

**M2 - Économétrie et Statistiques, parcours Économétrie
Appliquée**
IAE Nantes – Université de Nantes

Econométrie Spatiale

**Thème : Etude de la part des actifs salariés de
2014 dans les communes de la Gironde.**

ASSINE Ida Siting, TESSON Yohann et FOMBA Souleymane

Année universitaire : 2023-2024

Sommaire

I-Introduction	3
II-Analyse exploratoire	4
III-Méthodologie économétrique	26
IV-Estimation des modèles spatiaux.....	28
V-Conclusion.....	41
VI-Annexes.....	42
VII-Bibliographie	48
VIII-Table des matières	49

I. Introduction

La Gironde est le plus vaste département de France métropolitaine situé dans le Sud-Ouest de la France, en région Nouvelle-Aquitaine. ¹Cette région se plaçait en deuxième position concernant les créations d'emplois en 2014, derrière Midi-Pyrénées et devant l'Ile-de-France. En cinq ans, de 2009 à 2014, l'emploi y ait progressé de 2,2 %. En effet, une hausse significative de la masse salariale a été noté sur cette période avec 96.900 établissements implantés en Aquitaine pour 812.600 salariés. Ils déclarent 21,5 milliards de masse salariale, soit un salaire moyen par tête de 2.213 euros bruts mensuels. En cinq ans, la masse salariale a bondi de 12,7 %. Cette dynamisme de l'économie était portée de plus en plus par la Gironde et les Pyrénées-Atlantiques tandis que la Dordogne connaît les plus grosses difficultés, affichant 10, 9 % de taux de chômage.

C'est pourquoi dans le cadre de notre étude, suite à ce dynamisme économique de Aquitaine portée par la Gironde et les Pyrénées-Atlantiques, il est intéressant d'étudier la part des actifs salariés des communes de la Gironde afin de voir si cette croissance de l'emploi et de la masse salariale sur cette période précisément en 2014 était profitable à l'ensemble des communes de la Gironde où il y avait des disparités communales. En utilisant les outils puissants de l'économétrie spatiale, cette étude s'engage à cartographier et à analyser les dynamiques sous-jacentes qui façonnent cette distribution, éclairant ainsi les relations spatiales et les facteurs socio-économiques qui influent sur le tissu économique local. À travers cette approche novatrice, nous aspirons à dévoiler les intrications subtiles qui se cachent derrière les données, offrant ainsi des perspectives précieuses pour comprendre le développement économique dans les communes.

Les facteurs socio-économiques retenus dans cette étude et pouvant influencer la part des actifs salariés sont : le nombre d'habitants dans la commune, du revenu médian, du nombre de retraités dans la commune, la densité médicale, le nombre d'entreprises implanté dans la commune, le nombre de mineurs, le nombre d'étudiants, le score d'urbanité, le score d'équipement santé, le prix moyen au m² des biens immobiliers vendus et s'il s'agit d'une commune rurale ou pas.

¹ <https://www.20minutes.fr/bordeaux/1649011-20150709-aquitaine-deuxieme-region-plus-creatrice-emplois-2014>

Ainsi, pour bien effectuer cette étude spatiale de la part des actifs salariés des communes de la Gironde, nous disposons d'une base de données de 12 variables y compris la variable à expliquer et des données géoréférencées. Ces données nous permettront de définir les différentes relations spatiales entre les communes, d'estimer des modèles économétriques spatiales et de vérifier l'autocorrélation spatiale des résidus.

Pour une étude bien structurer, nous allons d'abord effectuer une analyse exploratoire de la base de données, ensuite expliciter la méthodologie économétrie à suivre et enfin faire l'estimation des modèles.

II. Analyse exploratoire

Notre étude portant sur la part des actifs salariés dans les communes de la Gironde de 2014, nous avons eu alors à notre disposition une base de données comportant 540 communes et des variables pouvant impacter la part des actifs salariés ces communes. Il s'agit des variables population qui représente le nombre d'habitant par commune et rurale qui est une variable binaire à savoir si la commune est considérée comme rurale ou non. Etant donné que l'objectif de l'étude est d'étudier les relations spatiales des individus et d'effectuer des modèles spatiaux, nous avons jugé intéressant de rajouter des variables explicatives. Ainsi, nous avons ajouté 9 variables dans la base de données. Il s'agit alors du revenu médian, du prix moyen au m² des biens immobiliers vendus, la population de 15 ou plus retraité, de la densité médicale, du score d'équipement santé, du nombre de mineurs, du nombre d'étudiants, du nombre d'entreprises dans chaque commune et du score d'urbanité. Le jeu de données de toutes ces variables a été collecté sur :

- Revenu médian : <https://www.insee.fr/fr/statistiques/3126432#consulter>
- Nombre de retraités : <https://www.insee.fr/fr/statistiques/2862200>
- Prix moyen au m² : <https://www.data.gouv.fr/fr/datasets/indicateurs-immobiliers-par-commune-et-par-annee-prix-et-volumes-sur-la-periode-2014-2023/>
- Reste des variables : <https://www.data.gouv.fr/fr/datasets/data-insee-sur-les-communes/>

Avec toutes ces variables susceptibles d'influencer la part des actifs salariés dans les communes de la Gironde, il convient alors de comprendre leurs caractéristiques et leurs relations potentielles afin de guider le processus de la modélisation. Pour ce faire, nous disposons d'une couche spatiale permettant de cartographier les données dont le référentiel est Lambert-93 EPSG 2154 et d'une base sous format csv où nous avons rajouté les variables.

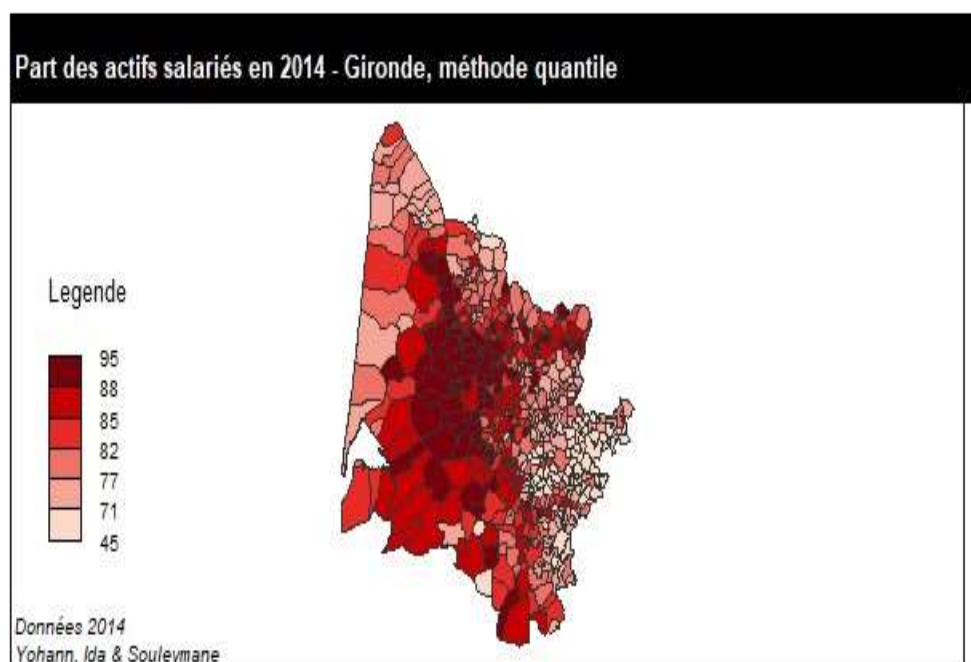
A. Présentation des données

a) La part des actifs salariés dans les communes de la Gironde

La part des actifs salariés représente la proportion des personnes actives dans une population donnée dans notre cas celle des communes de la Gironde de 2014, qui sont employées en tant que salariés, c'est-à-dire qu'elles perçoivent un salaire fixe ou variable en échange de leur travail au sein d'une entreprise ou d'une organisation. Cette mesure exclut généralement les travailleurs indépendants ou ceux qui sont employés dans le cadre de contrats temporaires ou informels. La part des actifs salariés est souvent utilisée comme indicateur du niveau d'emploi salarié dans une économie ou une région donnée, et elle peut être utile pour évaluer la santé économique, la structure de l'emploi et le niveau de protection sociale des travailleurs.

Cette variable est la variable à expliquer de l'étude.

Figure n°1 : La part des actifs salariés par commune en 2014

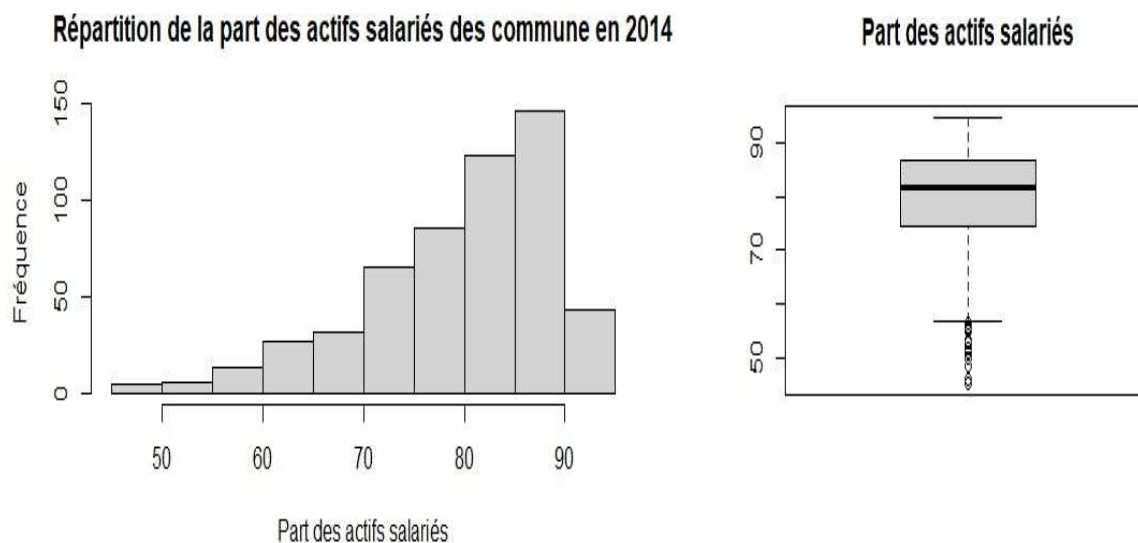


Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

La représentation cartographique ci-dessus de la part des actifs salariés des communes de la Gironde permet de visualiser spatialement la répartition de cette variable sur le territoire. Nous constatons que la part des actifs salariés varie d'une commune à l'autre. Et que les communes dont plus de 88% de la population active sont des actifs salariés, se situent au centre de la Gironde précisément les communes se trouvant sur Bordeaux métropole par exemple Bordeaux, Bruges, Le Bouscat, Mérignac, Le Haillan, Blanquefort,

Nous remarquons que plus la commune est loin de Bordeaux métropole plus la part des actifs salariés est petite. Cela suppose qu'il y a plus de zones d'activité industrielle dans la commune de Bordeaux et dans les communes aux alentours de ce dernier que les communes périphériques.

Figure n°2 : Boxplot et histogramme de la part des actifs salariés



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

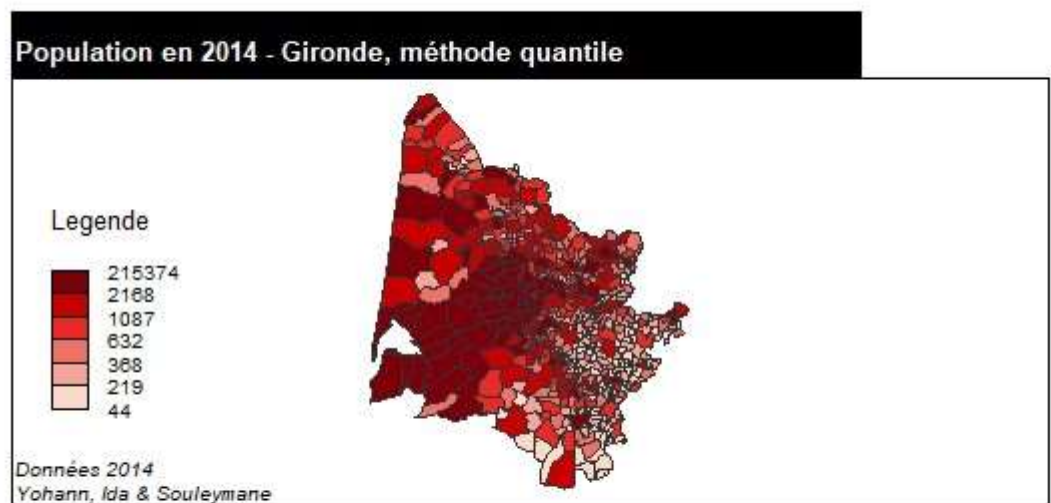
La répartition de la part des actifs salariés indique que la plupart des communes de la Gironde ont une part des actifs salariés comprise entre 80 et 90% par rapport à la population active. Et que cette distribution est susceptible d'avoir des individus atypiques. Mais après vérification si les individus extrêmes du boxplot sont atypiques ou pas par le test de Rosner, il s'est trouvé que la variable part des actifs salariés ne comporte pas d'individus atypique.

b) Population par commune de la Gironde

La population par commune fait référence au nombre total de résidents vivant dans une zone administrative spécifique délimitée par les frontières d'une commune. Cette mesure démographique est généralement basée sur les recensements de la population réalisés à intervalles réguliers par les autorités nationales ou locales.

