Aggregation Level Matters: Evidences from French Electoral Data

Russo Luana, Laurent Beauguitte

Aggregation Level Matters: Evidences from French Electoral Data

Luana Russo*, Laurent Beauguitte**

July 2012

*Sciences-Po, Cevipof **UMR IDEES, Rouen

This paper benefited from the inputs of the project Cartelec members: L. Beauguitte, S. Bourdin, M. Bussi, B. Cautrès, C. Colange, S. Freire-Diaz, A. Jadot, J. Rivière, and L. Russo. For further information: www.cartelec.net

Abstract: This working paper, issued from the Cartelec project, raises a methodological issue regarding ecological inference. Studying abstention on two recent French elections in the Parisian agglomeration, a linear regression is launched on three different levels of aggregation from the finest one (polling station) until cities and electoral districts (*circonscriptions*). Tests on multicollinearity clearly indicate that employing the finest level of aggregation allows more variables being significant, enhances the explanatory power of the model, and improves collinearity amongst dependant variables.

Key-words: Abstention, Ecological Inference, Linear Regression Model, Multicollinearity, Parisian Agglomeration, Political Geography

1 Introduction

When dealing with ecological data the level of aggregation is a key point (Oumlil and Balloun 1998[10]). It is well known, at least since Robinson (1950[13]), that using different levels of aggregation leads to different results. In this paper, we will show that the finest the level of aggregation, the better the quality of the estimates and the less the multicollinearity in the model. Indeed, it was already demonstrated that, when using individual data, increasing the sample size is a viable remedy for

collinearity (Leahy, 2001[**?**]). In this case we do not use individual data nor a sample, but all the polling station available for a given area of the Parisian agglomeration. Hence, the number of cases is only involving the aggregation level employed, and the higher aggregation levels are only a sum of the lowest aggregation level available. This is a crucial difference with respect to the individual data, because in this case having more observation does not mean adding additional observations *per se*, but simply descending at a lowest aggregation level, and then indirectly having more information about the demographic, economical and social context.

In order to verify whether using different aggregation levels when dealing with ecological data leads to a difference in the final estimates, we basically use a linear regression model in which the dependent variable is the abstention and the independent ones are key variables according to both theoretical assumptions and previous empirical works (all variables are precisely described and justified below).

The paper proceeds as follows: we start with a very inclusive linear regression model that we apply only at the polling station level. Then, we test the multicollinearity and, by using a stepwise regression, we try to eliminate the multicollinearity problem. Once obtained a final, parsimonious and multicollinearity-free model, we apply the same model at the same data again, but this time we aggregate the data at two different levels: cities and electoral districts (see map 1 on the following page). We do this same procedure for two French elections: the first round of the 2007 Presidential Elections and the first round of the 2010 Regional Elections.

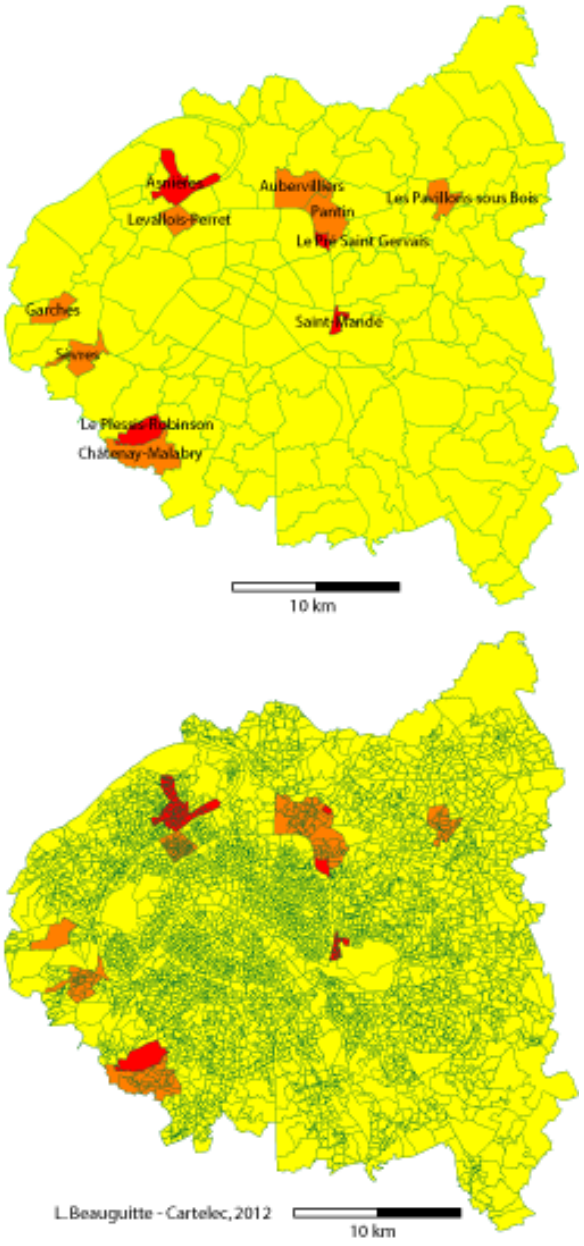
Since the goal of this paper is mainly methodological, we will not focus on the interpretation of the models in substantial terms. However, we try to build plausible and reliable models regarding abstention.

