



HAL
open science

Analyse écologique, modèles multi-niveaux et sociologie électorale

Joël Gombin

► **To cite this version:**

Joël Gombin. Analyse écologique, modèles multi-niveaux et sociologie électorale. Congrès de l'Association française de science politique, session commune aux ST 5 : " Etudier les comportements électoraux. Bilan de la recherche francophone " et ST 13 : " Variables, Individus, Contextes. Comment observer et analyser leurs interactions ? ", Sep 2009, Grenoble, France. <halshs-00422081>

HAL Id: halshs-00422081

<https://shs.hal.science/halshs-00422081v1>

Submitted on 19 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Congrès AFSP 2009

ST 5 : « Etudier les comportements électoraux : bilan de la recherche francophone »

ST 13 : « Variables, individus, contextes.

Comment observer et analyser leurs interactions ? »

Session de travail : Individus et contextes (approche écologique)

Joël Gombin

Université de Picardie-Jules Verne – CURAPP (UMR 6054)

joel.gombin@u-picardie.fr

Analyse écologique, modèles multi-niveaux et sociologie électorale : l'exemple des votes pour le Front national

Depuis son émergence électorale en 1983-1984, le Front national ne cesse d'intriguer de nombreux chercheurs en sciences sociales, qu'ils soient politistes ou sociologues, géographes, historiens ou encore anthropologues. Parmi les nombreuses questions que le « phénomène Le Pen » soulève, et qui restent débattues, figure en bonne place celle de savoir comment appréhender le vote en faveur du Front national. Des déterminants individuels, socioéconomiques ou psychosociologiques, suffisent-ils à rendre compte de ce comportement électoral ? La forte spatialisation du vote en faveur du Front national n'invite-t-elle pas à prendre en compte des facteurs non-individuels du vote ? Les difficultés rencontrées par les enquêtes d'opinion à rendre compte du vote FN n'appellent-elles pas une approche écologique de ce phénomène ?

Cette communication propose une analyse du vote en faveur du Front national aux élections présidentielles de 1995, 2002 et 2007 (soit quatre tours de scrutins) au moyen d'une méthodologie alliant analyse écologique et modélisation multiniveau, qui permet de tenir compte du caractère fortement spatialisé du vote FN tout en mettant en évidence les fondements sociaux des électorats frontistes – l'emploi du pluriel est volontaire, puisque notre méthodologie met l'accent sur l'hétérogénéité dans l'espace, mais aussi le temps, de ces électorats. En d'autres termes, l'hypothèse ici défendue est que les logiques sociales au fondement des votes frontistes pourront d'autant mieux être mises en évidence qu'on acceptera qu'elles puissent varier dans l'espace et le temps.

On commencera par expliciter ce qu'il faut entendre par analyse écologique et modélisation multiniveau et en quoi leur articulation peut être pertinente pour rendre compte des votes en faveur du Front national. Puis, après avoir présenté les données mobilisées et le modèle construit, on présentera les résultats obtenus. Enfin, on discutera ces résultats et la méthodologie employée.

Analyse écologique et modélisation multiniveau

Les deux notions qui sont abordées ici sont inégalement familières au politiste en général et à l'électoraliste en particulier. La première fait figure de grand classique de la science politique, tandis que la seconde évoque davantage une méthode statistique moderne.

La notion d'analyse écologique est mal définie et est utilisée pour renvoyer à des méthodes, des raisonnements et des pratiques très différents. Le qualificatif d'écologique a de plus parfois été utilisé pour caractériser non pas une méthode, mais un ensemble d'approches théoriques en sociologie électorale¹ - sans même évoquer ici l'Ecole de Chicago et l'écologie urbaine². De plus, elle est parfois

¹ Nonna Mayer, Pascal Perrineau, *Les comportements politiques*, Paris, Armand Colin, 1992.

² Yves Grafmeyer, Isaac Joseph, *L'école de Chicago. Naissance de l'écologie urbaine*, Paris, Flammarion, 2004.

confondue avec des notions voisines mais distinctes, comme celle d'analyse contextuelle³. Nous considérons ici que l'analyse écologique consiste à mettre en relation entre elles des données *caractérisant un collectif territorial* – par opposition à des données caractérisant des individus. La mise en relation de ces données peut prendre différentes formes : rapprochement de cartes, à la Siegfried ; utilisation d'outils statistiques simples, comme le coefficient de corrélation, ou plus complexes, qu'ils renvoient à une notion de causalité – régression linéaire, simple ou multiple – ou qu'ils s'en écartent – analyse géométrique des données (ACP, ACM). Les données en question peuvent être de différentes sortes, et soit caractériser le territoire *en tant que tel* – sa superficie, sa densité... – soit résulter de *l'agrégation de caractéristiques des individus* peuplant ce territoire – ainsi des résultats électoraux observés sur un territoire. Il convient d'insister sur le point suivant : la définition que nous donnons de l'analyse écologique, et les distinctions que nous opérons, sont strictement méthodologiques et non ontologiques. En d'autres termes, l'analyse écologique ne se caractérise, à notre sens, que par le niveau de recueil des données utilisées. Utiliser des données agrégées n'implique pas nécessairement de recourir à des modèles explicatifs contextuels ou environnementaux, de même que le recours à ce type de données n'est pas absolument essentiel pour étayer ce type de théories. Il n'en reste pas moins que, dans la littérature, il existe des affinités électives entre le type de données mobilisées et les modèles explicatifs privilégiés, et qu'on ne peut vérifier n'importe quelle hypothèse avec n'importe quelles données ; mais ces affinités électives ne sont pas des liens mécaniques. Les discussions, souvent passionnées, autour des mérites et des défauts de l'analyse écologique souffrent trop souvent de confusions entre le type de données utilisées, le type de méthode d'analyse mobilisée et le type de schéma explicatif proposé. Nous plaçons pour que ces trois éléments soient distingués et fassent l'objet de discussions distinctes.

Signalons d'emblée que l'analyse écologique, entendue en ce sens, a longtemps été dominante, voire exclusive, en géographie puis en sociologie électorales, pour une raison simple : l'accès à des données individuelles fut longtemps impossible ; puis, après l'apparition des sondages d'opinion, il demeura coûteux. Dès lors, les électoralistes n'eurent pendant longtemps d'autre choix que d'utiliser des données agrégées de diverses sortes et de les rapprocher entre elles, de manière plus ou moins sophistiquée. L'école dite française de sociologie électorale, de Siegfried à Dogan en passant par Klatzmann ou Goguel, repose essentiellement sur l'analyse écologique⁴, mais cette dernière n'était pas inconnue à l'étranger non plus, comme en témoignent les travaux de Herbert Tingsten⁵, politologue suédois à l'origine du terme de « comportement politique ».

On notera également que, entendue en ce sens, l'analyse écologique ne pose pas de problème particulier, dès lors que l'analyste s'en tient à des énoncés portant sur les collectifs territoriaux qu'il étudie. La controverse sur la valeur de l'analyse écologique renvoie, non pas au recours à des données agrégées, mais à la prétention d'inférer de ces énoncés sur des collectifs des énoncés sur les individus qui composent ces collectifs – ce qu'on appelle souvent l'inférence écologique⁶.

³ Sur cette dernière notion, voir Céline Braconnier, *Comprendre les comportements électoraux par les approches environnementales*, Habilitation à Diriger des Recherches, dir. Yves Déloye, Université de Cergy-Pontoise, 2009.

⁴ On notera au passage que les raisons qui justifiaient le recours continué à des données écologiques après que le recours aux données issues de sondages soit devenu courant sont exactement les mêmes que celles qui incitent à utiliser ce type de données concernant le Front national : hétérogénéité géographique et sous-déclaration systématique de ce vote dans les enquêtes d'opinion. « The regional diversity of France and Italy from the economic, social, religious, cultural, and political viewpoint facilitates the analysis of the social composition of the various parties by means of the ecological method [...] The ecological method seems particularly relevant for France and Italy, since opinion surveys are carried out there under much more difficult conditions than in countries like the United States or Great Britain. Many factors such as the greater regional diversity, multiparty systems, a peasant class that is in many respects psychologically impenetrable, and the reluctance of some voters, particularly Communists, to indicate their voting choice, make polling very difficult in France and Italy », Mattei Dogan, « Political Cleavage and Social Stratification in France and Italy », p. 130-131, in Seymour M. Lipset, Stein Rokkan (dir.), *Party Systems and Voter Alignments: Cross-National Perspectives*, New York, The Free Press, 1967, p. 129-195.

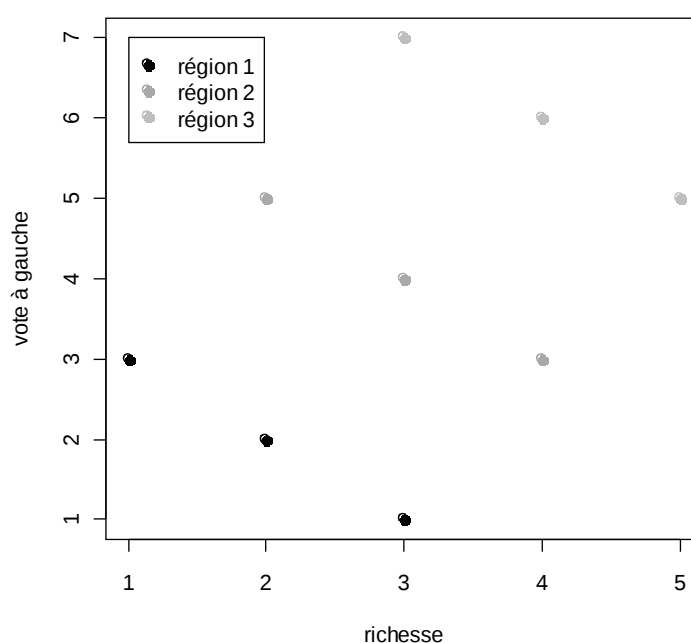
⁵ Herbert Tingsten, *Political behavior : studies in election statistics*, Totowa, N.J., Bedminster Press, 1963 (première édition : Londres, 1937).

⁶ On considère généralement que la référence séminale concernant la critique de l'analyse écologique, ou plus exactement de l'inférence écologique, est : W. S. Robinson, « Ecological correlations and the behavior of individuals », *American Sociological Review*, XV, 3, 1950, p. 351-357. Cependant, comme le note Raymond Boudon (« Propriétés individuelles et propriétés collectives : un problème d'analyse écologique », *Revue française de sociologie*, IV, 3, 1963, p. 275-299), le « paradoxe de Robinson » avait été soulevé dès 1929 (E. L. Thorndike, « On the Fallacy of Imputing the Correlations Found for Groups to the Individuals or Smaller Groups Composing Them », *American Journal of Psychology*, 52, 1929, p. 122-124). Gary King estime même que le problème de l'inférence écologique a été soulevé pour la première fois en 1919 (même si c'est

La modélisation multiniveau désigne une technique statistique de traitements de données. Ces données sont *explicitement* structurées de manière particulière : elles sont, au moins partiellement⁷, organisées selon une structure hiérarchique, c'est-à-dire que les individus statistiques, les unités de niveau 1 pour reprendre une formulation usuelle – les électeurs par exemple – sont inclus dans des ensembles ou des groupes, les unités de niveau 2 – e.g. la commune dans laquelle les électeurs résident. Nous insistons sur la structuration *explicite* des données : en effet, les données auxquelles ont affaire les sciences sociales, et la sociologie électorale en particulier, sont très fréquemment structurées de manière hiérarchique, sans pour autant que cette structuration soit explicitée et prise en compte dans le traitement des données. Il s'agit là d'une des origines de la supposée *fallace écologique*, ou plutôt de l'écart pouvant exister entre des relations analysées au niveau individuel et au niveau agrégé. Écart, car ainsi qu'on va le voir, la concurrence entre ces deux niveaux d'explication, popularisée par Robinson, est elle-même fallacieuse.

Prenons un exemple simple. Soit une enquête portant sur la relation entre niveau de richesse et vote à gauche, réalisée sur neuf individus résidant dans trois territoires différents⁸. La figure 1 présente les résultats de l'enquête.

Figure 1. Richesse et vote à gauche



uniquement dans une note de bas de page et qu'il est, en l'espèce, considéré comme bénin par les auteurs) dans : William Ogburn, Inez Goltra, « How Women Vote », *Political Science Quarterly*, XXXIV, 3, 1919, p. 413-433, cité in Gary King, *A Solution to the Ecological Inference Problem. Reconstructing Individual Behavior from Aggregate Data*, Princeton, N.J., Princeton University Press, 1997, p. 3. Parmi les contributions classiques au débat, on notera en particulier : Leo Goodman, « Ecological Regressions and the Behavior of Individuals », *American Sociological Review*, XVIII, 6, 1953, p. 663-664 ; Leo Goodman, « Some Alternatives to Ecological Correlation », *The American Journal of Sociology*, LXIV, 6, 1959, p. 610-625 ; O. D. Duncan *et al.*, *Statistical geography : problems in analyzing areal data*, Glencoe, Free Press, 1961 ; R. Boudon, « Propriétés individuelles et propriétés collectives », *op. cit.* ; Hayward R. Alker, « A Typology of Ecological Fallacies », in Mattei Dogan, Stein Rokkan (dir.), *Quantitative Ecological Analysis in the Social Sciences*, Cambridge and London, The MIT Press, 1969, p. 69-86 ; D. A. Freedman *et al.*, « Ecological Regression and Voting Rights », *Evaluation Review*, XV, 6, 1991, p. 673 ; C. H. Achen, W. P. Shively, *Cross-level inference*, Chicago, University of Chicago Press, 1995 ; Gary King, *A Solution to the Ecological Inference Problem*, *op. cit.*

⁷ Au moins partiellement parce que chaque unité de niveau 1 peut éventuellement être comprise dans plusieurs unités de niveau 2 non emboîtées entre elles.

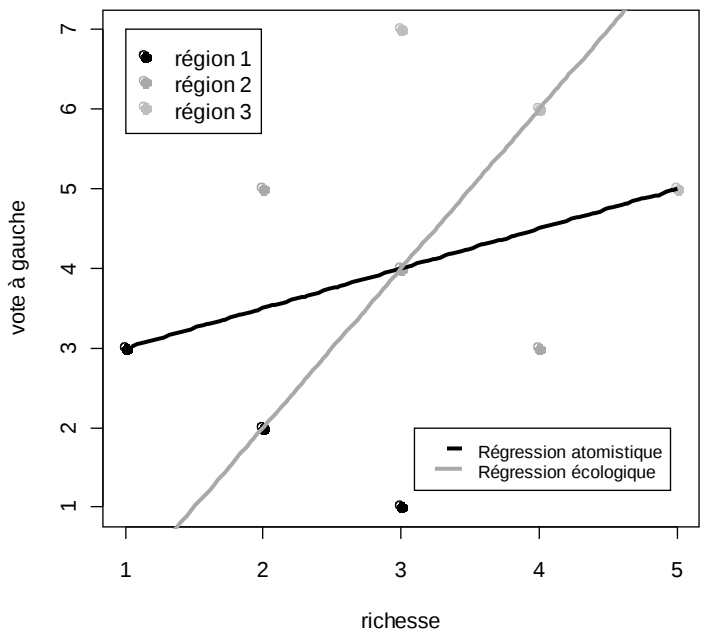
⁸ Cet exemple est très librement inspiré du débat américain sur les *Red States* et *Blue States*, et notamment de Andrew Gelman, *Red state, blue state, rich state, poor state: why Americans vote the way they do*, Princeton, Princeton University Press, 2008. Schématiquement, les États les plus riches des côtes votent davantage en faveur des Démocrates, alors que les sondages montrent qu'au niveau individuel le vote démocrate décroît en fonction du revenu, en particulier lorsqu'on analyse la relation État par État.

La figure 1 laisse perplexe : d'un côté, le nuage de points est ascendant, laissant penser qu'il existe une relation croissante entre richesse et vote à gauche. De l'autre, si l'on considère chaque région séparément, la relation est nettement décroissante.

Face à ces données, on peut dans un premier temps effectuer une régression linéaire classique, sur l'ensemble des neuf individus. C'est le traitement statistique standard pour des données issues d'un sondage national. La droite de régression résultante est figurée en noir sur la figure 2. C'est ce qu'on pourrait appeler une « régression atomistique » : les individus enquêtés sont considérés comme des atomes, indépendants les uns des autres. La relation qui en résulte est croissante (l'équation de la droite de régression est : $y = 2,5 + 0,5x$) mais rend faiblement compte des données observées ($R^2 = 0,1$ – la régression rend compte de 10 % de la variance seulement).

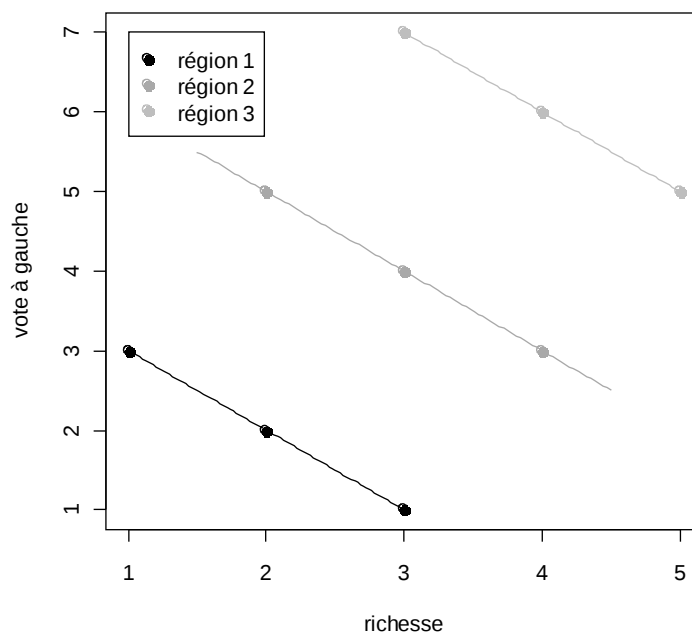
Imaginons maintenant qu'on ne dispose pas des données individuelles, mais seulement des moyennes agrégées par région. On est ici contraint à mettre en œuvre une analyse écologique, puisqu'on ne dispose que de données agrégées. On peut ainsi pratiquer la « régression écologique », dont le résultat est figuré en gris sur la figure 2. On constate que la relation mise en évidence entre richesse et vote à gauche est plus intensément positive que dans le cas de la régression atomistique ($y = -2 + 2x$), et qu'elle rend parfaitement compte des données, puisque la droite de régression passe par les trois points moyens ($R^2 = 1$). On est ici confronté au paradoxe de Robinson : lorsque le niveau d'agrégation croît, l'intensité de la corrélation croît (parfois !) également.

Figure 2. Régressions atomistique et écologique



Enfin, on peut effectuer une régression séparée par région. La figure 3 illustre le résultat : la relation entre richesse et vote à gauche est négative dans chacune des régions, seule la constante change ($y = \alpha_i - x$). De plus, pour chaque région, la régression est parfaite ($R^2 = 1$). De plus, si on fait la moyenne de ces droites de régression, on obtient la droite de régression de la région 2, également décroissante donc ($y = 7 - x$).

Figure 3. Régressions régionales



Alors, à quel saint, ou plutôt à quelle régression faut-il se vouer ? Quel est le niveau correct qu'il convient d'étudier ? Il nous semble clair que l'exemple présenté suggère qu'il n'y a pas de « bonne » régression ou de « bon » niveau auquel travailler. C'est sans doute pour cela que la querelle entre tenants de l'analyse écologique et défenseurs de l'analyse individuelle n'a jamais trouvé de solution : à notre sens, fallace écologique et fallace individuelle sont deux figures d'un même problème, l'ignorance du caractère (le plus souvent géographiquement, mais pas seulement ni nécessairement) structuré des données analysées. Ainsi, dans l'exemple présenté, des arguments existent aussi bien pour affirmer que les riches votent à gauche que pour défendre que les pauvres votent à gauche. En réalité, les trois approches présentées apportent chacune quelque chose. Au final, dans un territoire pertinent donné, la richesse est effectivement liée à un moindre vote à gauche ; mais il se trouve également que, *par ailleurs*, les régions en moyenne les plus riches sont aussi celles qui votent le plus à gauche. Cela peut être lié à une tierce variable, non observée, caractérisant les régions ou tous les individus qui les composent – par exemple la pratique religieuse.

Les modèles multiniveau servent précisément à modéliser ce genre de situations, où la prise en compte de la structure hiérarchique des données est nécessaire. Ils permettent en effet d'exploiter l'ensemble de l'information présente dans les données. Sans trop entrer dans les détails techniques⁹, la caractéristique d'un modèle multiniveau est d'estimer des paramètres à plusieurs niveaux – pour reprendre les termes introduits plus haut, aux niveaux 1 et 2 (et éventuellement 3, etc.). Les coefficients du niveau 1 (constantes et/ou pentes) peuvent ainsi varier d'une unité de niveau 2 à l'autre, et cette variation fait elle-même l'objet d'une modélisation et ainsi, potentiellement, d'une explication par des variables se rapportant aux unités de niveau 2. Dans le cas de l'exemple proposé, un modèle multiniveau permet d'estimer les paramètres pour chacune des régions ainsi que pour l'ensemble des enquêtés. Le cas

⁹ Le lecteur désireux de mieux comprendre la modélisation multiniveau pourra se reporter, en français, à Pascal Bressoux, *Modélisation statistique appliquée aux sciences sociales*, Bruxelles, De Boeck, 2008, ou en anglais à Andrew Gelman, Jennifer Hill, *Data analysis using regression and multilevel/hierarchical models*, Cambridge, New York, Cambridge University Press, 2007. Ce dernier manuel est plus complet et présente l'avantage d'avoir été rédigé par deux politistes. Dans la présente communication, nous avons, à la suite de Gelman et Hill, utilisé le logiciel libre R (R Development Core Team, *R: A language and environment for statistical computing*, Vienne, R Foundation for Statistical Computing, 2008, URL : <http://www.R-project.org>) avec les packages *arm* (Andrew Gelman, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman and Tian Zheng, *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*, R package version 1.2-9, URL : <http://www.stat.columbia.edu/~gelman/software/>), *lme4* (Douglas Bates, Martin Maechler et Bin Dai, *lme4: Linear mixed-effects models using S4 classes*, R package version 0.999375-28, URL : <http://lme4.r-forge.r-project.org>) et *rgrs* (Julien Barnier, *rgrs: Functions to make R usage in social sciences easier (in french)*, R package version 0.2-11, <http://alea.fr.eu.org/j/rgrs.html>).

échéant, on pourrait également intégrer au modèle une variable de niveau 2, permettant de rendre compte des variations des constantes (et/ou des pentes – dans l'exemple proposé, seules les constantes varient). On pourrait ainsi imaginer que ce soit la pratique religieuse moyenne d'une région qui rende compte des différentes constantes, et intégrer cette variable au modèle¹⁰.

La modélisation multiniveau s'est répandue à partir des années 1980, en particulier en sciences de l'éducation et en épidémiologie¹¹. En science politique, elle est devenue une technique relativement répandue dans la littérature anglophone ; en revanche, en sociologie et en science politique francophone, elle demeure quasi-inconnue, Anne Jadot étant la seule, à notre connaissance, à « pratiquer » le multiniveau dans la science politique française¹². A titre d'exemple, on ne trouve trace du mot « multiniveau » ni dans le *Bulletin de méthode sociologique*, ni dans la *Revue française de science politique*, tandis que la *Revue française de sociologie* a publié un article présentant les modèles multiniveaux¹³.

