Creating Methods to Predict

Future European Parliamentary Elections

Michael Ryvin

Introduction

Over the course of 40 years, the European Union has held nine elections to the European Parliament. While initially this body was not very powerful, the EU Parliament has gained power over time, as Europe has slowly federalized. An institution that once dealt primarily with trade now has the power to control the currency of most European countries and regulate numerous industries. Given the growing importance of the EU Parliament, it is becoming increasingly important to follow European Parliamentary elections, as these elections will only become more important with time. Given that, the question arises of if there are ways to predict the results of these elections in advance and potentially manipulate them.

Although there has been extensive research regarding what types of voters may vote for certain political party groups, could it be possible that certain types of countries vote for certain political party groups as well? For example, while it is fairly well-documented that younger people in many nations tend to vote for more left-leaning parties, does that mean that a country with more younger voters or a younger population as a whole also sends more left-leaning politicians to the European Parliament? Likewise, are there certain stable and unstable statistics that are better at predicting the voting patterns and results of these elections, and if so, are there ways for them to be manipulated?

This research intends to fill in the gaps that previous research has left. Although models made based solely on volatile data, such as polling or inflation, may be more accurate in the short term, these models cannot be used years in advance. Any volatile data that is used more than a year before the election is bound to change, and it is often difficult to predict where these volatile factors such as 'how the economy is doing' will be by the time of the election. For example, it is impossible to predict how much the Covid-19 pandemic will affect the 2024 elections, as it is difficult to determine when the pandemic will end, and whether some voters will still hold resentment towards certain parties two years from now. By including or only using relatively stable factors, the model will have more useability when projecting from further back in time. Furthermore, relatively stable factors tend to be much easier to predict in the long-term, as factors such as demographics and GDP usually follow certain trends over the course of decades, regardless of short-term crises that occur.

In order to test the research question, data will be collected about the 2014 and 2019 European Parliamentary elections, as well as data about relatively stable and unstable factors of each nation in the European Union around the time of those elections. This data will then be cleaned and analyzed using programs such as R and Python in order to determine any factors that seem to hold a correlation with election results of each party group in the European Parliament. The factors will be chosen through a combination of observing the data visually and using the mathematical values given when testing for correlation. Upon completion, a number of factors will have been kept and included in six different models, one for each political ideology represented in the European Parliament.

Overall, this study will use a mixture of inductive and deductive reasoning. While the independent variables collected will be based on prior research, the creation and attempted use of the models will allow for the introduction of new theories regarding the European elections. As a result, this study will develop its own theories, but will require prior research as well.

Other than simply making prediction models, the purpose of the study will be to give political parties, the media, and the general public a better understanding of the direction that the European Union is heading and the potential results of European elections. By learning how these elections will play out, parties will be able to gain more knowledge on how they should change their platforms to either better appeal to voters, or to help sustain their parties' future prospects. Likewise, the media and general public would be able to avoid many surprising results in these elections, as well as know how their votes in national elections may change the political environment when European elections come up. Hopefully, if the models created are strong enough, it may even be possible for people to know the rough results of the elections long before campaigns even begin.

Theory

Before explaining which variables would be needed, there is a question as to why making a predictive model would be beneficial. Rather than creating a model, it would also be possible to see how individual variables correlate with the vote share of each party and make conclusions about that. However, this approach would be inferior to making a predictive model that can incorporate multiple variables. First off, there may be variables that are not correlated with the vote share on their own, but are once other variables are included. For example, perhaps a variable of age only correlates with the vote share when the education of people that age is taken into account.

Secondly, and more importantly, a model with multiple variables would allow for a comparison of different variables and their relative importance to one another. Some variables may be more significant to the model than others, and understanding which variables are more significant is important to the useability of the model. Otherwise, it will be difficult to differentiate between variables that have a coincidental relationship versus variables that potentially could have a causational relationship.

Finally, creating a model allows for a more nuanced approach to the variables being used. By making a model, all the independent variables that could affect vote share could be split into multiple independent variables. For example, having a predictive model would make it easier to separate an age variable into multiple age brackets rather than simply using the average age of each country. While technically many variables could be used without making a predictive model, it would be extremely inefficient and far more difficult to parse through.

Another aspect of this study to discuss is why six different models are being made rather than one singular model for vote share. The primary reason for this is to make sure that only one dependent variable is included in each model. To be able to analyze how different factors affect the vote share of each party group in the EU Parliament, there must be multiple dependent variables. Each party group's vote share counts as a separate dependent variable and thus must be analyzed separately. Six party groups were chosen to represent the six overarching political ideologies present in the EU Parliament: right-wing, center-right, center, green, center-left, and left-wing.

Understanding the reasoning behind using a predictive approach and why six models are being made, it is now important to understand which kind of independent variables will be needed for the models. This will be determined based on prior research regarding each of the party groups and overall factors.

One major factor that has been correlated with changes in vote share from multiple groups is economic data. Prior studies have found that the appeal of far-left and far-right parties has generally grown with the share of people who are not economically well-off (Hudgens). It does seem that poorer voters or ones that may be in the lower middle class are far more likely to be convinced to vote for the parties on the fringes of the political spectrum due to these parties promising the most radical changes to society. However, there are also studies that suggest that far-right parties can also attract more affluent voters as well (Rooduijn et al.). As a result, it is possible that for the far-right these economic factors may cancel out. Regardless, they should be included as they can be useful for the far-left or other parties.

Economic data is also useful for parties closer to the center. The center-left parties have generally been shown to perform best with voters in the middle class (Bandau). This is to be expected, as these parties tend to have emerged from working-class movements in the 20th century, and as such their support lies primarily in this demographic. In contrast to these parties, however, are the green parties of Europe. Although green parties are traditionally center-left as well, they have been shown to perform better in countries that are more economically affluent (Pearson and Rüdig). This is due to the fact that lower-income people, even if they understand the importance of climate change, are more likely to prioritize policies that will help them in the short term, rather than hypothetical issues that will arise in the future. Meanwhile, voters who may still sit on the left of the political spectrum but are more affluent will be able to vote for green parties that focus on these more long-term issues like climate change.

Another important category of variables to obtain is education variables. First off, there is generally a correlation between the educational and economic factors of a country. People who go to higher education will tend to earn more money than those who do not, and as such education is an important factor to obtain. By doing so, it may be able to cover for small things that the economic variables may not find.

A more important reason for including education as a variable is its relation to voting for parties on the extremes of the political spectrum. There have been studies that suggest that people with lower education tend to be far more likely to be swayed by far-left and far-right parties (Rooduijn et al.). However, it is difficult to tell whether this is a byproduct of these voters also being less affluent, or these voters being more likely to be convinced of arguments from parties that are more likely to make promises of change. By having both educational and economic factors in the model, this difference could potentially be determined once the models are created.

The next category of variables that will need to be included is variables to do with immigration. These variables will need to be included due to their potential relationship with right-wing and far-right parties. Generally, the differentiating factor between the success of left-wing and right-wing parties is the variable of immigration. Right-wing parties have been shown to succeed more when immigration levels are higher (Edo et al.). As a result, immigration levels may end up being an important factor when determining the vote share of right-wing or left-wing parties, as well as potentially center-right or center parties. Ideally, immigration should both be viewed in terms of the change as well as a total. This is primarily so that it can be determined whether increases or decreases in immigration affect vote share, or if the total immigration is what affects vote share.

Another category of variables that should be included is age. The primary reason for this variable's inclusion is the fact that age tends to have a strong influence on voting patterns in most countries. However, it is difficult to tell whether age will be a good variable for any of the party groups as these age trends may differ widely depending on the nation. While younger voters in some countries may be associated with the center-left, in others they may be associated with the far-left or maybe even right-wing. Although these trends are difficult to predict, it is a potentially useful variable that should be included.

All of the remaining variables to be included will be based on polling data. This is because these variables will depend on opinion-based data. The most important category of variables of this nature to include is variables to do with levels of Euroscepticism. This category is included due to how it affects nearly every party group. First off, left-wing and right-wing parties have both been shown to attract Eurosceptical voters. Although not entirely for the same reasons, voters on either side tend to be distrustful of institutions such as the EU due to them being multinational. Right-wing parties tend to be distrustful due to them espousing nationalist views (Bolin et al.). Left-wing parties may also espouse nationalist views, but they primarily are eurosceptic due to their view of institutions like the EU as benefiting the rich more than they benefit the poor (Wagner).

While parties on the extremes tend to be Eurosceptical, parties closer to the center tend to attract people who like the EU. The centrist parties, which typically are liberals, primarily appeal to europhiles and hope for the EU to succeed (Bolin et al.). While the center-left, greens, and center-right do not necessarily specifically appeal to europhiles, the metric is still useful to differentiate them from their more extreme counterparts. As a result, Euroscepticism can potentially be used as a variable in all of the models.

The final variables that must be found are opinions to do with concern for the environment and climate change. As would be expected, these variables are correlated with vote share for green parties (Pearson and Rüdig). These variables must be included as they may potentially be important in differentiating green parties from the other parties on the left. While other variables may be sufficient, it remains to be seen whether they will be useful enough without including opinions on environmental issues.

As long as all of these categories of variables are included, models should be able to be made about each party group. However, other variables may be included should they be easy to obtain for all the countries and could potentially serve as ways of making cleavages in the parties. Ideally, if any other variables are included, they will be other opinion-based variables.

Data and Methods

To conduct the analysis, sufficient data first had to be collected. This data could be grouped into two primary categories: the dependent and independent variables. The dependent variables would consist of the actual election results of the years being analyzed, while the independent variables would consist of all the possible variables that could affect the election results, depending on the European Parliamentary Party.

First off, a decision had to be made for which election years to analyze for the dependent data. European elections occur once every five years, with the last election occurring in 2019 and the next one occurring in 2024. Any future elections obviously could not be used, as even if the independent variables could be predicted, there would be nothing to compare them to. Looking back, it did not make sense to use elections from before the 2014 election. This is due largely in part to the differences in parliamentary groups as well as the members of the European Union. The further back in European Parliamentary history, the less that the parliamentary groups resembled the ones of today, and the fewer countries that could be used as data points in the set. By using 2014 and 2019, all 28 EU member states (EU-27 and UK) could be included, and the party groups remained roughly consistent. Importantly, this also allowed 2014 to be used as the basis for the model being created, and 2019 as a way to test the model.

At first glance, it may have seemed simple to find and collect the election results for any given election year. While it is true that these election results are readily available for the public to access, there existed no dataset with the extensive results necessary to complete this analysis. It was possible to find the vote share of parties that ended up making it into the European Parliament. For some countries, such as Germany, this would not have been an issue. Only 3.5% of votes in the 2019 EU election in Germany went to parties that did not end up having seats in the EU Parliament, largely due to the large number of seats that Germany has and the low vote threshold necessary to obtain a seat. Many other countries in the EU are not so lucky. Greece, for example, had 21.01% of the votes cast in the 2019 EU election go to parties that did not end up entering the EU Parliament.

Because of the number of parties that did not enter the EU parliament, many election datasets simply lumped together many parties into an ‘Other’ category, that could consist of parties from anywhere in the political spectrum. More importantly, however, is that even if individual parties were listed, it would be difficult to gauge what parliamentary group they would have joined had they made it into the EU Parliament.

The first issue could be solved simply by going to each country in the EU and individually aggregating the results of EU elections in each country. Parties that otherwise would have been classified as ‘other’ had their votes counted separately, so as to not lump together parties of vastly different ideologies. The party name, the number of votes, and the country in which they were based were collected.

In some cases, the names of candidates or independents were counted separately rather than as a part of another party. This was necessary as there existed cases where individuals that were part of a party list did not join the parliamentary group that their party was part of. The most notable example of this was MEP Brian Crowley from Ireland, who as a member of Fine Fáil chose to join the ECR parliamentary group rather than ALDE, the party group that Fine Fail was part of. Independents also had to be counted as they often got a substantial amount of the vote and would otherwise skew any results were they to be ignored.

A couple of other caveats existed when collecting the votes for some specific countries. Both Ireland and Malta use a form of single transferable voting, meaning that voters had to rank candidates when casting their votes. This would be difficult to display in the dataset, and so only the first preference votes were used. This would not necessarily perfectly reflect voter intentions, but it was the best way to show the votes in these nations. In addition to this, an accurate list of all the parties by number of votes in the 2019 Austrian European elections could not be found, but the percentages of the votes were counted instead, with 0.8% going to an 'unknown' category. This 0.8% is too small to have an overwhelming effect on the model, and so this was not a large setback.

Once all of these votes were combined, the vote shares were found by adding up the votes by country and determining what percentage of the votes came from each party. Data on votes cast for 'no party' was not collected, and so the results were slightly different from those reported in other places. This choice was made deliberately as not all countries reported these protest votes, and so they were ignored for all countries for consistency.

The next step in finalizing the dependent variables was to assign each party to a certain parliamentary group. For some parties, this was a simple process. If all members of a certain party joined a certain EU parliamentary group, then the votes for that party would count for that group. The only exception to this was the aforementioned Ireland case with Brian Crowley, who was counted separately from the rest of Fine Fáil.

In cases where the national party was split among multiple European parties, the votes were split proportionally by the proportion of MEPs that went to each party. This was common among voting coalitions, where multiple national parties ran under one list in the election, but split up their European allegiances once they joined the European Parliament. The parties and coalitions that needed to be split up are listed below.

Year	Country	Party	First Proportion	Second Proportion
2014	Croatia	HDZ-HSS-HSP AS-BUZ-ZDS-HDS	5/6 EPP	1/6 ECR
2014	Croatia	SDP-HNS-IDS-HSU	1/2 S&D	1/2 ALDE
2014	Spain	La Izquierda Plural	5/6 GUE/NGL	1/6 Greens/EFA
2014	Spain	Coalicion por Europa	2/3 ALDE	1/3 EPP
2019	Germany	Die Partei	1/2 GUE/NGL	1/2 Greens/EFA
2019	Greece	Laiki Enotita - Metopo Anatropis	1/2 GUE/NGL	1/2 Greens/EFA
2019	Netherlands	Christian Union - Reformed Political Party	1/2 EPP	1/2 ECR
2019	Poland	European Coalition	17/22 EPP	5/22 S&D
2019	Slovakia	Progressive Slovakia-SPOLU	1/2 EPP	1/2 RE
2019	Spain	Podemos-IU-Unidas Podemos Cambiar Europa	5/6 GUE/NGL	1/6 Greens/EFA
2019	Spain	Ahora Republicas	2/3 Greens/EFA	1/3 GUE/NGL

While the vast majority of votes were cast for parties that ended up joining the EU Parliament, many also were not. Thus, simply ignoring these votes would skew the models more in favor of the parliamentary groups that contain fewer, larger parties. As a result, a decision also had to be made about every party, coalition, or individual that did not gain any MEPs.

In order to do this, each party without an EU party designation was researched, and a judgment call was made for each regarding its potential alignment with a party group or ideology. In many cases, the parties were already part of a party group that existed within the European Parliament. These included not just the six party groups that existed in 2014 and 2019, but also smaller party groups that were part of these larger ones. For example, the 'European Pirate Party' group is its own distinct European party group, but it is part of the Greens/EFA party group as well. All national parties that were part of the European Pirate Party were thus classified as Greens/EFA.

Even with these metrics, many parties did not have known party groups. For most of these parties, the party platforms or websites were analyzed to determine what ideology best described them. These were categorized into rough ideological categories, consisting of either the left-right spectrum or other categories that would better describe the parties, such as liberal or regionalist. Regionalist parties without a clear alignment to Greens/EFA were classified as unknown, otherwise, they were grouped with the rest of EFA in Greens/EFA. A party was generally determined to be further from the center when it was more eurosceptic, with the direction away from the center being determined by policy positions. All of these classifications would be split up into party groups later.

This finally left only parties without an accessible party platform. The majority of these parties were classified as unknown. However, some decisions were made for parties with names that easily implied a certain political party group or ideology. For example, a party with the term 'communist' in its name could be reasonably assumed to be left-wing or far-left in ideology.

With all of the parties now categorized, the data had to be further simplified to keep the dependent variables for each country consistent and to a minimum. In order to do this, all of the previously mentioned ideological categorizations had to be designated to certain party groups. Parties categorized as far-left or left-wing were designated as GUE/NGL, center-left as S&D, green or regionalist as Greens/EFA, center or liberal as ALDE or RE, center-right as EPP, and right-wing or far-right as ECR/EFDD or ECR/ID. A decision was made to combine both of the right-wing groups into one larger group due to the difficulty in determining the distinction between the two for all of the unknown parties.

All of the Non-inscrit parties were also split into these categories depending on their own ideology. Although non-inscrits by definition rejected joining one of these larger parliamentary groups, it was thought to be more important to treat these groups as groupings of ideologically similar parties rather than purely as a party structure.

Upon completing all of this, the percentages of the votes were added up for each party group by country, and the party names were thrown out. The result was the dependent variable for the study.

The process for obtaining the independent variables was far easier, as it simply involved searching for data that the literature review suggested could have an influence on the vote share of certain European party groups. In total, 144 variables were obtained.

The first variables that were necessary for the study were the economic variables. These were represented in the form of GDP (PPP) per capita (*World Economic Outlook Database, April 2022*). This was chosen as it is an easily obtainable metric that can give a good judgment of how the average person is doing in terms of their financial situation in the country. The years obtained for this variable were 2013, 2014, 2018, and 2019. This was done as the European Parliamentary elections do not occur at the end of the year, and so it may be important to factor in the previous year as well. Furthermore, this would allow for additional variables describing the change in

GDP per capita from the year preceding the election to the year of the election. These percent changes from 2013 to 2014 and 2018 to 2019 were added as additional economic variables.

The next variables obtained had to do with age. For this set, data was found grouping the percentages of the population by age ranges (*Population by Age Group*). These blocs were ages 0-14, 15-24, 25-49, 50-64, 65-79, and 80 & over. It can be noted that these blocs are not identical in size, but were deemed sufficient enough as they roughly represent age groupings with differing enough experiences that they may vote differently. In the case of the ages 0-14, while being unable to vote, the number of young children could still have an effect on how adults vote, whether they are parents or not.

