



MÉTHODES DE SIMULATION ET MODÉLISATION ÉCONOMIQUE

**Projet Économétrie Spatiale:
Étude économétrique et spatiale des dynamiques
régissant la criminalité du Grand Paris**

Auteurs:

Axel GARDAHAUT
Robin LOUÉRAT

À l'attention de :
Véronique THELEN

June 28, 2020

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 1.1 | Problématique et Contextualisation | 2 |
| 1.2 | Données utilisées | 3 |
| 2 | Démarche | 4 |
| 3 | Modélisation | 4 |
| 3.1 | Analyse descriptive | 4 |
| 3.2 | Typologie de communes | 7 |
| 3.3 | Création des matrices de voisinage | 14 |
| 3.4 | Cartographie du taux de vol automobiles en 2015 | 15 |
| 3.5 | Graphique de Moran | 16 |
| 3.6 | Modèles spatiaux avec une matrice de distance | 17 |
| 3.7 | Comparaison des modèles | 18 |
| 3.8 | Analyse des effets directs et indirects | 20 |
| 4 | Conclusion | 21 |

1 Introduction

1.1 Problématique et Contextualisation

Si Paris est la ville préférée mondiale des touristes (Travelers's Choice 2018 de TripAdvisor), la région parisienne reste une région associée à la criminalité et la délinquance dans l'imaginaire collectif (français).

D'après Durkheim, une société sans crimes ne doit être considérée comme normale (1966).

Il est donc plutôt normal d'observer des crimes et d'avoir des criminels dans une société.

L'important est de déterminer les variables qui causent la criminalité dans une société, les éléments qui provoquent une baisse ou une hausse du taux de crime et ce qui pousse un individu au crime.

En effet la criminalité a un effet négatif sur l'activité économique et sur la qualité de vie des citoyens qui se sentent moins en sécurité chez eux ou dans l'espace public.

De nombreuses études ont été réalisées sur le lien entre la criminalité et des variables économiques en particulier le taux de chômage, cependant les études faites dans le cadre de différentes sociétés semblent ne pas parvenir à un consensus.

Dans le cadre de ce projet nous souhaitons étudier les dynamiques qui régissent la criminalité du Grand Paris.

Pour ce faire, nous nous sommes appuyés sur le travail de Alison Oliver : "The Economics of Crime: An Analysis of Crime Rates in America".

L'étude cherche à déterminer les facteurs qui déterminent de manière significative les variations du taux de crime aux USA au cours du temps.

L'étude porte sur des données nationales sur la période 1960-1998.

Elle se base sur la théorie du "Rational Choice" de Becker qui exprime ce qui pousse l'individu au crime. En effet, la criminalité est la somme de crimes individuels.

Selon Becker, le bénéfice d'un crime est égal à :

$$b - (pc + o)$$

où b le gain lié au crime, p la probabilité de se faire prendre, c le coût de la punition et o le coût d'opportunité.

L'individu étant rationnel, un crime sera commis seulement si :

$$b - (pc + o) > 0$$

L'idée centrale est donc d'étudier les variables qui influent sur b , p , c et o .

La théorie nous pousse donc à penser que les variables favorisant b vont entraîner une hausse du taux de crime et inversement pour p , c et o .

'Crime reduction can occur through reducing the benefits of crime, raising the probability of being caught, or increasing the costs of punishment conditional upon being caught' (A. Oliver). Ehrlich (1973) énonce que l'éducation influe sur la commission des crimes dans la mesure où un individu mieux éduqué saura mieux évaluer les bénéfices à tirer des activités légales et illégales.

Un investissement de l'individu dans l'éducation peut être associé avec un plus grand accès à des revenus légaux donc le coût d'opportunité o aurait tendance à être plus grand. On s'attend donc à ce que le taux de population éduquée évolue dans le sens inverse à celui du taux de crime.

Il exprime également que le taux de chômage peut être vu comme un indicateur des opportunités de revenu disponibles dans le marché du travail légal.

Donc si le taux de chômage croît, les opportunités dans le marché du travail traditionnel diminuent, donc plus susceptible de générer du revenu via des activités illégales, donc o diminue.

Ainsi d'après la théorie et les recherches existantes, on s'attend à ce qu'une variation du taux de chômage entraîne une variation du taux de crime dans le même sens.

Ehrlich exprima également que la médiane du revenu moyen dans une communauté a un impact sur le taux de crime, en particulier sur le taux de crime à la propriété, car il mesure les bénéfices à commettre un crime : b ('opportunities provided by potential victims of crime').

Selon les criminologues, la distribution en âge de la population a une grande influence sur le taux de crime. Il semble que le taux de crime est impacté principalement par la proportion de jeunes hommes.

On s'attend donc à ce que le taux de jeunes soit en relation directe (positive) avec le taux de crime. Les travaux de Becker (1968) et Ehrlich (1973) montrent également l'effet négatif sur le taux de crime de la présence policière, du nombre d'arrestations, et de la sévérité des condamnations.

Des variables sociétales comme la structure des ménages ou bien l'appartenance à certaines communautés semblent également avoir un impact sur le taux de crime mais ce ne sera pas étudié ici.

Dans l'étude de A. Oliver, la variable dépendante est le taux de crime national et les variables explicatives sont des variables économiques, sociales et liées à la répression (Deterrence) au niveau national.

Plusieurs modèles sont construits avec plus ou moins de variables.

Peu d'entre elles semblent significatives.

Cependant cette étude montre que l'impact du taux de jeunes et de la répression semble significatif sur le taux de crime. L'auteur nous invite également à étudier les crimes par genres pour étudier plus finement les dynamiques en place.

Dans le cadre de ce projet nous souhaitons étudier les dynamiques qui régissent la criminalité du Grand Paris. Pour ce faire nous avons prolongé le projet d'économétrie réalisé par Axel Gardahaut et Jacky Mpoy par une approche spatiale.

1.2 Données utilisées

Nous avons récoltés les données sur les communes du Grand Paris pour l'année 2015.

Nous avons réalisé une jointure entre les dossiers complets l'INSEE pour les différentes communes et arrondissements de 2015, le dossier complet concernant les ménages français en 2015 (issu du recensement de la population) et les données de l'étude "Crimes et délits commis dans les communes du Grand Paris en 2015" de l'ONDRP (Observatoire National de la Délinquance et des réponses pénales).

Nous avons récoltés nos données pour 135 communes/individus de la région parisienne correspondant à la zone du Grand Paris.

Nous avons les variables suivantes :

- Nom: Nom de la commune ou arrondissement
- Pop: Population de la commune/arrondissement
- NB_VPNC: nb de faits de violences physiques non crapuleuses
- VPC: violences physiques crapuleuses
- VSP : vols simples contre les particuliers
- Log : nb de logements
- MED15: niveau de vie médian pour 2015 (en k)
- Paris: indicatrice Paris
- CR: nb Cambriolages
- CRP: cambriolages en résidences principales
- CRL: cambriolages (principale+secondaire)
- VA : vols d'automobiles
- DD: destructions et dégradations
- ILS: infractions à la législation sur les stupéfiants
- y_i: somme de tous les crimes confondus
- taux_delit: taux tous crimes confondus pour 1000 habitants
- taux_etabl_soc: nb d'établissements sociaux pour 1000 habitants
- taux_chom: taux de chômage (en %)
- taux_sco: % des actifs ayant au moins le BEPC
- taux_pauv: taux de pauvreté des habitants de la commune
- taux15_29: % des 15-29 ans dans la population de la commune

L'indicateur des violences physiques non crapuleuses (VNPC) est composé : des homicides (hors homicides pour voler et règlements de compte), des coups et blessures volontaires mortels et non mortels, des violences à dépositaire de l'autorité, des mauvais traitements à enfants et des séquestrations prises d'otages.

