

Spatial Autocorrelation in Voting Turnout

Mahdi-Salim Saib*

CEVIPOF-Centre de recherches politiques de Sciences Po 98, rue de l'Université - 75007 Paris, France

Abstract

The presence of spatial autocorrelation in the data can yield biased or inconsistent point estimates when Ordinary Least Squares (OLS) model is used inappropriately. Therefore, in this paper we try to assess the fit of the model taking into account the autocorrelation in analyze of voting behavior in the 2007 French Presidential Elections and the 2010 French Regional Elections. We find that the voter turnout in the Ile de France region is spatially structured and that the Simultaneous Auto-Regressive (SAR) model clearly improves the quality of adjustment compared with the OLS model for the both elections.

Keywords: Spatial autocorrelation; Elections; Socioeconomic indicator

Introduction

Spatial analysis of electoral outcomes is at a nascent stage in France. Since 2002, the Interior Ministry provides electoral data in a digitized format, allowing cartographic treatment to previously prohibited scales: thousands of cantons, or even several tens of thousands of counties [1]. Nowadays, it is not surprising that the majority of studies that compared these data with socio-economic data to analyze the voting behavior were limited to either observe and stack several layers, or to use either often inappropriate methods to spatial data [2,3]. One of the main features of these data is spatial autocorrelation, which measures the degree of interaction and interdependence between observations spatially localized according to the law of Tobler [4]: "Everything is related to everything else, aim near things are more related than far things".

Though several studies pointed out that many theories in political science predict spatial clustering of similar behaviors among neighboring regions and emphasizes the importance of using proper diagnostic tools to determine the type of spatial autocorrelation within the data [2,5]. It is unsurprising that the majority of aggregate analyses of voting patterns in the France are conducted at the department scale. However, few aggregate analyses, that employ regression techniques to analyze voting patterns explicitly, take into account spatial effects.

Spatial regression analysis through creating a specific-contiguity weight allows taking this autocorrelation into account and allows examining the relationship between the attributes of interest and explanatory variables that can interpret the observed spatial pattern [6]. In this study, conducted over 1300 counties of the Ile-de-France region, we applied statistics based Moran to highlight the existence of turn out inequality spatially structured through space, and we used simultaneously Autoregressive (SAR) regression models [7] to identify the share of regional inequalities of participation that stems directly from the specific socioeconomic composition of the studied areas.

The choice to focus the analysis on the Ile de France region is primarily justified by the fact that the research results presented here are part of the 3rd Axis of earth policy program, including Axis interested to "Dynamics of critical areas and urban conflicts" in this region. Another the Ile-de-France region is characterized by very high socio-economic disparities, marked between departments, but also notable at intra-departmental or even intra-county levels, this gradient forms in itself a relevant scale of analysis for electoral and socio-economical works [8]. This article deals only with the methodological asp; we will not focus

on the interpretation of the résultat in substantial terms. However, we try to assess the fit of the model taking into account the autocorrelation in analyze of voting behavior.

Methods

Election data

The election data came from the Ministry of Interior of two elections: the 2007 French Presidential Elections and the 2010 French Regional Elections (first round in both cases). We chose in purpose a low (Regionals) and a high intensity election (Presidentially) to check if the abstention in these both different political configurations presents the same spatial structure. We consider the first round in both cases, as all political parties, from extreme left to extreme right, are present. Which allows us a comparison without partiality of political part Figure 1 shows spatial distribution of the Abstention: A) 2007 French Presidential Elections and B) 2010 French Regional Elections?

Socioeconomic indicator

To characterize accurately the socio-economical level of a county in the region, we used: The deprivation indicator (FDep) developed by Rey [9,10]. The concept of the urban unit developed by the National Institute of Statistics and Economic Studies (INSEE) was used to define the degree of urbanicity. There are five categories of urban unit: rural (less than 2,000 people), quasi-rural (population 2000 to 9999), quasi-urban (population of 10,000-99,999), urban (population of 100,000 to 1,999,999) and-suburban (population >2,000,000). The indicator was built at the county using the following socioeconomic variables census data from the 2008 population: median household income, percentage of high school graduates in the population aged 15 and over, the percentage blue-collar workers in the active population and the unemployment rate. The socio-economic index (SI) was defined as the weighted sum of these four variables by the first principal component of PCA and stratified in four degrees of district classes of urbanicity [11].