Il convient toutefois de noter que, si la technique statistique (ou plus exactement sa stabilisation et sa routinisation) est relativement récente, l'inspiration méthodologique et théorique est ancienne¹⁴. Les travaux de Mattei Dogan, souvent associé à Daniel Derivry¹⁵, en apportent la preuve : dans leurs articles de 1971 et plus encore 1986¹⁶, Dogan et Derivry déploient exactement la logique de la modélisation multiniveau – sauf qu'ils ne possèdent pas, à l'époque, les outils mathématiques (ni, *a fortiori*, informatiques) pour la mettre en œuvre, et sont ainsi réduits à multiplier les analyses par départements, croisées avec d'autres variables de niveau départemental.

Ce qui est remarquable, toutefois, c'est qu'alors que les travaux de Derivry et Dogan s'appuient intégralement sur des données agrégées (soit au niveau cantonal, soit au niveau départemental), l'immense majorité des travaux contemporains mobilisant la modélisation multiniveau prennent comme unités de niveau 1 des individus. Le plus souvent, ce sont des données d'enquête par questionnaire ou de sondages qui sont ainsi modélisées. On voit immédiatement tout le profit qu'on peut tirer d'une telle analyse : elle permet d'échapper à certaines critiques dirigées contre les sondages, en replaçant l'individu dans son contexte (ou, plus exactement, en reconstruisant un contexte à travers des données contextuelles). Toutefois, pour qu'un tel dessein de recherche produise son plein rendement, il faut que les données y soient adaptées – et en particulier, que les unités de niveau 2 soient a) significatives par rapport à la question de recherche ; b) suffisamment nombreuses. Or, on retrouve ici l'une des limitations classiques des sondages : leur coût, qui interdit ou rend difficile un échantillonnage dans un nombre suffisamment grand d'unités géographiques. Ainsi, à titre d'exemple, dans le Panel électoral français 2007, qui est l'une des plus grandes enquêtes électorales menées récemment en France, la stratification géographique est effectuée par région administrative, ce qui fait varier les effectifs régionaux de 11 enquêtés pour la Corse (et 21 pour l'Auvergne) à 292 pour l'Île-de-France. Le nombre d'individus ayant répondu aux quatre vagues de l'enquête s'élève au total à 1846, et il y a donc 22 unités de niveau 2. Ces données permettent a priori de recourir à une modélisation multiniveau (encore qu'il ne faut pas que les

¹⁰ Il serait toutefois délicat statistiquement d'introduire autant de variables et de niveaux dans un modèle ne comptant au total que neuf unités de niveau 1 et trois unités de niveau 2.

¹¹ Il est intéressant de noter que le *International Journal of Epidemiology* a consacré une partie de l'un de ses numéros récents (XXXVIII, 2, 2009) à republier l'article de Robinson de 1950 ainsi que quatre commentaires, qui insistent à des degrés divers sur l'importance des modèles multiniveaux comme réponse aux critiques de Robinson.

¹² Anne Jadot, Marcel van Egmond, « Réconcilier l'individuel et le contextuel ? L'intérêt de la méthode multiniveaux en recherche électorale », *Revue de la Maison française d'Oxford*, I, 1, 2003 ; Anne Jadot, « (Ne pas) être un électeur Européen. Une analyse multiniveaux des déterminants individuels et contextuels de l'abstention en 1999 », *Revue internationale de politique comparée*, IX, 1, 2002, p. 31-45.

¹³ Pascal Bressoux, Paul Coustère, Christine Leroy-Audouin, « Les modèles multiniveau dans l'analyse écologique : le cas de la recherche en éducation », *Revue française de sociologie*, XXXVIII, 1, 1997, p. 67-96.

¹⁴ Cf. la note de blog d'Andrew Gelman, qui rapporte que certains précurseurs construisirent des modèles multiniveaux dès la fin des années 1940 : Andrew Gelman, « The benefits of making a statistical method more ordinary », blog « Statistical Modeling, Causal Inference, and Social Science », URL : http://www.stat.columbia.edu/~cook/movabletype/archives/2009/05/the_benefits_of.html.

¹⁵ Les différentes publications issues du programme d'analyse écologique d'une base de données cantonales constituée par Mattei Dogan, qui s'étend de 1960 à 2000, sont listées dans : Mattei Dogan, « La France découpée en dix tranches sociologiques. La méthode de la décilisation », communication au Congrès de l'AFSP, Toulouse, 2007.

¹⁶ Daniel Derivry, Mattei Dogan, « Unité d'analyse et espace de référence en écologie politique. Le canton et le département français », *Revue française de science politique*, XXI, 3, 1971, p. 517-570 ; Daniel Derivry, Mattei Dogan, « Religion, classe et politique en France. Six types de relations causales », *Revue française de science politique*, XXXVI, 2, 1986, p. 157-181.

variables qualitatives utilisées dans le modèle comptent au total un nombre trop élevé de modalités), mais la véritable question est de savoir si la région administrative est une unité pertinente. En effet, on sait le peu d'ancrage historique et de sens politique que cet échelon possède. A n'en pas douter, l'échelon départemental eût été préférable – mais il aurait nécessité un échantillon beaucoup plus important.

C'est pourquoi l'utilisation de la modélisation multiniveau sur des données agrégées – justifiant ainsi pleinement le rapprochement opéré dans le titre de cette communication entre analyse écologique et modèles multiniveaux – nous semble conserver un très grand intérêt. Paradoxalement, malgré le potentiel extrêmement important de ce type de dessein de recherche, il demeure – à notre connaissance – rarissime. Le seul exemple que nous ayons rencontré d'une véritable modélisation multiniveau de données agrégées n'a pas été publié à ce jour¹⁷. Pourtant, le recours à des données agrégées permet d'avoir accès à des données de très bonne qualité, car enregistrant des pratiques sociales réelles, peu onéreuses, disponibles à des échelons territoriaux fins, tandis que la modélisation multiniveau permet d'exploiter au maximum l'information qu'elles contiennent tout en minimisant le risque de fallace écologique. Il y a fort à parier que ces données ont encore beaucoup à nous apprendre – en tout cas, le risque vaut d'être couru.

Le cas du Front national semble particulièrement propice au recours à ce type de dessein de recherche. En effet, cette marque électorale a sans doute été celle qui a été la plus étudiée depuis le milieu des années 1980, et ce au moyen de méthodes diverses et par des spécialistes ressortissant de plusieurs disciplines (politistes, sociologues et géographes en particulier). Au-delà des controverses qui opposent les auteurs, il ressort de cette littérature que le « phénomène Le Pen¹⁸ » ne peut être ramené à un facteur explicatif unique, et que l'électorat du Front national, ou plus exactement le « conglomerat électoral¹⁹ » constitué, à un moment donné, par les électeurs ayant glissé un bulletin de vote estampillé « Front national » dans l'urne, est extrêmement fluctuant dans le temps et dans l'espace. De plus, une des caractéristiques fondamentales de ce conglomerat est la très grande réticence de ceux qui le constituent à avouer face à un enquêteur leur vote. Dès lors, le recours à des données agrégées apparaît extrêmement utile, de même qu'une méthode permettant de tirer parti de leur haut degré d'hétérogénéité.

Données et modèles

Cette communication propose donc une analyse fondée sur les résultats électoraux obtenus par Jean-Marie Le Pen aux élections présidentielles de 1995, 2002 (premier et deuxième tours) et 2007, agrégés par commune et mis en relation avec des données diverses. En premier lieu, on a utilisé les données relatives à la composition socioprofessionnelle de la population de chaque commune. Le recensement général de la population (RGP) fournissant celle-ci en 42 positions, on a voulu exploiter cette finesse statistique. Pour ce faire, on a effectué une analyse en composantes principales²⁰ (ACP) de la composition socio-professionnelle en 1999 de chacune des 35 653 communes de notre base de données²¹, dont on a retenu les cinq premières composantes (qui résument 22,9 % de l'inertie totale ; ce chiffre qui peut sembler faible se comprend aisément compte tenu du grand nombre d'individus statistiques

¹⁷ Piero Ignazi, Spencer Wellhofer, « Class, Property, and Religion : Long-Run Trends on Secularization and Politics in Italy », communication présentée au 1st ISA Forum of Sociology, Barcelona, 5-8 septembre 2008, atelier « The persistent impact of class and religion in contemporary party politics » organisé par le Committee on Political Sociology. Lors de ce même atelier, nous avons présenté un papier intitulé « Class, religion and vote in contemporary France. An ecological analysis » qui mobilisait également un modèle multiniveau sur des données agrégées. Les deux papiers, non publiés pour l'heure, sont disponibles sur le site du Committee on Political Sociology : http://www2.spbo.unibo.it/cps/virtual_paper_room/2008_09_05_barcelona/.

¹⁸ Jacques Le Bohec, *Sociologie du phénomène Le Pen*, Paris, La Découverte, 2005.

¹⁹ Selon l'expression proposée par Daniel Gaxie, « Des penchants vers les ultra-droites », in Annie Collovald et Brigitte Gaïti (dir.), *La démocratie aux extrêmes. Sur la radicalisation politique*, Paris, La Dispute, 2006, p. 223-245.

²⁰ Appartenant à la famille des analyses factorielles (ou de l'analyse géométrique des données), l'ACP permet de dégager les principales dimensions structurant un ensemble de données. Jean-Paul Benzécri, *L'Analyse des données : leçons sur l'analyse factorielle et la reconnaissance des formes et travaux du laboratoire de statistiques de l'Université Paris VI*, Paris, Dunod, 1973 ; Brigitte Le Roux, Henri Rouanet, *Geometric Data Analysis : from correspondence analysis to structured data analysis*, Dordrecht, Kluwer Academic Publishers, 2004.

²¹ Ce nombre de 35 653 est dû à l'exclusion de la base des communes de la Corse, des DOM-TOM et des communes pour lesquelles certaines données étaient manquantes.

concernés). Par commodité, on se contente ici de restituer notre interprétation de ces cinq dimensions. La première dimension porte, sans surprise, sur l'opposition urbain-rural. La deuxième opposition, sans guère de surprise non plus, porte sur la quantité de capitaux détenue par les habitants d'une commune (ou, si l'on préfère, leur position sociale, basse ou élevée) et oppose les communes fortement ouvrières aux communes dans lesquelles les cadres, professions libérales et professeurs sont surreprésentés. La troisième dimension, à l'interprétation peut-être plus délicate, correspond à l'opposition entre « économie résidentielle²² » et « économie productive ». La quatrième composante principale renvoie au rapport au public (au sens de clients), en opposant très nettement les employés de commerce aux employés administratifs d'entreprise. Enfin, la cinquième dimension renvoie au poids de l'économie publique (en se plaçant dans le même cadre d'analyse, issu de la théorie de la base économique, que précédemment).

Afin d'ajouter une composante dynamique à ces données synchroniques, on a également utilisé les données fournies par l'INSEE concernant l'évolution intercensitaire de la population active (PCS en six positions) depuis 1968. Afin de rendre ces données exploitables dans le cadre de cette analyse, on a dans un premier temps construit un modèle multiniveau de croissance par PCS. Ce modèle permet de modéliser, commune par commune, l'évolution pour chacune des PCS. Afin de demeurer simple, on a modélisé l'évolution entre 1968 et 1999 de manière linéaire, même s'il est clair qu'une modélisation quadratique aurait sans doute été plus efficace – mais elle aurait également été beaucoup plus difficile à estimer, à interpréter et à réutiliser. Pour chaque commune et pour chaque PCS, deux paramètres – l'ordonnée à l'origine et la pente temporelle – sont donc retenues, soit au total 12 variables par commune. Sur celles-ci, une ACP est menée, permettant ainsi de réduire fortement le nombre de variables à prendre en compte. Les trois premières composantes principales, résumant respectivement 42, 18 et 16 % de la variance totale, sont retenues. Le premier axe oppose les communes anciennement rurales, dans lesquelles le déclin des agriculteurs a été le plus rapide, aux communes davantage urbaines. Le deuxième axe oppose les communes ayant le plus subi la désindustrialisation à celles dans lesquelles la proportion d'ouvriers n'a que peu diminué, voire a augmenté. Enfin, la troisième composante concerne les indépendants (CSP 2) – là où ils étaient autrefois le plus nombreux, leur diminution a été la plus rapide.

En troisième lieu, afin de tester une hypothèse récurrente dans le débat public, et dans une moindre mesure dans la littérature scientifique²³, on utilise comme variable indépendante la proportion de logements HLM dans la commune.

En quatrième lieu, le revenu imposable moyen par commune (en 2006) est également utilisé. Cette donnée est sujette à caution, dans la mesure où la moyenne n'a pas nécessairement beaucoup de sens, s'agissant d'une distribution non normale et asymétrique²⁴ – c'est pourquoi on retient souvent plutôt la médiane. De plus, une mesure d'inégalité – comme le coefficient de Gini, ou l'écart interdécile – peut s'avérer plus pertinente. On a toutefois fait le choix de retenir la moyenne, car compte tenu des règles de secret statistique mises en œuvre par la direction générale des finances publiques (DGFip, qui diffuse les données), la moyenne est la donnée disponible pour le plus grand nombre de communes – sans pour autant l'être pour les toutes petites communes. La distribution de cette variable étant très asymétrique, on en a retenu une transformation logarithmique. De plus, il conviendra de se rappeler, au stade de l'interprétation, que l'essentiel des communes possède un revenu fiscal moyen compris entre 7 500 et 30 000 €²⁵. On verra plus loin qu'afin de vérifier que cette variable n'est pas redondante avec la composition socioprofessionnelle de la commune, on a construit un modèle d'interaction avec le caractère ouvrier de la commune.

En cinquième lieu, le type de commune, par référence au zonage en aires urbaines et aires d'emploi de l'espace rural (ZAUER²⁶) qui distingue les centres urbains, les communes monopolarisées et les communes multipolarisées qui constituent les communes périurbaines, et les communes rurales, est utilisé. Cette distinction semble importante, parce que le simple continuum urbain-rural mis en évidence par la première dimension de l'ACP évoquée plus haut ne permet pas de saisir ce qui se joue aux marges

²² Sur cette notion, et la critique de son opérationnalisation par l'INSEE, cf. Laurent Davezies, « L'économie locale 'résidentielle' », *Géographie, économie, société*, XI, 1, 2009, p. 47-53 ; du même, *La République et ses territoires : la circulation invisible des richesses*, Paris, Seuil (coll. La République des idées), 2008.

²³ Pour un exemple récent, cf. Bertrand Alidières, *Géopolitique de l'insécurité et du Front national*, Paris, Armand Colin, 2006.

²⁴ Un revenu imposable ne peut être inférieur à 0, en revanche il n'a pas de borne supérieure.

²⁵ Soit approximativement, en base logarithmique, entre 9 et 10,5.

²⁶ Thomas Le Jeannic, Joëlle Vidalenc, « Pôles urbains et périurbanisation - Le zonage en aires urbaines », *Insee Première*, 516, 1997.

des villes, et qui est l'un des objets les plus importants de la géographie sociale contemporaine²⁷. De plus, certains auteurs, et en particulier Jacques Lévy²⁸, avancent l'hypothèse que les votes « tribunitiens », et le vote FN en particulier, trouveraient leur meilleur terreau dans les communes périurbaines – Jacques Lévy en rendant compte par le choix délibéré que feraient les individus les plus intolérants à la diversité sociale et ethnique de se regrouper dans un entre-soi peu propice à l'ouverture politique. Il importe donc de pouvoir tester cette théorie.

Une dernière donnée qui serait intéressante à mobiliser serait la distance d'une commune à l'aire urbaine de plus de 200 000 habitants la plus proche. Il s'agirait ici encore de tester une hypothèse fréquemment avancée par des géographes, et qui n'est pas sans lien avec la précédente, selon laquelle le vote Front national serait organisé selon un « gradient d'urbanité » – appréhendé cette fois-ci par la distance géographique au centre et non par le statut périurbain de la commune²⁹. Toutefois, l'effet de cette variable n'étant pas linéaire, sa modélisation devient extrêmement complexe dans le cadre du multiniveau³⁰ ; aussi n'avons nous pas retenu cette donnée.

Rappelons à ce stade que l'objet de cette communication n'est pas d'aboutir à des conclusions définitives quant au vote Front national à partir de ces données, mais bien de suggérer le potentiel offert à la sociologie électorale par la modélisation multiniveau de données agrégées. Aussi les analyses qui sont présentées ici ne sont-elles qu'exploratoires.

Afin que l'interprétation des résultats demeure dans les limites d'une complexité raisonnable, nous avons mené cinq régressions multiniveau successives, portant respectivement sur les cinq variables ou ensembles de variables présentées ci-dessus. L'idéal aurait sans doute été de construire un seul modèle, intégrant l'ensemble de ces variables ; mais si cela eût été aisé dans le cadre d'une régression multiple ordinaire, lorsqu'il s'agit d'un modèle multiniveau il n'est guère conseillé d'accroître outre mesure le nombre de régresseurs, car outre que cela rend l'estimation du modèle très coûteuse en ressources informatiques³¹, son interprétation devient également très complexe.

De plus, chaque tour de scrutin considéré a été modélisé séparément. Là encore, il est théoriquement possible de construire un modèle intégrant l'évolution temporelle ; mais, à nouveau, cela en accroîtrait déraisonnablement la complexité, d'autant que la variable temporelle devrait sans doute être modélisée non pas linéairement mais quadratiquement.

Au total, ce sont donc les résultats de vingt-huit modèles multiniveaux différents³² qui sont présentés ci-après. On conçoit que, dans ces conditions, seuls les résultats les plus saillants soient mis en valeur. Pour chaque modèle – outre les résultats du modèle présentés en annexe – un graphique et une carte sont présentés, le but étant à chaque fois de mettre en valeur les variations spatiales et, autant que possible, temporelles des logiques sous-jacentes au vote en faveur du Front national. A l'issue de cette présentation, une tentative de synthèse sera menée, permettant d'au moins partiellement pallier l'impossibilité de construire un seul « super-modèle » intégrant toutes les données.

²⁷ Cf. parmi de très nombreux exemples, Samuel Arlaud, Yves Jean et Dominique Royoux (dir.), *Rural-Urbain. Nouveaux liens, nouvelles frontières*, Rennes, Presses universitaires de Rennes, 2005.

²⁸ Jacques Lévy, « Vote et gradient d'urbanité. L'autre surprise du 21 avril », *EspacesTemps.net*, 05 juin 2003, <http://espacestemps.net/document432.html> ; Dominique Andrieu, Jacques Lévy, « L'archipel français. Les votes de 2007, indicateurs d'une nouvelle configuration de l'espace français », *EspacesTemps.net*, 27 juillet 2007, <http://www.espacestemps.net/document2861.html>. Voir aussi la controverse : Fabrice Ripoll, Jean Rivière, « La ville dense comme seul espace légitime ? », *Annales de la recherche urbaine*, 102, 2007, p. 121-130 ; Jacques Lévy, « Regarder, voir. Un discours informé par la cartographie », *Annales de la recherche urbaine*, 102, 2007, p. 131-140.

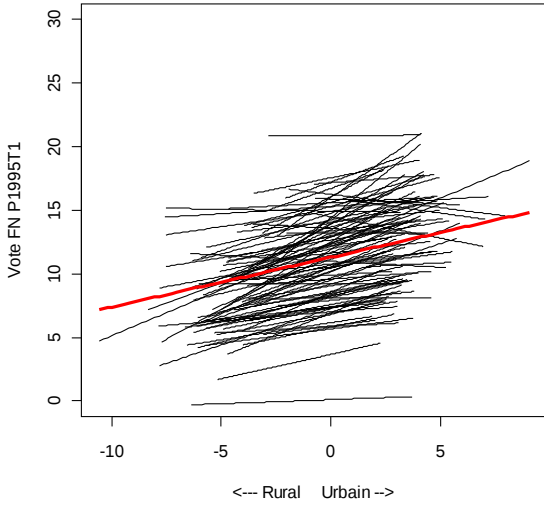
²⁹ Loïc Ravenel, Pascal Buléon et Jérôme Fourquet, « Vote et gradient d'urbanité : les nouveaux territoires des élections présidentielles de 2002 », *Espace, populations, sociétés*, 3, 2003, p. 469-482 ; Michel Bussi, Jérôme Fourquet, « Élection présidentielle 2007. Neuf cartes pour comprendre », *Revue française de science politique*, LVII, 3, 2007, p. 411-428 ; Jérôme Fourquet, « L'érosion électorale du lepénisme », in Pascal Perrineau (dir.), *Le vote de rupture. Les élections présidentielle et législatives d'avril-juin 2007*, Paris, Presses de Sciences Po, 2008.

³⁰ Compte tenu du profil de la courbe, il faudrait la modéliser par une équation du quatrième degré. L'estimation d'un tel modèle multiniveau devient alors extrêmement complexe d'un point de vue computationnel, et l'interprétation des résultats plus qu'ardue.

³¹ Le temps nécessaire à l'estimation d'un modèle multiniveau croît de manière géométrique par rapport au nombre d'effets aléatoires inclus dans le modèle.

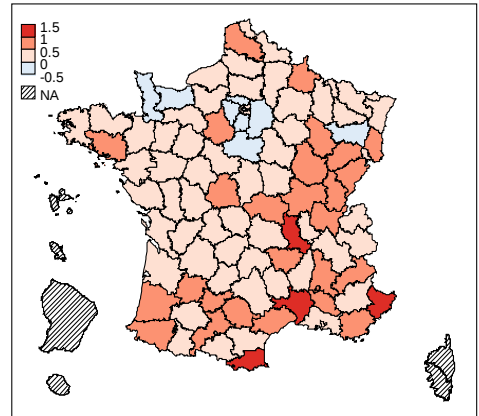
³² Modèles vides inclus, qui ne servent qu'à mesurer la significativité des autres modèles.