Dans notre cas, il s'agit de population par commune de la Gironde du recensement de 2014. La variable population est essentielle pour comprendre les dynamiques du marché du travail et de l'emploi, ainsi que pour informer les politiques de développement économique et d'emploi. Dans l'étude de la part des actifs salariés, la taille et la composition de la population jouent un rôle crucial dans la détermination de la disponibilité et de la demande de main-d'œuvre salariée dans une commune donnée.

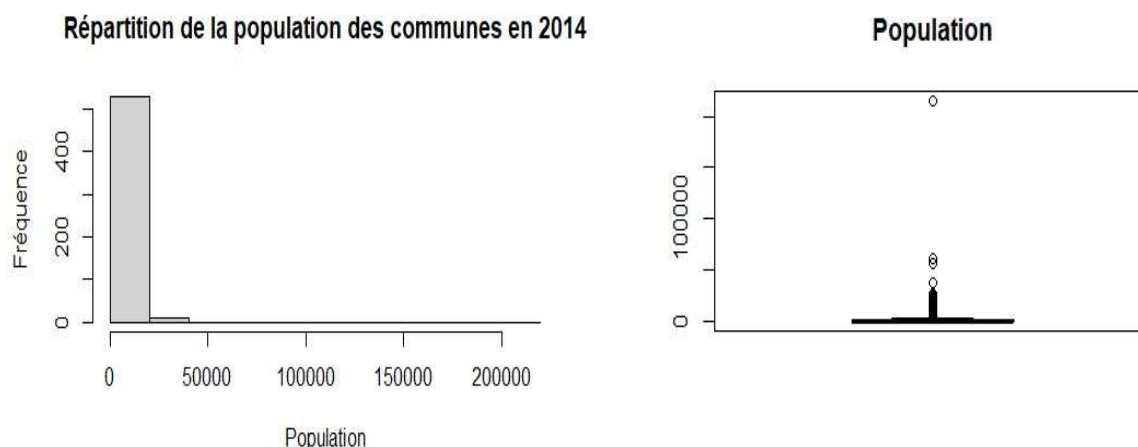
Figure n°3 : Population par commune de la Gironde



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Cette carte représente la population par commune de la Gironde. Nous observons qu'une grande partie des communes de la Gironde ont un nombre d'habitant supérieur à 2168. Certes, nous remarquons une disparité du nombre d'habitant d'une commune à l'autre, mais la population est plus concentrée de Bordeaux métropole qui est le centre vers l'ouest de la carte. Ce qui n'est pas le cas au Sud-Est car la majeure partie des communes ont une population comprise 44 et 632.

Figure n°4 : Boxplot et histogramme de population



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

La figure ci-haut montre que les communes de la Gironde n'ont pas une population qui excède 100 000 habitants sauf une seule. Cette commune est bien évidemment la commune de Bordeaux qui est composée de 215 374 habitants en 2014 et qui est représentée par le point extrême du boxplot.

c) Rurale

Une zone rurale est caractérisée par une faible densité de population et une prédominance d'activités agricoles, forestières ou naturelles par rapport aux activités urbaines ou industrielles. Les zones rurales sont généralement situées en dehors des grandes agglomérations urbaines et sont souvent associées à des paysages naturels, des villages ou des petites villes. Les principales caractéristiques des zones rurales incluent un environnement moins densément peuplé, une économie dominée par l'agriculture et d'autres activités primaires, ainsi qu'une infrastructure et des services publics moins développés par rapport aux zones urbaines.

Dans notre étude, l'importance de cette variable réside dans sa capacité à fournir un contexte économique et social unique qui influence la disponibilité, la nature et la distribution de l'emploi salarié dans la région. Comprendre ces dynamiques est essentiel pour élaborer des politiques et des stratégies de développement économique efficaces pour les zones rurales.

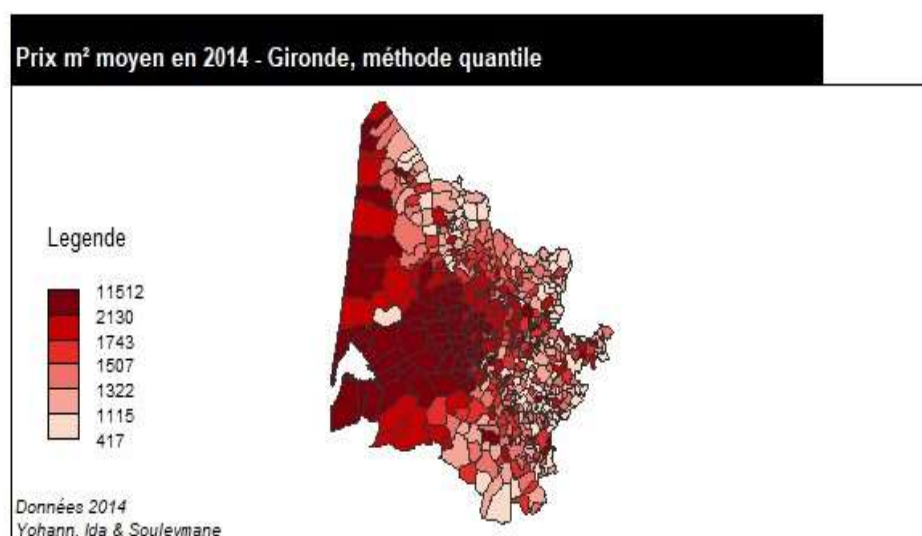
d) Le prix moyen au m² des biens immobiliers vendus

Le prix moyen au mètre carré (m²) des biens immobiliers vendus est une mesure qui indique la valeur moyenne par unité de surface des propriétés vendues dans une région donnée sur une période de temps spécifique. Cette mesure permet d'évaluer le niveau général des prix immobiliers et de suivre les tendances du marché immobilier.

Le choix de cette variable dans notre étude permet fournir des informations importantes sur le coût de la vie, l'accessibilité au logement, la mobilité de la main-d'œuvre et le développement économique local, tous des facteurs qui peuvent influencer la part des actifs salariés dans une commune donnée.

Le jeu de données de cette variable a été collecté sur site data.gouv.fr pour l'année 2014 puis rajouter à notre base de données.

Figure n°5 : Prix moyen au m² des biens immobiliers vendus par commune de la Gironde

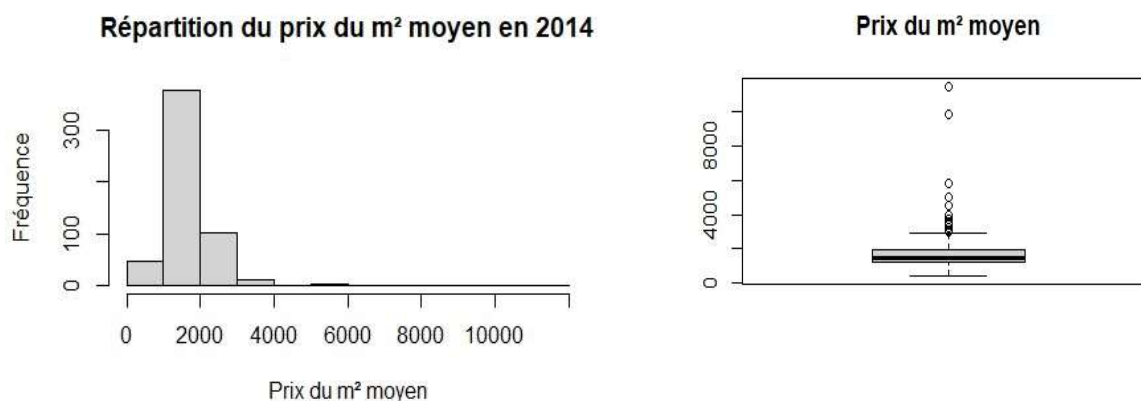


Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

La représentation spatiale du prix moyen au m² des biens immobiliers vendus en 2014 par commune de la Gironde permet de constater que le prix moyen au m² est supérieure à 2130 au niveau des communes de Bordeaux métropole, du Bassin d'Arcachon et du Médoc Atlantique. Ce montant élevé à Bordeaux Métropole s'explique par la concentration de l'activité

économique dans ces communes mais aussi du nombre d'habitant qui y réside. Pour ce qui est du prix au Bassin d'Arcachon et au Médoc Atlantique, la cherté des biens immobiliers s'explique du fait qu'il s'agit d'une zone littorale et touristique.

Figure n°6 : Boxplot et histogramme du prix moyen au m² des biens immobiliers vendus



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

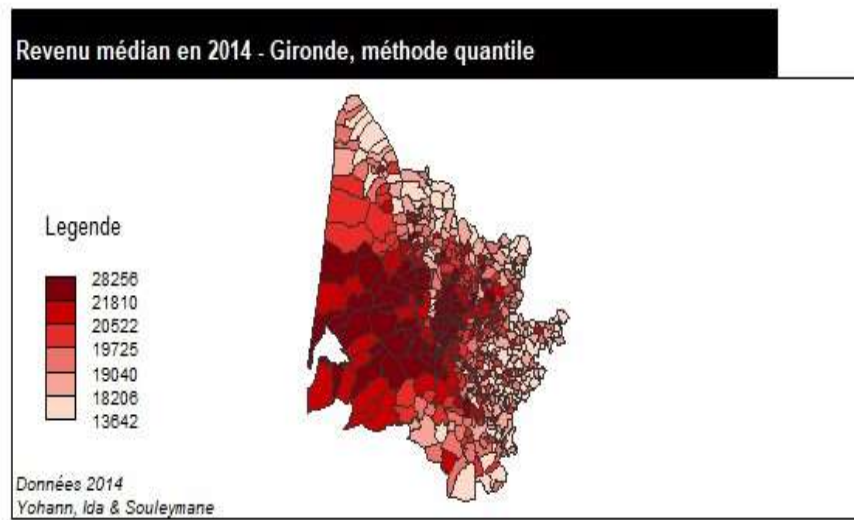
A partir de la figure ci-dessus, nous remarquons que le prix moyen au m² des biens immobiliers vendus en 2014 dans les communes de la Gironde était dans la majeure partie des communes entre 1000 et 2000. Le boxplot indique qu'il est possible d'avoir des individus atypiques dans la distribution. Puisqu'il s'agit d'une étude spatiale, nous avons effectué l'étude des valeurs atypiques que sur la variable à expliquer afin d'éviter la modification de la carte par la suppression des individus atypiques.

e) Le revenu médian

Le revenu médian est le revenu qui divise la population en deux groupes égaux : la moitié de la population gagne plus que le revenu médian et l'autre moitié gagne moins. C'est une mesure statistique centrale qui est souvent utilisée pour représenter le niveau de revenu typique au sein d'une population donnée.

Le revenu médian est un indicateur important dans l'étude de la part des actifs salariés car il fournit des informations sur le niveau de rémunération typique, les disparités de revenus, l'impact sur le marché du travail et les politiques d'emploi et de rémunération. Ces facteurs peuvent influencer la disponibilité, la qualité et les conditions des emplois salariés dans une commune donnée.

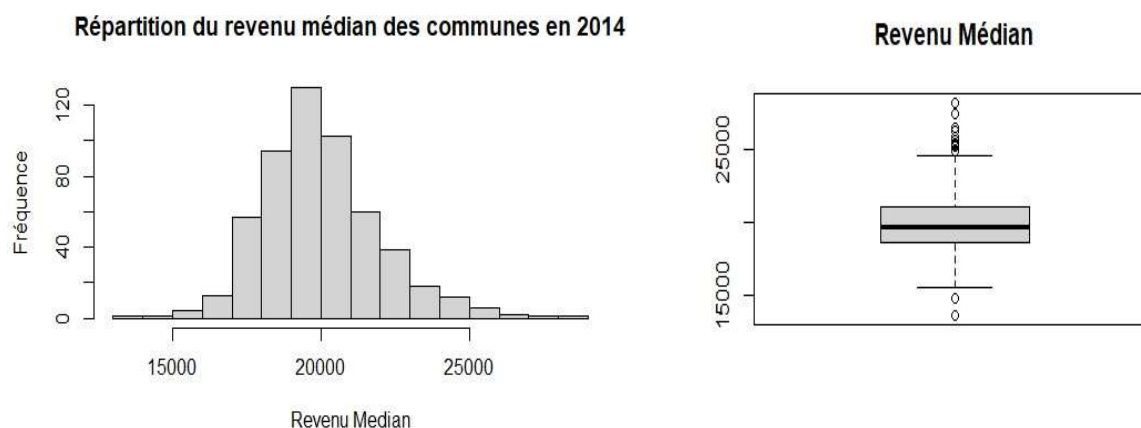
Figure n°7 : Revenu médian par commune de la Gironde



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Nous pouvons remarquer à partir de la carte que le revenu médian entre 19725 et plus de 21810 dans les communes de la Gironde situant au centre vers l'ouest de la carte. Il s'agit presque des mêmes communes où le prix moyen au m² des biens immobiliers vendus est plus élevé. C'est-à-dire des communes se situant à Bordeaux Métropole et sur le littoral. Ce qui est logique car il y a plus une concentration de l'activité économique sur ces zones que dans les communes du Sud-Est.

