Exploring the effects of socioeconomic factors on voter preferences: A case-study of France 2022

Niloufar Pourshir Sefidi*[1] and Peter Mooney†[1]

[1]Hamilton Institute, Maynooth University
[2]Department of Computer Science, Maynooth University

GISRUK 2023

Summary

Understanding the reasons behind voter preferences in elections is time-consuming and resource intensive using traditional methods such as surveys. In this paper, we describe a flexible statistical modelling approach to investigate the effect of different socioeconomic factors on voters' preferences for the 2022 French presidential election. Our results show that socioeconomic factors have a significant effect on the voters' preferences in this election. The methodology developed can be used to investigate these effects for elections in other countries where appropriate socioeconomic and spatial data is available.
KEYWORDS: French elections, Spatial modelling, Voter behaviour, Socioeconomic factors.

1 Introduction

Recognizing the drivers of voter preferences offers clear indications not only for predicting the election's outcome but also for understanding voter behaviours and their influences. Many studies have sought to understand the factors influencing the voting behaviours of citizens (Leigh, 2005; Sigelman and Sigelman, 1982; Mutz, 2018) with two main approaches emerging. The first approach is to conduct surveys with individual voters and ask them about their preferences and the reasoning behind their choices (Guth et al., 2006; Branton, 2003; Mutz, 2018). This approach is at an individual level, is very time-consuming, and is an expensive process to repeat. Sample size is limited by the budget (financial and time) available. This approach is also limited by the fact that the determinants of voters' preferences vary geographically. The second approach, which utilizes large-scale census data and election results, can overcome these limitations (Kim et al., 2003; Scala et al., 2015; Miller and Grubesic, 2021). Socioeconomic factors that influence voting preferences vary by location (Gelman et al., 2005). While the identification of these factors has grown over the years, it is incorrect to assume that their spatial distribution is known. Political science researchers have emphasized that these effects should be considered exogenous and unknown. Instead, their spatial

*niloufar.pourshirsefidi.2022@mumail.ie
†peter.mooney@mu.ie

distribution should be estimated for each dataset rather than assumed a priori (Calvo and Escolar, 2003; Darmofal, 2008; Mansley and Demšar, 2015). Darmofal (2008) demonstrated spatial variation in the effects of socioeconomic factors on voting behaviour during the Democratic realignment period (1928-1936). Miller and Grubesic (2021) investigated local halo effects and spatially varying effects of socioeconomic factors on Republican support in the 2016 U.S. presidential election.

Many models have been proposed for spatial modelling such as geographically weighted regression (Brunsdon et al., 1996), Bayesian spatially varying coefficient model (Gelfand et al., 2003), Multinomial logit model (Dow and Endersby, 2004) and Poisson regression model with spatial random effects (Bernardinelli et al., 1995). In this work, we utilize the Poisson log-linear model as a less computationally expensive alternative to the Multinomial logit model for modelling candidate votes. We aim to investigate how various socioeconomic factors impact voters' preferences for the 2022 French presidential election. To our knowledge, this is the first study to use a Poisson log-linear model with spatial random effects to examine the influence of socioeconomic factors on voters' preferences for the French presidential election. Our findings reveal that socioeconomic factors significantly shape voters' preferences for the French presidential election and that their impact varies spatially.

2 Methodology and Data

2.1 Data description

We investigate the effect of several socioeconomic factors (immigration rate, poverty rate, higher education rate, average life expectancy, unemployment rate and white-collar rate) on voters' preferences for the 2022 election at French department[1] level. We utilised three data sources:

- **Election data:** Results of the first round of the 2022 French presidential election. This data consists of 12 candidates and their votes in 96 departments in Metropolitan France.
- **Census data:** Census data is recorded annualy by INSEE (Institut National de la Statistique et des Etudes Economiques). We used either 2019 or 2021 at the department level. A summary of relevant variables and their definition is shown in Table 1.
- **French department boundaries:** We used an openly available ESRI Shapefile for the French department boundaries.