After this, the largest number of variables were obtained to find information about education level. For this category, data was collected showing the percentage of people of various age groups to attain a certain level of education (*Population by educational attainment level, sex and age (%) - main indicators*). The levels of education were split into three groups: levels 0-2, corresponding to people who received somewhere between less than primary education to lower secondary education, levels 3-4, corresponding to upper secondary education to post-secondary non-tertiary education, and levels 5-8, corresponding to short-cycle tertiary education up to doctoral level or equivalent. For each of these education levels, data was used from the following age blocs: ages 15-64, 20-24, 25-34, 25-64, 30-34, 35-44, 45-54, 45-64, and 55-64. Each of these was found for both 2014 and 2019. In total, there were 54 variables included as a result.

The next variables included dealt with immigration. First off, data was collected relating to the number of immigrants that entered each country depending on where they were from (*Immigration by broad group of country of previous residence*). This included the total number of immigrants, immigrants from EU27 countries, immigrants from non-EU27 countries, and immigrants from unknown origins. Data using EU28 could not be found, but it was deemed to be negligible due to the fact that the amount of emigration away from the UK was negligible. This data was found for the years 2013, 2014, 2018, and 2019. The reason for this is because the raw number of immigrants may not have been a very accurate number to base a model off of, as each of these countries had wildly different populations. As a result, these variables were used to find the percent change in number of immigrants from the year preceding the election to the year of the election.

In addition to these immigration variables, data was also collected on the number of immigrants from each of the aforementioned categories as a percentage of the total population of the country (*Recent immigrants by sex, age and country of birth*). This data was also found for the years of the elections and the years preceding them, but in this case the change was not calculated. Instead, both were found simply as a precaution for the fact that the elections occurred in the middle of the election years, as noted before.

All of the remaining variables included were opinion-based variables. As a result, rather than using data about the countries, the data was obtained from polling (*Standard Eurobarometer 91 - Spring 2019*) (*Standard Eurobarometer 81 - Spring 2014*). The first of these categories dealt with the levels of euroscepticism. This consisted of three separate questions in the polling; the first question asked whether or not the person felt like a citizen of the EU, the second simply asked whether the person had a positive or negative opinion about the EU, and the third asked whether the person thought that the country would be better off if it left the EU. The reason for including all three is due to the fact that none of the three fully covers the idea of euroscepticism.

For example, someone could have a negative opinion of the EU but still not think that their country should leave it.

The next category of questions dealt with people's opinions on the climate and environment. This included two questions: what are the two most important issues you are facing at the moment, and what are the two most important issues the EU is facing at the moment. For the question asking the issues that the people are personally facing, the percentage of answers relating to environment, climate, and energy issues were included as one variable per year, as this was how they were reported. For the question asking about issues facing the EU, answers of 'the environment,' 'energy supply,' and 'climate change' were kept as separate variables, as this was how they were reported. Both questions were included as it is possible that the success of certain party groups depends either on how people personally perceive issues relating to the environment or how people perceive them as European issues as a whole.

A couple of other variables were included that could be included in the models that did not fit into any of the previous categories. First was the percentage of people that were satisfied with their life. This was included as people who are less happy with their lives may be more inclined to vote for certain party groups that they think will change their unhappiness. The second variable was the percentage of people who identified themselves as someone with left, center, or right political views. This potential correlation is clearer, as a country with more people with political views leaning in a certain direction could potentially have a higher percentage of votes for parties of that same political leaning. With both of these variables added, all of the independent variables were included.

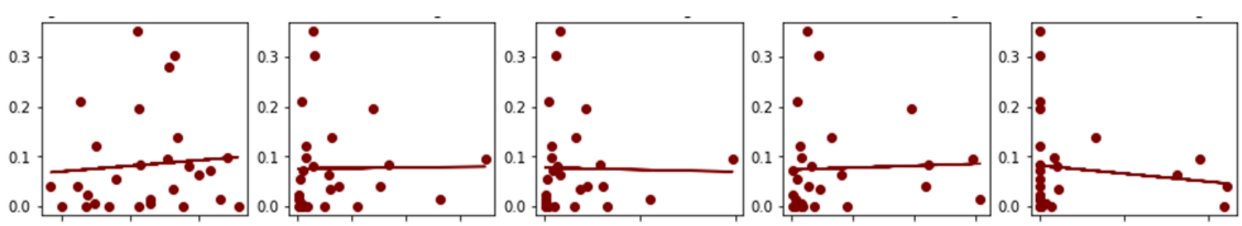
With all of the data sorted out, the models could now be made. In total, there would be six models made, one for each of the dependent variables. These were GUE/NGL, or the left-wing to far-left parties, S&D, or the center-left parties, Greens/EFA, or the green and many regionalist parties (as well as others), ALDE and RE, or liberal and centrist parties, EPP, or center-right parties, and finally ECR and EFDD/ID, or right-wing to far-right parties.

All of the models made would be in the form of linear regressions. The reason for this was to better understand to what degree each of the independent variables influenced the vote shares. Given that all of the variables included were numerical, it was deemed that a linear model would do a better job of treating all of the variables as continuous, unlike say a decision tree that would split the data at arbitrary intervals. Most importantly, a linear regression allowed easy manual manipulation in order to alter the models to prevent overfitting.

Initially, a stepwise regression was performed to create a base model for each of the party groups. Any final model would have to be at least better than this stepwise regression, otherwise the model outputted by the stepwise regression would be used. The adequateness of the model was determined by a combination of three factors. First was the p-value, which tended to be incredibly low for all of the initial stepwise regressions. This would mainly be looked at to see if any manual alterations significantly decreased the effectiveness of the model. Second was the Adjusted R^2, which determined to what extent the variation in the model could be explained by the independent variables being used. In many of the models, the Adjusted R^2 was quite low, and so manual manipulation would definitely be required for those models. Finally, the average residual of the model compared to the actual vote share was used to see how the model actually played out on the data. In each case, the average residual was not too large on the 2014 data, but was always larger on the 2019 data, often to the point of making the model unusable. Effectively, the 2014 was treated as the training data, while the 2019 data was treated as the testing data. By looking at these factors, better models than the stepwise regressions could potentially be found.

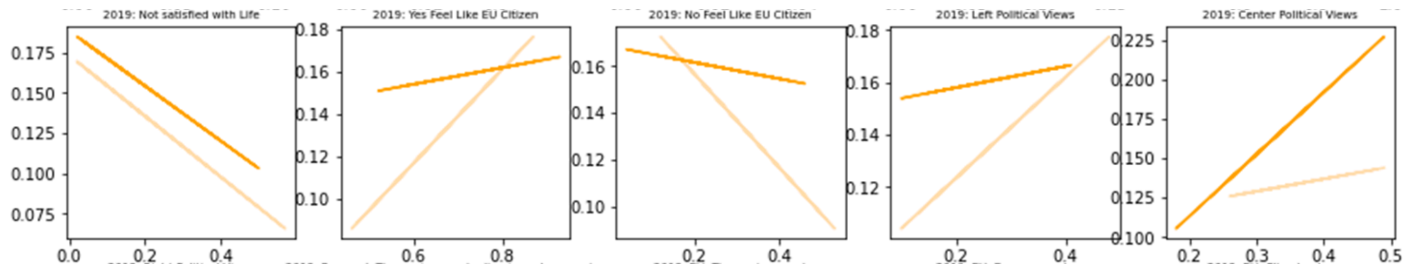
When manually manipulating the regressions, variables were primarily taken away, rather than added. The main issue with many of the stepwise regressions was the overfitting of the models due to too many variables. By removing some variables, while the average residual would increase on the 2014 data, it could decrease when applied to the 2019 data - thus creating a model that would be better at predicting other European elections. In some cases, some variables were not removed or added, but rather altered. This happened with variables relating to immigration, age, and education, where different variables were tested to see if they would improve the model.

There were also a few cases where completely different variables were tested. Rather than having to test every single variable combination for every model, the way to determine which additional variables to test out was determined by creating many scatterplots containing lines of regression between the vote share of a party and a particular independent variable. Some examples are shown here:



These results could also be used to determine where a variable should be attempted to be logarithmic, as some of the scatterplots above seem to suggest.

Scatterplots were also helpful in determining which variables to add or remove in terms of how they affected the residuals for the 2019 data. By plotting the lines of regression of each 2014 variable and comparing them to each 2019 variable, it was easy to spot which variables diverged more and should be removed as well as which lined up similarly and should be inspected more. This can be seen in some of the variables being compared between ALDE vote



share in 2014 and RE vote share in 2019 below.

The leftmost graph would be one that could be inspected assuming that the scatterplots previously mentioned had points that seemed to show a linear relationship. Meanwhile, the other graphs would be ignored or potentially suggest the need to remove those variables, as the variables do not have a consistent relationship from 2014 to 2019.

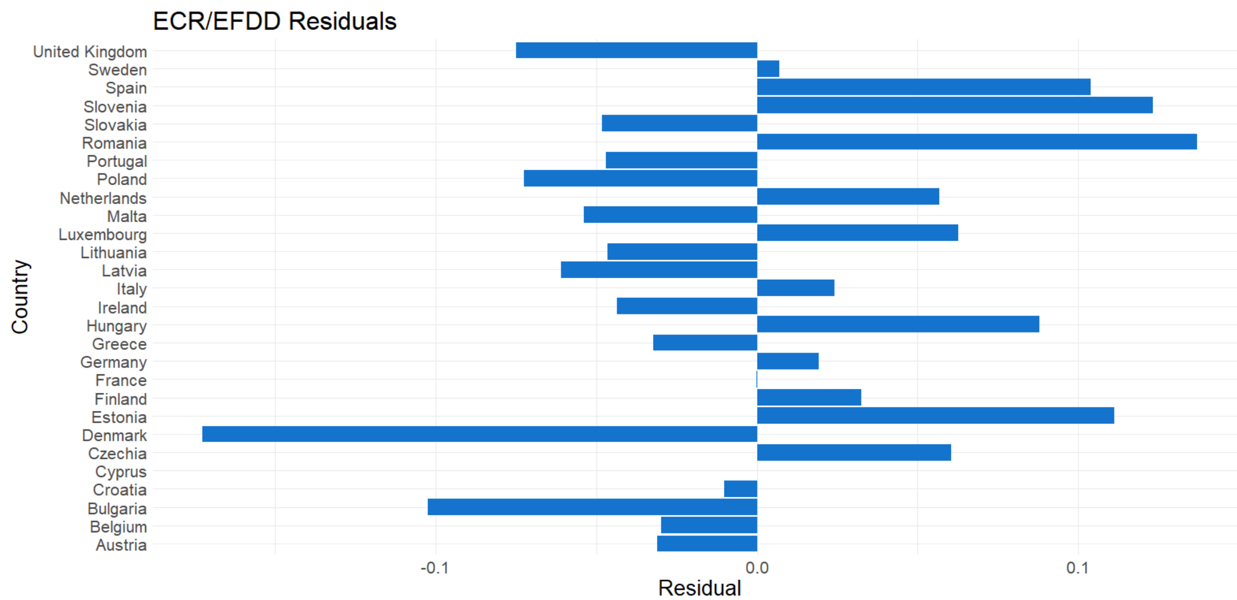
In the end, 29 variables were used across the six models. 115 of the 144 variables were not used in the models. While this may make it seem like these variables should not have been gathered in the first place, their omission can be analyzed later.

Analysis of ECR/EFDD and ECR/ID

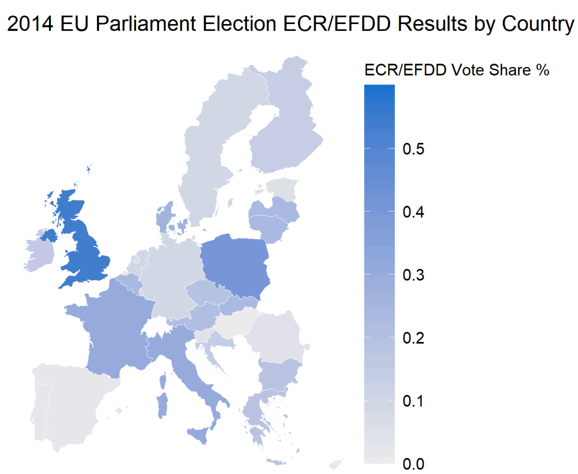
The first model to be examined is the one describing the generally right-wing and far-right political party groups in the European Parliament. In both 2014 and 2019, this includes the European Conservatives and Reformists, or ECR. However, the other party group differs in 2014 and 2019, as both only existed following their respective elections. For 2014 this is Europe of Freedom and Direct Democracy, or EFDD, and for 2019 this is Identity and Democracy, or ID. This model includes five different variables, across two categories of variables: education and immigration. These variables are the logarithm of the percentage of people aged 15 to 64 with an education level of 0 to 2, the non-EU27 immigration in the year preceding the election, the EU27 immigration in the year of the election, the total immigration in the year preceding the election, and the total immigration in the year of the election. The model with all of these variables can be seen below.

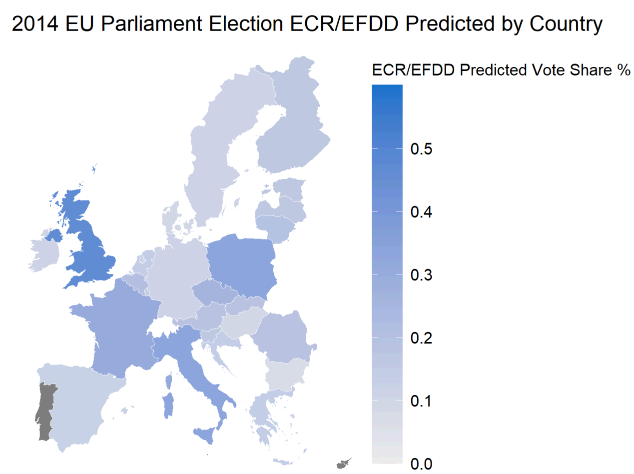
Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	-0.1148	0.06820	0.107190	
Log of people aged 15 to 64 with an education level of 0-2	-0.1606	0.04365	0.001390 **	
Non-EU27 immigration in previous year	0.000009866	0.000002361	0.000424 ***	
EU27 immigration in current year	0.00001151	0.000002985	0.000914 ***	
Total immigration in previous year	-0.000001810	0.000001366	0.199352	
Total immigration in current year	-0.000007872	0.000001807	0.000276 ***	
Adjusted R^2				0.6101
p-value				0.000098

In order to test out this model, the 2014 independent variables can be inserted, using the coefficients given above. When this occurs, there is an average error of about 5.7%. This means that the model tended to estimate the combined vote share of ECR, EPDD, ID, and other right-wing parties within 6.12% of the actual combined vote share within each country. The error did not seem to specifically overestimate or underestimate the vote share, as the average direction of the error was it overestimating the data by $1.85 * 10^{-16}$, meaning that the errors were evenly distributed. The actual and predicted vote shares of this voting bloc can be seen in the following bar plot:



The largest error on this data can clearly be seen to be Denmark, where the model severely underestimates the actual vote share. Several countries also have a significant overestimation of the vote, such as Romania and Slovenia. There does not appear to be any geographic biases, as indicated below.

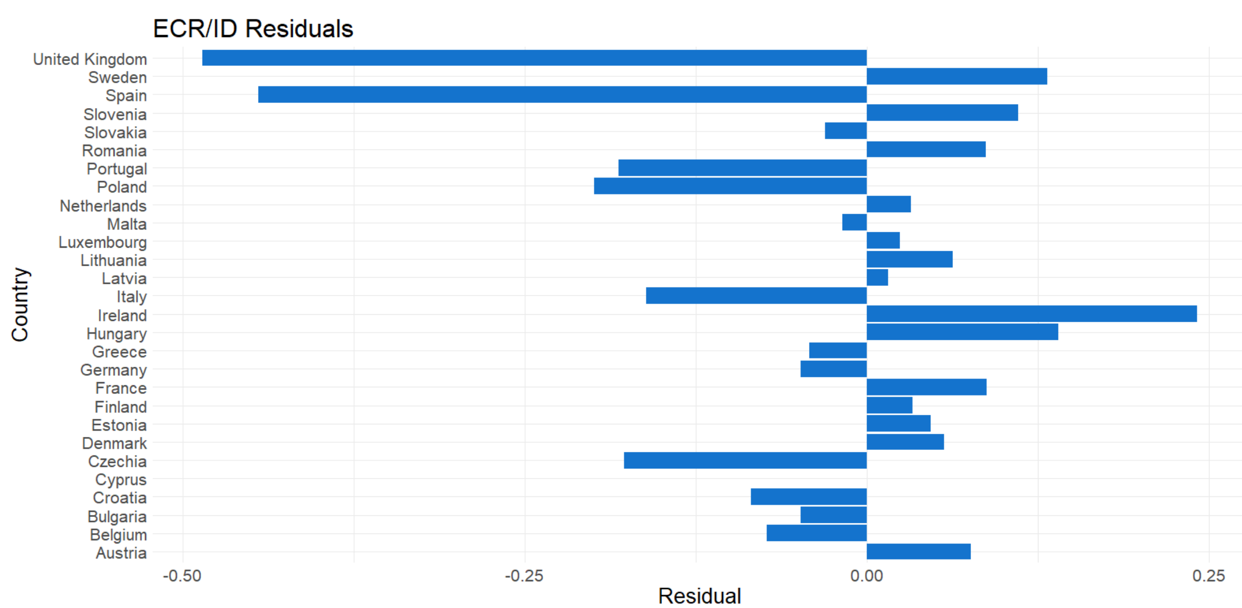




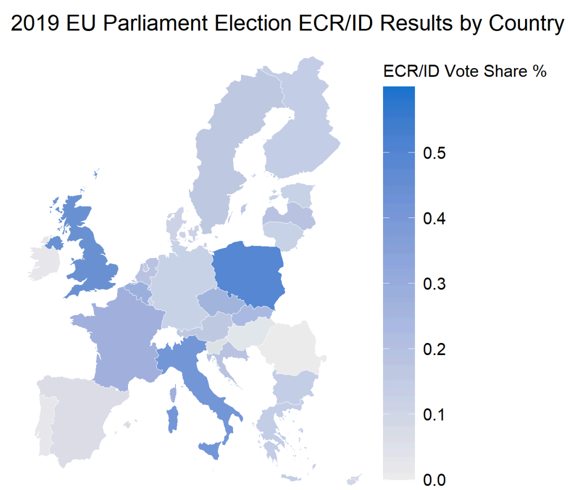
However, Portugal and Cyprus can be seen to be grayed out in the map. This is for two different reasons. Portugal is grayed out because the vote share that the model gives it is negative. Given that Portugal had a very low vote share for ECR/EFDD, as indicated by the map on the left, this is not that large of an error. Cyprus, meanwhile, is grayed out due to some variables in the model not being found for Cyprus. As a result, the model cannot be applied to that country.