Figure n°8 : Boxplot et histogramme du revenu médian



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

Quant à la répartition du revenu médian, nous remarquons que la distribution suit une loi normale et que le pic est atteint à 20000. Cette variable également est susceptible d'avoir des individus atypiques.

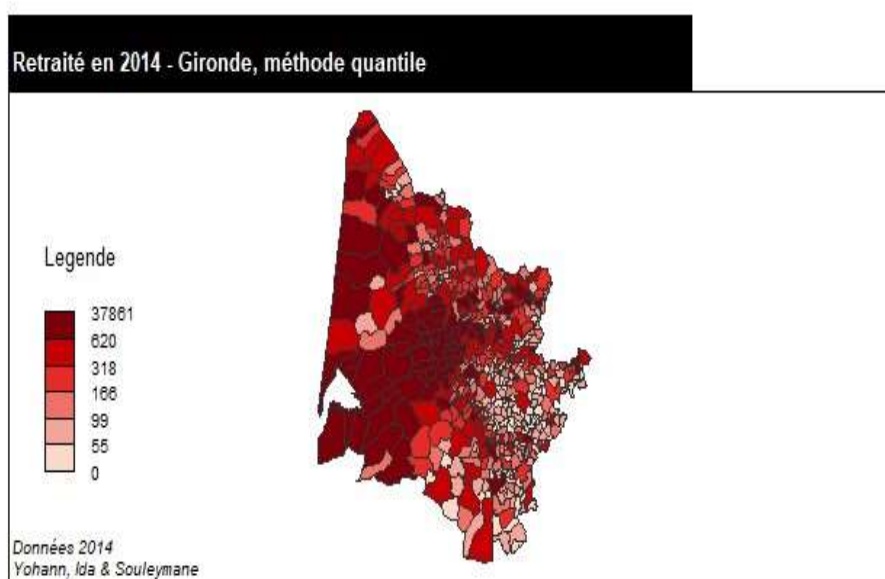
f) Le nombre de retraités

La retraite est une période de cessation de l'activité professionnelle rémunérée, généralement associée à l'âge avancé, où une personne cesse de travailler et commence à recevoir un revenu régulier, souvent sous la forme de pensions ou d'autres prestations, généralement versées par un régime de retraite, un fonds de pension ou un système de sécurité sociale. C'est une phase de la vie où l'individu cesse de participer activement à la force de travail et commence à profiter de ses années de travail en bénéficiant de revenus prévus pour lui assurer un niveau de vie confortable après sa carrière professionnelle.

L'importance de cette variable est que le nombre de retraités peut avoir un impact significatif sur la dynamique du marché du travail et sur la disponibilité des emplois salariés dans les communes de la Gironde. Comprendre cette variable démographique est essentiel pour évaluer les tendances de l'emploi et informer les politiques de développement économique et social au niveau local.

Ce jeu de données est collecté sur le site de l'INSEE.

Figure n°9: Nombre de retraité par commune de la gironde

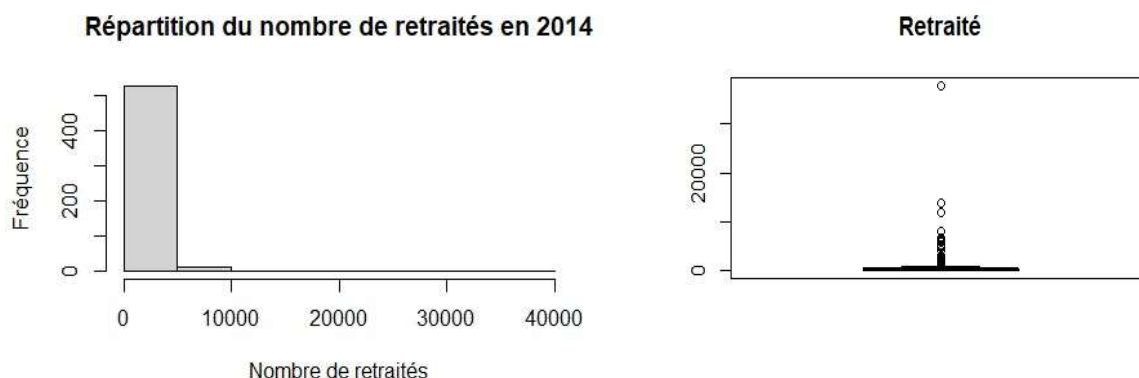


Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Au regard de la carte de la gironde par rapport au nombre de retraités, nous constatons qu'il y a plus de 620 retraités dans les communes se situant au centre de la carte et dans celle se trouvant à l'Est c'est-à-dire sur le littoral. Ce nombre élevé de retraités sur les communes littorales et au niveau celles de bordeaux métropole, est dû au fait qu'après la retraite, ces derniers peuvent contribuer à l'économie locale en tant que consommateurs de biens et services, ce qui peut stimuler certaines industries comme le tourisme, les loisirs et les services de santé.

Alors qu'au Sud-Est de la carte, le poids des retraités n'est pas élevé.

Figure n°10 : Boxplot et histogramme du nombre de retraité



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

Le nombre de retraités dans majeure partie des communes de la Gironde est comprise entre 0 et 5000, cela montre que certaines communes de la Gironde sont confrontées à une population vieillissante qui peut entraîner un besoin de main d'œuvre. Sur le boxplot, nous remarquons que le nombre de retraités sur une commune dépasse 20000. Bien évidemment, d'après la carte et les analyse précédente, nous pouvons déduire qu'il s'agit de la commune de Bordeaux où le nombre de retraités est à 37861.

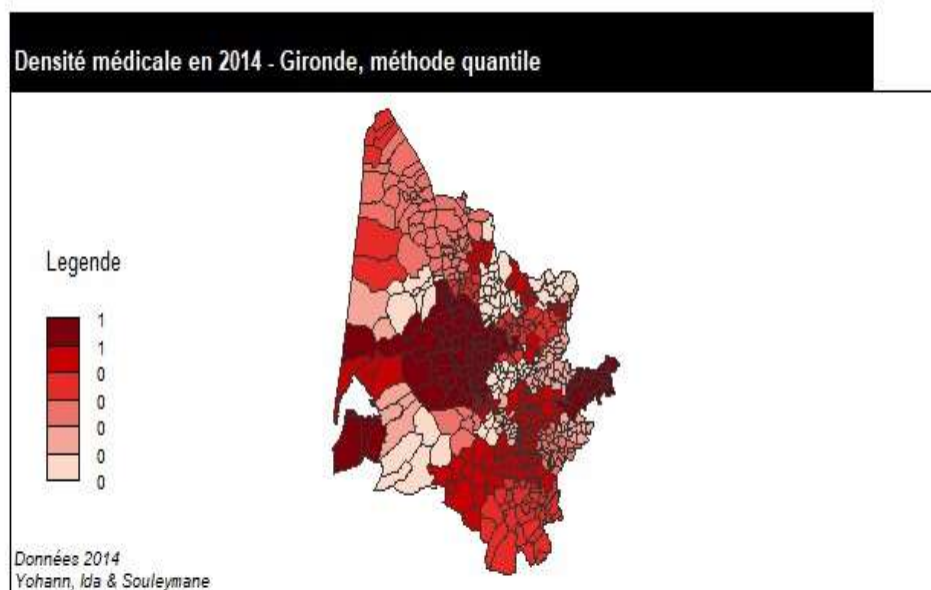
g) La densité médicale

La densité médicale est une mesure qui indique le nombre de professionnels de la santé, tels que les médecins, les infirmières ou les spécialistes de la santé, par rapport à la population d'une région donnée. Elle est généralement exprimée en fonction du nombre de professionnels de la santé pour une certaine population, souvent par millier d'habitants. Cette mesure permet d'évaluer la disponibilité des services de santé dans une région donnée et peut influencer l'accès aux soins de santé pour la population locale. Dans notre cas, elle est mesurée par commune.

Une densité médicale élevée peut contribuer à améliorer la qualité de vie des habitants en assurant un accès facile et rapide aux services de santé. Cela peut favoriser la santé et le bien-être des travailleurs, ce qui peut influencer leur capacité à maintenir un emploi salarié.

La densité médicale peut jouer un rôle important dans l'étude de la part des actifs salariés en influençant la santé et le bien-être des travailleurs, l'attrait des communautés pour les nouveaux résidents et entreprises, ainsi que les coûts associés aux soins de santé pour les individus et les entreprises.

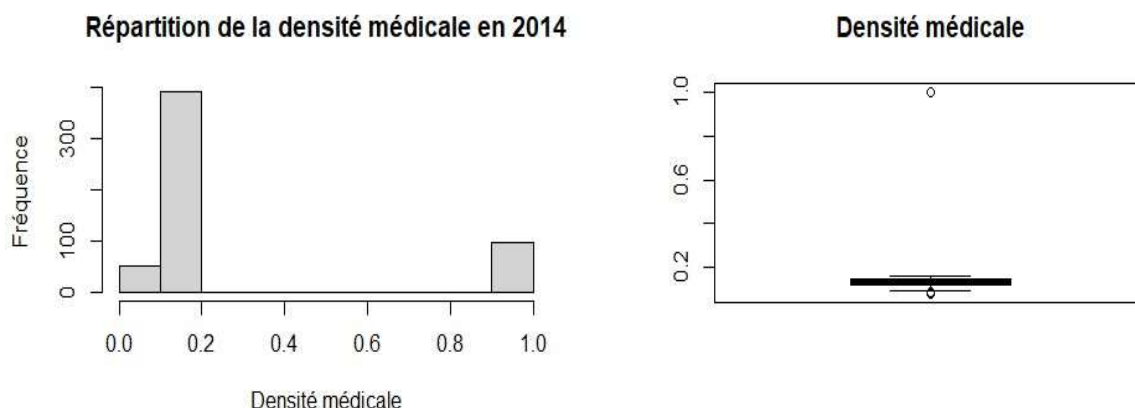
Figure n°11 : Densité médicale par commune de la Gironde



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

La représentation cartographique indique une forte densité médicale dans les communes situant à Bordeaux métropole, au Bassin d'Arcachon, dans les communautés de commune du Sud de la Gironde, du Bazadais et du Pays Foyen. La légende ne montre que des chiffres entiers mais les zéros sont des valeurs décimales (voir la figure ci-dessous).

Figure n°12 : Boxplot et histogramme de la densité médicale



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

La figure ci-haut montre qu'il y a un faible ratio de professionnels de la santé par rapport à la population dans la majeure partie des communes de la Gironde car le pic de l'histogramme est entre 0,1 et 0,2. Cela peut avoir des implications sur l'accessibilité et la disponibilité des services de santé dans ces communes.

h) Le score équipement santé

Le score d'équipement santé est une mesure qui évalue la disponibilité et l'accessibilité des services de santé dans une commune donnée. Il prend en compte divers facteurs tels que le nombre de professionnels de la santé, les établissements médicaux, les équipements médicaux et les infrastructures de santé disponibles dans la commune. Ce score permet d'évaluer la qualité et la quantité des services de santé accessibles à la population locale.

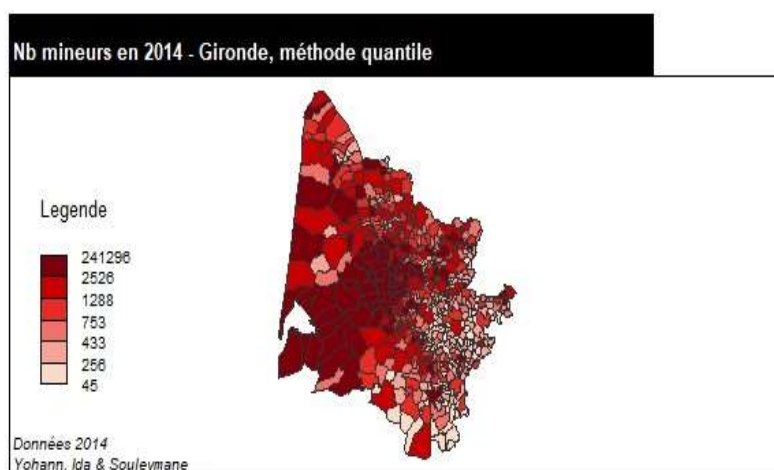
Un score d'équipement santé élevé indique généralement une meilleure disponibilité des services de santé dans la commune. Il s'agit d'une variable qualitative.

i) Le nombre de mineurs

Un mineur est une personne qui est en dessous de l'âge légal de la majorité, généralement définie par la loi. Dans de nombreux pays, un mineur est une personne âgée de moins de 18 ans, bien que cette définition puisse varier selon les juridictions. Les mineurs sont souvent considérés comme nécessitant une protection et une supervision spéciales en raison de leur jeune âge et de leur manque d'expérience, et leurs droits et responsabilités peuvent être limités par la loi en conséquence.

