Populism On The Rise: Micro Analysis of Italian Election Results (2018 - 2019)

Chiara Marino[1] and Alessandro Valfre'[2]

[1]Department of Economics, Sant'Anna School of Advanced Studies, Pisa
[2]Department of Economics, Sant'Anna School of Advanced Studies, Pisa

May 17, 2024

Abstract

This project contributes to the existing literature on the study of election results, especially regarding the impact of socio-economic and demographic features of Italian municipalities on the parties' vote shares. The analysis is centred on two specific Italian elections, the general one of 2018 and the European one of 2019, allowing for a deep understanding of the rise of populist parties in Italy. To answer the research question we have employed several unsupervised and supervised techniques to finally estimate Ordinary Least Squares Regression models. Overall, the results of such comprehensive analysis show an opposite pattern for the two analysed parties: Lega's vote share is positively related to income, middle education and employment rate, while M5S's vote share is negatively linked to the same. Moreover, Lega received a higher percentage of votes from less fragile municipalities, while M5S received it from the most fragile ones, especially in the southern regions.

Keywords: Elections, Italian Municipalities, PCA, Classification, populisms.

1 Introduction

This project started with investigating Italian election results from a truly micro perspective i.e. employing municipalities as units of analysis. Doing research at the municipality level stems mainly from the importance of examining localised dynamics, especially considering the Italian territory. Therefore, this project can provide a useful contribution to the study of elections, offering insights that might be overlooked in broader regional or national analysis (Levi & Patriarca, 2020). Moreover, by selecting the general election of 2018 and the European election of 2019 this project aims to investigate the rise of populist parties in Italy. Characterised by the triumph of populist parties, the above-mentioned elections are indeed crucial in identifying the features allowing for the rise of this new political wave. Indeed, clarifying populist traits and features is still quite necessary as the term populism has been applied to many political movements and, as Rooduijn and Pauwels (2011) put it, it lacks conceptual clarity. Thus, within this project populism is defined as a thin ideology consisting of two main components: people-centrism and anti-elitism (Mudde, 2004). The two analysed Italian populist parties seem indeed to possess such characteristics as they have focused on a confined range of issues i.e. showing to have a thin ideology, they put great emphasis on the centrality of the people and their conflict with the status quo. After having briefly presented the theoretical grounding from a political science perspective, we now focus on the statistical part, applying most of the tools and techniques learned throughout the course. The employed methods are subject to our research question i.e. finding the best predictors for the vote share of the two main populist parties, Lega and Movimento 5 Stelle (M5S). The following section gives an overview of the dataset employed for the analysis, describing the most relevant variables for our study. Section 3 is dedicated to unsupervised learning, while Sections 4 and 5 detail the supervised learning techniques employed and the OLS modelling.

2 Methodology

The dataset is composed of 7903 observations and 58 variables. Part of the pre-processing of variables included identifying and recognising unions and mergers of municipalities, summing up values or computing averages depending on the variables in question. The final number of municipalities is 7903 as in the fragility index ISTAT dataset. The procedures mentioned above were necessary since data were retrieved from several datasets and the number of observations differed from one source to another (ISTAT, Ministero dell'Interno). The following tables summarised the most relevant variables for our analysis (out of the total 58 features).

Name of the variable	**Description**
Fragility Index	Composite index designed to identify areas most susceptible to specific risk factors and to facilitate the territorial analysis in a historical series. It it formed by 12 elementary indicators that describe the main dimensions (territorial, environmental and socio-economic) depicting the fragility of municipal territories. Scores range from 1 (least fragile) to 10 (most fragile) for each municipality.
Income per capita	Average income in the municipality for working age individuals; depending on the analysis, it is reported in euros or thousands of euros
Employment rate	Employed people over the total working age population, expressed as percentage
Middle Education	Proportion of individuals aged 25-64 people with middle-school title relative to the total population in the same age range
Abstention Rate	Percentage of eligible voters who choose not to cast their vote in an election over the total electorate

The following table details the parties used for the analysis.

Name of the party	Description
Lega	Right-wing party, one of the two considerable as "populist party" and object of our analysis, present both in 2018 and 2019
M5S (Movimento 5 Stelle)	Left-wing party, one of the two considerable as "populist party", present both in 2018 and 2019
PD (Partito Democratico)	Left-wing party, taken into account as control. We expect the received percentage of votes not to be affected by peaks (like we expect from populist parties). This party is present both in 2018 and 2019
Forza Italia	Right-wing party, taken into account as control. We expect the received percentage of votes not to be affected by peaks (like we expect from populist parties). This party is present both in 2018 and 2019
Fratelli d'Italia	Right-wing party, taken into account as control. We expect the received percentage of votes not to be affected by peaks (like we expect from populist parties). This party is present both in 2018 and 2019. This party proves to be an interesting study subject in the future, considering the great peak of votes in the general election of 2022.
Others	Sum of all those parties that are present only in one of the two elections, for which there is not a clear correspondence between 2018 and 2019, (like "La Sinistra " in 2019 and "Liberi e Uguali" in 2018) and the regional parties (like Union Valdotaine in Valle d'Aosta)

For representation purposes, we also employed specific library packages on R (like "sf" and "geojson") to create infographics of the Italian territory, where we can associate the value of the variable to the municipality. This proves to be useful, especially in looking at geographical patterns in the victories of parties.

3 Unsupervised Learning

3.1 Principal Component Analysis

Our project started with a form of unsupervised dimension reduction technique: Principal Component Analysis (PCA). This was done to perform a preliminary exploration of our dataset to capture its main patterns and structure. Indeed, PCA allows us to simplify the large dataset, while maintaining significant patterns and trends, identify the most relevant variables and ultimately use them as inputs for supervised analysis. From the loadings plots for both elections, the following five variables exhibit the highest contributions: employment rate, income per capita, fragility index, middle education and abstention rate.

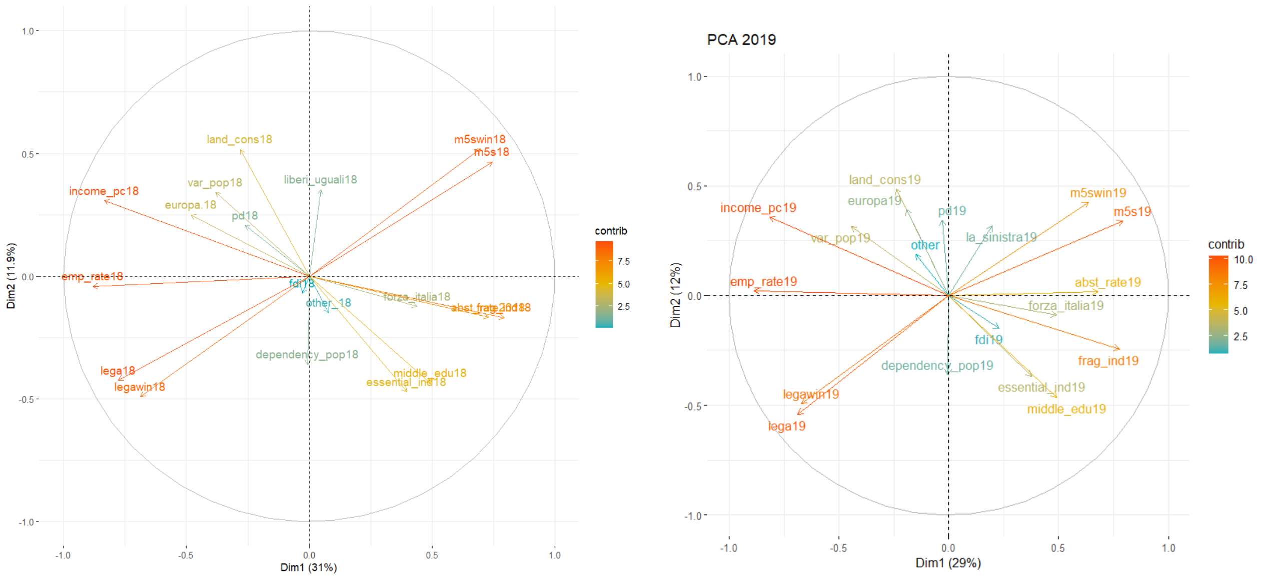

Fig. 1: PCA Loadings Plot 2018 (left) and 2019 (right)

However, looking more specifically at the biplot, it is difficult to capture a clear pattern for the data due to such a large number of observations. For this reason, we decided to first, create various subsamples of

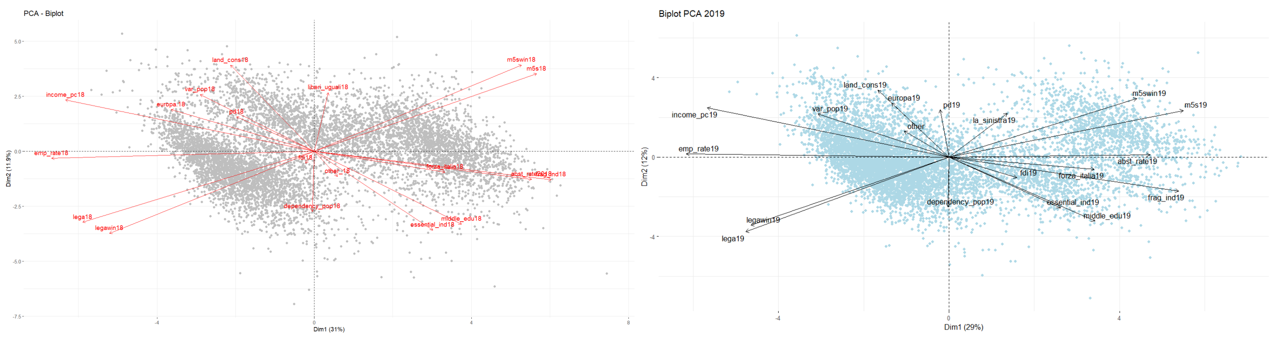

Fig. 2: PCA Biplot 2018 (left) and 2019 (right)

our observations (i.e. municipalities) and then apply PCA. Figure 3 below shows one particular result of subsampling for about 150 observations, coloured according to the level of fragility. From left to right, we observe an increase in the fragility level, suggesting that the pattern might be indeed geographical. To examine more closely whether the above-mentioned pattern is indeed geographical we moved back from subsampling to the whole dataset. By aggregating the municipalities by regions and provinces, we took the average position of PCA (1st and 2nd principal component position). Figure 4 below shows the results at the regional level.