*Corresponding author: Saib MS, CEVIPOF-Centre de recherches politiques de Sciences Po 98, rue de l'Université - 75007 Paris, France, Tel: 33(0)1 45 49 51 05; E-mail: saib.mahdi.s@gmail.com

Received October 05, 2017; Accepted October 26, 2017; Published October 30, 2017

Citation: Saib MS (2017) Spatial Autocorrelation in Voting Turnout. J Biom Biostat 8: 376. doi: 10.4172/2155-6180.1000376

Copyright: © 2017 Saib MS. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

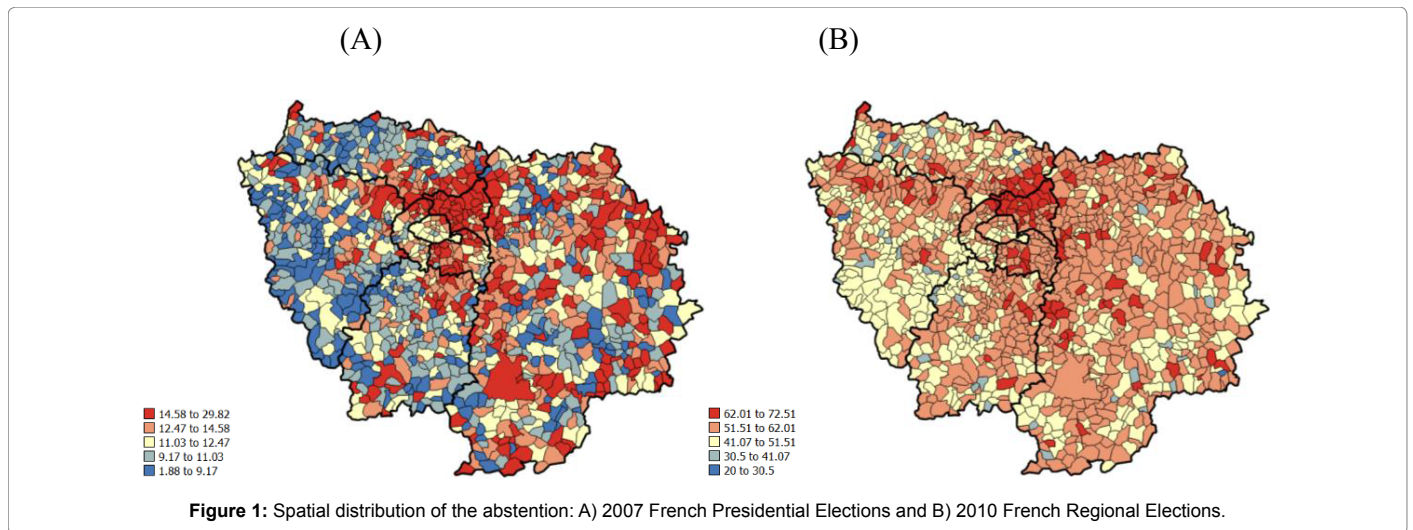


Figure 1: Spatial distribution of the abstention: A) 2007 French Presidential Elections and B) 2010 French Regional Elections.

Spatial autocorrelation

The analysis of local spatial autocorrelation was introduced by Anselin. The Moran index (Moran) makes it possible to measure the level of spatial autocorrelation of a variable and to test its significance. It is equal to the ratio of the covariance between contiguous observations (defined by the matrix of interactions) to the total variance of the sample. The local Moran index measures the degree of spatial correlation at the local level for each spatial unit. As in the case of the Moran index, one can calculate the Z-scores and test the significance of the degree of local spatial autocorrelation. Significant records can be represented as maps [12,13].

Spatial regression analysis

Many studies use linear regression analyses to determine the relationships between voting behavior and socio-economies datas [3,5,14].

$$\gamma_i = \beta_0 + \sum_k \beta_k x_{ki} + \varepsilon_i$$

where γ is the dependent variable for observation i , β_0 is the intercept, β_k is the regression coefficient (slope) of each factor x_k , and ε_i is the error term. However, for analysis of observational data with spatial dependence, the classical linear regression model with spatial autocorrelated residuals violates the independence assumption for error. Simultaneous autoregressive (SAR) models such as “SAR lag” and “SAR error” models are among the most commonly used. We based ourselves on the Lagrange multiplier test statistics developed by Anselin et al. [5,13-15] to select the best specification of the SAR model. This led us to choose a SAR lag model.

$$y = \rho W y + x \beta + \varepsilon; \text{ with } \varepsilon.$$

The SAR lag model is similar to the classical linear regression model in which a spatially lagged dependent variable W is included to control for spatial autocorrelation. W corresponds to a spatial weight matrix that defined the notion of neighborhood between geographic units, and ρ to a spatial autoregressive parameter that estimates the scale of interactions between the observations of the dependent variable [16,17].