La population des mineurs joue un rôle crucial dans l'économie et la société locales en tant que future main-d'œuvre potentielle, bénéficiaires de services sociaux et éducatifs, et membres de familles qui peuvent influencer les décisions de travail et d'emploi des adultes. La prise en compte de cette variable est importante pour une analyse complète de la dynamique du marché du travail dans les communes de la Gironde.

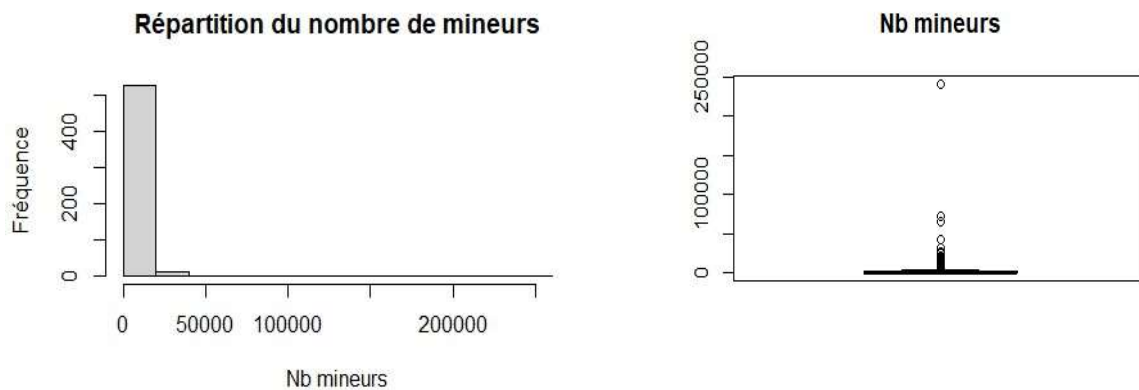
Figure n°15 : Nombre de mineurs par commune de la Gironde



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

La carte montre qu'il y a un nombre important de mineurs soit plus de 1288 un peu partout dans les communes de la Gironde mais il convient de préciser que cette population est nombreuse dans les communes de Bordeaux métropole, du Bassin d'Arcachon et sur les communes se trouvant sur le littoral. Le poids de cette population sur ces lieux suit la logique du fait qu'il s'agit des zones économiques où la part de la population active est plus élevée et que les mineurs dépendent des parents.

Figure n°16 : Boxplot et histogramme du nombre de mineurs



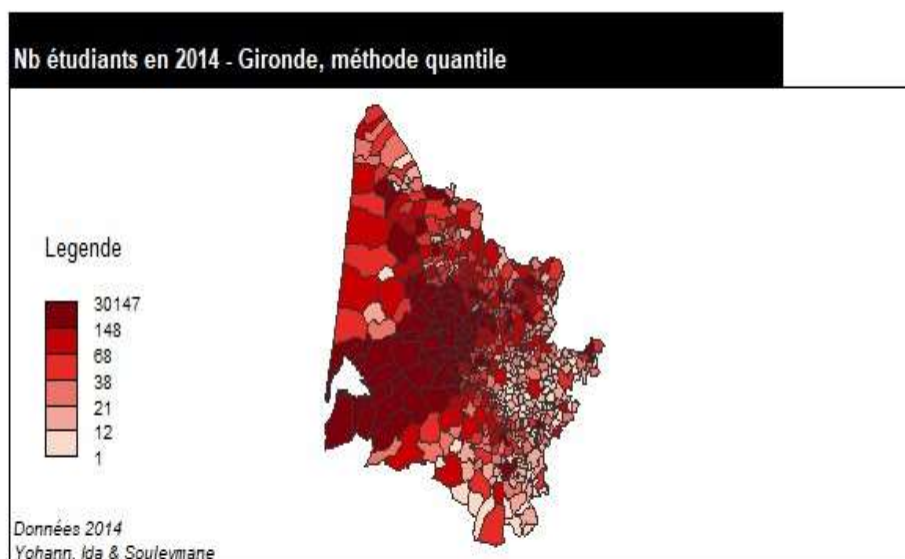
Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA*

La représentation ci-haut indique la répartition du nombre de mineurs sur les communes et que le nombre qui revient le plus est entre 1288 et plus de 2528.

j) Le nombre d'étudiants

Les étudiants représentent la future main-d'œuvre dans les communes. Leurs choix d'éducation et de formation peuvent influencer la disponibilité future de travailleurs qualifiés dans différents secteurs d'emploi. Leur participation future à la population active peut avoir un impact sur la disponibilité des emplois et les compétences recherchées par les employeurs.

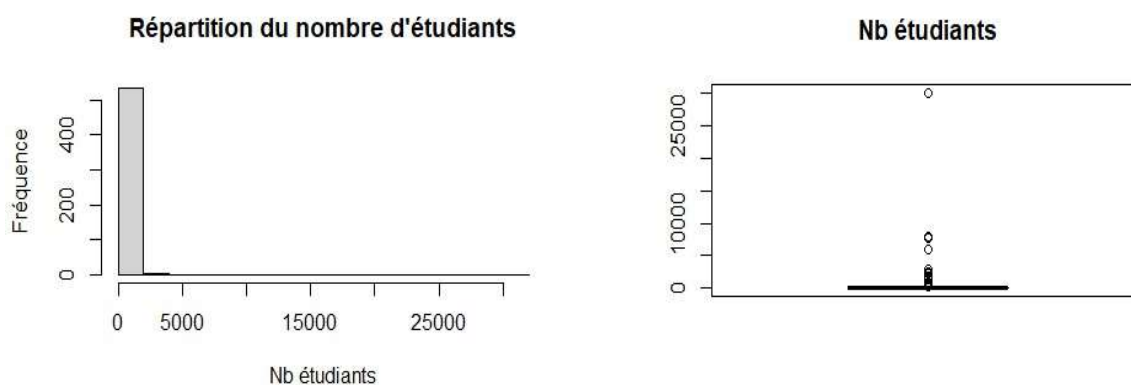
Figure n°17 : Nombre d'étudiants par commune de la Gironde



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Le nombre d'étudiants est plus élevé dans les communes de Bordeaux métropole et celles du Bassin d'Arcachon. Leur nombre élevé sur ces lieux reflète la disposition des offres de formations sur ces lieux.

Figure n°18 : Boxplot et histogramme du nombre d'étudiants



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

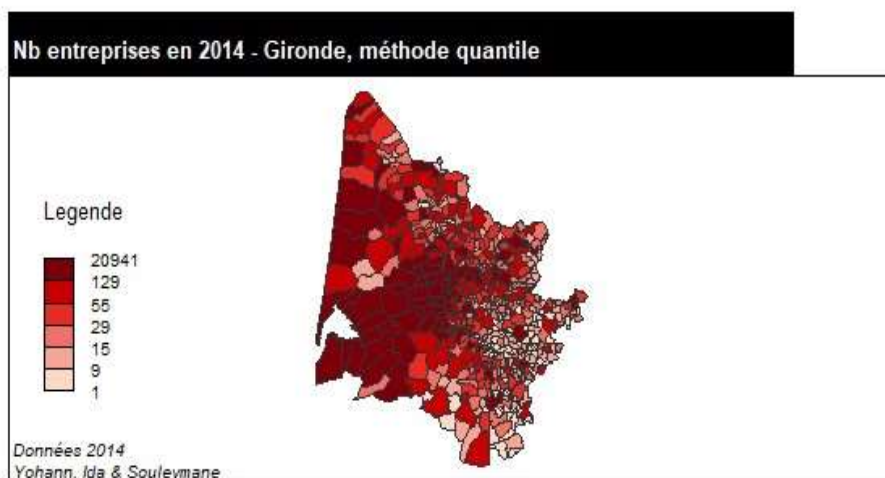
La répartition du nombre d'étudiants n'étant pas normal, montre les disparités des offres de formation d'une commune à l'autre. Ainsi, par rapport au boxplot, nous remarquons qu'il y a une commune où le nombre d'étudiants dépasse 25000.

k) Le nombre d'entreprises

Une entreprise est une entité économique qui s'engage dans des activités commerciales dans le but de réaliser un profit. Elle peut prendre différentes formes juridiques, telles que les sociétés anonymes, les sociétés à responsabilité limitée, les entreprises individuelles, etc. Les entreprises produisent des biens ou des services qui sont ensuite vendus sur le marché pour répondre aux besoins ou aux désirs des consommateurs. Elles emploient également du personnel pour effectuer les opérations commerciales et administratives nécessaires au fonctionnement de l'entreprise.

Les entreprises sont les principaux pourvoyeurs d'emplois salariés dans une commune donnée. Leur présence et leur activité économique contribuent à stimuler la demande de main-d'œuvre et à créer des opportunités d'emploi pour les habitants de la commune. Elles opèrent dans une variété de secteurs économiques, tels que l'industrie, le commerce, les services, etc. Leur diversité offre aux travailleurs salariés un large éventail d'opportunités professionnelles et contribue à la robustesse de l'économie locale.

Figure n°19 : Nombre d'entreprises par commune de la Gironde

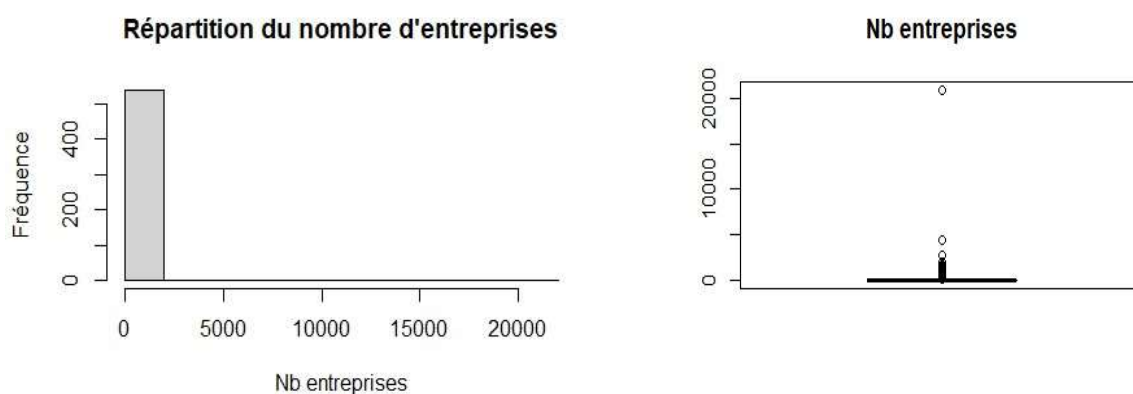


Source :

Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMB

La cartographie permet de voir que dans l'ensemble les communes de la Gironde sont caractérisées par une forte activité commerciale et industrielle.

Figure n°20 : Boxplot et histogramme du nombre d'entreprises



Source : *Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMB*

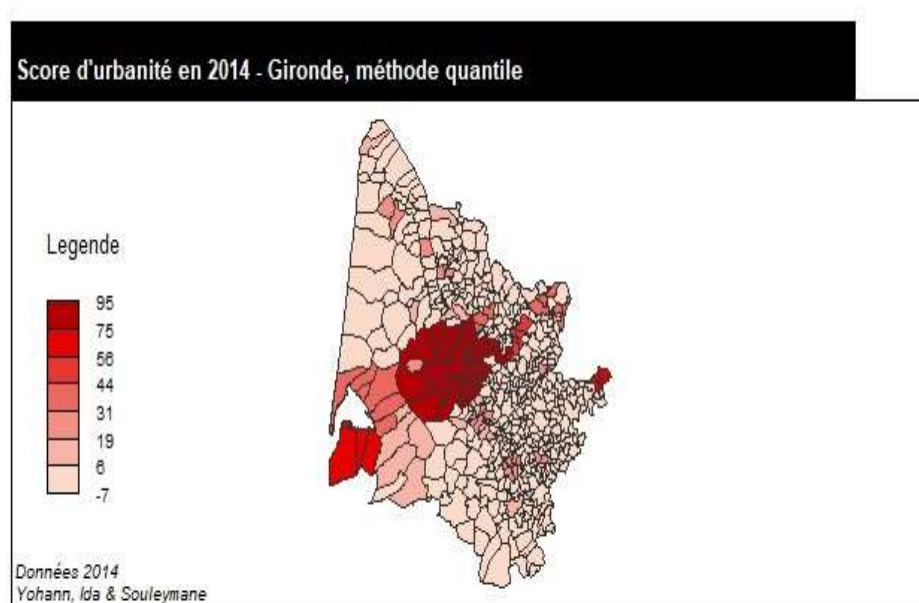
Nous pouvons remarquer que la distribution n'est pas normale et que le pic de l'histogramme est atteint plus de 129 entreprises par commune. Il y a une commune qui a plus de 20000 entreprises.

l) Le score d'urbanité

Le score d'urbanité est une mesure qui évalue le degré d'urbanisation d'une zone géographique donnée. Il prend en compte divers facteurs tels que la densité de population, la taille de la commune, la présence d'infrastructures urbaines telles que les transports en commun, les services publics, les équipements commerciaux, etc.

