What Predicts Ideology in China?

Weihong Qi
wqi3@ur.rochester.edu
University of Rochester

Abstract

This project investigates the predictors of economic, political, and social ideology in China. This project uses the classical linear regression model, LASSO regression model, and random forest regression to find the predictors of ideology in China with the 2018 China Family Panel Survey data. The results show that information accessibility and psychological status are the most important predictors of ideology. Demographics and socioeconomic status are less influential. The three models show similar prediction accuracy, while the random forest regression takes a significantly longer running time.

1 Introduction

Political ideologies in western democracies are well studied due to their importance in predicting election results [1, 3, 4, 7]. But existing literature pays relatively little attention to ideology in authoritarian regimes, despite its significance in influencing policy-making in these states. [10] present China's ideological spectrum using an online survey and find a multi-dimensional ideology spectrum that is quite different from the left-right division in the United States. However, there is no further research on what shapes people's ideologies nor the implications on the governance style of authoritarian regimes witnessing how public preference can be shaped.

This project intends to discover the predictors of ideology in China using data mining techniques. According to [2], [10], and [11], I study three dimensions of ideology in China: the attitude toward authoritarian rule, the attitude toward intervention in the economy, and the attitude to traditional social values. The three dimensions capture the political, economic, and social ideology of Chinese citizens. Regarding the predictors, I will focus on four categories based on current findings in political science literature[3, 6] and the available data: demographics, socioeconomic status, psychological status, and information accessibility.

Objective and subjective features in deciding ideology have long been an important topic in political science research [5]. Some scholars find that demographics and socioeconomic status have significant effects on ideology [8, 9], while others find values, principles, and psychological status is influential features [3, 6]. Demographics and socioeconomic status are mostly related to individual experiences and social identity, which form individuals' views of the world and how they treat others. While information accessibility and psychological status influence the content of information an individual obtains and how she interprets such information. Based on the literature, this project restricts attention to whether objective features (demographics and socioeconomic status) or subjective features (information accessibility and psychological status) are more important in determining ideology in China.

2 Data processing

2.1 Data source

The data to be used by this project is the 2018 China Family Panel Studies (CFPS). CFPS was launched by Peking University in 2010. The survey collects individual-, family-, and community-level longitudinal data in contemporary China every two years. The 2018 CFPS survey questionnaire includes questions capturing 32,669 individuals' attitudes that will be ideal to use in this project. The three measurements of attitude are: "(rate of) severity of government corruption", "(whether the interviewee believes) economy prosperity must induce greater inequality" and "the importance of having a son to carry on the family name". Table 1 presents the summary statistics of the three measurements.

2.2 Attribute selection

First, I removed redundant attributes in the data set. By redundant, it is particularly referred to the attributes that are unrelated to the research purpose of this project or attributes that are not applied to most of the respondents (and thus coded as missing). Within the 2018 CFPS adult survey data, there are many attributes that capture the information of the survey process. For example, "whether the survey was completed", "whether answered the information about the first full-time job" and "interview year and month". As the attributes do not contain information related to this project, I removed the attributes from the data. Besides, there are many attributes that are not applied to most of the respondents. For instance, the question asking about the relationship between the interviewee and his or her fifth child is mostly coded as "not applied" or missing because most of the interviewees do not have that many children. As this project restricts attention to the ideology of Chinese adults, the questions asked to children are also removed. In the end, there are 77 attributes including two identification variables under study. The detailed list of attributes under study is in Appendix A.

2.3 Missing data

Some attributes bring significant attrition in data. Fortunately, the CFPS survey provides the family ID to identify interviewees from the same family. Considering the spillover effects within households, it is reasonable to fill the missings with the family average. For attributes that have more than 3000 missings (around 10% of the sample size), I fill the missings with family means. However, even though filling missings with the family means, there are two attributes that still cause significant attrition to the data: "Get political information on the internet" and "Make political comments on the internet". The possible cause of the large-scale missing is that the questions are considered "sensitive" in China and people are reluctant to answer the questions. To avoid the attrition of the data while reflecting their importance in predicting ideology, I remove the two attributes in the main models.

2.4 Data Transformation

First, I transformed all the Boolean variables to be 0 or 1 format. For example, the "whether believe god or not" variable is initially coded as 5 for "no" and 1 for "yes". I transformed it to be 0 for "no" and 1 for "yes". Then I encoded all the categorical variables, such as "province", to be Boolean variables. To be specific, for each province, I created a Boolean variable with the province name and coded it as 1 if the interviewee lives in the province and 0 if not.

3 Models and algorithms

This project uses the supervised learning method to find the predictors of the ideologies. To be specific, I use the linear regression model (OLS), LASSO regularized model, and random forest regression (RF) to find the predictors and intends to compare the performance of the models. The OLS models are the most frequently used models in political science research. However, when having high dimensional data, a traditional linear regression model can induce the over-fitting problem. Thus I will also use the LASSO regression models to reduce the dimensionality and better address the bias-variance trade-off. However, LASSO regression models still require all the assumptions of linear regression to make precise predictions, including the assumptions of data distributions. Random forest, on the contrary, does not have assumptions of data distribution. Note that although the ideological variables are discrete, they are also ordinal. Thus I regression models rather than classification.

3.1 Linear regression

The OLS model can be written as:

$$y_i = x_{i1}\beta_1 + ... + x_{iK}\beta_K + \epsilon_i$$

The OLS model aims to minimize the sum of squared residuals (SSR), which is the squared deviation of predicted values from the true values. In other words, OLS fitting results minimize:

$$\Sigma_i = (y_i - \hat{y}_i)^2$$

To make the estimation unbiased, the classical OLS model requires the following assumptions:

- Full rank: the sample data matrix has full column rank
- Exogeneity: $E[\epsilon_i|X_j] = 0,\ i, j = 1, ..., n$
- Homoscedasticity: Each ϵ_i has the same finite variance and is uncorrelated with every other ϵ_j conditional on $\mathbf{X}$.
- Normal distribution: the ϵ are normally distributed.

