The Effect of Religious Affiliation on Political Beliefs

Yun-Tsz Tsai

Introduction

This study aims to explore the relationship between religious beliefs and voting behavior. The data being used is from General Social Survey in 2016, which contains 2538 samples and 26 variables.

What will be examine in the analysis is whether religious affiliation affect the possibility of individuals voting for Barack Obama, the candidate from Democratic Party. Traditionally, Democrats are associated with relatively liberal values and it is therefore reasonable to assume that there might be a correlation between individuals' political stance and religious beliefs, which are usually an indicator of the former. In addition to religion, recent studies suggest that women are more liberal in terms of their political ideology. Therefore, I propose two hypotheses as follows:
1) Individuals with less religious affiliation are more likely to support candidates from a more liberal party (In this study, Obama)
2) Women are more likely to support candidates from a more liberal party.

To see how religions and sex interact with voting behavior, I start with the data in 2016 and verify my hypothesis using the same dataset.

Descriptive Table on Sample Attributes

The original dataset contains 33 variables and 2867 observations. However, after ruling out obervations with missing values, there are 1522 obervations and 4 attributes, sex, income, religion and age are used to conduct this analysis. The religion variable contains 4 categories, with Jewish being clustered into `Other`. The income variable is a nominal variable with three levels. Individuals are labeled based on the reported income. The table illustrates individual level attributes and the results are group by their votes. Approximately 63% of the observations voted for Obama and on all recorded attributes, the number of individuals voting for Obama

		Non-Obama Voters			Obama Voters		
		mean	SD	Percent	mean	SD	Percent
	Age	55.49	16.44	37.19	51.74	16.36	62.81
Sex	Male			18.66			25.30
	Female			18.53			37.52
Religion	None			3.75			15.90
	Catholic			8.02			13.40
	Other			1.84			5.65
	Protestant			23.59			27.86
Income	high			5.52			5.91
	low			5.78			16.89
	middle			25.89			40.01
	All			37.19			62.81

is higher than that of those who did not. Interestingly, the ratios of sex among non-Obama voters is rather balanced while among Obama voters, we see ~ 12% more female voters.

Logistic Regression

In this section, I constructed three logistic regression models. Firstly, in Model A, there is religion as the only variable. None is used as the reference category, and this is the same for Model B and Model C. Figures in the table is presented in odd ratios. At first glance, it seems that both Catholic and Protestant are less likely to vote for Obama compared to individuals without any religious affiliation.

In Model B, income is added as another independent variable. The coefficient for Catholic remains the same as in Model A but the coefficient for Protestant decreases slightly from 0.28 to 0.26. Individuals with low and middle income are more likely to vote for Obama compared to individuals with high income.

Finally, I added sex and age into the model and the results are shown under the column Model C. Similar to the results obtained from Model B, the coefficient does not change at all for Catholic. The coefficient for Protestant and dropped slightly while the coefficients for individuals with both income categories increased. The increase is noticeable with individuals with low income. Compared to men, women are more likely to vote for Obama. As for the age effect of voting for Obama, for every one-year increase in age, one becomes less likely to vote for Obama.

Table 1: Probabillity of Voting for Obama

	Model A	Model B	Model C
Intercept	4.25***	2.51***	3.84***
	(0.63)	(0.51)	(1.01)
Catholic	0.39***	0.39***	0.39***
	(0.07)	(0.07)	(0.08)
Other	0.72	0.71	0.64
	(0.19)	(0.19)	(0.17)
Protestant	0.28***	0.26***	0.24***
	(0.05)	(0.04)	(0.04)
Low Income		3.29***	3.41***
		(0.67)	(0.71)
Middle Income		1.63**	1.65**
		(0.28)	(0.29)
Female			1.70***
			(0.19)
Age			0.99***
			(0.00)
Num.Obs.	1522	1522	1522
AIC	1937.6	1901.3	1869.4
BIC	1958.9	1933.3	1912.1
Log.Lik.	−964.781	−944.664	−926.721

+ p $<$0.1, * p $<$0.05, ** p $<$0.01, *** p $<$0.001
high: $>$ 130000

Figure 1: Estimated Odds Ratios from Model C
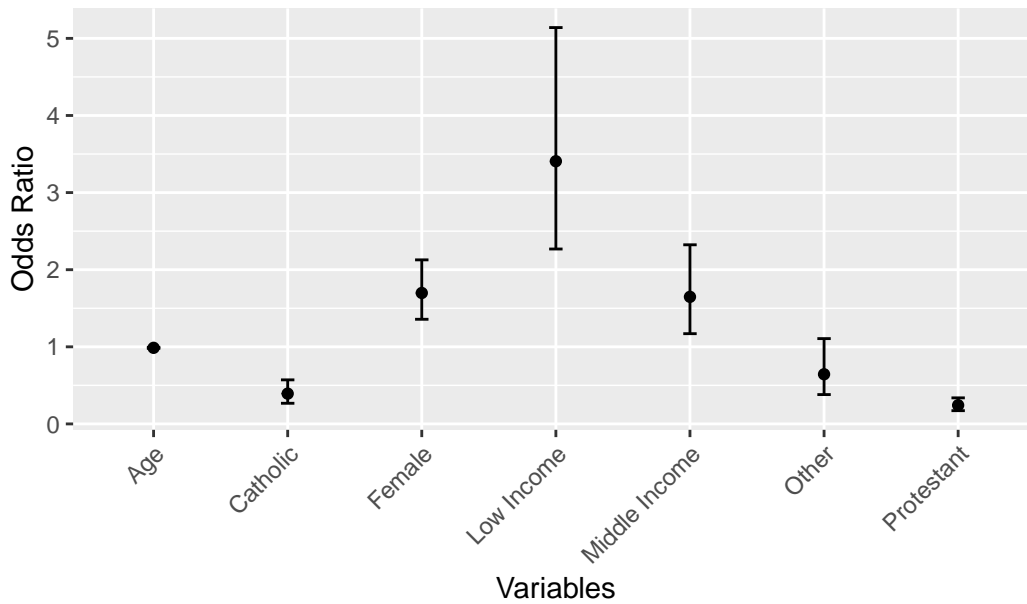
Predicted Probability