2 Method and Data

In order to verify if the level of aggregation was affecting the final estimates, we run the same linear regression model on three different aggregation levels for two elections: the 2007 French Presidential elections and the 2010 French Regional elections (first round in both cases). We choose in purpose a low (Régionales) and a high intensity election (Présidentielles) to check if the variables produce the same effect on abstention in these both different political configurations[1]. As we consider the first round in both cases, the political offer can be compared without any bias as all political parties, from extreme left to extreme right, are present.

[1]The abstention, measured in percentage of the *inscrits* was 16.22% (national level) in 2007 *vs* 53.64 in 2010. For the space studied here, the gap is even higher (14% in 2007 and 55.79 in 2010).

Figure 1: Cities - top - and polling stations in *la petite couronne*



In red: cities and polling stations without any data for 2007 and 2010; in orange, cities and polling stations without electoral data for 2007.

The data are referring to the area around Paris called *la petite couronne*, which includes four departments: the 75 (Paris), the 92 (Hauts-de-Seine), the 93 (Seine-Saint-Denis), and the 94 (Val-de-Marne). These four departments are part of the Parisian agglomeration and, even if it presents strong contrasts (the 92 being the wealthiest department in France while the 93 is one of the poorest), this urban system forms in itself a relevant unit of analysis, especially for electoral and socio-economical works (Rivière, 2012[12]).

We will employ a weighted regression were the weight will be the voters who have the right to vote per each polling station (*inscrits*[1]). The weighted least squares regression has the advantage of taking into account the behaviour of the random errors in the model. In fact, by optimizing the weighted fitting criterion, it is possible to find the parameter estimates. Hence, the parameter estimates allow the weights to determine the contribution of each observation to the final parameter estimates.

With regard to the R-squared (R^2), we will report only the adjusted R^2. In fact, while the R^2 increases when a new variable is added, the adjusted R^2 increases only if the new term improves the model more than would be expected by chance. This feature will be crucial when determining which regression model to employ.

The electoral data are provided at the polling station level by the Minister of the Interior, and in this specific case they consist in the absolute numbers of the abstention per polling station. These numbers were converted in percentages (obtained on the base of the number of members of the polling station, the *inscrits*) in order to make the analysis clearer, more readable and effective.

The independent variables[2] firstly regard the age of the population divided in three classes (18-24, 25-65 and more than 65 – thereafter named Young, Working age and Old). Several studies, starting with Lancelot in 1968[5], showed that old adults were more reluctant than young ones to abstain. So we expect to get a higher level of abstention where the rate of 18-24 years is higher (which of course does not mean that a relation on an aggregate level is true at the individual one).

We tested the importance of variables regarding nationality, even if only French citizens are able to vote. We considered three variables: percentage of French persons, percentage of foreigners (people living in *la petite couronne* but not owning French nationality), and finally percentage of immigrants (people born abroad but which now get French nationality).