L'indicateur des violences physiques crapuleuses (VPC) est composé: des homicides pour vol et des règlements de compte, des vols avec violences avec arme (à feu ou blanche) et les vols avec violences commis sans arme.

L'indicateur MED15 correspond à la médiane du niveau de vie. À titre d'informations, le niveau de vie correspond au revenu disponible divisé par le nombre d'unité de consommation.

Les contours des agglomérations étudiées sont obtenues grâce à leur code Insee via l'API Carto de www.data.gouv.fr.

2 Démarche

Notre objectif dans le cadre de ce projet est d'utiliser une modélisation spatiale de la criminalité pour essayer d'améliorer le modèle retenu.

Notre but est d'étudier les variables qui déterminent la criminalité du Grand Paris, et plus précisément le taux de vol automobile par commune.

Pour ce faire, nous avons utilisé plusieurs modèles différents:

1. Le modèle OLS de la première étude
2. Modèle SEM
3. Modèle SAR
4. Modèle SLX
5. Modèle SDM
6. Modèle SDEM
7. Modèle Manski

Nous allons ensuite via une stratégie de tests, retenir les modèles qui correspondent à nos problématiques et en tirer des conclusions.

3 Modélisation

3.1 Analyse descriptive

Sans revenir dans le détail, quelques analyses que nous pouvons retenir de l'analyse descriptive de nos données.

Total tous crimes confondus (2015)

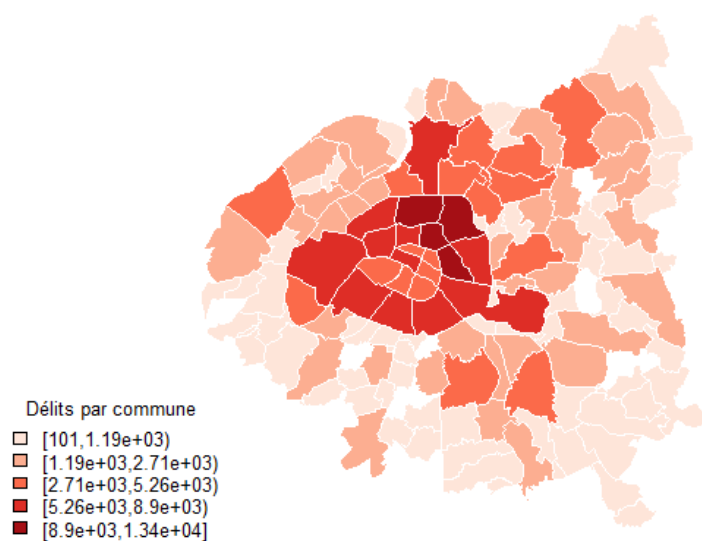


Figure 1: Cartographie de la criminalité tous types confondus

On observe de fortes variations pour le taux de délit sur l'ensemble du Grand Paris.
On observe en effet le schéma suivant de répartition pour le total de crimes commis:

Taux de delit (2015)

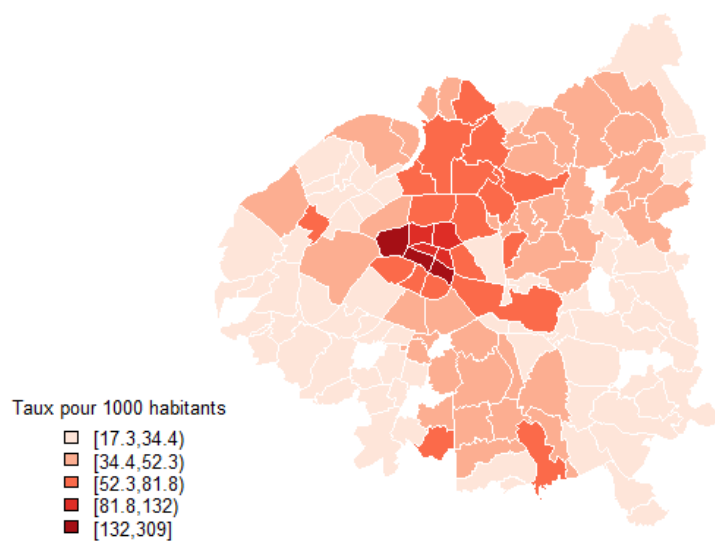


Figure 2: Statistiques descriptives

Pour filtrer l'effet de la population, on étudiera les taux de criminalité pour 1000 habitants. Tandis que 7 observations possèdent un taux de délit supérieur à 100 délits pour 1000 habitants, 6 sont des arrondissements de Paris, avec un maximum à 309 crimes pour 1000 habitants pour le 1er arrondissement. De l'autre côté du spectre, 60% des communes possèdent un taux de délit inférieur à 40 délits pour 1000 habitants. Ce constat appuie notre intuition d'une possible discrimination entre Paris et le reste du Grand Paris donc potentiellement explicable par un schéma si les variables de la théorie économique ne parviennent pas à capter l'information expliquant ce décalage.

De même on observe, de très fortes disparités au niveau de l'éducation, on observe une forte taux d'éducation (72%) sur la majeure partie du Grand Paris mais certaines communes accusent un retard certain. Ces mêmes communes accusent généralement un taux de pauvreté et de chômage élevé. Cependant, il ne semblerait pas que cela ait un impact sur le taux de jeunes entre 15 et 29 ans. Une forte densité semble généralement associée avec une population jeune. On observe des écarts qui vont du simple au double (de 15% à 30%). Les communes (hors Paris) comptabilisant les plus hauts niveaux de vie médian (MED15) sont bien moins denses que les autres communes.

Ce ne sont pas les communes ou arrondissements les plus sujets aux crimes et possèdent de faibles taux de jeunes. Sans surprise, on observe une forte corrélation entre le taux de chômage et le taux de pauvreté, il sera pris en compte dans le modèle. Le taux de délit n'est pas fortement corrélé avec les autres variables à l'exception du taux d'établissements sociaux, ceci sera discuté plus loin.

La densité est fortement corrélée avec les établissements sociaux et le taux de jeunes, on observe surtout pour ces points l'influence du statut particulier de PARIS.

On observe une forte corrélation entre le niveau de vie médian, il faudra étudier si il y a colinéarité. On observe une très faible corrélation entre le taux de jeunes et le taux de scolarisation. On observe une corrélation négative entre le taux de pauvreté et le taux d'établissements sociaux. Enn Paris est très corrélée avec le taux de scolarisation et le taux d'établissements sociaux. Il est intéressant d'observer que les différents types de crimes n'opèrent pas dans les mêmes typologies de communes. Nous allons réaliser notre étude sur le fond de carte du Grand Paris ci-dessous :



Figure 3: Fond de carte du Grand Paris

3.2 Typologie de communes

Pour illustrer la diversité des typologies de criminalité, voici pour chaque commune la typologie la plus grande dans la commune.