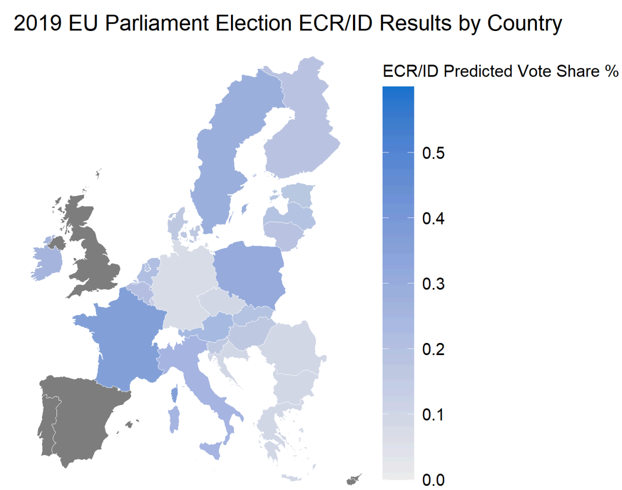
Another way to test the data is to apply it to the 2019 data. To do this, the independent variables that were found for the 2019 data that correspond to the same variables used based on the 2014 data in the model are input into the model. As an example, one of the variables when applied to the 2014 data is the total immigration in the year preceding the election, which is the total immigration in 2013. When applying the 2019 data, this would mean using the variable of total immigration in 2018.

When the 2019 data is applied to the model, the model is significantly worse. When doing so, the average error doubles to 11.63%. This causes a significant decrease in the usability of the model. Another issue with the model is that it underestimates the actual results, with the average direction of the error being -3.16%. The true nature of the model's issues can be seen when viewing a residual bar plot.



Looking at the x-axis, it can be seen that the errors in this model are massive. Spain and the UK have their vote shares underestimated by nearly 50%, with several other countries having very large underestimations as well. Fewer countries overestimate the vote share by a large margin, though the vote share in Ireland is overestimated by nearly 25%. As a result, although the error of about 12% is already bad on its own, the individual errors on the model deem it nearly useless as a predictor.





The only good thing about the model is that there is no geographic region that appears to have a specific bias in the model. However, it can be seen that the UK, Portugal, Spain, and Cyprus are all grayed out in the model. For Cyprus, this is the same case as before with some data missing. However, Spain, Portugal, and the UK are all gray due to having negative vote shares according to the model. This is especially true with the case of the UK, where it went from being one of the countries with one of the highest vote shares for this political group in reality to one of the lowest in the model. Were this the only exception it may have been something unique to the United Kingdom given the special nature of the 2019 European Parliamentary Election for them, but since this problem seems to be systematic, there are likely more pressing problems with the model.

The first potential issue with the model could stem from the educational variable, the logarithm of the percentage of people aged 15 to 64 with an education level of 0 to 2.

Traditionally, right-wing parties tend to appeal to people with less educational background. However, the model gives this variable a negative relation, meaning that a higher percentage of lower educated people would seem to decrease the right-wing vote. This would suggest that either conventional wisdom is incorrect, or that having more people with lower education affects how people with higher education levels vote. Another possible explanation would be that the coefficient is only negative due to other variables being included in the model.

Because the model includes multiple variables, each of these variables may be affecting one another. For example, if a linear regression model were made that contained only the variable of the logarithm of the percentage of people aged 15 to 64 with an education level of 0 to 2, then the coefficient for this variable would be -0.11 rather than -0.16, as the model above indicates. Every time a new variable is added or removed from the model, the coefficients of all of the other variables change, including the coefficient of the intercept.

Another thing to note is that the education level is the variable that is least likely to have changed significantly between 2014 and 2019, so it is not likely to have affected the lack of predictive value in the model. Although education levels can definitely change within 5 years, the percentage of people aged from 15 to 64 with this education level is highly unlikely to have changed in such a small time span. With a smaller age range it may be possible, though still unlikely, but because the age range is so large, the variable is unlikely to have changed enough to contribute to the massive errors in the model.

Although the education likely had some effect on the lackluster model, the immigration variables almost certainly had an outsized effect on the predictability of the model. There are four immigration variables, but they can be thought of as all adjusting one overarching variable. For example, EU27 immigration in the year of the election increases the vote share, while Total immigration in the year of the election seems to decrease the vote share. However, the increase is larger than the decrease, so the overall combination of immigration variables in the year of the election has an increase in vote share.

While it may be true that immigration has an outsized effect on right-wing share of the vote, the model may have been taking immigration as too large of a factor. The reason why this is a problem is that the immigration level from the 2014 to 2019 election changed by quite a lot. At the time of the 2014 election, the Syrian migrant crisis was in its early stages, and as such there was a larger influx of immigrants coming to Europe. By 2019, the Syrian migrant crisis had largely passed, and so even if immigration were still a factor, it was not as much on the forefront of voters' minds as it had been 5 years previous. As a result, countries where immigration was the large determining factor of the model for 2014, such as the UK and Germany, had very inaccurate predictions for 2019. By decreasing immigration, the model had fewer variables increasing the vote share, thus the variable of education, which remained roughly consistent, created negative vote shares.

Given the potential volatility of immigration as a variable, the best variable to look at as a person looking to manipulate elections in their favor is the education variable. According to the education variable, it would be expected that ECR/ID parties should promote policies that promote at least achieving an educational level of 3-4, meaning a high school or trade school degree. However, given prior research, it does seem odd to decrease the number of low education voters. As a result, this may be the one model that parties cannot make many conclusions from due to the large errors it produces.

Overall the model was not great. However, it still managed to teach how right-wing voters do not vote based on single issues, but are complex voters just like any other political grouping.

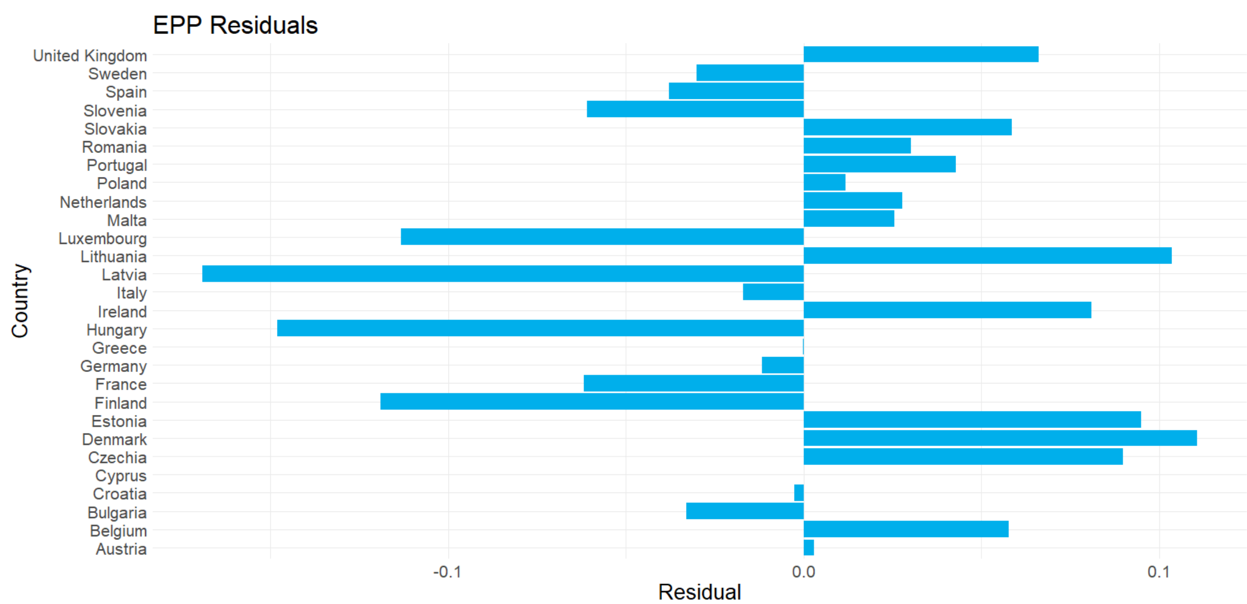
Analysis of EPP

The next model to be examined is the one describing the generally center-right party group in the European Parliament, that being the European People's Party, or EPP. The model consists of five predictors relating to four different categories of variables. These predictors are the percentage of people with a negative opinion of the EU, the logarithm of the percentage of people aged 25 to 49, the percentage of people aged 55 to 64 that achieved an education level from 5 to 8, and the non-EU 27 immigration in 2013 and in 2014. The full model can be seen below.

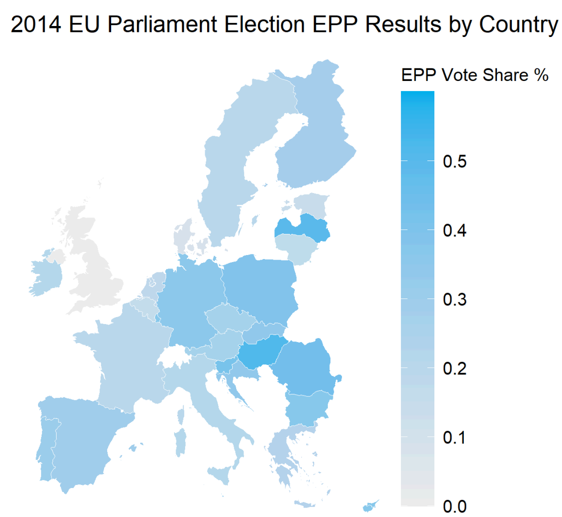
Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	1.024	0.3446	0.00725 **	
Age 55-64, Education Levels 5-8	-0.7507	0.2397	0.00503 **	
Non-EU27 in year preceding election	-0.000002532	0.000001090	0.03031 *	
Non-EU27 Immigration in year of election	0.000001977	0.0000009114	0.04171 *	
Logarithm of percent aged 25-49	0.4608	0.3383	0.18759	
Percent negative opinion of the EU	-0.3224	0.1999	0.12178	
Adjusted R^2				0.514
p-value				0.0008472

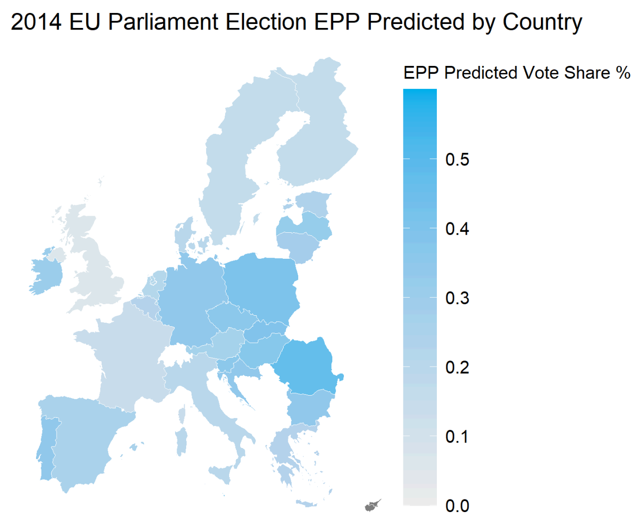
The main thing to note from the model is that the p-value is incredibly low, at 0.0008472, but the Adjusted R^2 is relatively bad, at just 0.514. This suggests that there is a lot of randomness in the data that the model does not account for. This is not ideal, and could suggest that the model could fail if applied to future elections. Despite this, the model seems to perform well on both the 2014 and 2019 data.

When applied back to the 2014 data, the model has an average error of about 5.96%. This means that the model tended to predict the vote share of EPP in each country within 5.96% of the actual vote share. This error did not seem to be biased in either direction, as the model tended to underestimate the actual vote share by just $3.731 * 10^{-16}$. This error is negligible, suggesting that there is no systematic bias on the 2014 data. The actual and predicted vote shares of this voting bloc on the 2014 data can be seen in the following residual bar plot:



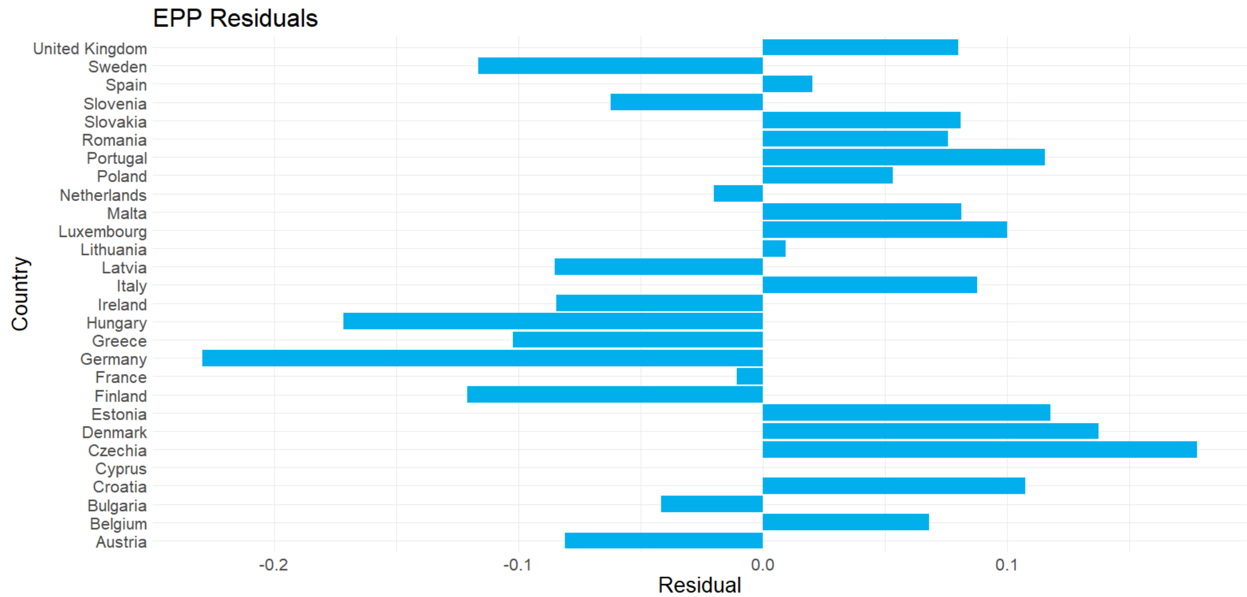
For the most part, the model seems to do pretty well on the 2014 data. It seems to overestimate the share of the vote in countries like Denmark and Lithuania, while underestimating in Latvia and Hungary, but overall it is quite good. There also does not appear to be any geographic bias based on the maps below.



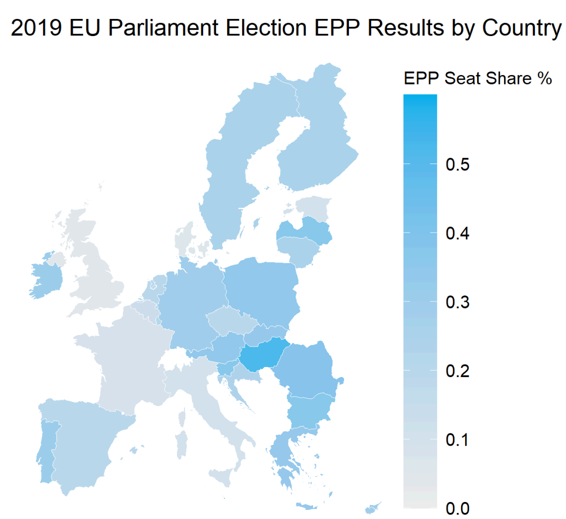


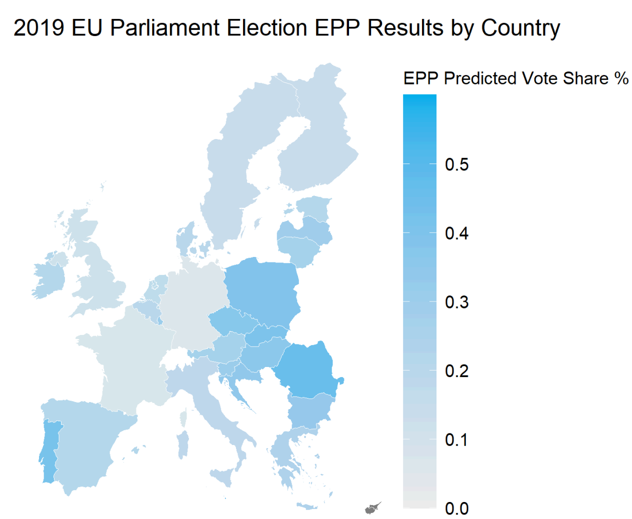
The only exception to this is Cyprus, as some of the variables used in the model did not have data for Cyprus, and as such the model could not be applied there. It can also be seen that the model seems to generally give a higher vote share to Eastern European countries, though this is consistent with the actual vote share.

The model becomes more error prone when it is applied to the 2019 data. When doing so, the average error increases to about 9.03%. While this increase is not ideal, it is not too large of an increase from the average error in the 2014 data. This suggests that the model could have a relatively strong predicting power in future elections, especially compared to the previous right-wing party group. The model also benefits from not having a strong bias in either direction, with it only overestimating the EPP vote share by an average of 0.70%. While the model is even less biased when applied to the 2014 data, this bias is still incredibly small, further suggesting the possibility of this model being relatively strong. The actual vote share of the EPP in 2019 and the predicted vote share of the EPP in 2019 based on the model can be seen in the following two maps and residual bar plot.



As expected, the model is worse when applied to 2019. This can be noted by the fact that the range on the x-axis has expanded compared to the 2014 residual plot. However, most of the countries on the map still appear to be relatively close to their actual vote share. Notably, Germany and Hungary are severely underestimated, while Czechia and Denmark continue to be overestimated.