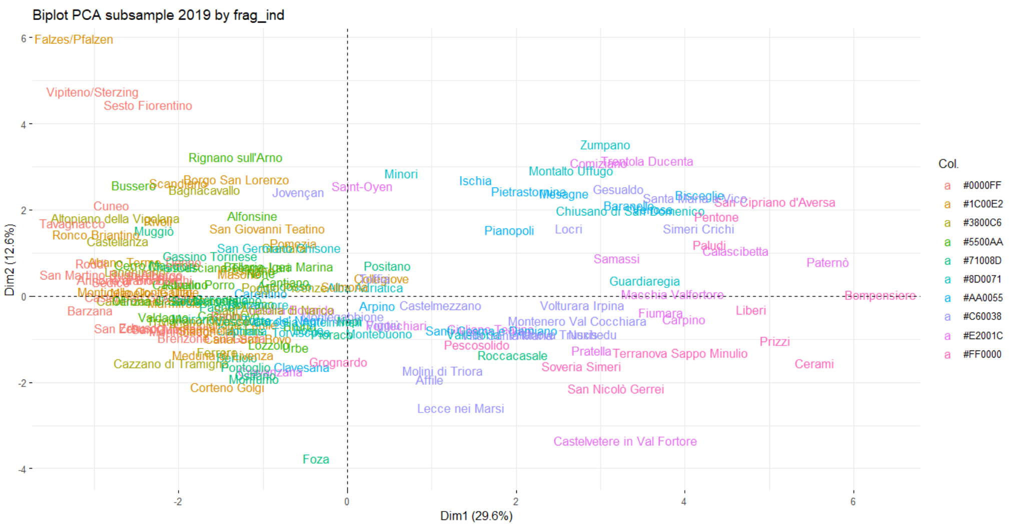

Fig. 3: PCA subsampling by fragility index

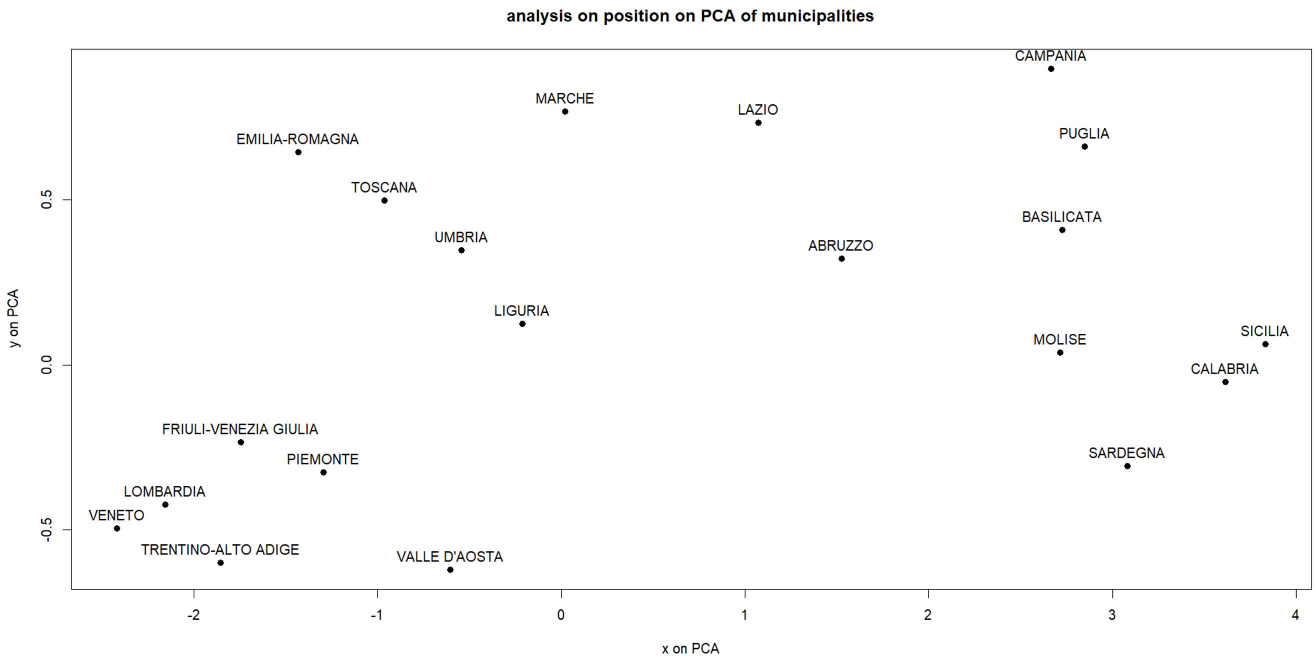

Fig. 4: Representation of the position of PCA depending on the region

From this picture, the emergence of three types of groups is clear:

- On the bottom left part of the plot, there are the Northern regions (Valle d'Aosta, Trentino Alto-Adige, Lombardia, Veneto, Friuli, Piemonte), characterized by both negative average values of x and y
- On the top centre part of the plot, there are the Central regions (Emilia-Romagna, Toscana, Umbria, Liguria, Marche, Abruzzo e Lazio), characterized by positive values of y
- On the right side of the picture, there are the Southern regions (Campania, Puglia, Basilicata, Molise, Sardegna, Calabria and Sicilia), characterized by positive values of x

This suggests a significant impact of the municipality's geographical location on the position of PCA. Figure 5 below, instead shows PCA coordinates aggregating at the province level, thus at a meso-level of aggregation.

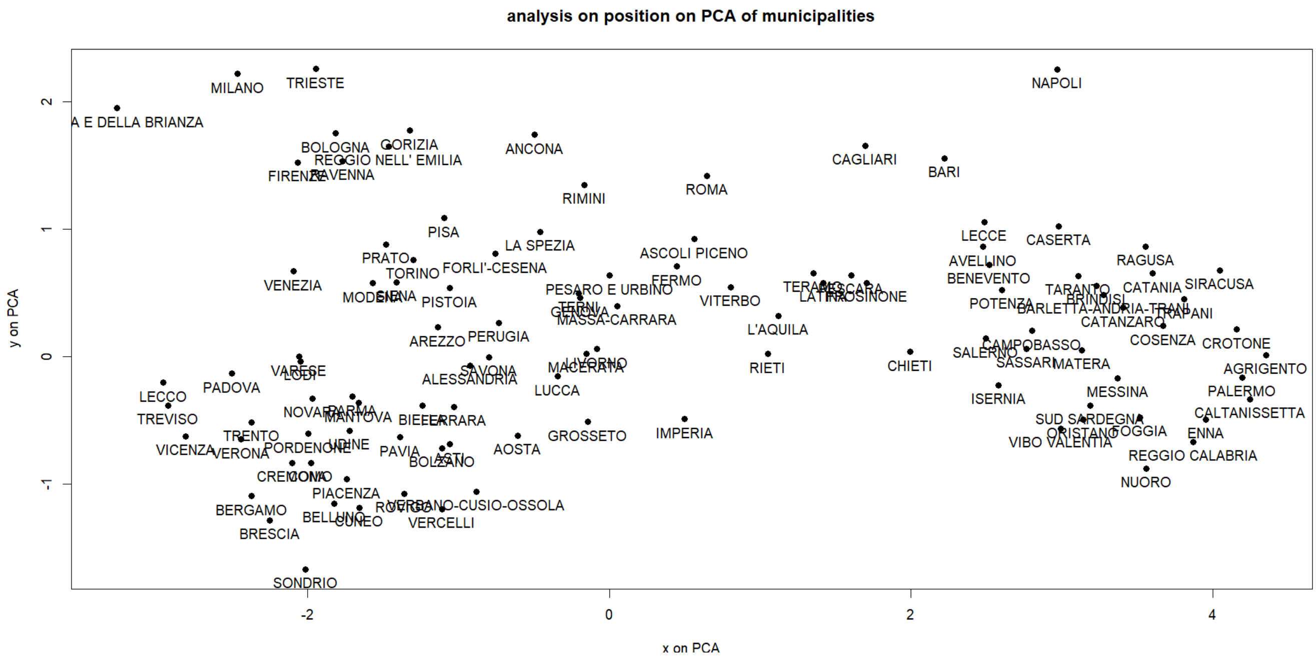

Fig. 5: Representation of the position of PCA depending on the province

The results observed at the regional level are generally confirmed: on the bottom left side of the plot, we see provinces from Northern regions, in the centre we see the Central regions, while on the right there are only provinces from Southern Italy. PCA analysis showed the relevance and necessity of data aggregation. While from a municipality level of analysis, no pattern emerged and the corresponding plot exhibited an indistinct cloud of data points, aggregation showed that a geographical pattern is indeed underway.