In this study we tested two types of models. In the first, the SE indicator was introduced directly in the model. In the second model, the SE indicator was considering as a categorical variable which we

| Elections | Moran's I |
|-------------------------------|-----------|
| French Presidential elections | 0.28** |
| French Regional elections | 0.31** |
| **p<0.01. | |

Table 1: Global Moran's of abstention French Presidential elections and abstention French Regional elections.

divided into 5 categories (approximately equivalent to quintiles). The first category comprised the least deprived counties which served the reference class. Finally, we used the Akaike Information Criterion to compared the goodness-of-fit of different regression models.

Results

The results of the global Moran's I analyses are summarized in Table 1. Table 1 shows the positive and statistically significant spatial autocorrelation for 2007 abstention French Presidential elections and 2010 French Regional elections. Figure 2 displays a map showing the geographical distribution of high cluster in these both different political configurations. For these both French elections, almost the same aggregation areas of high cluster of abstention, located in the North East of Paris (capital of the region) were observed centered on STAINS, SAINT-DENIS and AULNAY-SOUS-BOIS counties.

The spatial autocorrelation in residuals for ordinary least squares (OLS) regression was found by the Moran's I statistic. Moreover, Lagrange multipliers (LM) and Robust LM for spatial lag were both statistically significant in favor of conducting SAR lag model regression. Table 2 shows the estimation results of non-spatial and SAR lag model regressions for abstention French Presidential elections and the French Regional elections.

For non-spatial regression, the socioeconomic index (SE) is significantly and positively spatially associated with abstention in each French election. The estimation of all SLM models had significantly positive values for spatial effect. In addition, the percentages of variance explained (R^2) by the SAR lag models were greater than that in the non-spatial regressions, indicating that spatial regression model was successful in accounting for spatial correlation.

The socioeconomic index is also significantly and positively associated with abstention in the two SAR lag models regressions for these both different elections. The significant regression coefficients (β) for the socioeconomic index were 0.61 and 1.04, and the percentages of

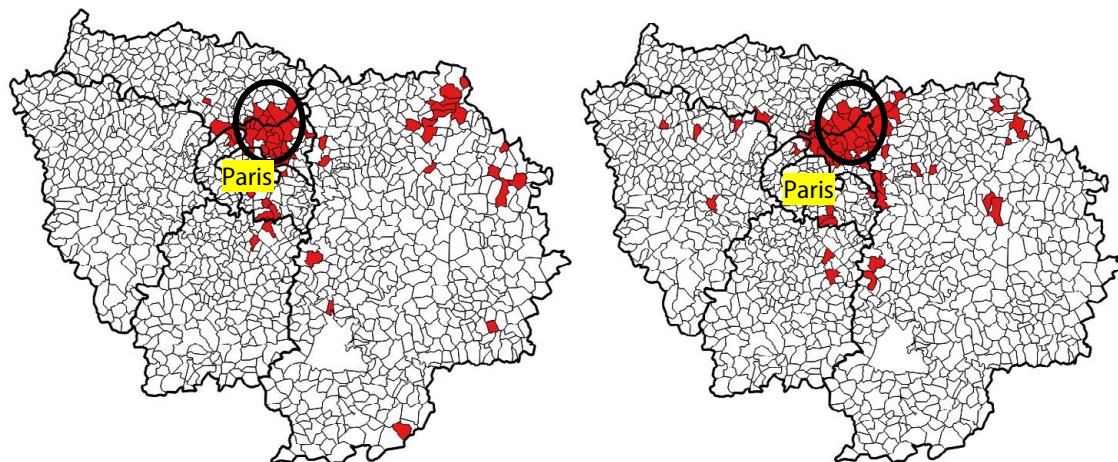


Figure 2: High cluster of abstention: A) 2007 French Presidential Elections and B) 2010 French Regional Elections.

| Variable | Non-spatial regression (OLS) | | SAR lag model regression | |
|--------------------------|------------------------------|--------------------|--------------------------|--------------------|
| | Presidential | Regional | Presidential | Regional |
| Constant | 13.52 ^{**} | 5.93 ^{**} | 9.73 ^{**} | 3.92 ^{**} |
| Socioeconomic-index (SE) | 0.76 [*] | 1.36 [*] | 0.61 ^{**} | 1.04 ^{**} |
| Rho (ρ) | / | / | 0.32 ^{**} | 0.42 ^{**} |
| R ² | 0.24 | 0.18 | 0.37 | 0.34 |
| AICc | 6438.25 | 8369.38 | 6056.21 | 8197.05 |
| I residual | 0.19 [*] | 0.21 [*] | 0.03 | 0.05 |

Rho (ρ) denotes the spatial autoregressive coefficients. R² (the percentage of variation explained) is not directly provided for spatial model and model fit is thus assessed with a pseudo-R² value calculated as the squared Pearson correlation between predicted and observed values (18). ^{*}Significance at 0.05; ^{**}Significance at 0.01.