Vote FN en 1995 et clivage urbain-rural



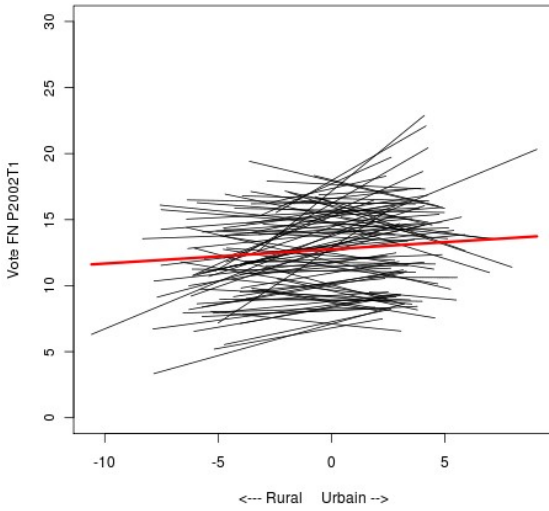
Graphique 1

Vote FN des villes et vote FN des champs en 1995



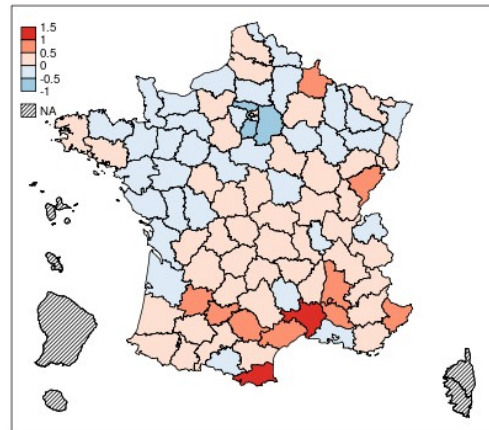
Carte 1

Le vote FN et le clivage urbain-rural en 2002 - 1er tour



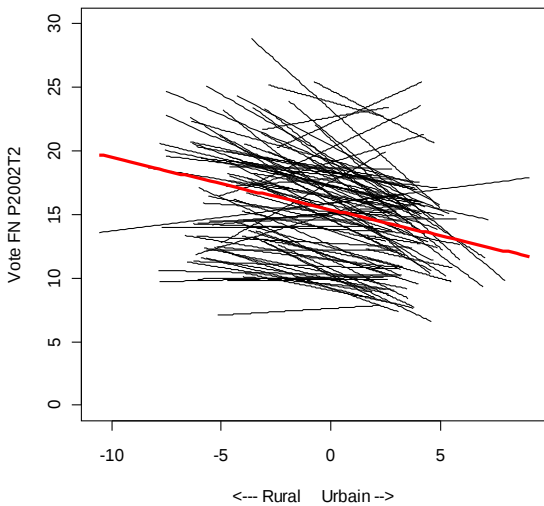
Graphique 2

Vote FN et le clivage urbain-rural en 2002 - 1er tour



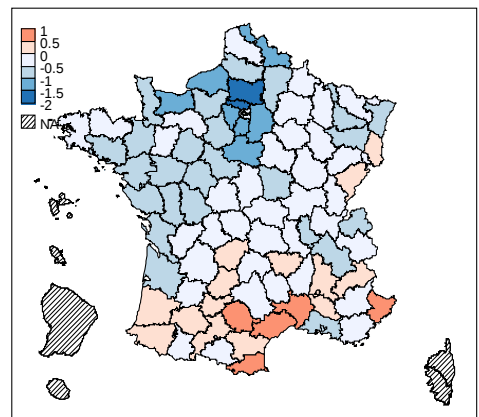
Carte 2

Vote FN et le clivage urbain-rural en 2002 au 2ème tour

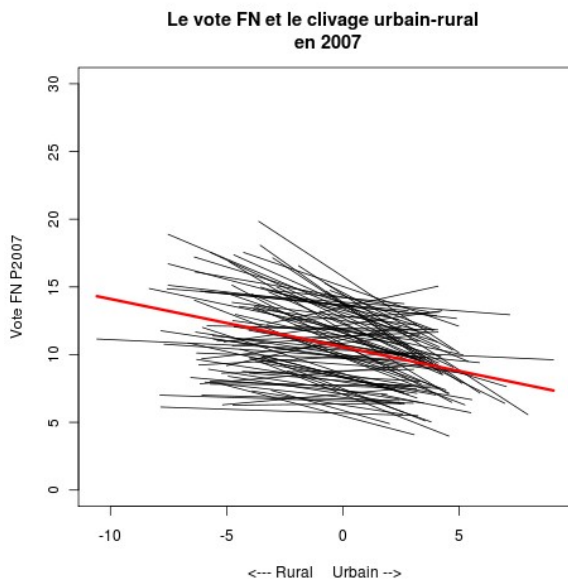


Graphique 3

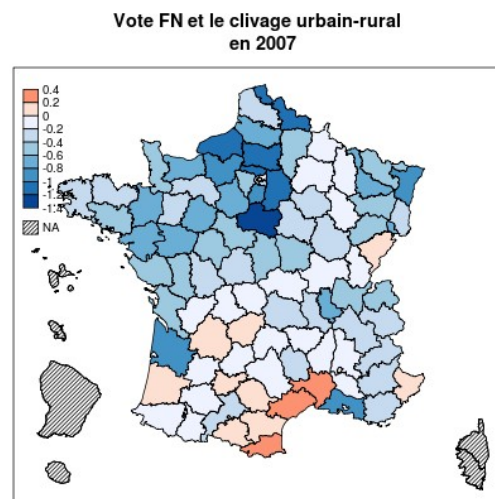
Vote FN et le clivage urbain-rural en 2002 au 2ème tour



Carte 3



Graphique 4



Carte 4

Résultats

Commençons par exposer les résultats des modèles multiniveaux portant sur la composition sociodémographique des communes en 1999. Ce modèle possède une certaine pertinence, mais moins sans doute que ce à quoi on aurait pu s'attendre³³. De plus, les cinq dimensions qui avaient été retenues se révèlent diversement à même de rendre compte des variations du vote en faveur du FN.

La première dimension, qui renvoie au clivage urbain-rural, est intéressante car elle offre un bon exemple de la manière dont une même variable peut être corrélée différemment au même résultat – ici le vote en faveur du FN – dans le temps et dans l'espace. Le graphique 1 représente la relation entre cette dimension et le vote en faveur de Jean-Marie Le Pen au premier tour de l'élection présidentielle de 1995 par département, compte tenu des autres dimensions (ou, si l'on préfère, toutes choses égales par ailleurs). Chaque ligne noire représente l'estimation de cette relation pour un département ; la ligne rouge correspond à l'estimation pour l'ensemble de la France. Ce type de graphique permet ainsi de visualiser la manière dont une telle relation varie d'un département à l'autre. Dans le cas présent, on observe que dans la plupart des départements, le vote Le Pen varie comme une fonction croissante de la première dimension – ce qui invalide ainsi la thèse de Jacques Lévy sur le vote frontiste comme vote non-urbain ou suburbain. Toutefois, dans quelques cas, cette relation est décroissante et correspond ainsi à ce schéma. Comme on le voit sur la carte 1, sur laquelle est représentée la pente de chacune des droites du graphique 1, il s'agit essentiellement de départements du bassin parisien ; et encore faut-il noter que ces pentes ne sont que faiblement négatives.

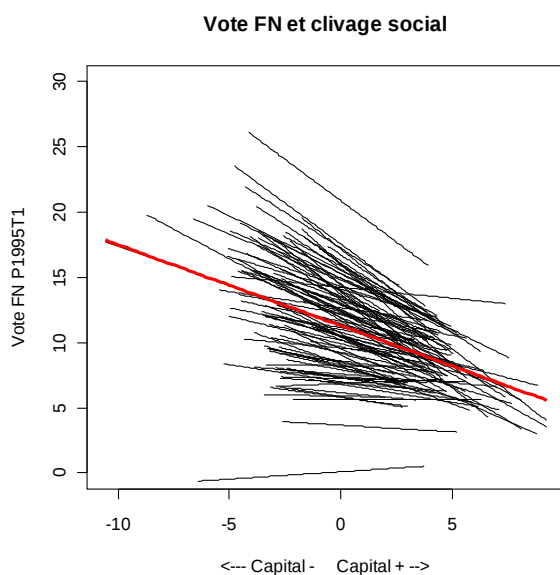
Comme le montrent le graphique 2 et la carte 2, cette dimension urbain-rural renvoie à une opposition plus forte lors du premier tour de l'élection présidentielle de 2002. En particulier, on trouve davantage de départements dans lesquels le vote frontiste est plus fort hors des villes : dans le bassin parisien à nouveau, mais aussi dans l'Ouest, tandis que dans le Sud, de Nice à Toulouse, le vote frontiste est davantage un fait urbain (à l'exception, notable, des Bouches-du-Rhône).

Le graphique 3 et la carte 3 montrent comment, entre les deux tours de l'élection présidentielle de 2002, se produit un basculement dans le rapport entre vote FN et urbanité (ou ruralité) : en l'espace de quinze jours, le vote Le Pen devient dans une majorité de départements non-urbain, situés pour la plupart au nord d'une ligne Bordeaux-Strasbourg. Cette tendance se confirme lors de l'élection présidentielle de 2007 (graphique 4 et carte 4) ; si le vote Le Pen demeure plutôt urbain dans quelques départements, notamment en Languedoc-Roussillon, l'intensité de cette liaison n'a cessé de diminuer de 1995 à 2007.

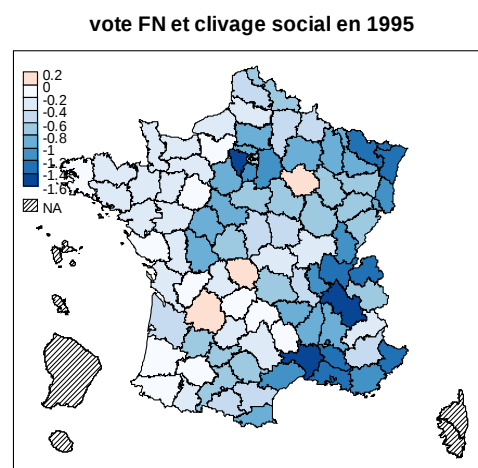
³³ La performance d'un modèle multiniveau est mesurée grâce à la réduction de la déviance par rapport au modèle vide. Contrairement aux modèles de régression par les moindres carrés ordinaires, il est difficile d'apprécier cette performance de manière absolue. Il s'agit donc surtout de comparer entre elles les performances relatives de plusieurs modèles multiniveaux.

La deuxième dimension concernant l'influence de la composition socio-professionnelle des communes concerne la position dans la hiérarchie socio-professionnelle, et en particulier la proportion d'ouvriers dans la population active. Comme on le voit sur le graphique 5, en 1995, il existe une forte relation entre le caractère ouvrier d'une commune et le score que recueille le président du Front national ; cependant, cette relation, si on l'observe dans quasiment tous les départements, n'y possède pas la même force partout. Elle est particulièrement accentuée sur le pourtour méditerranéen, le long du sillon rhodanien, en Alsace et en Moselle, en Ile-de-France (cf. carte 5). Dans le centre de la France, sur la façade atlantique, dans tout le grand Ouest, et même de la Normandie au Nord, la relation entre caractère ouvrier d'une commune et vote FN est faible ou inexistante. Lors du premier tour de l'élection présidentielle de 2002, cette relation est particulièrement accentuée sur le pourtour méditerranéen, en particulier dans le Vaucluse, le Gard et l'Hérault, tandis qu'ailleurs elle tend à être moins nette qu'en 1995 (graphique et carte 6). Lors du second tour de l'élection présidentielle de 2002, cette géographie tend à se confirmer, même si dans quelques départements de l'Ouest et du Sud-Ouest, l'audience de Jean-Marie Le Pen est inversement proportionnelle au caractère ouvrier de la commune (graphique et carte 7). En 2007 en revanche, le lien entre le caractère ouvrier d'une commune et le vote frontiste semble plus étroit que jamais – sans doute parce que les franges les moins populaires du conglomerat électoral de 2002 se sont tournées vers Nicolas Sarkozy.

La controverse scientifique qui a eu, et a toujours, cours au sujet du caractère ouvrier ou non de l'électorat du Front national doit ainsi trouver une réponse nuancée : il l'a été à des degrés changeants dans le temps, et surtout, d'un point de vue synchronique, il l'est très inégalement dans l'espace. De cette manière, on comprend mieux la physionomie de la carte électorale du FN : elle est si différente de la carte de l'implantation ouvrière parce que les ouvriers ne votent pas partout davantage pour ce parti que le reste de la population. Cette approche ne résoud pas tous les problèmes – reste à savoir en fonction de quoi varie le lien entre caractère ouvrier d'une commune et vote FN – mais elle a le mérite d'éviter de recourir à des facteurs explicatifs ad hoc pour expliquer les « déviations » par rapport à un supposé modèle national du vote³⁴.



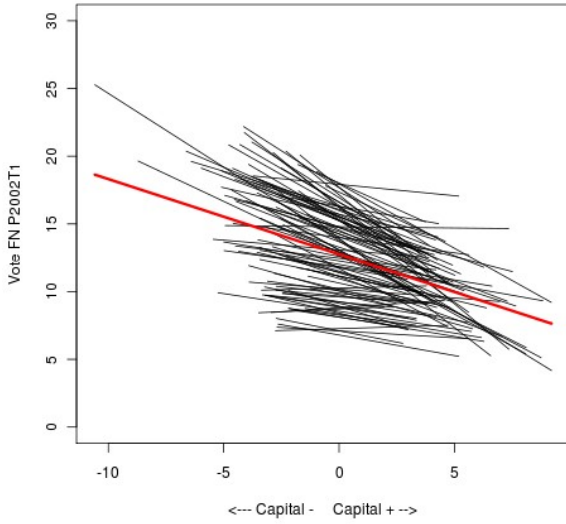
Graphique 5



Carte 5

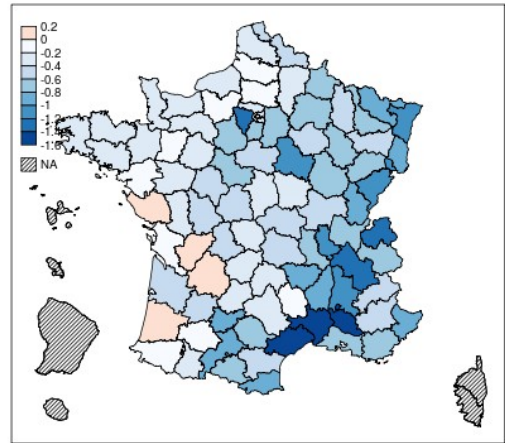
³⁴Cette tentative semble affecter particulièrement certains géographes, qui ont certes intégré certains acquis de la sociologie électorale, mais pour immédiatement rechercher les « écarts » ou les « déviations » par rapport au « modèle sociologique », écarts ou déviations qui relèveraient alors de la géographie et non de la sociologie. Un exemple symptomatique en est : Michel Bussi *et al.*, « Modéliser la diffusion spatiale de l'extrême droite : une expérimentation sur le Front national en France », communication au colloque « L'extrême droite en Europe », Paris, AFSP, 2004 : www.cevipof.msh-paris.fr/publications/notes_etudes/afsp2004-bussi.pdf.

Le vote FN et le clivage social en 2002 - 1er tour



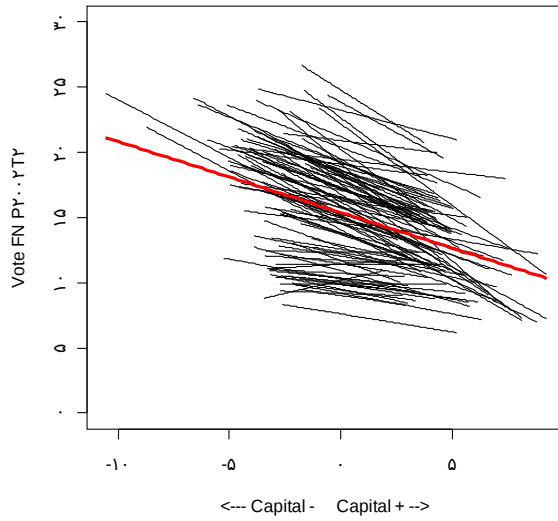
Graphique 6

Vote FN et le clivage social en 2002 - 1er tour



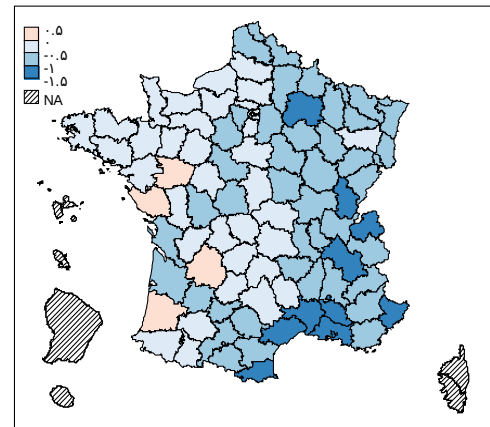
Carte 6

Vote FN et le clivage social en 2002 - 2ème tour



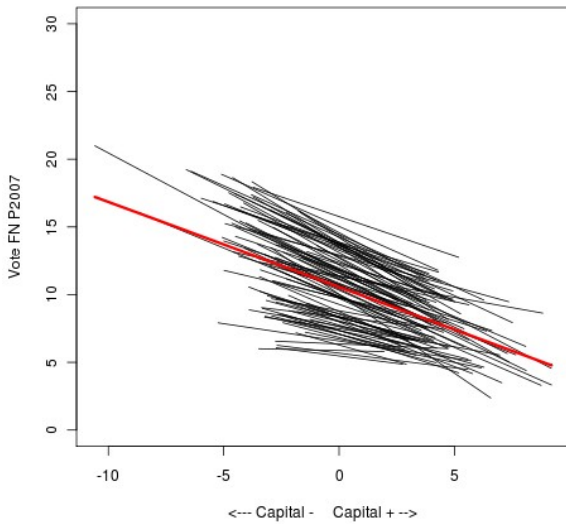
Graphique 7

Vote FN et le clivage social en 2002 - 2ème tour



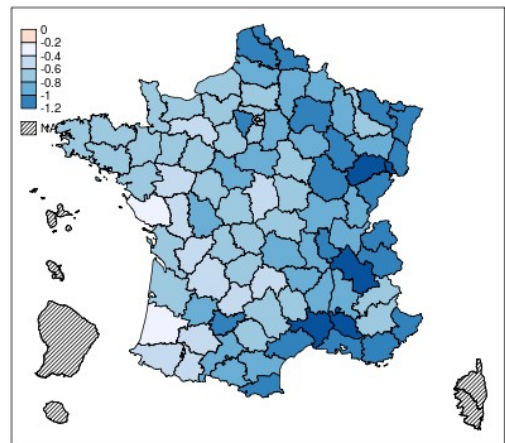
Carte 7

Le vote FN et le clivage social en 2007



Graphique 8

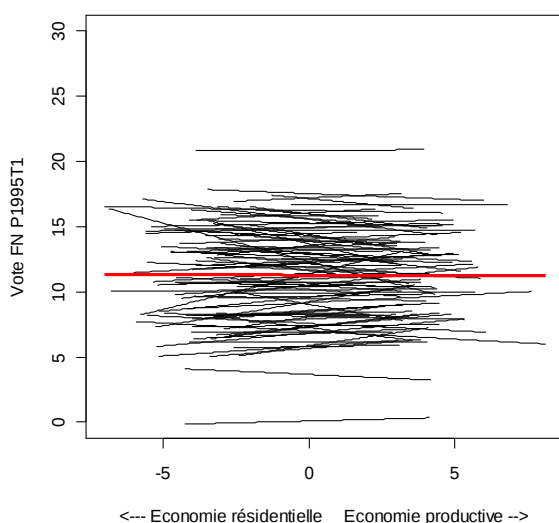
Vote FN et le clivage social en 2007



Carte 8

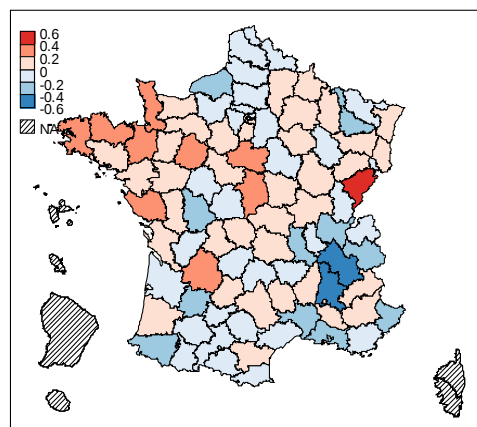
L'influence de l'importance respective de l'économie résidentielle et de l'économie productive dans une commune sur le vote en faveur du Front national est dans l'ensemble très faible, comme le montrent les pentes très douces des droites figurées sur les graphiques 9 à 12, et son sens varie d'un département à l'autre. Cependant, cette influence a progressé dans le temps jusqu'au deuxième tour de l'élection présidentielle de 2002, et n'est pas négligeable dans certains départements. De plus, la structure géographique de cette corrélation est intéressante (cartes 9 à 12) : grossièrement, dans les terres de mission du Front, la corrélation est positive entre économie productive et vote Front national, tandis que dans ses zones de force, c'est plutôt dans les territoires où l'économie résidentielle joue un rôle important que l'audience électorale du FN est la plus élevée. Cette dimension demeure cependant assez secondaire dans l'explication du vote FN, en tout cas à ce stade.

Vote FN et économie résidentielle



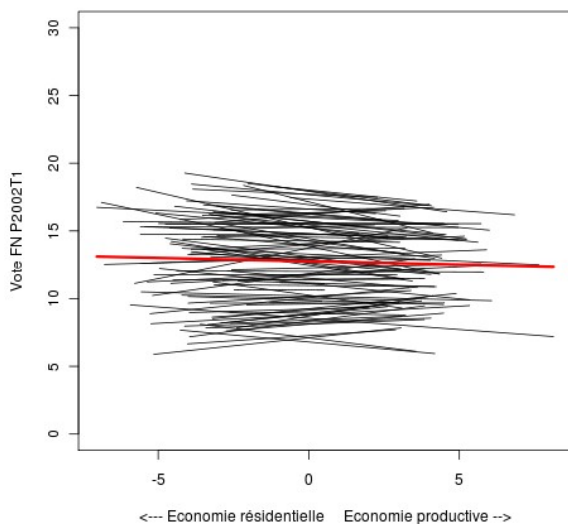
Graphique 9

vote FN et économie résidentielle en 1995



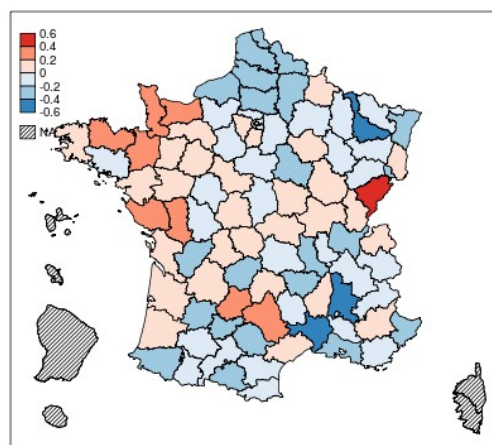
Carte 9

Le vote FN et l'économie résidentielle en 2002 - 1er tour

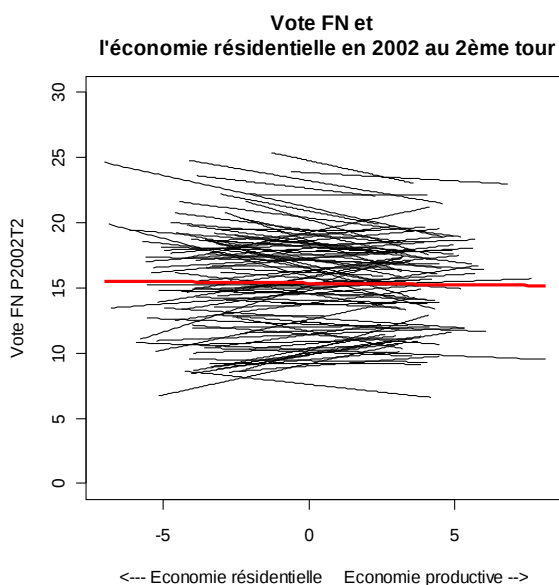


Graphique 10

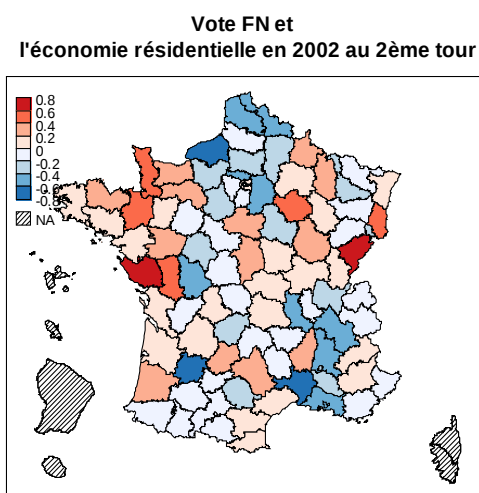
Vote FN et l'économie résidentielle en 2002 - 1er tour



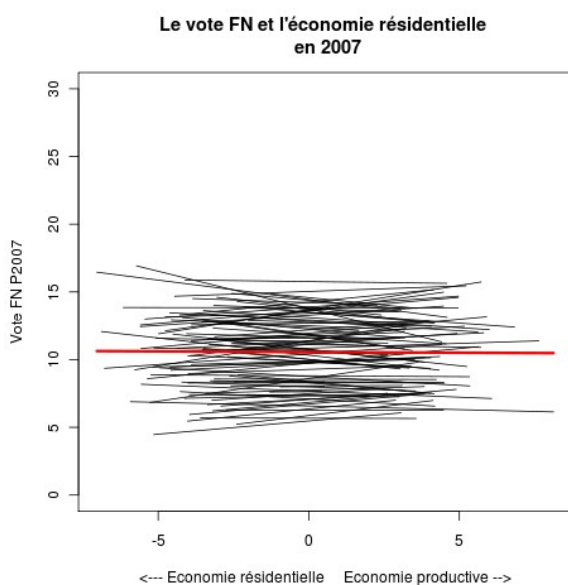
Carte 10



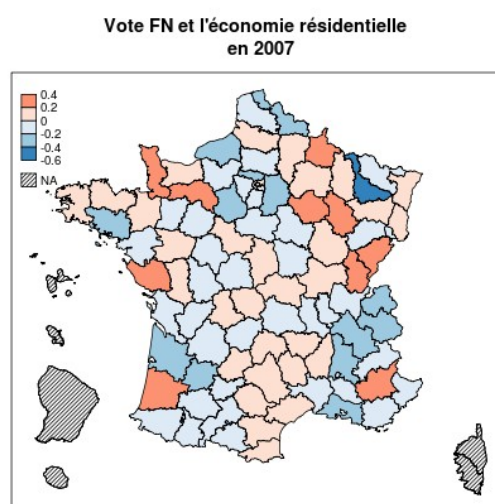
Graphique 11



Carte 11



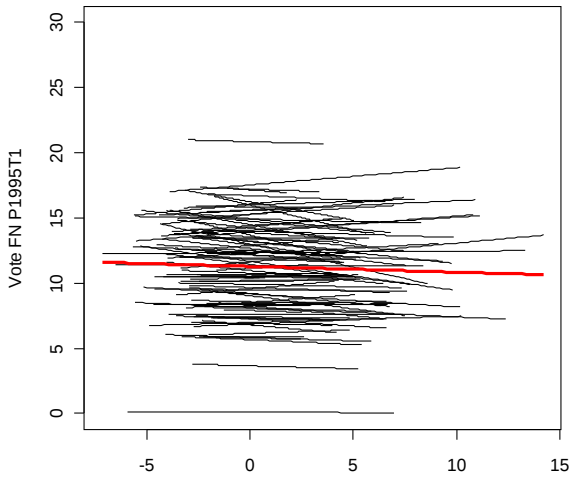
Graphique 12



Carte 12

La quatrième dimension, qui caractérise les territoires en fonction de la proportion de la population active en contact avec le public (usagers, clients...), ne semble également posséder qu'un rapport lâche avec le vote en faveur du Front national. En 2007, faible exposition au public et vote FN semblent légèrement corrélés positivement, tandis qu'aux scrutins précédents, des corrélations positives et négatives existaient, mais dans tous les cas les pentes sont faibles et ne peuvent rendre compte, au mieux, que de variations de quelques points de pourcentage.