2.2 Modelling procedure

Candidate votes can be modelled using a Multinomial Logit Model (MLM). Accordingly, $Y_i = \{y_{i1},\ ..., y_{ij},\ ...\}$ is the vector of votes in the department $i = 1,\ ...,\ I$ and for the candidates $\{j;\ 1\ , \ldots,\ J\}$. We can assume:

$$\mathrm{Y_i} \sim Multinomial\,(\pi_\mathrm{i}) \tag{1}$$

where $\pi_\mathrm{i} = \{\pi_\mathrm{i1}, \ldots, \pi_\mathrm{ij}, \ldots\}$ is the vector of vote proportions of the candidates. The proportions can then be regressed on the independent variables (e.g., poverty rate). However, as this model

[1] `https://en.wikipedia.org/wiki/Departments_of_France`

Table 1: Socioeconomic variables and their definition

Variable	**Definition**
Immigration rate	The percentage of the population classified as immigrants.
Poverty rate	The proportion of individuals considered monetary poor.
Higher education rate	The percentage of individuals over 18 years old with a higher education degree.
Average life expectancy	The average life expectancy of people.
Unemployment rate	Percentage of unemployed people in the total labour force.
White-collar rate	Percentage of population with a white-collar job (i.e., a job that requires a university degree).

specification is computationally expensive we do not use it. Instead, we use Poisson log-linear model that has been shown to be equivalent to MLM (Baker, 1994; Lee et al., 2017) and is supported by R-INLA (Rue et al., 2017) in R for spatial modelling. For more information on MLM see Dow and Endersby (2004). The Poisson log-linear model used here is defined as follows:

$$\begin{aligned} y_{ij} &\sim Poisson\left(\lambda_{ij}\right) \\ \log\left(\lambda_{ij}\right) &= \phi_i + X_i B_j + s_{ij} \end{aligned} \tag{2}$$

where y_{ij} is the number of votes for candidate j in the department i, ϕ_i is the intercept for department i, X_i is the vector of covariates for department i, B_j is the vector of coefficients for candidate j, and s_{ij} is the spatial random effect for candidate j in department i. We assume the same structure for all the candidates and drop the "j" subscript for notation simplicity in the following descriptions. We use the Besag model (Besag, 1974) for s_i that is:

$$s_i \mid s_{k \neq i} \sim N\left(\frac{1}{n_i}\sum_{k \sim i} s_i, \frac{\sigma_s^2}{n_i}\right) \tag{3}$$

where $k \sim i$ refers to all neighbors of the department i, with n_i representing the total number of neighbors and σ_s^2 is the variance of the spatial random effect.

3 Results

With the Poisson transformation used the effect of each predictor on the candidates' votes is not identifiable. However, the difference in the effects is identifiable and we can interpret these differences. From equation 2, we can calculate the exponentiated difference of coefficients. This indicates the percentage change in the vote ratio for one candidate compared to another. Our analysis focused on how each predictor affected the vote ratio of candidates. As a result, we presented the effects of six predictors: higher education rate, unemployment rate, white-collar rate, poverty rate, immigration rate, and average life expectancy in figure 1. We focus on Macron, Le Pen, and Melenchon as the primary winners during the first round of the 2022 election.

We report that the vote ratio of Macron to Le Pen decreased by 4.98% for every unit increase in the unemployment rate in a department. This figure is a 5.30% increase for a unit increase in the white-collar rate and a 3.82% decrease for a unit increase in the poverty rate. Effects for higher education rate, immigration rate, and average life expectancy were not significant.

The vote ratio of Melenchon to Macron showed unit increases of 16.63%, 10.80% and 2.23% for higher education rate, poverty rate, and immigration rate respectively. However, this ratio showed a decrease of 8.33%, 1.04% and 3.34% for the unemployment rate, white-collar rate, and average life expectancy in a department.