[1]Being *inscrit* on electoral list is mandatory to be able to vote. The inscription is theoretically an obligation for all French citizens, but the constraint has never being used, and the non inscription rate is estimated around 7 to 9% of the French adult population (Braconnier and Dormagen, 2007a[1]).

[2]Source: INSEE National census, 2007.

Socio-economic variables used are from three different types: percentage of unemployed persons, level of education (no diploma, low level of diploma, high level of diploma[1]) and finally housing status (owners *vs* renters – the category free housing was introduced as a control variable). The literature regarding abstention generally highlights the relation between a high level of abstention and a low level of diploma, a high level of unemployment and the rate of renters [2][14].

In all cases, the absolute number where converted in percentages obtained on the base of the total amount of the population in each polling station.

The following list sums up the dependant variables used for the different models

- Age:

  young: 18-24 years old;

  working age: 25-64 years old;

  old: 65 onward years old.

- National origin:

  French;

  strangers;

  immigrants.

- Occupational status:

  unemployed.

- Education:

  any diploma;

  from CAP to BEP (secondary education);

  all types of BAC (end of the secondary education /university).

- Housing:

  owners;

  renters;

  free housing.

[1]For people knowing the French educational system: low level means diploma from CAP to BEP and high level - the name being a little abusive for the youngest generations - means *Baccalauréat* and above.

Starting from the polling station level (which is the smallest level of aggregation for electoral ecological data), we aggregated the data at other two levels: cities and electoral districts (*circonscriptions législatives* in French). Table 1 summarizes the number of observations for each level.

Table 1: Number of observations per aggregation level.

	Polling stations	**Cities**	**Electoral districts**
Presidential Election 2007	3089	124	59
Regional Election 2010	3268	136	59

As you can notice, the number of observations is different with regard the polling station and the city levels for the two elections considered. Actually, for both the elections, the real number of the polling stations was 3326, but it was necessary to exclude some cities from the analysis because some of them had a re-design of the polling stations and it was not possible to match the electoral and the socio-demographic data. As you can see, the problem was not severe, as for the Presidential Election we were able to maintain the 92.87% of the polling stations, and for the Regional Elections elections the 98.25%.

With regard to the city level, since Paris was significantly bigger than the other cities in the dataset[1], following the already administrative division, we maintained the twenty smaller areas called *arrondissements*. Then, in the dataset we kept each *arrondissement* as a separate observation.

Since we are dealing with ecological data, it is useful to underline how to interpret the dataset. Each unit of observation is a polling station, and for each variable we have a percentage. This percentage is the total amount of a given variable in a given polling station.

3 Findings

In order to build a reliable and effective model, we first checked the correlations (see appendix). We noticed that some of the independent variables were medium/highly correlated among them. Since this might cause some problem with the multicollinearity (Lewis-Beck 1980[7]; Upton and Cook 2008[15]), we will use a specific diagnostic to check the multicollinearity and try to keep it under control.

[1]In 2008, Paris counts 2 590 000 inhabitants and the second largest city of *la petite couronne* is Boulogne-Billancourt with 109 000 inhabitants only. The largest Parisian *arrondissement* regarding population is the 15^{th} with 238 000 persons.

We started with a very inclusive regression model, with all the thirteen variables selected. The model employed is a weighted linear regression (in which the weights are the *inscrits*). Tables 2 and 3 show the first model applied to both the elections considered.

Table 2: First model: observations and adjusted R^2.

Election	**Obs.**	**Adj.** R^2
Presidential Election 2007	3089	0.5011
Regional Election 2010	3268	0.2645

Table 3: First model: standardized β coefficients.

Variables	Std. β coeff.		Std. Err.	
(ID: Abstention)	PE	RE	PE	RE
Young	-.0261346	-.0617661**	.0236746	.0831016
Working age	-.0080677	-.3182642***	.0219815	.0794442
Old	.0935859**	-.2617109***	.0195017	.0696243
French	.0935859	-.4335976	.6670382	2.403678
Strangers	.-.6190991	-.2436914	.6675262	2.405822
Immigrants	-.0302352	-.0581329	.0242692	.0868779
Unemployed	.1655555***	.0841649*	.0417229	.1514527
No diploma	.1507885*	.2397054**	.0280403	.1006813
CAP/BEP	.1150329*	.3921551***	.0354464	.1275432
BAC (all types)	.1150329*	.3941463**	.0226946	.0818271
Owners	.1438963	.0389029	.0136184	.0490349
Renters	.1735346*	.0319292	.0138451	.0497786
Free housing	0528821*	0585268*	.0213972	.0760094