Although the classical OLS model is the easiest to understand and operate, two risks cast OLS models into problems. First, the assumptions about data distribution are quite strong and we do not have methods to well test the distribution for each attribute with a large number of predictor candidates. Second, adding so many attributes into OLS regression is possible to induce the over-fitting problem. In addition to using the cross-validation method to avoid the problem, other models are necessary for robustness checks for results and model comparison.

3.2 LASSO regression

LASSO regression performs feature selection and regularization to improve prediction accuracy. Instead of minimizing the SSR, LASSO regression minimizes:

$$\Sigma_i = (y_i - \hat{y}_i)^2 + \lambda\Sigma_j|\beta_j|$$

Where λ is the tuning parameter to decide which coefficients are shrunk to 0 and avoid the problem of over-fitting.

3.3 Random forest regression

Random forest is an ensemble learning method, including regression and classification. RF algorithm first picks K random data points from the training data and builds a decision tree associated with the points. Then repeat the previous step by N times to construct N trees. For each of the new data points, predict the value of the dependent variable (which is ideology in this project) based on each of the trees and assign the mean of the predictions to the new data point as the predicted value. RF does not make assumptions as classical OLS and avoids the risk of throwing true predictors by accident. But this algorithm requires more computational resources because the tree construction process is more computationally costly than the other two methods.

To make the best predictions, I combine the results of the three models by looking at the overlap identified predictors.

4 Data analysis and results

4.1 Summary statistics

As discussed earlier, I study three dimensions of ideology in China and the survey questions are:

- Economic ideology: "(whether the interviewee agrees that) economy prosperity must induce greater inequality"
- Political ideology: "(rate) the severity of corruption of the government "
- Social ideology: "the importance of having a son to carry on the family name "

The economic ideology variable captures people's attitudes toward intervention in the economy. If the individual believes economic prosperity causes severe inequality, she may support more intervention in the economy, which is a conservative economic ideology.

The political ideology variable captures people's subjective feelings about government corruption. Although the true level of corruption is not observable, how people evaluate the same problem while living in the same society reveals their attitude toward the government. If the individual perceives more corruption, she tends to oppose authoritarian rule, which is a liberal political ideology.

In a society with the tradition of primogeniture, whether a family treats sons and daughters differently can be a signal of their conservatism. Conservative people regard having a son carrying the family name as an important issue in their life. Thus more emphasis on having a son is related to a conservative social ideology.

Table 1 presents the summary statistics of the three ideological variables. I keep the initial code of the three variables. The ranges of economic ideology and social ideology are [1,5] with a higher number meaning the issue is more important to the interviewee. Political ideology ranges from 0 to 10 with a higher number meaning severer corruption.

Figure 1-3 plots the distribution of the ideological variables. It is easy to see that economic ideology is a uni-modal and symmetric distribution, while social ideology reveals a uni-modal and left-skewed distribution. This implies most people are moderate in economic ideology but more conservative in social ideology. The political ideology is a bit more complex. It reveals bi-modal with the first peak around 5 and the second around 10. There could be two large groups of people: while a moderate group considers corruption in China not to be a severe problem, the other group considers it to be extremely severe.

Table 1. Summary statistics of ideological attributes

Statistic	N	Mean	St. Dev.	Min	Max
Economic ideology	32,387	2.589	0.810	1	5
Political ideology	29,556	6.647	2.796	0	10
Social value ideology	30,127	4.115	1.124	1	5

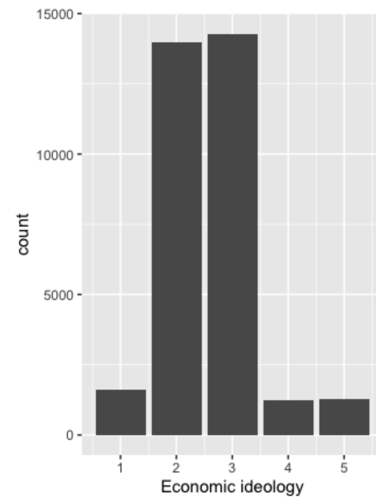

Figure 1. The distribution of economic ideology score
Note: The economic ideology score ranges from 1 to 5, with 1 being most liberal and 5 being most conservative.

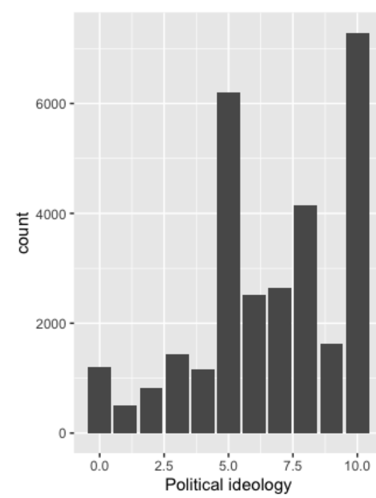

Figure 2. The distribution of political ideology
Note: the political ideology score ranges from 0 to 10, with 0 to be most conservative and 10 to be most liberal.

4.2 Predictors of ideology

I present the OLS regression results in the Appendix. The OLS model reports some strong predictors of ideologies, such as whether to use a mobile phone, whether to use the internet, whether trust other people and educational level.

Table 2 presents the results of the LASSO regression model. This algorithm drops a lot of attributes by making the coefficient shrink to 0. The model consistently shows some of the attributes having a significant influence on ideology, such as gender, the use of computers, and marital status.