4 Supervised Learning

Building on the results of unsupervised learning, from data exploration and dimension reduction, the analysis progresses towards supervised variable selection methods. This is done to ultimately construct the appropriate OLS model to predict the vote shares of the main winning parties.

4.1 Ridge and Lasso

Ridge and Lasso regressions are part of the so-called "Shrinkage methods" which allow for the regularisation of coefficient estimates, providing guidance in the selection of best predictors for the construction of OLS models (Gareth et al., 2013). This is done through the introduction of a penalty term. In particular, the Ridge regression minimises the quantity below:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = RSS + \lambda\sum_{j=1}^{p}\beta_j^2$$

However, with this type of penalty term the Ridge regression does not set any coefficient equal to zero, i.e. it never results in the exclusion of any variable. On the other hand, the Lasso regression, setting the norm of betas as the penalty term, practically forces some of the coefficients to zero when λ is sufficiently large. (Gareth et al, 2013). It is important to recall that Lasso regression minimises the following quantity:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j|$$

Consequently, for practical purposes (i.e. variables selection) it would be more helpful for the discussion to study the results of Lasso. As shown by Figures 6 and 7 below, the coefficients of the employment rate, fragility index and middle education are the most significant and persistent in predicting vote shares for the two populist parties. In particular, the coefficient for employment rate is quite persistent for both parties, positive for Lega and negative for M5S. This is in line with the expectations as M5S campaigned for unemployment

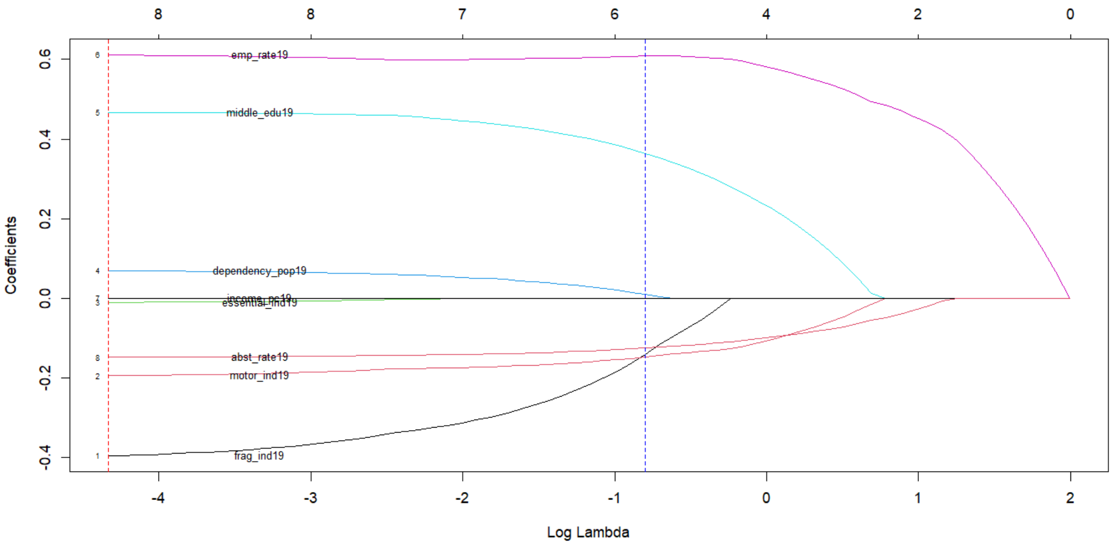

Fig. 6: Lasso Results for LEGA 2019

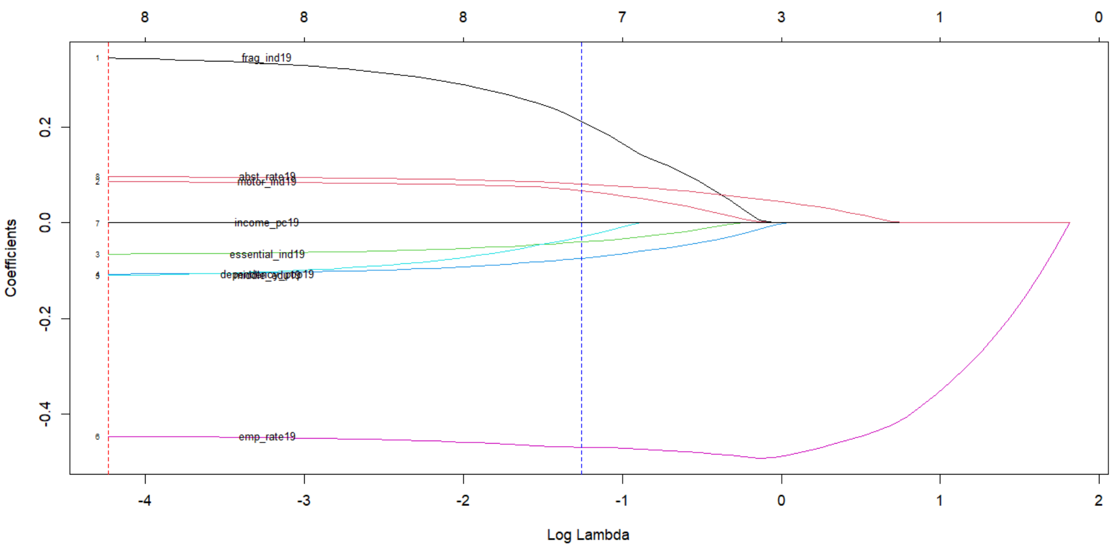

Fig. 7: Lasso Results for M5S 2019

benefits policies, and also when in government it was the promoter of the so-called "Reddito di Cittadinanza". Secondly, the fragility index coefficient seems quite persistent for both parties, positive for M5S and negative for Lega. This means that for increasing levels of fragility of municipalities, there is support for M5S. The opposite is true for Lega i.e. decreasing levels of fragility correspond to higher vote shares for Lega. Lastly, the coefficient of middle education is positive, significant and quite persistent for Lega, denoting that the presence of large share of middle-educated people is a good predictor for Lega vote shares.

4.2 Feature Selection

This subsection focuses on the application of feature selection. This is done to control for the selection of features obtained through Ridge and Lasso regressions. The results are in line with the Ridge and Lasso, confirming the list of features employed for classification and OLS modelling.

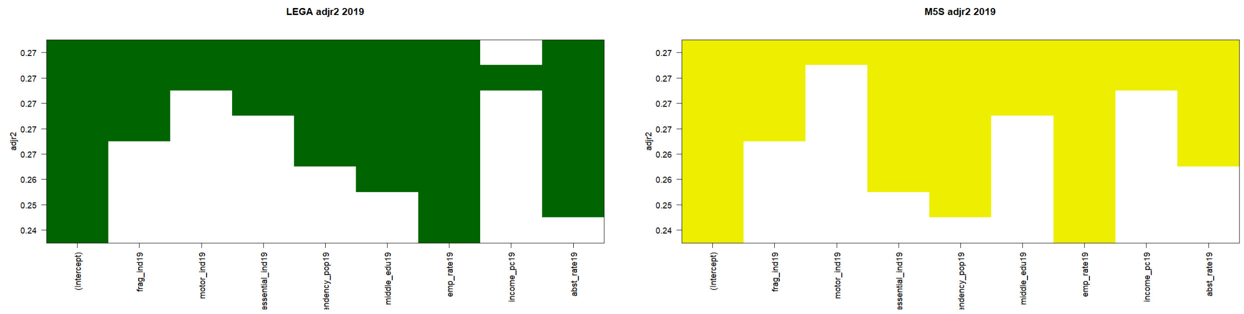

Fig. 8: Feature selection results

4.3 Classification

Supervised classification is an attempt to build prediction using a categorical label. This instrument aims to use the data available, train a model that will be useful for prediction, and allow for the prediction of a new unit label based on the values in its feature vector. For our purposes, we created the training set and test set, with a 75-25 proportion, so to train the municipalities on the results of the elections on the training set and to evaluate the quality of the same on the test set. Classification is very sensitive to changes in the dependent variables, so maintaining the percentage of a party in the municipalities was not meaningful. Each decimal deviation from the true value led to the emergence of an error in the test, so the accuracy was 0. For this reason, we created two dummy variables, called "legawin" and "m5swin", to assess whether the party wins in the municipality (value of the dummy equal to 1) or not (value of the dummy equal to 0). We then computed the probability of winning in the municipality and we tested the model. The below model uses the dummy for the victory as the dependent variable, and income per capita, middle education, employment rate and abstention rate as independent variables. i.e. it is a logistic model. Figure 9 below presents the confusion matrices for the logistic model.

$$legawin = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i + \beta_4 * AbstRate_i$$
$$m5swin = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i + \beta_4 * AbstRate_i$$

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 890 274
         1 151 660

               Accuracy : 0.7848
                 95% CI : (0.766, 0.8028)
    No Information Rate : 0.5271
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5654

 Mcnemar's Test P-Value : 3.261e-09

            Sensitivity : 0.8549
            Specificity : 0.7066
         Pos Pred Value : 0.7646
         Neg Pred Value : 0.8138
             Prevalence : 0.5271
         Detection Rate : 0.4506
   Detection Prevalence : 0.5894
      Balanced Accuracy : 0.7808

       'Positive' Class : 0
```

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 947 223
         1 256 549

               Accuracy : 0.7575
                 95% CI : (0.7379, 0.7762)
    No Information Rate : 0.6091
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.4946

 Mcnemar's Test P-Value : 0.1437

            Sensitivity : 0.7872
            Specificity : 0.7111
         Pos Pred Value : 0.8094
         Neg Pred Value : 0.6820
             Prevalence : 0.6091
         Detection Rate : 0.4795
   Detection Prevalence : 0.5924
      Balanced Accuracy : 0.7492

       'Positive' Class : 0
```

Fig. 9: Confusion matrices for Logistic regression, Lega (left) and M5S (right),2018

From the confusion matrices, we observe high accuracy of the test: near 75 % for Lega, meaning that we have the right prediction for 3 over 4 municipalities and 78 % for M5S. Moreover, the p-value points at a statistically significant relation and the confidence interval is not so large, with a margin of error at 95 % of significance of 2 percentage points. The results are similar for both parties in 2018, while for 2019 M5S appears to lose several percentage points with respect to 2018, thus the municipalities where M5S won were lower ;let's see what it happens for 2019:

The confusion matrices for 2019 changed a bit: Lega maintains the accuracy and the property described above (estimate accuracy rate at 77%), while Movimento 5 Stelle has a higher accuracy but, considering that it wins in a smaller number of municipalities, it is less able to predict the victories in the municipality.