****= $t \geq 4$; **= $t \geq 3$; *= $t \geq 1.96$*

weights: inscrits

As you can see, this model is more effective in terms of explained variance when applied to the 2007 Presidential Election 2007. Indeed, in the 2010 Regional Election the abstention rate hits 53.64 percent[1] in other words, it seems that for the 2010 election, the abstention has different dynamics that this model does not capture.

[1]In France, if we consider all elections on a national scale, excluding referendums, the highest rate of abstention was in the 2009 Européennes election with 59.37% of the *inscrits*.

In order to check the multicollinearity for this model, we employ the "estat vif" procedure in STATA (Table 4).

As Hamilton (2009) explains: "1/VIF (variance inflation factor) (or $1/R^2$) tells us what proportion of an x variable's variance is independent of all the other x variables. A low proportions [...] indicates potential trouble".

Table 4: First model: multicollinearity

Variables	**VIF**		**1/VIF**	
	PE	RE	PE	RE
Strangers	10844.34	11417.53	0.000092	0.000088
French	10819.49	11391.43	0.000092	0.000088
BAC (all types)	46.42	47.57	0.021540	0.021020
Renters	43.16	43.35	0.023168	0.023068
Owners	41.69	42.01	0.023986	0.023803
Immigrants	18.30	19.33	0.054636	0.051746
No diploma	17.82	18.59	0.056113	0.053798
CAP/BEP	11.21	11.17	0.089190	0.089527
Working age	4.69	4.74	0.213166	0.210754
Unemployed	4.37	4.54	0.229086	0.220261
Old	3.70	3.67	0.270191	0.272833
Free housing	2.40	2.34	0.416697	0.426763
Young	1.66	1.70	0.600729	0.588033
Mean VIF	1681.48	1769.84		

Chatterjee *et al.* (2000[3]) suggest that there is multicollinearity when
"*the largest VIF is greater than 10; or*
the mean VIF is considerably larger than 1."

With our largest VIF higher than 10000 and the mean VIF higher than 1600 in both the elections, we meet both criteria.

In order to fix the problem of the multicollinearity, there are two main possible solutions. We can either try to fix the multicollinearity by employing the "centring" procedure (see Hamilton 2009, p. 226), or we can decide to eliminate some variables.

To get some indications on which variables need to be eliminated, we decided to use the step-wise procedure. Table 5 shows the variables that the procedure suggests to maintain and to drop. Table 5 summarizes the results of the step-wise regressions on both the elections (*PE* stands for 2007 Presidential Election and *RE* stands for 2010 Regional Election).

Table 5: Step-wise regression

Variables	Coefficients		Std. Err.		P≥t		Removing.	
(ID: Abstention)	PE	RE	PE	RE	PE	RE	PE	RE
Young	–	-.246053	–	.06564	0.050	0.000	YES	NO
Working age	–	-.7187066	–	.0683442	0.050	0.000	YES	NO
Old	.0813572	-.6389446	.0130077	.0589315	0.000	0.000	NO	NO
French	-.0699387	–	.0138716	–	0.050	0.791	NO	YES
Strangers	–	.1523258	–	.0416564	0.050	0.000	YES	NO
Immigrants	–	–	–	–	0.050	0.372	YES	YES
Unemployed	.2550422	.4125384	.0418275	.1157916	0.000	0.000	NO	NO
No diploma	.0945562	.4533225	.0244067	.0870589	0.000	0.000	NO	NO
CAP/BEP	.099413	.7758603	.0256073	.1069413	0.000	0.000	NO	NO
BAC (all types)	-.0791843	.2587817	.0134074	.0703127	0.000	0.000	NO	NO
Owners	.0267003	–	.0128846	–	0.038	0.050	NO	YES
Renters	.034293	–	.0132596	–	0.010	0.050	NO	YES
Free housing	.0632967	.1535383	.0214196	0505448	0.003	0.002	NO	NO

PE: obs = 3089; Adj R^2 *= 0.5136*

RE: obs = 3268; Adj R^2 *= 0.3530*

Let's read the step-wise regression carefully: what this procedure suggests is basically taking out two out of three variables from one category (age and housing in the case of the Presidential Election, and national origin and housing for the Regional Election).

