Effect of Education on Extreme Right Vote in Europe

Adrian M. Goedeckemeyer

Minerva Schools at KGI

Abstract

In this paper, we investigate possible causes for voting for an extreme right (ER) party in Europe. We start by replicating some regression results from “Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002” by Arzheimer (2009) which created a regression model of personal and contextual variables that might increase the probability of voting for the ER and found that immigration and unemployment both have positive effects on ER vote but a negative interaction effect. To investigate possible causal relationships, we use Matching to estimate an effect of university education on likelihood to vote for the ER of -2%.

Effect of Education on Extreme Right Vote in Europe

There has been a large amount of research into reasons that voters vote for the extreme right (ER) (Lubbers & Scheepers, 2000) (Evans & Ivaldi, 2008) (Sprague-Jones, 2011) (Cornelis & Van Hiel, 2015) each identifying that there is a combination of causes including socioeconomic factors of the country and situational factors about the individual. Many studies identified national sentiments as an important variable. But the questions remain important, especially as right wing populism becomes a burning topic in countries outside of Europe. We revisit a study by Kai Arzheimer (2009) that aimed to identify the most important contextual factors and how they interact with each other. In addition to the logistical regression model we create a random forest model to assess predictive power of this data and importance of different variables. In our replication, we identify important demographic and contextual factors that increase the likelihood of an ER vote. University education has been identified to reduce the likelihood of ER vote in the regression model. We perform a Matching analysis to investigate the effect of university education on likelihood to vote for the ER in comparison to education levels below university.

The findings in this study are aimed towards statisticians in interested government agencies in any country, but especially Europe, and can be used to make decisions regarding where to invest resources to prevent votes for the extreme right, to make predictions based on the current situation and to guide further studies and models in this field. We recommend applying the methods used in this paper to more recent data when available.

The Data

The dataset used in our study covers the years 1980 to 2002 and comes from biannual Eurobarometer surveys conducted by the European commission in the 18 EU countries before the eastern enlargement (Arzheimer, 2009). The survey data includes demographic data, sentiments towards politics and democracy as well as a vote intention question that has been

coded into our outcome variable of support for extreme right parties. The data was combined with contextual data about the country of the respondent, including unemployment numbers, number of asylum seekers, unemployment benefits and a quantified representation of the political climate and the acceptance of right positions in moderate politics. We also added data on the percentage of foreign population and migration rates as potential contextual variables with a higher explanative power than the number of asylum seekers (Gidron & Javed, 2011). In total, we have 175.000 rows of data, of which many had to be removed as they lack relevant variables. In our analysis, we look at 35.000 individual responses with full data from 14 countries, with 4.5% of respondents voting for the ER.

Predictive Models

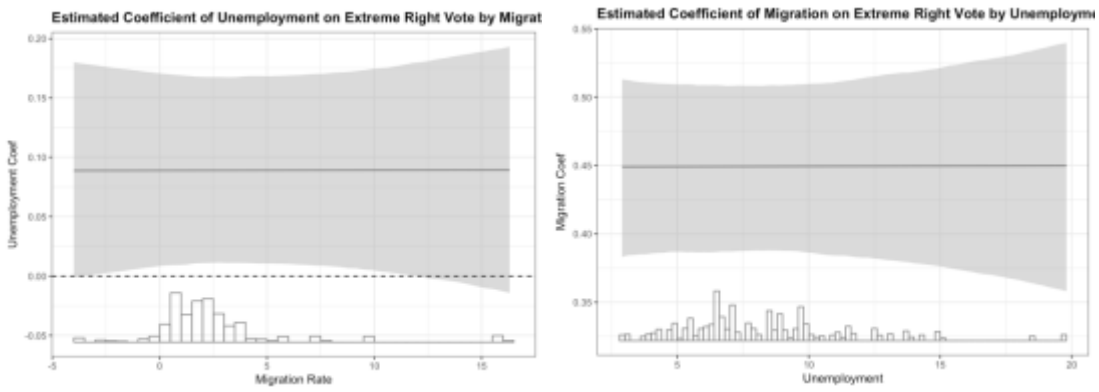
Logistic Regression

Like Arzheimer's study (2009) we perform a logistical regression to build a predictive model of vote intention for the ER. The model includes the interaction between migration and unemployment, unemployment and benefits, migration and benefits as well as salience of extreme politics with variance of political statements. The created model has a 90% prediction accuracy on the withheld test data set. Here it is important to note that we would have 95.5% prediction accuracy if we simply classified every observation as non-ER vote. Our model predicts more than half of ER voters as such and thus has adequate performance but is not as amazing as 90% might suggest, due to the overwhelm of non-extreme voters in the dataset.