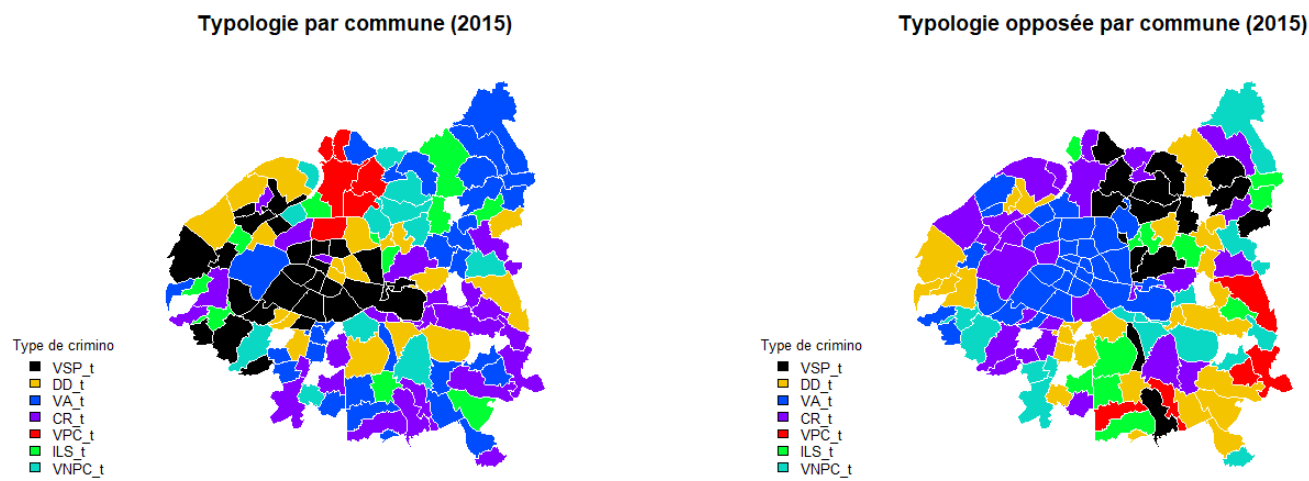


Figure 4: Carte des types de crimes les + et - présents par commune

On observe bien des clusters pour un certain nombre de typologies (VA,VSP,CR,DD) que ce soit en forte ou faible valeurs ce qui laisse supposer des externalités spatiales et/ou des effets d'entraînement.

On observe aussi une corrélation entre les différentes typologies:

| | CR_pct | VA_pct | DD_pct | ILS_pct | VSP_pct | VPC_pct | VNPC_pct |
|----------|---------|---------|---------|---------|---------|---------|----------|
| CR_pct | 1.0000 | 0.6370 | 0.0390 | -0.0772 | -0.483 | -0.4113 | 0.0398 |
| VA_pct | 0.6370 | 1.0000 | 0.1567 | 0.0761 | -0.694 | -0.1983 | 0.3079 |
| DD_pct | 0.0390 | 0.1567 | 1.0000 | -0.0273 | -0.445 | -0.0699 | 0.1748 |
| ILS_pct | -0.0772 | 0.0761 | -0.0273 | 1.0000 | -0.420 | -0.0103 | 0.2212 |
| VSP_pct | -0.4832 | -0.6936 | -0.4447 | -0.4205 | 1.000 | -0.1604 | -0.6444 |
| VPC_pct | -0.4113 | -0.1983 | -0.0699 | -0.0103 | -0.160 | 1.0000 | 0.2213 |
| VNPC_pct | 0.0398 | 0.3079 | 0.1748 | 0.2212 | -0.644 | 0.2213 | 1.0000 |

Figure 5: Matrice de corrélation des typologies

On observe une corrélation positive forte et guère suprenante entre les cambriolages et les vols d'automobiles et la corrélation négative de celles-ci avec les VSP comme en atteste les deux cartographies plus précédentes.

On observe une opposition Paris-Couronne assez nette pour ces typologies.

Ce qui est plus notable est la faible corrélation entre les dégradations et le reste des criminalités : on peut simplement observer que les dégradations sont négativement corrélée avec les VSP qui sont généralement dans les communes aisées. Cependant on peut constater que les communes avec un fort taux ILS sont proches des communes avec fort taux de dégradations.

