République Algérienne Démocratique et Populaire

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENNE

FACULTÉ DES MATHÉMATIQUES

MEMOIRE

Présenté pour l'obtention du diplôme de MAGISTER

En : MATHEMATIQUES

Spécialité : Recherche Opérationnelle : Méthodes Stochastiques

Par : OUALI Fouzia

Thème

Modèles Econométriques Spatiaux Méthodes et Applications

Soutenu publiquement, le 03/07/2008 à 8h:30, devant le jury :

Mr M. BENTARZI	Professeur	Président.
Mr A. BOUKHEBOUZE	Chargé de Cours	Directeur de thèse.
Mme H. GUERBYENNE	Maître de Conférence	Examinateur.
Mr A. AKNOUCHE	Maître de Conférence	Examinateur.
Mr H. BELBACHIR	Maître de Conférence	Examinateur.

REMERCIMENTS

En premier lieu, je remercie Allah le tout puissant de m'avoir prêté main pour la réalisation de ce travail.

A Mr Boukhabouze, j'exprime toute ma gratitude et reconnaissance pour la patience avec la quelle il ma guidé durant les travaux de recherches.

Mes sincères remerciements vont aussi à Mr Bentarzi, professeur à l'U.S.T.H.B, qui m'a fait l'honneur de présider le jury.

Mes remerciements s'adressent également à Mme Guerbyenne, maître de conférences à l'U.S.T.H.B, Mr Aknouche, maître de conférences à l'U.S.T.H.B et Mr Belbachir maître de conférences à l'U.S.T.H.B pour leur gratitude et leur dévouement qui ont permis d'honorer le jury.

Je tiens à présenter ma gratitude en vers toute personne ayant cru en moi, soutenu et apporté tout appui dont j'avais besoin, à citer :
Mme Guerbyenne, Mr Messaci, Mr Aknouche, Mr Hamdi, Melle Merzougui, Melle Djeddou, Melle Kahoul, Mme Benouaret et à toute personne que je n'ai pas cité qui ont fait de même.

DEDICACES

Ả mon mari Lotfi qui ma soutenu et mes filles,

Ả maman qui m'a apprit à être patiente pour surmonter les difficultés,

Ả mon père qui m'a appris que vouloir c'est pouvoir,

Ả la mémoire de Réda et Moncef,

Ả mes sœurs et frères,

Ả mes beaux parents,

Ả mes Amis,

Je dédie ce modeste travail.

Résumé

La plupart des disciplines sont confrontées à l'étude des données spatiales, c'est-à-dire, des observations d'une variable mesurée en différentes localisations réparties dans l'espace. Il est souvent admis que ces données observées en coupe transversale sont indépendantes alors que cette hypothèse est rarement justifiée et devrait être systématiquement testée. L'un des deux effets spatiaux que l'économétrie spatiale prend en compte est la dépendance spatiale (l'auto corrélation spatiale), considérée comme une relation fonctionnelle entre ce qui arrive en un point de l'espace et ce qui arrive ailleurs. Dans le cadre de l'économétrie spatiale telle que les modèles de régression spatiale, La dépendance spatiale se modélise grâce aux matrices de poids. Nous avons défini ces derniers et rapporté quelques méthodes aidants à leurs constructions telle la méthode du pointeur proposée par Boukhebouze et Dreosbeke. La dépendance spatiale est multidirectionnelle, ce qui entraîne la perte des propriétés des estimateurs obtenus par la méthode des moindres carrés ordinaires. Pour cela, nous avons développé quelques méthodes d'estimation, particulièrement celle du maximum de vraisemblance, ainsi que les tests de spécification. Une simulation des données a été faite, afin d'utiliser des tests de spécification pour la validation du type du modèle étudié, ainsi qu'un programme basé sur la méthode du pointeur proposée par Boukhebouze et Dreosbeke a été développé pour construire une matrice de poids.

SOMMAIRE

CHAPITRE IV LES TESTS

CHAPITRE V SIMULATION

Introduction générale

La plupart des disciplines scientifiques sont confrontées à l'étude des données spatiales. Autrement dit des observations d'une variable mesurée en localisations différentes réparties dans l'espace. Ces données spatiales peuvent être analysées en faisant recourt à des processus spatio-temporels. Les données spatiales possèdent souvent des propriétés particulières et doivent être analysées différemment des donnée a-spatiales (Anselin, 1988). Vu que l'espace est multidirectionnelle, nous verrons que certaines méthodes statistiques ne sont pas fiables.

L'introduction de l'espace dans les modèles économiques n'est ni neutre, ni immédiate et les techniques de l'économétrie spatiale visent à prendre en compte la présence de deux effets spatiaux important : l'autocorrélation spatiale qui fait l'objet de notre travail et qui se réfère à l'absence d'indépendance entre observations et l'hétérogénéité spatiale qui est liée à la différentiation des variables et des comportements.

A la fin des années 60 et au début des années 70, une série d'articles et un ouvrage de Cliff et Ord présentent de manière synthétique l'état de l'art concernant la statistique et économétrie spatiale (Le gallo, 2000). Alors qu'à la fin des années 70 jusqu'aux années 80s, le développement de la théorie de l'estimation et des tests a connu un progrès incomparable, dont Paelincke et Klaassen qui ont définis le terme de l'économétrie spatiale par cinq caractéristique (Anselin, 1999). Jusqu'au début des années 90, les techniques permettant de spécifier, d'estimer et de tester la présence de l'autocorrélation spatiale dans les modèles économétrique étaient principalement publiées dans des revues spécialisées et appliqué à des problèmes d'économie régionale, spatiale ou urbaine. Alors que l'application de ces méthodes s'est étendue à d'autres thèmes tels que l'analyse de la demande, l'économie international, l'économie rurale ou encore les phénomènes de croissance et de convergence (Baumont, Ertur, Le gallo, 2000; Ertur et Kock, 2004; Fingleton et Lopez-Bazo, 2006 ; Fischer et stirbock, 2006 ; Lundrg, 2006), l'agriculture et environnement (Blackman, Albers, Avalos-sartorio, Murphy, 2008) ou aussi en criminologie (Anselin, Cohen, Cook, Gorr, Tita, 2000). En effet, ils sont applicables à toutes les études empiriques nécessitant l'utilisation des données spatiale.

L’une des raisons principale à l’attention portée sur les effets spatiaux est la disponibilité croissante des données spatialisées, le fort développement actuel des logiciels des systèmes d’information géographique (SIG) et l’apparition des logiciels tels que SpaceStat (Anselin, 1999), GeoDa (Anselin, 2006), logiciels R (Bivand, 2002) proposant l’estimation des principaux modèles spatiaux. Aussi, la librairie pour Matlab (Le sage, 2005) que nous l’avons utilisées pour estimer les paramètres de notre modèle dans le chapitre 4.

L’objectif de ce travail et de présenter les outils nécessaires à une démarche économétrique visant à prendre en compte l’un des effets spatiaux « **l’autocorrélation spatiale** ». Le travail et organisé de la façon suivante:

Dans le premier chapitre, nous définissons l’économétrie spatiale qui vise à traiter les deux grandes particularités des données spatiales (l’autocorrélation spatiale et hétérogénéité spatiale). Suivi d’un rappel sur des notions statistiques permettant l’apport de l’information sur les différentes valeurs prises par une variable aléatoire, ainsi que leur répartition dans l’espace, pour les deux cas qualitatif et quantitatif.

Certaines notions du chapitre 1 ont donné naissance aux nouvelles définitions présentées dans le chapitre deux, telles que la notion de connectivité dans l’espace et ces principaux outils opérationnelles qui entourent les effets spatiaux en économétrie: les matrice de poids, les opérateurs spatiaux. Nous verrons en suite les problèmes du concept spatial.

Le troisième chapitre est consacré aux méthodes de traitement de l’autocorrélation spatiale dans les modèles de régression linéaire. Nous soulignons les sources et les données concernant l’autocorrélation spatiale (la dépendance spatiale). Puis une définition de cette dernière et ces principaux types de modèles. Ces modèles se caractérisent par une corrélation entre les erreurs et les variables explicatives. Par conséquence, nous verrons que l’inférence statistique basée sur les moindres carrés ordinaires (MCO) n’est plus fiable en présence de la dépendance spatiale. D’autres méthodes d’estimations doivent être utilisées : Méthode du maximum de vraisemblance, les variables instrumentales et l’approche bayesien.

Les différents types de modèles vu dans le chapitre trois seront détecter par des tests de spécification développés dans le chapitre quatre. Ces tests sont basés sur le principe du maximum de vraisemblance. Une règle de décision proposée par des chercheurs peut être appliquée afin de décider le modèle adéquat aux données proposées.

Une simulation est abordée dans le chapitre cinq. En utilisant la méthode du pointeur vue dans le chapitre II. Cette étude prouve que les tests développés dans le chapitre quatre sont efficaces pour la validation du type du modèle étudié, en utilisant la règle vue dans le chapitre précédant. Nous avons aussi remarqué que la méthode du maximum de vraisemblance estime bien notre paramètre spatial.

En conclusion générale, nous apportons une nouvelle technique qui est une amélioration de ses antécédentes étudiées. Nous jugeons que les majorités des contraintes ont été prises en considération, d'où une meilleure fiabilité des résultats obtenus.

CHAPITRE I

DEFINITIONS PRELIMINAIRES

I-1. L'économétrie spatiale

Le terme de l'économétrie spatial a été évoqué par Jean Paelink, au début des années 70, destiné à la science régionale qui traite les problèmes d'estimation et des tests rencontrés dans la mise en œuvre des modèles économétriques multirégionaux (Anselin, 1999 ; Le gallo, 2000).

Les différentes techniques qui traitent les particularités causées par l'espace dans l'analyse statistique des modèles en science régionale sont considérées dans le domaine de l'économétrie spatiale (Anselin, 1988). En effet, l'introduction de l'espace n'est ni neutre, ni immédiate et les techniques de l'économétrie spatiale visent à prendre en compte les deux grandes spécificités des données géographiques : la dépendance spatiale, l'hétérogénéité (Le gallo, 2000 ; Le gallo, 2004). Elle est principalement utilisée lorsqu'on est en présence d'un ensemble fini (régulier ou non) de points ou de zones reliées entre eux par des relations de voisinage (Anselin, 1999 ; Le gallo, 2000 ; Le gallo, 2004). L'aspect spatial des modèles et des données en science régionale empêche l'application directe des méthodes économétriques standards (Anselin, 1988 ; Anselin, 1999 ; Le gallo, 2000 ; Le gallo, 2004).