Then, we also checked the results using the "partimat" to get a graphical representation of the area where we have boundaries that separate the two classes (win or lose in the municipality), with points associated with municipalities and whether they are correctly predicted. The next figures report the analysis from 2018 and 2019.

We can see that the number of results correctly predicted (the green signs) is higher than wrongly predicted (the red signs); the approximate error computed oscillates between 0.25 and 0.28, in line with the accuracy found in the previous confusion matrices (in the section "Appendix" we report partimat also for logistic model).

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0  154   81
         1  364 1376

               Accuracy : 0.7747
                 95% CI : (0.7556, 0.7929)
    No Information Rate : 0.7377
    P-Value [Acc > NIR] : 8.334e-05

                  Kappa : 0.2933

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.29730
            Specificity : 0.94441
         Pos Pred Value : 0.65532
         Neg Pred Value : 0.79080
             Prevalence : 0.26228
         Detection Rate : 0.07797
   Detection Prevalence : 0.11899
      Balanced Accuracy : 0.62085

       'Positive' Class : 0
```

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1697  243
         1   24   11

               Accuracy : 0.8648
                 95% CI : (0.8489, 0.8796)
    No Information Rate : 0.8714
    P-Value [Acc > NIR] : 0.8183

                  Kappa : 0.0464

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.98605
            Specificity : 0.04331
         Pos Pred Value : 0.87474
         Neg Pred Value : 0.31429
             Prevalence : 0.87139
         Detection Rate : 0.85924
   Detection Prevalence : 0.98228
      Balanced Accuracy : 0.51468

       'Positive' Class : 0
```

Fig. 10: Confusion matrices for Logistic regression, Lega (left) and M5S (right),2019

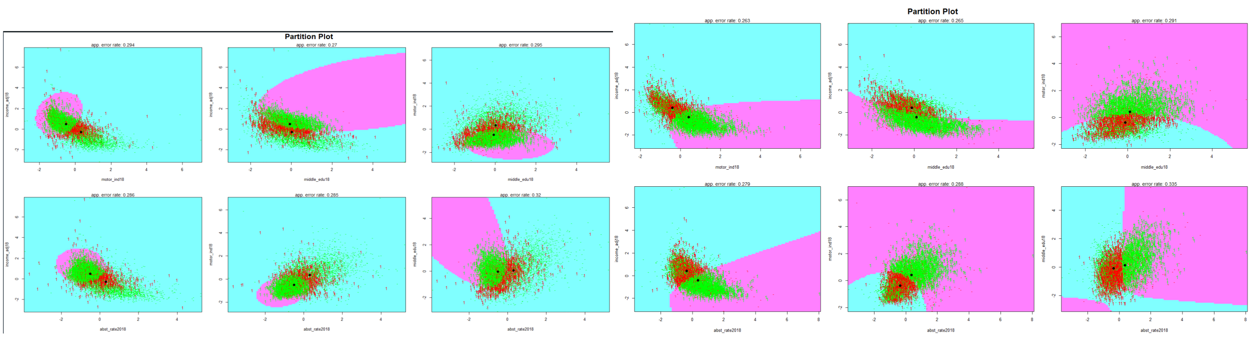

Fig. 11: Partimat of Lega (left) and M5S (right), with QDA, 2018

5 Ordinary Least Squared (OLS) Models

As a result of the supervised learning of the previous sections, we then used the best predictors of party vote share to build several Ordinary Least Squared models. Starting from a simple but reliable model, i.e. benchmark model, we constructed more complex and complete models.

5.1 Base Model

To run this benchmark model, which we have called base, we have used three variables as independent: income per capita, employment rate and percentage of middle education; in this analysis, the dependent variable is the percentage of the party, so or Lega or Movimento 5 Stelle; these are the results.

$$y_i = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i$$

Table 1: results for ols models parties, lega (left) and M5S (right), 2018

	Dependent variable:	
	lega18perc	m5s18perc
	(1)	(2)
income_adj18	0.489*** (0.040)	−0.186*** (0.038)
emp_rate18	0.782*** (0.014)	−0.735*** (0.014)
middle_edu18	0.385*** (0.014)	−0.137*** (0.013)
Constant	−54.454*** (1.082)	83.404*** (1.029)
Observations	7,903	7,903
R^2	0.479	0.450
Adjusted R^2	0.479	0.450
Residual Std. Error (df = 7899)	8.490	8.080
F Statistic (df = 3; 7899)	2,422.504***	2,158.237***
Note:		*p<0.1; **p<0.05; ***p<0.01

We can see how the two parties behave differently with the same model. On the one hand, the percentage of votes for Lega is affected in a positive and significant way by income per capita, employment rate and middle education, on the other hand, the same variables negatively affect the percentage of votes for Movimento 5 Stelle (M5S). This is coherent with what was previously observed: Lega won generally in Northern Italy, where the employment rate and income per capita are higher compared to southern regions.

5.2 Abstention Rate Model

In the so-called "Abstention Rate model" we have used the previous base model, adding the abstention rate, so socio-economic variables are associated to election one. Table 2 below shows the results for this model.

$$y_i = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i + \beta_4 * AbstRate_i$$

We maintained the results of the previous model, in terms of sign and significance of the results. In addition to that, the coefficient of the variable "abstention rate" for both parties is negative and significant. This is in line with the expectations that if people do not go to the polls, we can expect parties to receive fewer votes.

5.3 Fragility model

In this case, we added the fragility index to the previously estimated models. This was done to add a geographical and urbanistic feature to our model. The results for the 2018 general elections are shown in Table 3.

$$y_i = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i + \beta_4 * AbstRate_i + \beta_5 * FragIndex_i$$

Table 2: results for abst.rate models 2018

	Dependent variable:	
	lega18perc	m5s18perc
	(1)	(2)
income_adj18	0.334*** (0.039)	−0.201*** (0.038)
emp_rate18	0.649*** (0.015)	−0.748*** (0.015)
middle_edu18	0.413*** (0.013)	−0.135*** (0.013)
abst_rate2018	−0.421*** (0.018)	−0.041** (0.017)
Constant	−33.148*** (1.377)	85.461*** (1.356)
Observations	7,903	7,903
R^2	0.514	0.451
Adjusted R^2	0.514	0.451
Residual Std. Error (df = 7898)	8.202	8.077
F Statistic (df = 4; 7898)	2,087.637***	1,620.943***
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3: results for fragility models 2018

	Dependent variable:	
	lega18perc	m5s18perc
	(1)	(2)
income_adj18	0.266*** (0.040)	−0.142*** (0.040)
emp_rate18	0.593*** (0.016)	−0.716*** (0.016)
middle_edu18	0.423*** (0.014)	−0.145*** (0.014)
abst_rate2018	−0.398*** (0.018)	−0.040** (0.018)
frag_ind182	0.530 (0.413)	1.697*** (0.408)
frag_ind183	1.093*** (0.417)	1.300*** (0.412)
frag_ind184	0.508 (0.421)	1.826*** (0.416)
frag_ind185	−0.032 (0.426)	2.091*** (0.421)
frag_ind186	0.669 (0.434)	1.453*** (0.429)
frag_ind187	−0.783* (0.456)	2.262*** (0.451)
frag_ind188	−1.960*** (0.478)	2.805*** (0.473)
frag_ind189	−2.713*** (0.507)	3.408*** (0.500)
frag_ind1810	−2.390*** (0.555)	2.433*** (0.549)
Constant	−28.843*** (1.545)	80.877*** (1.526)
Observations	7,903	7,903
R^2	0.520	0.455
Adjusted R^2	0.519	0.454
Residual Std. Error (df = 7889)	8.154	8.053
F Statistic (df = 13; 7889)	657.951***	506.279***
Note:		*p<0.1; **p<0.05; ***p<0.01

We can observe several noteworthy features. First of all, the relation and the significance of the previous two models are maintained. Secondly, with the inclusion of the fragility index variable we observe a different pattern for these municipalities. On the one hand, for Lega, the percentage of votes is higher for the lowest classes (1,2, 3 and 4, so the classes with lower problems, fewer risks and fragility), while lower votes are expected for higher levels of fragility index. On the other hand, for Movimento 5 Stelle, the situation is reversed: less fragile municipalities lead to lower percentage of votes, while higher fragility levels relate to a higher percentage of vote for this party.