This result was predictable. In fact, we have three different variables for each category (except for "work", which only has the variable "Unemployment"). Since these are ecological data expressed in percentages, to keep just one of the variables in a given category means that the other two variables make up the remaining proportion that the chosen variable does not cover. That is, if in a given polling station the percentage of the owners in the population is 60%, it means that the remaining 40% is composed by renters and free housing.

Hence, when deciding which variable to maintain for each category, it seems more convenient to choose either the more significant or one which was an outer.

Then, for the national origin we keep French, since in this way we'll have kind of a dichotomy French/not French (and we would not have this advantage by using one of the two other categories) [1] since the *immigrants* are part of the *French*: mostly, they were born in another country and then became French citizen..

The two outer variables for "education" are "no diploma" or "BAC(s)" (end of the secondary education /university), and we decided it was indeed better to take "BAC(s)" for two reasons: 1) when looking at our regression model, the β coefficients for this variable is the highest, 2) the percentage of population with a BAC (or more) was higher than the percentage without diploma (about 38% *versus* 12%) [2].

With regard to the "housing", we decided to keep the variable "owners" over "renters" and "free housing", because of the aforementioned advantage: by keeping the variable "owners", we have a net distinction in any given area between the percentage of people who own their house and the ones who do not [3].

Finally, we decided to take all the three variables for the category "age", since in this case the middle category "Working age" is the more inclusive in terms of years, and also the most numerous. Furthermore, this is the only demographic variable we have [4].

Tables 6 and 7 show the second regression model, which is a more parsimonious

[1] The percentage of *French* in the population is 84.84, the *strangers* are the 15.16%, and the *immigrants* the 20.48%. The total is more than 100%

[2] The percentage of *no diploma* in the population is 13.01, the percentage of people holding a title form a CAP to a BEP are 18.91, and the ones olding a BAC or more are the 38.20%. All the remaining population out of this percentages is still studying.

[3] The percentage of *owners* in the population is 17.45, the *renters* are 25.73%, and the people benefiting of the *free housing* the 1.81%.

[4] The percentage of *young* in the population is 9.86, the people in their *working age* are the 55.60%, and the seniors, that is the *old* the 12.84%.

one, and it includes only one variable per category, which the only exception of the "age". This model includes only seven variables, while the first one included thirteen variables.

Table 6: Second model: observations and adjusted R^2

Election	**Obs**	**Adj** R^2
Presidential Election 2007	3089	0.4989
Regional Election 2010	3268	0.2496

Table 7: Second model: standardized β coefficients and standard error

Variables	Coefficients		Std. Err.	
(ID: Abstention)	PE	RE	PE	RE
Young	-.0424543*	-.0685306**	.0216045	.0753659
Working age	.0409709*	-.1482356***	.0158999	.0580375
Old	.131871***	-.1428987***	.0154836	.0556107
French	-.1524588***	-.0910477**	.0097089	.0346454
Unemployed	.1834103***	.0649041*	.0383116	.1399367
BAC (all types)	-.5578792***	-.2789387***	.0063275	.0228775
Owners	-.017734	.0400247*	.0027952	.0101924

****= $t \geq 4$; **= $t \geq 3$; *= $t \geq 1.96$*
weights: inscrits

As you can notice, with six variables less, the adjusted R^2 does not drop that much. In fat it is only 0.0022 less with regard to the Presidential Election, and 0.0149 less with regard to the Regional Election. Furthermore, when applying the model to the Regional Election, all the variables are significant. When considering the Presidential Election, the only non significant variable is "owners".

In order to definitively accept this model and test it on the other two aggregation levels, we have to check whether there is still a multicollinearity problem.

Table 11 shows the multicollinearity diagnostic for the second regression model.