Trois raisons majeures donnant l'importance à l'utilisation des méthodes d'effets spatiaux dans les modèles économétriques (Le gallo, 2000 ; Le gallo, 2004) :

- La nouvelle attention portait au rôle de l'espace et l'interaction spatiale à la théorie des sciences sociales et qui est aussi la base du développement des nouveaux travaux théoriques interdisciplinaires, pour traiter les interactions de l'environnement humanitaire global.
- Disponibilité croissante d'un grand ensemble de donnée socio-économique où les observations sont repérées géographiquement (des données avec des informations sur le plan régionales et locales).
- Le développement des logiciels de Systèmes d'Information Géographiques (SIG).

Remarque

Tout domaine donnant l'importance aux régions et la répartition géographique doit être considéré dans le rayon de la science régionale.

I-1-1. Définition de l'hétérogénéité spatiale

L'instabilité dans l'espace des relations économiques est appelée hétérogénéité spatiale. Ce phénomène se trouve à plusieurs échelles: les comportements et les phénomènes économiques ne sont pas les même dans le centre d'une ville et dans sa périphérie ; dans une région urbaine et dans une région rurale, dans le Nord d'un continent et le Sud etc... . Deux aspects sont liés à cette notion : l'instabilité et l'hétéroscédasticité.

Le premier, l'instabilité provient de l'absence de stabilité dans l'espace des comportements ou d'autres relations étudiées: les formes fonctionnelles et les paramètres varient selon leurs localisations et ne sont donc pas homogènes. Il est donc nécessaire de mobiliser des modélisations prenant en compte les caractéristiques particulières de chaque localisation de l'échantillon.

Le deuxième aspect de l'hétérogénéité spatiale est l'hétéroscédasticité. Dans les modèles économétriques, elle peut venir de variables manquantes ou de toute autre forme de mauvaise spécification. Par exemple, les unités spatiales elles-mêmes ne sont généralement ni de formes régulières, ni homogène : des régions peuvent avoir des formes et des aires différentes, des niveaux de développement technologique variables, de Populations plus ou moins importantes, etc… .

I-1-2. Définition de L'autocorrélation spatiale (Dépendance spatiale)

En terme général, La dépendance spatiale est considérée comme une relation fonctionnelle entre ce qui arrive en un point de l'espace et ce qui arrive ailleurs. Comme TOBLER l'avait souligné en 1979, en suggérant la première loi de la géographie (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2004) :

"Every thing is related to every things else, but near things are more related than distant things"

Elle est spécialement vérifiée quand les données d'une coupe transversale sont recueillies pour les unités spatiales collectives et contiguës (Anselin, 1988). Elle est nommée aussi par l'autocorrection spatiale qui veut dire: la corrélation d'une variable avec elle-même provenant de la disposition géographique des données. Donc elle peut être définies positive ou négative (Le gallo, 2000 ; Le gallo, 2004).

Une autocorrélation spatiale positive indique le regroupement géographique d’observations de classe voisine (des lieux proches se ressemblent davantage que les lieux éloignés).

Une autocorrélation spatiale négative indique le regroupement géographique d’observations dissemblables (des lieux proches sont plus différents que les lieux éloignés).

Une absence d’autocorrélation spatiale veut dire que la répartition spatiale des observations est aléatoire (aucune relation n’existe entre la proximité des lieux et leur degré de ressemblance).

Ainsi I'autocorrection spatiale mesure le degré auquel un attribut en une localisation est similaire aux attributs en une localisation voisine, sachant qu’on économie les localisations dans l'espace se divisent en trois catégories. Elles peuvent être des points représentant: exemple d'aires urbaines... . Ces points sont mesurés par leur latitude et leur Longitude, elles peuvent être aussi des lignes connectées entre elles ou non ou parfois les données sont fournies pour des aires géographiques comme des régions ou des pays (Le gallo, 2000 ; Le gallo, 2004)..

I-1-2-1. Données et sources de L’autocorrélation spatiale (Dépendance spatiale)

La présence de l’autocorrélation spatiale dans les modèles économétriques était principalement exposée dans les revues spécialisées et appliquées à des problèmes d’économie régionale, spatiale ou urbaine. En pratique, les données concernent ce type de problèmes sont obtenues à partir des observations ordonnées dans l’espace ou dans l’espace et le temps. Ces observations sont caractérisées par leur localisation absolue, en utilisant un système coordonné ou par leur localisation relative, basées sur une distance métrique particulière (Anselin, 1988 ; Le gallo, 2000). Les exemples familiers où on peut trouver ce type de situation sont l’emploi des données sur la population localisées dans l’espace géographique, tel que le recrutement, les contrats de recensements et d’autres activités économiques recueillies pour des unités administratives telles que les régions, les pays.

On souligne deux sources pour la provenance de l’autocorrélation spatiale. Premièrement, elle peut dériver des erreurs de mesures des observations dans les unités spatiales contiguës ou manque de données. Dans ce cas, elle est considérée comme un outil de diagnostic d'une mauvaise spécification (chapitre 3). Le deuxième facteur qui cause une autocorrélation spatiale est plus fondamental, les théories des interactions spatiales, les processus diffuseurs et hiérarchies spatiales produisent une structure de dépendance entre phénomène au différent lieu de localisations dans l'espace (chapitre 3).

I -1-2-2 Mesure d'autocorrélation spatiale

Lorsque les valeurs prises par une variable aléatoire, discrète ou continue, ne sont pas disposées au hasard sur une carte, mais sont proches pour deux régions voisines, nous dirons qu'il y a autocorrélation spatiale (Le gallo, Jayet, 2001). Pour mesurer cette autocorrélation spatiale *globale*, on dispose de plusieurs statistiques permettant d'apporter une information non seulement sur les différentes valeurs prises par la variable aléatoire mais également sur la façon dont ces valeurs sont réparties dans l'espace. On distingue le cas des variables qualitatives et celui des variables quantitatives, pour lesquelles deux indicateurs existent : le coefficient de Moran et le coefficient de Geary (Jayet, 2001).

I -1-2-2-1 Variables qualitatives

Dans la littérature, le cas où la variable étudiée prend deux valeurs, on leurs associes deux couleurs distinctes: blanc et noir. Dans ces conditions, pour savoir s'il y a autocorrélation spatiale, nous considérons les frontières qui relient deux observations voisines. Ces frontières peuvent être de trois types : noir-noir (NN) quand les deux régions qu'elles séparent sont noires, blanc-blanc (BB) quand les deux régions sont blanches, noir-blanc (NB) si les deux régions sont de couleurs opposées. Ainsi, intuitivement, pour une autocorrélation spatiale positive, les frontières NN et BB prédomineront de telle sorte que des concentrations de régions de même couleur apparaissent, alors que pour une autocorrélation négative, les frontières NB seront plus nombreuses et il y aura alternance quasi-régulière entre les régions noires et les régions blanches (Jayet, 2001).

Formellement, considérons la variable aléatoire x prenant la valeur l si la $i^{ème}$ observation est noire et 0 sinon. Soit w une matrice de poids (elle sera développée dans le chapitre 2) quelconque (symétrique ou non). Dans ces conditions, NN, NB et BB sont respectivement égaux à :

Soit δ_{ij} la matrice de connexion binaire entre éléments de G, dont le terme général est donné par :

$$\delta_{ij} = \begin{cases} 1 & \text{si le site i est connecté avec le site j} \\ 0 & \text{si non} \end{cases}$$

Si on définit la variable aléatoire x_i par :

$$x_i = \begin{cases} 1 \text{ si le site i est de couleur noire.} \\ 0 \text{ si le site i est de couleur blanche} \end{cases}$$

Les statistiques de comptage des connecteurs sont définies par :

$$NN = (1/2)\sum_i \sum_j \delta_{ij} x_i x_j$$

$$NB = (1/2)\sum_i \sum_j \delta_{ij} \left(x_i - x_j \right)^2$$

$$BB = (1/2)\sum_i \sum_j \delta_{ij} \left(1 - x_i\right)\left(1 - x_j\right) = (1/2)\sum_i \sum_j \delta_{ij} - NN - BB$$

Puisque la statistique *BB* est une fonction linéaire de *NN* et *NB*, elle ne fournit pas d'information supplémentaire et ne sera pas donc considérée par la suite. Les lois de probabilité suivies par *NN* et *NB* sous l'hypothèse nulle d'absence d'autocorrélation spatiale sont usuellement données pour deux cas différents :

- **Hypothèse l** : on suppose que chaque région est indépendamment noire ou blanche selon des probabilités p et q données (avec p + q = l). La coloration de la carte peut alors être assimilée à un tirage avec remise. La variable aléatoire définie par le nombre de fois où une région noire (ou blanche) est tirée.

- **Hypothèse 2** : on utilise un tirage aléatoire sans remise dans une urne contenant N_i régions noires et N_i régions blanches correspondant aux valeurs observées. La variable aléatoire définie par le nombre de fois où une région noire (ou blanche) est tirée.

Considérons la statistique centrée réduite suivante :

$$z = \frac{NN - E(NN)}{V^{1/2}(NN)}$$

Où E (NN) et V (NN) Les moments d'ordre l et 2 de variable aléatoire NN

Cliff et Ord (1981) ont montré que si le nombre d'observations est suffisamment grand, la distribution de cette statistique *z* est correctement approchée par une loi normale d'espérance nulle et de variance unitaire (Jayet, 2001). La même propriété vaut pour la statistique NB. Il est alors possible de tester les hypothèses l ou 2. En outre, il est utile de faire les tests pour plusieurs ordres de contiguïté pour avoir une idée de la structure de l'espace.

I -1-2-2-2 Variables quantitatives

I -1-2-2-2-1 La statistique de Moran

La statistique I de Moran (1948) est la statistique la plus connue d'autocorrélation spatiale. Cette statistique, qui apparaît comme une extension de la statistique *NN* aux variables quantitatives, s'écrit de la façon suivante (Baumont, Ertur, Le gallo, 2000 ; Jayet, 2001) :

$$I = \frac{N}{S_0} \cdot \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$$

Où x_i est l'observation dans la région i, $\overline{x}$ est la moyenne des observations des régions, N est le nombre de régions. w_{ij} est l'élément de la matrice de poids, S_0 est un facteur d'échelle égal à la somme de tous les éléments de w.

L'espérance mathématique de I est égale à $-\frac{1}{N-1}$. Par conséquent, il y a autocorrélation spatiale positive lorsque I est supérieur à cette espérance et il y a autocorrélation spatiale négative lorsque I est inférieur.