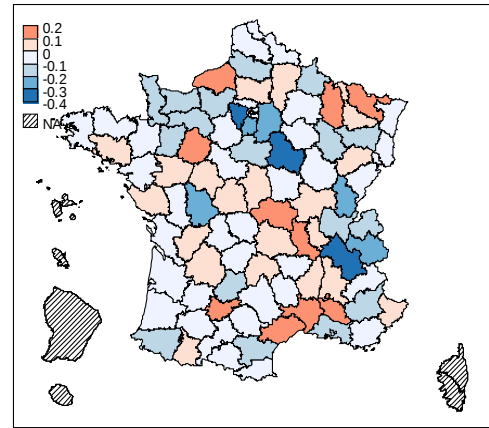
Vote FN et rapport au public



<--- Peu de contacts avec le public Contacts fréquents avec le public --->

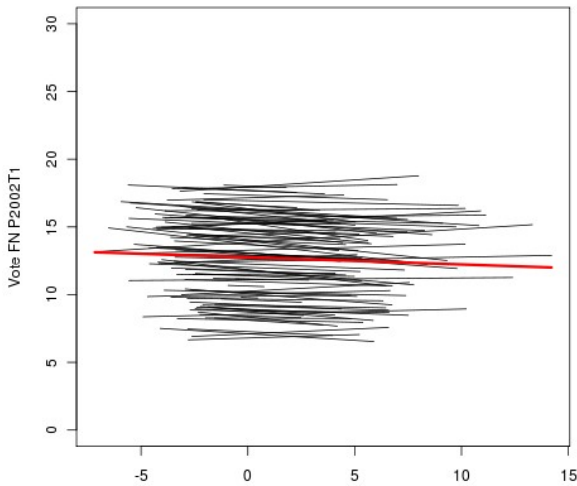
Graphique 13

vote FN et rapport au public en 1995



Carte 13

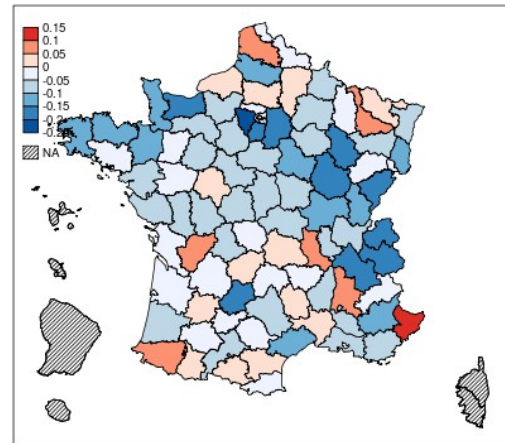
Le vote FN et le rapport au public en 2002 - 1er tour



<--- Peu de contacts avec le public Contacts fréquents avec le public --->

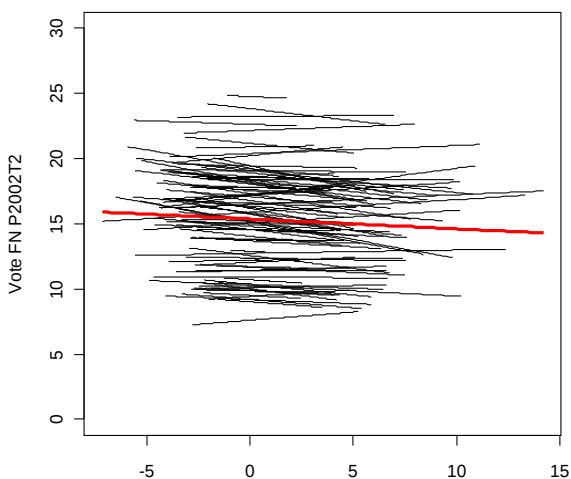
Graphique 14

Vote FN et le rapport au public en 2002 - 1er tour



Carte 14

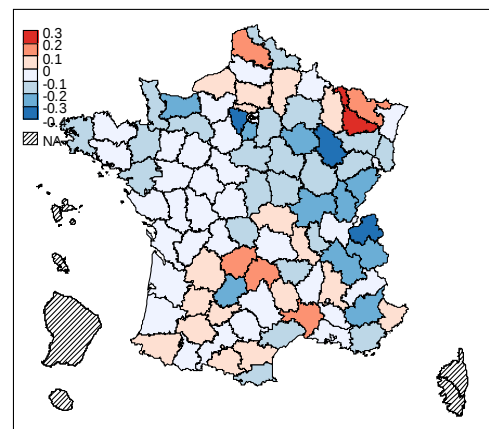
Vote FN et le rapport au public en 2002 au 2ème tour



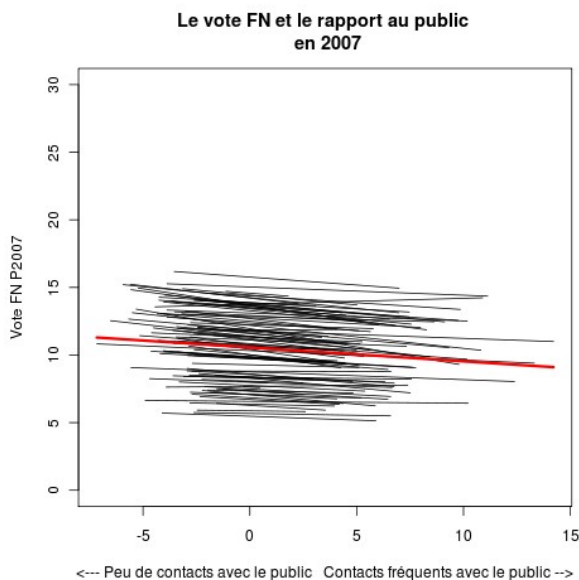
<--- Peu de contacts avec le public Contacts fréquents avec le public --->

Graphique 15

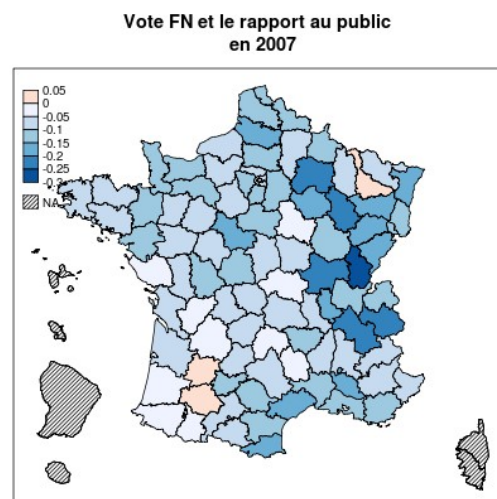
Vote FN et le rapport au public en 2002 au 2ème tour



Carte 15



Graphique 16



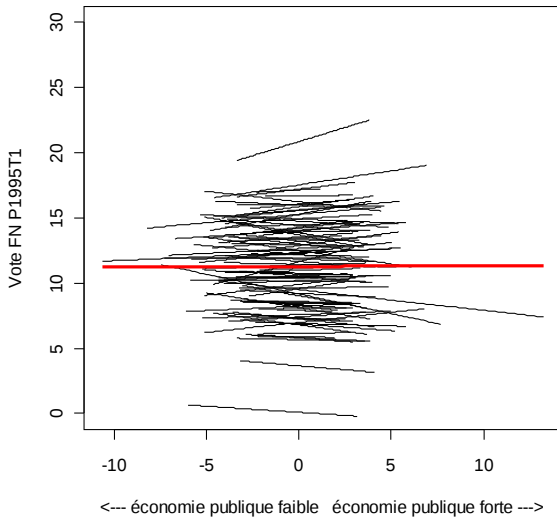
Carte 16

Enfin, la dernière dimension tirée de la composition socioprofessionnelle des communes est liée à l'importance de l'emploi public parmi les habitants d'une commune. Au niveau individuel, le clivage public-privé est devenu l'un des déterminants du vote les plus importants³⁵, les salariés du secteur public manifestant une propension à voter à gauche significativement plus importante que les salariés du privé. De plus, les salariés du public votent en général moins pour le Front national que les autres. On pourrait donc s'attendre à retrouver cette relation au niveau agrégé.

Sans être totalement négligeable, cette dimension ne contribue pourtant pas grandement à la structuration du conglomérat électoral frontiste. Il est intéressant de noter en revanche que si, en moyenne nationale, il n'y a pas de lien entre vote FN et économie publique (sauf lors du deuxième tour de l'élection présidentielle de 2002, où le vote FN est inversement proportionnel au poids de l'économie publique), en revanche, dans certains départements une telle corrélation existe, positive ici, négative là. L'interprétation n'en est pas aisée, en particulier parce que cette dimension peut renvoyer à des réalités socioéconomiques complexes (par exemple, l'emploi public est-il « par défaut », lié à une désindustrialisation, ou bien au contraire le signe d'un processus de métropolisation synonyme d'emplois fortement qualifiés ?). Des études de cas seraient sans doute nécessaires pour pouvoir pleinement interpréter cette dimension.

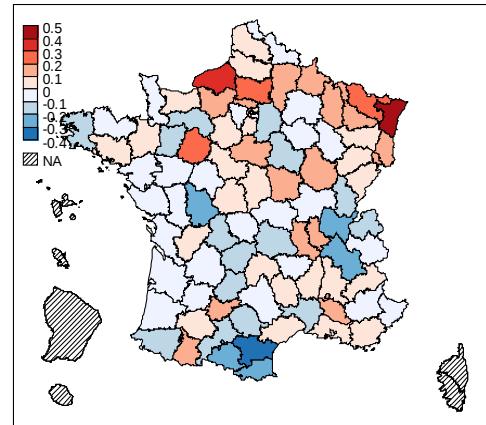
³⁵François de Singly, Claude Thélot, *Gens du privé, gens du public : la grande différence*, Paris, Dunod, 1989.

Vote FN et économie publique



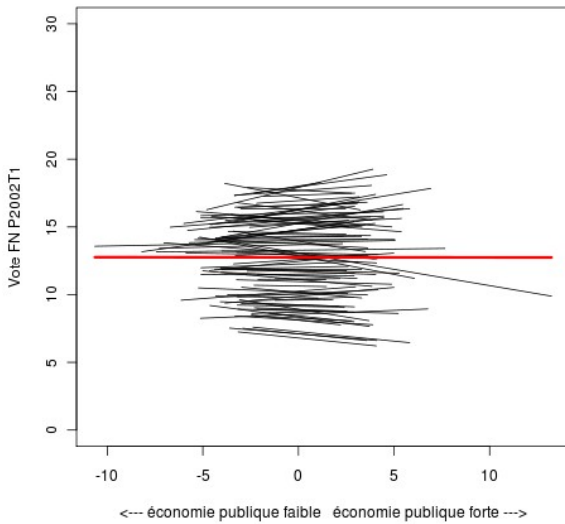
Graphique 17

vote FN et économie publique en 1995



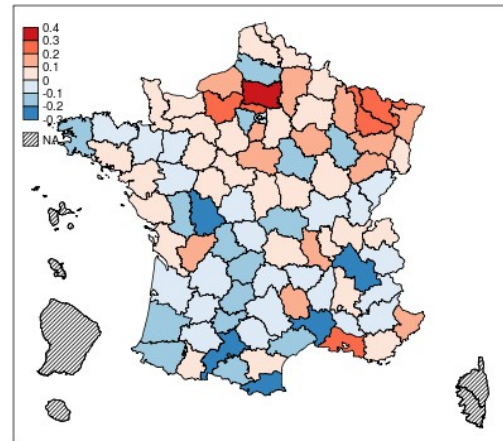
Carte 17

Le vote FN et l'économie publique en 2002 - 1er tour



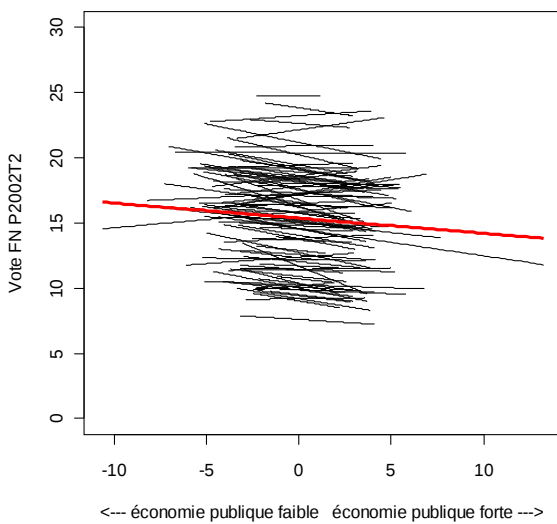
Graphique 18

Vote FN et l'économie publique en 2002 - 1er tour



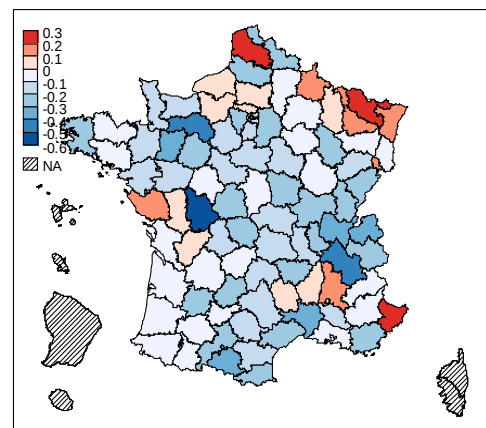
Carte 18

Vote FN et l'économie publique en 2002 au 2ème tour

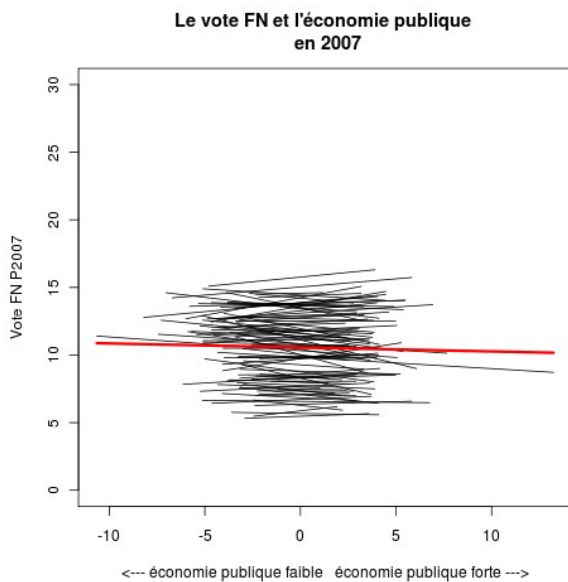


Graphique 19

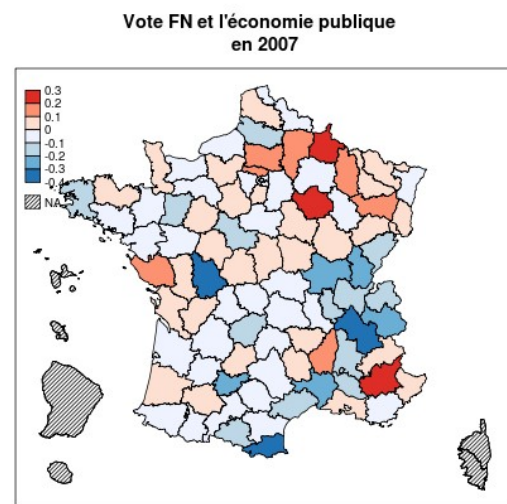
Vote FN et l'économie publique en 2002 au 2ème tour



Carte 19



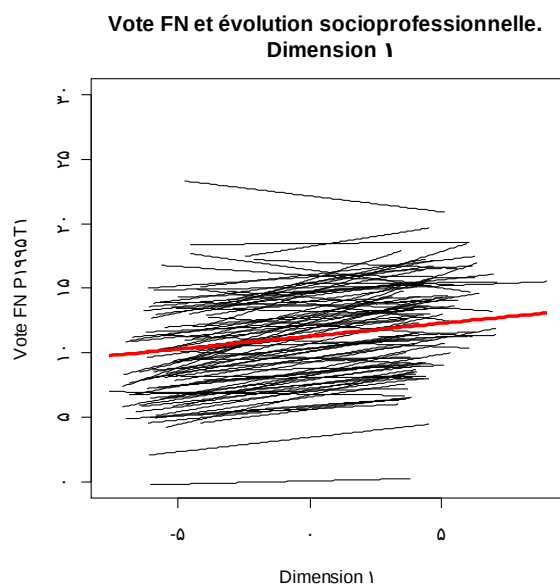
Graphique 20



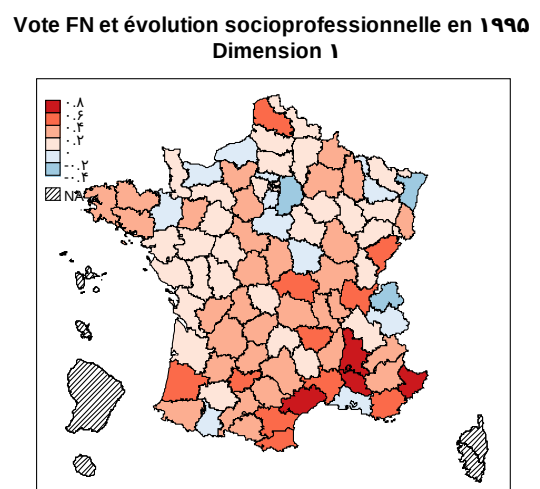
Carte 20

Une part de la difficulté rencontrée pour interpréter certains des résultats qui viennent d'être présentés provient de ce que les facteurs explicatifs envisagés sont purement statiques : ils renvoient à des niveaux, enregistrés par le recensement de 1999, et non à des évolutions, sans doute parfois plus significatives d'un point de vue sociologique. Aussi peut-on tenter de voir ce qu'il en est si l'on intègre cette dimension de l'évolution socio-démographique des territoires considérés.

Comme l'indiquent les résultats des régressions multiniveaux menées (cf. annexe), cette approche possède en 1995 et au premier tour de l'élection présidentielle de 2002 une capacité explicative moindre que celle reposant sur des données statiques. Pour ces deux scrutins, c'est principalement la deuxième composante principale dégagée, qui renvoie à la désindustrialisation, qui doit être considérée. En revanche, lors du deuxième tour de l'élection présidentielle de 2002 ainsi que lors de la présidentielle de 2007, ce modèle dynamique rend mieux compte du vote en faveur de J.-M. Le Pen que le modèle statique, et de plus la première dimension – qui concerne l'urbanisation – joue un rôle important.