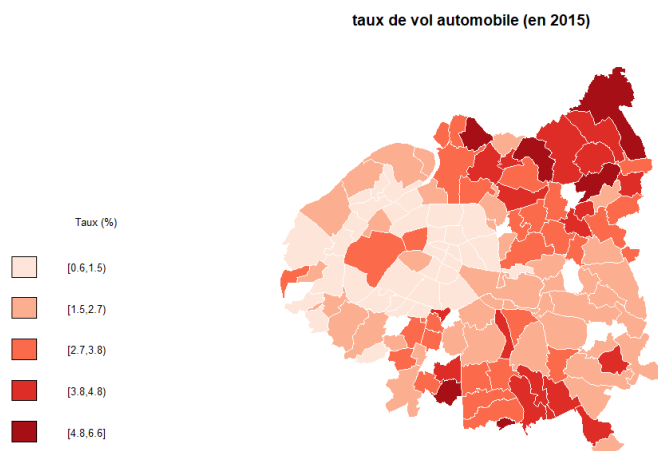


Figure 6: Taux de vols automobiles par commune

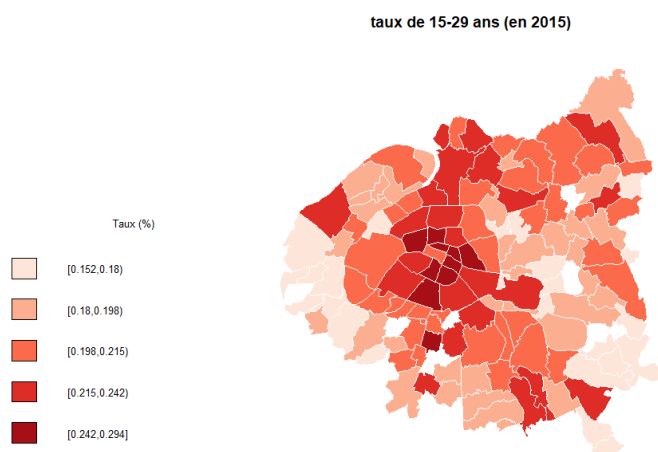


Figure 7: Taux de jeunes par communes

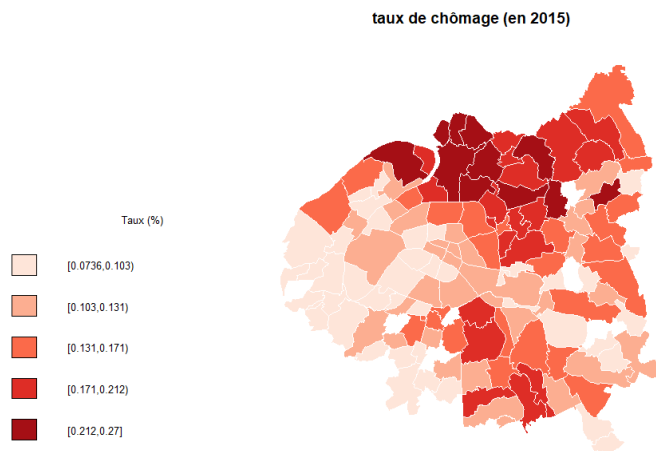


Figure 8: Taux de chômage par commune

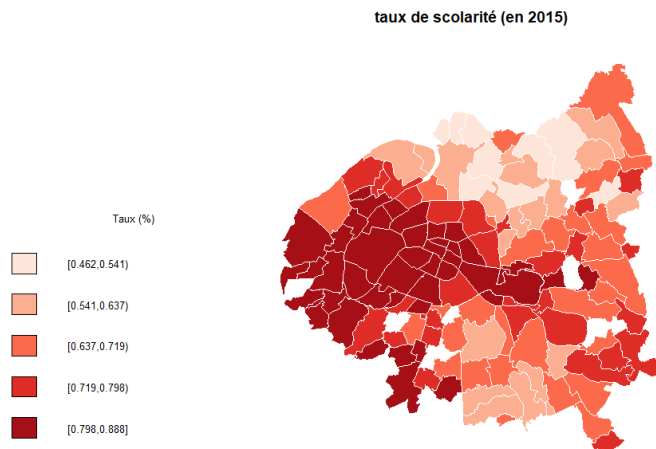


Figure 9: Taux de scolarité par commune

Pour étudier les typologies, on pourra s'intéresser aux pourcentages d'une typologie parmi les crimes totaux de la commune plutôt qu'aux taux qui ne prennent en compte la variabilité des autres car une commune peut avoir un nombre de crimes élevés dans une typologie mais ceci n'est pas forcément significatif dans le cas d'une commune avec beaucoup ou peu de crimes en totalité par rapport à la moyenne.

On observe les quantiles des pourcentages plutôt que les t-stats étant donné la non-normalité de certaines distributions.

La majorité des crimes commis à Paris, qui concentre les plus hauts taux de délits de la région étudiée sont des VSP (vols simples de particuliers), ce qui pourrait ne pas être sans lien avec la densité, le niveau de vie (social et économique) et le statut particulier de la ville de Paris.

Par ailleurs, les communes ayant un taux élevé de VSP (50-60%) sont très proches de Paris.

Les villes de densité plus faibles et moins aisées semblent beaucoup moins sujettes aux VSP (10-20%).

Typologie VSP (2015)

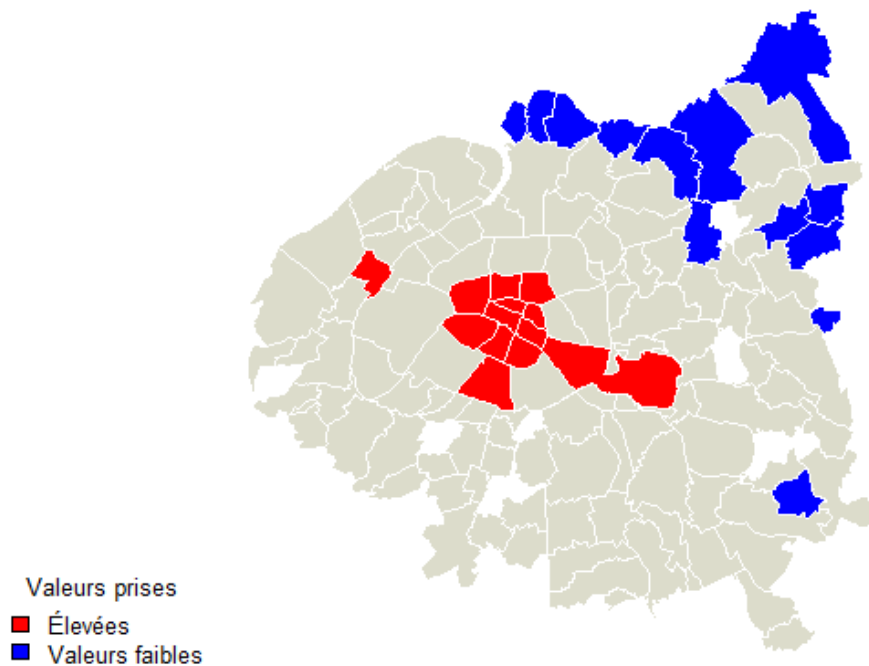


Figure 10: Carte des types des VSP

Les crimes liés aux stupéants (ILS), aux dégradations (DD) sont davantage localisés en périphérie (nord) de Paris, principalement dans les communes avec des taux de chômage élevés et une scolarisation faible, même si certaines communes et arrondissements frontaliers en concentrent également.

Typologie ILS (2015)

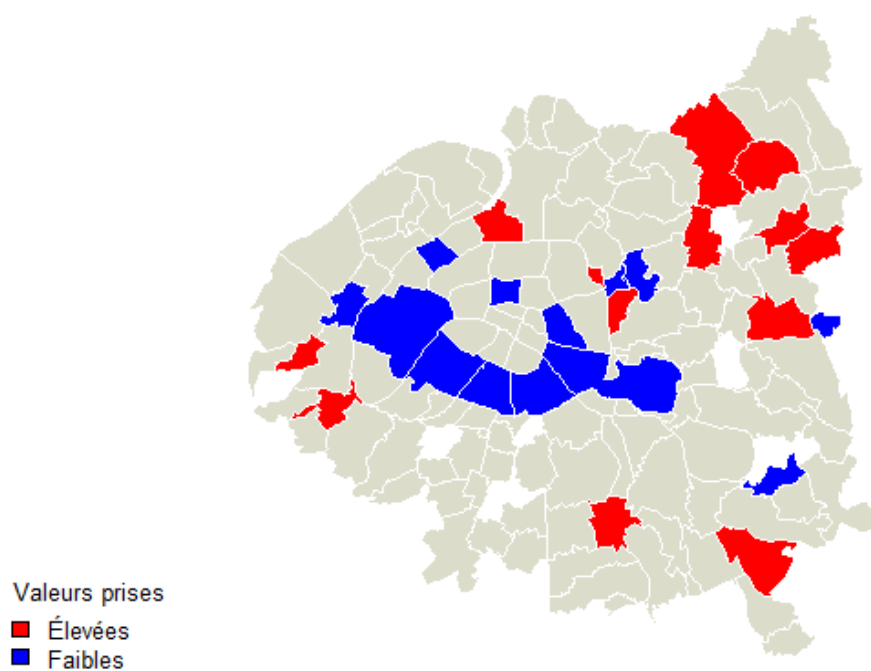


Figure 11: Carte des types des ILS

Typologie DD (2015)