I -1-2-2-2-2 La statistique de Geary

Un deuxième indicateur est le coefficient de Geary (1954) qui dérive plutôt de la statistique NB. Au lieu d'être construite à partir des produits croisés entre les déviations à la moyenne, la statistique de Geary utilise les carrés des différences entre les valeurs prises par la variable sur les régions voisines.

Elle s'écrit de la façon suivante (Jayet, 2001) :

$$c = \frac{N-1}{2S_0} \cdot \frac{\sum_i \sum_j w_{ij} \left(x_i - x_j \right)^2}{\sum_i \left(x_i - \overline{x} \right)^2}$$

Le terme du numérateur est au facteur $1/2$ près, la variance pondérée des différences x_i - x_j entre observations contiguës. Cette variance est faible pour une autocorrélation positive et très élevée en présence d'une autocorrélation négative. En l'absence d'autocorrélation spatiale, le numérateur est approximativement égal à l'estimateur de la variance de l'échantillon, c'est-à- dire le dénominateur. Le coefficient de Geary est alors proche de I.

Comme pour les statistiques NN et NB, les statistiques de Moran et Geary centrées et réduites sont asymptotiquement distribuées selon des lois normales (0, l).

En conclusion, pour décrire des phénomènes quantitatifs ou qualitatifs ayant une dimension spatiale, les interactions spatiales sont fréquentes. Ainsi, il est nécessaire de considérer non seulement les dimensions et les structures des observations mais aussi leurs positions relatives. Ceci s’effectue à l’aide des matrices de poids.

CHAPITRE II

LES MATRICES DE POIDS ET LES OPERATEURS SPATIAUX

II-1. Introduction

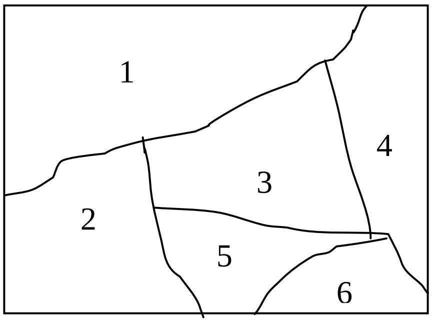
La notion de l'économétrie spatiale implique le besoin de déterminer la position relative des observations les unes par rapport aux autres, en plus de leurs dimensions et de leurs structures. Ces relations dépendent essentiellement de la répartition de ces individus dans un espace à m dimensions. Cette représentation met en lumière le processus d'interdépendance où le choix d'une métrique adéquate joue un rôle fondamental. L'analyse statistique en différents points d'un ensemble dénombrable G et à un instant donné, offre un instrument aux statisticiens et aux praticiens pour détecter les déplacements aléatoires d'un phénomène et étudier la structure spatio-temporelle de G à l'aide des interactions mutuelles entre ses individus (Boukhhbouze, 1999,2000). Les mesures de ces interactions spatiales doivent refléter les propriétés physiques ou autres de l'ensemble G relatives au phénomène en question. Ainsi dans le cas où G est un ensemble de sites géographiques, on remarque que la mesure de l'interaction spatiale entre deux sites quelconques peut dépendre de plusieurs facteurs. En particulier, des caractéristiques spatiales des lieux (position spatiale de chaque site, distance entre leurs centres géographiques ou démographiques, longueurs de la frontière commune barrière naturelles comme les montagnes, les rivières etc…), les caractères socio-économiques ou politiques des lieux (activités économiques, situation politique, vie sociale, accessibilité aux différentes facilités existante etc…) (Boukhhbouze, 1999).

A partir de cette détermination nous construisons une matrice dite de poids, afin de spécifier une topologie d'un modèle économétrique spatiale. Ces matrices sont définies à priori par le modélisateur en tenant compte des relations et interactions entre les unités spatiales. Elles ne contiennent pas d'éléments inconnus. Ces matrices sont aussi interprétées comme des matrices de transition dans les chaines de Markov (Bavaud, 1993 ; Bavaud, 2003). II y a deux types de matrices de poids:

- Matrices de contiguïté.
- Matrices de poids générales.

II-2. Les matrices de contiguïté

Moran et Geary (1954) définissent la contiguïté entre deux régions par l'existence d'une frontière commune (Anselin, 1988 ; Le gallo, 2004). Dans le cas d'un dispositif irrégulier, la contiguïté des unités spatiales est définie directement. Dans le cas contraire, c'est-à-dire les unités spatiales sont référées par une grille régulière ou par un ensemble de points rangés irrégulièrement, la détermination de contiguïté n'est pas unique, comme on peut voir dans la figure 1et la figure 2 respectivement, le cas irrégulier et une grille régulière.



La figure 1 (Cas irrégulier)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

La figure 2 (la grille régulière)

La matrice de contiguïté est une matrice carrée, symétrique, ses coefficients sont définis comme suit :

$$\delta_{ij} = \begin{cases} 1 & \text{si les régions i et j sont contiguë d'ordre 1.} \\ 0 & \text{si non.} \end{cases}$$

L'élément δ_{ij} indique l'existence ou non d'une frontière commune entre les régions *i* et *j*. Par convention, une région n'est pas contiguë avec elle-même, c'est-à-dire $\delta_{ii} = 0, \forall i$. On note aussi que deux régions non contiguës peuvent être relativement proche l'une de l'autre, dans un sens ou dans un autre. Cela nécessite une généralisation de la matrice de contiguïté par des matrices à plusieurs niveaux dîtes matrices de séparation spatiale. Deux régions sont contiguës d'ordre s (s étant un entier positif) quand il faut traverser au moins s frontières pour passer de l'une à l'autre. Alors, si s est le nombre minimal de frontières à traverser pour aller de i à j, on dit que les régions i et j sont contiguës à l'ordre s (Boukhebouze, 1999). Cela n'est pas égal à la matrice de contiguïté d'ordre 1, élevée à la puissance s. Dans la figure 1, la région 1 est contiguë d'ordre 1 aux régions 2, 3 et 4, par contre elle est contiguë d'ordre 2 pour aux régions 5 et 6.

Les matrices des séparations spatiales généralisent les matrices de distances de contiguïtés, $w^{(s)}$ étant la matrice de séparation spatiale d'ordre s. Le terme général $w_{ij}^{(s)}$ représente la quantification de l'influence relative du point j sur le point i à l'ordre s. Ces coefficients sont non négatifs, normalises et généralement asymétriques.

II-2-1 Etude des matrices de séparation spatiale

Dans cette étude nous allons élargir la perspective et les définitions des matrices de séparation spatiale, ce qui nous permet de les utiliser dans plusieurs domaines d'applications. Toutes les définitions relatives à ces matrices avec les propriétés qui en découlent ont été données (Boukhebouze, 1999; Boukhebouze et Droesbeke, 2000).

II-2-1-1. Le voisinage d'ordre s

Soit G un ensemble fini ou dénombrable d'éléments. A chaque individu i de G, on peut associer des « voisins s » plus au moins éloignés et qui peuvent être regroupés dans des voisinages. On associe à i un nombre (entier non négatif) plus il est élevé plus le concept de voisinage est faible. Ainsi, $v_i^{(1)}$ regroupe les individus de G ayant le plus d'affinité (ressemblance, d'attraction,….) avec i. Nous supposons qu'il est possible de définir une suite de sous ensemble $v_i^{(s)}$ (s=0,1,2 ,…), regroupant les voisins dont l'affinité avec i est la même.

Si nous supposons une suite réelle $a_1, a_2, \cdots$, telle que $a_0 = 0 < a_1 < a_2 < \dots\dots$, alors nous désignons dès lors des voisinages associés à *i* de G par :

$$V_i^{(s)} = \begin{cases} \{i\} \; si \; s = 0 (pour \; tout \; i = 1, 2, ...) \\ \{ j \in G \; telque \; a_{s-1} < d_{ij} \leq a_s \} (i \neq j), pour \; tout \; s = 1, 2, ..., et \; i = 1, 2, ... \\ \phi \; si \; non \end{cases}$$

Où d_{ij} étant une distance numérique entre *i* et *j* de G**.**

On dit que l'individu *j* exerce une influence spatiale (attraction) d'ordre *s* sur l'individu *i*, si et seulement si l'individu *j* est voisin *i*.

L'ensemble $V_i^s \; (i = 1, 2, ...)$ est le voisinage spatial d'ordre s de *i* de G.

II-2-1-2. Le taux d'importance d'un individu

Le taux d'importance q_j d'un individu *j* de G, relative à un problème donné, est défini par :

$$q_j = \frac{\lambda_j}{\lambda}$$

Où

$$\lambda_j = \sum_{k=1}^{m} x_j(k) c_k \quad \text{et} \quad \lambda = \sum_{j=1}^{n} \lambda_j$$

m : étant le nombre de critères.

c_k : le coefficient d'importance du $k^{ème}$ critère (k=1,2,……,*m*).

$x_j(k)$: est le nombre de fois que l'individu *j* a satisfait le critère k.

II-2-1-3. L'influence spatiale

L'influence spatiale d'ordre s, exercée par l'individu *j* sur l'individu *i*, notée $u_{ij}^{(s)}$, mesure la force d'attraction d'ordre s, exercée par l'élément *j* sur l'élément *i*. Elle est définie par

$$u_{ij}^{(s)} = \begin{cases} q_j \times \left[1 + D_{ij}\right]^{-(s+1)} \ \ si_{(i,j)\in G^2}, \ j \in v_i^{(s)}, (pour \ tout \ s = 1, 2, ..., et \ i = 1, 2, ...) \\ 0 \ \ si \ non \end{cases}$$

Où

$$D_{ij} = d_{ij} \times \left[\left\{ \min_{i\in G} \min_{j\in G} \left(d_{ij} \right) \right\} \right]^{-1} \quad \left(i \neq j \right)$$

Avec $\min_{i\in G} \min_{j\in G} \left(d_{ij} \right) \neq 0$

II-2-1-4. Le taux d'influence spatiale

Le taux d'influence spatiale de l'individu *j* sur l'individu *i* d'ordre s, noté $W_{ij}^{(s)}$ est le rapport entre la force d'attraction d'ordre s, exercée par l'individu *j* sur l'individu *i* et la force totale d'ordre s exercée par tous les individus de G sur l'individu *i*; il est donné par :

$$W_{ij}^{(s)} = \begin{cases} 1 \ \ si \ i = j \ \ s = 0 \\ u_{ij}^{(s)} \times \left[u_{ij}^{(s)} \right]^{-1} \ \ si \ i \neq j \ et \ \ j \in v_i^{(s)}; (pour \ tout \ s = 1, 2, ... \ et \ i = 1, 2, ...) \\ 0 \ \ si \ non \end{cases}$$

Où $u_{ij}^{(s)}$ est la force d'attraction d'ordre *s*, exercée par l'individu *j* sur l'individu *i* et $u_{i\bullet}^{(s)}$ la force totale d'ordre s exercée par tous les individus de G sur l'individu *i*, donnée par :

$$u_{i\bullet}^{(s)} = \sum_{j=1}^{n} u_{ij}^{(s)} \quad \text{avec } \ n = \text{card(G)}.$$

II-2-1-5. L'ordre maximum de séparation spatiale

L'ordre maximum de séparation spatiale d'un individu *i* de l'ensemble G, relativement à un problème donné, est un entier η_i tel que :

$$V_i^{(\eta_i)} \neq \phi \quad et \quad V_i^{(s)} = \phi \quad pour \ tout \ s = \eta_i + 1, \eta_i + 2, ...$$

$\eta_G = \min_{i \in G}(\eta_i)$ et $\eta_G * = \max_{i \in G}(\eta_i)$ sont respectivement l'ordre minimum et l'ordre maximum de séparation spatiale de G.