Table 8: Second model: multicollinearity

Variables	**VIF**		**1/VIF**	
	PE	RE	PE	RE
Unemployed	3.68	3.84	0.271595	0.260290
BAC (all types)	3.51	3.54	0.284638	0.282233
French	2.48	2.56	0.403516	0.390102
Working age	2.43	2.46	0.411882	0.406643
Old	2.31	2.27	0.433036	0.440716
Owners	1.69	1.71	0.591151	0.585970
Young	1.39	1.39	0.717100	0.719406
Mean VIF	2.50	2.54		

As the Table 11 demonstrates, by selecting one variable for each category (with the exception of "age") we meet the criteria suggested by Chatterjee et al (2000[3]), since even if our mean VIF is slightly larger than 1, the largest VIF is not greater than 10.

3.1 City and Electoral Districts aggregation levels

In order to test whether the aggregation level affects the quality of the analysis, we test the second regression model on the same data, but using two different aggregation levels: cities and electoral districts.

With regard to the city level, as showed in Table 1, we have 124 unities for the Presidential Election and 137 for the Regional one. For both elections, 20 of those units represent the twenty smaller areas (*arrondissements*) in which Paris is divided.

Tables 9 and 10 show the linear regression model applied to the city level.

Table 9: Second model, cities: observations and adjusted R^2

Election	**Obs**	**Adj** R^2
Presidential Election 2007	124	0.7000
Regional Election 2010	136	0.8003

Table 10: City level: standardized β coefficients and standard error.

Variables	Coefficients		Std. Err.	
(ID: Abstention)	PE	RE	PE	RE
Young	-.1205111*	-.1508704*	.1064167	.178327
Working age	-.0209643	-.3545781**	.0838745	.1406654
Old	.0609746	-.2686493**	.0962709	.1584229
French	-.313109*	-.2835766**	.054503	.0902889
Unemployed	.0522401	.1456201	.2432053	.4178619
BAC (all types)	-.6178728**	-.237419*	.0329872	.0551368
Owners	-.0572686	.044503	.0161465	.0278532

****= $t \geq 4$; **= $t \geq 3$; *= $t \geq 1.96$*
weights: inscrits

As you can see, for both elections, the model has a high adjusted R^2 value. However, only three variables are significant for the Presidential Election ("French" and "BAC", and one variable from the "age" category). With regard to the Regional Election five variables are significant, but the adjusted R^2 is even higher. These are usually signs of the possible presence of multicollinearity. In fact, it was shown that when multicollinearity occurs, the variances are large and thus far from the true value (Morris 1982[9]; Pagel and Lunneberg 1985[11]; Mooney and Duval 1993[8]).

Table 11 shows the diagnostic for the multicollinearity applying the regression model at the city level.

Table 11: City level: multicollinearity

Variables	**VIF**		**1/VIF**	
	PE	RE	PE	RE
Unemployed	10.49	11.30	0.095365	0.088493
BAC (all types)	6.74	5.33	0.148320	0.187692
French	5.35	5.50	0.186881	0.181974
Working age	3.39	2.61	0.294693	0.382832
Owners	3.36	3.49	0.297212	0.286929
Old	3.11	2.70	0.321340	0.370344
Young	1.50	1.46	0.668535	0.684407
Mean VIF	4.85	4.63		

The diagnostic shows that at the city level there is a problem with multicollinearity. The problem is not that severe, but in both cases the biggest VIF is larger than 10 and the mean VIF is almost doubled with respect to the polling station level.

This outcome was indeed predictable, since it is well known that a remedy to the multicollinearity problem is the increase of the number of the observations. However, this usually applies to individual data and specifically refers to the increasing of the sample. In this case, we are not dealing with a sample but with the entire population available for the area considered (the *petite couronne*), and the data analysed are exactly the same than at the previous level: the only difference is that they have been aggregated differently, that is at a higher level (cities).

Finally, we apply the linear regression model at the electoral district level (tables 12 and 13).