Table 2: Regression analysis.

variance explained (R²) were 0.37 and 0.34; therefore, the above results implied a positive spatial correlation between the socioeconomic indicator and abstention French elections.

Table 3 reports the results of non-spatial and SAR lag models regressions according to Socioeconomic Status categories. We find the strongest significantly positive associations with abstention in categories 4 and 5, in the two OLS regressions for each French Election. The positive spatial autocorrelation detected in the residuals of the OLS model justifies the application of the SAR model. Categories 4 and 5 are always most strongly associated with Presidential elections and Regional elections and this relation is monotonic and linear, from the privileged to the deprived areas:

- Presidential elections β for category 2=0.60 and for β category 5=3.56
- Regional elections β for category 4=1.75 and β for category 5=6.26.

The percentages of variance explained (R²) by the SAR lag model, and significantly positive values for spatial effect indicating also that spatial regression model was successful in accounting for spatial correlation abstention. The spatial regression results showed that socioeconomic index is positively associated with abstention in the two-French elections, and this significant association is the strongest in the most deprived county. The introduction of the spatially lagged variable in the model makes it possible to control the presence of spatial autocorrelation (I residual presidential elections=-0.04 and I residual regional elections=0.06). Its inclusion in the SAR model improves significantly the quality of the adjustment compared to the

| | Non-spatial regression (OLS) | | SAR lag models regression | |
|----------------|------------------------------|--------------------|---------------------------|--------------------|
| | Presidential | Regional | Presidential | Regional |
| Constant | 10.90 ^{**} | 8.94 ^{**} | 6.44 ^{**} | 6.30 ^{**} |
| C2 | 0.91 [*] | 2.65 [*] | 0.60 [*] | 1.75 [*] |
| C3 | 1.50 [*] | 4.07 [*] | 1.12 [*] | 2.73 [*] |
| C4 | 2.91 [*] | 6.05 [*] | 2.28 ^{**} | 4.38 ^{**} |
| C5 | 4.50 [*] | 8.24 | 3.56 ^{**} | 6.26 ^{**} |
| Rho (ρ) | / | / | 0.33 ^{**} | 0.44 ^{**} |
| R ² | 0.22 | 0.17 | 0.29 | 0.31 |
| AICc | 6468.22 | 8383.62 | 6180.26 | 8216.46 |
| I residual | 0.16 [*] | 0.18 [*] | 0.04 | 0.06 |

Category 1 was used as the reference category. Rho (ρ) denotes the spatial autoregressive coefficients. R (the percentage of variation explained) is not directly provided for spatial model and model fit is thus assessed with a pseudo-R² value calculated as the squared Pearson correlation between predicted and observed values (18). ^{*}Significance at 0.05; ^{**}Significance at 0.01.

Table 3: Results of non-spatial and SLM regressions according to Socioeconomic Status categories.

OLS model (Akaike information criterion reduced from 6468.22 to 6180.26 for presidential elections and from 8383.62 to 8216.46 for the regional elections).

Discussion

In this study, we have sought to highlight the existence of inequalities in voter turnout spatially structured in the Île de France region. In other words, identify the share of inequalities of participation that stems directly from the specific socio-economic composition of the studied areas, to distinguish specific neighborhood effect. We did find highly

significant spatial autocorrelation coefficients across all of our spatial models. Further, our spatial diagnostic test (Moran's I) imply strong evidence of spatial autocorrelation with the both elections. These findings are a similar line of Darmofal [5] that identifies many theories in political science predicts spatial clustering of similar behaviors among neighboring regions.