Figure 4-6 presents the top important variables identified by the RF algorithm. The importance is measured by Mean

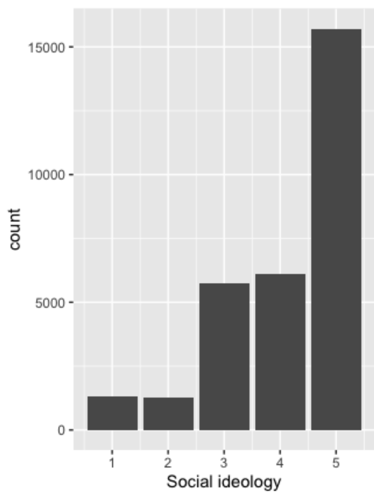

Figure 3. The distribution of social ideology
Note: The social ideology score ranges from 1 to 5 with 1 representing the most liberal and 5 the most conservative

Decrease Gini Index (IncNodePurity). The RF models indicate the importance of trust and individual health conditions in predicting ideology [1].

Note that while the models identify different predictors, the signs of coefficients are pretty robust. This suggests we can also infer more reliable information from the signs of frequently identified predictors.

To find the overlapping predictors identified by different models, I summarize the identified predictors in each ideology by at least two models as follows:[2]

OLS and LASSO:

- Economic: computer, reading
- Political: mobile phone, wordtest
- Social: gender, reading, wordtest

OLS and RF:

- Economic: education, income, trust cadre, trust doctor, trust neighbor
- Political: weight, height, wordtest, trust cadre, trust doctor, trust neighbor, interpersonal relation
- Social: age, confidence, the expected number of sons, wordtest, education, satisfaction, interpersonal relation

LASSO and RF:

- All three dependent variables overlap with the same attribute wordtest

[1]To test the robustness of the RF model, I report the full results using a different random seed in the Appendix. The different random seeds identify similar predictors, which confirms the robustness of this RF model
[2]For the predictors identified by the RF model, I focus on the top 10 important predictors to find the overlappings

Table 2. Predictors identified by LASSO regression model

Category	Attributes
Economic ideology	gender (-0.0063), mobile phone (-0.0689) computer (-0.1999), reading(-0.1841) wordtest(-0.0043)
Political ideology	gender(0.1330), mobile phone (0.6850) computer (0.3454), reading (0.0660) wordtest(0.0120) never married (-0.4676), married (0.0821) common-law partner (-0.3117) divorced (0.2728) widowed (-0.2349)
Social ideology	gender(0.1802), mobile phone (-0.0960) computer (0.2504), reading (-0.3074) wordtest (-0.0040) never married (-0.3271), married (0.0453) common-law partner (-0.2900) divorced (-0.3414) widowed (0.0472)

Note: This table presents the significant contributors predicted by the LASSO regression model. The coefficients are in parentheses.

To find the strongest predictors among all ideologies, I focus on the attributes identified by at least two models and show up at least two times:

- Demographics: education (2)
- Socioeconomic status: null
- Information accessibility: reading (2), wordtest (4)
- Psychological status: trust cadre (2), trust doctor (2), trust neighbor (2), interpersonal relation (2)

Surprisingly, the most frequently identified predictor among all categories is the word test score in 2018. The word test score ranges from 0 to 34, giving 1 credit to the interviewee when he or she read the word correctly. This attribute is positively correlated with educational level, but as the educational level does not show the same strong power of prediction, there are other factors associated with the word test score that influences people's ideology. Notice that the word test score is positively associated with political ideology (more corrupted) and negatively associated with social ideology (less important to have a son). This is a signal that people with higher word test scores are more liberal.

There is at least one factor associated with the word test score that makes people more liberal: the accessibility of information. People who have the better reading ability are likely to be exposed to more information flows, including books, journals, the internet, and other information channels. While the information channels are available to all people, it