Table 12: Second model, electoral districts: observations and adjusted R^2

Election	**Obs**	**Adj** R^2
Presidential Election 2007	59	0.8240
Regional Election 2010	59	0.8490

Table 13: Electoral districts level: standardized β coefficients and standard error

Variables	Coefficients		Std. Err.	
(ID: Abstention)	PE	RE	PE	RE
Young	-.1522416*	-.1471945*	.1243341	.2556732
Working age	-.0130935	-.2113426	.1115796	.235215
Old	.1017578	-.1214201	.1285218	.2628305
French	-.3678302*	-.3531195*	.0669929	.1368121
Unemployed	.0904198	.0671718	.293413	.6330592
BAC (all types)	-.6500518**	-.4319581*	.0426144	.0898281
Owners	-.0520452	.0376641	.0193139	.0417978

****= $t \geq 4$; **= $t \geq 3$; *= $t \geq 1.96$*

weights: inscrits

As you can see, running the regression model at the electoral district level, the adjusted R^2 increases even more, as the significant variables decrease: only three per

elections (the same in both elections), and with lower levels of significance. This suggests that the multicollinearity could be even increased.

Furthermore, the model looses more explanatory power.

Table 14 shows the multicollinearity diagnostic at the electoral district level.

Table 14: Electoral district level: multicollinearity

Variables	**VIF**		**1/VIF**	
	PE	RE	PE	RE
BAC (all types)	14.32	15.40	0.069812	0.064933
Unemployed	13.07	14.95	0.076504	0.066888
Old	7.34	6.96	0.136323	0.143775
Working age	7.25	7.64	0.137885	0.130924
French	6.73	6.97	0.148630	0.143407
Owners	2.71	2.84	0.368551	0.351632
Young	1.73	1.70	0.578233	0.589295
Mean VIF	7.59	8.07		

With respect to the city level, the multicollinearity is increased, both with regard to the larger VIF and the mean VIF, showing that the more aggregated are the data the more the collinearity problem increases.

4 Conclusions

The main goal of this paper was to show that when using ecological data, the higher the aggregation level employed, the lower the explanatory power of the regression model.

In order to demonstrate this, we took two elections (first round of the 2007 Presidential Election and first round of the 2010 Regional Election) and we applied the same model at the *petite couronne* (Departments 75, 92, 93, 94), on the same data aggregated at three different levels.

The higher levels (cities and electoral districts) are obtained by aggregating the polling station data. Upon the successful design of a linear regression model at the polling station level (high explanatory power and absence of multicollinearity), we applied that same model on the city and electoral district levels.

We found out the the higher the aggregation level, the more the multicollinearity and the less the explanatory power of the model. Hence, we can conclude that the

aggregation level has an impact on the quality of the estimates and on the explanatory power of the model.

References

[1] Braconnier, C., and Dormagen, J.-Y.: Ségrégation sociale et ségrégation politique, *Non inscrits, mal inscrits et abstentionnistes*, Centre d'analyse stratégique, 11, 6-61 (2007a).

[2] Braconnier, C. and Dormagen, J.Y.: *La démocratie de l'abstention. Aux origines de la démobilisation électorale en milieu populaire*. Gallimard, Folio actuel, Paris (2007b).

[3] Chatterjee, S., Hadi, A.S., and Price, B.: *Regression Analysis by Example*. John Wiley and Sons, New York (2000).

[4] Hamilton, L.C.: *Statistics with STATA*. Books/Cole, Canada (2009).

[5] Lancelot A.: *L'abstentionnisme électoral en France*. Paris, Armand Colin (1968).

[6] Leahy K.: Multicollinearity: when the solution is the problem. In O.Parr Rud (Eds) *Data Mining Cookbook*, p. 106-108, John Wiley and Sons, Inc, New York (2001).

[7] Lewis-beck, M.S.: *Applied Regression. An introduction*. Sage Publications, California (1980).

[8] Mooney, C.Z., Duval, R.D.: *Bootstrapping: A nonparametric approach to statistical inference*. Sage Publications, Newbury Park, CA (1993).

[9] Morris, J. D.: Ridge regression and some alternative weighting techniques: A comment on Darlington. *Psychological Bulletin*, 91, 203-210 (1982).

[10] Oumlil A.B., Balloun J.L.: Levels of Aggregation: A Conceptual Model. *Quality & Quantity*, 32, 109-117 (1998).