Many auteurs claim that voting is a scholarly activity which takes place in a number of different contexts and through a range of mechanisms at a variety of different units [18-21]. Such contextual effects complement the living environment effects and can result from people interacting with their material environment, social networks [22]. In this work we used the FDep index although this index offers several advantages: it is one dimensional, maximizing the representation of the heterogeneity of its components and strongly associated with the components stratified in different urban criteria to better integrate the rural/urban gradient [11]. In order to approach social situations on the basis of geo-referenced information, we selected the FDep indicator due to the properties it offers: it is one-dimensional, maximizing the representation of the heterogeneity of its components and strongly associated with the components stratified in different urban criteria to better integrate the rural [10]. This indicator takes into account into account the socioeconomic context but do not take into account the context of the living environment, other indices are needed for a more detailed analysis.

In this study we used aggregated data at the county. This level of analysis is not the thinnest available. Indeed, the results of ANR Cartelec presented notably by Russo and Beauguitte [1] show that the scale processing polling station enables more robust results and finer. However, the use of such detailed data is not without cause other problems. The non-superposition of the electoral units and offices include using a ventilation procedure results that weakens matching. This study suffers also from potential issues of ecological inference [23]; i.e., problems of inferring individual-level behavior from aggregate (county level) data. Although, as noted by several recent studies point out, the problem of ecological fallacy is far less severe with county-level data as opposed to state-level data, and county-level data are the smallest spatial unit of analysis that allows for the inclusion of macroeconomic variables such as the unemployment rate.

Conclusions

Finally, our paper, in accord with other studies, sends a message to analysts who may want to use aggregated data to analyze the voting behavior. Accounting the spatial autocorrelation may produce better fit of model clearly improves and more robust conclusions.

References

- Luana R, Beauguitte L (2017) Aggregation Level Matters: Evidences from French Electoral Data.
- Burnett JW, Lacombe DJ (2012) The Review of Regional Studies Accounting for Spatial Autocorrelation in the 2004 Presidential Popular Vote: A Reassessment of the Evidence 42: 75-89.
- Kim J, Elliott E, Wang DM (2003) A spatial analysis of county-level outcomes in US Presidential elections: 1988-2000. *Elect Stud* 22: 741-61.
- Tobler WR (2014) A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ Geogr* 46: 234.
- Darmofal D (2006) *Spatial Econometrics and Political Science*.
- Kosfeld R, Eckey HF, Türck M (2007) LISA (Local Indicators of Spatial Association). *Wirtschaftswissenschaftliches Stud*.
- Kissling WD, Carl G (2007) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob Ecol Biogeogr*.
- Rivière J (2012) Vote et géographie des inégalités sociales: Paris et sa petite couronne.
- Rey G, Jouglu E, Fouillet A, Hémon D (2009) Ecological association between a deprivation index and mortality in France over the period 1997-2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 9: 33.
- Saib MS, Caudeville J, Beauchamp M, Carré F, Ganry O, et al. (2015) Building spatial composite indicators to analyze environmental health inequalities on a regional scale. *Environ Heal* 14: 68.
- Windenberger F, Rican S, Jouglu E, Rey G (2012) Spatio temporal association between deprivation and mortality: Trends in France during the nineties. *Eur J Public Health* 22: 347-353.
- Anselin L, Tedesco C, McTighe TJ, Austin C, Kitron U, et al. (2010) Local Indicators of Spatial Association-LISA. *Geogr Anal*.
- Anselin L (1995) Local Indicators of Spatial Association-LISA. *Geogr Anal*.
- Lacombe DJ, Shaughnessy TM (2007) Accounting for Spatial Error Correlation in the 2004 Presidential Popular Vote. *Public Financ Rev*.
- Anselin L (1988) Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geogr Anal*.
- Anselin L, Getis A (1992) Spatial statistical analysis and geographic information systems. *The Annals of Regional Science* 26: 19-33.
- Kissling WD, Carl G (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob Ecol Biogeogr*.
- Enelow JM, Hinich MJ (1984) The spatial theory of voting: an introduction.
- Johnston RJ, Pattie CJ (1998) Composition and context: Region and voting in Britain revisited during labour's 1990s' revival. *Geoforum*.
- John A, Françoise Subileau; Marie-France Toinet. Les chemins de l'abstention : Une comparaison franco-américaine. *Rev Fr Etud Am*. 58: 436-436.
- Johnston RJ, Pattie CJ (1995) The Impact of Spending on Party Constituency Campaigns at Recent British General Elections. *Party Polit*.
- Pattie C, Johnston R (1995) It's not like that round here: Region, economic evaluations and voting at the 1992 British General Election. *Eur J Polit Res*.
- Shively WP (1969) "Ecological" Inference: The Use of Aggregate Data to Study Individuals. *Am Polit Sci Rev* 63: 1183-1196.