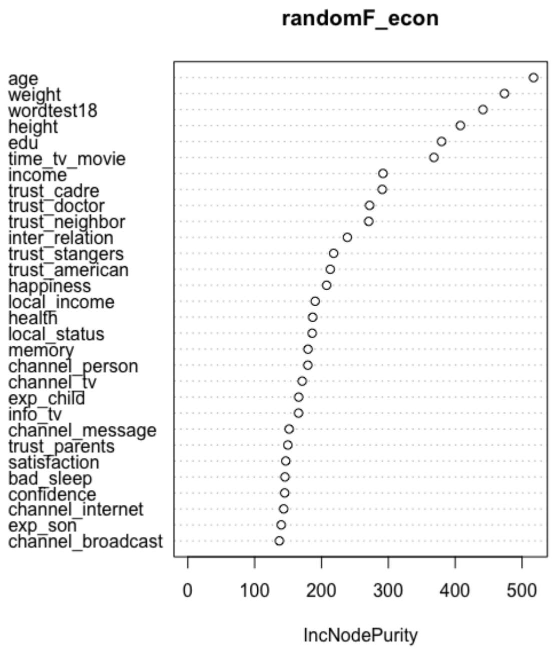

Figure 4. Random forest variable importance of economic ideology

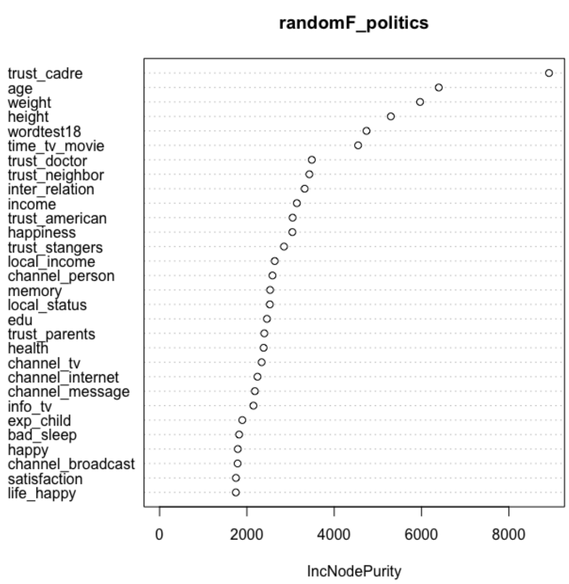

Figure 5. Random forest variable importance of political ideology

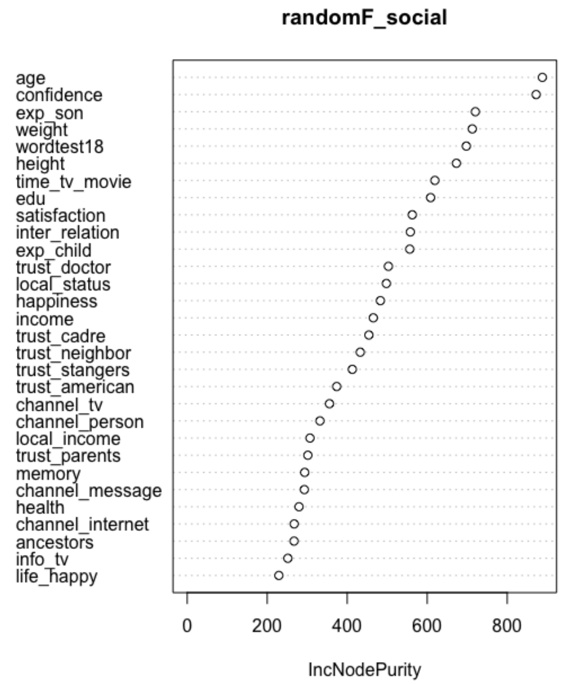

Figure 6. Random forest variable importance of social ideology

requires certain reading and searching ability to access and digest information effectively. Further, although the word test is conducted in Chinese, it is reasonable to believe that reading ability in Chinese is positively related to the ability to read foreign languages. In an authoritarian regime where censorship is prevalent, foreign languages, especially English, are likely to provide more liberal and critical perspectives.

Another influential information accessibility feature is reading habits. If an individual has reading habits, she is more likely to have a liberal economic and social ideology.

The other set of predictors is the subjective evaluation of interpersonal relations quality and the trust of non-family members. The trust of non-family members, such as neighbors, cadres, and doctors, predicts a more conservative economic ideology and social ideology.

The self-evaluated interpersonal relation quality shows a slightly different tendency. The interpersonal relation quality is positively associated with political and social ideology. This indicates that better interpersonal relation quality predicts a more liberal political ideology and a more conservative social ideology.

Educational level is the most powerful demographic attribute. A better educational level predicts a more liberal economic and social ideology but has no significant influence on political ideology.

It is interesting to notice that weight and height are identified as predictors of political ideology. If this is not a noise, one explanation is that people who are in good nutritional conditions are more likely to be politically conservative.

The takeaway from all the results is that psychological status and information accessibility are strong predictors of ideology in China, while demographics and socioeconomic status are not good indicators. Better information accessibility is more likely to form a liberal ideology but better psychological status and interpersonal quality are more likely to form a conservative ideology[3].

5 Model evaluation

I use the root mean squared errors (RMSE) to measure the accuracy of prediction. The RMSE is given by:

$$RMSE = \sqrt{\Sigma_i \frac{(\hat{y}_i - y_i)^2}{n}}$$

Table 3. Model evaluation

Model	RMSE	Time (seconds)
Economic (OLS)	0.7661	0.542
Political (OLS)	2.6672	0.510
Social (OLS)	0.9727	0.465
Economic (LASSO)	0.7935	0.096
Political (LASSO)	2.7688	0.106
Social (LASSO)	1.0830	0.095
Economic (RF)	0.7748	130.226
Political (RF)	2.6926	127.319
Social (RF)	0.9745	154.782

Table 3 presents the model evaluation results. Among the three models, OLS has the best performance. This indicates that LASSO and RF are not making better predictions than OLS. Further, although LASSO takes the least time, it over drops the candidate predictors and makes the performance worst among all the three models. In fact, LASSO has the least overlapping results with OLS and RF, suggesting it does not provide robust results. RF model has a significantly longer running time because the algorithm generates 100 trees to make the prediction.

6 Conclusions and discussions

In conclusion, psychological status and information accessibility are strong predictors of ideology in China. Among all the predictors, the word test score in 2018 is the most important attribute. The strong predicting power of word test score remarks the importance of the ability to access, digest, and interpret information in forming individual ideology. Demographics and socioeconomic status are not good indicators of ideology. The only strong predictor in this category is education. Educational level predicts a liberal ideology.

[3]The only exception is that better interpersonal relation predicts more liberal political ideology

This research uncovers the strong predictors of ideology in China. As there are several attributes identified, the specific mechanism that each attribute affects ideology is not revealed with empirical evidence. Future research can focus on revealing the mechanism by which psychological status and information accessibility are transformed into ideology.

References

[1] Sarah F. Anzia, Jake Alton Jares, and Neil Malhotra. 2022. Does Receiving Government Assistance Shape Political Attitudes? Evidence from Agricultural Producers. *The American political science review* 116, 4 (2022), 1389–1406.

[2] Davide Cantoni, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y. Jane Zhang. 2017. Curriculum and ideology. *Journal of political economy* 125, 2 (2017), 338–392.