The extent to which Germany is underestimated can be seen on the map, indicating that the model severely fails there. Despite this, one fascinating part of the model is its accurate prediction that the vote share of the EPP in France would shrink from 2014 to 2019. This seems to suggest that absent the understanding of the dramatic shift in the French political system during this time frame, the model was still able to anticipate a drop in the vote share for the center-right.

Given that this model seems to have some predicting potential, it is important to understand why the variables included may give it this predicting ability. The first variable to be examined is the percentage of people with a negative opinion of the EU. This variable has a negative relationship with the EPP vote share, meaning that the higher the percentage of people who hold a negative opinion of the EU, the lower the EPP vote share is. This is consistent with the general ideology of the EPP. The EPP consists mostly of center-right parties that support the European Union. While this grouping does contain some eurosceptic parties, such as Fidesz in Hungary (at the time of the 2014 and 2019 elections), the majority of the parties and thus likely the majority of the voters tend to hold more favorable views towards the EU. Although this variable does seem to make sense, it also does not have the strongest relationship with EPP vote share. It is an important contributing factor to the model, but at a p-value of about 0.122, it is not very significant.

The next variable included was simply the percentage of people aged 25 to 49. This relationship was logarithmic, meaning that this variable had a greater effect on the model at smaller percentages than larger ones. This variable was positive, implying that a higher percentage of people in this age range resulted in a higher vote share for the EPP. This variable is more difficult to explain, as it is difficult to determine whether center-right parties perform better with this age group across all of Europe. However, one potential explanation for this relationship is that this age range tends to be at a more stable point in their lives, with parenthood and consistent work, and thus may favor the perceived stability provided a party closer to the center, particularly ones rooted in tradition. Regardless, this is the least significant variable in the model, as the p-value for this variable is at about 0.188, far above the other variables.

The second and third most important variables are the two variables that deal with immigration. These are the number of immigrants from non-EU27 countries in both the year preceding the election and the year of the election. These variables are difficult to ascertain, as it would stand to reason that the countries with higher base populations would be the ones to accept the most immigrants as well. More interestingly, the immigration in the year preceding the election had a negative impact on the vote share, while the immigration in the year of the election had a positive impact on the vote share. This oddly suggests that the vote share decreases if immigration was high in the previous year, but increases if it was high in the year of the election.

While it may seem odd at first, further examination reveals a potential explanation for these variables being included in this way. The variable for immigration in the preceding year has a coefficient that is more negative than the coefficient for the year of the election is positive. This means that if immigration were consistent from year to year, that the overall effect of immigration would still be negative. However, this negative effect would be far larger were the immigration to severely decrease in the year of the election. This would suggest that a country where immigration is consistent from year to year would have less of a negative effect on the vote share than one where the immigration was unusually high in the year preceding the election.

Both of these variables have significant p-values, as the non-EU27 immigration from the year preceding the election has a p-value of about 0.03, while the p-value for non-EU27 immigration from the year of the election is about 0.04. It makes sense for these variables to have an overall negative effect on EPP vote share. Higher immigration, especially from non-European countries, has been associated with an increase in vote share for right-wing and far-right parties. Given that the EPP is also on the right of the political spectrum, it would be expected that some of the vote share going to these parties to the right of EPP came from EPP.

The final variable included in the model dealt with the percentage of people aged 55 to 64 that obtained an education level from 5 to 8. This is a very specific variable to be included in the model, and so it is surprising to note that it is the most significant variable, with a p-value of just 0.007. Also noteworthy is that this was a negative relationship, meaning that a higher percentage of people in this age group with higher education correlated with a lower vote share for the EPP. There is no singular explanation that can fully explain this relationship, but one potential theory is that this age group's educational attainment was important for trickling down certain political values associated with higher education from an earlier point in time. For example, nowadays higher education tends to be associated with values on the left of the political spectrum in Europe. If a higher percentage of older people received tertiary education, this would mean that these values would have existed generations ago, and potentially trickled down from one generation to the next. This theory could be entirely incorrect, but regardless the correlation between EPP vote share and this independent variable is strong.

Overall, this is a stronger model than the one for the right-wing party groups. While the independent variables could not all be explained very easily, it is undeniable that they have predictive power on the election results for EPP parties.

Analysis of ALDE and RE

The following model will be used to describe two slightly different European parties. The first is the Alliance of Liberals and Democrats for Europe, or ALDE, which was a primarily centrist and liberal group that existed following the 2014 European Parliamentary elections. The other is Renew Europe, or RE, which represents parties of the same ideologies as ALDE, but was formed following the 2019 EU Parliament elections. Although these are technically different party blocs, RE is considered to be the direct successor to ALDE, and so they are treated as the same for this model.

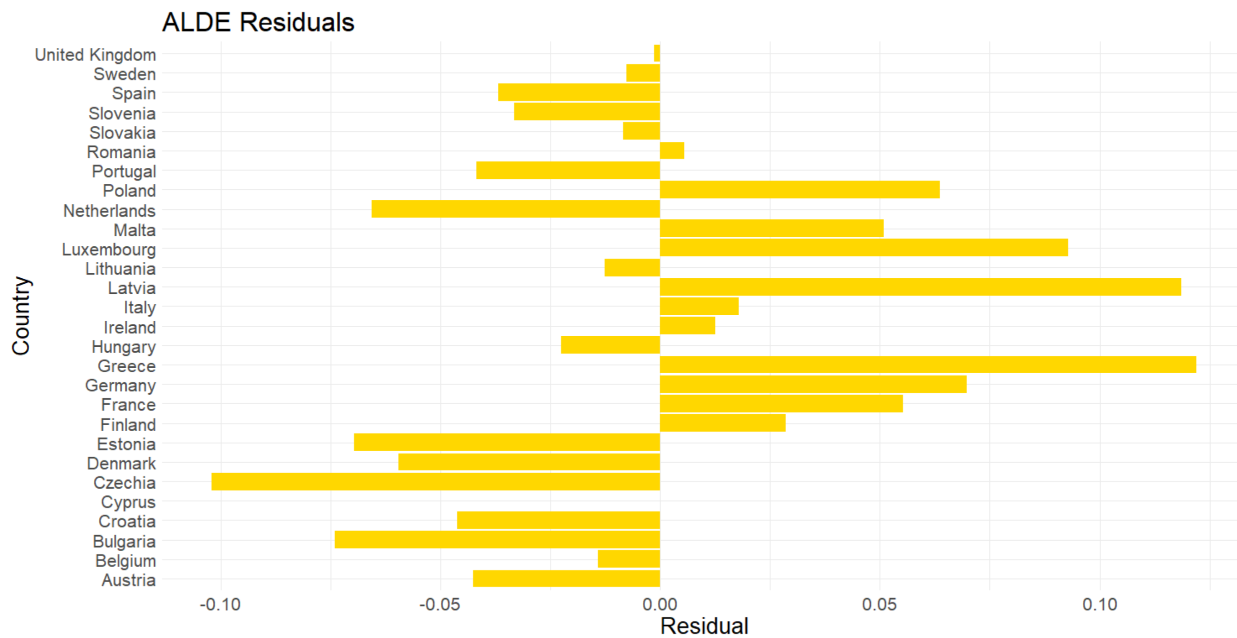
The model consists of six predictors relating to five different categories of variables. These predictors are the percentage of people aged 55 to 64 who achieved an education level from 5 to 8, the percentage of people who hold a negative opinion on the EU, the percentage of people who disagree with the notion that their country would be better off if it left the EU, the percent change in non-EU27 immigration from the year preceding the election to the year of the election, the percentage of people who are satisfied with life, and the percentage of the population that is aged between 25 and 49. The full model can be seen below.

Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	0.25121	0.30984	0.42704	
Age 55-64, Education Levels 5-8	0.84326	0.21921	0.00101 **	
Percent negative opinion about the EU	-0.38672	0.15620	0.02236 *	
Percent disagree that their country would be better off	0.38992	0.15531	0.02076 *	

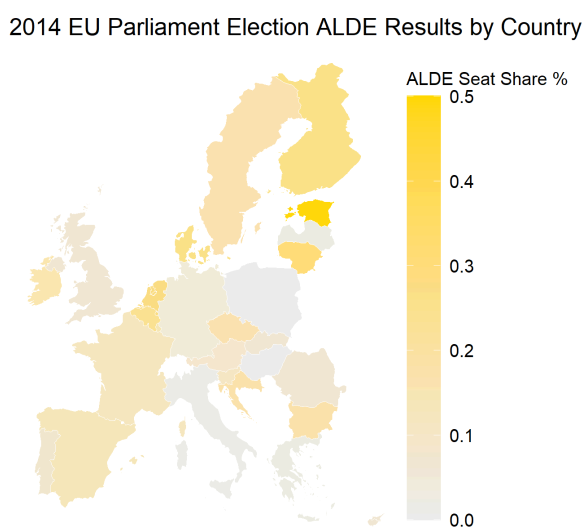
if it left the EU				
Non-EU27 immigration change from year preceding election to year of election	-0.29619	0.07817	0.00115 **	
Percent satisfied with Life	-0.15943	0.09594	0.11216	
Percent age 25-49	-0.81299	0.76598	0.30117	
Adjusted R^2				0.6627
p-value				0.00005327

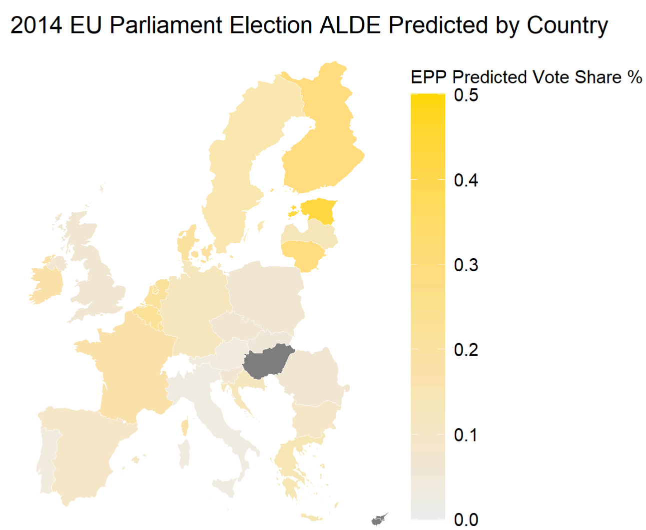
The p-value on this model is extremely low, at $5.327 * 10^{-5}$. This is effectively a p-value of 0, showcasing just how strong the model is on the 2014 data. The Adjusted R^2 is also relatively strong, at 0.6627. Although there is still a lot of unexplainable randomness in the data, this R^2 is still one of the strongest and thus suggests that these variables have more predictive power for this model than the predictive variables of the previous models.

Applying this model back to the 2014 data produces an average error of about 4.73%, meaning that the true ALDE vote share tended to be within 4.73% of the vote share that the model predicted. This is already a fairly strong result, but the important thing is that the error did not seem to have any bias. The model tended to overestimate the ALDE vote share by $1.05 * 10^{-16}$, effectively meaning that the model was not biased in either direction for the vote share. This can be seen in the following residual bar plot.



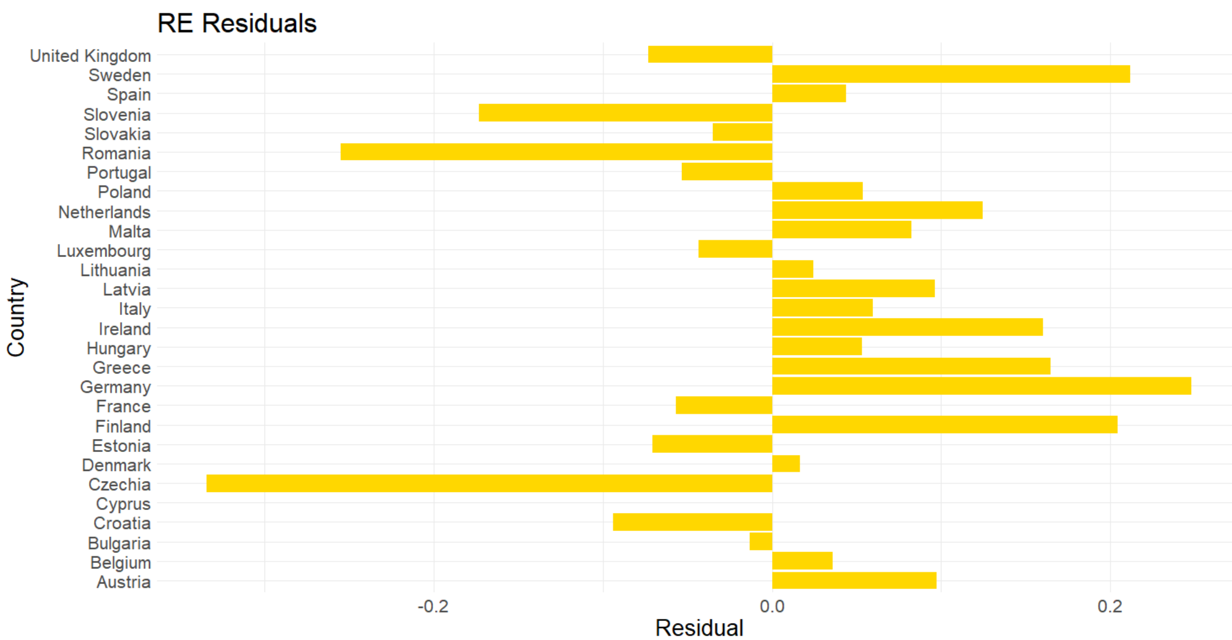
Some countries seem to have further off predictions than others. Notably, Greece, Latvia, and Luxembourg have their ALDE vote shares severely overestimated by the model. The maps below seem to indicate a lack of geographical bias in the model as well,



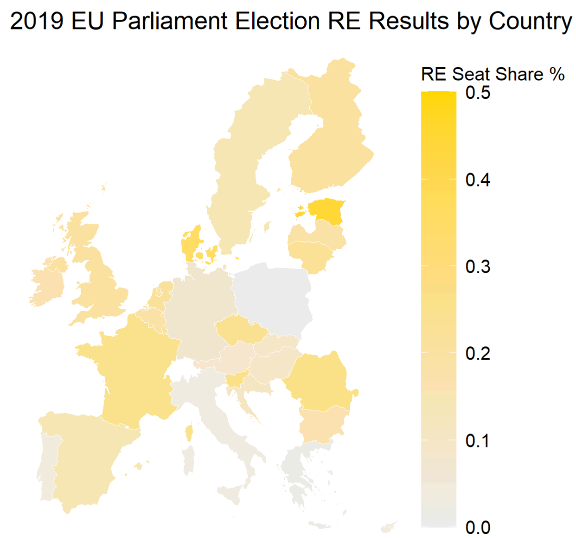


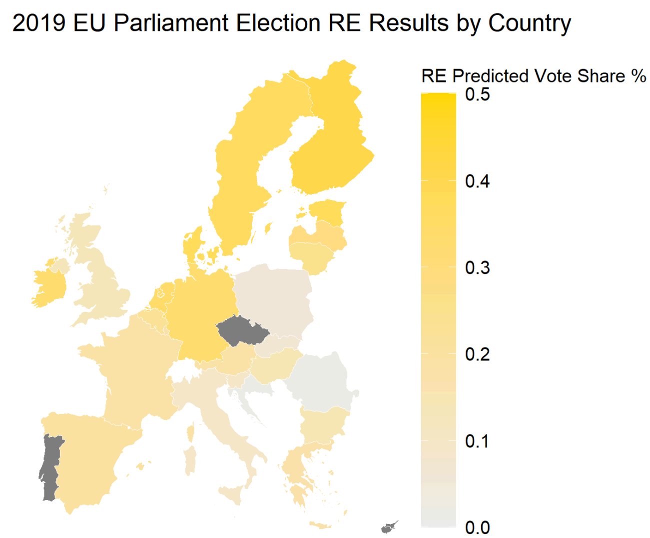
Overall, the maps seem to indicate that the predicted votes are more smoothed out, but the predictions still seem to be fairly strong. The only issues that can be seen on the map are Cyprus and Hungary. Cyprus is not colored in as at least one of the variables in the model does not have data for Cyprus. As a result, the model cannot be applied to Cyprus. The model does work on Hungary, but it gives Hungary a negative ALDE vote share. In order to keep both maps consistent, limits were kept the same on both maps, and so Hungary's datapoint cannot be mapped. However, the model does still seem to be fairly strong for Hungary, as is indicated by the low error for Hungary in the bar plot.

When applied to the 2019 data, the model now predicts RE vote share rather than ALDE vote share. Unfortunately, the errors do increase compared to the predictions on the 2014 data. The average error on the model increases to 10.67%, meaning that the model tends to predict the actual RE vote share only within 10.67%. Luckily, the model remains relatively unbiased. It only tends to overestimate the RE vote share by 1.76%. This is worse than the bias on the 2014 data, but it is still very small and not a significant enough bias to cause any worry about a systematic error. These errors and biases can be seen in the following residual bar plot.



The model produces its biggest errors for Romania and Czechia, where the model severely underestimates the RE vote share. Without these outliers, the model becomes far stronger, though the data points must be kept to avoid causing a larger bias in the model. The model again does not seem to have any strong geographic biases, as seen below.





However, the model again has issues with countries receiving negative vote share. Both Portugal and Czechia are given negative RE vote shares according to the mode. For Portugal, this issue is minor, as Portugal did not receive a high RE vote share and so the difference with the predicted vote share is small. However, the issue with Czechia is very concerning, as Czechia received a very large RE vote share and so it definitely should not have been given a negative vote share. The most likely explanation for this is that the parties categorized as RE in Czechia may hold different traits or have different supporters compared to RE parties in many other countries.

Any of these outliers can also potentially be explained by understanding the relationship between the independent and dependent variables. The first variable to understand is the percentage of people aged 25 to 49. This variable has a negative relationship with ALDE/RE vote share, meaning that a higher percentage of the population in this age range is related to a decreased vote share. It is difficult to explain fully why this variable is included in the data. One possible explanation is that this age range tends to vote more for EPP, as indicated by the fact that this same variable had a positive correlation in that model. Regardless, this is not the most predictive variable in the model, with it having a fairly large p-value at 0.301.

