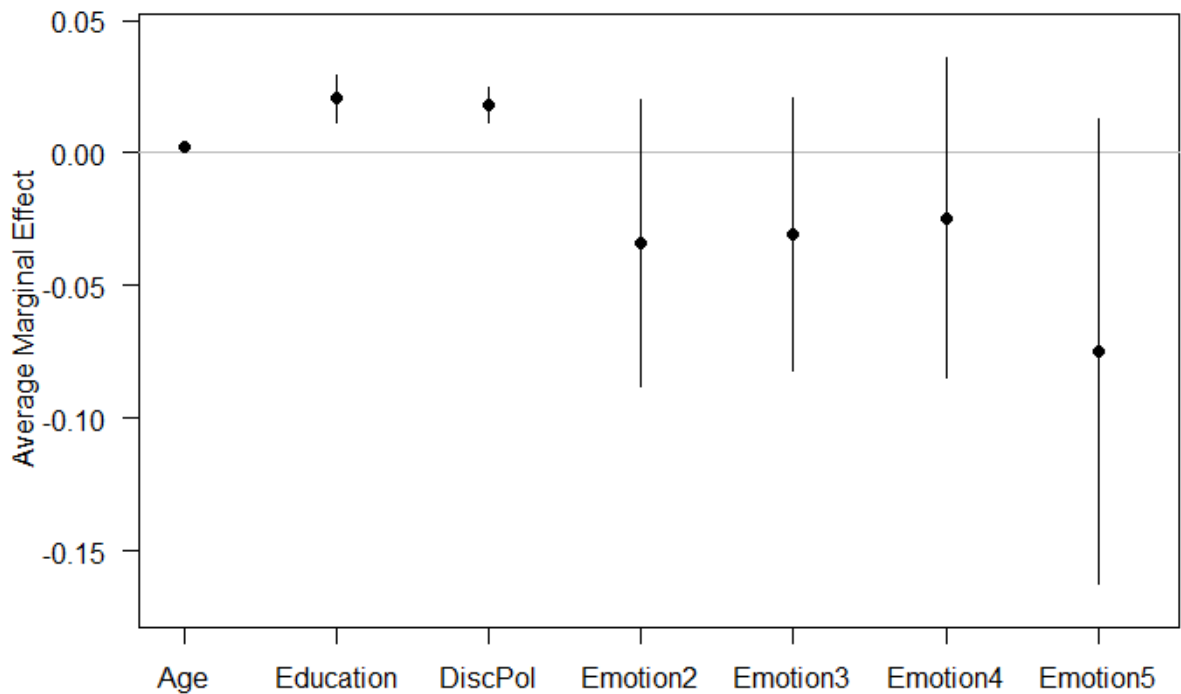
Figure 3. 2020 vote marginal effects.

Following these first two models. I estimated my third model which includes police positivity as the dependent variable. Figure 4 includes the model's estimates. The adjusted R2 value of this model is .138. This means that my model's variables account for 13.8% of the overall variation. It is not a great number, but likely would be improved significantly with the introduction of income, gender, and region as control variables in the model.

Police Feeling 0-100 (Negative to Positive)

	Dependent variable:
	PoliceFeeling
Age	0.464***
	(0.047)
Education	-0.547**
	(0.271)
DiscussPolitics	-0.247
	(0.320)
PartyAffiliation	1.969***
	(0.720)
InstitutionConfidence	-4.443***
	(1.235)
EmotionProblems2	-5.585**
	(2.169)
EmotionProblems3	-5.447**
	(2.120)
EmotionProblems4	-4.851**
	(2.455)
EmotionProblems5	-9.791***
	(3.277)
Constant	60.203***
	(6.030)
Observations	1,084
R^2	0.145
Adjusted R^2	0.138
Residual Std. Error	24.333 (df = 1074)
F Statistic	20.245*** (df = 9; 1074)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 4. Police feeling model.

This model was estimated using OLS linear regression. As can be seen, police feeling (or police positivity) increases with age, decreases with education, and decreases with party affiliation. Specifically, a one-year increase in age is associated with a 0.464 unit increase in police positivity; a one-unit increase in education is associated with a 0.547 unit decrease in police positivity; and a one unit increase in party affiliation (Dem/Indp./Rep.) is associated with a 1.969 unit increase in police positivity. This all logically makes sense.

Next, the model predicts that those with the greatest institutional confidence will like the police the most. The question answers read from 1 (Full trust) to 3 (No trust). Finally, the model predicts that emotionally distressed people are less likely to approve of the police. Specifically, those with the greatest distress are associated with a 9.791 unit decrease in police positivity. This is a really cool finding. I have not seen this in political science literature, but it could be in criminology or psychology literature.

After I created these models, I tested their multicollinearity by using the variance inflation factor (VIF). Below are the VIF tests for each model respectively. As is seen,

the VIF estimations do not reach above five. Therefore, it is not necessary to use a mixed-effects model or alter the methodology. Below are the un-stargazed tables.

```
                                  GVIF
EmotionProblems          1.161644
Age                      1.229723
Education                1.034426
DiscussPolitics          1.043126
PartyAffiliation         1.019055
InstitutionConfidence 1.066762
PoliceFeeling            1.161448
```

Table 2. VIF for 2020 vote model.

```
                                  GVIF
Age                      1.134669
Education                1.062620
DiscussPolitics          1.056828
PartyAffiliation         1.006770
InstitutionConfidence 1.072909
EmotionProblems          1.141663
```

Table 3. VIF for police positivity.

```
                          vif.modelDepression.
Age                                   1.123519
Education                             1.072807
DiscussPolitics                       1.047591
PartyAffiliation                      1.014268
InstitutionConfidence                 1.081344
PoliceFeeling                         1.126612
```

Table 4. VIF for the ordinal depression model.

In sum, this analysis makes use of the ANES and GSS combined data and explores the effects different variables have on political and human behavior. I tested models to predict voting in the 2020 election, depression, and police positivity. Next, I used best practices to assess the data's linearity and the extent of the multicollinearity present. Finally, I extrapolated the model data and visualized it using an R plot.