Les zones plus urbaines ont tendance à offrir une plus grande variété d'emplois salariés dans différents secteurs économiques en raison de leur concentration d'entreprises et d'infrastructures commerciales. Par conséquent, le score d'urbanité peut fournir des indications sur la disponibilité des emplois dans une commune donnée.

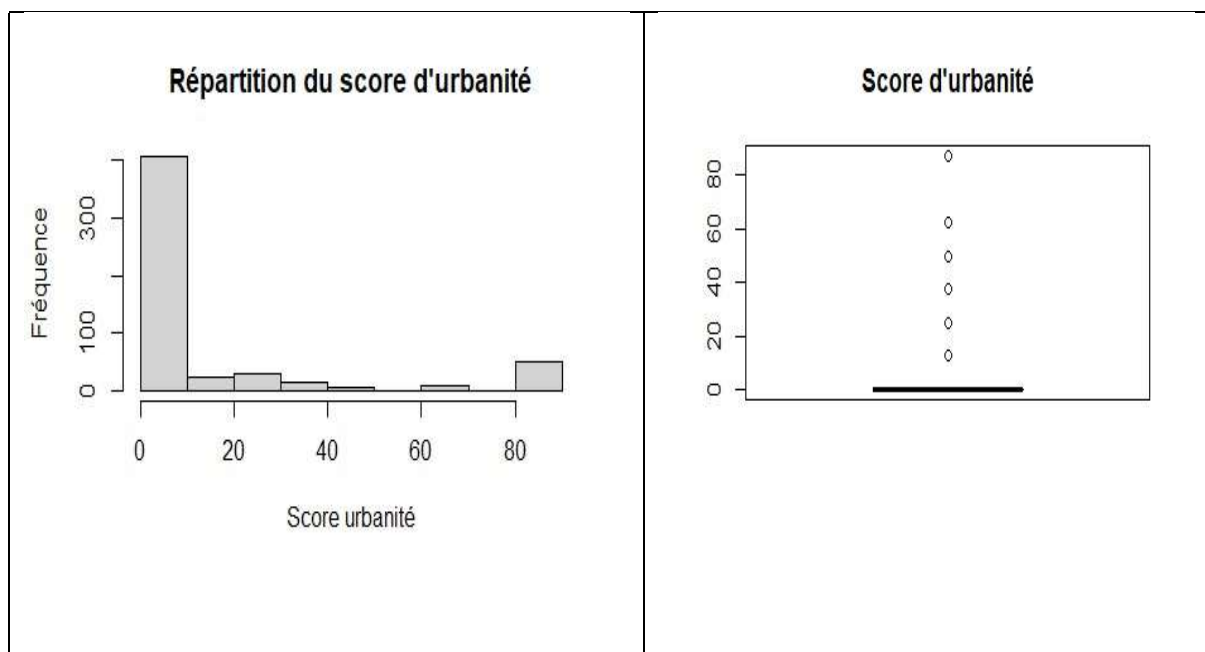
Figure n°21 : Score d'urbanité par commune de la Gironde



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMB

La représentation spatiale du score d'urbanité montre qu'il y a plus de commune rurale en Gironde. Les communes ayant un score supérieur à 44 se trouve à Bordeaux métropole et au niveau du Bassin d'Arcachon.

Figure n°22 : Boxplot et histogramme du score d'urbanité



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMB

La distribution du score d'urbanité n'est pas normale. Le score fréquemment obtenu par les communes de la Gironde est entre 0 et 10.

B. Statistiques descriptives

Tableau1 : Statistiques descriptives

	Min	1er Quartile	Médiane	Moyenne	3ème Quartile	Max
Part actif salaries	45,16	74,58	81,77	79,65	86,61	94,84
Population	44	276	632,5	2375,5	1517,2	215374
Prixm2Moyen	417	1214	1509	1651	1927	11512
Revenu médian	13642	18653	19728	19981	21072	28256
Retraite	0	76,31	166,17	599,4	408,48	37860,96
Densité médicale	0,081	0,124	0,137	0,284	0,145	1
Nb mineurs	45	324,2	753	2772,4	1783,2	241296
Nb étudiants	1	16	38	218,7	95	30147
Nb entreprises	1	12	29	156,2	80	20941
Score urbanité	0	0	0	12,64	0	87,5

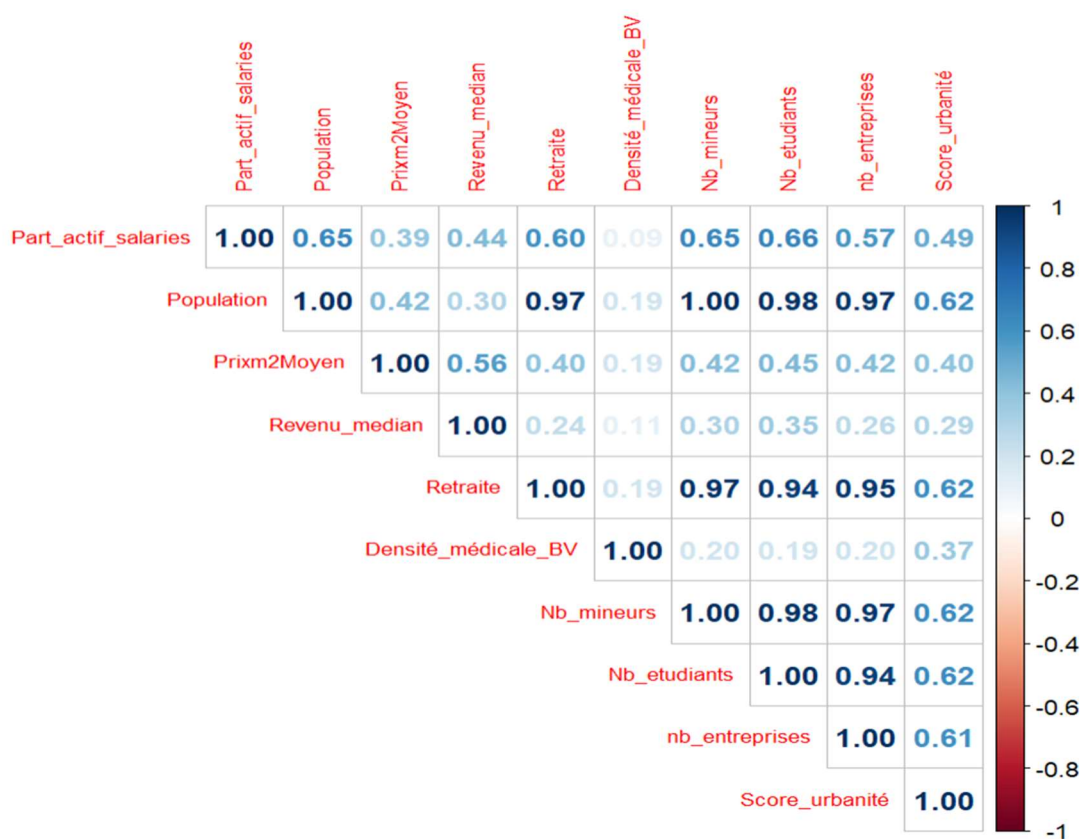
Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Le tableau ci haut présente les statistiques descriptives de l'ensemble de nos variables quantitatives. Nous pouvons remarquer que sur toutes les variables, il y a une grande différence entre le minimum et le maximum. La part moyenne des actifs salariés des communes de la Gironde est de 79,65. La population moyenne est de 2375,5. Le prix moyen au m² des biens immobiliers vendus est de 1651. Le revenu médian est en moyenne de 19981. En moyenne, le nombre de retraités en Gironde est 599,1. La densité médicale de la Gironde est en moyenne de 0,284. Le score d'urbanité est en moyenne de 12,64. Le nombre moyen d'entreprises est de 156,2.

Nous constatons que pour la plupart des variables, il y a une grande différence entre la médiane et la moyenne. Et que c'est la moyenne qui est plus grande. Cela indique que la distribution de ces variables est asymétrique et les données sont fortement influencées par un petit nombre de valeurs extrêmement élevées.

C. Analyse des corrélations

Figure n°23 : Matrice de corrélation



Source : Dossier économétrie spatiale, Yohann TESSON, Ida ASSINE & Souleymane FOMBA

Au regard de la matrice de corrélation, nous constatons qu'il y a une forte corrélation entre certaines de nos variables explicatives et la variable à expliquer. La part des actifs salariés a une corrélation positive avec population, nombre de retraités, nombre de mineurs, nombre d'étudiants et nombre d'entreprises. La variable population a une corrélation positive et significative avec nombre de retraités, nombre de mineurs, nombre d'étudiants et nombre d'entreprises. Ces 4 dernières variables sont fortement corrélées entre elles. Elle convient alors de prêter une attention dans la suite de l'analyse à ces variables afin d'éviter un problème de multicolinéarité et que nos modèles ne soient pas biaisés. Pour ce faire, un choix de variables à introduire dans les modèles sera fait.

III. Méthodologie économétrique

Notre étude consiste à étudier la part des actifs salariés dans les communes de la Gironde. Etant donné qu'il s'agit d'une étude de la proportion des salariés de façon géographique, la méthodologie la plus adaptée alors est celle de l'économétrie spatiale. L'économétrie spatiale désigne l'ensemble des méthodes économétriques permettant de traiter les effets spatiaux ou interactions spatiales dans les modèles de régression en coupe transversale. Il faut distinguer deux types d'effet spatiaux : une hétérogénéité spatiale et une autocorrélation spatiale. Dans, notre étude c'est l'autocorrélation spatiale que nous cherchons à identifier. C'est-à-dire, nous allons vérifier si la part des actifs salariés observée dans une commune de la Gironde dépend-elle de la part des actifs salariés observée dans une autre commune voisine ?

Pour ce faire nous allons d'abord définir les différentes relations spatiales entre nos individus en construisant les matrices de poids du critère de contiguïté d'ordre 1 : du plus proche voisin, de type Reine, type Tour et de 3 voisins.

Ensuite, nous allons estimer un modèle de régression linéaire multiple par la méthode des moindres carrés ordinaires (MCO).

Puis vérifier en utilisant la fonction VIF (Variance Inflation Factor) si le modèle estimé à un problème de multicolinéarité entre les variables explicatives. Si le modèle ne fait pas l'objet de multicolinéarité, nous calculons l'indice de Moran sur les résidus de l'estimation afin de voir s'il y a une autocorrélation spatiale des résidus. Ce qui nous permettra de tester l'hypothèse :

H0 : Absence d'autocorrélation spatiale des résidus

Si on observe une autocorrélation spatiale des résidus par le rejet de **H0**, alors il convient d'effectuer le test de Lagrange afin de déterminer le modèle adéquat. C'est-à-dire entre le modèle **RLMerr** et celui **RLMlag** lequel choisir.

La sortie du résultat du test de Lagrange, nous permet d'avoir les résultats du **LMerr**, **LMlag**, **RLMerr** et **RLMlag**. Mais, nous nous intéresserons qu'aux résultats des modèles robustes c'est-à-dire les deux derniers. Ainsi, si la p-value d'un ou des deux modèles est significative, on estime un modèle SDM (Spatial Durbin Model) puis le comparé avec le modèle SEM (Spatial Error) puis le modèle SAR (Spatial Autoregressive).

Le meilleur modèle à choisir est celui qui minimisera l'AIC et dont le LM test est significatif.

IV-Estimation des modèles spatiaux

A- Mesure de l'autocorrélation spatiale

a) Mesure de l'indice de Moran (autocorrélation spatiale globale)

Nous allons, dans un premier temps mesurer l'indice de Moran des différentes matrices de poids. Cette indice permet d'évaluer la dépendance spatiale globale d'un ensemble de données. Elle mesure la similarité des valeurs d'une variable entre différentes unités spatiale (ici les communes) et leur proximité géographique. Si les valeurs similaires tendent à être proches les unes des autres dans l'espace, l'indice de Moran sera élevé, indiquant une forte autocorrélation spatiale. A l'inverse, un indice de Moran faible ou négatif suggère une distribution spatiale aléatoire ou dissemblable. Le tableau ci-dessous nous montre les résultats obtenus pour les 4 matrices de poids.

Tableau 2 : Résultat des indices de Moran pour les 4 matrices de poids

Matrice de poids	Indice de Moran	P-value
Type Reine	0.63577	2.2 ^e -16
Type Tour	0.63630	2.2 ^e -16
Plus proche voisin	0.68945	2.2 ^e -16
3 plus proche voisin	0.65138	2.2 ^e -16

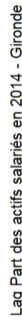
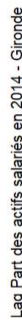
Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

On pose comme hypothèse nulle H_0 , le fait qu'il n'y ait pas d'autocorrélation spatiale globale, suggérant une dispersion aléatoire entre les communes. D'après le tableau des résultats, les p-values sont toutes inférieures à 0.05. On rejette donc H_0 . Il existe donc une relation entre le part des actifs salariés des différentes communes de Gironde. Les valeurs des indices étant positive et variant de 0.63 à 0.69, les valeurs similaires de la variable en question se regroupent donc géographiquement de manière globale. Il est ainsi nécessaire de tester des modèles d'autocorrélation spatiale pour les 4 matrices de poids.