This relation is in line with the electoral campaign of that period as Movimento 5 Stelle was quite strong about welfare policies, like the famous "Reddito di Cittadinanza" (amount of money to reintegration in the labour market). Indeed, this policy helped to receive a considerable number of votes in the most fragile areas with high levels of unemployment and low levels of income per-capita.

5.4 Difference model

We have observed how the two studied populisms behaved with OLS models, yet we would like to verify whether the characteristics of the municipalities can have an impact on the variation of the votes across years.

To do so, we took the difference between the percentage of votes obtained in 2019 and 2018, and we checked whether this difference is affected by the conditions of 2019. In this further model, the percentage of immigrants in each municipality is added. Data on the percentage of immigrants were available for 2019 but not for 2018, so in our previous analysis this variable was shelved.

We build the model as follows:

$$y_i = \beta_0 + \beta_1 * Income_i + \beta_2 * MiddleEdu_i + \beta_3 * Employ_i + \beta_4 * AbstRate_i + \beta_5 * FragIndex_i + \beta_6 * Immigrants_i$$

The model is based on the fragility model, with the addition of the percentage of immigrants. The results are as follows:

Table 4: results for models with difference 2018-2019

	Dependent variable:	
	diff_lega	diff_m5s
	(1)	(2)
income_pc19	−0.225*** (0.026)	0.043* (0.022)
emp_rate19	0.010 (0.010)	0.137*** (0.009)
middle_edu19	0.085*** (0.009)	0.038*** (0.007)
abst_rate19	−0.009* (0.005)	0.011*** (0.004)
factor(frag_ind19)2	0.975*** (0.249)	−0.575*** (0.212)
factor(frag_ind19)3	1.400*** (0.257)	−0.344 (0.219)
factor(frag_ind19)4	1.622*** (0.258)	−0.520** (0.219)
factor(frag_ind19)5	1.386*** (0.264)	−0.059 (0.225)
factor(frag_ind19)6	1.159*** (0.266)	−0.048 (0.227)
factor(frag_ind19)7	1.353*** (0.276)	−0.044 (0.235)
factor(frag_ind19)8	0.982*** (0.295)	0.078 (0.251)
factor(frag_ind19)9	0.176 (0.314)	−0.264 (0.267)
factor(frag_ind19)10	−0.553 (0.347)	0.336 (0.296)
immigrants	0.122*** (0.015)	0.024* (0.013)
Constant	15.892*** (0.958)	−24.143*** (0.815)
Observations	7,903	7,903
R^2	0.062	0.071
Adjusted R^2	0.060	0.069
Residual Std. Error (df = 7888)	5.140	4.373
F Statistic (df = 14; 7888)	36.974***	42.772***
Note:		*p<0.1; **p<0.05; ***p<0.01

We can see that the model gives us interesting results. First of all, the percentage of middle education and employment rate coefficients are positive and statistically significant. Secondly, changes in the percentage of votes have a huge impact on the fragility index. In particular, for Lega, we observe that the percentage increases significantly for all the possible values, a symptom of the increased consensus of the right-wing party. From what we observe from the infographics, Movimento 5 Stelle lost consensus and won mainly in the southern region, while Lega conquered the central region, which on average is characterised by a medium fragility index. This factor indicates the shift of these municipalities' support from one party to the other.

Last but not least, the inclusion of the variable 'immigrants' proved to strengthen the support for Lega: the impact on the differential is positive and statistically significant, so a higher differential with a higher concentration of immigrants. This fact is coherent with the main policies that were discussed during the electoral campaign in 2018-2019. While Movimento 5 Stelle focused on welfare policies, Lega's main argument was centred on border security policy, to contrast the illegal immigration from Northern Africa through the Mediterranean Sea, and the system of repatriation of illegal immigrants.

Indeed, for the European election of 2019, the focus of Lega on security policy seems to have been more appealing to the electorate when compared to the welfare policies proposed by Movimento 5 Stelle. It is important to recall that these two parties, after 2018, created the so-called "Governo Giallo-Verde", where the two parties allied to form the new Italian government.

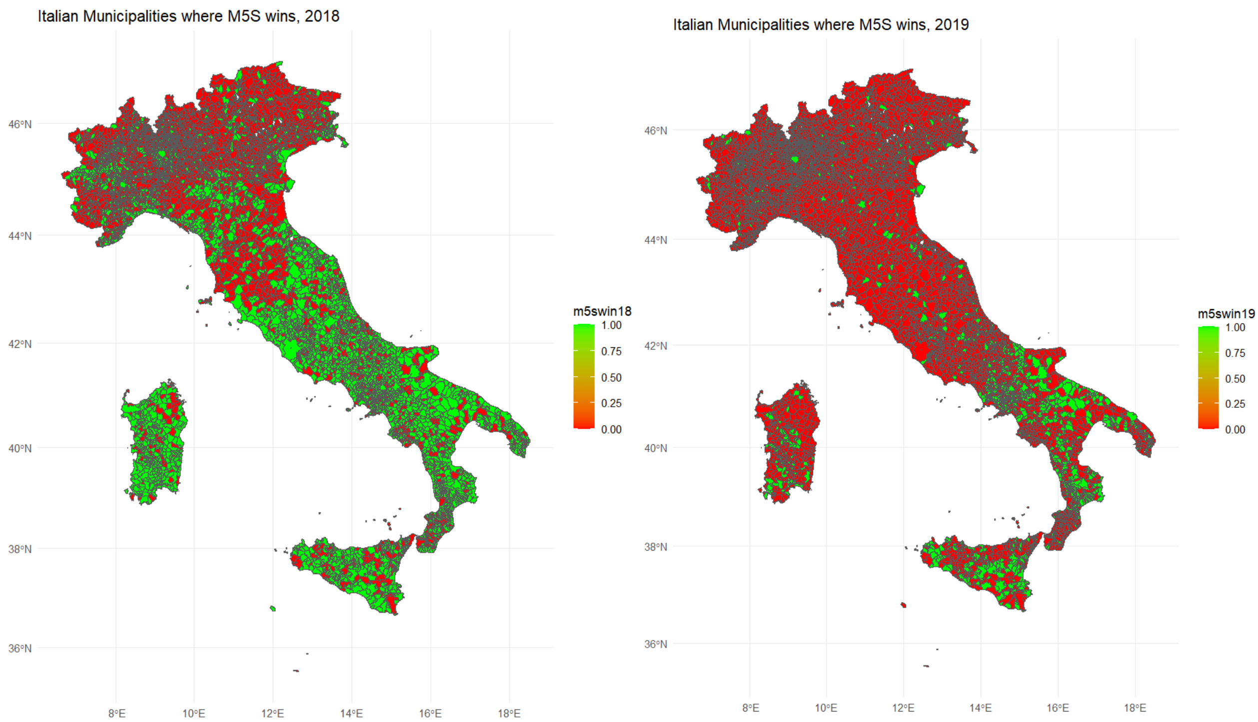

Fig. 12: Municipalities where Movimento 5 Stelle won (in green), 2018 and 2019

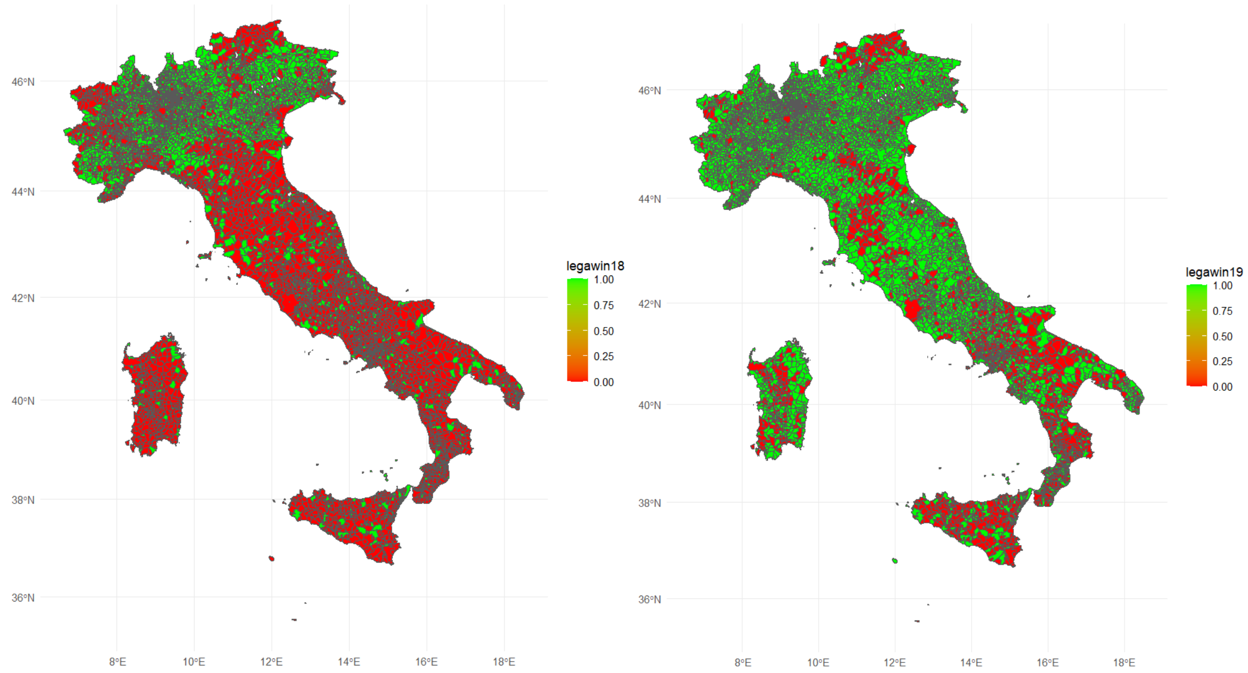

Fig. 13: Municipalities where Lega won (in green), 2018 and 2019

Figure 12 shows that in 2018 Movimento 5 Stelle won mainly in Central and Southern regions (and a few of Northern municipalities), while Lega was concentrated in the northern regions.