Remarque

* La matrice des taux d'influence spatiale d'ordre *s* est également appelée la matrice de séparation spatiale d'ordre *s* (matrice de poids, matrice de pondération spatiale).

* $\sum_{j=1}^{n} w_{ij}^{(s)} = 1$ et $w_{ij}^{(s)} = 0 \ \forall i \in G \ s \rangle 0$; si s = 0, on prend $w_{ij}^{(0)} = 1$; ce qui signifie que $W^{(0)} = \mathrm{I}_n$ (matrice identité).

* dans certains cas, les matrices de séparation spatiale de différents ordres sont données par :

$$W_{ij}^{(s)} = \begin{cases} \left| v_i^{(s)} \right|^{-1} & j \in v_i^{(s)} \\ 0 \quad si \ non \end{cases}$$

II-2-1-6. Propriétés

Les structures de voisinage peuvent aider l'utilisateur, dans plusieurs domaines d'application, à choisir une configuration adéquate, afin de générer systématiquement les matrices de pondération spatiale.

Nous désignons une structure de voisinage associée à *i* de G par :

$$V_i = \left\{ V_i^1, V_i^2, ..., V_i^{(s)}, ... \right\} \quad ………. (1)$$

Où $V_i^{(s)} \subset G$ (pour tout s = 0,1,2,…… et *i* = 1,2,……).

Nous désignons aussi une structure de voisinage associée à G par :

$$V = \left\{ V_1, V_2, ..., V_i, ... \right\} \quad ………….. (2)$$

La structure (2) est dite strictement admissible sur G si et seulement si les conditions suivantes sont satisfaites :

(*i*) $V_I^{(0)} = \{i\}$ (pour tout *i*=1,2,…..)

(*ii*) $\left|V_i^{(s)}\right| < x$ (pour tout *i*=1,2,….. et *s* =1,2,……)

(*iii*) Le système $\{V_i^{(1)}, V_i^{(2)}, \ldots\ldots, V_i^{(\eta_i)}, \ldots.\}$ forme une partition de l'ensemble G (pour tout *i* = 1,2,…..)

La structure (1) est dite la *plus fine structure* si et seulement si les sous-ensembles $V_i^{(s)}$, contiennent un nombre d'éléments de G, réduit à son minimum (*i*=1,2,….. et *s* =1,2,……). Ce qui veut dire que, η_i atteint son maximum (pour tout *i*=1,2,…..), tout en gardant les conditions (*i*), (*ii*) et (*iii*).

La structure (2) est dit *optimale sur G* si et seulement si elle est strictement admissible et la structure (1) est la plus fine structure.

La structure (2) est dite *optimale conditionnelle sur G* si et seulement si elle est optimale et les matrice de séparation spatiale $W^{(s)}$ sont irréductibles pour tout *s*.

La littérature contient peu d'informations pour spécifier les poids des matrices de séparation spatiale sur les applications des modèles spatiaux. La plupart des articles rencontrés laissent le soin à l'utilisateur de choisir des poids qu'il estime convenables à son cas. Pour cela, nous considérons une nouvelle approche proposée par Boukhebouze et Droesbeke qui déterminent systématiquement les poids des matrices de séparation spatiale en essayant de développer des structures de voisinages et de classes indépendantes pour élargir les notions liées à ces matrices et de réduire davantage leurs dimensions.

II-2-2. Méthode du pointeur

Cette méthode consiste à construire des structures de voisinages optimales et optimales conditionnelles, au sens défini par Boukhebouze et Dreosbeke. Elle se base sur des transformations bornées dans un espace de dimension fini et ensuite sur des différentes distances entre individus de l'ensemble G dans cet espace. Cette méthode a été proposée dans deux cas. Le premier est la méthode des positions spatiales, elle consiste à construire une structure de voisinage optimale. Le deuxième est la méthode des rapports spatiaux, elle consiste à construire une structure de voisinage optimale conditionnelle. Nous verrons dans ce travail seulement la première méthode; qui serra utilisée dans l’application (chapitre 5).

II-2-2-1. Méthode des positions spatiales

Cette méthode consiste à construire une structure de voisinage optimale. Le mécanisme de cette méthode est de calculer un certain nombre de paramètres de séparation spatiale $\alpha_1, \alpha_2, \ldots\ldots, \alpha_m$, avec lesquels on construit les différents voisinages autour de chaque point, en minimisant le cardinal de chaque voisinage, sous la condition que pour chaque élément, la structure de voisinage correspondante soit fortement admissible.

Supposons qu’il y ait $m_i^{(1)}$ différentes distances de chaque individu *i* de l’ensemble G dont $|G| = n$ et $\max\limits_{1\le i\le n}\left[\max\limits_{1\le j\le n}\left(d_{ij}\right)\right] < x$. Donc, on aura *n* suites ordonnées (pour la première fois (le (1) représente la 1ére fois)) de la manière suivante :

$$\begin{cases} d_1^{(1)}\left(1\right) < d_2^{(1)}(1) < \ldots\ldots. < d_{m_1^{(\circ)}}^{(1)}(1) \\ d_1^{(1)}\left(2\right) < d_2^{(1)}(2) < \ldots\ldots. < d_{m_é^{()1}}^{(1)}(2) \\ \ldots \\ \ldots \\ d_1^{(1)}\left(n\right) < d_2^{(1)}(n) < \ldots\ldots. < d_{m_n^{(1)}}^{(1)}(n) \end{cases}$$

On prend

$$\alpha_1 = \max_{1\le i\le n_1}\left[d_1^{(1)}(i)\right]$$

Pour chaque suite ordonnée liée à l’individu *i*, $d_1^{(1)}\left(i\right) < d_2^{(1)}(i) < \ldots\ldots. < d_{m_{1i}}^{(1)}(i)$, avec *i*=1,…,*n*, on déplace l’origine à droite avec une translation de α_1 et on considère uniquement les éléments de chaque suite à droite de la nouvelle origine. Ce qui signifie qu’on considère iniquement les éléments positifs de chaque suite par rapport à la nouvelle origine.

Supposons qu'après cette translation de α_1, on aura n_1 suites ordonnées (pour la 2ème fois (2)) de la manière suivante:

$$\begin{cases} d_1^{(2)}(1) < d_2^{(2)}(1) < .... < d_{m_1^{(2)}}^{(2)}(1) \\ d_1^{(2)}(2) < d_2^{(2)}(2) < .... < d_{m_2^{(2)}}^{(2)}(2) \\ ... \\ ... \\ d_1^{(2)}(n_1) < d_2^{(2)}(n_1) < .... < d_{m_{n_1}^{(2)}}^{(2)}(n_1) \end{cases}$$

$m_i^{(2)}$: est le nombre de différentes distances restantes de la suite liée à l'individu i de l'ensemble G après la translation α_1, tel que : $m_i^{(2)} < m_i^{(1)}$; n_1 le nombre de suites restantes (nombre d'individus) après la translation α_1.

On prend

$$\alpha_2 = \max_{1 \le i \le n_1} \left[d_1^{(2)}(i) \right]$$

Pour chaque suite ordonnée liée à l'individu *i*, $d_1^{(2)}(i) < d_2^{(2)}(i) < .... < d_{m_i^{(2)}}^{(2)}(i)$, avec $1 \le i \le n_1 \le n$; on déplace l'origine à droite avec une translation de α_2 et on considère uniquement les éléments de chaque suite à droite de la nouvelle origine. Ce qui signifie qu'on considère uniquement les éléments positifs de chaque suite par rapport à la nouvelle origine.

Supposons qu'après cette translation de α_{k-1}, on aura n_{k-1} suite ordonnées (pour la $k^{ème}$ fois (k)) de la manière suivante:

$$\begin{cases} d_1^{(k)}(1) < d_2^{(k)}(1) < .... < d_{m_1^{(k)}{}_1}^{(k)}(1) \\ d_1^{(k)}(2) < d_2^{(k)}(2) < .... < d_{m_2^{(k)}}^{(k)}(2) \\ ... \\ ... \\ d_1^{(k)}(n_{k-1}) < d_2^{(k)}(n_{k-1}) < .... < d_{m_{n_{k-1}}^{(k)}}^{(k)}(n_{k-1}) \end{cases}$$

$m_i^{(k)}$: est le nombre des différentes distances restantes de la suite liée à l'individu *i* de l'ensemble G après la translation α_{k-1}, tel que : $m_i^{(k)} < m_i^{(k-1)} < .... < m_i^{(1)}$; n_{k-1}: est le nombre de suites restantes après la translation α_{k-1}.

On prend

$$\alpha_k = \max_{1 \le i \le n_{k-1}} \left[d_1^{(k)}(i) \right]$$

Pour chaque suite ordonnée liée à l'individu *i*, $d_1^{(k)}(i) < d_2^{(k)}(i) < .... < d_{m_i^{(k)}}^{(k)}(i)$, avec $1 \le i \le n_{k-1} \le n_{k-2} \le .... \le n$; on déplace l'origine à droite avec une translation de α_k et on considère uniquement les éléments de chaque suite à droite de la nouvelle origine.