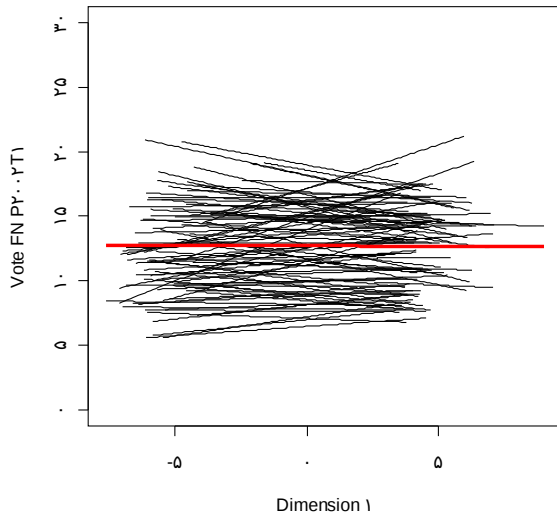


Graphique 21



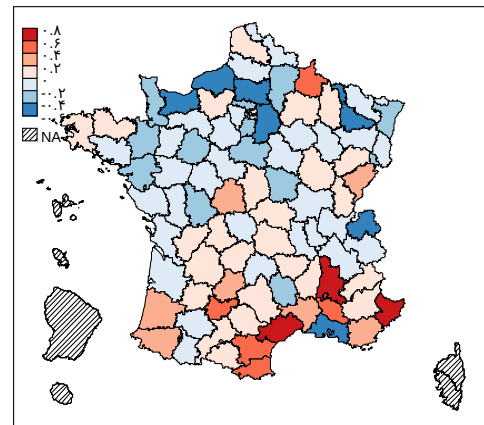
Carte 21

**Vote FN et évolution socioprofessionnelle.
Dimension 1**



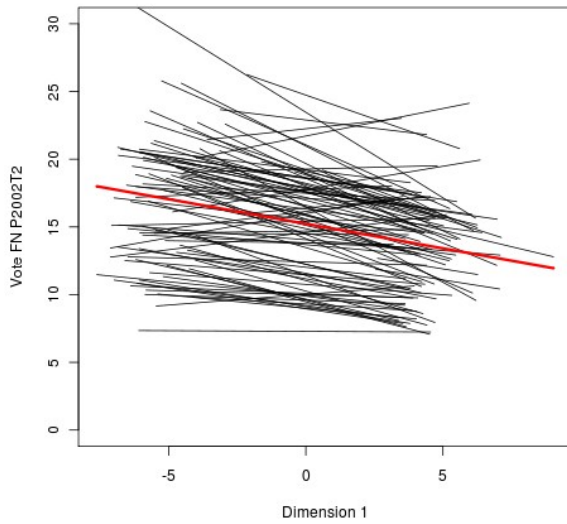
Graphique 22

**Vote FN et évolution socioprofessionnelle en 2002
1er tour - Dimension 1**



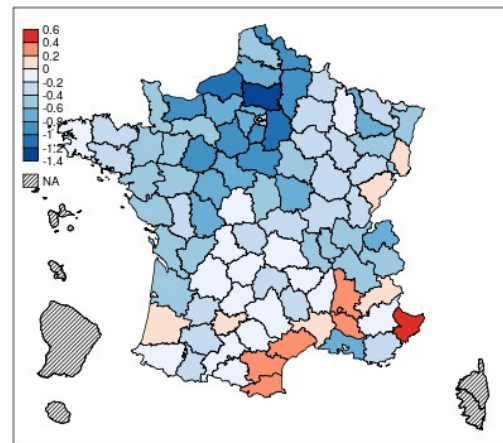
Carte 22

**Vote FN et évolution socioprofessionnelle
Dimension 1 en 2002 au 2ème tour**



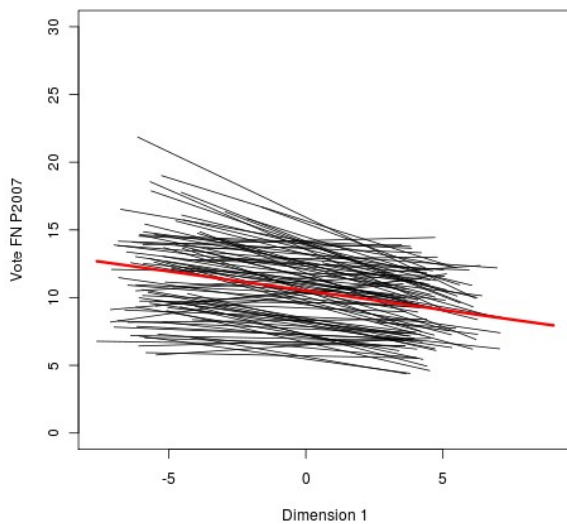
Graphique 23

**Vote FN et évolution socioprofessionnelle
Dimension 1 en 2002 au 2ème tour**



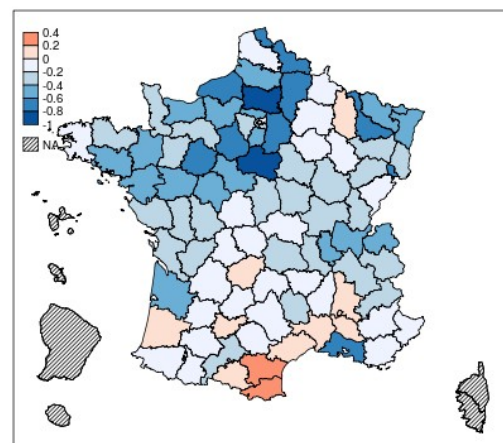
Carte 23

**Vote FN et évolution socioprofessionnelle
Dimension 1 en 2007**



Graphique 24

**Vote FN et évolution socioprofessionnelle
Dimension 1 en 2007**



Carte 24

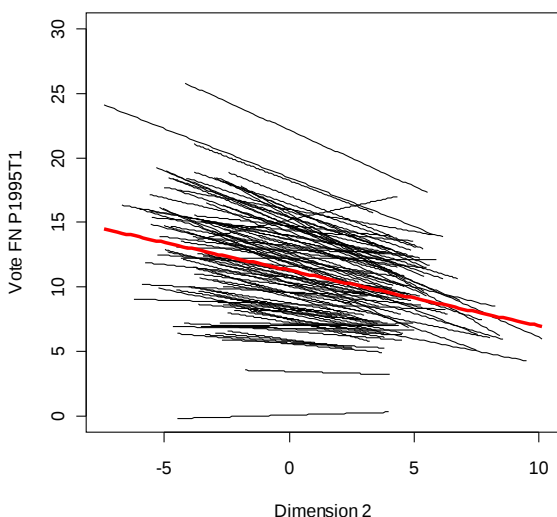
Les graphiques et cartes 21 à 24 illustrent la première dimension (du côté négatif de cette dimension, figurent les communes anciennement rurales, tandis que du côté positif figurent les communes anciennement urbaines). Les graphiques comme les cartes sont très proches de ceux réalisés à partir des données statiques de 1999 – et, de fait, le coefficient de corrélation entre les coefficients par département estimés par ces deux modèles est très élevé (de 0.85 pour 1995 à 0.91 pour 2007). On observe la même évolution – un vote FN de moins en moins urbain – et la même organisation spatiale – opposition entre un grand Nord et Nord Ouest où le vote FN est plutôt extra-urbain et un Sud méditerranéen où il est davantage urbain.

La deuxième dimension de l'évolution sociodémographique – qui concerne la désindustrialisation et la présence d'ouvriers – entretient, elle aussi, des liens avec le vote Front national très comparables à ceux observés pour la deuxième composante socio-professionnelle statique, même si la corrélation est ici un peu moins élevée – elle varie de 0.82 à 0.74. En ce qui concerne ces deux dimensions, l'analyse longitudinale n'apporte donc guère par rapport à une analyse fondée sur des données « en coupe » - tout au plus, elle permet de la confirmer.

La troisième dimension de l'évolution sociodémographique cependant ne correspond à aucune des cinq premières composantes principales issues de l'ACP menée sur les données de 1999 et possède donc une réelle « valeur ajoutée » par rapport à celles-ci. Cette variable connaît peu de variations dans la manière dont elle est liée au vote FN – dans quasi tous les départements, le vote FN est d'autant plus élevé que la commune était autrefois composée d'indépendants et que la part de ceux-ci s'est réduite dans la population active. Cette relation est illustrée par les graphiques et cartes 29 à 32. L'impact de cette variable est relativement important, puisqu'on peut estimer qu'il peut aller, *ceteris paribus*, jusqu'à environ 5 à 10 points de pourcentage – soit une influence comparable à celle de la désindustrialisation. Le vote Front national apparaît ainsi particulièrement élevé dans les territoires touchés par le déclin de certains groupes sociaux – ouvriers, indépendants.

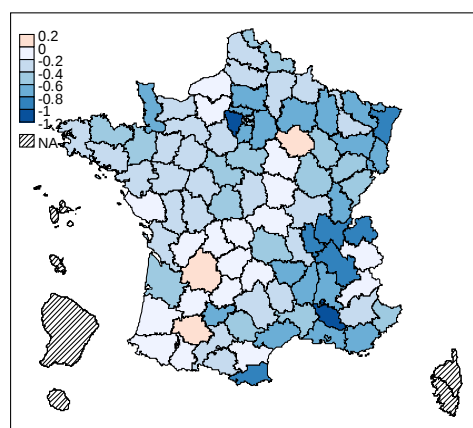
Pour conclure sur l'utilisation de données relatives à la composition et à l'évolution socioprofessionnelle de la population, elles permettent clairement – qu'elles soient saisies à un instant t ou de manière diachronique – de rendre compte d'une composante importante du vote Front national. La composante ouvrière/industrielle est sans doute la plus importante, mais d'autres ne sont pas à négliger non plus, comme le déclin des indépendants, ou l'opposition entre les villes et les zones rurales, d'autant plus intéressante qu'elle joue très différemment dans le temps et l'espace. Il apparaît ici clairement que l'utilisation de modèles multiniveaux, permettant aux coefficients de varier dans l'espace, conduisent à nettement réévaluer la portée explicative de ces variables socioprofessionnelles.

**Vote FN et évolution socioprofessionnelle.
Dimension 2**



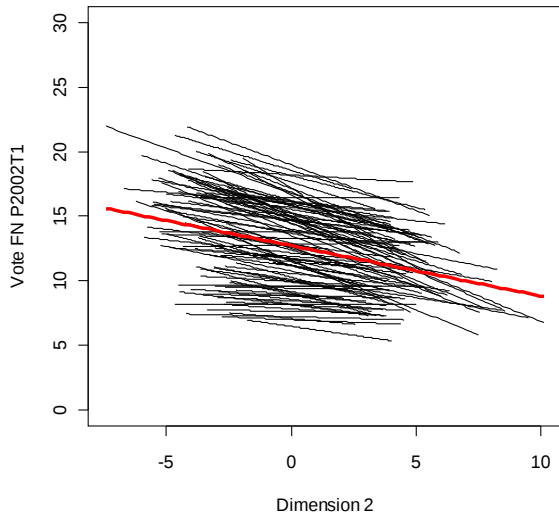
Graphique 25

**Vote FN et évolution socioprofessionnelle en 1995
Dimension 2**



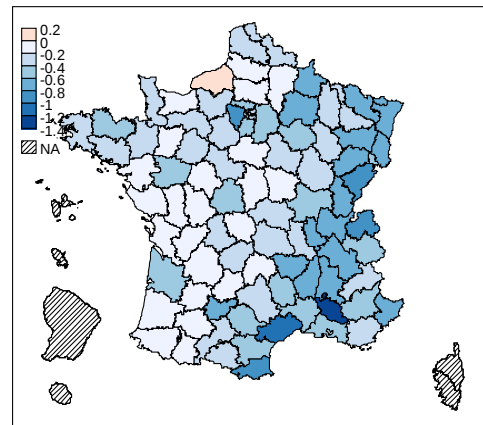
Carte 25

**Vote FN et évolution socioprofessionnelle.
Dimension 2**



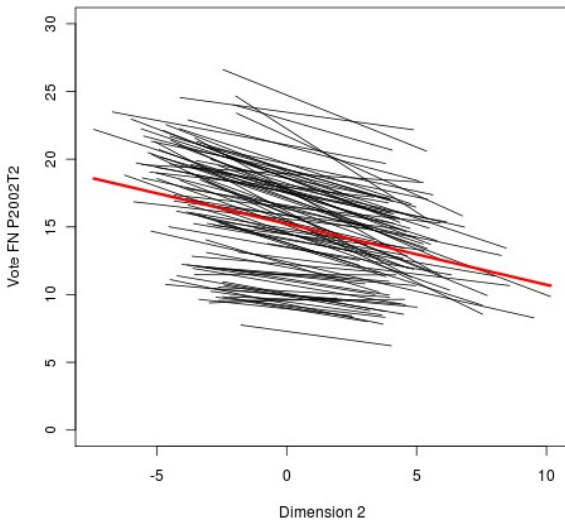
Graphique 26

**Vote FN et évolution socioprofessionnelle en 2002
1er tour - Dimension 2**



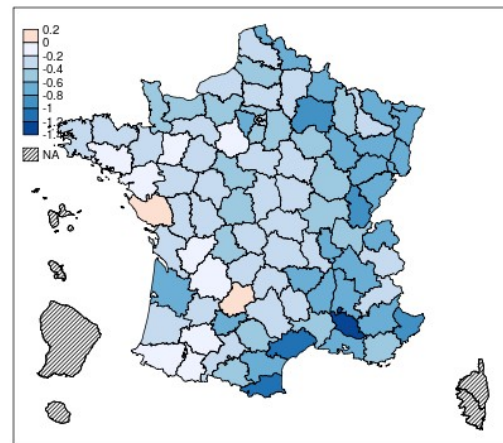
Carte 26

**Vote FN et évolution socioprofessionnelle
Dimension 2 en 2002 au 2ème tour**



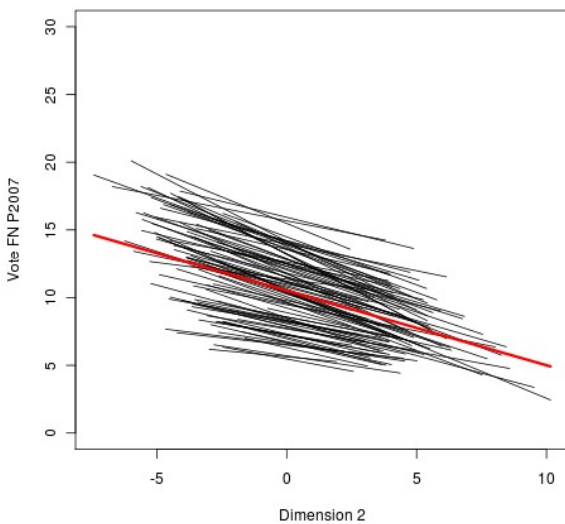
Graphique 27

**Vote FN et évolution socioprofessionnelle
Dimension 2 en 2002 au 2ème tour**



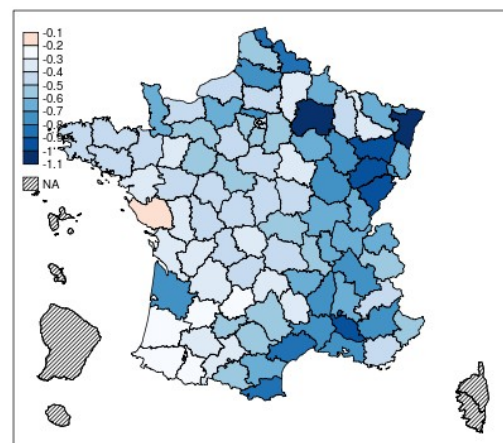
Carte 27

**Vote FN et évolution socioprofessionnelle
Dimension 2 en 2007**



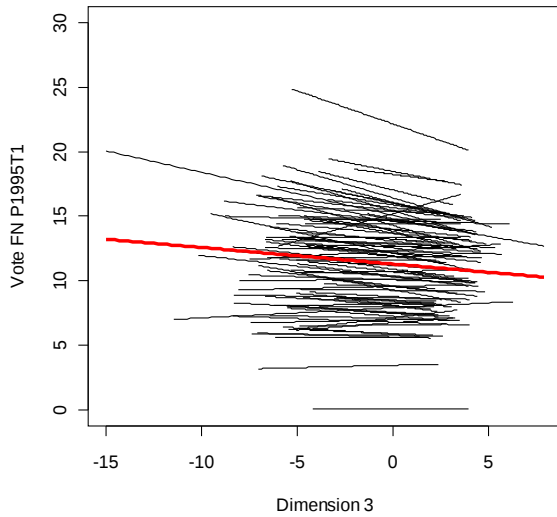
Graphique 28

**Vote FN et évolution socioprofessionnelle
Dimension 2 en 2007**



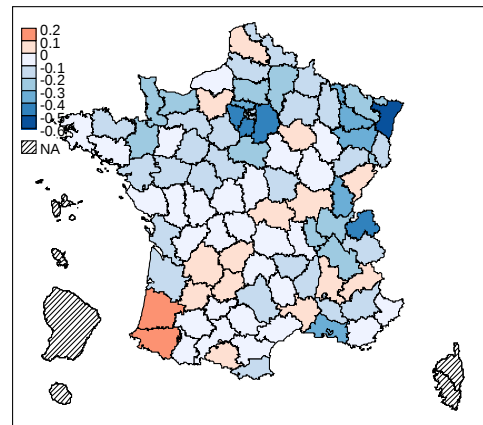
Carte 28

**Vote FN et évolution socioprofessionnelle.
Dimension 3**



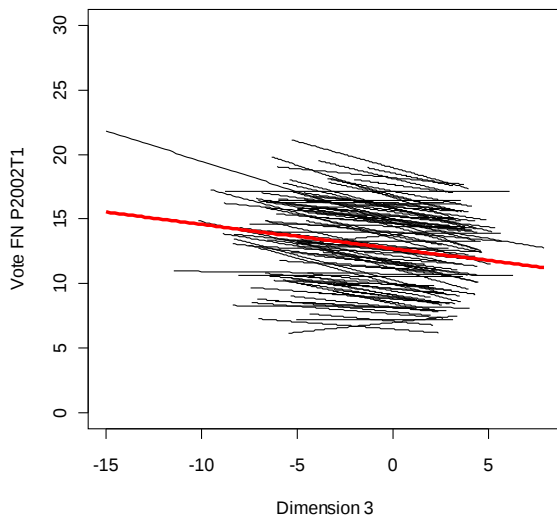
Graphique 29

**Vote FN et évolution socioprofessionnelle en 1995
Dimension 3**



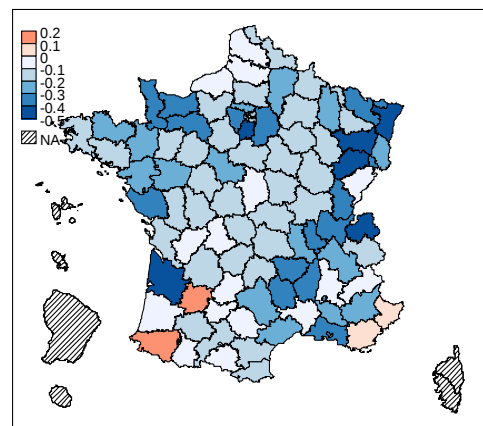
Carte 29

**Vote FN et évolution socioprofessionnelle.
Dimension 3**



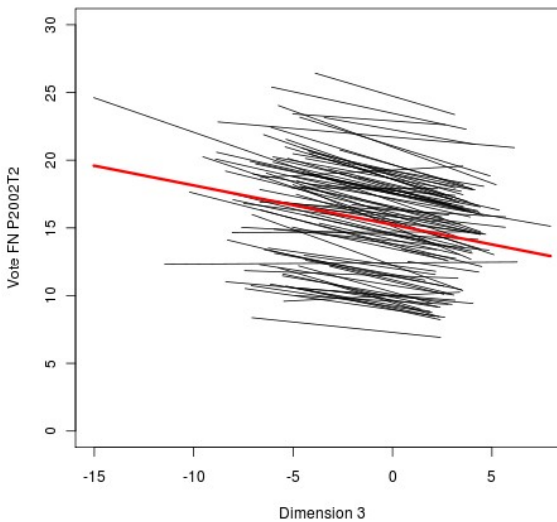
Graphique 30

**Vote FN et évolution socioprofessionnelle en 2002
1er tour - Dimension 3**



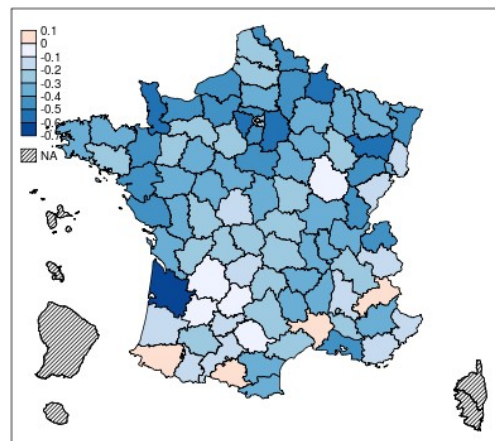
Carte 30

**Vote FN et évolution socioprofessionnelle
Dimension 3 en 2002 au 2ème tour**

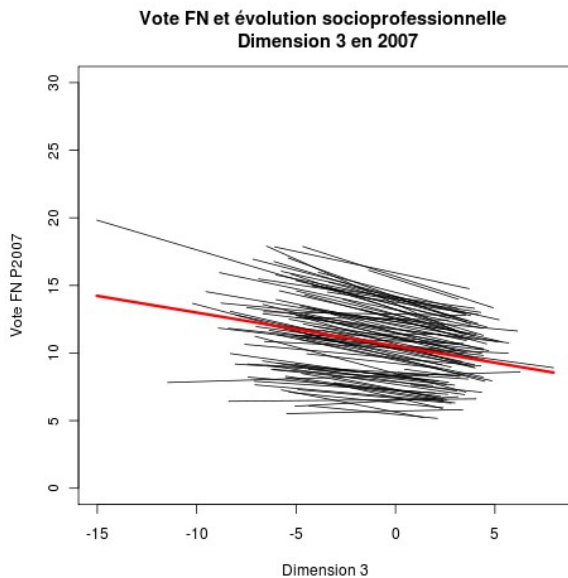


Graphique 31

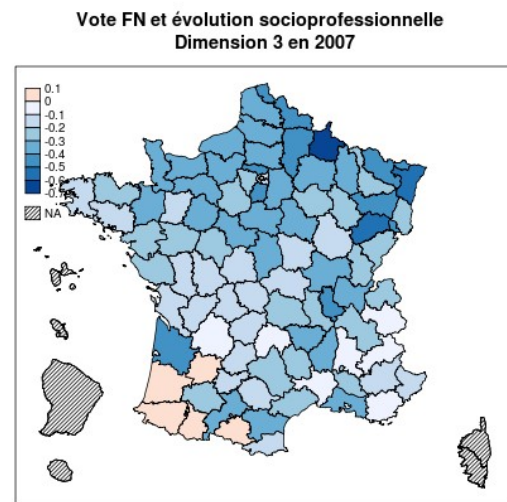
**Vote FN et évolution socioprofessionnelle
Dimension 3 en 2002 au 2ème tour**



Carte 31



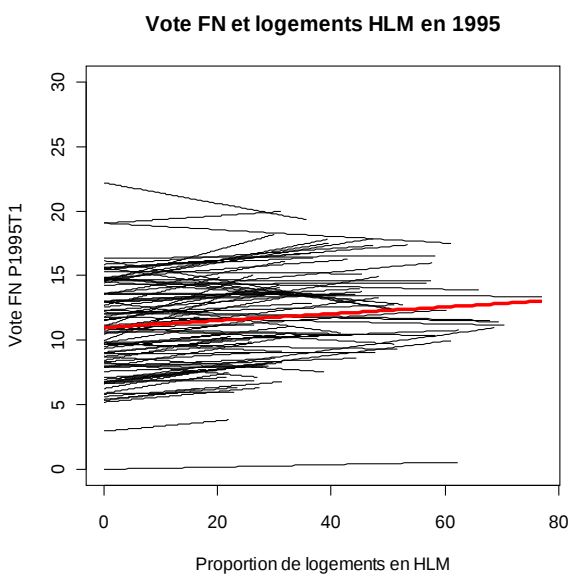
Graphique 32



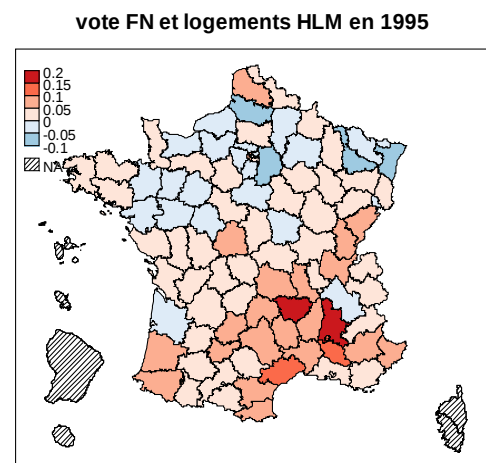
Carte 32

Examinons maintenant les autres variables retenues. Il conviendra de les interpréter avec plus de prudence, car elles sont étudiées séparément, et ne sauraient donc faire l'objet d'interprétations invoquant la clause *ceteris paribus*.