In 2019, a change happened: Lega maintained control of Northern Italy, while it gained consensus in the central (and some southern municipalities); on the other hand, Movimento 5 Stelle won at the very tip of southern Italy.

Conclusion

This project aimed to investigate the determinants of the rise of Italian populist parties, starting from micro units of analysis, i.e. the Italian municipalities. Thus, looking specifically at the general election of 2018 and the European one of 2019, this project has shown the emergence of relevant patterns within municipalities in the vote share of the two main populist parties, Lega and Movimento 5 Stelle (M5S). From the analysis, the relation between the geographical location of municipalities and voting decisions is clear, northern regions tended to vote more for Lega, while southern regions opted for Movimento 5 Stelle. Only two regions faced a winning party different from those mentioned before (Emilia-Romagna and Toscana), where Partito Democratico won, being the leading party for 75 years. Moreover, concerning the predictors for the parties' vote shares at the municipality level we have found interesting results, mostly in line with our politically-based expectations. On the one hand, the support for Lega is positively related to income, the percentage of people with middle school diploma, and the employment rate. On the other hand, Movimento 5 Stelle's (M5S) vote share is negatively related to income and the employment rate. Furthermore, Lega received support mostly from less fragile municipalities while M5S electorate was composed mainly of people from most fragile municipalities, especially from the South of Italy.

6 Appendix

In this appendix, we would like to discuss, first of all, what are the potential future paths that this type of research can offer and secondly, we would like to specify some of the statistical techniques that we have used but we have not included in the main text.

6.1 Future Paths

Regarding possible future paths of research, we noticed that a new extreme-right populism rose after 2018-2019, i.e. Fratelli d'Italia. Indeed, after a long history of low support, Fratelli d'Italia won the general elections of 2022, becoming the first party in Italy. This is especially interesting as with the analysed dataset we can see that for both elections (2018-2019) this party has obtained an extremely low share of votes. Figure 14 shows the results obtained by Fratelli d'Italia at the municipality level in the general election of 2018.

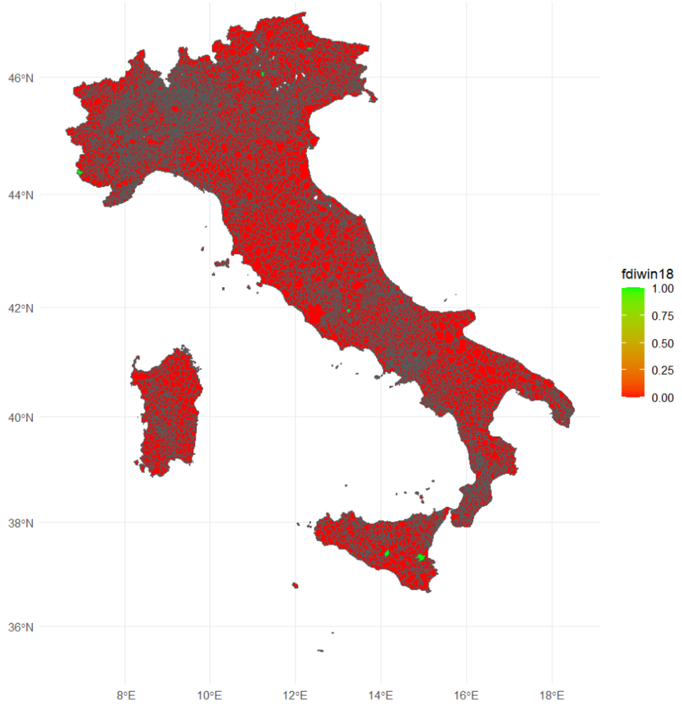

Fig. 14: Map of municipalities where Fratelli d'Italia won, 2018

From Figure 12 it is clear that Fratelli d'Italia won only in 5 municipalities over 7903, literally a drop in the ocean. Considering the result in 2018 and the victory in the general election of 2022, it might be interesting to look at the predictors for such a large difference i.e. repeating the analysis we conducted comparing the 2018 and 2022 general elections at the municipality level. Certainly, the Covid-19 pandemic might have played a role in increasing its consensus, also recalling that Fratelli d'Italia had never been in government before 2022.

Turning now again to the main populist parties we would like to explore more in detail the differences in vote shares for Lega and M5S across elections. Figure 15 below shows the winning probability for Lega at the municipality level by fragility index for 2018.

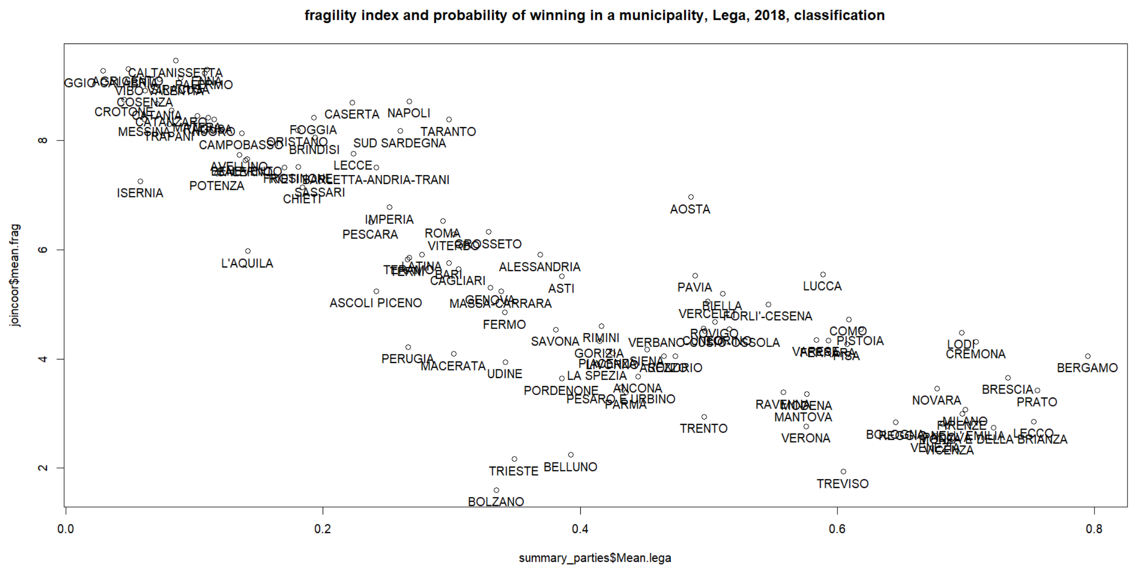

Fig. 15: Fragility index and probability of winning in a municipality, Lega, 2018

From Figure 15 a pattern indeed emerges. Lega has a higher probability of winning in provinces of northern Italy, which also have a low fragility level, while for southern regions we see a probability between 0 and 0.2 of winning in a municipality.

Figure 16 explores the same looking at M5S always for 2018.

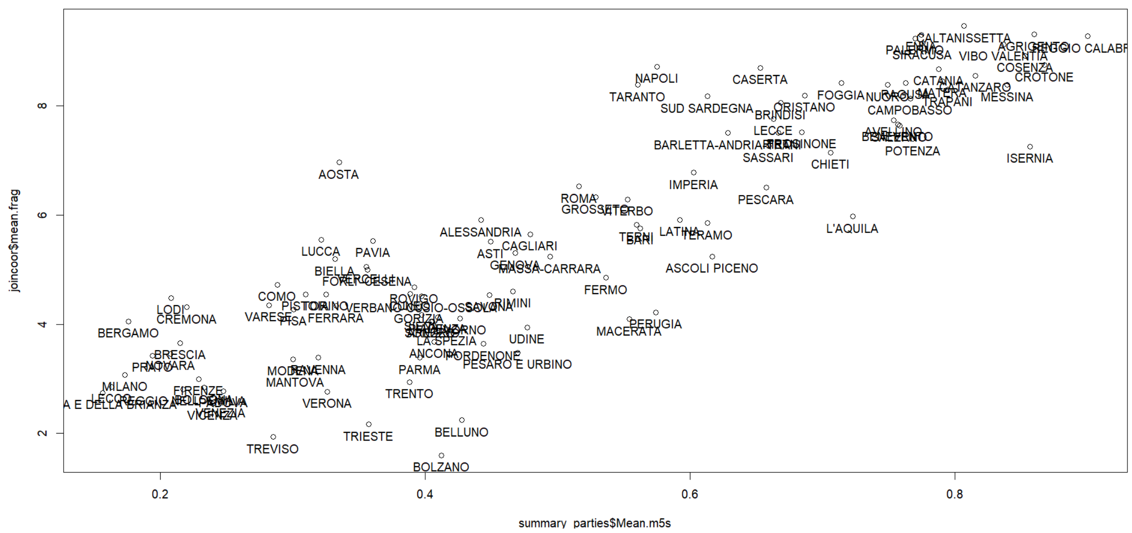

Fig. 16: Fragility index and probability of winning in a municipality, M5S,2018

Coherently with the models seen before, Movimento 5 Stelle has a higher probability of winning in the southern region, associated with a higher fragility index, while a lower probability is observed for the northern region. In the case of 2018, comparing the x-axis, Movimento 5 Stelle had a higher probability of winning in a random municipality, only a few provinces here face a probability lower than 0.2, while for lega the majority of districts of southern Italy had a probability lower than 0.2.