The next variable to understand is the percentage of people that are satisfied with life. The interesting part about this variable is that it has a negative correlation, meaning that a higher percentage of people satisfied with their lives corresponds to a lower vote share. This is difficult to fully explain, as generally dissatisfaction with life has been seen to be correlated with an increase in vote share for parties on the extreme ends of the political spectrum. Given that this variable only has a p-value of 0.112, it is possible that this variable does not have much predictive power on its own, but is useful for altering the other variables in the model.

One of the more predictive variables in the model was the non-EU27 immigration change from the year preceding the election to the year of the election. At a p-value of 0.001, the variable appears to be strongly correlated. This variable has a negative correlation with the vote share. This means that the higher the percent increase in non-EU27 immigration from the year preceding the election to the year of the election, the lower the vote share is for ALDE/RE. This is completely consistent with what would be expected, as increasing immigration, especially from non-European countries, tends to be correlated with an increase in right-wing views. Given that many people to the right of the political spectrum may vote for the centrist parties in their respective countries, a further polarization of their views could make them more likely to vote further right than they would otherwise.

The next two variables both deal with opinions relating to the European Union. First is the percentage of people who disagree with the idea that their country would be better off if it left the EU. This variable, expectedly, has a positive correlation. This means that a higher percentage of people that don’t think that their country would be better off if it left the EU is correlated with a higher vote share for ALDE/RE. These political groups are generally considered to be europhilic, so it is not surprising that if fewer people want to leave the EU, the more likely the vote share for these centrist and liberal parties is to increase. In many countries, such as in France with Macron's REM, this trait is the defining characteristic of the ALDE/RE parties. As a result, it is also unsurprising that the variable is significant, with a p-value of 0.021.

The other variable relating to opinions about the EU is also easy to explain. There exists a negative correlation between the percentage of people who hold a negative view of the EU and the vote share. This implies that the higher percentage of people who hold a negative opinion on the EU, the lower the vote share of ALDE/RE parties will be. This variable is roughly just as significant, with a p-value of 0.022. Interestingly enough, the coefficients of both EU opinion variables are very similar, though this may just be a coincidence.

The final variable involved in the model is the percentage of people aged 55-64 that achieved an education level from 5 to 8. This variable has a positive correlation, meaning that the more people in this age range that achieve tertiary education, the higher the vote share is for ALDE/RE parties. Unlike the education variables in the previously discussed models, this can be more easily explained. While not always true, people who achieve these higher forms of education, especially those at an older age, are more likely to hold liberal political views. These people are far more likely to hold more left-leaning social and cultural views, turning them away from right-leaning parties. They are also more likely to be more affluent and thus hold more right-leaning economic views, turning them away from left-leaning parties. As a result, many may settle for liberal parties in the center of the political spectrum, if they have an option to do so. Furthermore, many of these older members of society are far more likely to be in positions of authority, whether in personal, educational, or professional interactions. By having them hold so much influence, the liberal views that they hold may trickle down to younger generations.

Based on the model, the best thing for RE parties to do now is to promote the European Union. By improving how people view the EU, people may be more inclined to vote for the centrist and europhilic RE parties rather than the parties further from the center that tend to be more eurosceptic. However, this is something that is difficult to exploit, as it is difficult to change the opinions of people while other parties are trying to do the same.

Ironically, one thing that RE parties may want to do is oppose immigration from non-EU27 countries. Although this may not necessarily be in line with their policies, it is clear that increasing immigration from these countries is correlated with a decrease in their vote share. It may not be in their best interest to allow many non-EU27 immigrants into the EU, or else they may leak votes to parties that are more anti-immigrant or xenophobic.

As a whole, the model to predict the centrist and liberal parties is fairly strong, but not perfect. It still has a large error when applied to future elections, which may make it not ideal. However, it still has predictive power, and has statistical significance, so perhaps the model could still be used in the next European elections.

Analysis of Greens/EFA

The model describing the Greens/EFA political grouping is one describing a very heterogeneous party bloc. While the party is primarily for the Green parties of the European Parliament, many other party groups are included. Some of the other groups included are the European Free Alliance, or EFA, which is made up of regionalist parties, various Pirate Parties, as well as Volt, a European Federalism party. Although this group is fairly diverse, most of the parties tend to range from center-left to left-wing in ideology.

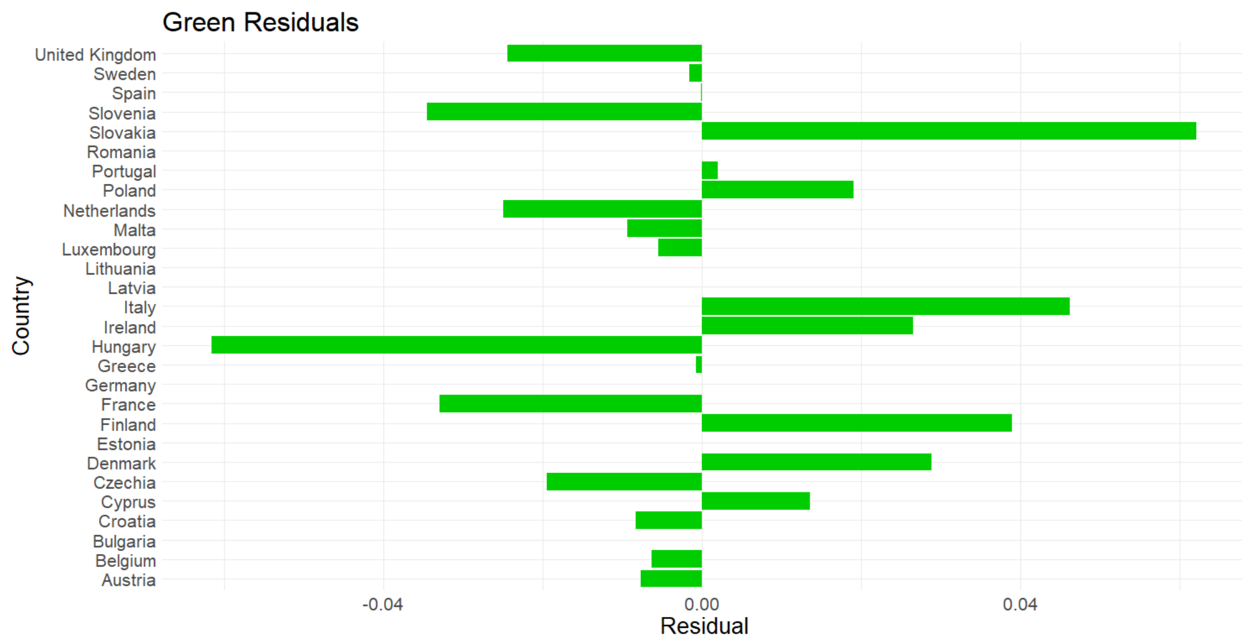
The model consists of eight different variables, but these variables are only part of three categories. The two variables included in the model that don't relate to education are the percentage of people who believe that climate change is one of the biggest issues facing the EU and the EU28 immigration as a percentage of the population in the year of the election. Most of the education variables have to do with the percentage of people who achieved an education level of 5 to 8, with the age groups of 55-64, 45-64, 45-54, 35-44, and 25-64 included. There also exists a single education related variable for other education levels, that being the percentage of people who have an education level from 0 to 2 that are aged 35 to 44. The full model containing all eight variables can be seen below.

Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	0.17248	0.05449	0.00745 **	
Belief that climate change is one of the top two issues facing the EU	0.49254	0.21983	0.04315 *	
Immigration from EU28 countries as percentage of population	0.51847	0.65219	0.44090	
Age 55-64, Education Levels 5-8	-5.74917	2.96966	0.07492 .	
Age 45-64, Education Levels 5-8	12.68917	6.04590	0.05593 .	
Age 45-54, Education Levels 5-8	-5.98786	3.15705	0.08031 .	
Age 35-44,	1.11395	0.48784	0.03986 *	

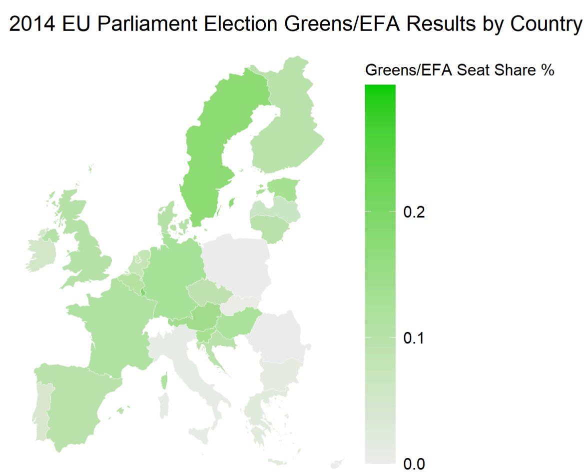
Education Levels 5-8				
Age 35-44, Education Levels 0-2	-0.17722	0.08189	0.04965 *	
Age 25-44, Education Levels 5-8	-2.34616	0.81133	0.01261 *	
Adjusted R^2				0.5791
p-value				0.007511

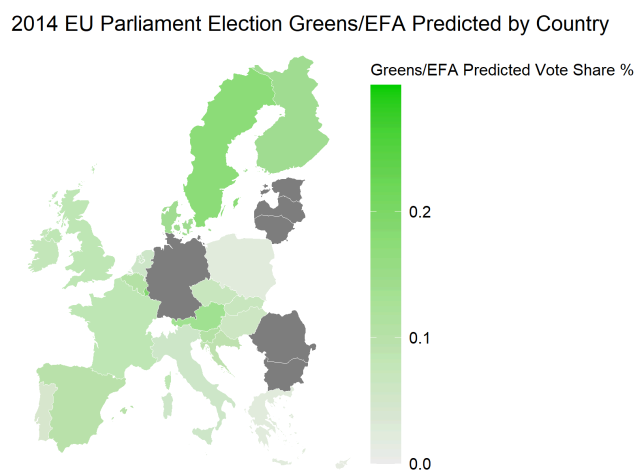
The p-value on this model is sufficiently low, at 0.007511. While higher than the previous models discussed, it is still well below the threshold needed to declare a model statistically significant. The Adjusted R^2 is 0.5791, which is not ideal but is relatively consistent with the previous models. While other models existed with lower p-values, this was the best model that could be found that both had a low p-value and a sufficiently high R^2.

When the model is applied to the 2014 data, it produces an error of just 2.16%. This is by far the lowest error of any of the models, suggesting that even though the p-value and R^2 values are not ideal, the actual predictive value of the model is stronger. Furthermore, the model produced essentially no bias in these errors. On average, the model overestimated the vote share for Greens/EFA by $4.41 * 10^{-16}$. This number is so incredibly small in this context as to make it reasonable to say that no bias exists. This lack of bias can be further seen in the residual bar plot.



The thing to note here is that the x-axis is far narrower than in the previous bar plots, with only two errors being greater than 6%. Those two errors occur in Hungary, where the vote share is underestimated, and Slovakia, where the vote share is overestimated. Still, neither of these is an incredibly large error in relation to the previous models seen. The errors also seem to be evenly dispersed in either direction, also consistent with the nonexistent bias noted before. Despite all of this, the model does have some issues, as can be seen on the following maps.

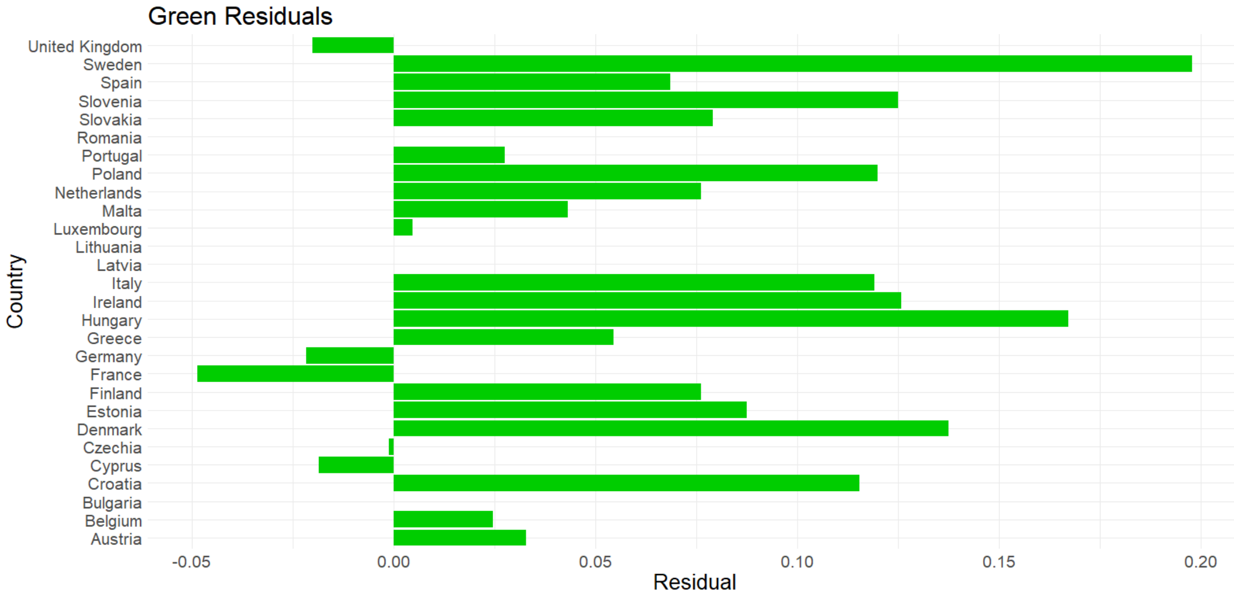




While geographic bias is not a particular issue here, there is an issue with lack of data for some of these countries. One of the variables included in the model had data for most of the countries in the dataset, but not all. As a result, Germany, Romania, Bulgaria, and the Baltic States are not included in the model, and cannot have their vote shares predicted. This is especially bad for Germany and the Baltic countries, all of which had a substantial portion of their votes go to Greens/EFA parties. If the data for the independent variables missing could be found then the model could be more accurately evaluated, but as it stands it can only be evaluated by the remaining 22 countries in the dataset.

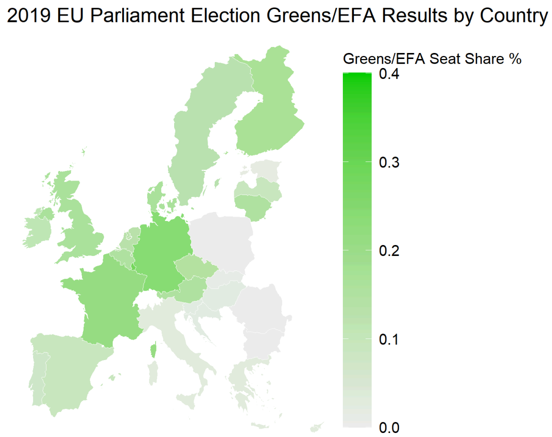
Although these issues are concerning, what makes this model especially good compared to all of the other ones is the fact that it does the best at predicting the 2019 elections. When the model is inputted with data from 2019, it produces an average error of 7.47%. While again not ideal, it is still a much better error than the other models. One potential explanation for this is the fact that Green parties are likely to be the ones with the most consistent agenda across all of Europe. While every other party group tends to have a variety of policies that may attract voters, the Green parties are primarily ones interested in combating climate change and protecting the environment. That is not to say that these are single-issue parties, but that their support may be somewhat more predictable.

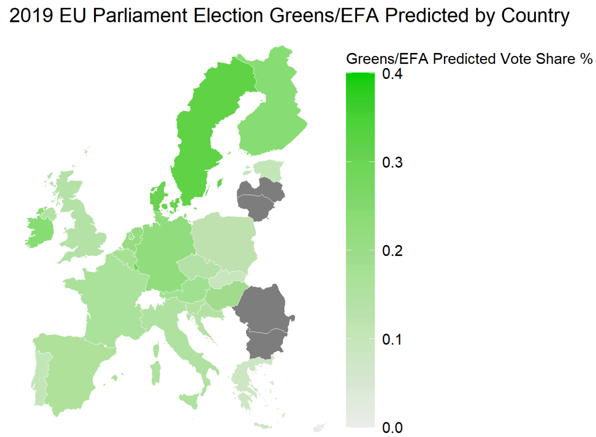
While the error may be small, the bias on the 2019 data is quite large. The model tends to overestimate the vote share by about 6.54%. This is the most concerning aspect about this model, and can be easily visualized in the following residual plot.



Given the dramatic change in bias from 2014 to 2019, the only explanation would be that a highly volatile variable was included in the model. Of the possible variables, the only one fitting this description is the one dealing with whether or not climate change is one of the biggest issues facing the EU at the moment. This explains why the model becomes as biased as it does, as the level of environmental concern likely increased in nearly all countries from 2014 to 2019, but not all countries had a green party emerge as a result. Instead, many of these countries may have had the policies of green parties adopted by other left-leaning parties, or potentially even parties in the center or right. If people voted directly for European Parliamentary parties then maybe the model may have rung true, but since people vote for national parties, there may simply have been a lack of prominent Greens/EFA parties to vote for.

Luckily, the maps had an improvement from the 2014 to the 2019 data.





By using the 2019 data, the model can now be applied to both Germany and Estonia. While the other countries still cannot be analyzed, Germany is a crucial country to have the model work on due to how many MEPs it has and how influential it is in the EU. Despite this, the map also clearly shows how the model severely overestimates the vote share in places that had low vote share for the Greens/EFA. This is particularly noticeable in Eastern Europe and Italy, which are given far higher vote shares than they received in reality. This is further proof that the issue with the model is the volatility of how many people view climate change as one of the most important issues in the EU. Given that the model did not severely overestimate the countries which actually had a prominent vote share from these parties, it could stand to reason that the countries with the highest vote share are also the ones that held the highest concern for climate change in the 2014 election. Meanwhile, the other countries had a sharp increase in concern from 2014 to 2019, resulting in the bias of the model.

Although it is faulty, concern for climate change is still one of the more significant variables in the model. The percentage of people who believe that climate change is one of the biggest issues facing the EU has a p-value of 0.043. While this is not very significant, it is still statistically significant at a confidence level of 95%. The explanation for why this variable has a positive correlation with vote share is fairly self-explanatory - countries where climate change is more of a concern are also countries where voters will vote for parties that will address those concerns. The main fault of the inclusion of this variable in the model is that there was a sharp increase in concern from 2014 to 2019, as climate change became a more salient issue. Perhaps if the coefficient were adjusted and the level of concern did not increase dramatically throughout the EU, this variable could still be used to predict future elections.