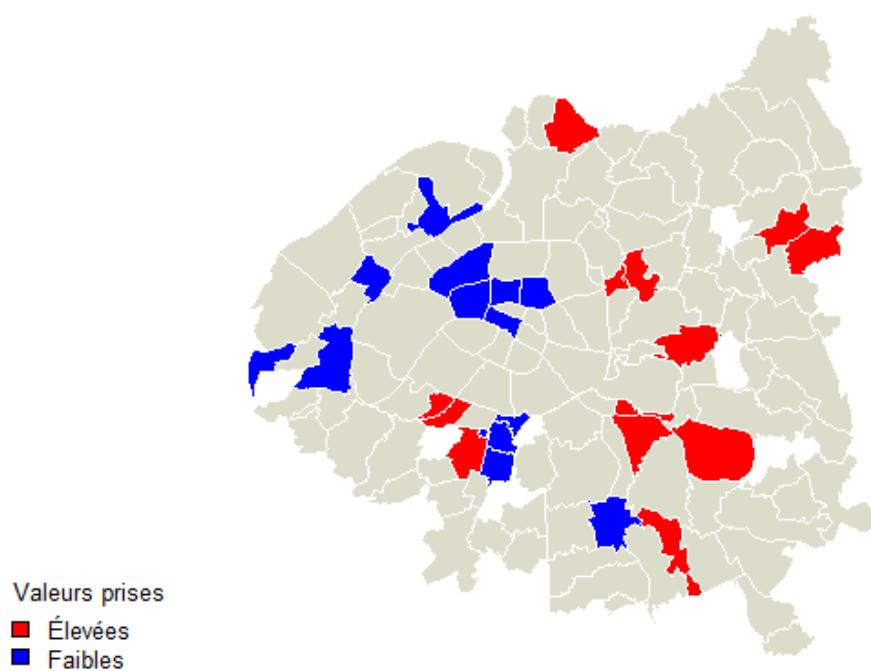


Figure 12: Carte des types des DD

Les cambriolages (CRL) sont davantage recensés dans les villes aux revenus moyens, de faibles taux de chômage et quelque peu éloignées de Paris.

Typologie CR (2015)

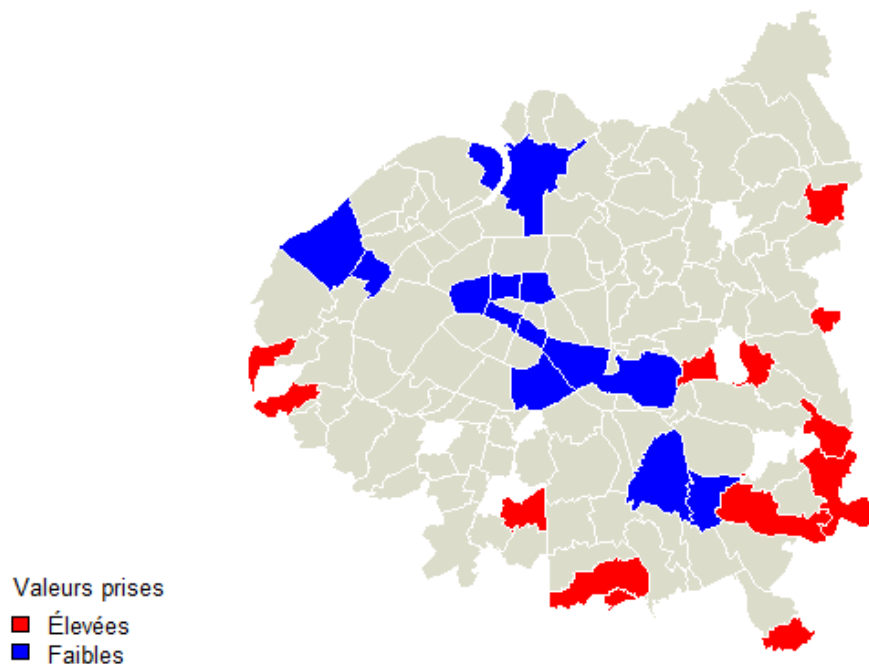


Figure 13: Carte des types des CR

De même pour les vols d'automobiles (VA), ce sont les villes situées à la périphérie du Grand Paris qui sont les plus touchées.

Typologie VA (2015)

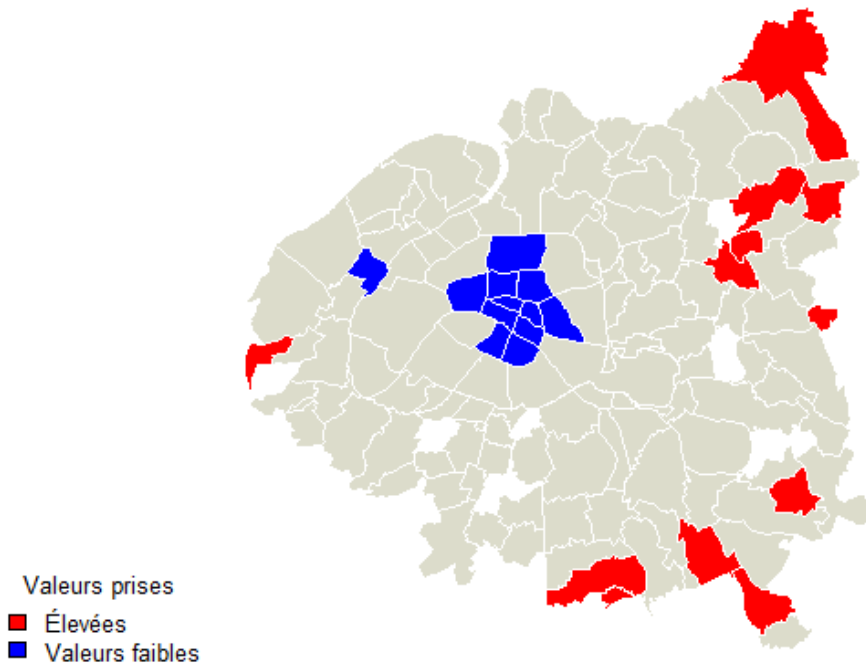


Figure 14: Carte des types des VA

La plus faible densité, donc une plus faible congestion et l'éloignement à Paris peut laisser penser que cela rend ce type de crime plus aisé à réaliser.

Ainsi si l'omission d'une variable prenant en compte le nombre de voitures par habitant ou équivalent cad "offertes" au vol, on peut s'attendre à ce que le modèle spatial incorpore ceci dans la mesure où le besoin d'un véhicule dépend à minima de l'arbitrage entre la localisation des lieux de services et de travail ainsi que des alternatives de transport disponibles.*

Nous allons donc étudier cette typologie de crime (VA) pour essayer de mieux expliquer les phénomènes observés en prenant les composantes spatiales de cette branche du crime.

3.3 Création des matrices de voisinage

Les méthodes incorporant la composante spatiale d'un problème économétrique sont basées sur la mesure de proximité entre les différentes régions observées.

Pour modéliser les proximités relatives entre les différentes régions étudiées, on dispose de différentes méthodes possibles.

Nous allons initialiser plusieurs matrices de voisinage, la matrice de contiguité, des matrices de voisinage par la distance des k plus proches voisins pour $k = 2$, $k = 5$ et $k = 10$, une matrice de voisinage basée sur la distance euclidienne (entre les centroides des régions).