On répète ce procédé jusqu'à ce que nous trouvions un entier *m* tel que:

$$\sum_{i=1}^{m} \alpha_i = \max_{1 \le i \le n} \left[d_{m_i^{(1)}}^{(1)}(i) \right] = \max_{1 \le i \le n} \left[\max_{1 \le j \le n}(dij) \right]$$

Ce qui est possible car le nombre des différentes distances restantes de la suite ordonnée liée à l'individu *i* de G, après chaque translation α_{k-1} diminue d'après la relation

$$m_i^{(k)} \le m_i^{(k-1)} \le .... \le m_i^1 \quad \text{(pour tout } i\text{=1,2,…..)} .$$

Nous déterminons des structures de voisinages $\left\{ V_i^{(0)}, V_i^{(1)}, ..., V_i^{(\eta_i)} \right\}$, pour tout *i* de G. Nous distinguons trois cas possibles.

(1) Soit $\forall i, j \in G, i \ne j, d_{ij} \ne 0$. Dans ce cas on prend

$$a_0 = 0, \qquad a_h = \sum_{i=1}^{h} \alpha_i, 1 \le h \le m$$

et

$$V_i^{(s)} = \begin{cases} \{i\}\ si\ s = 0 (pour\ tout\ i = 1, 2, ...) \\ \left\{ j \in G\ telque\ a_{s-1} < d_{ij} \le a_s \right\}\ si\ (i \ne j), 1 \le s \le \eta_i \le m. \\ \phi\ si\ non \end{cases}$$

(2) Soit $\forall i \in G, \exists j \in G, i \neq j, d_{ij} = 0$. Dans ce on prend,

$$a_0 = 0, \qquad a_{h+1} = \sum_{i=1}^{h} \alpha_i, 1 \leq h \leq m$$

et

$$V_i^{(s)} = \begin{cases} \{i\} \; si \; s = 0 (pour\ tout\ i = 1, 2, ...) \\ \{j \in G \; telque \; a_{s-1} < d_{ij} \leq a_s\} \; si \; (i \neq j), 2 \leq s \leq \eta_i \leq m \\ \{j \in G / d_{ij} = 0\} \; si \; i \neq j, \; s = 1 \\ \phi \; si \; non \end{cases}$$

(3) Supposons que

$\exists B \subset G, \forall j \in B, i \neq j, d_{ij} \neq 0 \; et \; \forall i \in \overline{B}, \forall j \in \overline{B}, i \neq j, d_{ij} = 0 \;\; B \bigcup \overline{B} = G$ avec $B \bigcup \overline{B} = G$

Dans ce cas, on prend $\delta_1 = \max_{i \in B} \left[\min_{j \in B} d_{ij} \right]$

(a) Si $\delta_1 = \alpha_1$,même cas que (1)

(b) Si $\delta_1 < \alpha_1$,dans ce cas on prend

$$a_1 = \delta_1 \quad , \qquad a_{h+1} = \sum_{i=1}^{h} \alpha_i + \delta_1, \; 1 \leq h \leq m$$

et $$V_i^{(s)} = \begin{cases} \{i\} \; si \; s = 0 (pour\ tout\ i = 1, 2, ...) \\ \{j \in G \; telque \; a_{s-1} < d_{ij} \leq a_s\} \; si \; (i \neq j), 2 \leq s \leq \eta_i \leq m \\ \{j \in G / d_{ij} \leq a_1\} \; si \; i \neq j, \; s = 1 \\ \phi \; si \; non \end{cases}$$

Si on donne le même poids à chaque individu du même voisinage $V_i^{(s)}$, la matrice de séparation spatiale d'ordre s $W^{(s)}$ est donné par son terme général :

$$W_{ij}^{(s)} = \begin{cases} \left| v_i^{(s)} \right|^{-1} & j \in v_i^{(s)} \\ 0 \;\; si \; non \end{cases}$$

II-3. Les matrices de poids générales

Le simple concept de contiguïté binaire a été agrandi par Cliff et Ord (1973-1981) en définissant une mesure générale pour l'intensité de l'interaction entre les régions. Leurs suggestions consistent à combiner entre les mesures de distances et de longueur relative de la frontière commune entre les régions. Cette matrice de poids est asymétrique et définie comme suit (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2004):

$$w_{ij} = \left(d_{ij}\right)^{-a} \left(B_{ij}\right)^{b}$$

d_{ij} : est la distance entre les régions *i* et *j*;.

B_{ij} : est la proportion de la frontière intérieure de l'unité *i* qui est en contacte avec l'unité *j*.

a et *b* sont des paramètres détermines à priori.

Un autre cas où l'interaction entre deux régions sera seulement exprimée par la distance

$$w_{ij} = \frac{1}{d_{ij}}$$

où d_{ij} est la distance entre la région *i* la région *j*.

$f\left(d_{ij}\right)$ Cette fonction est décroissante ; traduisant ainsi la baisse des interactions spatiales quand l'éloignement augmente, elle a été utilisé dans (Lichstein, Simons, Shriner et Franzreb ; 2002). Plusieurs formes fonctionnelles sont disponibles, certains auteurs utilisent des matrices de poids dans lesquelles les éléments ont perdus toute référence à la localisation géographique (Le gallo, 2000 ; Le gallo, 2004).

Les matrices de poids sont souvent standardisées de telle sorte que la somme des éléments de chaque ligne est égale à 1. Les éléments d'une matrice standardisée sont égaux à:

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j}^{n} w_{ij}}$$

La détermination d'une spécification propre aux éléments de la matrice w_{ij} est l'une des plus difficiles étapes en économétrie spatiale.

II-4. Les opérateurs spatiaux décalés

L'objectif de l'utilisation de la matrice de poids, dans la spécification des modèles économétriques spatiaux, est de relier l'observation de la variable pour un point de l'espace avec les observations de cette variable pour d'autres lieux du système.

Pour n régions, wY est appelé variable spatiale décalée associée à *Y*, son $i^{ème}$ élément est (Anselin, 1988 ; Anselin, 1999):

$$[wy]_i = \sum_{j=1}^{n} w_{ij} y_j$$

Cet $i^{ème}$ élément est la moyenne pondérée des observations des régions voisines à la région *i*, si la matrice de poids est standardisée.

Le décalage spatial n'a pas le même sens que le décalage connu en série chronologique, cela dû au fait que l'espace est multidirectionnelle. Elle permet la comparaison entre la valeur de y associée à une localisation *i* et ses voisines (Anselin, 1988 ; Anselin, 1999 ; Le gallo, 2000 ; Le gallo, 2004).

Le concept d'opérateur spatial, qui en découle est particulièrement important car il permet d'introduire l'autocorrélation spatiale dans les modèles économétriques : l'autocorrélation spatiale est modélisée comme une relation fonctionnelle entre une variable *Y* ou un terme d'erreur ε et son décalage spatial associé, wY pour une variable endogène décalée ou $w\varepsilon$ pour une erreur spatialement décalée. Les spécifications qui en résultent sont étudiées dans le chapitre prochain (chapitre 3).

II-5. Le problème du concept spatial

A partir d'un champ de travail, on peut avoir une variété de matrice de poids spatiale et d'agrégation spatiale. Cette variété crée des problèmes méthodologiques en économétrie spatiale. L'un des problèmes important est : l'unité superficielle modifiable et le manque de critère pour le choix de la matrice de poids spatiale (Anselin, 1988).

II-5-1. L'unité superficielle modifiable

Le problème de l'unité superficiel modifiable est dû au fait que les mesures statistiques pour les données d'une coupe transversale sont sensible au chemin auquel les unités spatiales sont organisées. Ce problème est considéré comme une combinaison de deux problèmes : l'agrégation et l'identification.

L'agrégation des unités spatiales est significative, si le phénomène est homogène à travers les unités d'observations, sinon elle sera considérée comme une mauvaise couverture. Par conséquence, cet aspect du problème doit être considéré comme une spécification de la forme hétérogénéité spatiale (Anselin, 1988).

Le deuxième aspect du problème appartient à l'identification propre de la structure de la dépendance spatiale. On a déjà vu qu'une analyse des associations spatiales est déduite en reliant une variable avec ces décalages spatiaux (les opérateurs spatiaux décalés) et elle est indiquée par un coefficient de corrélation ou de régression. Ce qui implique que les différents choix de la matrice de poids résulteront différents coefficients, alors la mesure de l'association spatiale est indéterminée (Le gallo, 2000 ; Le gallo, 2004).

II-6. Conclusion

Les matrices de poids permettent de modéliser la structure spatiale des données et d'aboutir à la notion de variable spatiale décalée. Cette dernière est d'une importance fondamentale pour introduire l'autocorrélation spatiale dans les modèles économétriques tels que les modèles de régression qui est l'objectif de notre travail ; introduite aussi dans les modèles spatio-temporels, citons par exemple les modèles STARIMA. Ils s'appliquent à un vecteur aléatoire de dimension n représentant les observations d'un phénomène aléatoire en n endroits fixes de G, à l'instant t et ces observations exhibent une autocorrélation spatiale.

Les modèles STARIMA

Suivant Pfeifer et Deutsch (1979), la classe des modèles STARIMA (Space-Time Autoregressive Integerated Moving) multivarié aux différents points de l'ensemble G, notée STARIMA ($p_{\alpha 1;\alpha 2;.....;\alpha p}, d, q_{\lambda 1;\lambda 2;......;\lambda q}$) est exprimée par :

$$(1-B)^d\, z(t) = \sum_{k=1}^{P}\sum_{s=0}^{\alpha_K} \phi_{ks} W^s\,(1-B)^d\, z(t-k) - \sum_{k=1}^{q}\sum_{s=0}^{\lambda_K} \theta_{ks} W^s \varepsilon(t-k) + \varepsilon(t)$$

Où B^K : est un opérateur temporel de retard k, définit de la manière suivante :

$$B^k\left(Z(t)\right) = Z(t-k)$$

$Z(t)$: est le vecteur des observations aux différents points de l'ensemble G à l'instant t, il est de dimension n (card(G)=n).

d : est le nombre des différences nécessaires, pour rendre la série stationnaire.

p : est l'ordre temporel de la partie autorégressive $p \in \mathbb{N}$.

α_k : est l'ordre spatial du $k^{ème}$ terme de la partie autorégressive, avec $\alpha_k \leq \eta^*_{(G)}$.

$\eta^*_{(G)}$: est l'ordre maximum de séparation spatial de l'ensemble G.

ϕ_{ks} : sont des paramètres de la partie autorégressive, d'ordre temporel k et d'ordre spatial s.

q : est l'ordre temporel de la moyenne mobile $q \in \mathbb{N}$.

λ_k : est l'ordre spatial du $k^{ème}$ terme de la moyenne mobile, avec $\lambda_k \leq \eta^*_{(G)}$.