En premier lieu, la proportion de logements HLM. Les discours de sens commun les associent souvent au vote en faveur du Front national, en arguant notamment de la cohabitation entre « autochtones » des classes populaires et « immigrés » qui serait de nature à provoquer des tensions propices au vote frontiste. En réalité, comme nous le montrent les résultats du modèle multiniveau construit, ainsi que les graphiques et cartes 33 à 36, cette variable ne rend compte que dans une faible mesure du vote frontiste. De plus, si en 1995 l'association est effectivement positive entre proportion de HLM et vote FN, en particulier dans la moitié sud de la France, la tendance s'inverse à partir de 2002, en particulier au second tour, et le vote FN devient alors négativement corrélé aux logements HLM, dans quasiment toute la France à l'exception de quelques départements du Sud-Est. C'est d'ailleurs à partir du deuxième tour de l'élection présidentielle de 2002 que l'association entre logements HLM et vote FN devient véritablement importante.

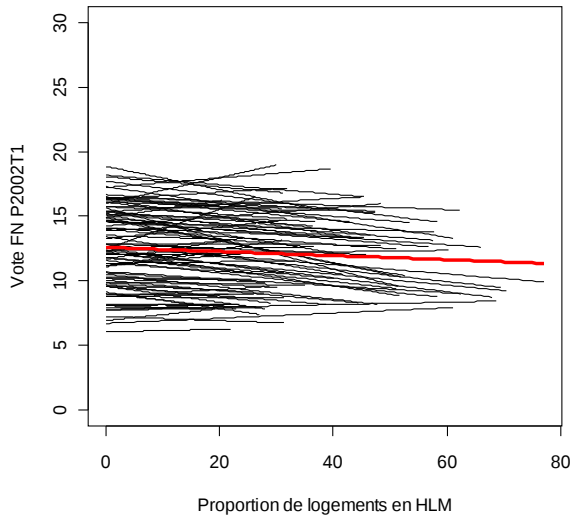


Graphique 33



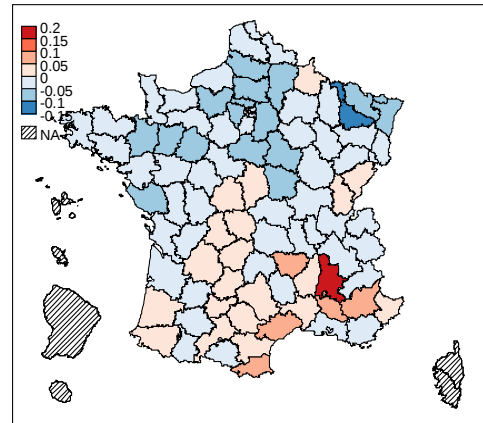
Carte 33

Vote FN et logements HLM en 2002 - 1er tour



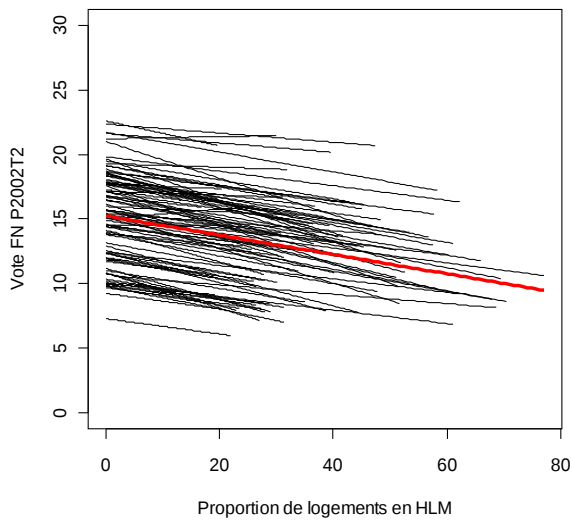
Graphique 34

vote FN et logements HLM en 2002 - 1er tour



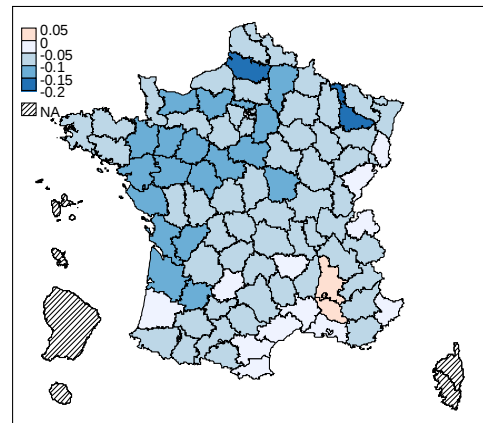
Carte 34

Vote FN et logements HLM en 2002 - 2ème tour



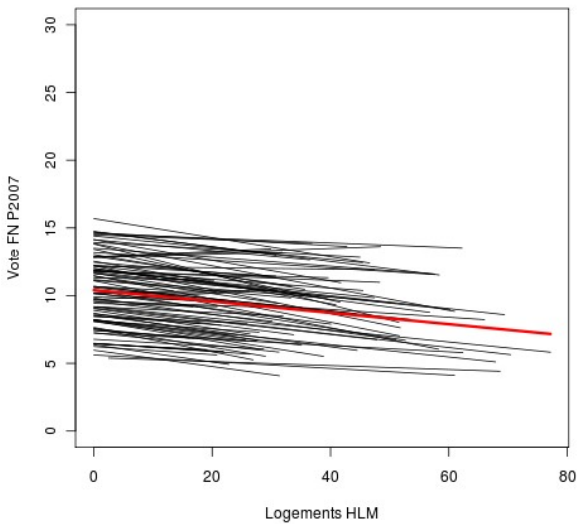
Graphique 35

vote FN et logements HLM en 2002 - 2ème tour



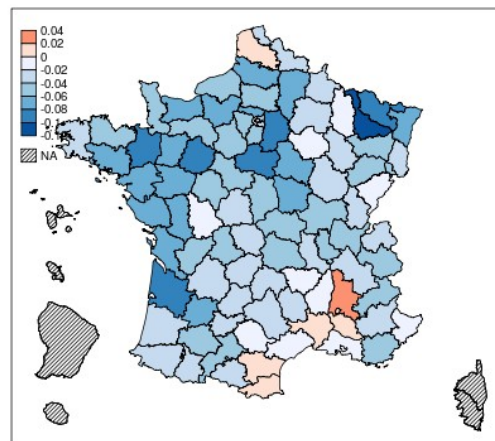
Carte 35

Vote FN et logements HLM en 2007



Graphique 36

Vote FN et logements HLM en 2007

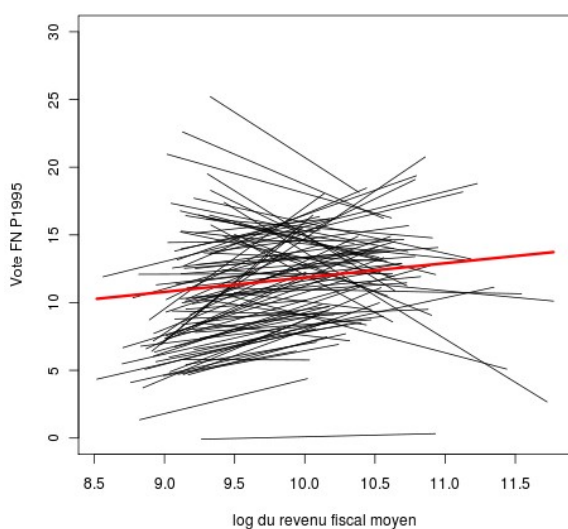


Carte 36

Le revenu moyen imposable par commune est assez rarement utilisé dans la littérature française électoraliste, pourtant il s'agit sur le plan théorique d'une variable importante puisqu'elle est liée à l'hypothèse du « vote du porte-monnaie³⁶ » - même s'il est vrai qu'il s'agit ici typiquement d'une variable contextuelle : l'agrégation de valeurs individuelles prend probablement un sens qui dépasse la simple moyenne de ces valeurs. En d'autres termes, l'inférence du niveau agrégé au niveau individuel ne doit pas se faire sans éléments causaux supplémentaires. Mais en tout état de cause, il s'agit d'une variable contextuelle extrêmement intéressante.

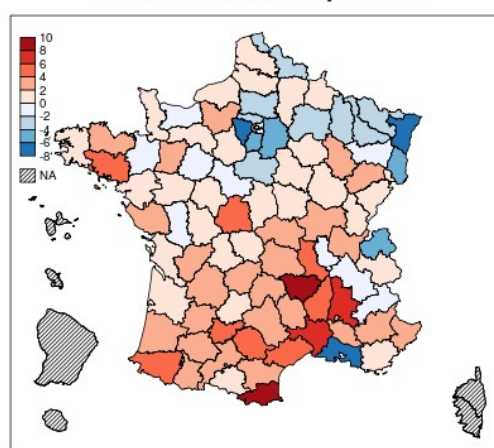
Si on considère l'effet de cette variable seule (graphiques et cartes 37 à 40), on constate qu'il est loin d'être négligeable. Deux observations s'imposent. D'abord, l'influence apparente moyenne du revenu fiscal moyen évolue fortement dans le temps : si en 1995 il semble y avoir une légère corrélation positive entre cette variable et le vote FN (ligne rouge sur le graphique 37), cette relation s'inverse progressivement, pour devenir nettement négative lors du deuxième tour de l'élection présidentielle de 2002 et en 2007.

Vote FN et revenu fiscal moyen en 1995



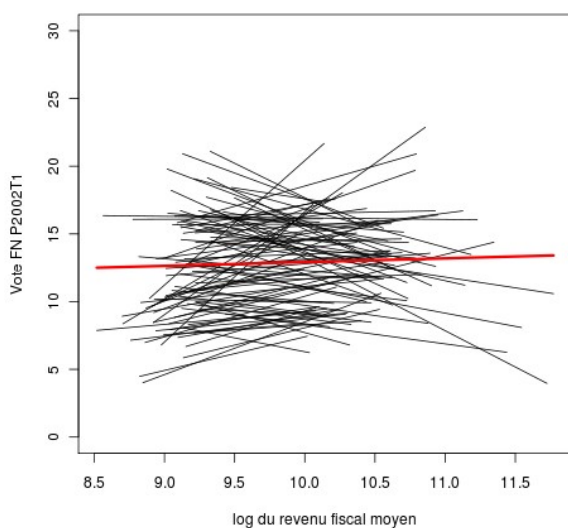
Graphique 37

Vote FN et revenu fiscal moyen en 1995



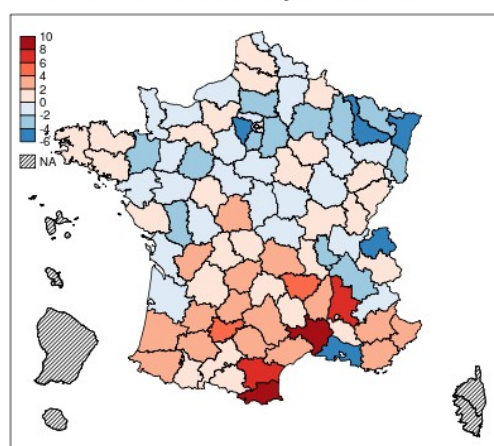
Carte 37

Vote FN et revenu fiscal moyen en 2002 - 1er tour



Graphique 38

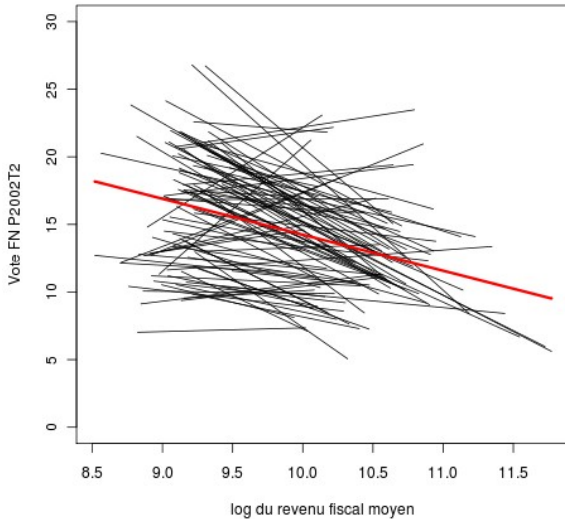
Vote FN et revenu fiscal moyen en 2002 - 1er tour



Carte 38

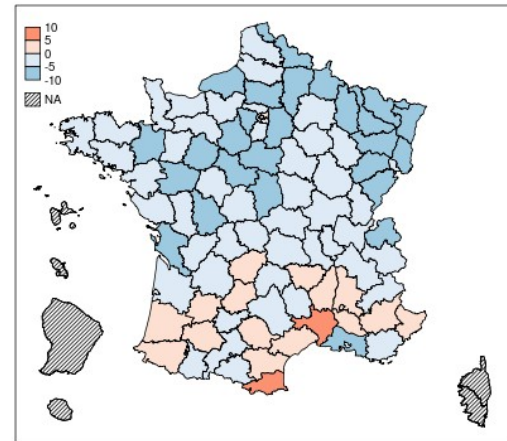
³⁶Michael Lewis-Beck, « Le vote du 'porte-monnaie' en question », in Daniel Boy, Nonna Mayer (dir.), *L'électeur en questions*, Paris, Presses de la FNSP, 1997, p. 239-261.

Vote FN et revenu fiscal moyen en 2002 - 2ème tour



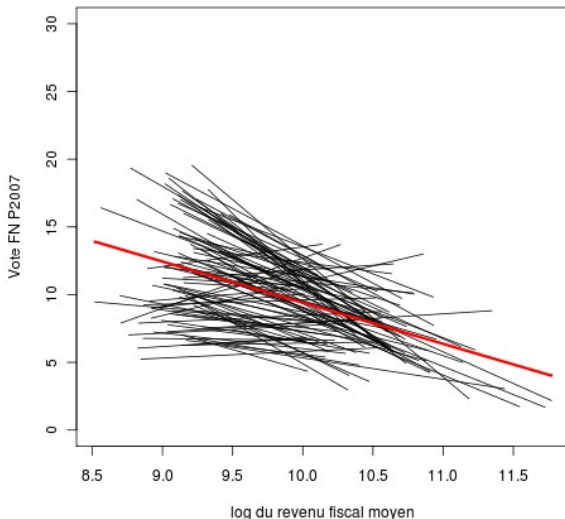
Graphique 39

Vote FN et revenu fiscal moyen en 2002 - 2ème tour



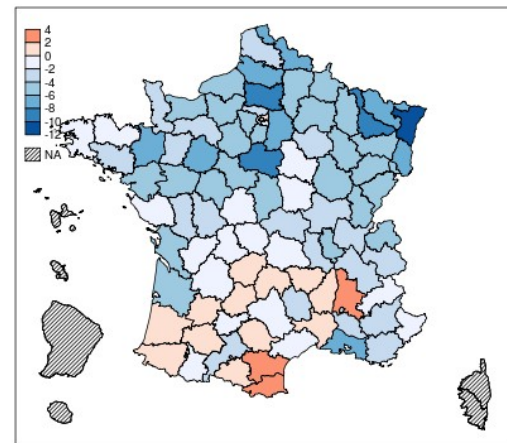
Carte 39

Vote FN et revenu fiscal moyen en 2007



Graphique 40

Vote FN et revenu fiscal moyen en 2007

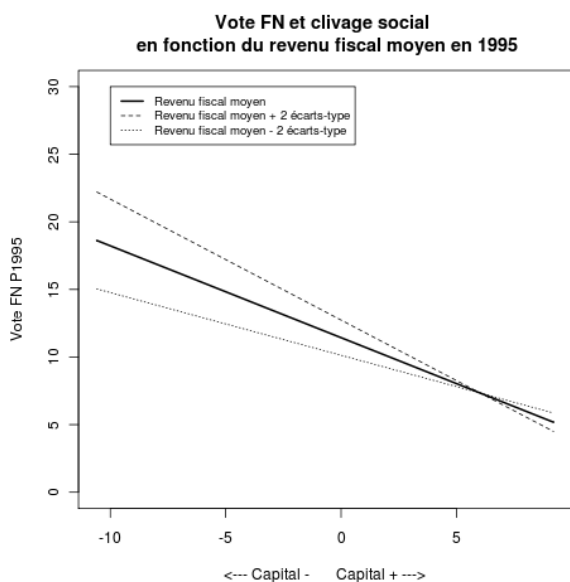


Carte 40

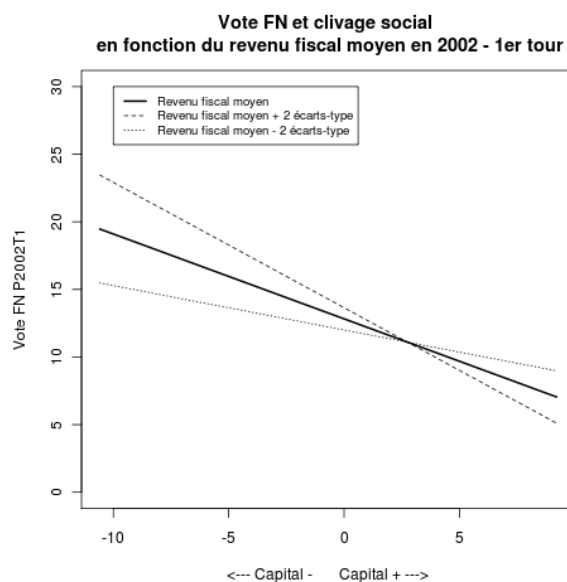
Ensuite, il convient de relever que ces moyennes apparentes recouvrent en réalité des variations très importantes et des changements de signe d'un département à l'autre. De plus ces variations sont très spatialisées : le vote FN décroît en fonction du revenu moyen dans les départements de l'Est de la France et du bassin parisien, quand il varie de manière inverse dans le Sud de la France et notamment sur le pourtour méditerranéen (à l'exception des Bouches-du-Rhône). Cette structure spatiale est assez stable : entre 1995 et 2007 le coefficient de corrélation est de 0.71.

On peut toutefois se demander dans quelle mesure le revenu moyen par commune ne recouvre pas sa structure socio-professionnelle, et comment ces deux éléments interagissent le cas échéant. Sans épuiser cette question ici, on a construit un modèle prenant en compte la deuxième composante principale issue de l'ACP menée sur la composition socio-professionnelle des communes en 1999 (celle relative au caractère plus ou moins ouvrier d'une commune), le revenu fiscal moyen et l'interaction entre ces deux termes. Cela permet non seulement d'isoler les effets propres de chacune de ces variables, mais aussi d'analyser comment elles interagissent.

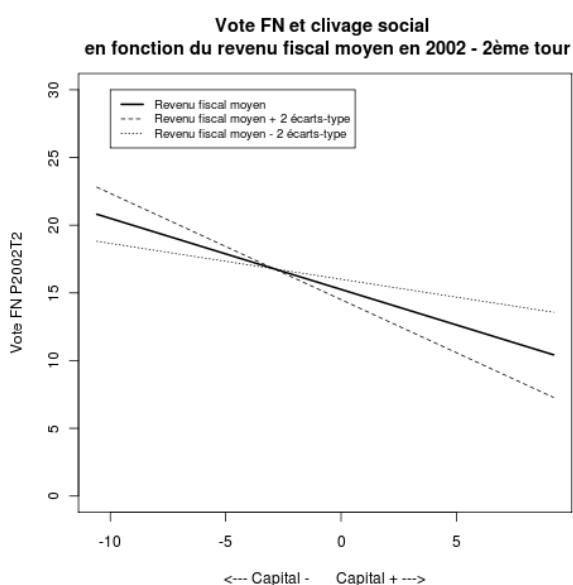
Les résultats de ce modèle montrent que les deux variables exercent chacune une influence propre, mais aussi que les effets d'interaction entre elles sont très importants. En fait, cela permet de mieux comprendre comment ces deux variables s'articulent et comment leurs interactions évoluent dans le temps.



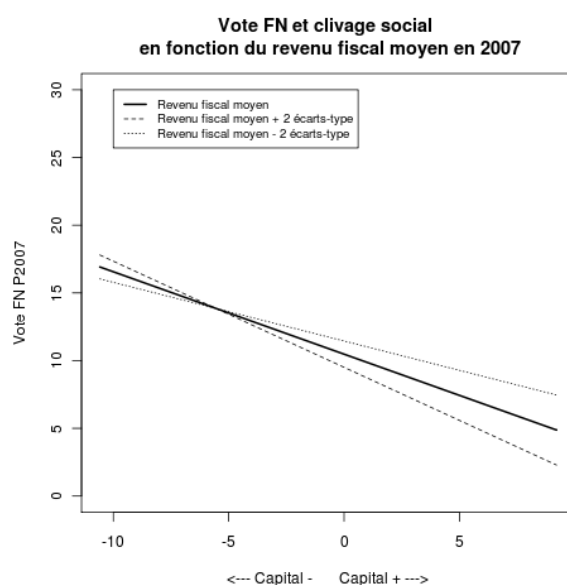
Graphique 41



Graphique 42

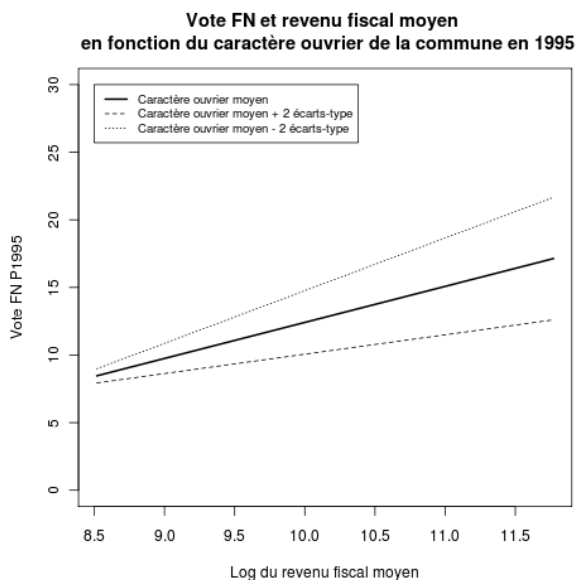


Graphique 43

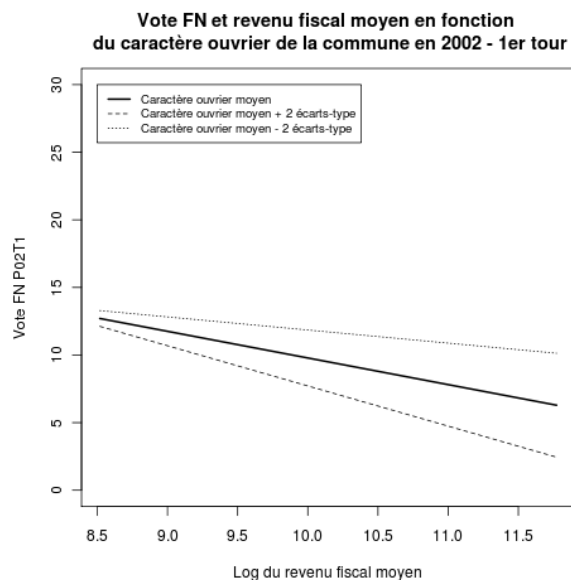


Graphique 44

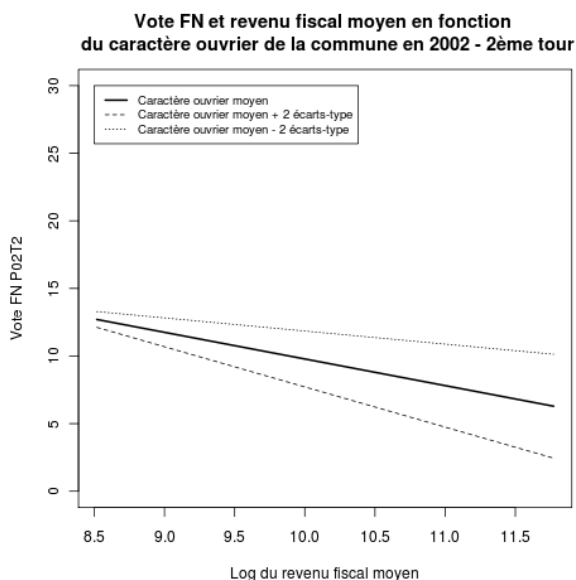
Les graphiques 41 à 44 figurent l'effet d'interaction entre le caractère plus ou moins ouvrier d'une commune et son niveau de revenu fiscal moyen. Comme on le voit, la corrélation entre caractère ouvrier de la commune et vote FN est d'autant plus forte que le revenu fiscal moyen est élevé. Dès lors, passé un certain seuil, dans les communes les moins ouvrières, à caractère ouvrier maintenu égal, le vote FN devient une fonction décroissante du revenu fiscal moyen. En d'autres termes, dans les communes les plus ouvrières, le vote FN croît en même temps que le revenu moyen, tandis que dans les communes les moins ouvrières, le vote FN décroît quand le revenu moyen augmente. On voit en quoi le niveau de revenu joue ici comme une variable contextuelle : il permet de caractériser les communes en fonction de leur niveau d'inégalité socio-économique. On peut ainsi penser que les ouvriers votent d'autant plus en faveur du Front national qu'ils vivent dans un contexte de fortes inégalités socio-économiques. Mais le seuil de basculement (le point où les droites se croisent, sur les graphiques) évolue au cours du temps : de 1995 à 2007, il se déplace vers les communes les plus ouvrières. Cela signifie que l'effet lié aux inégalités socio-économiques ne joue que dans des communes de plus en plus uniformément ouvrières.