In contrast, the least significant variable used in the model is the EU28 immigration as a percentage of the population in the year of the election. This variable has a p-value of 0.441, which is not close to being significant at all. Though this may be true, this variable does still make sense as a predictor given that it has a positive correlation. Countries with higher immigration from other countries in the European Union are also countries that are likely to have better opportunities for higher education. Given the correlations between higher education and vote share, this would explain the addition of the immigration variable.

All of the remaining variables deal with education. All of the variables to do with education are identical, though they correspond to different age ranges and educational levels. Not all of them are statistically significant on their own, but they all are close to one another in significance: the lowest p-value for an educational variable being 0.012, and the highest being 0.08. The education variables are difficult to explain individually, as the negative and positive correlations play off of each other.

The simplest showcase of how these variables work in tandem is by looking at the age range of 45-64 with education levels 5 to 8. At first, it may seem that this age range has a positive correlation with a coefficient of 12.689. However, two more age ranges exist for the same education level, those being 45-54 and 55-64. Both of these have negative correlations of -5.988 and -5.749, respectively. This negative correlation seems paradoxical at first, but must be added with the previous 12.689. When combined together, this means that the coefficients are actually equal to 6.701 (12.689 - 5.988) and 6.94 (12.689 - 5.749), respectively. This applies to the combination of all of the higher education variables as a whole, all playing off of each other.

As a whole, the large coefficient for the age range of 45-64 means that in most cases, a higher percentage of tertiary education is correlated with higher vote share for the Greens/EFA. This is consistent with what would be expected, as being concerned enough about the environment in order to vote for green parties is generally something that requires a degree of affluence and knowledge associated with higher education. While anyone can be concerned with the environment, not everyone can afford to have this be their main concern when voting, and so people who are less affluent and less educated have been seen to prioritize other issues.

This also explains the final educational variable, that being the percent of people aged 35 to 44 that earned an educational level between 0 and 2. This variable has a negative correlation as having more people with lower educational background will likely cause votes to be pulled away from the green parties and towards other parties on the left that concern themselves more with economic policy.

Based on this model, Greens/EFA parties would greatly benefit from promoting policies that would encourage people to seek out higher education. Overall, it appears that having a higher percentage of the population with a university degree will cause the Greens/EFA parties to be larger as a result. Other than that, the only other thing that these parties can do that would have a strong chance of increasing their vote share is by increasing the salience of climate change as an issue. Oddly enough, this would likely be less effective, as this could simply increase the vote share for parties on the left as a whole, as Greens/EFA parties may only exist when there are a sufficient number of highly educated voters.

The Greens/EFA model in some ways has some large problems due to its biases, but it is also one of the most promising due to the small errors it produces. Perhaps if these biases were to be accounted for, then this party group would be able to have the strongest model of the 6 party groups being examined.

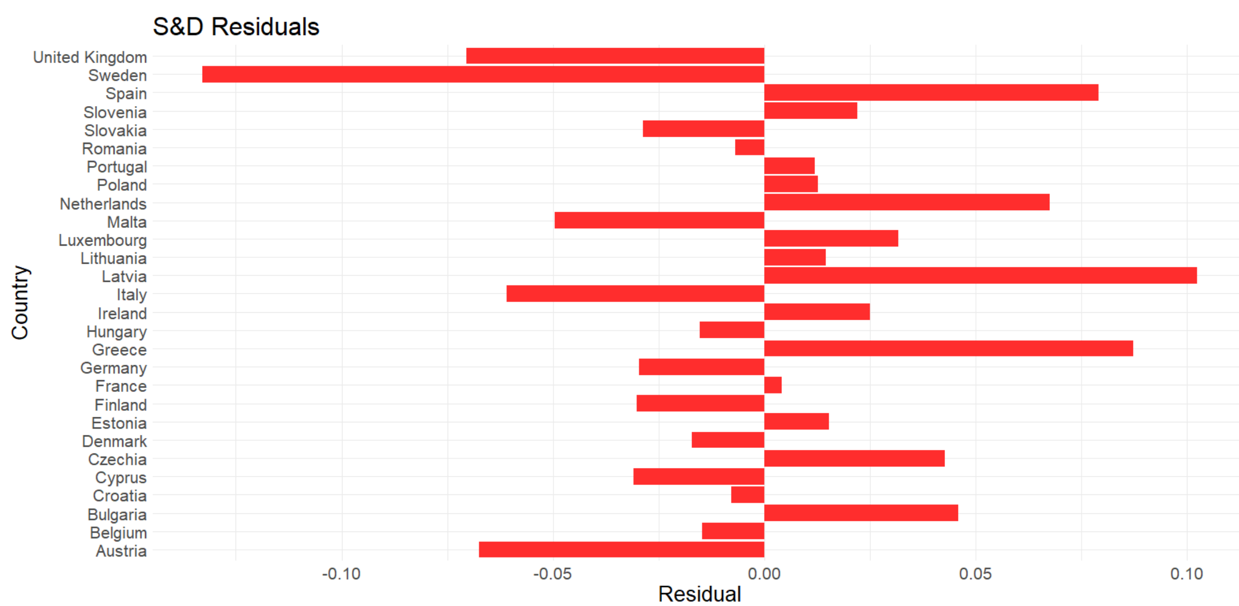
Analysis of S&D

The next model is used to predict the vote share of the Progressive Alliance of Socialists and Democrats, otherwise known as S&D. This group is made up of center-left parties across the European Union, primarily ones that identify with the social democratic ideology. The model to predict their vote share involves just three variables, part of only two categories. These variables are the percentage of people who hold right-leaning political views, the percentage of people aged 15 to 64 that achieved an education level from 3 to 4, and the percentage of people aged 30-34 who have an education level from 0 to 2. The statistics behind this simple model can be seen below.

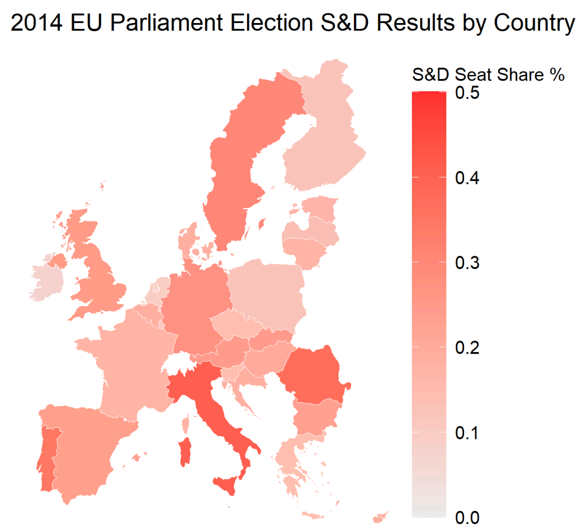
Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	-0.18741	0.09044	0.049154 *	
Age 30-34, Education Levels 0-2	1.19364	0.16232	0.000000136 ***	
Age 15-64, Education Levels 3-4	0.54578	0.13353	0.000422 ***	
Right-leaning political views	-0.23410	0.16786	0.175917	
Adjusted R^2				0.7103
p-value				0.0000002995

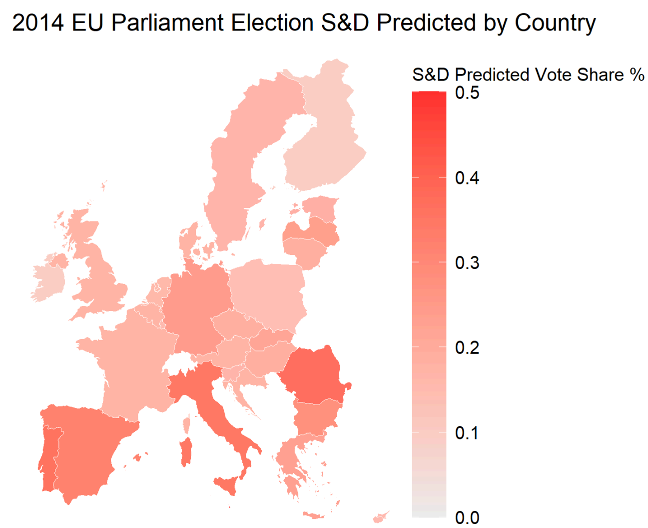
The p-value in the model is very low, at 2.995 * 10^{-7}. This is the lowest p-value of any of the models which is especially notable by the fact that this model has the fewest variables involved. This model also has the highest Adjusted R^2, at 0.7103. Both of these metrics seem to suggest that this model is the strongest of all the European party blocs.

When the model is applied onto the 2014 data, the average error is 4.02%. This may be larger than the average error for the Greens/EFA model, but it is still very low. Furthermore, the model also has no bias on the 2014 data. The model underestimates the vote share by an average of 5.45 * 10^{-17}, which is essentially a bias of zero. These small errors can be seen more clearly when looking at the residual bar plot.



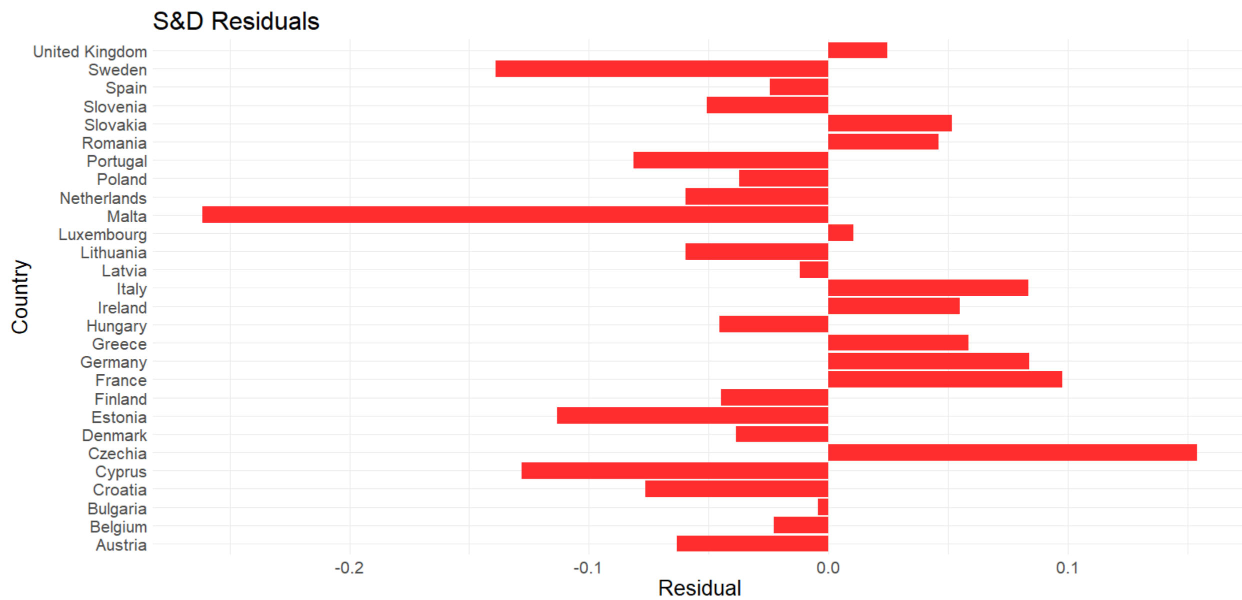
For the most part, the model seems to have small errors. While larger errors exist, they generally counterbalance one another in either direction. The only unusually large error is in Sweden, where the vote share is severely underestimated. This may be partially due to the fact that Sweden has a very large and established Social Democratic party that attracts voters of different demographics compared to Social Democratic parties in other nations. Overall, there also did not seem to be any geographic biases in the model.



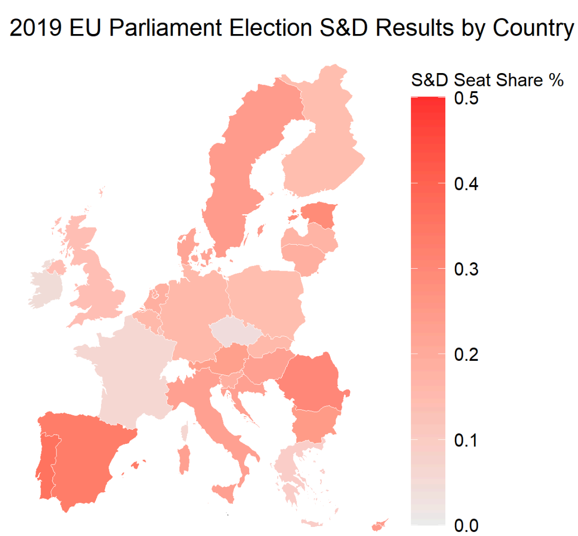


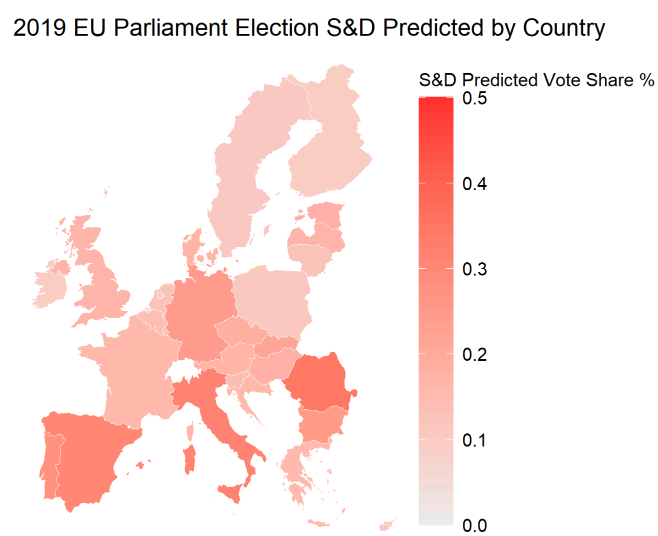
Other than Sweden, every other country visually appears to have a similar real and predicted vote share. A benefit with this model is that every country in the dataset had a substantial vote share for S&D, so no countries had a severe error caused by social democratic parties simply not existing. Also, unlike the previous models, this model can be applied to every country in the EU28, as the data is available for every country.

The true superiority of this model is showcased when it is applied to the 2019 data. When the 2019 data is input into the model, the resulting error is only 6.88%. While this is an increase on the 2014 data, it is the smallest increase in error from 2014 to 2019 of any model, and is also the smallest error on the 2019 data of any model. The model does have a slight bias, with it underestimating the true vote share by an average of 2.13%. While this should be noted, it is still quite small. This bias may also be partially explained by the results the model gives for Malta, as shown in the residual bar plot.



Malta is by far the largest error in the model, with the underestimation of the vote exceeding 25%. Sweden is again underestimated by a large margin, while Czechia is overestimated by a large margin. Despite all of this, the majority of the predictions are fairly accurate. The model also fails to show any geographic bias, as shown in the following maps.





There are individual countries with noticeable errors, such as Czechia, but nothing systematic for a certain region of Europe. Given that the error was not super large in 2014 for Czechia, one possible explanation for the error is dissatisfaction with the actual party or parties that represent S&D, rather than the social democratic ideology as a whole.

The simplest explanation for why this model is so good is the fact that there exists variables that are able to predict the vote share without overfitting. By having only three variables involved, the model is more likely to stay consistent from one election to the next, and thus may have more predicting power. However, having a model with these few variables is not possible for every party bloc, so it is important to understand why the three variables are so predictive of the vote share.

The first variable, the percentage of people with right-leaning views, is the simplest to understand. This variable has a negative correlation with the vote share, implying that if more people claim to have right-leaning views in a country, then that country will have a lower vote share for the S&D. Given that S&D is a left-leaning party group, this makes sense. If a country is more right-leaning as a whole, it would be inferred that more of the votes in that country would go to right-leaning parties rather than left-leaning ones. However, this variable has the weakest relationship with vote share of the three, as it has a p-value of only 0.176.

Both of the other variables deal with education, but of different educational levels. The model has a positive correlation between the percentage of people aged 15 to 64 who have an educational level of 3 to 4 and the vote share for S&D. This relationship is highly significant, with a p-value of 0.0004. This is the largest age range available in the dataset, implying that were it to be available, this variable may be able to be generalized as the percentage of the total population that achieved an education level of 3 to 4. This education level essentially includes people who achieved an education up to high school or a trade school. As such, this demographic is more likely to be working class, and more likely to have lower wages as a result. Given the policies of social democratic parties, it would make sense that these types of people would be inclined to vote for S&D.

These same ideas apply to the other education variable, though more dramatically. The model has a positive correlation between the percentage of people aged 30 to 34 who have an educational level of 0 to 2 and the vote share for S&D. This educational category is equivalent to no higher than middle school in the US, and so the people classified in this educational category would be even more likely to hold working class jobs. It is difficult to determine why this specific age group was chosen, but perhaps it is simply because this age group serves as a generalization of the whole population. Regardless, this is the most significant variable, with a p-value of $1.36 * 10^{-7}$.

As a whole, this model indicates that it would be difficult to manipulate the vote share of S&D parties. While decreasing the number of people with right-leaning political views would be helpful, this is an improbable task given that every other party is also trying to sway voters to their own political ideologies. At best, S&D parties should help promote living standards for the working class in order to decrease the need for attending higher education. By doing so, they may be able to prevent some of the bleeding of their vote share to the Greens/EFA, who seem to perform better with left-leaning higher education voters. However, promoting working class living standards could also risk the S&D losing vote share to the traditionally left-wing GUE/NGL. as a result, the S&D is in a precarious position where no matter what they do, they are at risk of losing vote share to one of the two other major left-leaning party groups

Although the use of the model is difficult, it cannot be overstated just how predictive the S&D model is. By every metric, it is the strongest model of the six political groupings. If S&D parties can figure out a way to manipulate the variables in this model to their benefit, it would be the most likely model to actually reflect on the real world. If a way to use the model is not found, then it will simply remain a predictive tool for future elections.