[3] Edward G Carmines and Nicholas J D'Amico. 2015. The New Look in Political Ideology Research. *Annual review of political science* 18, 1 (2015), 205–216.

[4] Christopher Ellis. 2012. *Ideology in America / Christopher Ellis, James A. Stimson.* Cambridge University Press, Cambridge.

[5] Paul Gill, John Horgan, and Paige Deckert. 2014. Bombing Alone: Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorists. *Journal of forensic sciences* 59, 2 (2014), 425–435.

[6] Peter K Hatemi and Rose McDermott. 2016. Give Me Attitudes. *Annual review of political science* 19, 1 (2016), 331–350.

[7] Stephen A Jessee. 2012. *Ideology and Spatial Voting in American Elections.* Cambridge University Press, New York, NY. xiii–xiii pages.

[8] Alan B Krueger and Jitka Maleckova. 2003. Education, poverty and terrorism: Is there a causal connection? *The Journal of economic perspectives* 17, 4 (2003), 119–144.

[9] Alexander Lee. 2011. Who Becomes a Terrorist?: Poverty, Education, and the Origins of Political Violence. *World politics* 63, 2 (2011), 203–245.

[10] Jennifer Pan and Yiqing Xu. 2018. Chinaâs Ideological Spectrum. *The Journal of politics* 80, 1 (2018), 254–273.

[11] Yang You, Weihong Zeng, David Y Yang, Alberto F Alesina, and Marlon Seror. 2020. Persistence Despite Revolutions. *NBER Working Paper Series* (2020).

A Appendix

Table 4. OLS Regression Results

	Dependent variable:		
	ideology_economic	ideology_political	ideology_social
	(1)	(2)	(3)
hukou	-0.050^{***}	0.056	-0.008
	(0.013)	(0.045)	(0.016)
gender	0.019	0.052	0.128^{***}
	(0.019)	(0.065)	(0.024)
age	0.0005	0.003	-0.001^{*}
	(0.001)	(0.002)	(0.001)
mobile_phone	0.010	0.174^{**}	0.038
	(0.024)	(0.084)	(0.030)
mobile_net	-0.033^{*}	0.216^{***}	-0.075^{***}
	(0.017)	(0.060)	(0.022)
computer	-0.035^{*}	-0.070	-0.033
	(0.020)	(0.068)	(0.025)
channel_tv	0.003	0.024	0.031^{***}
	(0.005)	(0.017)	(0.006)
channel_internet	-0.007	0.089^{***}	-0.010
	(0.005)	(0.018)	(0.007)
channel_paper	0.013^{*}	-0.043^{*}	-0.007
	(0.007)	(0.023)	(0.008)
channel_broadcast	-0.001	0.050^{**}	0.016^{**}
	(0.006)	(0.021)	(0.008)
Observations	18,053	17,878	18,047
R^2	0.080	0.116	0.191
Adjusted R^2	0.075	0.111	0.186
Residual Std. Error	0.764 (df = 17952)	2.657 (df = 17776)	0.970 (df = 17945)
F Statistic	15.612^{***} (df = 100; 17952)	23.060^{***} (df = 101; 17776)	41.923^{***} (df = 101; 17945)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5. OLS Regression Results (Continued)

	Dependent variable:		
	ideology_economic	ideology_political	ideology_social
	(1)	(2)	(3)
channel_message	−0.002	0.055***	0.005
	(0.005)	(0.017)	(0.006)
channel_person	−0.005	0.020	0.023***
	(0.005)	(0.017)	(0.006)
inter_relation	−0.009**	0.080***	0.032***
	(0.003)	(0.012)	(0.004)
happiness	−0.002	0.027**	0.010**
	(0.003)	(0.012)	(0.004)
altruistic	−0.020	−0.204***	0.010
	(0.014)	(0.048)	(0.017)
trust	−0.027**	−0.275***	−0.102***
	(0.013)	(0.045)	(0.016)
trust_parents	−0.014***	0.123***	0.032***
	(0.004)	(0.015)	(0.006)
trust_neighbor	0.006*	0.023**	0.013***
	(0.003)	(0.011)	(0.004)
trust_american	0.002	0.023**	−0.004
	(0.003)	(0.009)	(0.003)
trust_stangers	−0.006**	0.010	−0.029***
	(0.003)	(0.011)	(0.004)
trust_cadre	0.007***	−0.236***	0.001
	(0.003)	(0.009)	(0.003)
trust_doctor	0.006**	0.021**	0.007*
	(0.003)	(0.010)	(0.004)
satisfaction	0.010	−0.076***	0.066***
	(0.007)	(0.026)	(0.009)
confidence	0.016**	0.046*	0.119***
	(0.007)	(0.026)	(0.009)
Observations	18,053	17,878	18,047
R^2	0.080	0.116	0.191
Adjusted R^2	0.075	0.111	0.186
Residual Std. Error	0.764 (df = 17952)	2.657 (df = 17776)	0.970 (df = 17945)
F Statistic	15.612*** (df = 100; 17952)	23.060*** (df = 101; 17776)	41.923*** (df = 101; 17945)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6. OLS Regression Results (Continued)