Ce fond de carte illustre bien les liens entre les voisins du Grand-Paris :



Figure 15: Représentation des liens entre les voisins

3.4 Cartographie du taux de vol automobiles en 2015

On représente les zones du Grand Paris en fonction de leurs taux de vol automobile :

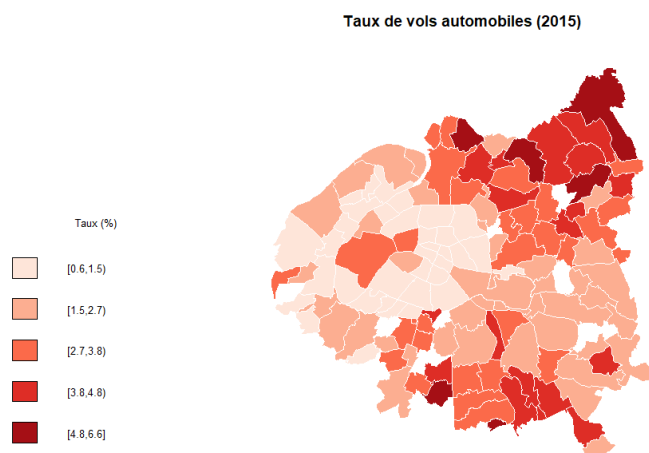


Figure 16: Cartographie du taux de vol automobile en 2015

La littérature nous donne 2 types de voleurs automobiles (étude réalisée au Canada):

1. Les voleurs professionnels organisés en réseaux qui ciblent des voitures de valeur plus élevée en général (vol de revente)
2. Les voleurs amateurs, pauvres, désorganisés, qui volent pour jouir d'un bien qu'ils ne peuvent obtenir d'une autre manière; ils ciblent des voitures de faible valeur moins sécurisées (vol d'usage)

Ce qui peut expliquer la présence de vols dans Paris 16e et les quartiers environnants, résidentiels et huppés par rapport à la localisation des autres vols plutôt réalisés dans les communes au delà de 15 km de Paris Centre.

3.5 Graphique de Moran

L'indice de Moran permet de mesurer le niveau d'autocorrélation spatiale d'une variable et de tester sa significativité. Il est égal au ratio entre la covariance des observations contigües, définies par la matrice d'interactions spatiales, et la variance totale de l'échantillon. Ses valeurs sont comprises entre -1 et 1. Une valeur nulle signifie que la répartition spatiale de la variable étudiée est aléatoire. Une valeur négative indique une autocorrélation spatiale négative, et inversement pour une valeur positive.

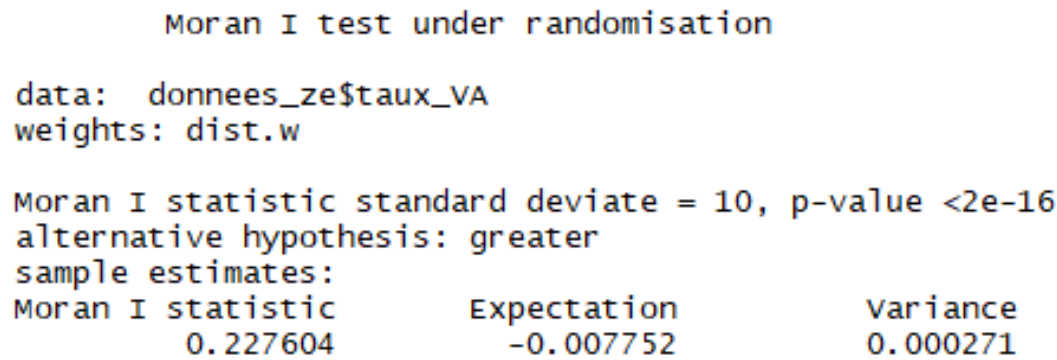


Figure 17: Statistique I de Moran

Notre I de Moran est de 0.22 avec une p-value significative au seuil 5%. On note donc une autocorrélation spatiale positive ainsi on s'attend à visualiser des clusters de hautes valeurs et basses valeurs en terme de vol d'automobiles.

Le diagramme de Moran permet une visualisation de ces résultats. Ce diagramme montre le nuage de points des couples associant la valeur (indiquée en abscisse) de la variable dans une unité spatiale et la moyenne des valeurs de la variable dans les zones en interaction avec cette unité spatiale, d'après la matrice d'interactions spatiales. Le graphique est complété par la droite de régression linéaire, dont la pente est égale à l'indice de Moran.

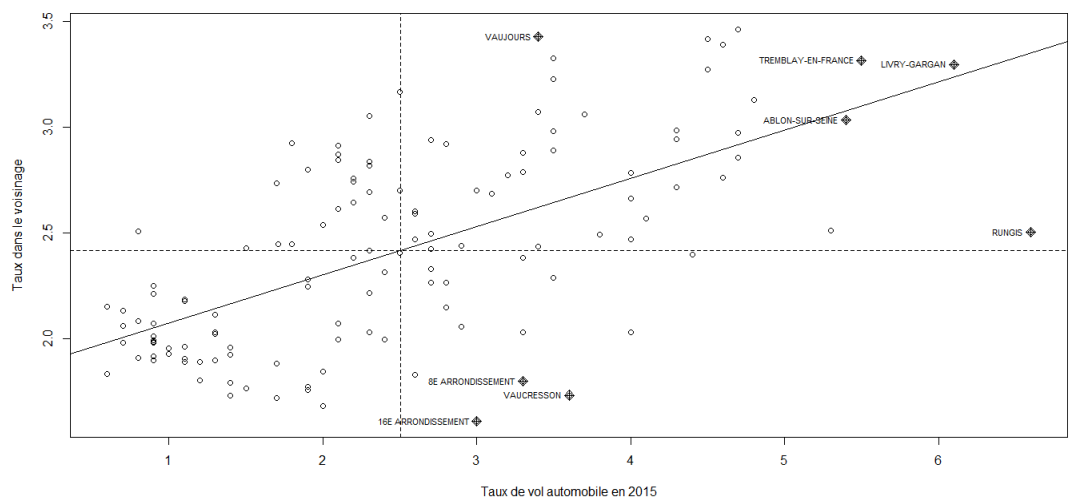


Figure 18: Graphique I de Moran

La figure 6 présente le diagramme de Moran du taux de vol automobile dans le Grand-Paris, calculés en utilisant la matrice euclidienne.

À la vue du nuage de points, on comprend pourquoi la valeur apparaît statistiquement significative.

Au-delà de la représentation de l'autocorrélation spatiale globale, ce graphique a aussi l'intérêt de signaler les unités spatiales où l'autocorrélation spatiale locale est significative.

Ces unités spatiales sont représentées par des losanges et sont libellées avec leur nom, tandis que les autres sont indiquées simplement par des cercles.

Les unités spatiales avec une autocorrélation spatiale locale significative peuvent être visualisées sous forme cartographique.