The vote ratio of Melenchon to Le Pen increased by 13.98% for every unit increase in the higher education rate, 4.20% for a unit increase in the white-collar rate, 6.56% for a unit increase in the poverty rate, and 2.47% for a unit increase in immigration rate. On the other hand, the vote ratio decreased by 12.90% for a unit increase in the unemployment rate and by 2.71% for one year increase in average life expectancy in a department.

Furthermore, the spatial random effects indicate that spatial patterns are not explained by the predictors in our model. Figure 2 show the map of spatial random effects for Macron and Le Pen. A pattern exists showing overall positive values in the east and negative values in the western parts of France for Le Pen while for Macron positive values are in the north and negative values in the south.

4 Conclusions and Future Work

Our study involved developing a model to predict the vote ratio of candidates in the 2022 French presidential election and evaluating how various socioeconomic variables influenced voting patterns. We discovered that there were distinct geographic patterns in the way people voted which were tied to different socioeconomic factors. Unemployment rate, white-collar rate, poverty rate, and average life expectancy all played a significant role in determining the winner. We also found that higher education rate, unemployment rate, white-collar rate, and poverty rate were all significant factors in determining whether Melenchon would win over Macron and Le Pen. There are several avenues for further exploration in this research. Firstly, we could employ the datasets from the 2012 and 2017 presidential elections to draw comparisons with the 2022 election results. Secondly, we could apply the same model to forecast the vote ratio of candidates in the 2022 election, using data from the 2012 and 2017 elections. Lastly, we could test the effectiveness of our model in other countries to determine whether the findings align with those observed in the French presidential election.

Acknowledgements

This work has emanated from research conducted with the financial support of the Science Foundation of Ireland (SFI) under Grant Number 18/CRT/6049. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Science Foundation of Ireland.