θ_{ks} : sont des paramètres de la moyenne mobile, d'ordre temporel k et d'ordre spatial s.

$W^{(s)}$: est la matrice de séparation spatiale d'ordre s, de l'ensemble G, elle est de dimension n*n.

$\varepsilon(t)$: est le vecteur des innovations aux différents poids de l'ensemble G à l'instant t, il est de dimension n, (dit aussi le vecteur de bruit blanc), tel que :

$$\mathrm{E}\left(\varepsilon(t)\right) = 0 \quad \text{et} \quad \mathrm{E}\left(\varepsilon(t)\varepsilon'(t-h)\right) = \begin{cases} \sigma_\varepsilon^2 \mathrm{I}_n & si \quad h = 0 \\ 0 & si \quad h \neq 0 \end{cases}$$

CHAPITRE III

FORMES ET ESTIMATIONS

III-1. Introduction

Dès 1914, Student suspecte la présence d'une relation entre différentes observations géographiques. Une idée qui entraîne l'abondant de l'hypothèse statistique fondamentale d'observations indépendantes (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2002). Ce phénomène, appelé autocorrélation spatiale, peut être relié à différentes sources. La spécification de l'autocorrélation spatiale nécessite les outils destinés à modéliser l'interdépendance entre les régions, à savoir la matrice de poids et les opérateurs spatiaux (chapitre 2), ces outils permettant alors de tester la présence d'autocorrélation spatiale dans une série spatiale, à l'aide de la statistique globale de Moran (chapitre 1).

Nous verrons dans ce chapitre, qu'à partir de quelques conditions, l'autocorrélation spatiale apparaît sous deux formes (les Variables spatiales décalées et Autocorrélation spatiale des erreurs) et vu que la nature de dépendance en espace est multidirectionnelle, les estimateurs obtenus par la méthode des moindres carrés ordinaires perdent leurs propriétés. Nous développerons d'autres méthodes d'estimation.

III-2. Les principaux types de dépendances spatiales

Paelincke et Klaassen soulignent cinq caractéristiques concernant le champ d'étude à respecter dans la formulation de modèles économétriques spatiaux (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2002).

- Le rôle de l'interdépendance spatiale dans les modèles spatiaux.
- L'asymétrie des relations spatiales.
- L'importance des facteurs explicatifs localisés dans d'autres espaces.
- L'apparition explicite de l'espace dans les modèles.
- La distinction entre interaction ex-ante et ex-post.

Ces principes conduisent à l'utilisation de matrices de poids (asymétrie et apparition de l'espace) et l'emploi de variables spatiales décalées endogènes ou exogènes et d'une autocorrélation spatiale des erreurs ou la combinaison de ces deux aspects (interdépendance spatiale et l'importance des facteurs explicatifs) (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2002). A partir de ces conditions nous remarquons que l'autocorrection spatiale apparaît sous deux différentes formes :

- Les variables spatiales décalées
- Autocorrélation spatiale des erreurs

Comme point de départ, nous considérons le modèle de régression linéaire classique suivant:

$$Y=X\beta+\varepsilon$$

Sous les conventions d'écriture:

Y : est le vecteur (N, 1) des observations de la variable dépendante.

N : est le nombre d'observations.

X : est la matrice (N, K) des observations des variables exogènes, K étant le nombre de paramètres inconnus.

β : est le vecteur (K, 1) des paramètres inconnus.

ε : est le vecteur (N, 1) des erreurs.

Sauf indication contraire, on a les hypothèses suivantes :

H_1 : Hypothèse sur les variables explicatives.

X : est une matrice non stochastique de rang complet $K \leq N$ et lorsque la taille de l'échantillon devient infiniment grand, $\lim_{N\to\infty}(1/N)X'X = Q$ où Q est une matrice finie et non singulière

H_2 : Hypothèse sur les erreurs.

Le vecteur d'erreur consiste en des erreurs non observables qui satisfont les propriétés $E(\varepsilon)=0$ et $E(\varepsilon\varepsilon')=\sigma^2 I_N$, où $E(\cdot)$ désigne l'espérance mathématique et I_N la matrice identité d'ordre N .

L'introduction de l'autocorrélation peut s'effectuer de plusieurs manières :

III-2-1. Les variables spatiales décalées

Dans cette forme l'autocorrélation spatiale donne une substantive interprétation, car elle provient des données affectés par des processus qui relient des lieux différents et qui sont à l'origine d'une organisation spatiale (Anselin, 1988 ; Le gallo, 2000 ; Le gallo, 2002), on peut avoir deux modèles pour ce cas.

III-2-1-1. Le Modèle autorégressif spatial

La première façon d'incorporer l'autocorrélation spatiale est le modèle autorégressif spatial (SAR), où la variable d'intérêt en une localisation est jointement déterminée par ces valeurs en d'autres localisations (Le gallo, 2000 ; Le gallo, 2002; Le sage, 1998; Mulkay, 2007):

$$Y = \rho wY + X\beta + \varepsilon \qquad (3-2-1-1)$$

ρ : est le paramètre spatial autorégressif indiquant l'ampleur de l'interaction existant entre les observations de Y.

wY : est la variable endogène décalée pour la matrice de poids w.

ε : est le vecteur des erreurs iid (0, I σ^2).

Ce modèle peut être écrit sous une autre façon Y^*, qui s'interprète comme la variable dépendante « Filtrée », où les effets de l'autocorrélation spatiale sont éliminés (Le gallo, 2000).

$$Y^* = (\mathrm{I} - \rho w)Y = X\beta + \varepsilon \qquad (3-2-1-2)$$

Supposons que la matrice $(\mathrm{I} - \rho w)$ est inversible, alors on a une troisième forme de (3-2-1-1).

$$Y = (I - \rho w)^{-1} X\beta + (I - \rho w)^{-1}\varepsilon$$

III-2-1-2. Le Modèle régressif croisé spatial

Une autre façon de prendre en compte l'autocorrélation spatiale est d'introduire une ou plusieurs variables exogènes décalées (Le gallo, 2000 ; Le gallo, 2002):

$$Y = X\beta + wZ\delta + \varepsilon$$

Z : est une matrice de dimension (N, L), L des variables correspondantes ou non aux variables incluses dans X.
wZ : est la variable exogène décalée pour la matrice de poids *w*.
δ: est le vecteur (L, 1) des paramètres spatiaux indiquant l'ampleur de la corrélation spatial existant entre les observations de Y et celles de Z.

III-2-2. Autocorrélation spatiale des erreurs

Dans ce type, I'autocorrélation spatiale provient d'une mauvaise spécification du modèle, elle est considéré comme une nuisance. Elle touche le terme d'erreur, c'est-à-dire le terme d'erreur peut être spécifiée suivant une forme autorégressif spatial ou moyenne mobile spatiale (Anselin et Rey, 1991).

III-2-2-1. La forme autorégressive spatiale dans les erreurs

La forme autorégressive spatiale dans le terme des erreurs (SEM) est souvent utilisée en pratique (Le gallo, 2000 ; Le gallo, 2002; Le sage, 1998 ; Mulkay, 2007):

$$Y = X\beta + \varepsilon \quad où \quad \varepsilon = \lambda w\varepsilon + \mu \qquad (3-2-2-3)$$

λ : est le paramètre autorégressif spatial indiquant l'intensité de l'autocorrélation spatiale entre les résidus de la régression.
μ : est le vecteur des erreurs iid (0, $I\sigma^2$).

Si la matrice $(I-\lambda w)$ est non-singulière, le modèle (3-2-2-3) se réécrit sous la forme suivante (Le gallo, 2000):

$$Y = X\beta + (I-\lambda w)^{-1}\mu \qquad (3-2-2-4)$$

A travers la transformation inverse spatiale $(I-\lambda w)^{-1}$ qui est une matrice pleine, on dit qu'on a un choc aléatoire dans une région `*i*` spécifique qui affecte la valeur de *Y* de cette région et également les valeurs de *Y* dans les autres régions (Le gallo, 2000).

En multipliant les deux membres de (3-2-2-4) par $\left(I-\lambda w\right)$, on obtient :

$$\left(I-\lambda w\right)\,Y=\left(I-\lambda w\right)X\,\beta+\mu \qquad (3-2-2-5)$$

C'est un modèle où les variables expliquées et les variables explicatives sont filtrés spatialement, le terme μ étant homoscédastique (Le gallo, 2000 ; Le gallo, 2002).

On peut avoir un autre modèle dit de DURBIN SPATIAL (SDM) en développant le modèle (3-2-2-2-3) (Le gallo, 2000 ; Mulkay, 2007).

$$Y=\lambda wY+X\,\beta-\lambda wX\,\beta+\mu$$

Parfois des effets qui ne sont pas pris en compte dans les variables explicatives se retrouvent dans les erreurs sous forme d'une autocorrélation. Dans ce cas, autocorrélation spatiale permet de repérer des variables significatives mais non prises en compte dans le modèle.

III-2-2-2. La forme moyenne mobile spatiale dans les erreurs

Cette forme est moins fréquente, elle est spécifiée de la façon suivante (Le gallo, 2000 ; Le gallo, 2002 ; Le sage, 1998):

$$Y=X\,\beta+\varepsilon \quad où \quad \varepsilon=\lambda w\mu+\mu$$

γ : est le coefficient moyen mobile.

μ : est le terme d'erreur homoscédastique (iid).

En pratique, on obtient rarement un modèle avec les deux types. On cherche à modéliser la dépendance spatiale par l'autocorrélation des erreurs ou par une variable spatiale décalée et non les deux. Il y a d'autres types de dépendance spatiale en combinant les deux principales formes et à l'ordre de dépendance supérieure a 1. Quoiqu' Anselin et Bara considèrent les processus d'ordre supérieur comme une mal spécification de la matrice de poids (Anselin, 1988; Le gallo, 2000).

III-3. Estimation des modèles spatiaux

Nous allons voir maintenant des méthodes d'estimation, en particulier l'approche la plus fréquente le maximum de vraisemblance (MV). Vue la nature de dépendance en espace est bidimensionnelle et multidirectionnelle, les estimateurs obtenus par la méthode des moindres carrés ordinaires (MCO) perdent leurs propriétés, en présence d'une variable endogène décalée (il n'y a pas convergence) et /ou d'une autocorrélation spatiale des erreurs (elle ne nous donne pas un estimateur efficace) (Anselin, 1988).

III-3-1. Limitation de MCO pour les modèles spatiaux

III-3-1-1. En présence d'une variable dépendante spatialement décalée

Nous voulons vérifier la non convergence de MCO en présence d'une variable dépendante spatialement décalée indépendamment des propriétés des erreurs.