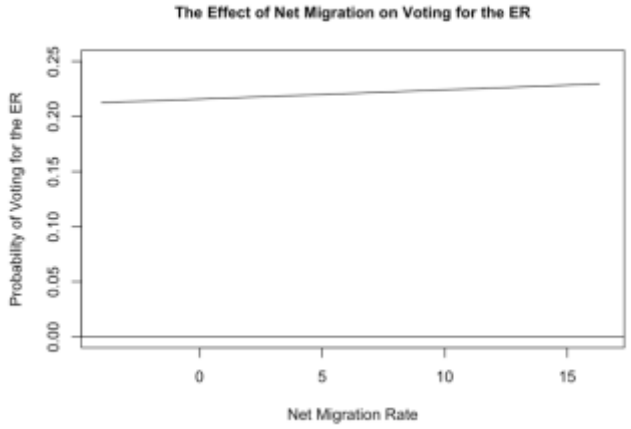
	Estimate	Std. Error	z value	Pr (>|z|)	Sig
(Intercept)	-5.35E+00	4.40E-01	-12.164	< 2e-16	***
malenein	4.57E-01	3.05E-02	15.01	< 2e-16	***
age1	3.98E-01	4.33E-02	9.178	< 2e-16	***
age2	1.65E-01	4.11E-02	4.022	5.77E-05	***
age4	-1.18E-01	5.70E-02	-2.059	0.039476	*
mye1	1.19E-01	3.83E-02	3.119	0.001817	**
mye2	-2.69E-01	4.49E-02	-5.984	2.17E-09	***
farmerown	6.52E-03	5.28E-02	0.124	0.901621	

worker	4.05E-01	4.11E-02	9.845	< 2e-16	***
retired	2.54E-02	5.48E-02	0.463	0.643605	
unemployed	4.73E-01	5.95E-02	7.949	1.87E-15	***
zlrs	5.50E-01	7.54E-03	73.029	< 2e-16	***
euschlecht	7.54E-01	3.72E-02	20.31	< 2e-16	***
zsatisdmo	6.25E-01	1.89E-02	33.034	< 2e-16	***
disp	1.20E-02	5.74E-03	2.099	0.035789	*
lfed1	2.10E-01	6.33E-02	3.319	0.000904	***
migration	4.50E-01	3.72E-02	12.098	< 2e-16	***
sur	8.81E-02	4.16E-02	2.119	0.034093	*
replacementrate	3.29E-02	9.32E-03	3.53	0.000416	***
rmax	2.39E-02	1.08E-02	2.221	0.026351	*
salienzmean	6.81E-02	1.10E-02	6.208	5.38E-10	***
rvar	-1.32E-02	3.49E-03	-3.787	0.000152	***
countryBE	-2.22E+00	1.05E-01	-21.119	< 2e-16	***
countryDE-E	-3.72E+00	1.82E-01	-20.409	< 2e-16	***
countryDE-W	-3.07E+00	1.07E-01	-28.759	< 2e-16	***
countryDK	-1.46E+00	1.81E-01	-8.108	5.16E-16	***
countryES	-5.18E+00	3.94E-01	-13.129	< 2e-16	***
countryFI	-3.84E+00	3.08E-01	-12.477	< 2e-16	***
countryFR	-1.55E+00	2.45E-01	-6.314	2.73E-10	***
countryIT	-1.83E-01	2.66E-01	-0.687	0.491816	
countryNL	-3.22E+00	2.11E-01	-15.289	< 2e-16	***
countryNO	-1.07E+00	2.08E-01	-5.119	3.08E-07	***
countryPT	-3.02E+00	3.52E-01	-8.569	< 2e-16	***
countrySE	-4.27E+00	4.32E-01	-9.879	< 2e-16	***
migration:sur	-4.19E-05	2.72E-03	-0.015	0.987684	
sur:replacementrate	-6.50E-04	1.02E-03	-0.637	0.524016	
migration:replacementrat	-1.21E-02	1.05E-03	-11.511	< 2e-16	***
salienzmean:rvar	3.15E-04	1.28E-04	2.46	0.013884	*

Arzheimer found a negative interaction effect of asylum seekers and unemployment on ER. We picked the net migration rate for each country as a predictor that might be more related with people's sentiments and do not find the counterintuitive interaction there. In our model the interaction has a very minimal positive effect but appears to not be crucial to the model. The effect found in our model is more moderate as in the replication by Gidron and Javed:



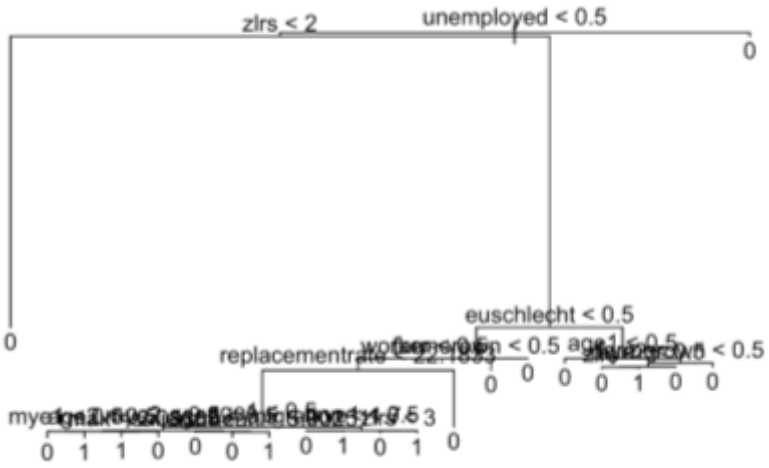
To illustrate the relevance of the migration variable we also simulate outcomes for the full range of net migration values while all continuous variables are held at mean and categorical variables held at their level with highest positive influence on ER. We find a nearly linear relationship due to the weak interaction terms that cancel each other out where increasing immigration from slight emigration up to 16% immigration yields a 3% increase in ER vote likelihood[1].



Classification Trees

To further investigate the possibility of vote prediction we create a classification tree model. Due to the large number of predictors we pruned the tree using cross validation (James et. al., 2015) and receive a slightly more interpretable model with the same test set accuracy of ~95% (more details in r markdown file).

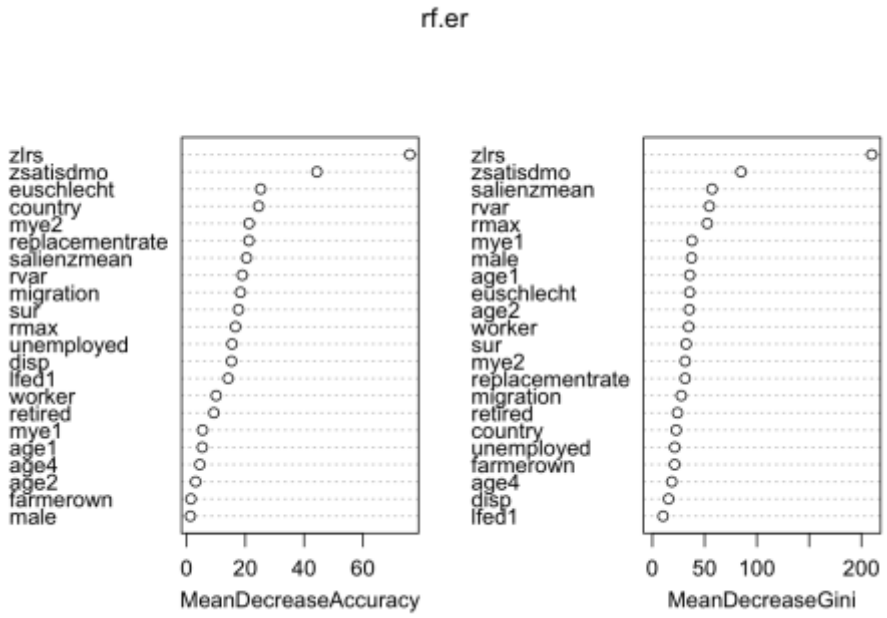
[1] #replication: Replicates findings from another study to check for the results and improve on them and extends on it in a replicable manner.



The low error rate comes at the cost of a very high false negative rate. Only 4% of extreme right votes are predicted as such. A future improvement here could entail adjusting the pruning function to validate for false negative rate rather than misclassification rate overall.

Random Forest

We follow with a random forest model to avoid the bias of our previous model. The random forest does indeed perform marginally better and picks up twice as many right extreme votes correctly. To interpret the random forest model, we are especially interested in the variable importance, showing the influence of changing variables on prediction.



Interpretation

The predictive power of our random forest was highest, but the logistic regression performed a lot better on classifying extreme votes correctly, with 57.5% classified correctly. We recommend logistic regression for further modelling approaches. Across all models we found variables such as personal unemployment, personal beliefs, education to have high power on the prediction results[2]. It also appears that there are strong country level effects, making the country variable important in our tree and giving it high coefficient estimates in our regression model. Thus, it might be appropriate to follow up with more country level studies and not treat all European countries as one.