Analysis of GUE/NGL

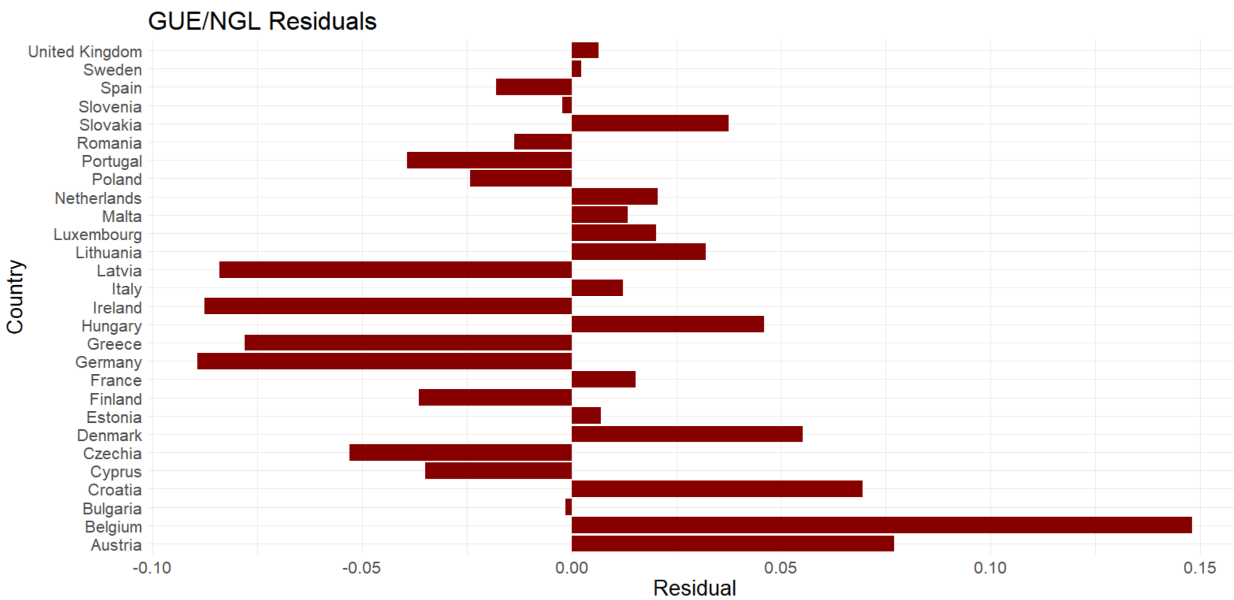
The final model is used to predict the vote share of The Left in the European Parliament, otherwise known as GUE/NGL. This group includes left-wing and far-left parties, generally associated with the ideologies of socialism or communism. The model involves seven variables, but all of them relate to education. These education variables can be put into two categories, those to do with the percentage of people who have an education level from 0 to 2, and those to do with the percentage of people who have an education level from 3 to 4. For the former, the age ranges are 25-64 and 45-64, while for the latter the age ranges are 15-64, 25-64, 55-64, 25-34, and 45-54. This education-based model can be seen below.

Model				Value
Variable	Coefficient	Standard Error	p-value	
Intercept	0.6754	0.1450	0.000152 ***	
Age 55-64, Education Levels 3-4	-2.0780	0.5752	0.001736 **	
Age 45-64, Education Levels 0-2	1.3158	0.5718	0.032269 *	
Age 45-64, Education Levels 3-4	-1.7876	0.8664	0.052307 .	
Age 25-64, Education Levels 3-4	7.3121	1.9886	0.001495 **	
Age 25-64, Education Levels 0-2	-2.2183	0.6486	0.002714 **	

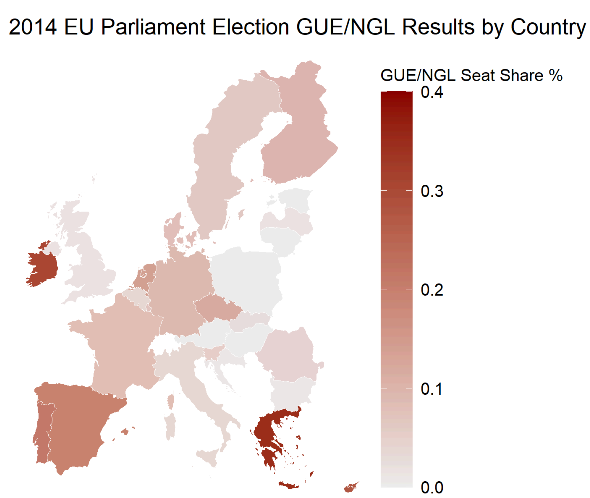
Age 25-34, Education Levels 3-4	-1.8467	0.6643	0.011556 *	
Age 15-64, Education Levels 3-4	-2.6108	1.1414	0.033198 *	
Adjusted R^2				0.6011
p-value				0.0003302

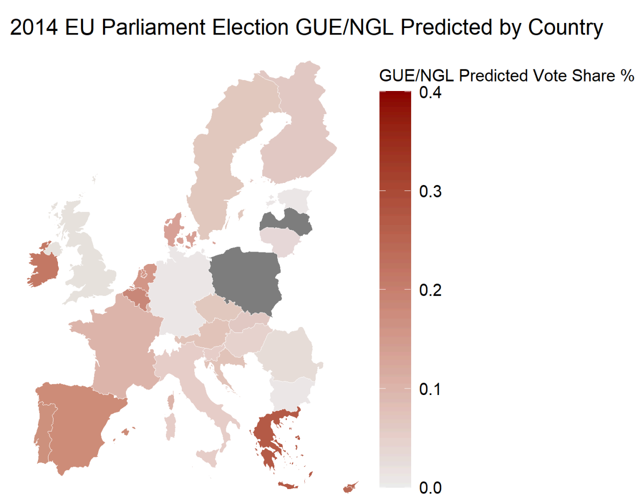
The p-value in the model is fairly low, at 0.0003302. This is not the lowest p-value of the models, but is still well below the threshold necessary at a 95% confidence level. The Adjusted R^2 is not great, at only 0.6011, but it is the best that could be obtained without compromising the model's accuracy. While much of the randomness in the data could not be explained by the model, at least a majority of it could.

Although the above metrics are not ideal, the model seems to do well when it is applied to the 2014 data. The average error of the model on the 2014 data is just 4.02%, effectively equalling the smallest average error of the models. The model also lacks any bias, as on average it underestimates the GUE/NGL vote share by just $5.50 * 10^{-16}$, essentially zero. These errors can be seen in the following bar plot.



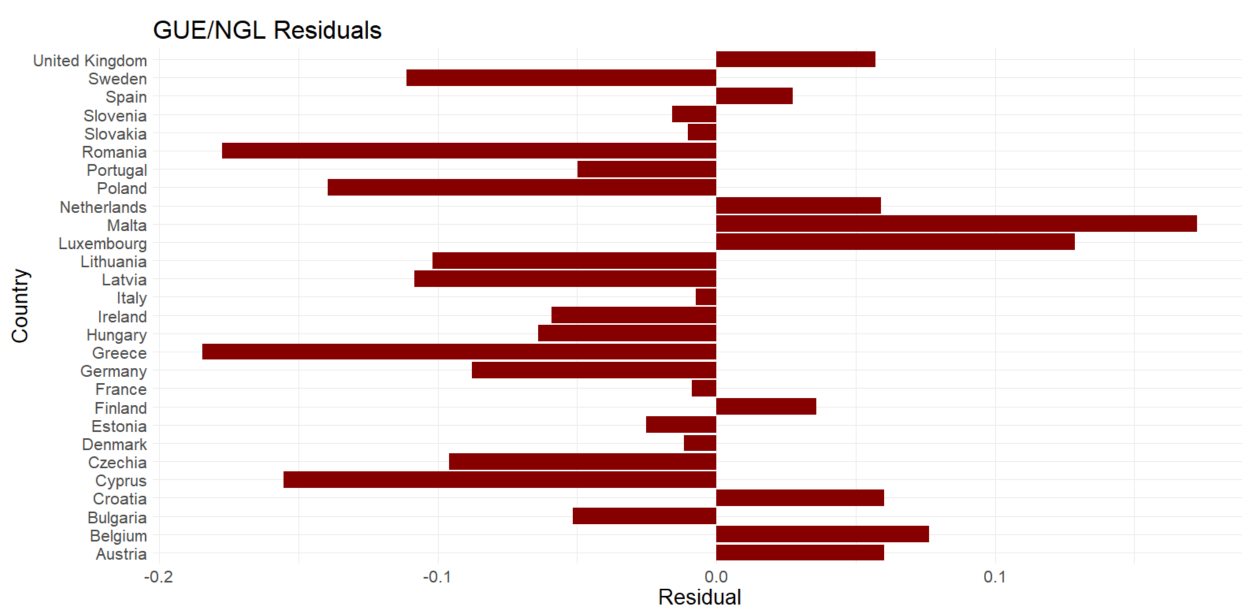
Many of the countries have very small errors, though some larger errors exist. The one that sticks out the most is the overestimation of the vote in Belgium. While numerous explanations may exist for this, one possibility may simply be due to the uniqueness of how European elections are conducted in Belgium due to it being a multinational country. By splitting the elections by nationality, there may simply be too many political parties of other ideologies, thus reducing how much more fringe ideologies receive. Another explanation, however, may be geographic bias.





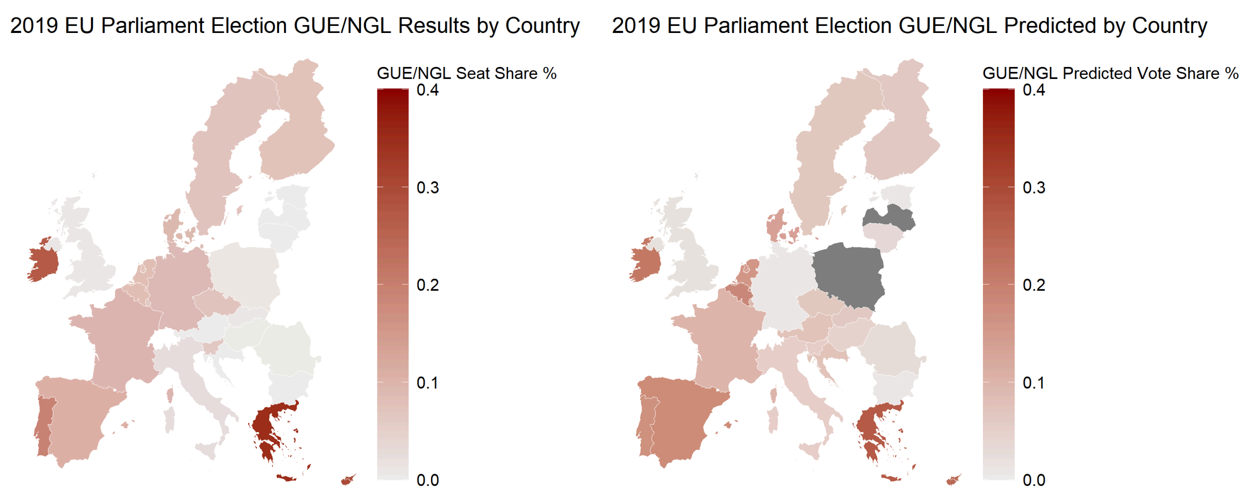
While it is difficult to tell from a lack of data points, the model does appear to overestimate the low countries region of Europe. It is difficult to come to any conclusion from this as the region only contains three countries, but further studies could potentially see if the model does have a bias towards overestimation in that region. Aside from potential biases, the other thing to note in the map is that Poland and Latvia are not colored. This is due to the fact that both were given negative predicted vote shares. While this is an issue for displaying purposes, both countries actually had a low percentage of votes go to GUE/NGL, so this is not actually a large issue with the model.

The good thing about this model is that it remains relatively consistent when applied to the 2019 data. On the 2019 data, the model has an average error of 7.65%. Just like with every other model, this is an increase on the 2014 data, but luckily for this model it is not a very large increase in error. There is also an increase in bias, with the model underestimating the actual vote share by an average of 2.81%. This is something to take note of, but it is not an overwhelmingly large bias. This bias can be seen in the following bar plot.



There are no errors that are dramatically larger than all of the other errors. Instead, several errors are relatively large, with the vote share in Romania, Greece, and Cyprus being overestimated, while the vote share in Malta is overestimated. Malta is notable here, as the model for S&D severely underestimated the vote share of Malta. Given that the two models were created independently, and that both of the models apply to parties on the left of the political spectrum, it could be surmised that certain aspects of Malta cause many votes that would otherwise go to GUE/NGL to go to S&D. Furthermore, given that Malta for the most part has a two-party system, it is possible that the party on the left, which is part of S&D, also receives votes from people who would otherwise vote for a GUE/NGL party.

Just like for when the model is applied to the 2014 data, the 2019 data appears to show a bias towards overestimating the vote share in the low countries.



Here the bias is more clear, as most countries actually have their vote shares underestimated. This is further evidence that there is something systematic about the low countries region that has to do with the variables, which in turn causes this bias to exist.

In order to determine that, the variables must be understood. This is the only model to involve only a singular category of variable. It is also the model that is closest to having every variable be statistically significant, with the only one that is not significantly barely missing the mark at a 95% confidence level as it has a p-value of 0.052. As was explained in the Greens/EFA model, the education variables are difficult to explain individually, as they all play off of one another. However, the variables can be analyzed in groups depending on the education level.

First off there are the variables dealing with the percentage of people in a certain age group that have an education level from 3 to 4. At a base level, these variables will tend to be positive overall, as the coefficient for the age range of 25 to 64 is very large, at 7.3121. This will of course be decreased depending on the other age ranges, but will tend to stay positive overall. It could become negative if there is an overwhelming number of people with education level from 3 to 4 in any of the other age ranges included in this model.

This aspect of the model seems to make sense for the same reason why it made sense in the S&D model - a higher percentage of people who achieved no higher than high school or trade school level education will result in a higher percentage of people who are working class. Given the policies espoused by left-wing parties, it makes sense that working class voters would vote for parties that promote economic policies that would benefit them.

Where the model becomes more difficult to explain is in its relation to education levels 0 to 2. This category overall has a negative correlation, implying that more people with these lower education levels will decrease the vote share of GUE/NGL. This is due to the fact that the coefficient for the age range of 25 to 64 is -2.22, while the coefficient for the age range of 45 to 64 is 1.32. In most cases, since the latter age range is encapsulated by the former, this will result in an overall negative correlation. This relationship seems odd, as one would expect an increase in lower education voters to correlate with an increase in working class voters, and thus also an increase in vote share for GUE/NGL parties. It is possible that countries which have this many uneducated people cause other people to vote differently. Perhaps by having so many voters who did not graduate high school, other issues could arise that cause different voters to want to vote for non-left-wing parties.

One thing to note with the variables is that this model shares with the S&D model a reliance on educational variables. Both models too seem to have increasing vote shares when fewer people have a university education. As such, it is difficult to determine what exactly makes these models fundamentally different other than the different coefficients. The defining difference is the inclusion of right-leaning political views as a variable in the S&D model. While it may seem paradoxical, voters for parties in GUE/NGL often are more likely to have certain right-wing views than voters of parties in S&D.

This is not true for all parties or all political views. Parties in GUE/NGL tend to be to the left of S&D on economic issues. However, the fundamental difference that can cause there to be more right-wing views in GUE/NGL parties is views on globalism, free trade, and the European Union. Although left-wing and right-wing parties may be more skeptical on these issues for different reasons, ultimately it is this skepticism that can cause right-leaning and left-leaning voters to cross over and vote for a party of the opposite end of the political spectrum. As such, its inclusion in the S&D model is one of the defining differentiating factors between S&D vote share and GUE/NGL vote share.

It is difficult to see how exactly this model could be used by left-wing parties to help boost their vote. In theory, they would need to decrease the amount of people who go into higher education. While they could decrease funding to higher education, this would be somewhat odd given their general policies of government spending. Instead, similarly to S&D, advocating for wage increases for the working class could disincentivize some people from going to university.

By making it easier to be working class, GUE/NGL parties would simultaneously be promoting the policies that they would be supporting anyway whilst also ensuring that their parties continue to exist in the future.

Likewise, parties that oppose GUE/NGL could advocate against helping the working class in order to incentivize going to university. However, this would be likely to backfire, as this would cause the current members of the working class to vote for left-wing parties in droves. Instead, funding could be increased to universities, and the cost of universities could be reduced, thus lowering the barriers for going to university. By increasing the ease in which people can go into higher education, there may end up being a decrease in the number of people with a high school education or less. As a result, parties that benefit from people who went to higher education, such as the Greens/EFA, would benefit.

While the S&D model was stronger, the use of education seems to have made a strong model for GUE/NGL as well. However, due to the small size of this party group, it remains to be seen if the model would hold well if a surge in these parties occurred. Perhaps these models indicate that a surge cannot occur, but that change can happen as gradually as education levels change. Regardless, the best way to evaluate this model will be once the next election occurs.

Discussion

In order to look at all of the models as a whole, the categories of variables that were expected to be used should be looked at individually to see whether they had an actual effect on the models.

The most used category was education. This category of variables was used in all six models. As a whole, numerous variables within this category were used. This may be due to the fact that this category had the most variables, or simply because these variables produced the most predictive results. These variables were most useful in the parties on the left of the political spectrum, where they made up the bulk of the variables. However, education variables were still very useful for the ALDE/RE and the right-leaning parties as well. This is likely due to the fact that education served simultaneously as a measure on the relative levels of knowledge between countries as well as the relative class structures in different countries.

By contrast, the least used category was economic variables. No model used any economic variables in their model. However, this does not necessarily mean that economics has no effect on the vote shares. Rather, it means that the economic variable used, GDP per capita, is not as strong a predictor as other variables were. Perhaps, the education variables did the work of the economic variables by determining the relative affluence of a country by the percentages of people by different age groups and education level.

The education variables may have also covered most of the work of age variables, though not entirely. Age variables were used in two models, the ones for EPP and ALDE/RE. It may not be a coincidence that these are the two parties for which the age variable was used. In both models, the age group of 25-49 was used. However, EPP had a positive relationship with this variable, while ALDE/RE had a negative relationship. This would suggest that age may have acted as one of the separating factors between EPP and ALDE/RE. Perhaps a country with more people in this age group would be more likely to have centrist or center-right votes go to EPP, while a country with fewer people in this age group would have more votes go to ALDE/RE.