For the purspose of demonstration, `Figure 2` shows the predicted values of a female with middle income voting for Obama. The overall tread reamins roughly consistant for all religion categories. Older females are less likely to vote for Obama, and among all religions, Protestant are least likely to vote for Obama. As mentioned earlier, religion might be a proxy for one's political ideology. Take Christianity for example, Catholics are relatively more on the conservative end while Protestants are more on the liberal end. Thus, it seems completely reasonalble to anticipate that being a Protestant makes someone more likely to vote for Obama. However, the predicted results show the opposite.

Figure 2: Probability of A Female with Middle Income
Voting for Obama

Results from Model C

The shadow around the line indicats the confidence interval
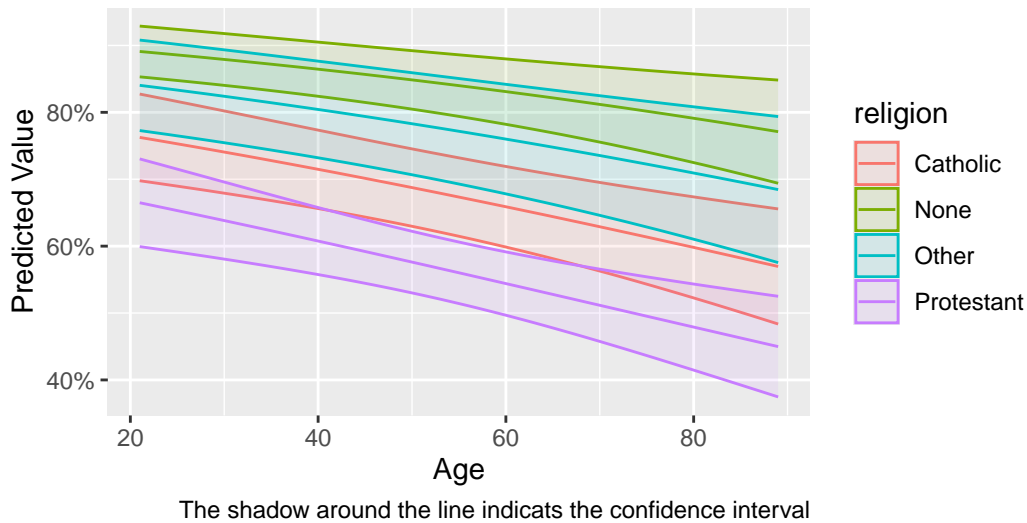
Model Fit

In total, there are three models in this article, Model A, Model B and Model C. Model A is nested in Model B and Model B is nested in Model C. In the following section, I will examine the goodness of fit of the models by looking at three aspects, namely **likelihood ratio test**, **Negelkerke's pseuso's** R^2 and **the share of observations correctly predicted by the models**.

Likelihood Ratio Test

The test is based on comapring two model and the the model being compared is the one that entails the other. The null hypothese is the two model being compared are the same in terms of their ability to predict the data. When comparing Model B to Model A and Model C to Model B, both results are statistically significant, indicating that the former is better than the latter.

Pseudo R^2

The idea of Rsedo R^2 is similar to R^2 used to evaluate linear regression model. There are more than one type of pseudo R^2 and the pseudo R^2 used in this article is the one by Nagelkerke.

Table 3 Model Evaluation

Metrics	Model A	Model B	Model C
Likelihood Ratio Test	NA	40.23***	35.89***
Pseuso R²	0.07	0.10	0.13
Share of Correct Prediction (%)	62.81	64.91	67.35

The Likelihood ratio test for Model is not applicable as it is the most basic model in this analysis. *** P < 0.01

The value can range from 0 to 1, with higher valuing suggesting better performance of the model on prediction. In this study, Model C has the highest score among the three models.

Share of Correct Prediction

Share of correct predition is calculated as the following equation shows:
$P = \frac{the number of correct prediction}{total of observation}$
Model C has the best performance on predicting outcomes.

Conclusion

The impact of religion remain prevelant in all the models. Except for religion, sex, income and age also appear to affect one's voting behavior. Among those, individuals with low income are much more likely to vote for Obama. However, the difference between Catholics and Protestants, despite both being statistically significant, and its impact on voting behavior is realtively small compared to other variables being investigated in this essay. Further studies is needed to clarify the relatioship between religion and actual voting behavior.