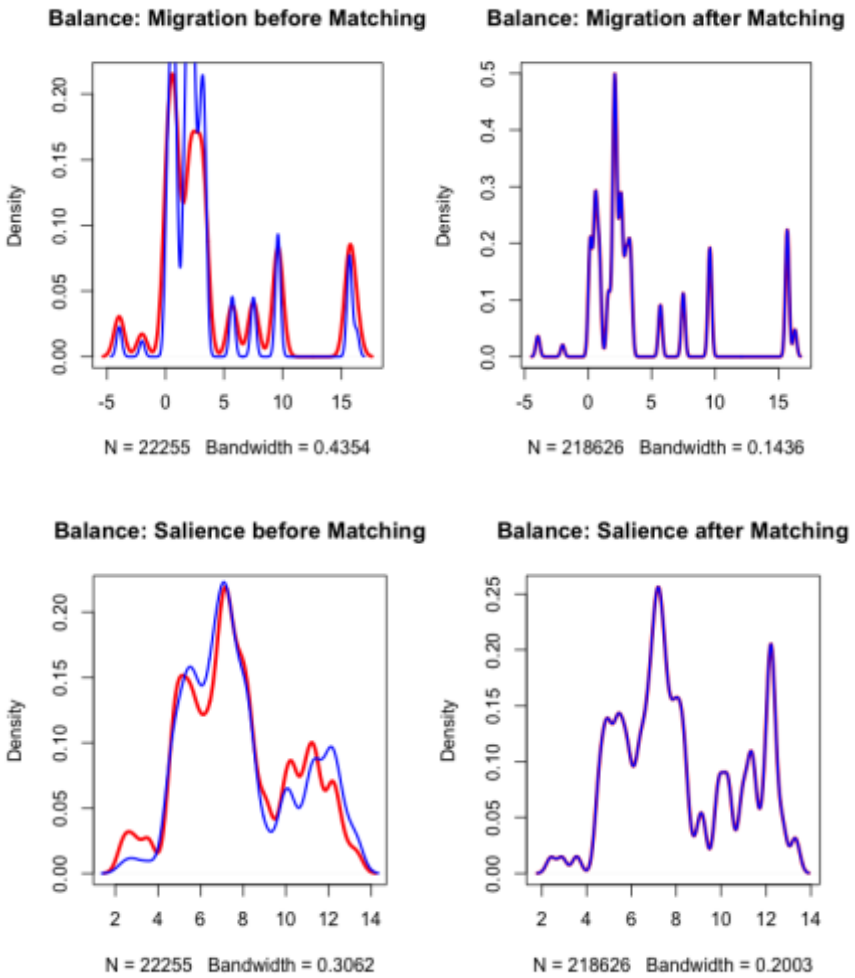
The Effect of Higher Education

To make inferences about in how far the variables with a high explanatory power in our regression influence vote intention, we perform matching on our dataset. In our regression model university education lead to a decrease in likelihood of voting for the ER and in our random forest model, the variable carried high importance as well. Matching on other confounding variables allows us to make a comparison of means in the absence of observed confounders. As university education is our treatment variable (received by 35.3% of respondents), we eliminate many variables that can be directly related to it from the analysis, this includes years of education and general dissatisfaction with the EU and democracy. These variables have strong explanative power in our model as well, but can be influenced by the education one received, and should thus not be matched on. We assume that SUTVA holds in this case, but admit that there could be effects of education on one's neighbors/relatives/etc.

Genetic Matching allows us to identify the optimal weights to use for each variable when Matching our observations (Diamond & Sekhon, 2013). Due to the large size of our

[2] #regression

dataset this is extremely expensive and as there are many categorical variables that we need to match exactly on, such as country, age, gender, occupation to achieve balance we needed to use a caliper. Through the caliper we drop observations if they are too far away from all other observations in our dataset. By dropping less than .3% of observations (in which none of the outcomes is overrepresented) through the caliper argument we can match the variables with full balance on all covariates. Genetic Matching showed us that this can be done through coarsened exact matching as well. These are fortunate circumstances as we have an incredibly large dataset and can afford to drop some rows that contain outliers (A full breakdown of Matching Balance can be found here). Countries with a low number of full responses such as Sweden, Greece, France, Finland and Austria only included members of one group even before matching.