There were still more variables included only in the ALDE/REmodel that would differentiate it from EPP and other parties. One such category of variables was the euroscepticism variables. This is somewhat surprising, as it would have been expected for these variables to at least be used for the right-wing party group as well. However, its use in ALDE/RE was still expected. Including these variables helps to differentiate ALDE/RE as the party of europhiles, one of the distinctive features of this party. Similarly, this was the only model to include the variable dealing with satisfaction with life. This would also make sense, as theoretically the countries that have the most people satisfied with the way things are going would vote for the parties that are more centrist and promise more of the status quo rather than radical change.

Another category of variables that was used only in one model was environmental concern variables. This kind of variable was only used in the Greens/EFA model. However, as noted in the analysis of the model, this may not have been ideal. While this variable could in theory be used as a good model, the increasing salience of climate change as an issue has caused this variable to vary far more in recent years. Perhaps if this variable were to stabilize, it would be stronger in the future. Regardless, as would have been expected, this was one of the key variables used in differentiating the Greens/EFA model from other models.

The variable of political views was also used only once as a way to differentiate models. The model for S&D was the only one to include this as a variable, that being the percentage of people with right-leaning political views. This was used as one of the main differentiating features between the model for S&D and GUE/NGL. It is not too surprising that these variables were not used in any other models, as the other models all had other factors to differentiate them from one another.

The primary differentiating factor for a lot of the models was the immigration category. Immigration variables were used for all of the models other than S&D and GUE/NGL. Generally, immigration was used to put cleavages between ALDE/RE, EPP, and the right-wing parties. Higher immigration tended to hurt ALDE/RE and EPP due to the fact that much of their vote share was likely being lost to the right-wing parties. Immigration was used in Greens/EFA in a different way due to it focusing on EU-28 immigration. Nonetheless, this was still a useful variable for that model as well.

Overall, the category of variables that tended to produce the strongest models was clearly education. First off, as noted, it was used in every model. Were every other variable removed from the models, there would still remain six models to analyze, albeit likely less strong models. Secondly, the strongest models were the ones that relied more on education than any other variables. All of the strongest models were from parties on the political left, which used far more education variables. GUE/NGL had a model consisting exclusively of education variables, and S&D was mostly using education variables. Greens/EFA had other variables included, though this is probably why the model produced a larger bias on the 2019 data.

One of the reasons for education being one of the strongest predictors was the fact that it was one of the most stable variables used in the dataset that still had predictive power. Educational attainment is generally slow to change, and as such these variables can be relied upon further to not have any dramatic changes. Age and GDP per capita are also unlikely to change very quickly, but were less predictive variables. Immigration and the opinion-based variables were more predictive, but also far more likely to undergo quick changes that can alter their ability to predict.

As a result, it can be concluded that stable variables, if the correct ones are found, seem to be able to have a far stronger predictive quality than unstable ones. While there were strong models for some of the party groups not on the left, their models may not function as well in future elections in 2024 and 2029. Meanwhile, the parties on the left which rely more on educational variables for their models may have models that would be strong in predicting these future elections. As a result, these models are also likely better for manipulation of data over longer periods of time.

Were a similar study to be conducted again, the models should be attempted again using exclusively factors that are relatively stable. Better economic variables could be found, and immigration variables could be found relating to the number of foreign-born residents rather than the immigration that occurs from year to year. This way, not only will the models likely be better at predicting, but also the predictions will be able to be made from further away from the election. Ideally if these changes were made, then far stronger models could be made for the six party groups.

Bibliography

'. “Eiropas Parlamenta Vēlēšanas 2014.” *CVK*, 30 May 2014, http://www.ep2014.cvk.lv/index_rez.html.

“2019 European Election Results .” *Europarl*, https://www.europarl.europa.eu/election-results-2019/en/european-results/2019-2024/.

“ALDE Member Parties.” *ALDE Party*, https://www.aldeparty.eu/alde_member_parties.

“Alliance of European Republican Movements.” *Alliance of European Republican Movements*, https://www.aerm.org/.

“Autoritatea Electorală Permanentă.” *Alegeri*, 2014, http://alegeri.roaep.ro/?alegeri=alegeri-parlamentul-european-2014.

“Az Európai Parlamenti Választás Összesített Eredménye.” *Valasztas*, 25 May 2014, https://static.valasztas.hu/dyn/ep14/szavossz/hu/eredm.html.

“Balsavimo Rezultatai Lietuvoje.” *VRK*, 1 June 2014, https://www.vrk.lt/statiniai/puslapiai/2014_ep_rinkimai/output_lt/rezultatai_daugiamand_apygardose/rezultatai_daugiamand_apygardose1turas.html.

“Balsavimo Rezultatai Lietuvoje.” *VRK*, 3 June 2019, https://www.vrk.lt/2019-europos-parlamento/rezultatai?srcUrl=%2Frinkimai%2F904%2F2%2F1548%2Frezultatai%2Flt%2FrezultataiEpDaugmVrt.html.

Bandau, Frank. “The Electoral Crisis of Social Democracy: Postindustrial Dilemmas or Neoliberal Contamination?” *Political Studies Review*, vol. 20, no. 3, 2021, pp. 493–503., https://doi.org/10.1177/14789299211032461.

“Biroul Electoral Central.” *Europarlamentare*, 26 May 2019, http://europarlamentare2019.bec.ro/wp-content/uploads/2019/04/DA_PPR_CRETU_CORINA.pdf.

Bundeswahlleiter, Der. “Europawahl 2014.” *Ergebnisse Deutschland - Der Bundeswahlleiter*, Der Bundeswahlleiter, 2014, https://www.bundeswahlleiter.de/europawahlen/2014/ergebnisse/bund-99.html.

Edo, Anthony, et al. “Immigration and Electoral Support for the Far-Left and the Far-Right.” *SSRN Electronic Journal*, 2018, https://doi.org/10.2139/ssrn.3167728.

"Eesti Vabariik." *Euroopa Parlamendi Valimised*, 5 June 2014, http://ep2014.vvk.ee/voting-results-en.html.

"EIROPAS PARLAMENTA VĒLĒŠANAS 2019." *CVK*, https://epv2019.cvk.lv/pub/velesanu-rezultati.

"Elecciones Europeas 25 De Mayo De 2014." *HistoriaElectoral*, 25 May 2014, https://www.historiaelectoral.com/eu2014.html.

"Eleições Para o Parlamento Europeu 2014." *Comissão Nacional De Eleições*, 25 May 2014, https://www.cne.pt/content/eleicoes-para-o-parlamento-europeu-2014.

"Eleições Para o Parlamento Europeu 2019." *Comissão Nacional De Eleições*, 2019, https://www.cne.pt/content/eleicoes-para-o-parlamento-europeu-2019.

"Eligendo Archivio - Ministero Dell'interno Dait." *Eligendo*, 26 May 2019, https://elezionistorico.interno.gov.it/index.php?tpel=E&dtel=26/05/2019.

"Euroflections: Leading Academics on the European Elections 2019." Edited by Niklas Bolin et al., *EuroFlections*, Mittuniversitetet, https://www.researchgate.net/publication/333967891_Euroflections_Leading_academics_on_the_European_elections_2019.

"European Alliance for Freedom and Democracy." *EAFD Party*, https://eafdparty.eu/.

"European Elections 2014." *European Elections 2014: Official Results*, http://results.elections.moi.gov.cy/English/EUROPEAN_ELECTIONS_2014/Islandwide.

"European Elections 2014." *Tulospalvelu*, Ministry of Justice, 2 Nov. 2021, https://tulospalvelu.vaalit.fi/EPV-2014/en/tulos_kokomaa.html.

"European Elections 2019." *European Elections 2019: Official Results*, http://results.elections.moi.gov.cy/English/EUROPEAN_ELECTIONS_2019/Islandwide.

"European Elections 2019." *Tulospalvelu*, Ministry of Justice, 22 Sept. 2020, https://tulospalvelu.vaalit.fi/EPV-2019/en/tulos_kokomaa.html.

"European Parliament Election 2019." *The Federal Returning Officer*, The Federal Returning Officer, 2019, https://bundeswahlleiter.de/en/europawahlen/2019/ergebnisse/bund-99.html.

European Parliament Election Results 2014. Environment, Community, and Local Government, 2014, https://www.europarl.europa.eu/ireland/resource/static/files/Elections%202014/results-of-2014-elections-dept.-envi-brochure.pdf.

"European Parliament Election – May 2019." *European Parliament Election – May 2019*, Ministry of Interior, May 2019, https://dev-ekloges.singularlogic.eu/current/e/home/en/parties/.

"EUROPEAN PARLIAMENT ELECTIONS 2014." *Volitve*, Republika Slovenija, 2014, http://volitve.gov.si/ep2014/en/index.html.

"Europees Parlement 22 Mei 2014." *KiesRaad*, 22 May 2014, https://www.verkiezingsuitslagen.nl/verkiezingen/detail/EP20140522.

"Europees Parlement 23 Mei 2019." *KiesRaad*, 23 May 2019, https://www.verkiezingsuitslagen.nl/verkiezingen/detail/EP20190523.

"Európai Parlamenti Képviselők Választása 2019 - Összesített Eredmények." *Nemzeti Választási Iroda*, 21 July 2019, https://www.valasztas.hu/osszesitett-eredmenyek_ep2019.

"Final Results." *Elections to the European Parliament 2019*, Volby, 2019, https://volby.statistics.sk/ep/ep2019/en/data02.html.

"Frekwencja Wyborcza." *Pkw*, 2014, https://pe2014.pkw.gov.pl/pl/.

"Home." *Europe Elects*, Europe Elects, 2 Nov. 2022, https://europeelects.eu/.

Hudgens, Austin. "The Quiet Rise of the Far-Left." *Claremont-UC Undergraduate Research Conference on the European Union*, vol. 18, no. 1, 2018, pp. 59–68., https://doi.org/10.5642/urceu.201801.07.

"Immigration by Broad Group of Country of Previous Residence." *European Commission*, 30 Mar. 2022, https://ec.europa.eu/eurostat/databrowser/view/MIGR_IMM12PRV__custom_3387909/default/table?lang=en.

Izbori Članova U Europski Parlament Iz Republike Hrvatske. REPUBLIKA HRVATSKA DRŽAVNO IZBORNO POVJERENSTVO REPUBLIKE HRVATSKE, 9 June 2014, https://www.izbori.hr/arhiva-izbora/data/euparlament/2014/izabrani/i_14_000_0000.pdf.

"Izidi Glasovanja Za Celotno Slovenijo." *Državna Volilna Komisija*, Republika Slovenija, 5 June 2019, https://volitve.gov.si/ep2019/#/rezultati.

"Live Results from the 2019 Local Elections and European Elections." *RTE.ie*, 2019, https://www.rte.ie/news/elections-2019/results/#/european/.

"Member Parties." *EFA - European Free Alliance*, https://e-f-a.org/member-parties/.

"Member Parties." *European Greens*, https://europeangreens.eu/map.

"Members." *The Party of European Socialists*, 17 Oct. 2022, https://pes.eu/members/.

"MEP ELECTION." *Electoral Commission of Malta*, 2019, https://electoral.gov.mt/ElectionResults/MEP.

"Number and Share of Valid Votes for Political Parties for the SR." *Elections to the European Parliament 2014*, Volby, 2014, https://volby.statistics.sk/ep/ep2014/EP-dv/Tabulka3_en.html.

"Our Members and Associates." *ECPM*, https://www.ecpm.info/members-and-associates.html.

"PARLAMENTO EUROPEO." *Consulta De Resultados Electorales*, May 2019, https://web.archive.org/web/20200912203324/http://www.infoelectoral.mir.es/min/busquedaAvanzadaAction.html?vuelta=1&codTipoEleccion=7&codPeriodo=201905&codEstado=99&codComunidad=0&codProvincia=0&codMunicipio=0&codDistrito=0&codSeccion=0&codMesa=0.

"Participating Parties." *Initiative of Communist & Workers' Parties* , https://www.initiative-cwpe.org/en/participating-parties/.

"Parties and Partners." *EPP*, https://www.epp.eu/parties-and-partners.

Pearson, Mitya, and Wolfgang Rüdig. "The Greens in the 2019 European Elections." *Environmental Politics*, vol. 29, no. 2, 2020, pp. 336–343., https://doi.org/10.1080/09644016.2019.1709252.

"Population by Age Group." *Eurostat*, 22 June 2022, https://ec.europa.eu/eurostat/databrowser/view/tps00010/default/table?lang=en.

"Population by Educational Attainment Level, Sex and Age (%) - Main Indicators." *European Commission*, 29 Sept. 2022, https://ec.europa.eu/eurostat/databrowser/view/EDAT_LFSE_03__custom_3387832/default/table?lang=en.

"Proclamation Des Résultats De L'élection Des Représentants Au Parlement Européen." *Legifrance*, 30 May 2019, https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038528275.

"A Progressive Movement for Europe." *DiEM25*, https://diem25.org/en/.

"Recent Immigrants by Sex, Age and Country of Birth." *European Commission*, 27 Sept. 2022, https://ec.europa.eu/eurostat/databrowser/view/LFST_RIMGPCGA__custom_3387949/default/table?lang=en.

"Resultater - Hele Landet." *EUROPA-PARLAMENTSVALG SØNDAG*, 25 May 2014, https://dst.dk/valg/Valg1475795/valgopg/valgopgHL.htm.

"Resultater - Hele Landet." *EUROPA-PARLAMENTSVALG SØNDAG*, 26 May 2019, https://dst.dk/valg/Valg1684426/valgopg/valgopgHL.htm.

"Results in Figures." *Elections 2019 Belgium*, Ibz, 4 June 2019, https://elections2019.belgium.be/en/results-figures?el=EU&id=EUR00000.

REZULTATE IZBORA ČLANOVA U EUROPSKI PARLAMENT IZ REPUBLIKE HRVATSKE. REPUBLIKA HRVATSKA DRŽAVNO IZBORNO POVJERENSTVO REPUBLIKE HRVATSKE, 27 May 2019, https://www.izbori.hr/site/UserDocsImages/2019/Izbori_clanova_u_EU_parlament_iz_RH/Rezultati/2019-06-05_14_30_izvjesce.pdf.

Rooduijn, Matthijs, et al. "Radical Distinction: Support for Radical Left and Radical Right Parties in Europe." *European Union Politics*, vol. 18, no. 4, 2017, pp. 536–559., https://doi.org/10.1177/1465116517718091.

"Résultats Des Élections Européennes 2014." *Interieur.Gouv*, 25 May 2014, https://www.interieur.gouv.fr/Elections/Les-resultats/Europeennes/elecresult__ER2014/(path)/ER2014//FE.html.

"Standard Eurobarometer 81 - Spring 2014." *Eurobarometer*, European Commission, July 2014, https://europa.eu/eurobarometer/surveys/detail/2040.

"Standard Eurobarometer 91 - Spring 2019." *Eurobarometer*, European Commission, Aug. 2019, https://europa.eu/eurobarometer/surveys/detail/2253.

"Stimmenstärkste Partei pro Bundesland." *Bundesministerium Fur Inneres*, https://web.archive.org/web/20160702062927/http://euwahl2014.bmi.gv.at/.

"The UK's European Elections 2019." *BBC News*, BBC, 2019, https://www.bbc.com/news/topics/crjeqkdevwvt/the-uks-european-elections-2019.

"Val till Europaparlamentet - Röster." *Valmyndigheten*, 30 May 2014, https://historik.val.se/val/ep2014/slutresultat/E/rike/index.html.

"Val till Europaparlamentet - Röster." *Valmyndigheten*, 31 May 2019, https://historik.val.se/val/ep2019/slutresultat/E/rike/index.html.

Vera, Prima. "Primaveraeuropea.eu." *Primaveraeuropea.eu*, 10 May 2018, https://primaveraeuropea.eu/.

"Volby Do Evropského Parlamentu Konané Na Území České Republiky Ve Dnech." *Volby.cz*, 24 May 2014, https://www.volby.cz/pls/ep2014/ep11?xjazyk=CZ.

"Volby Do Evropského Parlamentu Konané Na Území České Republiky Ve Dnech." *Volby.cz*, 25 May 2019, https://www.volby.cz/pls/ep2019/ep11?xjazyk=CZ.

"Vote 2014." *BBC News*, BBC, 2014, https://www.bbc.co.uk/news/events/vote2014/eu-uk-results.

"Voting and Election Result." *Valimised*, 29 May 2019, https://ep2019.valimised.ee/en/election-result/index.html.

Wagner, Sarah. "Euroscepticism as a Radical Left Party Strategy for Success." *Party Politics*, vol. 28, no. 6, 10 Aug. 2021, pp. 1069–1080., https://doi.org/10.1177/13540688211038917.

"Who We Are? // ECR Group." *ECR Group*, https://ecrgroup.eu/ecr.

"World Economic Outlook Database, April 2022." *IMF*, https://www.imf.org/en/Publications/WEO/weo-database/2022/April.

Wybory Do Parlamentu Europejskiego 2019, 2019, https://pe2019.pkw.gov.pl/pe2019/pl/wyniki/pl.

"Élections Européennes 2014." *France-Politique*, 25 May 2014, https://www.france-politique.fr/elections-europeennes-2014.htm.

"Élections Européennes 2019 : Résultats Officieux." *Elections.public.lu*, Le Gouvernement Du Grande-Duche De Luxembourg, 2019, https://elections.public.lu/fr/elections-europeennes/2019/resultats.html.

ΝΕΑ, ΙΚΑΡΙΩΤΙΚΑ. "Ευρωεκλογές 2014: Η Πρώτη Νίκη Του Τσίπρα." *ΙΚΑΡΙΩΤΙΚΑ ΝΕΑ*, 25 May 2015, https://ikarianews.gr/archives/47320.

"Избори Заевропейски Парламент." *ЦЕНТРАЛНА ИЗБИРАТЕЛНА КОМИСИЯ*, 26 May 2019, https://results.cik.bg/ep2019/rezultati/index.html.

"Резултати От Избори За Европейски Парламент." *Избори За Европейски Парламент 2014*, 25 May 2014, https://results.cik.bg/ep2014/rezultati/.