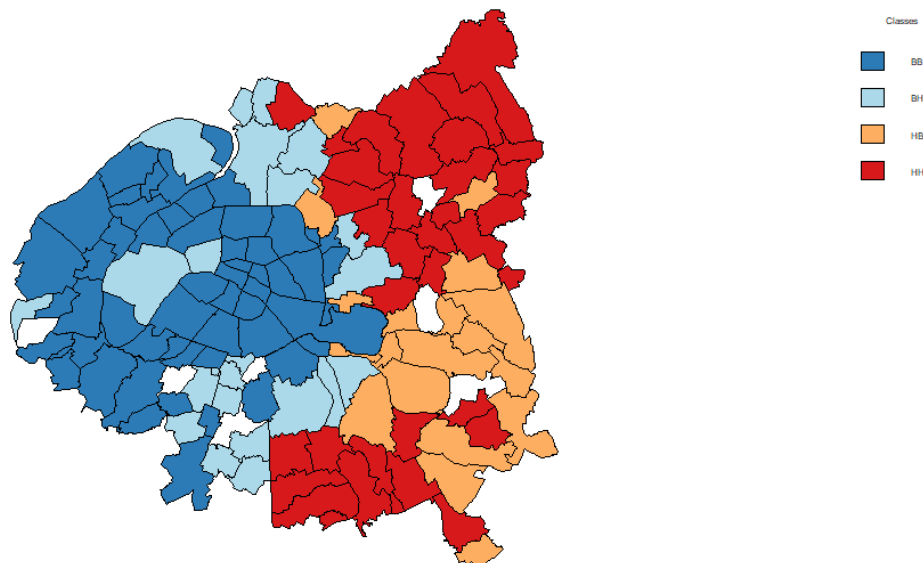


Figure 19: Cartographie de l'autocorrélation spatiale locale

3.6 Modèles spatiaux avec une matrice de distance

On utilise le modèle économétrique de base auquel on ajoute des modèles pour capturer les schémas spatiaux qui régissent le vol de voitures dans le Grand Paris.

On rappelle l'équation de base ($R^2=0.46$) :

$$\log(\text{taux_VA}) = \log(\text{taux15_29}) + \log(\text{MED15}) + \log(\text{taux_sco}) + \log(\text{taux_chom})$$

Dans un premier temps, on réalise une régression sur le modèle économétrique travaillé dans l'étude précédente.

Ensuite on teste la présence d'un schéma spatial dans les erreurs, le cas échéant, on cherchera à modéliser le schéma spatial dans des modèles spatiaux.

On fait le test de Moran sur les résidus.

Global Moran I for regression residuals

```
data:
model: lm(formula = modele, data = donnees_ze)
weights: matrice

Moran I statistic standard deviate = 6, p-value = 7e-10
alternative hypothesis: greater
sample estimates:
Observed Moran I      Expectation      Variance
      0.078439         -0.011830         0.000221
```

Figure 20: Résultats du test de Moran sur les résidus

On rejette H_0 , on en conclut qu'il y a un schéma spatial dans les résidus qu'on captera avec les modèles.

3.7 Comparaison des modèles

On applique la stratégie de test pour la sélection de modèle.

On peut s'attendre d'après l'analyse descriptive spatiale à des interactions exogènes.

On teste donc les statistiques LM_{err} et LM_{LAG} .

Ici les 2 sont significatifs donc on calcule les statistiques RLM_{err} et RLM_{LAG} .

Seul RLM_{LAG} est significatif donc on choisit le modèle LAG.

D'autre part, si le test RLM_{err} avait été significatif, nous aurions soit envisagé les deux modèles pour ensuite comparer leur pouvoir explicatif soit introduit des effets fixes pour stabiliser les résultats et reestimer des modèles sans composantes spatiales dans les erreurs.

On fait un test de significativité globale, le modèle est globalement satisfaisant au seuil 5%.

Cependant, le modèle qui capte le mieux le schéma spatial est le modèle SLX pour la distance euclidienne.

```
call:
lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1], collapse = "+"))),
    data = as.data.frame(x), weights = weights)

Residuals:
    Min       1Q   Median       3Q      Max
-0.832 -0.255  0.005  0.193  1.179

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      29.522     21.538   1.37    0.1730
log.taux15_29.    -1.185      0.365  -3.24    0.0015 **
log.MED15.         0.601      0.384   1.57    0.1198
log.taux_sco.     -2.963      0.655  -4.52 0.0000143 ***
log.taux_chom.    -0.079      0.310  -0.25    0.7994
lag.log.taux15_29.  4.916      1.854   2.65    0.0091 **
lag.log.MED15.     -4.390      2.056  -2.14    0.0347 *
lag.log.taux_sco.  -5.464      4.238  -1.29    0.1998
lag.log.taux_chom. -6.016      1.276  -4.71 0.0000065 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.376 on 121 degrees of freedom
Multiple R-squared:  0.598,    Adjusted R-squared:  0.571
F-statistic: 22.5 on 8 and 121 DF, p-value: <0.0000000000000002
```

Figure 21: Résultats SLX inverse distance

Comparaison des modèles pour les différentes matrices :

| | Distance euclidienne | | KPPV5 | |
|------------|----------------------|------------|-----------|------------|
| | AIC | p-value | AIC | p-value |
| OLS | 104.54105 | 2.177e-15 | 104.54105 | 2.177e-15 |
| SAR | 91.15634 | 8.7695e-05 | 50.61132 | 7.5162e-14 |
| SLX | 89.08526 | < 2.2e-16 | 76.23429 | < 2.2e-16 |
| SDM | 90.89465 | 0.66241 | 46.26935 | 1.5698e-08 |

Figure 22: Résultats des différentes méthodes pour la matrice euclidienne et 5 plus proches voisins

On observe que le modèle SDM est celui qui permet d'expliquer le mieux l'information sur le taux de vol automobile pour la matrice de distance des 5 plus proches voisins cependant il semble que cela ne soit pas adapté étant la non-significativité de quasi tous les coefficients.

```
Call:spatialreg::lagsarlm(formula = formula, data = data, listw = listw,
    na.action = na.action, Durbin = Durbin, type = type, method = method,
    tol.solve = tol.solve, trs = trs, control = control)

Residuals:
    Min       1Q   Median       3Q      Max
-0.9727327 -0.1941848 -0.0078713  0.2335471  1.1202412

Type: mixed
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.12284    6.35820   0.6484 0.5167076
log(taux15_29) -1.21911    0.41718 -2.9223 0.0034748
log(MED15)      1.02482    0.39080  2.6223 0.0087328
log(taux_sco)   -2.07981    0.61398 -3.3874 0.0007055
log(taux_chom)  0.22856    0.29982  0.7623 0.4458822
lag.log(taux15_29) 1.44486    0.58132  2.4855 0.0129379
lag.log(MED15)   -1.73020    0.74400 -2.3255 0.0200432
lag.log(taux_sco) -1.13152    1.18253 -0.9569 0.3386355
lag.log(taux_chom) -1.57425    0.51219 -3.0736 0.0021152

Rho: 0.354, LR test value: 9.39, p-value: 0.0021817
Asymptotic standard error: 0.112
z-value: 3.15, p-value: 0.001628
wald statistic: 9.93, p-value: 0.001628

Log likelihood: -44.9 for mixed model
ML residual variance (sigma squared): 0.114, (sigma: 0.338)
Number of observations: 130
Number of parameters estimated: 11
AIC: 112, (AIC for lm: 119)
LM test for residual autocorrelation
test value: 5.02, p-value: 0.025091
```

Figure 23: Résultats SDM 5-PPV

Il n'est pas étonnant intuitivement que le meilleur modèle soit le SLX qui modélise la présence d'externalités spatiales donc des caractéristiques des communes environnantes en non pas un modèle LAG qui suppose des effets d'entraînements. Par ailleurs, il semble difficile d'imaginer que les vols soient connus des criminels sauf s'ils opèrent en bande organisée.