	Dependent variable:		
	ideology_economic	ideology_political	ideology_social
	(1)	(2)	(3)
local_income	0.033***	−0.073***	0.015*
	(0.006)	(0.023)	(0.008)
local_status	0.025***	−0.043*	0.062***
	(0.007)	(0.023)	(0.008)
buddha	−0.012	0.053	0.055***
	(0.015)	(0.053)	(0.019)
god	0.069***	0.017	−0.007
	(0.018)	(0.063)	(0.023)
Allah	0.045	−0.668***	−0.033
	(0.034)	(0.120)	(0.044)
christ	0.046	0.389***	−0.053
	(0.036)	(0.126)	(0.046)
jesus	−0.016	−0.142	0.071*
	(0.030)	(0.103)	(0.038)
ancestors	−0.004	−0.024	0.164***
	(0.013)	(0.046)	(0.017)
ghost	−0.015	0.058	−0.088***
	(0.021)	(0.074)	(0.027)
fengshui	0.034**	0.152***	0.048***
	(0.013)	(0.047)	(0.017)
info_tv	−0.007***	−0.001	−0.005*
	(0.002)	(0.008)	(0.003)
party_member	0.020	−0.322	0.063
	(0.068)	(0.237)	(0.086)
religion_member	0.036	0.310**	0.036
	(0.035)	(0.123)	(0.045)
union_member	−0.084***	−0.124	−0.089***
	(0.025)	(0.089)	(0.032)
individual_worker_union	0.021	−0.199**	0.078**
	(0.026)	(0.091)	(0.033)
donation	−0.024	0.137**	−0.005
	(0.015)	(0.053)	(0.019)
Observations	18,053	17,878	18,047
R^2	0.080	0.116	0.191
Adjusted R^2	0.075	0.111	0.186
Residual Std. Error	0.764 (df = 17952)	2.657 (df = 17776)	0.970 (df = 17945)
F Statistic	15.612*** (df = 100; 17952)	23.060*** (df = 101; 17776)	41.923*** (df = 101; 17945)

Table 7. OLS Regression Results (Continued)

	Dependent variable:		
	ideology_economic	ideology_political	ideology_social
	(1)	(2)	(3)
height	−0.001	0.010***	0.001
	(0.001)	(0.004)	(0.001)
weight	−0.0003	0.003**	0.001
	(0.0003)	(0.001)	(0.0004)
hand	−0.003	−0.029	−0.015
	(0.024)	(0.085)	(0.031)
health	−0.009*	0.055***	−0.018***
	(0.005)	(0.018)	(0.007)
smoke	0.039**	−0.030	0.036*
	(0.016)	(0.056)	(0.020)
drink	0.013	0.020	0.004
	(0.017)	(0.060)	(0.022)
nap	−0.027**	0.001	0.019
	(0.012)	(0.042)	(0.015)
memory	−0.011**	0.041**	−0.004
	(0.005)	(0.017)	(0.006)
time_tv_movie	0.0003	0.003	−0.001
	(0.001)	(0.002)	(0.001)
family_dinner	−0.001	−0.016	0.003
	(0.003)	(0.010)	(0.003)
reading	−0.080***	−0.051	−0.169***
	(0.017)	(0.058)	(0.021)
depression	0.002	0.115***	0.005
	(0.009)	(0.033)	(0.012)
difficult_life	0.005	−0.014	0.032***
	(0.008)	(0.029)	(0.011)
bad_sleep	0.014**	0.036	0.016*
	(0.007)	(0.024)	(0.009)
Observations	18,053	17,878	18,047
R^2	0.080	0.116	0.191
Adjusted R^2	0.075	0.111	0.186
Residual Std. Error	0.764 (df = 17952)	2.657 (df = 17776)	0.970 (df = 17945)
F Statistic	15.612*** (df = 100; 17952)	23.060*** (df = 101; 17776)	41.923*** (df = 101; 17945)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 8. OLS Regression Results (Continued)

	Dependent variable:		
	ideology_economic	ideology_political	ideology_social
	(1)	(2)	(3)
veteran	0.022	−0.088	−0.046
	(0.054)	(0.189)	(0.069)
fulljob	0.012	0.005	−0.007
	(0.014)	(0.048)	(0.017)
income	-0.00000^{*}	0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)
exp_child	0.001	0.034	0.052^{***}
	(0.008)	(0.030)	(0.011)
exp_son	0.023^{**}	−0.030	0.126^{***}
	(0.011)	(0.040)	(0.014)
happy	-0.018^{**}	−0.012	-0.020^{**}
	(0.008)	(0.027)	(0.010)
lonely	0.016^{*}	0.009	−0.016
	(0.010)	(0.033)	(0.012)
life_happy	0.006	−0.016	0.025^{**}
	(0.008)	(0.028)	(0.010)
sad	−0.014	0.073^{*}	0.007
	(0.011)	(0.037)	(0.014)
pessimistic	0.031^{**}	0.024	−0.001
	(0.012)	(0.043)	(0.016)
wordtest18	0.0001	0.006^{***}	-0.001^{*}
	(0.001)	(0.002)	(0.001)
edu	-0.064^{***}	0.074^{***}	-0.068^{***}
	(0.007)	(0.023)	(0.009)
Observations	18,053	17,878	18,047
R^2	0.080	0.116	0.191
Adjusted R^2	0.075	0.111	0.186
Residual Std. Error	0.764 (df = 17952)	2.657 (df = 17776)	0.970 (df = 17945)
F Statistic	15.612^{***} (df = 100; 17952)	23.060^{***} (df = 101; 17776)	41.923^{***} (df = 101; 17945)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