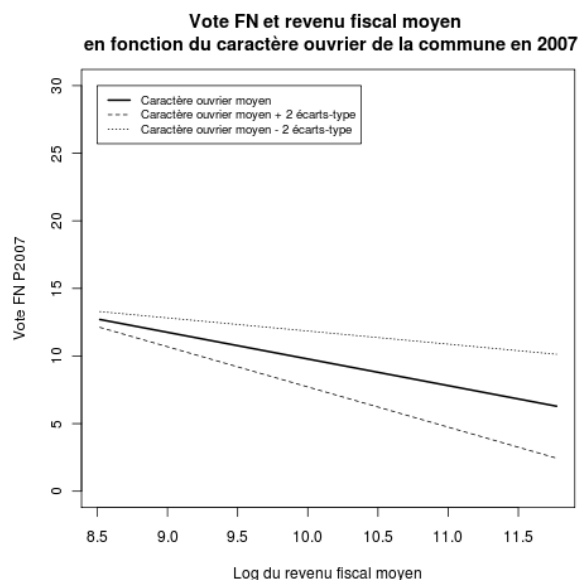
Graphique 45



Graphique 46



Graphique 47



Graphique 48

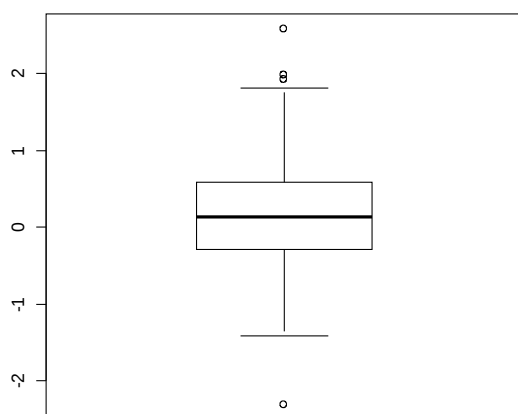
Une autre manière d'appréhender cette interaction est figurée sur les graphiques 45 à 48, qui figurent la relation entre revenu fiscal moyen d'une commune et vote FN en fonction du caractère ouvrier de celle-ci³⁷. La première confirmation est que, dans tous les cas de figure, à revenu fiscal moyen constant, les communes les plus ouvrières votent davantage en faveur du Front national. Cependant cette relation est d'autant plus forte que le niveau de revenu fiscal moyen de la commune est élevé : on retrouve bien ici l'effet de catalyseur que le niveau d'inégalité socio-économique de la commune joue. Enfin, à caractère ouvrier constant, le vote FN varie de manière croissante avec le revenu fiscal moyen en 1995 (mais ce, d'autant plus que la commune est ouvrière), tandis qu'à partir de 2002, cette variable n'exerce qu'un effet quasi-nul dans les communes les plus ouvrières, tandis que dans les communes les moins ouvrières, le vote FN décroît quand le revenu fiscal augmente. Cela permet ainsi de mieux comprendre les cartes 37 à 40 : les départements en rouge (ou en bleu pâle) sont, tendanciellement, ceux qui comptent le moins d'ouvriers. On pourrait en réalité encore complexifier le modèle d'interaction, ou plutôt son interprétation, puisqu'on n'a ici que pris en compte les effets fixes – la « tendance nationale », si on

³⁷Attention, la composante principale prise en compte étant négativement associée à la proportion d'ouvriers, les communes étiquetées « caractère ouvrier moyen – 2 écarts-type » sont en réalité les plus ouvrières.

préfère. Les choses se configurent encore différemment dans chaque département. Les résultats des modèles multiniveaux indiquent cependant que les variations des effets d'interaction sont faibles.

Abordons enfin le dernier type de modèles que nous avons construit, celui qui repose sur le type de communes (selon la nomenclature ZAUER de l'INSEE, qui distingue centres urbains, communes monopolarisées, communes multipolarisées et communes rurales). Ces modèles permettront d'affiner ce qui a déjà été montré *supra* concernant le clivage urbain-rural. Par souci de simplicité, et afin de tester le plus directement possible l'hypothèse de Jacques Lévy selon laquelle les communes périurbaines votent moins en faveur du FN que les centres urbains, on n'analyse ici que la différence entre le vote moyen estimé des centres urbains et des communes périurbaines³⁸ par département. Les graphiques 49 à 52³⁹ et les cartes 41 à 44 en présentent les résultats. Ceux-ci permettent d'affiner l'évolution présentée plus haut : en 1995, si le vote FN était davantage un vote urbain qu'un vote rural, en revanche au plan national il n'y avait pas de différence significative entre les centres urbains et les communes périurbaines. En Ile-de-France et dans l'Est toutefois, les communes urbaines votaient significativement plus pour le FN que les centres urbains. Cette tendance se confirme lors des scrutins suivants, et l'écart de vote FN entre communes périurbaines et centres urbains s'accroît dans presque tous les départements. Toutefois, le pourtour méditerranéen et le sillon rhodanien se distinguent de cette tendance ; les centre-villes y votent souvent davantage pour le Front que leurs couronnes périurbaines – ce qui renvoie aux multiples acceptions du périurbains et aux différents modèles urbains, des éléments curieusement absents des travaux de Jacques Lévy.

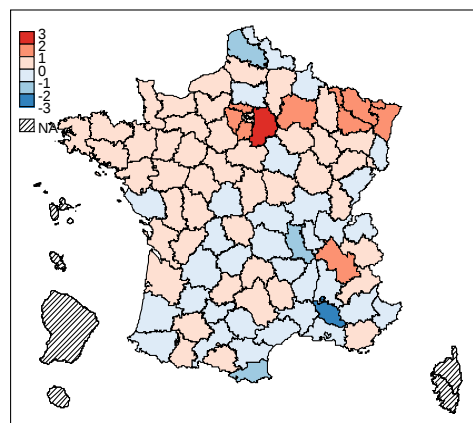
Le vote Le Pen dans les centres urbains et les communes périurbaines



Écart en point de %, par département, 1995

Graphique 49

Différentiel entre vote FN des communes périurbaines et des centres urbains

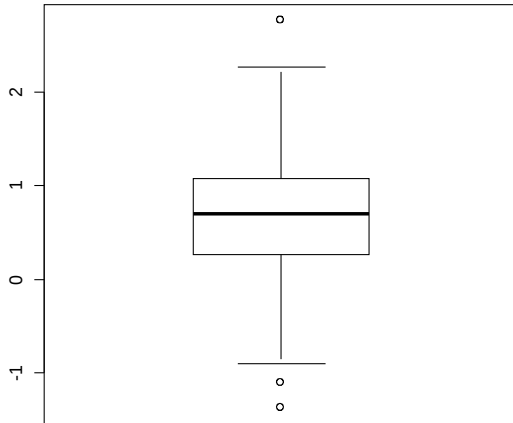


Carte 41

³⁸Ce terme étant ici restreint aux communes monopolarisées.

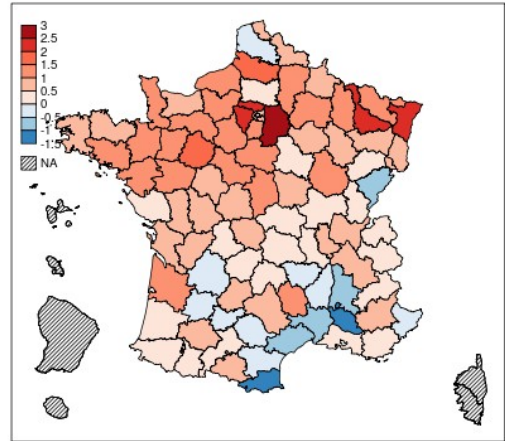
³⁹Sur ces « boîtes à moustaches », la ligne centrale représente la moyenne, les limites de la boîte centrale correspondent aux premier et troisième quartile, et les extrémités des moustaches sont situées à 1,5 fois la hauteur de la boîte centrale. Les points restants sont des *outliers*.

Le vote Le Pen dans les centres urbains et les communes périurbaines



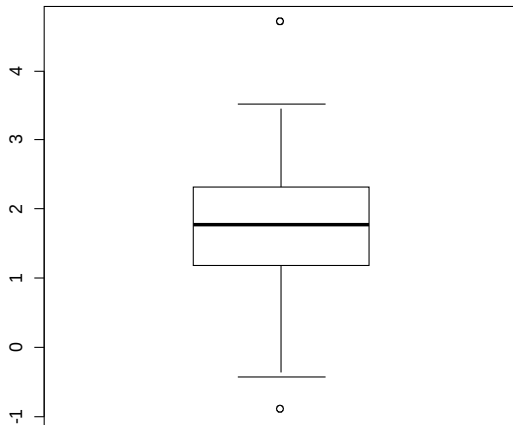
Ecart en point de %, par département, 2002, 1er tour
Graphique 50

Différentiel entre vote FN des communes périurbaines et des centres urbains en 2002 - 1er tour



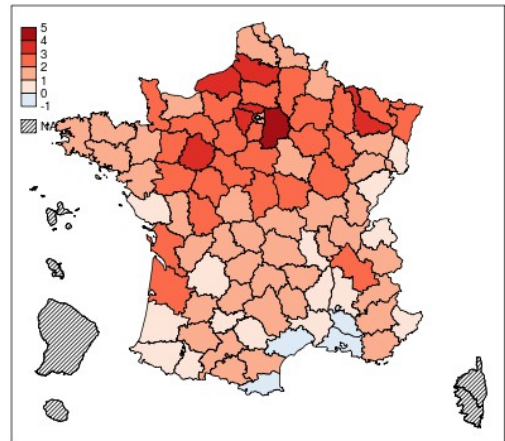
Carte 42

Le vote Le Pen dans les centres urbains et les communes périurbaines



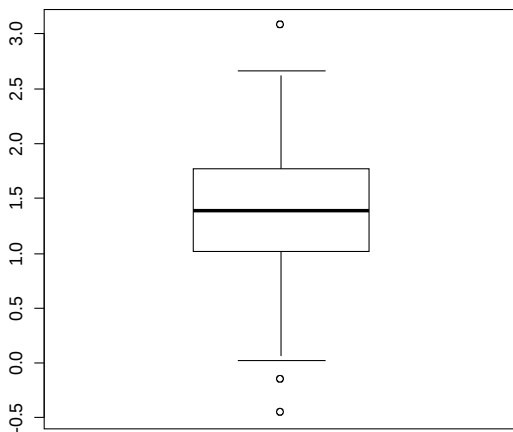
Ecart en point de %, par département, 2002, 2ème tour
Graphique 51

Différentiel entre vote FN des communes périurbaines et des centres urbains en 2002 - 2ème tour



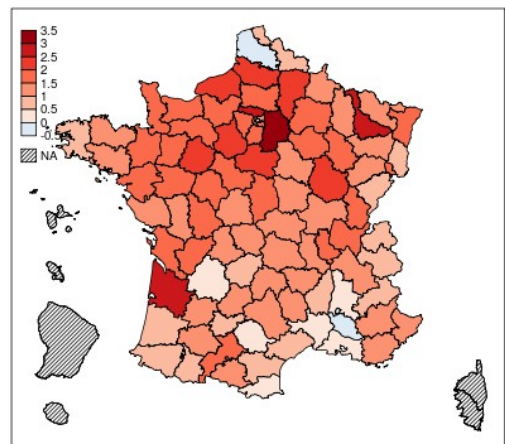
Carte 43

Le vote Le Pen dans les centres urbains et les communes périurbaines



Ecart en point de %, par département, 2007
Graphique 52

Différentiel entre vote FN des communes périurbaines et des centres urbains en 2007



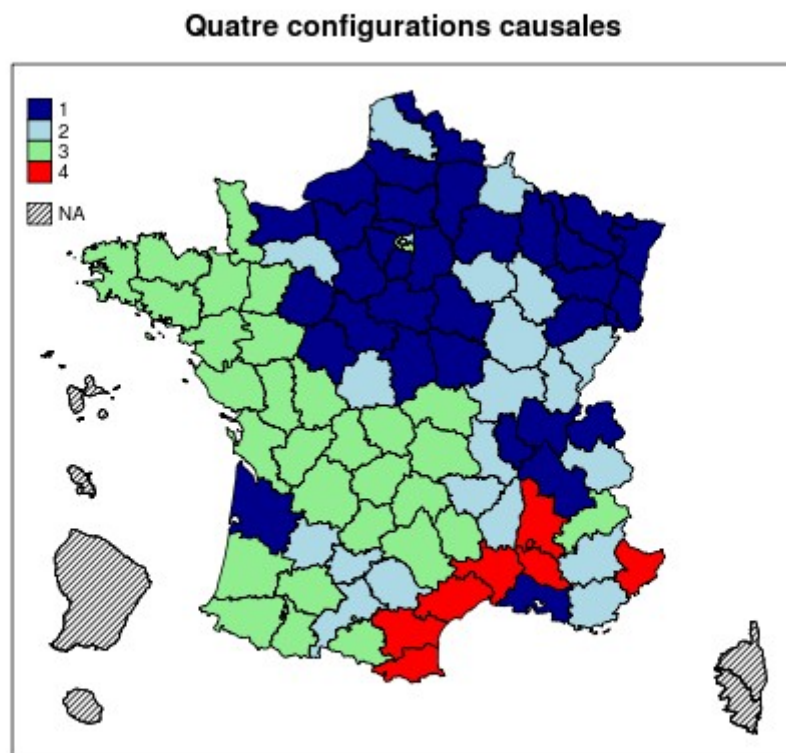
Carte 44

Au terme de cette recherche, tentons un instant de synthétiser ses résultats. De nombreux paramètres ont été estimés, de nombreux graphiques et cartes proposés. Qu'en retenir, qu'en conclure ?

Une analyse en composantes principales (ACP) menée sur la quasi-totalité⁴⁰ des coefficients obtenus permet d'organiser l'ensemble de l'information produite. Ses deux premières composantes résument 59 % de l'information contenue dans les coefficients estimés. Ces deux premières composantes permettent au final de distinguer, au moyen d'une classification automatique (K-means), quatre types de configurations causales menant au vote en faveur du FN dans une commune.

La première configuration causale caractérise des départements où les communes dans lesquelles le vote FN est le plus élevé sont plutôt rurales ou multipolarisées, mais néanmoins assez ouvrières, ayant connu le déclin des paysans et des indépendants, dans lesquelles le revenu fiscal moyen est faible, avec un parc HLM important. Dans ces départements, le vote FN y est plus élevé que la moyenne, et il s'agit de terres majoritairement déchristianisées⁴¹, à l'exception de l'Est alsacien. Les départements appartenant à cette classe sont situés dans la moitié Nord-Nord Est de la France ; il faut cependant y ajouter les Bouches-du-Rhône, la Gironde et une partie de la région Rhône-Alpes. Ces départements sont figurés en bleu foncé sur la carte 45. On pourrait dire, de manière lapidaire, qu'ici le vote frontiste est lié au déclin industriel et à l'effondrement de la France issue de la révolution industrielle.

La classe 2 constitue un type intermédiaire entre les trois autres ; ses principales caractéristiques sont moyennes par rapport aux autres.



Carte 45

⁴⁰Pour éviter des redondances avec d'autres coefficients, on n'a retenu des modèles d'interaction clivage social-revenu fiscal moyen que le coefficient de l'interaction.

⁴¹On a utilisé, comme variables illustratives – et ne rentrant donc pas dans l'analyse – la pratique religieuse telle que mesurée par le chanoine Boulard dans les années 1950 (François-André Isambert et Jean-Paul Terrenoire, *Atlas de la pratique religieuse des catholiques en France*, Paris, Presses de la FNSP/Éditions du CNRS, 1980) d'une part, la part des enfants scolarisés dans l'enseignement privé en 2007 d'autre part.

La classe 3 est constituée des départements dans lesquels le vote FN est, en moyenne, le plus faible. On reconnaît, en vert clair, sur la carte 45, la France de l'Ouest, rurale et classiquement conservatrice, de tradition chrétienne. Dans ces départements, le vote FN est le plus élevé là où la réalité sociale s'éloigne de cette image traditionnelle, là où l'économie « productive » est implantée, parfois s'est implantée au cours des quarante dernières années, là où les changements les plus profonds se sont fait sentir. Ici, le vote Front national apparaît donc comme étant lié à – une conséquence ? – la modernisation, au changement.

Enfin, le quatrième type de configuration causale, qui ne rassemble que sept départements, tous situés dans le Midi méditerranéen (Pyrénées-Orientales, Aude, Hérault, Gard, Vaucluse, Drôme et Alpes-Maritimes) – le Midi rouge déchristianisé en somme, à l'exception des Bouches-du-Rhône –, se caractérise par son vote FN plus élevé que la moyenne, un vote FN urbain ou périurbain mais pas rural, un vote FN des ouvriers résidant dans les zones les plus désindustrialisées, qui par le « miracle » de la spéculation foncière, se retrouvent être aussi des zones où vivent des ménages à hauts revenus. Dans cette configuration causale, le vote FN croît en même temps que le revenu fiscal moyen, c'est-à-dire qu'il croît en même temps que les inégalités socio-économiques dans une même commune deviennent plus insupportables. Ici, le vote frontiste n'est pas tant lié au déclin industriel, et encore moins à un développement qui serait venu déranger les équilibres traditionnels, mais plutôt à l'exacerbation des tensions dans des territoires qui connaissent les inégalités les plus criantes du pays⁴², comme l'avait senti Christophe Traïni⁴³.

On le voit, vouloir rapporter un phénomène politique complexe comme le vote en faveur du Front national à une seule cause, un seul facteur explicatif, revient à se lancer à la poursuite de la pierre philosophale. Une analyse certes provisoire mais néanmoins relativement poussée de données écologiques montre qu'il existe au moins trois grands types de configurations causales liées au vote frontiste, qui se distinguent parfaitement les unes des autres. Mais il y a fort à parier que s'il en est ainsi pour le Front national, il en est de même pour d'autres forces politiques, peut-être moins sollicitées par les travaux des politistes ces dernières années.

⁴²Olivier Pégaz-Blanc et Marjorie Martin, « Marseille-Aix-en-Provence, Nice, Toulon et Avignon : des revenus peu élevés et plus dispersés que dans les autres aires urbaines », *Sud INSEE l'essentiel*, 95, 2006, http://www.insee.fr/fr/insee_regions/provence/themes/sud_essentiel/sie95/sie95.pdf.

⁴³Christophe Traïni, « L'épicentre d'un 'séisme électoral'. Le vote Front national en région PACA », in Christophe Traïni (dir.), *Vote en PACA : les élections 2002 en Provence-Alpes-Côte d'Azur*, Paris, Khartala, 2004.