D'autre part, il semble que les communes du Grand-Paris ont en moyenne 5-6 voisins directs donc les 5 ppv sont des voisins directs mais que les communes les plus touchées par les vols auto ne possèdent pas 5 voisins directs du fait de leur localisation en périphérie ce qui peut expliquer les résultats obtenus via cette matrice pour cette typologie mais serait sans doute plus performante pour les typologies du centre comme les VSP.

3.8 Analyse des effets directs et indirects

On peut interpréter les coefficients des effets directs et indirects, dans un premier temps plus le taux de 15-29 ans est élevé dans l'arrondissement moins le taux de vol automobile est important, en revanche plus le taux de vol automobile chez les voisins est important.

A l'inverse pour la variable salaire médian, plus le revenu médian est important, plus le taux de vol automobile l'est aussi mais ce taux de vol décroît pour le voisinage.

En ce qui concerne le taux de scolarité, plus le taux de scolarité est important, moins le taux de vol automobile est important dans l'arrondissement ainsi que dans les arrondissements voisins.

Au total, cette variable impacte très fortement négativement le taux de vol automobile.

Le taux de chômage n'est pas significatif dans l'équation, il nous paraît cavalier d'en interpréter les effets d'autant que les résultats ne correspondent pas à l'intuition économique.

De manière générale, les résultats corroborent les premières intuitions que l'on pouvait avoir en pré-analyse.

Pour le modèle SLX:

```
Impact measures (SLX, estimable, n-k):
      Direct  Indirect  Total
log(taux15_29) -1.1045826  1.319197  0.2146143
log(MED15)      1.0814510 -2.336792 -1.2553408
log(taux_sco)   -2.2711306 -2.403168 -4.6742990
log(taux_chom)  0.1700614 -2.276118 -2.1060571
=====
Standard errors:
      Direct  Indirect  Total
log(taux15_29) 0.4519997 0.6303126 0.4713890
log(MED15)     0.4248356 0.8041096 0.6906335
log(taux_sco)  0.6626820 1.1909428 1.1506728
log(taux_chom) 0.3243588 0.5275877 0.4530932
=====
Z-values:
      Direct  Indirect  Total
log(taux15_29) -2.4437687  2.092925  0.4552807
log(MED15)     2.5455756 -2.906061 -1.8176657
log(taux_sco)  -3.4271803 -2.017871 -4.0622312
log(taux_chom)  0.5243004 -4.314199 -4.6481760

p-values:
      Direct  Indirect  Total
log(taux15_29) 0.01453474 0.0363559 0.648907
log(MED15)     0.01090977 0.0036601 0.069115
log(taux_sco)  0.00060988 0.0436047 4.8606e-05
log(taux_chom) 0.60006963 1.6018e-05 3.3488e-06
```

Figure 24: Effets directs et indirects du schéma spatial

On obtient des résultats très similaires pour le modèle SDM via la matrice 5-ppv.

| Impact measures (mixed, exact): | | | |
|---------------------------------|--------|----------|-------|
| | Direct | Indirect | Total |
| log(taux15_29) | -1.148 | 1.50 | 0.35 |
| log(MED15) | 0.928 | -2.02 | -1.09 |
| log(taux_sco) | -2.211 | -2.76 | -4.97 |
| log(taux_chom) | 0.123 | -2.21 | -2.08 |

Figure 25: Effets directs et indirects du schéma spatial

On obtient les intervalles de confiance empiriques par 1000 simulations MCMC.

On obtient les mêmes conclusions a priori cependant la volatilité rend les résultats non interprétables en partie pour les effets indirects et totaux.

4 Conclusion

Le prolongement spatial de l'étude sur la criminalité semble cohérent dans la mesure où un modèle économétrique simple ne parvient pas à capter toute l'information malgré la possible omission de variables explicatives également. Cela est justifié dans la mesure où un schéma spatial est détecté dans la criminalité du Grand Paris notamment pour certains type de criminalité tels que les vols à la tire et les vols d'automobile.

On observe une hausse du pouvoir explicatif, on obtient un R2 ajusté de 0.57 pour le modèle SLX contre 0.46 pour le modèle économétrique de base.

Il semble que les vols automobiles soient principalement localisés dans la couronne nord de Paris, dans des banlieues peu scolarisées, plus pauvres que Paris et moins jeunes.

Ceci s'explique vraisemblablement par la plus grande facilité à extraire les voitures de part le probable plus grand nombre de celle-ci, la circulation plus fluide et également la présence de voisins aux caractéristiques similaires.

On peut remarquer qu'étant donné la localisation des communes les plus touchées par les vols d'automobiles, on peut avoir de potentiels effets de bords dans notre échantillon.

On pourrait imaginer un modèle enrichi des données de la périphérie du Grand Paris et intégrer une variable explicatives de type "distance" à Paris intra-muros" pour tester cette hypothèse.

On a bien ici des phénomènes de "hot-cold spots" pour la criminalité de manière générale, que ce soit de manière agrégée ou par typologies, par ailleurs en faisant la distinction on peut voir qu'ils sont différents.

On pourrait également approfondir en étudiant des clusters mêlant différentes typologies étant donné les corrélations entre certaines d'entre elles. Les indices de repression policière pourrait être également intégrés afin de contrôler les pénalités qui balencent les bénéfices dans la théorie de Becker.

Une étude des outliers subsistants serait également riche d'information pour l'explication des mesures observées et l'affinage des modèles. Pour aller encore plus loin, on pourrait s'intéresser aux fonctions de demande qui régissent les comportements des voleurs automobiles définis auparavant, par ailleurs il semblerait que la tendance soit à une explosion du vol-revente par rapport au vol-utilisation.

Ce que l'on voit transparaître également dans les résultats des effets directs et indirects de notre modélisation.

Par ailleurs pour essayer de déterminer si et le cas échéant en quoi le Grand Paris est différent du reste des agglomérations françaises, on pourrait réaliser une étude comparative sur le reste des agglomérations françaises majeures.

On pourrait également améliorer la granularité pour écarter la possibilité de régression écologique et mieux observer les phénomènes et essayer d'écarter une hétérogénéité spatiale qui est possible.

References

- [1] ALISON.A : *The Economics of Crime: An Analysis of Crime Rates in America*
- [2] TREMBLAY PIERRE, CUSSON MAURICE, CLERMONT YVAN : *Contribution à une criminologie de l'acte: une analyse stratégique du vol de véhicules automobiles. Déviance et société. 1992 - Vol. 16 - N2. pp. 157-178.*

- [3] FRÉDÉRIC PUECH *Analyse des déterminants de la criminalité dans les pays en développement. Economies et finances. Université d'Auvergne - Clermont-Ferrand I, 2005*
- [4] FLACHAIRE.E :*Les méthodes du bootstrap dans les modèles de régression*, Economie et Prévision, Minefi - Direction de la prévision, 2001, 142, pp.183-194. halshs-00175894.
- [5] FLACHAIRE.E :*The Wild Bootstrap, Tamed at Last (February 2001). LSE STICERD Research Paper No. 58*
Available at SSRN: <https://ssrn.com/abstract=1094832>