Preuve :

Nous considérons un modèle autorégressif pur.

$$Y=\rho wY+\varepsilon$$

Où Y : la variable dépendante.

w : est une matrice de poids.

ρ : est le coefficient autorégressif spatiale.

ε : le terme d'erreurs iid.

On cherche l'estimateur de ρ en utilisant la méthode de MCO.

$$\varepsilon'\varepsilon=Y'Y-2\rho Y'w'Y+\rho^2 Y'w'wY$$

$$\frac{\partial\varepsilon'\varepsilon}{\partial\rho}=0 \quad\Leftrightarrow\quad -2Y'w'Y+2\rho Y'w'wY=0$$

$$\hat{\rho}=\left((wY)'wY\right)^{-1}(wY)'Y \;.$$

On pose $Y_l = w\,Y$ et on remplace Y_l par sa valeur, on obtient :

$$\hat{\rho} = \rho + \left(Y_l' Y_l\right)^{-1} Y_l' \;\varepsilon$$

On remarque que le deuxième terme est différent de zéro, ce qui nous donne un estimateur de MCO biaisé.

Asymptotiquement, la consistance de MCO dépend de deux conditions (Anselin, 1988; Le sage, 1998).

$P \lim_{N \to \infty} Y_l' Y_l / N = Q$ est une matrice finie et non singulière.

$P \lim_{N \to \infty} Y_l' \, \varepsilon / N = 0$.

Dans notre cas, la première condition peut être vérifiée en posant des conditions sur les valeurs de ρ et la structure des matrices de poids w (Anselin, 1988). Par contre, la deuxième condition ne peut pas être vérifiée dans le cas spatial. Car, On a :

$$Y_l = w\,Y \quad \text{et} \quad Y = (I - \rho w)^{-1} \varepsilon .$$

Alors

$$P \lim_{N \to \infty} Y_l ' \varepsilon / N = P \lim_{N \to \infty} \varepsilon' \left(I - \rho w'\right)^{-1} w' \varepsilon / N \neq 0$$

La présence de la matrice de poids spatiale dans cette expression donne une forme quadratique dans le terme d'erreur, donc excepté pour le cas $\rho = 0$ la limite en probabilité ne tend pas vers 0. D'où l'estimateur n'est pas asymptotiquement consistant. De là, les estimateurs obtenus par la méthode de MCO pour les paramètres des modèles spatiaux sont inconsistants indépendamment des propriétés des termes des erreurs (Anselin, 1988; Le sage, 1998).

III-3-1-2 En présence d'une autocorrélation des erreurs

Dans ce cas, les paramètres estimés sont sans biais mais inefficients car la structure de la matrice de variance des bruits est non diagonale.

Soit un modèle de régression linéaire simple.

$$Y = X\beta + \varepsilon \quad \text{tq} \quad \varepsilon = \lambda w \varepsilon + \mu$$

où :

w : est la matrice de poids spatiale

λ : est le coefficient autorégressif.

μ : est un terme d'erreur qui satisfait les hypothèses classiques iid avec une variance constante σ^2.

La variance des erreurs correspondante à ε est de la forme :

$$\Omega = \mathrm{E}\left(\varepsilon\varepsilon'\right) = \sigma^2\left(I - \lambda w\right)^{-1}\left[\left(I - \lambda w\right)^{-1}\right]'$$

Ce type de modèle peut être considéré comme une régression linéaire avec une matrice de variance paramétrée pour le terme d'erreur, d'où les propriétés usuelles de MCG et MCO existent. Or, dans le cas spatial, ces méthodes ne nous conduisent pas à des estimateurs consistants dus à la nature multidirectionnelle de la dépendance spatiale (Anselin, 1988).

III-3-2 Estimation par le maximum de vraisemblance

L'approche du maximum de vraisemblance pour les modèles spatiaux moyens mobiles et autorégressifs a été suggérée par Cliff et Ord (1973) et d'autres auteurs (Anselin, 1988 ; Le sage, 1998). Sous les conditions de régularités, les propriétés asymptotiques des estimateurs de vraisemblance pour les modèles économétriques spatiaux, tels que la convergence, l'efficacité et la normalité asymptotique ont été étudiés et vérifiés par Bates et White et d'autres auteurs pour les modèles à variables dépendantes spatialement décalées, Magnus et Rothemberg et d'autres auteurs pour les modèles de régression linéaire avec des effets spatiaux en terme d'erreur (Anselin, 1988).

Les conditions de régularités se traduisent par :

1. L'existence des paramètres sous considérations.
2. Differentiabilité continues de logvraisemblance (aux seconds et troisièmes ordres, et pour les valeurs des paramètres au voisinage de la vraie valeur).
3. Les dérives partielles des différents paramètres sont bornées.
4. L'existence de la matrice covariance non singulière et/ou définie positive.
5. Les différentes formes quadratiques sont finies.

Ces conditions sont importantes pour garantir une bonne fonction de vraisemblance. Ils sont essentiellement satisfaits par une structure non explosive des matrices de poids spatiales (Anselin, 1988, Le gallo, 2000).

III-3-2-1. La fonction de vraisemblance et le jacobien

Nous considérons un modèle spatial général où les deux principales formes de la dépendance spatiale existent.

$$Y = \rho w_1 Y + X\beta + \varepsilon \quad \text{où} \quad \varepsilon = \lambda w_2 \varepsilon + \mu \quad \text{et} \quad \mu \sim iid \quad N\left(0, I\sigma^2\right).$$

Sous la condition de normalité des résidus, nous retrouvons la fonction de vraisemblance de notre modèle.

$$L(\mu) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\mu'\mu\right).$$

Par un changement de variable, on veut avoir la fonction logvraisemblance de y .

D'après le modèle général μ est :

$$\mu = \left(I - \lambda w_2\right)\left(\left(I - \rho w_1\right)y - X\beta\right).$$

Le jacobien de la transformation est :

$$J = \det\frac{\partial\mu}{\partial y} = \left|I - \lambda w_2\right| \cdot \left|I - \rho w_1\right|.$$

On note : $\det(A) = |A|$

La fonction logvraisemblance de y , s'écrit donc:

$$Ln\left(L\left(y/\theta\right)\right) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) + \ln\left(\left|I - \rho w_1\right|\right) + \ln\left(\left|I - \lambda w_2\right|\right) - \frac{1}{2\sigma^2}\mu'\mu$$

où θ est le vecteur des paramètres $(\beta,\lambda,\sigma,\rho)$.

Les propriétés asymptotiques de l'estimateur du MV sont vérifiées seulement si les conditions de régularités de log vraisemblance sont satisfaites. Ces conditions sont prises pour le Jacobien (Anselin, 1988; Anselin, 1999; Le sage, 1998; Mulkay, 2007).

$\left|I - \lambda w_2\right| \rangle\, 0$.

$\left|I - \rho w_1\right| \rangle\, 0$.

Ces contraintes résultent une restriction sur les valeurs qui peuvent prendre les coefficients spatiaux.

En effet, la condition :

$$\left(\left|I-\alpha w\right| > 0 \;\; où \;\; \alpha = \lambda, \rho \quad et \quad w = w_1, w_2 \right) \;\Leftrightarrow\; \left(\prod_{i=1}^{N} \left(1-\alpha\omega_i\right) > 0 \quad \forall\, i = 1, \ldots, N \right)$$

où ω_i sont les valeurs propres de la matrice de poids w.

Alors, $\ln\left(\prod_{i=1}^{n} \left(1-\alpha\omega_i \right)\right) = \sum_{i=1}^{n} \ln\left(1-\alpha\omega_i \right)$.

La condition précédente revient à la suivante :

$$\left(1-\alpha\omega_i > 0 \,,\, \forall\, i = 1, \ldots, N \right) \;\Leftrightarrow\; \left(1 > \alpha\omega_i \,,\, \forall\, i = 1, \ldots, N \right).$$

Pour $\omega_i > 0$, on a:

$$\alpha < \frac{1}{\omega_i} \,, \forall\, i = 1, \ldots, N \text{ . Il suffit de prendre } \alpha < \min\left\{ \frac{1}{\omega_i}, \forall i = 1, \ldots\ldots, N. \right\}$$

d'où : $\alpha < \frac{1}{\omega_{max}}$.

Pour $\omega_i < 0$, on a:

$$\alpha > \frac{1}{\omega_i} \,, \forall\, i = 1, \ldots, N \text{ . Il suffit de prendre } \alpha > \max\left\{ \frac{1}{\omega_i}, \forall i = 1, \ldots\ldots, N. \right\}$$

d'où : $\alpha > \frac{1}{\omega_{min}}$.

Alors :

$$\frac{1}{\omega_{min}} < \alpha < \frac{1}{\omega_{max}}.$$

Avec ω_{max} est la valeur propre positive maximale et ω_{min} la valeur propre négative la plus grande en valeur absolue de w (Anselin, 1988; Le gallo, 2000; Le gallo, 2002; Le sage, 1998).

.

Dans le cas où les erreurs suivent un processus moyen mobile, notre modèle sera :

$$Y = \rho w_1 Y + X\beta + \varepsilon \quad \text{tel que } \varepsilon = \delta w_2 \mu + \mu$$

où $\mu \ iid\left(0, I\sigma^2\right)$.

Si $(I+\delta w_2)^{-1}$ est inversible,

$$\mu = \left(I + \delta w_2\right)^{-1} \varepsilon = \left(I + \delta w_2\right)^{-1}\left(\left(I - \rho w_1\right)Y - X\beta\right).$$

La fonction de vraisemblance de notre vecteur normale μ est :

$$L(\mu) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\mu'\mu\right).$$

Par un changement de variable, on veut avoir la fonction logvraisemblance de y , le jacobien de la transformation est :

$$J = \det\frac{\partial\mu}{\partial y} = \left|I+\delta w_2\right| \cdot \left|I-\rho w_1\right|$$

On note : $\det(A) = |A|$.

La fonction logvraisemblance de y s'écrit donc :

$$Ln\left(L\left(Y/\theta\right)\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) + \ln\left(\left|I - \rho w_1\right|\right) + \ln\left(\left|I + \delta w_2\right|\right) - \frac{1}{2\sigma^2}\mu'\mu$$

où θ est le vecteur des paramètres $(\beta, \delta, \sigma, \rho)$.

Les conditions de régularités de la fonction logvraisemblance sont satisfaites si :

$\left|I + \delta w_2\right| \rangle 0$.