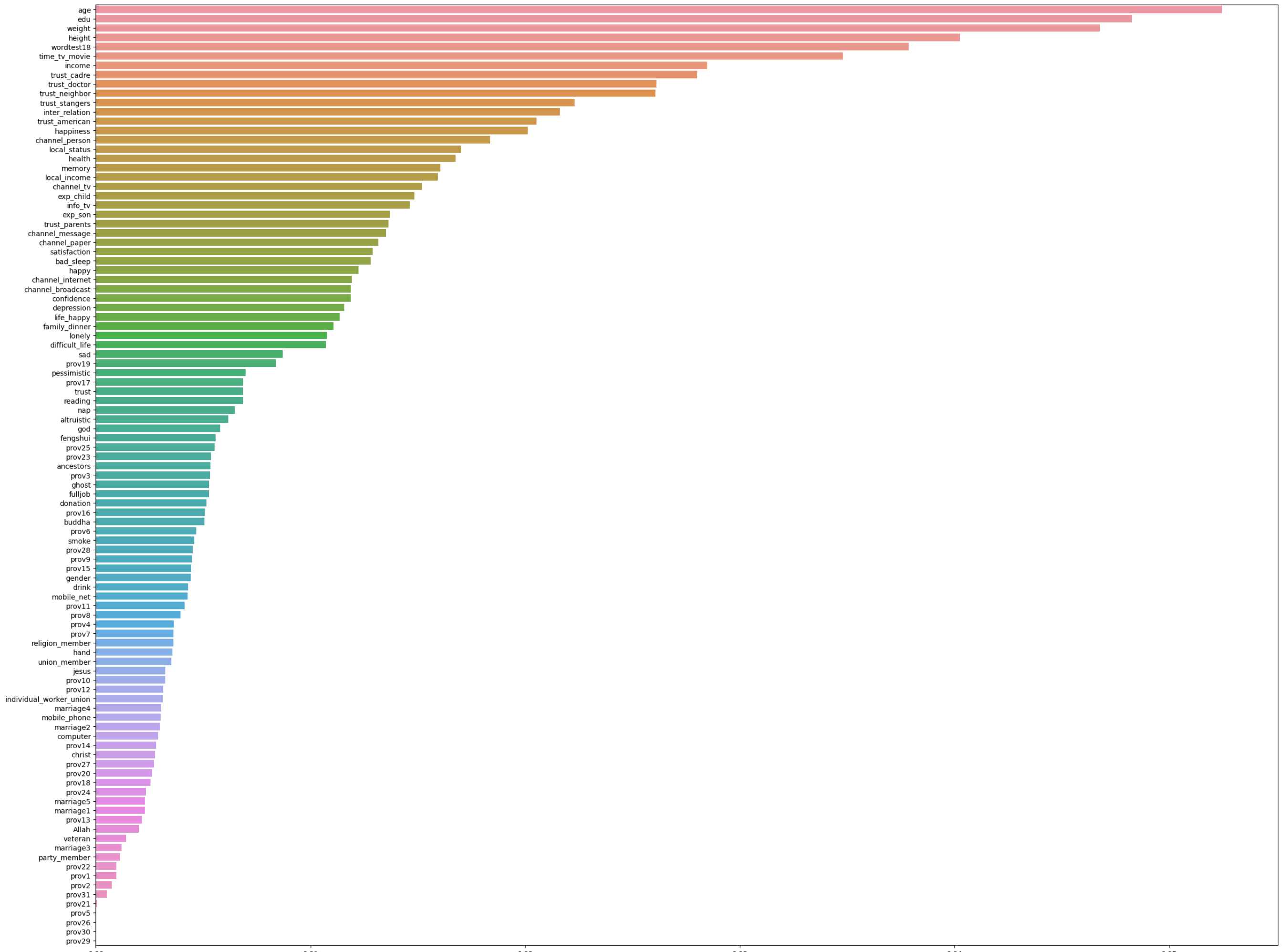

Figure 7. Random forest robustness check for economic ideology

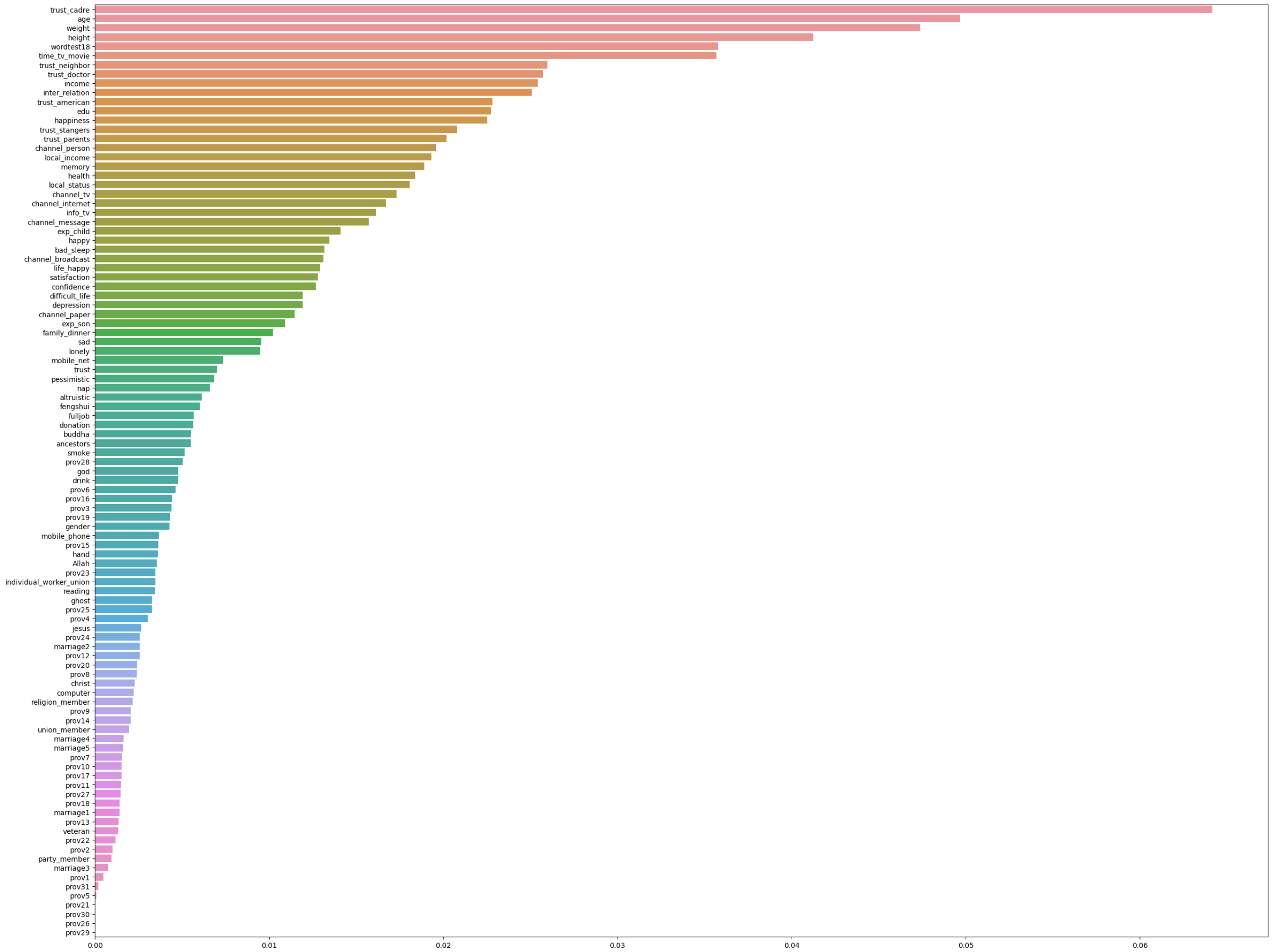