[11] Pagel, M.D., Lunneberg, C.E.: Empirical evaluation of ridge regression. *Psychological Bulletin*, 97,342-355 (1985).

[12] Rivière, J.: Vote et géographie des inégalités sociales : Paris et sa petite couronne, *Métropolitiques*, http://www.metropolitiques.eu/Vote-et-geographie-des-inegalites.html (2012).

[13] Robinson W.S.: Ecological Correlations and the Behaviour of Individuals. *American Sociological Review*, 15(3), p.351-357 (1950).

[14] Subileau F. and Toinet M.-F.: *Les chemins de l'abstention : une comparaison franco-américaine*. Paris, La Découverte (1993).

[15] Upton, G., Cook, I.: *Oxford Dictionary of Statistics*. Oxford University Press, New York (2008).

Table 15: Correlation table PE 2007 (obs=3089)

	Young	Working age	Old	French	Strangers	Immigrants	Unemployed	No diploma	CAP/BEP	BAC	Owners	Renters	Free housing
Young	1.00												
Working age	-0.13	1.00											
Old	-0.28	-0.25	1.00										
French	-0.32	0.08	0.43	1.00									
Strangers	0.32	-0.08	-0.43	**-1.00**	1.00								
Immigrants	0.33	-0.12	-0.45	**-0.96**	**0.96**	1.00							
Unemployed	0.22	-0.02	**-0.52**	**-0.75**	**0.75**	**0.78**	1.00						
No diploma	0.14	-0.32	-0.42	**-0.78**	**0.78**	**0.81**	**0.77**	1.00					
CAP/BEP	-0.13	-0.30	-0.25	-0.08	0.08	0.13	0.24	**0.52**	1.00				
BAC	-0.11	0.48	0.45	**0.52**	**-0.52**	**-0.58**	**-0.60**	**-0.87**	**-0.80**	1.00			
Owners	-0.33	-0.03	0.34	0.43	-0.43	**-0.50**	**-0.59**	-0.41	0.04	0.26	1.00		
Renters	0.25	0.01	-0.38	-0.42	0.42	0.49	**0.63**	0.45	0.04	-0.33	**-0.97**	1.00	
Free housing	0.14	0.20	0.21	0.07	-0.07	-0.12	-0.21	-0.32	-0.47	0.47	-0.01	-0.16	1.00
Abstention (%)	0.09	-0.26	-0.29	**-0.53**	**0.53**	**0.56**	**0.57**	**0.69**	0.46	**-0.67**	-0.30	0.34	-0.25

Table 16: Correlation table REG 2010 (obs=3268)

	Young	Working age	Old	French	Strangers	Immigrants	Unemployed	No diploma	CAP/BEP	BAC	Owners	Renters	Free housing
Young	1.00												
Working age	-0.16	1.00											
Old	-0.26	-0.26	1.00										
French	-0.30	0.08	0.43	1.00									
Strangers	0.30	-0.08	-0.43	-1.00	1.00								
Immigrants	0.31	-0.12	-0.45	**-0.96**	**0.96**	1.00							
Unemployed	0.21	-0.03	-0.52	**-0.76**	**0.76**	**0.79**	1.00						
No diploma	0.14	-0.32	-0.41	**-0.79**	**0.79**	**0.82**	**0.78**	1.00					
CAP/BEP	-0.11	-0.31	-0.24	-0.09	0.09	0.15	0.25	**0.52**	1.00				
BAC	-0.13	0.48	0.44	**0.54**	**-0.54**	**-0.59**	**-0.61**	**-0.88**	**-0.80**	1.00			
Owners	-0.33	-0.02	0.35	0.44	-0.44	**-0.50**	**-0.60**	-0.43	0.02	0.28	1.00		
Renters	0.24	0.01	-0.39	-0.42	0.42	**0.50**	**0.64**	0.46	0.06	-0.35	**-0.97**	1.00	
Free housing	0.13	0.20	0.20	0.07	-0.07	-0.12	-0.20	-0.31	-0.46	0.45	-0.01	-0.16	1.00
Abstention (%)	0.09	-0.27	-0.33	-0.40	0.40	0.43	0.43	**0.54**	0.40	**-0.56**	-0.21	0.24	-0.20