This is coherent with the results, especially considering M5S's focus on welfare policies, like "Reddito di Cittadinanza", which created great consensus, especially from possible beneficiaries located throughout the country.

Figure 17 explores the same for Fratelli d'Italia in 2018.

Figures 16 and 17 showed that the probability of winning in a municipality (averaged at the province level) is quite sensitive to geographical location for the main populist parties when they had their spike in

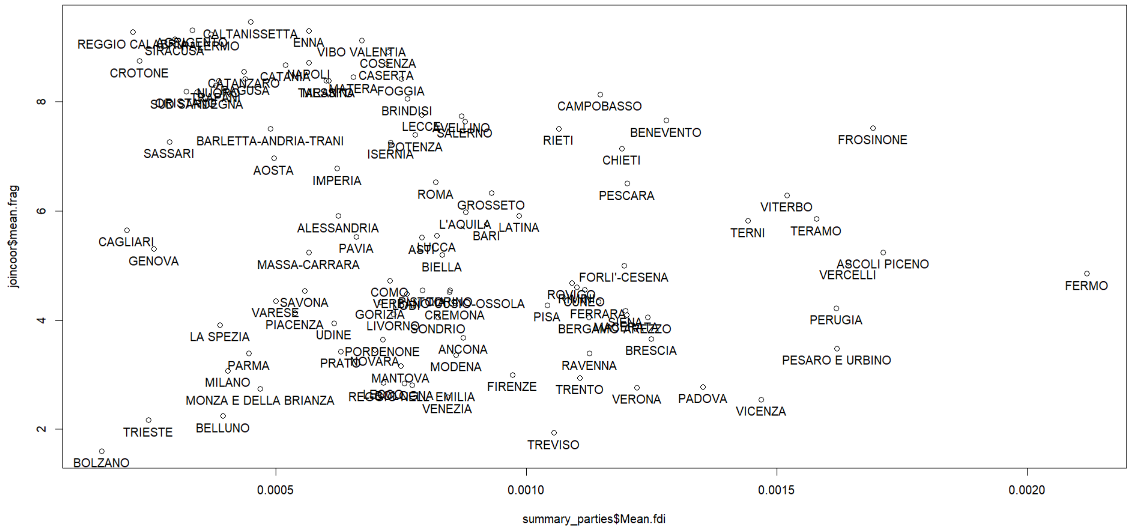

Fig. 17: Fragility index and probability of win in a municipality, Fratelli d'Italia, 2018

consensus. The winning probability for Fratelli d'Italia, instead, seems more heterogeneous at the national level and does not depend on location. This heterogeneity might also have had implications on the consistent and homogeneous growth of consensus of Fratelli d'Italia in the whole Italian peninsula.

6.2 Additional infographics on Partimat

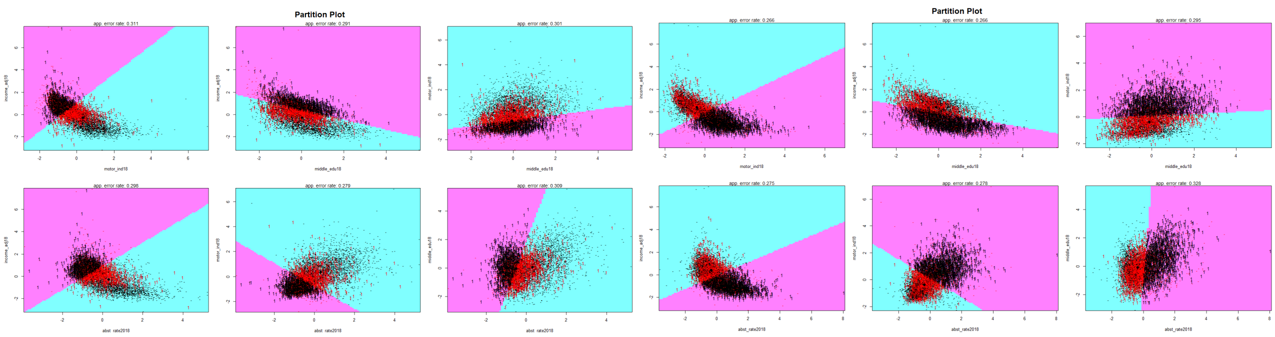

Fig. 18: Partimat of Lega (left) and M5S (right) on the logistic model, 2018

It is observed from these graphs that the pattern is very similar to the QDA case seen before. The approximated error is in the range [0.25,0.28], and the error in the prediction is concentrated on the boundary between the two regions.

6.3 Cross-Validation

The figures below explore the distribution of the real votes obtained by the parties: For Lega, we observe a spike in municipalities with a percentage of votes of 7 or 8, then a consistent number between 20 and 30 per cent. The spike can be attributed to southern regions where Movimento 5 Stelle took over.

In the figure below, we observe that the spike of votes is around 20 percent, and a consistent (cambia aggettivo) number of votes is still at 40-45 percent.

For PD, we have a distribution of votes with percentage mainly between 10 and 25; seems to be quite consistent in terms of votes, because we do not observe a spike like in Lega; the maximum density of votes is quite consistent for several percentage points.