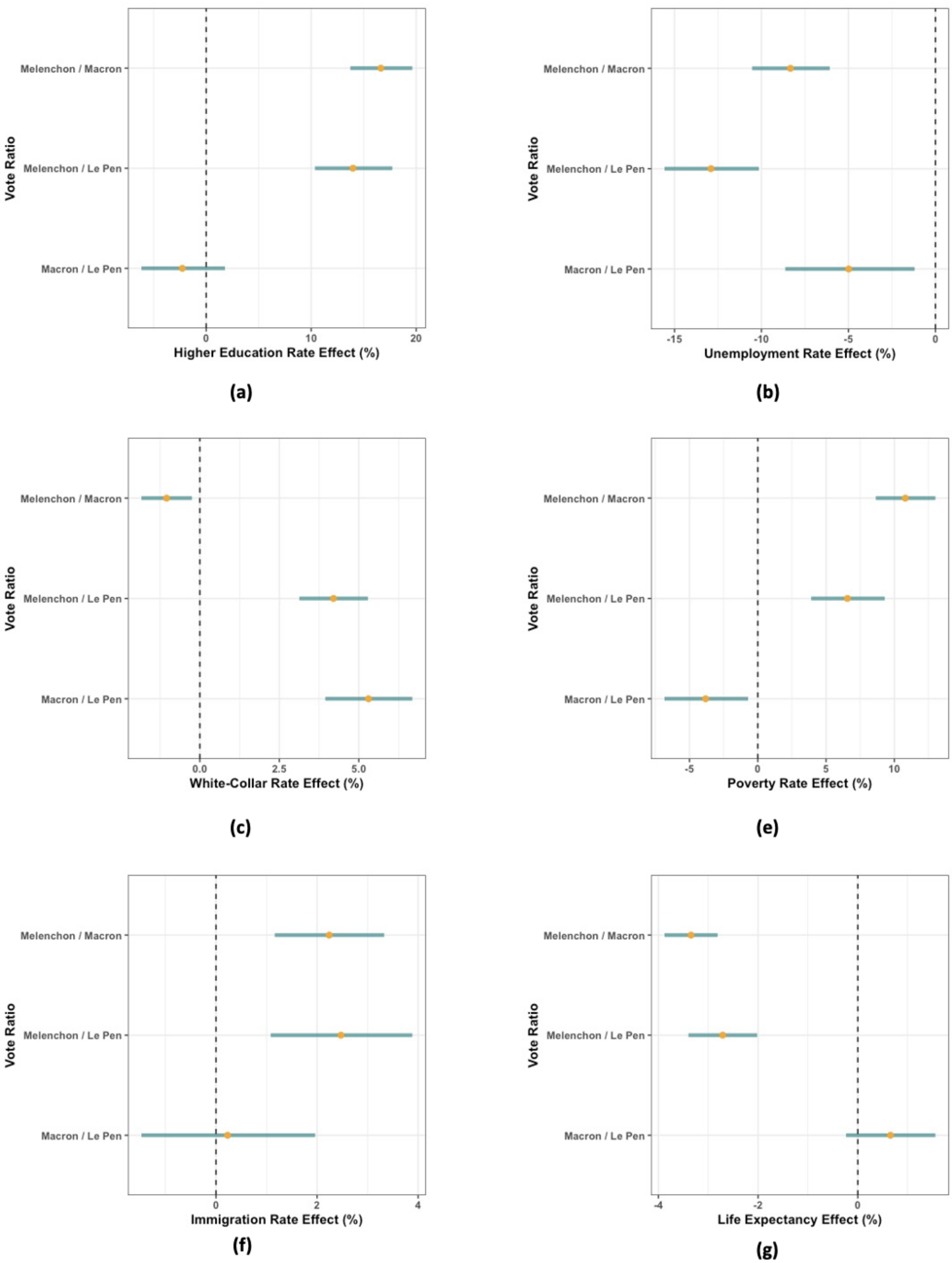
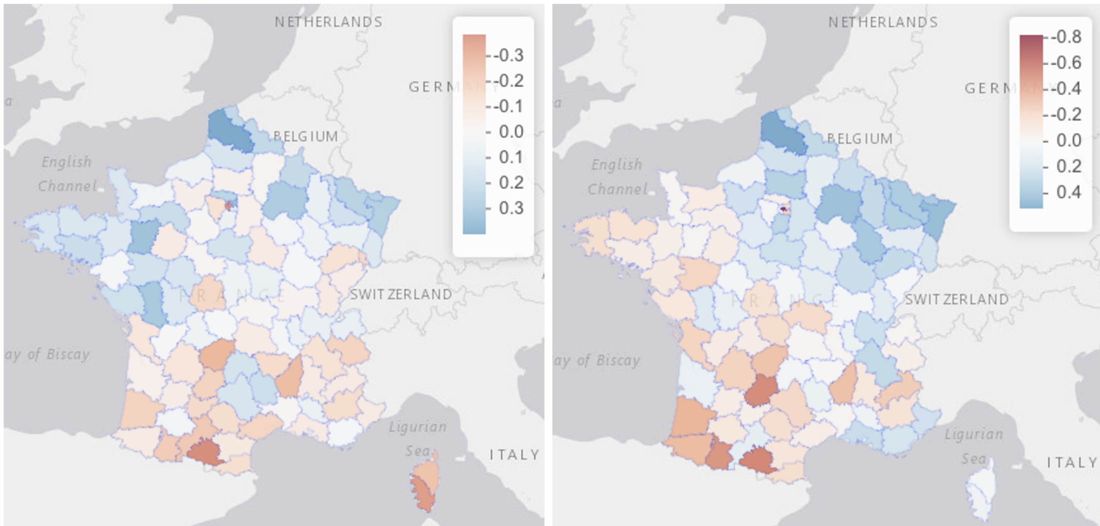


Figure 1: The effect of each predictor on the percentage changes of vote ratio for candidate i against candidate j is depicted on the x-axis and i/j is shown on the y-axis.