Figure n°24 : Diagramme de Moran selon les différentes matrices de poids



Matrice type Tour

**Matrice - PPV1****Matrice - PPV3**

Source : *Dossier Econométrie spatiale*, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

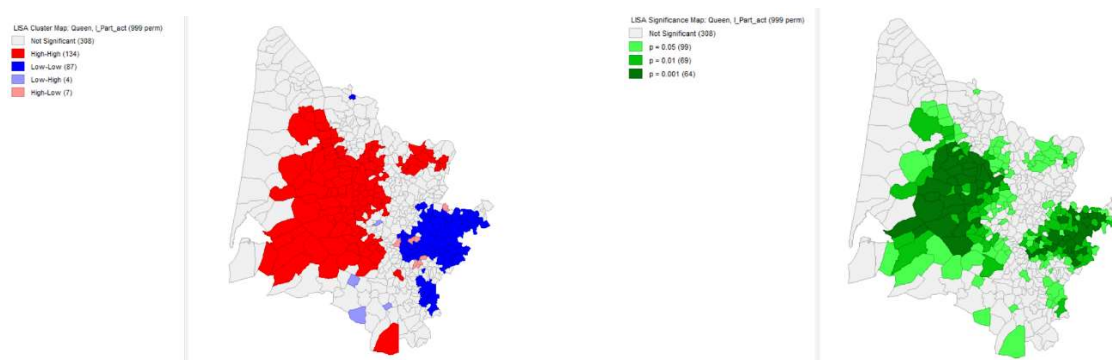
Pour les 4 matrices de poids, on voit que les points se regroupent bien le long de la ligne de régression qui dispose d'une pente positive assez importante. L'indice de Moran est donc positif et élevé. Ceci est également illustré par la présence de clusters dans les quadrants High-High et Low-Low. Très peu de valeurs sont représentées dans les quadrants supérieur gauche et inférieur droit.

b) LISA (Local Indicators of Spatial Association) – autocorrélation spatiale locale

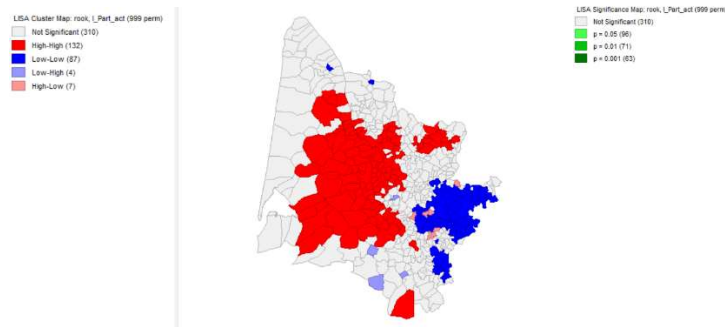
Nous pouvons dans un second temps mesurer l'autocorrélation spatiale locale de chaque commune par le calcul du LISA (Local Indicators of Spatial Association). Les LISA sont utilisés pour détecter les clusters spatiaux significatifs dans les données, tels que les « hot spots » (points chauds) et les « cold spots » (points froids). Les hot spots sont des unités spatiales où les valeurs de la variable étudiée sont significativement élevées et sont entourées d'autres unités spatiales ayant également des valeurs élevées. Cela indique une concentration spatiale de valeurs élevées. Les cold spots, quant à eux, sont des unités spatiales où les valeurs de la variable étudiée sont significativement faibles et sont entourées d'autres unités spatiales ayant également des valeurs faibles. Cela indique une concentration spatiale de valeurs faibles. À l'aide du logiciel Geoda, sont représentés graphiquement les LISA des communes ainsi que leur seuil de significativité.

Figure n°25 : Cartes des LISA ainsi que leur significativité des différentes communes de Gironde

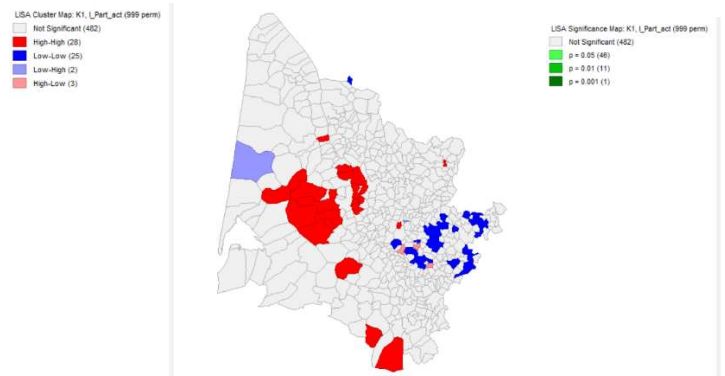
Queen :



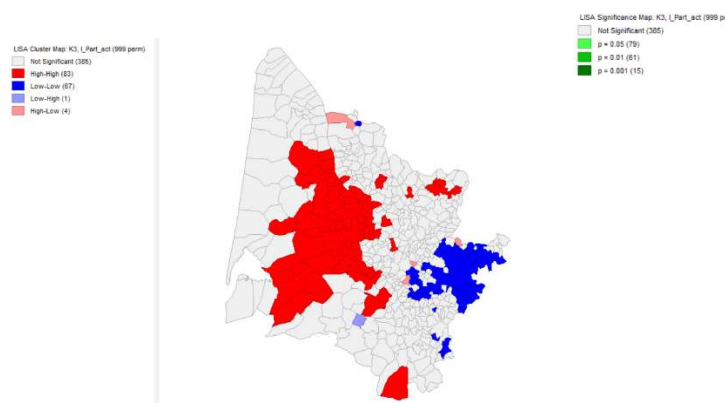
Tour :



PPV1 :



PPV3 :



Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

Les 4 graphiques nous montrent des cartes à peu près semblable mise à part celui du PPV1 où l'on constate qu'il y a moins de LISA significatives. Mais en règle générale, on observe les mêmes clusters. Sont représentés en rouge les « hot spots » et en bleu les « cold spots ». Il y a également des couleurs avec des teintes plus transparentes. Ces dernières représentent des communes ayant des dissemblances significatives avec ses voisins. Les cartes à droite représentent quant à elles, le seuil de significativité de chaque LISA. Cela va du vert clair (moins significatif = 0.05) au vert foncé (plus significatif = 0.01). Pour l'interprétation des cartes, on remarque un cluster High-high dans le centre de la Gironde qui correspond aux communes ayant des valeurs significatives élevées entourées d'autres communes ayant des valeurs élevées. La significativité tend à être plus importante à mesure que l'on se rapproche du centre du cluster. On observe également à l'est de la Gironde à « cold spots » regroupant les communes ayant des valeurs significatives faibles entourées d'autres communes ayant des valeurs faibles. La significativité des LISA est d'autant plus importante qu'on se rapproche du centre du cluster. Le cluster High-high englobe bien sûr la ville de Bordeaux et ces banlieues, tandis que le cluster Low-low correspond à des villes plutôt rurales.

B - Estimation de modèles spatiaux

Une fois les autocorrélations spatiales globales et locales mesurées, nous pouvons passer à l'estimation des modèles spatiaux. Cependant il faut tout d'abord vérifier si cela est faisable. Pour ce faire il faut dans un premier temps modéliser un modèle des moindres carrés ordinaires (MCO) puis ensuite calculer l'indice de Moran pour chaque matrice de poids sur les résidus. Suite à l'analyse de corrélation nous avons retenu 4 variables, celles de la population, de la ruralité des villes, du revenu médian ainsi que de la densité médicale. Le test de vif montrant s'il y a présence de multicolinéarité dans un modèle, nous affichent des valeurs pour les variables concernées autour de 1. Il n'y a donc pas de multicolinéarité et les 4 variables peuvent être retenues.

Tableau 3 : Résultat des indices de Moran sur les résidus du MCO

Matrice de poids	Queen	Tour	PPV1	PPV3
Indice de Moran	0.4078	0.4088	0.4719	0.4501
P-value	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

L'hypothèse nulle de ce test est la non présence d'autocorrélation spatiale dans les résidus. D'après le tableau des résultats les p-value sont toutes significatives au seuil de 1%. On rejette donc H_0 . Il y a présence d'autocorrélation spatiale pour les 4 matrices de poids. Le modèle MCO n'est donc pas le modèle le plus adapté. Il est nécessaire de tester des modèles d'autocorrélation spatiale.

Nous avons le choix entre deux modèle spatiaux principaux. Il y a tout d'abord le modèle SAR (Spatial Autoregressive). C'est un modèle qui suppose généralement que la variable dépendante dans une commune est influencé par les valeurs de cette même variable dans les communes voisines, ainsi que d'autres facteurs explicatifs. Cela se traduit par l'introduction d'un terme d'autorégression spatiale dans le modèle (ρ), qui capture cette dépendance spatiale. Le paramètre ρ indique dans quelle mesure les observations dans une commune sont corrélées avec les observations dans les communes environnantes. Si un modèle a un indicateur de dépendance spatiale non nulle, il capture donc les effets spatiaux endogènes. Le deuxième modèle est le modèle SEM (Spatial Error Model). Ce modèle tient en compte de la dépendance spatiale entre les observations par l'intermédiaire des erreurs. On suppose donc que les erreurs de chaque observation sont corrélées avec celles des observations dans les communes voisines. En d'autres termes, le modèle suppose que toute structure spatiale non expliquée par les variables explicatives est capturée par la structure d'erreur spatiale. Le paramètre de dépendance spatiale est dans ce cas-ci λ . Pour choisir entre le modèle SAR ou SEM, nous réalisons des tests de Lagrange sur les différentes matrices de poids. Les résultats des tests sont dressés dans le tableau ci-dessous.

Tableau 4 : Résultat des tests de Lagrange sur les différentes matrices de poids

Matrice de poids	Queen	Tour	PPV1	PPV3
Modèles	P-value			
LMerr	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16
LMLag	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16	2.2 ^e -16
RLMerr	0.2561	0.2395	0.1308	0.2694
RLMLag	2.2^e-16	2.2^e-16	2.359^e-06	1.4^e-11

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

Pour choisir le modèle, nous regardons la significativité des tests robustes (RMLag et RLMerr). On remarque que pour toutes les matrices, seul la p-value du RLMLag est significative. Le modèle le plus approprié dans tous les cas de figure est le modèle SAR.

Il y a deux autres modèles spatiaux qui peuvent venir en complément. Ce sont les modèles SDM (Spatial Durbin Model) et SLX (Spatial Lag of X). Ces modèles permettent la prise en compte des effets spatiaux exogènes. Pour ce qui est du modèle SDM, celui-ci combine à la fois les aspects d'un modèle SAR et ceux d'un modèle SEM. Il prend donc en compte à la fois la dépendance spatiale dans la variable dépendante et la dépendance spatiale dans les erreurs. Quant au SLX, ce modèle combine à la fois des termes de régression spatiale et des termes de régression classiques. Pour choisir entre ces deux modèles, on regarde également la significativité des tests robustes. Si au moins un des tests est significatif on choisit le modèle SDM, sinon on opte pour le modèle SLX. Dans notre cas, pour toutes les matrices, il y a au moins un test robuste significatif. On estime donc un modèle SDM. Ce modèle doit ensuite être comparé au modèle SAR pour vérifier s'il existe une différence et une préférence entre les deux. Pour ce faire nous utilisons le test de ratio de vraisemblance. Si la p-value du test est inférieure à 0.05, alors il existe une différence entre les deux modèles. Pour les départager on regarde ensuite les critères d'information (AIC). Celui qui a la plus petite valeur est retenu.

Tableau 5 : Résultat des tests du ratio de vraisemblance pour les différentes matrices

Matrice de poids	Queen		Tour		PPV1		PPV3	
P-value	0.00027		0.00033		0.1117		0.02504	
AIC	SAR : 3517	SDM : 3503.6	SAR : 3517.5	SDM : 3504.6	SAR : 3640.9	SDM : 3641.4	SAR : 3554.1	SDM : 3550.9
Modèle retenue	SDM		SDM		Pas de différence significative		SDM	

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

D'après le tableau ci-dessus, les tests du ratio de vraisemblance étaient tous significatifs sauf celle de la matrice du plus proche voisin. On a donc comparé les critères d'information entre les deux modèles. Pour les trois matrices en question, nous avons retenu le modèle SDM, modèle qui minimisait au mieux l'AIC. En ce qui concerne la matrice du plus proche voisin, il n'existe pas de différence significatif. Un modèle ne peut donc être préféré. On peut les départager à l'aide du LM test for residual autocorrelation. Cependant, pour les deux modèles, la p-value est inférieure à 0.05 signifiant la présence d'autocorrélation dans les résidus. On interprétera le modèle ayant la p-value la moins significative, c'est-à-dire le SDM (même si cette approche ne semble statistiquement pas la plus adaptée).