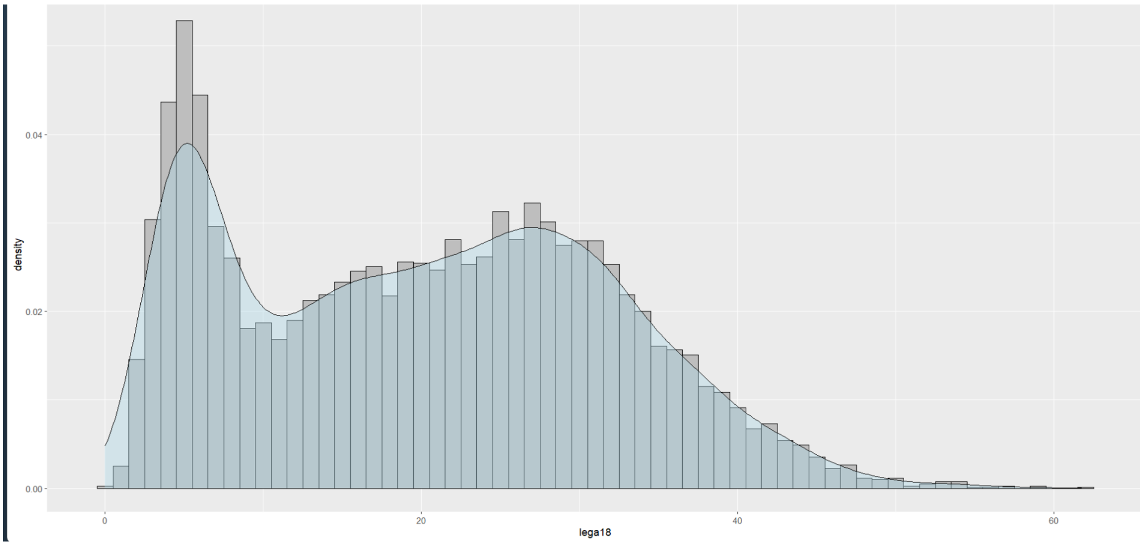

Fig. 19: Kernel estimation and density Lega 2018

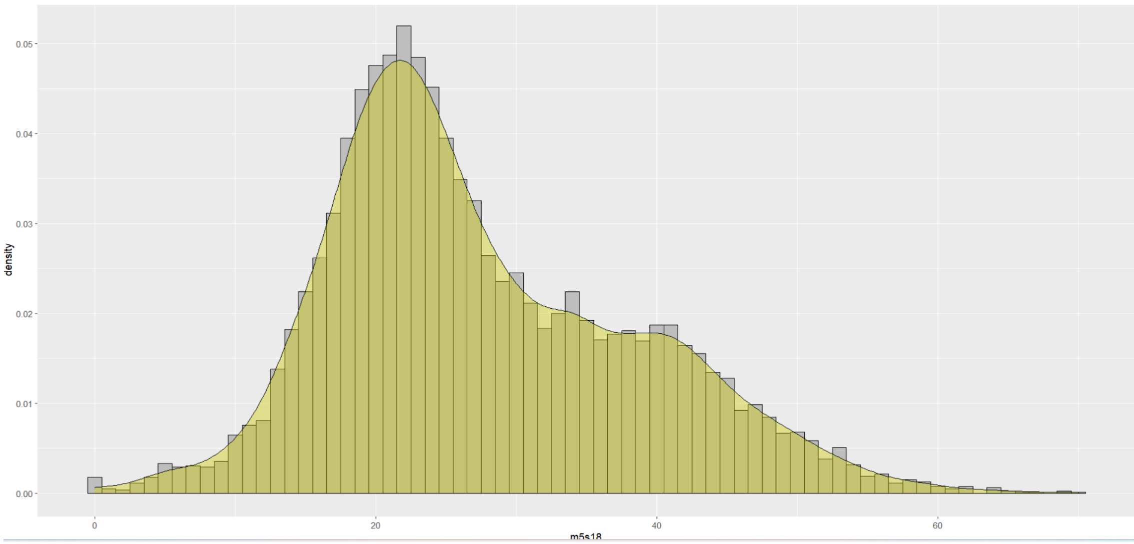

Fig. 20: Kernel estimation and density M5S 2018

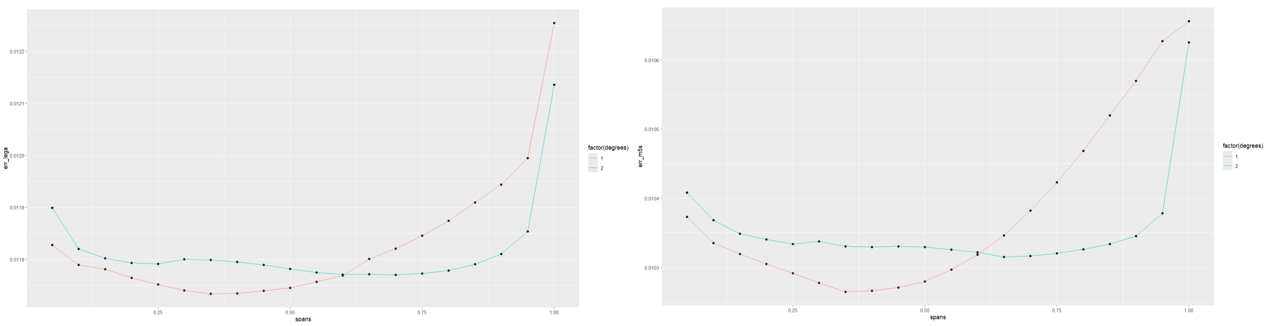

Fig. 22: K-folds cross validation Lega(left) and M5S(right), 2018

From the k-fold cross-validation representation, we see that both parties face the lowe(r) RMSE with only 1 degree, at a level of span of 0.3; the picture for both parties and both years are similar.

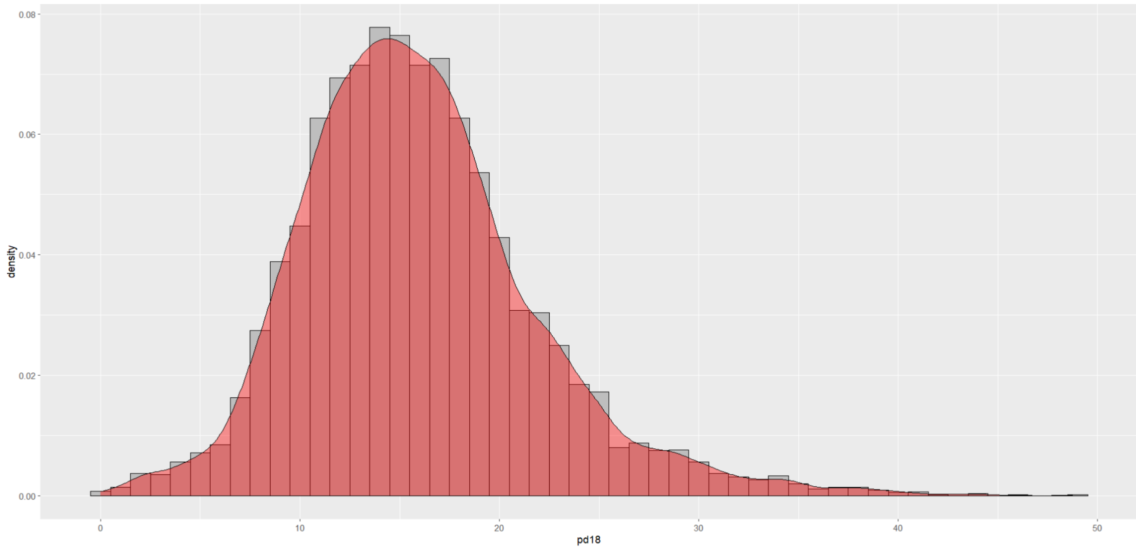

Fig. 21: Kernel estimation and density PD 2018

6.4 Additional infographics

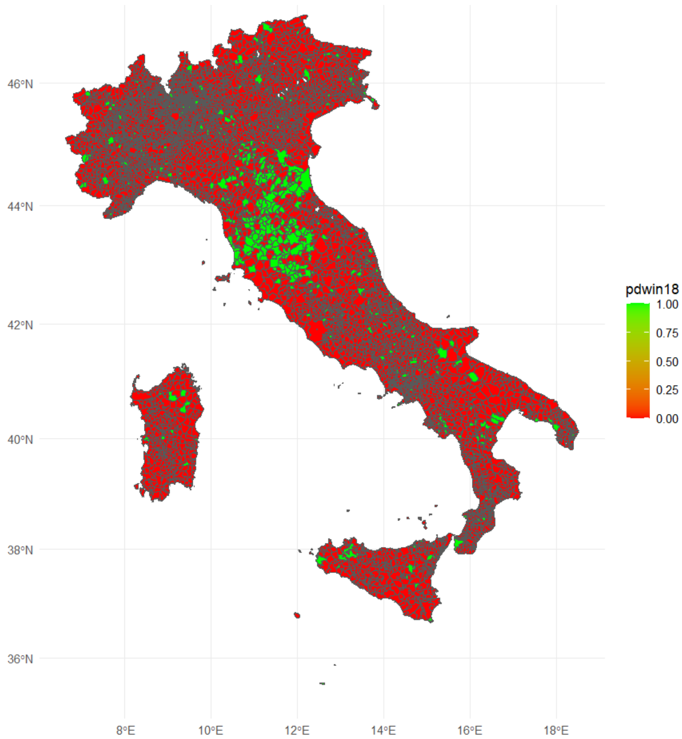

Fig. 23: Municipalities where PD won (in green),2018

From these infographics, we can see that PD won mainly in the municipalities located in Emilia-Romagna and Toscana, which historically tend to vote for this party, while in the other regions, the two populisms took over, not leaving space to PD.

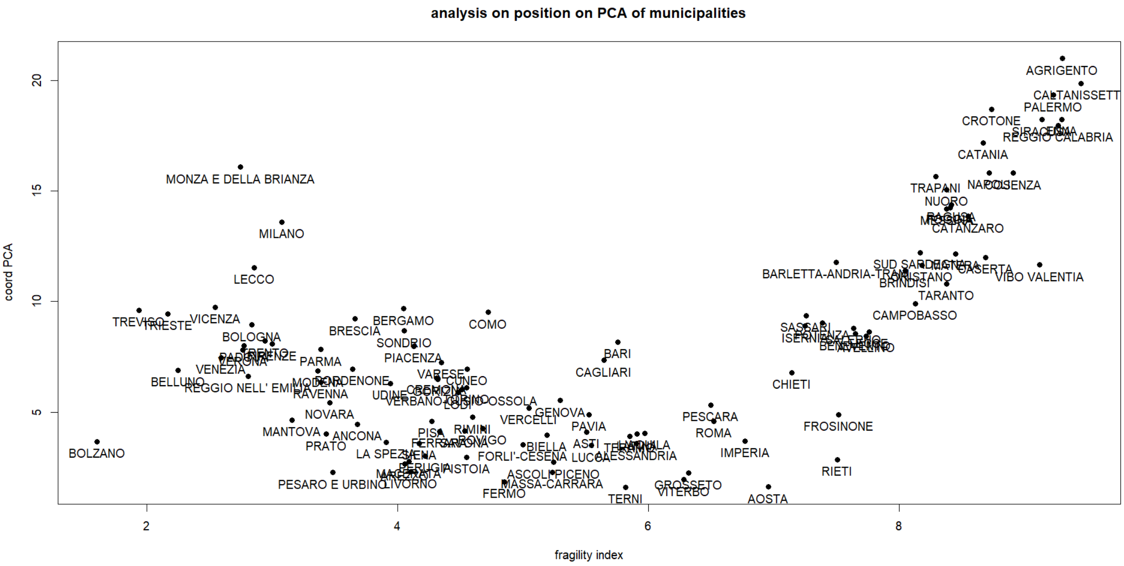

Fig. 24: Relation between coordinates in PCA and fragility index, aggregated by province, 2018

7 Bibliography

Borghese, S. (2021, May 15). Europee 2019: Un Primo Bilancio. YouTrend. https://www.youtrend.it/2019/05/27/elezioni-europee-2019-analisi-bilancio/

Chiaramonte, A., De Sio, L., Emanuele, V. (2020). Salvini's success and the collapse of the Five-star Movement: The European elections of 2019. Contemporary Italian Politics, 12(2), 140-154.

Gareth, J., Daniela, W., Trevor, H., Robert, T. (2013). An introduction to statistical learning: with applications in R. Spinger.

Istat Data. Data browser. https://esploradati.istat.it/databrowser// (IFC)

Levi, E., Patriarca, F. (2020). An exploratory study of populism: the municipality-level predictors of electoral outcomes in Italy. Economia politica, 37, 833-875.

Ministero dell′Interno. (n.d.). Eligendo - Ministero dell′interno DAIT. https://elezioni.interno.gov.it/

Mudde, C. (2004). The populist zeitgeist. Government and opposition, 39(4), 541-563.

Rooduijn, M., Pauwels, T. (2011). Measuring populism: Comparing two methods of content analysis. West European Politics, 34(6), 1272-1283.