(a) (b)

Figure 2: a) Values of the spatial random effect for Macron b) Values of the spatial random effect for Le Pen, mapped on the French departments.

Reproducibility

We provide the final version of the dataset we used for our study, and to ensure reproducibility and transparency to our model and results, the code for running the analysis is available online at `https://github.com/nilips70/French-Election-2022`. Raw census data files for departments, raw 2022 first-round French election result data and the French departments' boundaries data are extracted from INSEE[2], Open Data France[3] and Berkely library[4] websites.

Biography

Niloufar Pourshir Sefidi is a PhD student in the Centre for Research Training (CRT) Doctoral training programme in Data Science at Maynooth university. Niloufar's research interests are spatiotemporal analysis, big data analysis and machine learning.

Peter Mooney is an Assistant Professor of Computer Science at Maynooth university and is currently supervising Niloufar for her PhD. Peter has been involved in teaching and research related to spatial analysis and geographic data management for 15 years.

[2]`https://www.insee.fr/fr/accueil`
[3]`https://www.data.gouv.fr/`
[4]`https://geodata.lib.berkeley.edu/`

References

Baker, S. G. (1994). The multinomial-poisson transformation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *43*(4), 495–504.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, *14*(21-22), 2433–2443.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225.

Branton, R. P. (2003). Examining individual-level voting behavior on state ballot propositions. *Political Research Quarterly*, *56*(3), 367–377.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, *28*(4), 281–298.

Calvo, E. & Escolar, M. (2003). The local voter: A geographically weighted approach to ecological inference. *American Journal of Political Science*, *47*(1), 189–204.

Darmofal, D. (2008). The political geography of the new deal realignment. *American Politics Research*, *36*(6), 934–961.

Dow, J. K. & Endersby, J. W. (2004). Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Electoral studies*, *23*(1), 107–122.

Gelfand, A. E., Kim, H.-J., Sirmans, C., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*(462), 387–396.

Gelman, A., Shor, B., Bafumi, J., & Park, D. (2005). Rich state, poor state, red state, blue state: What's the matter with connecticut? *Poor State, Red State, Blue State: What's the Matter with Connecticut*.

Guth, J. L., Kellstedt, L. A., Smidt, C. E., & Green, J. C. (2006). Religious influences in the 2004 presidential election. *Presidential Studies Quarterly*, *36*(2), 223–242.

Kim, J., Elliott, E., & Wang, D.-M. (2003). A spatial analysis of county-level outcomes in us presidential elections: 1988–2000. *Electoral studies*, *22*(4), 741–761.

Lee, J. Y., Green, P. J., & Ryan, L. M. (2017). On the" poisson trick" and its extensions for fitting multinomial regression models. *arXiv preprint arXiv:1707.08538*.

Leigh, A. (2005). Economic voting and electoral behavior: how do individual, local, and national factors affect the partisan choice? *Economics & Politics*, *17*(2), 265–296.

Mansley, E. & Demšar, U. (2015). Space matters: Geographic variability of electoral turnout determinants in the 2012 london mayoral election. *Electoral Studies*, *40*, 322–334.

Miller, J. A. & Grubesic, T. H. (2021). A spatial exploration of the halo effect in the 2016 us presidential election. *Annals of the American Association of Geographers*, *111*(4), 1094–1109.

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences*, *115*(19), E4330–E4339.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with inla: a review. *Annual Review of Statistics and Its Application*, *4*, 395–421.

Scala, D. J., Johnson, K. M., & Rogers, L. T. (2015). Red rural, blue rural? presidential voting patterns in a changing rural america. *Political Geography*, *48*, 108–118.

Sigelman, L. & Sigelman, C. K. (1982). Sexism, racism, and ageism in voting behavior: An experimental analysis. *Social Psychology Quarterly*, (pp. 263–269).