Tableau 6 : Résultat des modèles SDM des différentes matrices

	SDM Queen	SDM Tour	SDM PPV1	SDM PPV3
Population	4.6392 ^e -06 (0.8637)	4.4270 ^e -06 (0.87)	3.1807 ^e -05 (0.2826)	1.7588 ^e -05 (0.5287)
Rural1	- 2.4610^e+00 (0.0037967)	- 2.4635^e+00 (0.0037874)	-3.9555^e+00 (1.719^e-05)	-3.0852^e+00 (0.0006264)
Revenu_median	-1.7712 ^e -04 (0.3716)	-1.7338 ^e -04 (0.3824)	5.6853^e-04 (0.004283)	7.3250 ^e +05 (0.7200421)
Densité_médicale_BV	1.2825 ^e +00 (0.4579)	1.1816 ^e +00 (0.505)	-1.5552 ^e +00 (0.340194)	-1.0534 ^e +00 (0.5950132)
Lag.Population	4.0201 ^e -05 (0.2962)	4.0504 ^e -05 (0.29215)	1.3888 ^e -05 (0.6559)	3.2669 ^e -05 (0.5993343)
Lag.Rural1	- 2.3258 ^e +00 (0.0924817)	- 2.2561 ^e +00 (0.0994658)	-1.3627 ^e +00 (0.149754)	-1.3802 ^e +00 (0.25953535)
Lag.Revenu_median	1.0181^e-03 (0.0002)	1.0133^e-03 (0.0002)	4.8223^e-04 (0.02)	7.8725^e-04 (0.002)
Lag.Densité_médicale_BV	- 5.2960^e+00 (0.01221)	-5.1011^e+00 (0.0177)	-1.1507 ^e +00 (0.498981)	-1.8131 ^e +00 (0.4324)
Wald Statistic	256.99 (2.22^e-16)	258.5 (2.22^e-16)	173.67 (2.22^e-16)	240.58 (2.22^e-16)
Rhô	0.64973 (2.22^e-16)	0.64898 (2.22^e-16)	0.38329 (2.22^e-16)	0.57995 (2.22^e-16)
LM test	16.339 (5.2974^e-05)	16.711 (4.3533^e-05)	12.811 (0.0003446)	40.649 (1.9974^e-10)

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

Le tableau ci-dessus nous montre les résultats des différents modèles SDM pour chaque matrice de poids. On remarque que ces résultats sont dans l'ensemble robuste puisque nous retrouvons à peu près les mêmes valeurs de coefficients et les mêmes significativités. Le paramètre ρ est

positif et significatif au seuil de 1% pour chaque modèle, ce qui indique que les valeurs d'une unité spatiale impactent les valeurs des entités spatiales voisines. L'intensité de l'autocorrélation spatiale varie de 0.38 à 0.65. Les variables Rural1 et Lag.Revenu_median (effet exogène) sont toutes deux significatives pour chaque modèle. Pour les modèles avec les matrices Queen et Tour on y ajoute la variable Lag.Densité_médicale_BV qui est significative au seuil de 5% (existence d'effet exogène) et pour le modèle PPV1, la variable Revenu_median est significative au seuil de 1%. Cependant, on observe pour tous les modèles SDM, une p-value significative au seuil de 1% pour le LM test qui indique qu'il reste des problèmes d'autocorrélation spatiale dans les résidus. Cela peut s'expliquer par un manque de variables explicatives et/ou un problème de forme fonctionnelle. Pour interpréter les coefficients des variables explicatives, il faut calculer leur impact direct et indirect.

Tableau 7 : Impacts des coefficients du modèle SDM (Queen)

Variables	Direct	Indirect	Total
Population	1.262235 ^e -05	1.153927 ^e -04	1.280151 ^e -04
Rurale1	-3.186057^e+00	-1.047998^e+01	-1.366604^e+01
Revenu_median	-1.030091 ^e -05	2.411305 ^e -03	2.401004 ^e -03
Densité_médicale_BV	4.580647 ^e -01	-1.191650 ^e +01	-1.145844 ^e +01

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

Pour la matrice « Queen », nous avons retenu la variable Rural comme effet endogène et Lag.Revenu_median comme effet exogène au seuil de 1%. Nous pouvons interpréter que le fait qu'une commune soit rurale baisse la part des actifs salariés de -3.19%. Le paramètre Rhô étant significatif au seuil de risque de 1%, cela indique la présence d'interaction entre communes. Le fait que la commune voisine soit considérée comme rurale a un impact sur la part des actifs salariés d'une commune. L'effet indirect est également négatif de l'ordre de -10%. Le LM test nous donne comme résultat qu'il y a une autocorrélation spatiale qui persiste dans les résidus. Il se peut qu'il manque des variables explicatives.

Tableau 8 : Impacts des coefficients du modèle SDM (Tour)

Variables	Direct	Indirect	Total
Population	1.254582 ^e -05	1.154566 ^e -04	1.280024 ^e -04
Rurale1	-3.185035^e+00	-1.026062^e+01	-1.344565^e+01
Revenu_median	-4.790697 ^e -06	2.397544 ^e -03	2.392753 ^e -03
Densité_médicale_BV	3.703628 ^e -01	-1.153639 ^e +01	-1.116603 ^e +01

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

On obtient la même interprétation pour la matrice de poids « Tour ».

Tableau 9 : Impacts des coefficients du modèle SDM (PPV1)

Variables	Direct	Indirect	Total
Population	3.731787 ^e -05	3.677523 ^e -05	7.409311 ^e -05
Rurale1	-4.563883^e+00	-4.059577^e+00	-8.623460^e+00
Revenu_median	7.164888^e-04	9.873088^e-04	1.703798^e-03
Densité_médicale_BV	-1.924357 ^e +00	-2.463278 ^e +00	-4.387635 ^e +00

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

Les variables retenus pour la matrice de poids PPV1 sont la variable Rural et la variable Densité_médicale_BV. Le fait qu'une commune soit rurale baisse de -4.56% la part des actifs salariés tandis qu'une augmentation du revenu médian de 1 euro augmente la part des actifs salariés de 0.0007%. Le paramètre Rhô étant significatif, il existe une dépendance spatiale entre les communes. L'effet indirect de la variable Rurale est également négatif de l'ordre de -4.1% tandis que le Revenu médian a un impact indirect de 0.0010%. Il persiste cependant une autocorrélation spatiale dans les résidus. Il se peut donc qu'il manque des variables explicatives au qu'il y ait un problème de forme fonctionnelle.

Tableau 10 : Impacts des coefficients du modèle SDM (PPV3)

Variables	Direct	Indirect	Total
Population	2.650919 ^e -05	9.385112 ^e -05	1.20360 ^e -04
Rurale1	-3.740237^e+00	-6.890549^e+00	-1.063079^e+01
Revenu_median	2.447233 ^e -04	1.803857 ^e -03	2.048580 ^e -03
Densité_médicale_BV	-1.554333 ^e +00	-5.269784 ^e +00	-6.824117 ^e +00

Source : Dossier Econométrie spatiale, Ida Siting Assine, Souleymane Fomba, Yohann Tesson

On obtient la même interprétation que pour la matrice « Queen » et « Tour », en ce qui concerne le cas de « PPV3 », si ce n'est que le fait qu'une commune soit rurale baisse de - 3.74% la part des actifs salariés avec un effet indirect de l'ordre de -6.89%.

V-Conclusion :

L'objectif de notre étude était d'expliquer la répartition de la part des actifs salariés en Gironde en 2014 et de déterminer s'il existe des dépendances intercommunales. Pour cela nous avons utilisé des méthodes issues de l'économétrie spatiale. Nous avons dans un premier temps, représenté à l'aide d'une carte la répartition de la part des actifs salariés. Nous avons observé une concentration d'actifs salariés dont l'épicentre est Bordeaux et dont cette part diminue au fur et à mesure que l'on s'éloigne de la ville centre. Les cartes des LISA, indicateurs d'autocorrélation spatiale ont également illustré ce phénomène où l'on retrouvait un cluster de forte valeurs avec une forte significativité sur Bordeaux et ses banlieues ainsi que des valeurs qui faiblissaient (de même en significativité) à mesure que l'on avançait dans la couronne périurbaine. Pour chaque matrice de poids que ce soit la Reine, la Tour, le plus proche voisin ou les 3 plus proches voisins, l'indice de Moran et les LISA nous indiquaient la présence de d'autocorrélation locale et globale en ce qui concerne la part des actifs salariés. L'indice de Moran sur les résidus nous a confirmé la nécessité d'utiliser des modèles spatiaux afin d'expliquer cette variable dépendante. Le test de Lagrange et du ratio de vraisemblance nous ont fait retenir pour l'ensemble des matrices de poids le modèle SDM. Les résultats obtenus nous ont indiqué qu'il existait bien une dépendance spatiale intercommunale et que la part des actifs salariés pouvait être expliqué par le fait qu'une commune soit rurale. Cette caractéristique a un effet négatif sur cette proportion de l'ordre de -3%/-4%. Pour ce qui est de la matrice du plus proche voisin, la part des actifs salariés pouvait être expliqué par le revenu médian moyen d'une commune. Celle-ci avait un effet positif sur la part. Cependant ces différents modèles ont des limites puisque de l'autocorrélation spatiale persiste dans les résidus. Il faudrait pour corriger cela, ajouter des variables explicatives ou vérifier s'il y a un problème de forme fonctionnelle. Il aurait été bien également de peaufiner l'étude en comparant la répartition des parts des actifs salariés et les résultats des modèles de 2014 à avec ceux post-covid pour voir s'il y a une tendance à s'excentrer du centre-ville.

VI-Annexes :

Modèle SAR (Queen) :

Residuals:

	Min	1Q	Median	3Q	Max
	-26.38694	-2.75164	0.71344	3.44672	16.45570

Type: lag

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.9874e+01	3.6158e+00	5.4965	3.875e-08
Population	1.4072e-05	2.5355e-05	0.5550	0.57891
Rurale1	-3.6884e+00	7.0219e-01	-5.2528	1.498e-07
Revenu_médian	3.2231e-04	1.4109e-04	2.2844	0.02235
Densité_médicale_BV	-1.3823e+00	8.8378e-01	-1.5641	0.11779

Rho: 0.7081, LR test value: 261.98, p-value: < 2.22e-16
Asymptotic standard error: 0.035393
z-value: 20.007, p-value: < 2.22e-16
Wald statistic: 400.28, p-value: < 2.22e-16

Log likelihood: -1751.502 for lag model
ML residual variance (sigma squared): 34.04, (sigma: 5.8344)
Number of observations: 540
Number of parameters estimated: 7
AIC: 3517, (AIC for lm: 3777)
LM test for residual autocorrelation
test value: 37.597, p-value: 8.6983e-10

Modèle SAR (Tour) :

Residuals:

	Min	1Q	Median	3Q	Max
	-26.39157	-2.68982	0.66009	3.49586	16.43664

Type: lag

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	19.92531610	3.61232382	5.5159	3.469e-08
Population	0.00001419	0.00002535	0.5598	0.57564
Rurale1	-3.68730192	0.70216387	-5.2513	1.510e-07
Revenu_médian	0.00032617	0.00014105	2.3124	0.02076
Densité_médicale_BV	-1.39011399	0.88367564	-1.5731	0.11569

Rho: 0.70651, LR test value: 261.48, p-value: < 2.22e-16
Asymptotic standard error: 0.035304
z-value: 20.012, p-value: < 2.22e-16
Wald statistic: 400.48, p-value: < 2.22e-16

Log likelihood: -1751.753 for lag model
ML residual variance (sigma squared): 34.03, (sigma: 5.8335)
Number of observations: 540
Number of parameters estimated: 7
AIC: 3517.5, (AIC for lm: 3777)
LM test for residual autocorrelation
test value: 36.97, p-value: 1.1998e-09

Modèle SAR (PPV1) :

Residuals:

	Min	1Q	Median	3Q	Max
	-28.56485	-3.24795	0.98148	4.39685	16.06916

Type: lag
Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.5017e+01	3.7107e+00	9.4368	< 2.2e-16
Population	4.1670e-05	2.8982e-05	1.4378	0.15050
Rurale1	-4.7811e+00	7.8923e-01	-6.0579	1.379e-09
Revenu_median	8.4612e-04	1.5851e-04	5.3381	9.394e-08
Densité_médicale_BV	-2.0493e+00	1.0114e+00	-2.0261	0.04275

Rho: 0.39985, LR test value: 138.07, p-value: < 2.22e-16
Asymptotic standard error: 0.027877
z-value: 14.343, p-value: < 2.22e-16
Wald statistic: 205.73, p-value: < 2.22e-16

Log likelihood: -1813.459 for lag model
ML residual variance (sigma squared): 44.554, (sigma: 6.6748)
Number of observations: 540
Number of parameters estimated: 7
AIC: 3640.9, (AIC for lm: 3777)
LM test for residual autocorrelation
test value: 14.935, p-value: 0.00011128

Modèle SAR (PPV3) :

Residuals:

	Min	1Q	Median	3Q	Max
	-26.84337	-2.66137	0.70214	3.77918	14.60828

Type: lag
Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.3933e+01	3.6571e+00	6.5443	5.979e-11
Population	2.4677e-05	2.6272e-05	0.9393	0.3475819
Rurale1	-4.0193e+00	7.2967e-01	-5.5084	3.621e-08
Revenu_median	5.1749e-04	1.4623e-04	3.5389	0.0004018
Densité_médicale_BV	-1.7668e+00	9.1748e-01	-1.9257	0.0541404

Rho: 0.61302, LR test value: 224.93, p-value: < 2.22e-16
Asymptotic standard error: 0.034383
z-value: 17.829, p-value: < 2.22e-16
Wald statistic: 317.88, p-value: < 2.22e-16

Log likelihood: -1770.028 for lag model
ML residual variance (sigma squared): 36.635, (sigma: 6.0527)
Number of observations: 540
Number of parameters estimated: 7
AIC: 3554.1, (AIC for lm: 3777)
LM test for residual autocorrelation
test value: 43.108, p-value: 5.1808e-11

Modèle SDM (Queen) :

Residuals:

Min	1Q	Median	3Q	Max
-24.8034	-2.6304	0.5386	3.5315	15.6119

Type: mixed
Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.5647e+01	4.5753e+00	3.4199	0.0006265
Population	4.6392e-06	2.7024e-05	0.1717	0.8637001
Rurale1	-2.4610e+00	8.5022e-01	-2.8946	0.0037967
Revenu_median	-1.7712e-04	1.9826e-04	-0.8934	0.3716444
Densité_médicale_BV	1.2825e+00	1.7279e+00	0.7422	0.4579451
lag.Population	4.0201e-05	3.8484e-05	1.0446	0.2962092
lag.Rurale1	-2.3258e+00	1.3824e+00	-1.6824	0.0924817
lag.Revenu_median	1.0181e-03	2.7380e-04	3.7184	0.0002005
lag.Densité_médicale_BV	-5.2960e+00	2.1134e+00	-2.5059	0.0122132

Rho: 0.64973, LR test value: 193.39, p-value: < 2.22e-16
Asymptotic standard error: 0.04053
z-value: 16.031, p-value: < 2.22e-16
Wald statistic: 256.99, p-value: < 2.22e-16

Log likelihood: -1740.813 for mixed model
ML residual variance (sigma squared): 33.498, (sigma: 5.7878)
Number of observations: 540
Number of parameters estimated: 11
AIC: 3503.6, (AIC for lm: 3695)
LM test for residual autocorrelation
test value: 16.339, p-value: 5.2974e-05

Test ratio de vraisemblance entre SAR et SDM (Queen) :

Likelihood ratio for spatial linear models

data:
Likelihood ratio = 21.377, df = 4, p-value = 0.0002666
sample estimates:
Log likelihood of sdm Log likelihood of sar
-1740.813 -1751.502

Modèle SDM (Tour) :

Residuals:

Min	1Q	Median	3Q	Max
-24.8013	-2.6470	0.5212	3.4924	15.6007

Type: mixed
Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.5650e+01	4.5587e+00	3.4329	0.0005972
Population	4.4270e-06	2.7029e-05	0.1638	0.8698979
Rurale1	-2.4635e+00	8.5085e-01	-2.8953	0.0037874
Revenu_median	-1.7338e-04	1.9851e-04	-0.8734	0.3824311
Densité_médicale_BV	1.1816e+00	1.7735e+00	0.6662	0.5052535
lag.Population	4.0504e-05	3.8450e-05	1.0534	0.2921534
lag.Rurale1	-2.2561e+00	1.3695e+00	-1.6474	0.0994658
lag.Revenu_median	1.0133e-03	2.7369e-04	3.7023	0.0002137
lag.Densité_médicale_BV	-5.1011e+00	2.1507e+00	-2.3718	0.0177009

Rho: 0.64898, LR test value: 193.86, p-value: < 2.22e-16
Asymptotic standard error: 0.040365
z-value: 16.078, p-value: < 2.22e-16
Wald statistic: 258.5, p-value: < 2.22e-16

Log likelihood: -1741.298 for mixed model
ML residual variance (sigma squared): 33.515, (sigma: 5.7892)
Number of observations: 540
Number of parameters estimated: 11
AIC: 3504.6, (AIC for lm: 3696.5)
LM test for residual autocorrelation
test value: 16.711, p-value: 4.3533e-05

Test du ratio de vraisemblance entre SAR et SDM (Tour) :

Likelihood ratio for spatial linear models

data:
Likelihood ratio = 20.91, df = 4, p-value = 0.00033
sample estimates:
Log likelihood of sdm Log likelihood of sar
-1741.298 -1751.753

Modèle SDM (PPV1) :

Residuals:
Min 1Q Median 3Q Max
-28.47919 -3.22610 0.96152 4.16400 15.75953

Type: mixed
Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.2815e+01	4.0884e+00	8.0264	1.110e-15
Population	3.1807e-05	2.9600e-05	1.0746	0.282576
Rurale1	-3.9555e+00	9.2020e-01	-4.2985	1.719e-05
Revenu_median	5.6853e-04	1.9903e-04	2.8565	0.004283
Densité_médicale_BV	-1.5552e+00	1.6306e+00	-0.9538	0.340194
lag.Population	1.3888e-05	3.1167e-05	0.4456	0.655892
lag.Rurale1	-1.3627e+00	9.4606e-01	-1.4404	0.149754
lag.Revenu_median	4.8223e-04	2.0729e-04	2.3264	0.020000
lag.Densité_médicale_BV	-1.1507e+00	1.7020e+00	-0.6761	0.498981

Rho: 0.38329, LR test value: 122.38, p-value: < 2.22e-16
Asymptotic standard error: 0.029085
z-value: 13.178, p-value: < 2.22e-16
Wald statistic: 173.67, p-value: < 2.22e-16

Log likelihood: -1809.709 for mixed model
ML residual variance (sigma squared): 44.257, (sigma: 6.6526)
Number of observations: 540
Number of parameters estimated: 11
AIC: 3641.4, (AIC for lm: 3761.8)
LM test for residual autocorrelation
test value: 12.811, p-value: 0.0003446

Test du ratio de vraisemblance entre SDM et SAR (PPV1) :

Likelihood ratio for spatial linear models

data:
Likelihood ratio = 7.4992, df = 4, p-value = 0.1117
sample estimates:
Log likelihood of sdm2 Log likelihood of sar2
-1809.709 -1813.459

Modèle SDM (PVV3) :

```
Residuals:
    Min       1Q   Median       3Q      Max
-25.81802  -2.72210   0.62669   3.51495  13.68568

Type: mixed
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.0264e+01  4.4238e+00  4.5806 4.636e-06
Population   1.7588e-05  2.7920e-05  0.6299 0.5287330
Rurale1      -3.0852e+00  9.0213e-01 -3.4199 0.0006264
Revenu_median 7.3250e-05  2.0438e-04  0.3584 0.7200421
Densité_médicale_BV -1.0534e+00  1.9816e+00 -0.5316 0.5950132
lag.Population 3.2969e-05  6.2756e-05  0.5254 0.5993343
lag.Rurale1   -1.3802e+00  1.2237e+00 -1.1279 0.2593535
lag.Revenu_median 7.8725e-04  2.5568e-04  3.0791 0.0020765
lag.Densité_médicale_BV -1.8131e+00  2.3093e+00 -0.7851 0.4323896

Rho: 0.57995, LR test value: 185.15, p-value: < 2.22e-16
Asymptotic standard error: 0.037391
z-value: 15.511, p-value: < 2.22e-16
Wald statistic: 240.58, p-value: < 2.22e-16

Log likelihood: -1764.458 for mixed model
ML residual variance (sigma squared): 36.405, (sigma: 6.0337)
Number of observations: 540
Number of parameters estimated: 11
AIC: 3550.9, (AIC for lm: 3734.1)
LM test for residual autocorrelation
test value: 40.469, p-value: 1.9974e-10
```

Test du ratio de vraisemblance entre SDM et SAR (PPV3) :

```
      Likelihood ratio for spatial linear models

data:
Likelihood ratio = 11.139, df = 4, p-value = 0.02504
sample estimates:
Log likelihood of sdm3 Log likelihood of sar3
      -1764.458          -1770.028
```

Modèle MCO avec les variables retenues :

```
Call:
lm(formula = Part_actif_salaries ~ Population + Rurale + Revenu.
    Densité_médicale_BV, data = shp_gironde)

Residuals:
    Min       1Q   Median       3Q      Max
-28.911  -4.162   1.054   5.098  15.466

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.659e+01  3.874e+00  14.608 < 2e-16 ***
Population    6.560e-05  3.442e-05   1.906  0.05719 .
Rurale1      -7.974e+00  9.246e-01  -8.625 < 2e-16 ***
Revenu_median 1.497e-03  1.857e-04  8.063 4.92e-15 ***
Densité_médicale_BV -3.556e+00  1.202e+00 -2.957 0.00324 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.939 on 535 degrees of freedom
Multiple R-squared:  0.279,    Adjusted R-squared:  0.2736
F-statistic: 51.76 on 4 and 535 DF,  p-value: < 2.2e-16
```

Test du VIF sur les variables du MCO avec les variables retenues :

```
vif(fit1)
      Population      Rurale      Revenu_median      Densité_médicale_BV
      1.127619      1.359599      1.174819      1.391834
```

Modèle MCO avec toutes les variables présentées (présence de multicolinéarité) :

```
Residuals:
    Min       1Q   Median       3Q      Max
-28.558  -3.919   1.167   4.903  15.605

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.482e+01  6.125e+00   8.951  < 2e-16 ***
Population     -1.148e-03  5.340e-03  -0.215  0.829816
Rurale1        -7.608e+00  1.437e+00  -5.295  1.75e-07 ***
Prxm2Moyen     -4.109e-04  4.696e-04  -0.875  0.381996
Revenu_median   1.548e-03  1.969e-04   7.862  2.16e-14 ***
Retraite       -4.040e-03  1.400e-03  -2.885  0.004071 **
Densité_médicale_BV -1.025e+01  3.048e+00  -3.362  0.000829 ***
Score_équipement_santé1 6.569e+00  5.026e+00   1.307  0.191784
Score_équipement_santé2 4.102e+00  4.707e+00   0.872  0.383867
Score_équipement_santé3 1.401e+00  4.557e+00   0.307  0.758662
Score_équipement_santé4 1.375e+00  4.496e+00   0.306  0.759807
Nb_mineurs      2.963e-03  4.071e-03   0.728  0.467129
Nb_etudiants    -5.992e-03  2.862e-03  -2.094  0.036763 *
nb_entreprises  -6.176e-03  4.372e-03  -1.413  0.158290
Score_urbanité  -9.988e-03  3.000e-02  -0.333  0.739296
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.718 on 525 degrees of freedom
Multiple R-squared:  0.3313,    Adjusted R-squared:  0.3135
F-statistic: 18.58 on 14 and 525 DF,  p-value: < 2.2e-16
```

Test du VIF avec toutes les variables présentées (présence de multicolinéarité):

```
> vif(fit1)
      GVIF Df GVIF^(1/(2*Df))
Population 28717.081407 1 169.461150
Rurale     3.473844 1 1.863825
Prxm2Moyen 1.341045 1 1.158035
Revenu_median 1.397813 1 1.182292
Retraite    73.611742 1 8.579729
Densité_médicale_BV 9.462024 1 3.076040
Score_équipement_santé 8.806718 4 1.312507
Nb_mineurs 21377.695055 1 146.211132
Nb_etudiants 150.997152 1 12.288090
nb_entreprises 157.647773 1 12.555786
Score_urbanité 5.911952 1 2.431451
>
```

VII-Bibliographie :

- Site de l'INSEE, <https://www.insee.fr/fr/accueil>
- Site gouvernemental de données, <https://www.data.gouv.fr/fr/>
- Cours d'Econométrie Spatiale, Muriel Travers

VIII-Table des matières

Sommaire.....	2
I-Introduction	3
II-Analyse exploratoire	4
A-Présentation des données	5
a) La part des actifs salariés dans les commune de la gironde.....	5
b) Population par commune de la Gironde	7
c) Rurale	9
d) Le prix moyen au m ² des biens immobiliers vendus	10
e) Le revenu médian	11
f) Le nombre de retraités	13
g) La densité médicale	15
h) Le score équipement santé.....	17
i) Le nombre de mineurs	17
j) Le nombre d'étudiants	19
k) Le nombre d'entreprises	21
l) Le score d'urbanité	23
B-Présentation des données	24
C-Analyse des corrélations	25
III-Méthodologie économétrique	26
IV-Estimation des modèles spatiaux.....	28
A-Mesure de l'autocorrélation spatiale	28
a) Mesure de l'indice de Moran (autocorrélation spatiale globale)	28
b) LISA (Local Indicators of Spatial Association)-autocorrélation spatiale locale	31
B-Estimations de modèles spatiaux	33
V-Conclusion.....	41
VI-Annexes.....	42
VII-Bibliographie	48