Figure 8. Random forest robustness check for political ideology

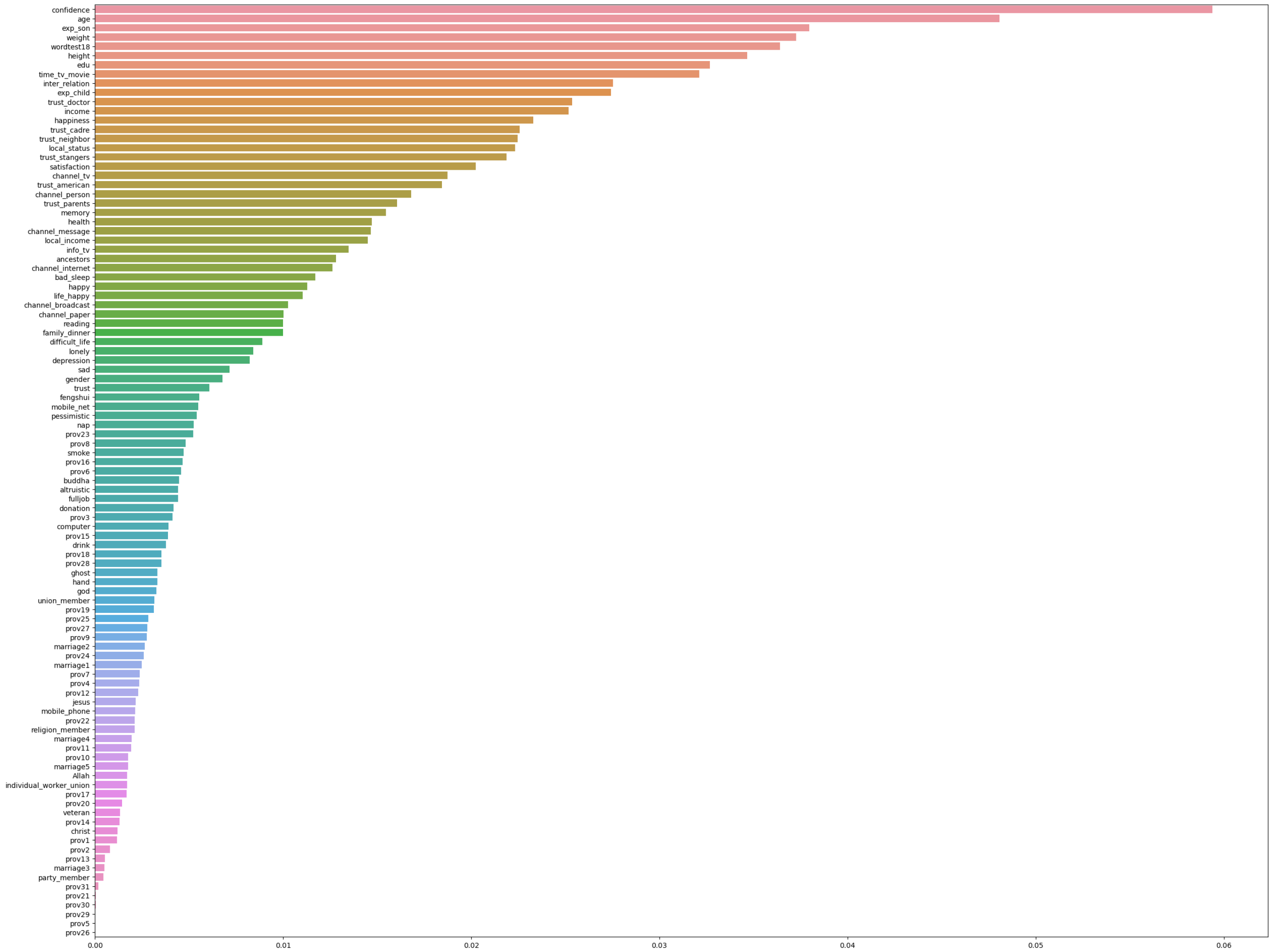

Figure 9. Random forest robustness check for social ideology