The estimated treatment effect for the treated is -2% and is highly significant with a p-value of 4.4118 * 10^-11. We check the robustness of these findings for unobserved confounding variates with a Rosenbaum sensitivity analysis (Rosenbaum, 2005). We find that for unobserved covariates of an explanatory magnitude of up to lambda = 1.5 the findings remain significant[3].

Interpretation

We can confidently infer an effect of attending university that reduces likelihood of voting for the ER. Considering that only 4.5% of the surveyed voted for the ER, a decrease in likelihood of 2% is highly relevant. We need to keep in mind that there can still be many confounders that obscure a direct causal relationship, such as socioeconomic status of parents or other earlier sources of political beliefs. As we have achieved balance on all covariates, including country, we are also able to infer the effects of education in each country and compare. Which could guide questions on what university systems most effectively reduce ER vote. We have shown that for causal question like this one, the large Eurobarometer dataset is a great resource to approach inference problems with matching and performing a similar analysis on other variables in the dataset can yield more interesting results.

Discussion

The Eurobarometer data is a valuable resource for government decision making. Towards a goal of understanding and combating extreme right vote, it allows us to identify the effects of national variables such as immigration and unemployment as well as personal demographics such as education, age or occupation. We confirmed many of the findings from Arzheimer's model regarding what variables carry explanatory power and found similar results of importance in tree based methods. Taking this analysis one step further we isolated the effect of university education on likelihood of ER and found a highly significant effect,

[3] #effectsize

confirming that it is an important variable and a focus on teaching some essential skills that reduce ER vote in lower education levels could an appropriate consequence. Especially interesting would be a further causal analysis of immigration on extreme right vote, as causal analysis on continuous variables is more challenging, it might be worthwhile to identify a specific immigration policy that changed intake of refugees in one country, while all others maintain a similar old policy and then perform a synthetic control analysis. A similar analysis is also possible for the effects of unemployment policies, which could perhaps be done on a state level in a country such as Germany where some states might introduce a program to dramatically reduce unemployment.

Code and Replication Data:

https://github.com/adriangoe/CS112/tree/master/Final contains a README

References

Arzheimer, K. (2009). Contextual Factors and the Extreme Right Vote in Western Europe, 1980-2002. *American Journal Of Political Science*, *53*(2), 259-275.

Arzheimer, K. (2009, May 4). Replication data for: Contextual factors and the extreme right vote in western Europe, 1980-2002 - Kai Arzheimer Dataverse. Retrieved December 17, 2016, from Harvard Dataverse, https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/12092

Cornelis, I., & Van Hiel, A. (2015). Extreme-Right Voting in Western Europe: The Role of Social-Cultural and Antiegalitarian Attitudes. *Political Psychology*, *36*(6), 749-760. doi:10.1111/pops.12187

Diamond A., & Sekhon J.S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review Of Economics And Statistics*, *95*(3), 932-945. doi:10.1162/REST_a_00318

Evans, J., Ivaldi, G., & Written symposium. (2008). Forecasting the Extreme Right Vote in France (1984-2007). (p.p. p. 137-151). Palgrave Macmillan; 2008.

Gidron, N., & Javed, J. (2011, May 8). Replication data for: The unholy Trinity: Immigration, unemployment and Extremism in Europe, 1980-2002 - Gov2001 Dataverse. Retrieved December 17, 2016, from Gov2011 Dataverse (Harvard University), https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/15902

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An introduction to statistical learning : With applications in R* ([Corrected at 6th printing 2015].). (Springer texts in statistics, 103; Springer texts in statistics). New York: Springer.

Lubbers, M., Gijsberts, M., & Scheepers, P. (2002). Extreme right-wing voting in Western Europe. *European Journal Of Political Research*, *41*(3), 345-378. doi:10.1111/1475-6765.00015

Lubbers, M., & Scheepers, P. (2000). Individual and contextual characteristics of the German extreme right-wing vote in the 1990s. A test of complementary theories. *European Journal Of Political Research*, *38*(1), 63-94. doi:10.1023/A:1007015715758

Rosenbaum, P.R. (2005). Sensitivity Analysis in Observational Studies. In Brian S. Everitt & David C. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science, vol. 4 (pp. 1809-1814). Chichester: John Wiley & Sons, Ltd. Retrieved March 10, 2016 from: Retrieved from https://service.sipx.com/service/php/inspect_document.php?id=cr-339348d5-5f29-11e6-a73e-22000b61898b

Sprague-Jones, J. (2011). Extreme right-wing vote and support for multiculturalism in Europe. *Ethnic And Racial Studies*, *34*(4), 535-555. doi:10.1080/01419870.2010.512665