Annexe : résultats des régressions multiniveaux

1°) Premier tour de l'élection présidentielle de 1995

a) Modèle vide

```
Linear mixed model fit by REML
Formula: dataset$"LePen%insP95" ~ 1 + (1 | codedpt)
Data: dataset
   AIC   BIC logLik deviance REMLdev
200291 200317 -100143  200285  200285
Random effects:
Groups   Name             Variance Std.Dev.
codedpt  (Intercept)  14.287   3.7798
Residual                    15.983   3.9979
Number of obs: 35611, groups: codedpt, 93
```

```
Fixed effects:
              Estimate Std. Error t value
(Intercept)  11.0771    0.3926   28.21
```

b) Modèle avec CSP statique

```
Linear mixed model fit by REML
Formula: dataset$"LePen%insP95" ~ Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs +
Dim5pcs + (1 + Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs | codedpt)
Data: dataset
   AIC   BIC logLik deviance REMLdev
197409 197646 -98676  197327  197353
Random effects:
Groups   Name             Variance Std.Dev. Corr
codedpt  (Intercept)  12.909453  3.59297
          Dim1pcs    0.125718  0.35457   0.082
          Dim2pcs    0.180098  0.42438  -0.748 -0.178
          Dim3pcs    0.054264  0.23295  -0.244 -0.226  0.339
          Dim4pcs    0.028045  0.16747  -0.055  0.247  0.115  0.084
          Dim5pcs    0.035774  0.18914   0.284 -0.116  0.016  0.049  0.408
Residual                    14.515123  3.80987
Number of obs: 35611, groups: codedpt, 93
```

```
Fixed effects:
              Estimate Std. Error t value
(Intercept)  11.289511    0.374620  30.136
Dim1pcs       0.390463    0.040967   9.531
Dim2pcs      -0.620365    0.048489 -12.794
Dim3pcs      -0.006895    0.031181  -0.221
Dim4pcs      -0.044187    0.026008  -1.699
Dim5pcs       0.002699    0.029542   0.091
```

```
Correlation of Fixed Effects:
      (Intr) Dm1pcs Dm2pcs Dm3pcs Dm4pcs
Dim1pcs  0.063
Dim2pcs -0.678 -0.162
Dim3pcs -0.194 -0.170  0.251
Dim4pcs -0.041  0.147  0.087  0.044
Dim5pcs  0.179 -0.100  0.038  0.041  0.231
```

c) Modèle avec CSP dynamique

Linear mixed model fit by REML

Formula: dataset\$"LePen%insP95" ~ 1 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 + (1 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
198381	198509	-99176	198335	198351

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	13.803380	3.71529	
	Dim1evol6899	0.054566	0.23359	-0.116
	Dim2evol6899	0.091268	0.30211	-0.603 -0.071
	Dim3evol6899	0.026432	0.16258	-0.521 0.565 0.573
Residual		15.008679	3.87410	

Number of obs: 35611, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	11.29443	0.38715	29.173
Dim1evol6899	0.19989	0.02761	7.239
Dim2evol6899	-0.43012	0.03647	-11.795
Dim3evol6899	-0.12799	0.02356	-5.433

Correlation of Fixed Effects:

	(Intr)	D16899	D26899
Dim1evl6899	-0.112		
Dim2evl6899	-0.519	-0.070	
Dim3evl6899	-0.372	0.409	0.369

d) Modèle avec logements HLM

Linear mixed model fit by REML

Formula: dataset\$"LePen%insP95" ~ 1 + HLM + (1 + HLM | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
200195	200246	-100092	200175	200183

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	14.4903577	3.806620	
	HLM	0.0029425	0.054245	-0.231
Residual		15.8865737	3.985797	

Number of obs: 35611, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.990480	0.396143	27.744
HLM	0.026422	0.007056	3.745

Correlation of Fixed Effects:

	(Intr)
HLM	-0.208

e) Modèle avec le revenu imposable moyen

Linear mixed model fit by REML

Formula: dataset\$"LePen%insP95" ~ log.revenu + (1 + log.revenu | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
193058	193109	-96523	193046	193046

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	1096.306	33.1105	

```

log.revenu    10.580  3.2526  -0.994
Residual      14.746  3.8401
Number of obs: 34782, groups: codedpt, 93

```

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  1.2577    3.6217  0.3473
log.revenu   1.0594    0.3579  2.9604

```

Correlation of Fixed Effects:

```

      (Intr)
log.revenu -0.995

```

f) Modèle avec interaction clivage social-revenu moyen

Linear mixed model fit by REML

Formula: dataset\$"LePen%insP95" ~ Dim2pcs + log.revenu + Dim2pcs * log.revenu + (1 + Dim2pcs + log.revenu + Dim2pcs * log.revenu | codedpt)

```

Data: dataset
      AIC      BIC logLik deviance REMLdev
191311 191438 -95641  191273  191281

```

Random effects:

```

Groups   Name                Variance Std.Dev. Corr
codedpt  (Intercept)              725.6661 26.9382
          Dim2pcs              44.0851  6.6397 -0.498
          log.revenu           7.9977  2.8280 -0.991  0.548
          Dim2pcs:log.revenu    0.5161  0.7184  0.512 -0.999 -0.565
Residual                    13.9638  3.7368

```

Number of obs: 34782, groups: codedpt, 93

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  -14.1509    3.1039 -4.559
Dim2pcs       3.5545    1.0082  3.526
log.revenu    2.6539    0.3252  8.160
Dim2pcs:log.revenu -0.4392    0.1069 -4.107

```

Correlation of Fixed Effects:

```

      (Intr) Dm2pcs lg.rvn
Dim2pcs    -0.423
log.revenu -0.992  0.454
Dm2pcs:lg.r 0.438 -0.999 -0.473

```

g) Modèle avec type de communes (ZAUER)

Linear mixed model fit by REML

Formula: dataset\$"LePen%insP95" ~ type_urbain + (1 + type_urbain | codedpt)

```

Data: dataset
      AIC      BIC logLik deviance REMLdev
198937 199064 -99454  198901  198907

```

Random effects:

```

Groups   Name                Variance Std.Dev. Corr
codedpt  (Intercept)              14.36408 3.79000
          type_urbainCommune multipolarisée  0.59084 0.76866  0.094
          type_urbainEspace à dominante rurale 1.46557 1.21061 -0.244  0.191
          type_urbainPôle urbain              1.04065 1.02012 -0.091  0.065 -0.350
Residual                    15.24341 3.90428

```

Number of obs: 35611, groups: codedpt, 93

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  11.79781    0.39626 29.773
type_urbainCommune multipolarisée -0.09183    0.12820 -0.716
type_urbainEspace à dominante rurale -1.27667    0.14275 -8.943
type_urbainPôle urbain -0.19027    0.14621 -1.301

```

Correlation of Fixed Effects:
 (Intr) typ_Cm ty_Edr
 typ_rbnCmmm 0.024
 typ_rbnEsdr -0.248 0.207
 typ_rbnPôlu -0.106 0.129 -0.141

2°) Premier tour de l'élection présidentielle de 2002

a) Modèle vide

Linear mixed model fit by REML
 Formula: Lepen1pciP2002T1 ~ 1 + (1 | codedpt)
 Data: dataset
 AIC BIC logLik deviance REMLdev
 201442 201468 -100718 201436 201436
 Random effects:
 Groups Name Variance Std.Dev.
 codedpt (Intercept) 9.8281 3.1350
 Residual 16.5173 4.0642
 Number of obs: 35614, groups: codedpt, 93
 Fixed effects:
 Estimate Std. Error t value
 (Intercept) 12.492 0.326 38.32

b) Modèle avec CSP statique

Linear mixed model fit by REML
 Formula: Lepen1pciP2002T1 ~ Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs +
 + Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs | codedpt) (1
 Data: dataset
 AIC BIC logLik deviance REMLdev
 199420 199658 -99682 199338 199364
 Random effects:
 Groups Name Variance Std.Dev. Corr
 codedpt (Intercept) 9.654868 3.10723
 Dim1pcs 0.167532 0.40931 0.072
 Dim2pcs 0.160194 0.40024 -0.609 -0.360
 Dim3pcs 0.060124 0.24520 -0.352 -0.175 0.166
 Dim4pcs 0.012024 0.10965 -0.050 0.203 0.163 -0.515
 Dim5pcs 0.028888 0.16997 0.317 -0.378 0.198 -0.166 0.327
 Residual 15.376512 3.92129
 Number of obs: 35614, groups: codedpt, 93
 Fixed effects:
 Estimate Std. Error t value
 (Intercept) 12.7537439 0.3247253 39.28
 Dim1pcs 0.1081039 0.0465571 2.32
 Dim2pcs -0.5556703 0.0466019 -11.92
 Dim3pcs -0.0498454 0.0325328 -1.53
 Dim4pcs -0.0527062 0.0218611 -2.41
 Dim5pcs -0.0007485 0.0283307 -0.03

Correlation of Fixed Effects:
 (Intr) Dm1pcs Dm2pcs Dm3pcs Dm4pcs
 Dim1pcs 0.052
 Dim2pcs -0.539 -0.318
 Dim3pcs -0.281 -0.126 0.115
 Dim4pcs -0.030 0.089 0.101 -0.250
 Dim5pcs 0.189 -0.253 0.144 -0.076 0.158

c) Modèle avec CSP dynamique

Linear mixed model fit by REML
 Formula: Lepen1pciP2002T1 ~ 1 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 +
 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 | codedpt) (1

Data: dataset
 AIC BIC logLik deviance REMLdev
 199814 199941 -99892 199768 199784
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	10.349501	3.21706	
	Dim1evol6899	0.086108	0.29344	-0.110
	Dim2evol6899	0.093501	0.30578	-0.474 -0.336
	Dim3evol6899	0.029866	0.17282	-0.201 0.370 0.381
Residual		15.607855	3.95068	

 Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	12.714990	0.336045	37.84
Dim1evol6899	-0.005252	0.033584	-0.16
Dim2evol6899	-0.389499	0.037047	-10.51
Dim3evol6899	-0.186721	0.024792	-7.53

Correlation of Fixed Effects:
 (Intr) D16899 D26899
 Dim1evl6899 -0.114
 Dim2evl6899 -0.405 -0.284
 Dim3evl6899 -0.144 0.279 0.252

d) Modèle avec HLM

Linear mixed model fit by REML
 Formula: Lepen1pciP2002T1 ~ 1 + HLM + (1 + HLM | codedpt)

Data: dataset
 AIC BIC logLik deviance REMLdev
 201302 201353 -100645 201281 201290
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	10.0205608	3.165527	
	HLM	0.0029172	0.054011	-0.214
Residual		16.3978264	4.049423	

 Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	12.58502	0.32998	38.14
HLM	-0.01587	0.00707	-2.24

Correlation of Fixed Effects:
 (Intr)
 HLM -0.198

e) Modèle avec revenu imposable moyen

Linear mixed model fit by REML
 Formula: Lepen1pciP2002T1 ~ log.revenu + (1 + log.revenu | codedpt)

Data: dataset
 AIC BIC logLik deviance REMLdev
 193797 193847 -96892 193784 193785
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	926.5905	30.4399	

```

log.revenu      9.4071  3.0671  -0.995
Residual        15.0712  3.8822
Number of obs: 34785, groups: codedpt, 93

```

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  10.1342    3.3640  3.0125
log.revenu    0.2782    0.3402  0.8177

```

Correlation of Fixed Effects:

```

      (Intr)
log.revenu -0.995

```

f) Modèle avec interaction clivage social-revenu moyen

Linear mixed model fit by REML

Formula: Lepen1pciP2002T1 ~ Dim2pcs + log.revenu + Dim2pcs * log.revenu + (1 + Dim2pcs + log.revenu + Dim2pcs * log.revenu | codedpt)

```

Data: dataset
      AIC      BIC logLik deviance REMLdev
192337 192464 -96154  192298  192307

```

Random effects:

```

Groups   Name                Variance Std.Dev. Corr
codedpt  (Intercept)              745.296591 27.30012
          Dim2pcs                2.422585  1.55647 -1.000
          log.revenu              8.319755  2.88440 -0.994  0.994
          Dim2pcs:log.revenu      0.037272  0.19306  0.980 -0.980 -0.989
Residual                               14.403612  3.79521

```

Number of obs: 34785, groups: codedpt, 93

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  -3.33426    3.12121  -1.068
Dim2pcs       5.25414    0.66968   7.846
log.revenu    1.67628    0.32880   5.098
Dim2pcs:log.revenu -0.61018    0.07037  -8.671

```

Correlation of Fixed Effects:

```

      (Intr) Dm2pcs lg.rvn
Dim2pcs    -0.362
log.revenu -0.995  0.361
Dm2pcs:lg.r  0.398 -0.997 -0.401

```

g) Modèle avec type de communes (ZAUER)

Linear mixed model fit by REML

Formula: Lepen1pciP2002T1 ~ type_urbain + (1 + type_urbain | codedpt)

```

Data: dataset
      AIC      BIC logLik deviance REMLdev
200576 200703 -100273  200539  200546

```

Random effects:

```

Groups   Name                Variance Std.Dev. Corr
codedpt  (Intercept)              10.15561  3.18679
          type_urbainCommune multipolarisée  0.72577  0.85192  0.128
          type_urbainEspace à dominante rurale 1.81525  1.34731 -0.259  0.382
          type_urbainPôle urbain              0.85359  0.92390  0.058  0.071 -0.628
Residual                               15.97512  3.99689

```

Number of obs: 35614, groups: codedpt, 93

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  13.01078    0.33439  38.91
type_urbainCommune multipolarisée -0.01749    0.13512  -0.13
type_urbainEspace à dominante rurale -0.68612    0.15600  -4.40
type_urbainPôle urbain -0.65880    0.13785  -4.78

```

Correlation of Fixed Effects:
 (Intr) typ_Cm ty_Edr
 typ_rbnCmmm 0.040
 typ_rbnEsdr -0.268 0.332
 typ_rbnPôlu -0.008 0.152 -0.312

3°) Deuxième tour de l'élection présidentielle de 2002

a) Modèle vide

Linear mixed model fit by REML
 Formula: LePenP2002T2pc ~ 1 + (1 | codedpt)
 Data: dataset
 AIC BIC logLik deviance REMLdev
 210564 210590 -105279 210558 210558
 Random effects:
 Groups Name Variance Std.Dev.
 codedpt (Intercept) 12.662 3.5584
 Residual 21.339 4.6195
 Number of obs: 35614, groups: codedpt, 93
 Fixed effects:
 Estimate Std. Error t value
 (Intercept) 14.94 0.37 40.39

b) Modèle avec CSP statique

Linear mixed model fit by REML
 Formula: LePenP2002T2pc ~ 1 + Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs +
 (1 + Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs | dpt)
 Data: dataset
 AIC BIC logLik deviance REMLdev
 207988 208225 -103966 207908 207932
 Random effects:
 Groups Name Variance Std.Dev. Corr
 dpt (Intercept) 15.060989 3.88085
 Dim1pcs 0.290425 0.53891 -0.180
 Dim2pcs 0.149698 0.38691 -0.569 -0.378
 Dim3pcs 0.124178 0.35239 -0.369 0.104 0.248
 Dim4pcs 0.029720 0.17240 -0.072 0.110 0.214 -0.317
 Dim5pcs 0.051681 0.22733 0.083 -0.039 0.165 0.064 0.351
 Residual 19.513843 4.41745
 Number of obs: 35614, groups: dpt, 93
 Fixed effects:
 Estimate Std. Error t value
 (Intercept) 15.37184 0.40543 37.91
 Dim1pcs -0.40925 0.06039 -6.78
 Dim2pcs -0.54560 0.04644 -11.75
 Dim3pcs -0.02241 0.04363 -0.51
 Dim4pcs -0.07565 0.02833 -2.67
 Dim5pcs -0.11572 0.03506 -3.30

Correlation of Fixed Effects:
 (Intr) Dm1pcs Dm2pcs Dm3pcs Dm4pcs
 Dim1pcs -0.181
 Dim2pcs -0.486 -0.333
 Dim3pcs -0.317 0.092 0.182
 Dim4pcs -0.048 0.055 0.150 -0.197
 Dim5pcs 0.045 -0.039 0.124 0.053 0.194

c) Modèle avec CSP dynamique

Linear mixed model fit by REML
 Formula: LePenP2002T2pc ~ Dim1evol6899 + Dim2evol6899 + Dim3evol6899 + (1 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 | codedpt)

Data: dataset
 AIC BIC logLik deviance REMLdev
 207845 207973 -103908 207800 207815
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	15.217798	3.90100	
	Dim1evol6899	0.137019	0.37016	-0.236
	Dim2evol6899	0.083971	0.28978	-0.481 -0.328
	Dim3evol6899	0.035323	0.18794	-0.178 0.431 0.264
Residual		19.553641	4.42195	

 Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	15.22951	0.40729	37.39
Dim1evol6899	-0.36320	0.04170	-8.71
Dim2evol6899	-0.45055	0.03667	-12.29
Dim3evol6899	-0.29083	0.02732	-10.64

Correlation of Fixed Effects:
 (Intr) D16899 D26899
 Dim1evl6899 -0.231
 Dim2evl6899 -0.392 -0.274
 Dim3evl6899 -0.125 0.317 0.171

d) Modèle avec HLM

Linear mixed model fit by REML
 Formula: LePenP2002T2pc ~ 1 + HLM + (1 + HLM | codedpt)

Data: dataset
 AIC BIC logLik deviance REMLdev
 210088 210138 -105038 210067 210076
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	12.5407911	3.541298	
	HLM	0.0023022	0.047981	-0.045
Residual		21.0005469	4.582635	

 Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	15.241586	0.369107	41.29
HLM	-0.074735	0.006854	-10.90

Correlation of Fixed Effects:
 (Intr)
 HLM -0.065

e) Modèle avec revenu imposable moyen

Linear mixed model fit by REML
 Formula: LePenP2002T2pc ~ log.revenu + (1 + log.revenu | codedpt)

Data: dataset
 AIC BIC logLik deviance REMLdev
 202076 202127 -101032 202064 202064
 Random effects:

Groups	Name	Variance	Std.Dev.	Corr
--------	------	----------	----------	------

```

codedpt (Intercept) 1232.702 35.1098
          log.revenu 12.519 3.5382 -0.994
Residual          19.115 4.3720
Number of obs: 34785, groups: codedpt, 93

```

Fixed effects:

```

          Estimate Std. Error t value
(Intercept) 40.8422    3.8700 10.554
log.revenu  -2.6606    0.3914 -6.799

```

Correlation of Fixed Effects:

```

          (Intr)
log.revenu -0.995

```

f) Modèle avec interaction clivage social-revenu moyen

Linear mixed model fit by REML

Formula: LePenP2002T2pc ~ Dim2pcs + log.revenu + Dim2pcs * log.revenu + (1 + Dim2pcs + log.revenu + Dim2pcs * log.revenu | codedpt)

Data: dataset

```

          AIC      BIC logLik deviance REMLdev
201197 201324 -100584 201159 201167

```

Random effects:

```

Groups   Name                Variance Std.Dev.  Corr
codedpt (Intercept)          1.2325e+03 35.106299
          Dim2pcs             5.3274e-01 0.729891 1.000
          log.revenu          1.3347e+01 3.653348 -0.994 -0.994
          Dim2pcs:log.revenu  3.8640e-03 0.062161 -0.753 -0.753 0.704
Residual 1.8561e+01 4.308301

```

Number of obs: 34785, groups: codedpt, 93

Fixed effects:

```

          Estimate Std. Error t value
(Intercept) 30.03835    3.92477 7.654
Dim2pcs      4.58115    0.74674 6.135
log.revenu  -1.53346    0.40829 -3.756
Dim2pcs:log.revenu -0.52968    0.07739 -6.844

```

Correlation of Fixed Effects:

```

          (Intr) Dm2pcs lg.rvn
Dim2pcs    -0.021
log.revenu -0.995 0.022
Dm2pcs:lg.r 0.061 -0.997 -0.065

```

g) Modèle avec type de communes (ZAUER)

Linear mixed model fit by REML

Formula: LePenP2002T2pc ~ type_urbain + (1 + type_urbain | codedpt)

Data: dataset

```

          AIC      BIC logLik deviance REMLdev
209423 209550 -104696 209387 209393

```

Random effects:

```

Groups   Name                Variance Std.Dev.  Corr
codedpt (Intercept)          13.4557 3.6682
          Commune multipolarisée 1.0018 1.0009 0.042
          Espace à dominante rurale 2.2636 1.5045 -0.228 0.447
          Pôle urbain          1.3492 1.1615 0.026 0.031 -0.580
Residual 20.4720 4.5246

```

Number of obs: 35614, groups: codedpt, 93

Fixed effects:

```

          Estimate Std. Error t value
(Intercept) 15.0644    0.3848 39.15

```

Commune multipolarisée	0.5948	0.1564	3.80
Espace à dominante rurale	0.2258	0.1748	1.29
Pôle urbain	-1.7151	0.1663	-10.31

4°) Premier tour de l'élection présidentielle de 2007

a) Modèle vide

Linear mixed model fit by REML
Formula: LePeninsP07 ~ 1 + (1 | codedpt)
Data: dataset
AIC BIC logLik deviance REMLdev
195378 195403 -97686 195371 195372
Random effects:
Groups Name Variance Std.Dev.
codedpt (Intercept) 6.7594 2.5999
Residual 13.9384 3.7334
Number of obs: 35614, groups: codedpt, 93

Fixed effects:
Estimate Std. Error t value
(Intercept) 10.2365 0.2705 37.84

b) Modèle avec CSP statique

Linear mixed model fit by REML
Formula: LePeninsP07 ~ Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs + (1 + Dim1pcs + Dim2pcs + Dim3pcs + Dim4pcs + Dim5pcs | codedpt)
Data: dataset
AIC BIC logLik deviance REMLdev
191671 191908 -95807 191587 191615
Random effects:
Groups Name Variance Std.Dev. Corr
codedpt (Intercept) 6.8949804 2.62583
Dim1pcs 0.1484584 0.38530 -0.330
Dim2pcs 0.0749480 0.27377 -0.580 -0.065
Dim3pcs 0.0480779 0.21927 -0.187 0.279 0.189
Dim4pcs 0.0087031 0.09329 -0.250 0.003 0.458 -0.272
Dim5pcs 0.0316065 0.17778 0.081 -0.219 0.308 0.195 0.371
Residual 12.3762136 3.51798
Number of obs: 35614, groups: codedpt, 93

Fixed effects:
Estimate Std. Error t value
(Intercept) 10.560280 0.274936 38.41
Dim1pcs -0.354682 0.043602 -8.13
Dim2pcs -0.628322 0.033866 -18.55
Dim3pcs -0.009885 0.029255 -0.34
Dim4pcs -0.102586 0.019303 -5.31
Dim5pcs -0.029528 0.027575 -1.07

Correlation of Fixed Effects:
(Intr) Dm1pcs Dm2pcs Dm3pcs Dm4pcs
Dim1pcs -0.316
Dim2pcs -0.483 -0.076
Dim3pcs -0.158 0.210 0.132
Dim4pcs -0.127 -0.012 0.234 -0.141
Dim5pcs 0.043 -0.157 0.209 0.128 0.173

c) Modèle avec CSP dynamique

Linear mixed model fit by REML

Formula: LePeninsP07 ~ Dim1evol6899 + Dim2evol6899 + Dim3evol6899 + (1 + Dim1evol6899 + Dim2evol6899 + Dim3evol6899 | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
191501	191629	-95736	191454	191471

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	7.144881	2.67299	
	Dim1evol6899	0.080216	0.28322	-0.343
	Dim2evol6899	0.051347	0.22660	-0.497 -0.025
	Dim3evol6899	0.030601	0.17493	-0.442 0.420 0.375
Residual		12.361852	3.51594	

Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.51845	0.27966	37.61
Dim1evol6899	-0.28461	0.03198	-8.90
Dim2evol6899	-0.55172	0.02890	-19.09
Dim3evol6899	-0.24713	0.02387	-10.35

Correlation of Fixed Effects:

	(Intr)	D16899	D26899
Dim1evl6899	-0.328		
Dim2evl6899	-0.403	-0.038	
Dim3evl6899	-0.333	0.324	0.243

d) Modèle avec HLM

Linear mixed model fit by REML

Formula: LePeninsP07 ~ HLM + (1 + HLM | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
195127	195178	-97558	195106	195115

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	6.6556955	2.579863	
	HLM	0.0013783	0.037126	-0.089
Residual		13.8065810	3.715721	

Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.406337	0.269207	38.66
HLM	-0.041930	0.005403	-7.76

Correlation of Fixed Effects:

	(Intr)
HLM	-0.101

e) Modèle avec revenu fiscal moyen

Linear mixed model fit by REML

Formula: LePeninsP07 ~ log.revenu + (1 + log.revenu | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
186618	186669	-93303	186605	186606

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	903.7348	30.0622	
	log.revenu	8.7202	2.9530	-0.997
Residual		12.2686	3.5027	

Number of obs: 34785, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	39.7988	3.2879	12.105
log.revenu	-3.0396	0.3248	-9.357

Correlation of Fixed Effects:

	(Intr)
log.revenu	-0.997

f) Modèle avec interaction clivage social-revenu moyen

Linear mixed model fit by REML

Formula: LePeninsP07 ~ Dim2pcs + log.revenu + Dim2pcs * log.revenu + (1 + Dim2pcs + log.revenu + Dim2pcs * log.revenu | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
185022	185149	-92496	184982	184992

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	6.3142e+02	25.128045	
	Dim2pcs	3.8287e-01	0.618765	1.000
	log.revenu	6.2952e+00	2.509020	-0.995 -0.996
	Dim2pcs:log.revenu	3.7195e-03	0.060987	-0.871 -0.869 0.836
Residual		1.1681e+01	3.417809	

Number of obs: 34785, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	29.57320	2.84487	10.395
Dim2pcs	2.83452	0.58793	4.821
log.revenu	-1.98119	0.28589	-6.930
Dim2pcs:log.revenu	-0.35718	0.06099	-5.857

Correlation of Fixed Effects:

	(Intr)	Dm2pcs	lg.rvn
Dim2pcs	-0.025		
log.revenu	-0.996	0.032	
Dm2pcs:lg.r	0.046	-0.998	-0.056

g) Modèle avec type de commune (ZAUER)

Linear mixed model fit by REML

Formula: LePeninsP07 ~ type_urbain + (1 + type_urbain | codedpt)

Data: dataset

AIC	BIC	logLik	deviance	REMLdev
194050	194178	-97010	194013	194020

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
codedpt	(Intercept)	6.37723	2.52532	
	Commune multipolarisée	0.57735	0.75984	-0.097
	Espace à dominante rurale	1.67303	1.29346	-0.028 0.645
	Pôle urbain	0.69694	0.83483	-0.069 -0.104 -0.477
Residual		13.30680	3.64785	

Number of obs: 35614, groups: codedpt, 93

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	10.3044	0.2659	38.75
Commune multipolarisée	0.7047	0.1197	5.89
Espace à dominante rurale	0.3019	0.1495	2.02
Pôle urbain	-1.3770	0.1261	-10.92