$\left|I - \rho w_1\right| \rangle 0$.

L'espace du paramètre ρ est le même que le modèle précédant, par contre l'espace du paramètre δ est $\left]\frac{-1}{\omega_{\max}}, \frac{-1}{\omega_{\min}}\right[$. (Anselin, 1988; Le gallo, 2000; Le gallo, 2002; Le sage, 1998).

III-3-3. Estimation par le maximum de vraisemblance

En pratique, obtenir un modèle général est un cas rare. On cherche souvent à trouver des modèles simples comme le modèle spatial autorégressif et le modèle à erreurs auto corrélées spatialement. On utilise des techniques d'optimisation non-linéaires pour maximiser la fonction logvraisemblance. Pour l'estimation des modèles simples, on utilise la fonction logvraisemblance concentrée, le principe de cette méthode est de résoudre une partie des équations associées aux conditions du premier ordre et d'introduire ensuite les solutions obtenues dans la fonction logvraisemblance. Cette dernière fonction est simplifiée concentrée, qui ne dépend que de quelques paramètres, un dans le meilleur des cas. Dès lors, on peut trouver des estimations de ces paramètres par balayage, c'est-à-dire, en évaluant la fonction de vraisemblance pour un petit intervalle de ces paramètres, on trouve alors un maximum local qui correspond à la solution (Anselin, 1998; Le gallo, 2000; Le sage, 1999). On peut aussi maximiser la vraisemblance par rapport à l'ensemble des paramètres comme le propose Mulkay (2007).

III-3-3-1. Cas d'un modèle autorégressif spatial

Dans le cas d'un modèle autorégressif spatiale fonction logvraisemblance (3-2-1-1) la fonction logvraisemblance est :

$$Ln\left(L\left(Y/\theta\right)\right)=-\frac{N}{2}\ln\left(2\pi\right)-\frac{N}{2}\ln\left(\sigma^2\right)+\ln\left(\left|I-\rho w\right|\right)-\frac{1}{2\sigma^2}\left(Y-\rho wY-X\beta\right)'\left(Y-\rho wY-X\beta\right)$$

où θ est le vecteur des paramètres (β,σ,ρ).

D'après la méthode de la fonction logvraisemblance concentrée, on cherche à résoudre une partie d'équations associées aux conditions du premier ordre. Pour cela, on choisit les paramètres β,σ^2.

Les dérivées premières

A partir des conditions du premier ordre usuelles

$$\frac{\partial \ln(L)}{\partial \beta}=\frac{1}{2\sigma^{2}}\left(2X'Y-2\rho X'wY-2X'X\beta\right)=0 \;.$$

$$\hat{\beta}=\left(X'X\right)^{1-}X'\left(I-\rho w\right)Y \;.$$

$$\frac{\partial \ln(L)}{\partial \sigma^{2}}=-\frac{N}{2\sigma^{2}}+\frac{1}{2\sigma^{4}}\left(Y-\rho wY-X\beta\right)'\left(Y-\rho wY-X\beta\right)=0.$$

$$\hat{\sigma}^{2}=\frac{\left(Y-\rho wY-X\hat{\beta}\right)'\left(Y-\rho wY-X\hat{\beta}\right)}{N}.$$

En remplaçant ces estimateurs dans la fonction logvraisemblance, on obtient la fonction logvraisemblance concentrée qui ne dépend que de ρ .

$$Ln\left(L_{C}\right)=-\frac{N}{2}\ln\left(2\pi\right)-\frac{N}{2}-\frac{N}{2}\ln\left(\left(Y-\rho wY-X\hat{\beta}\right)'\left(Y-\rho wY-X\hat{\beta}\right)/N\right)+\ln\left(\left|I-\rho w\right|\right).$$

A partir des conditions du premier ordre et de ces définitions, en posant $A=\left|I-\rho w\right|$, on a :

$$\frac{\partial \ln\left(\left|A\right|\right)}{\partial \rho}=Tr\left(A^{-1}\frac{\partial A}{\partial \rho}\right).$$

$$\frac{\partial A}{\partial \rho}=-w \;.$$

Alors $\frac{\partial \ln\left(\left|A\right|\right)}{\partial \rho}=-Tr\left(\left(I-\rho w\right)^{-1}w\right).$

On trouve

$$\frac{\partial \ln\left(L_{C}\right)}{\partial \rho}=\frac{\left(Y-\rho wY-X\hat{\beta}\right)'wY}{\hat{\sigma}^{2}}-Tr\left(\left(I-\rho w\right)^{-1}w\right)=0 \,.$$

L’estimateur de ρ est la valeur qui annule l’équation non-linéaire précédente, pour résoudre cette équation en fait recoure aux méthodes numériques telles que la méthode de newton.

D’après Anselin (1988), sous les conditions de régularités, les propriétés asymptotiques des estimateurs de vraisemblance pour les modèles économétriques spatiaux, tels que la convergence, l'efficacité et la normalité asymptotique ont été étudiés et vérifiés par Bates, White, Magnus et Rothemberg. Alors Les estimateurs obtenus par la méthode du MV sont asymptotiquement efficaces. Donc La matrice variance covariance asymptotique est donnée par l'inverse de la matrice d'information. Les éléments de la matrice d'information pour ce modèle général sont obtenus par les dérivées partielles secondes de la logvraisemblance, tout en respectant les éléments du vecteur paramètre θ. Pour les éléments de la diagonale :

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\beta\partial\beta'}=-\frac{X'X}{\sigma^2}.$$

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\sigma^2\partial\sigma^2}=\frac{N}{2\sigma^4}-\frac{(Y-\rho wY-X\beta)'(Y-\rho wY-X\beta)}{2\sigma^6}.$$

$$\frac{\partial^2(\ln L_c(Y/\theta))}{\partial\rho^2}=-\frac{(wY)'(wY)}{\hat{\sigma}^2}-Tr\left((I-\rho w)^{-1}w(I-\rho w)^{-1}w\right).$$

Remarque

$$\frac{\partial Tr(A)^{-1}w}{\partial\rho}=Tr\left(\frac{\partial(A)^{-1}w}{\partial\rho}\right)=Tr\left((A)^{-1}w(A)^{-1}w\right)\quad \text{avec}\quad A=(I-\rho w)$$

Pour les autres termes :

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\beta\partial\rho}=\frac{\partial^2 L_c(Y/\theta)}{\partial\rho\partial\beta}=-\frac{X'wY}{\sigma^2}.$$

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\beta\partial\sigma^2}=\frac{\partial^2(\ln L(Y/\theta))}{\partial\sigma^2\partial\beta}=-\frac{1}{\sigma^4}(X'Y-\rho X'wY-X'X\beta).$$

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\sigma^2\partial\rho}=\frac{\partial^2(\ln L_c(Y/\theta))}{\partial\rho\partial\sigma^2}=-\frac{(Y-\rho wY-X\beta)'wY}{\sigma^4}.$$

Pour calculer les éléments de la matrice d'information, nous avons besoin d'introduire à ce niveau un théorème et certains résultats qui se présentent comme éléments clé.

Théorème 1

Quand $\mathrm{X} \sim \mathrm{N}(\mu, V)$, on a:

$\mathrm{E}(\mathrm{X}' A \mathrm{X}) = Tr(AV) + \mu' A \mu$ (C'est aussi vrai si X n'est pas une normale).

Le cumulant d'ordre '*r*' de $\mathrm{X}' A \mathrm{X}$ est

$$\mathrm{K}_r(\mathrm{X}' A \mathrm{X}) = (2)^{r-1}(r-1)\left(Tr(AV)^r + r\mu' A (VA)^{r-1}\mu\right).$$

La variance covariance de $\mathrm{X}, \mathrm{X}' A \mathrm{X}$ est

$$\mathrm{cov}(\mathrm{X}, \mathrm{X}' A \mathrm{X}) = 2VA\mu .$$

Résultats

D'après le modèle considéré (3-2-1-1), nous avons :

$\mathrm{E}(\varepsilon) = 0$.

$\mathrm{E}(\varepsilon\varepsilon') = I\sigma^2$.

$\mathrm{E}(Y) = (I - \rho w)^{-1} X\beta$.

$\mathrm{E}(YY') = \left((I-\rho w)^{-1} X\beta\right)\left((I-\rho w)^{-1} X\beta\right)' + (I-\rho w)^{-1}\sigma^2\left((I-\rho w)^{-1}\right)'$.

$V(Y) = (I-\rho w)^{-1}\sigma^2\left((I-\rho w)^{-1}\right)'$.

En faisant appel au théorème1 et aux résultats décrits, nous trouvons :

$$(I(\theta))^{-1} = \begin{bmatrix} \dfrac{X'X}{\sigma^2} & 0 & \dfrac{X' w_A X\beta}{\sigma^2} \\ 0 & \dfrac{N}{2\sigma^2} & \dfrac{Tr(w_A)}{\sigma^2} \\ \dfrac{(X' w_A X\beta)'}{\sigma^2} & \dfrac{Tr(w_A)}{\sigma^2} & Tr(w_A w_A') + Tr\left((w_A)^2\right) + \dfrac{(w_A X\beta)'(w_A X\beta)}{\sigma^2} \end{bmatrix}^{-1}.$$

Avec $w_A = w(I - \rho w)^{-1}$.

III-3-3-2. Cas d'un modèle d'une autorégression spatiale des erreurs

C'est-à-dire

$$Y = X\beta + \varepsilon \quad \text{tq} \quad \varepsilon = \lambda w\varepsilon + \mu \quad \text{où} \quad \mu \sim \mathrm{N}\left(0, I\sigma^2\right) \text{ iid.}$$

La fonction de logvraisemblance de ce modèle est :

$$Ln\left(L\left(Y/\theta\right)\right) = -\frac{N}{2}\ln\left(2\pi\right) - \frac{N}{2}\ln\left(\sigma^2\right) + \ln\left(\left|I-\lambda w\right|\right) - \frac{1}{2\sigma^2}\left(Y - X\beta\right)'\Omega\left(\lambda\right)\left(Y - X\beta\right)$$

où $\Omega\left(\lambda\right) = \left(I-\lambda w\right)'\left(I-\lambda w\right)$ et $\theta = \left(\beta, \sigma^2, \lambda\right)$ le vecteur des paramètres.

On procède de la même manière que le cas précédant, on cherche les estimateurs de β et σ^2 conditionnellement à λ .

A partir des conditions du premier ordre pour β et σ^2 .

$$\frac{\partial \ln\left(L\right)}{\partial\beta} = \frac{1}{2\sigma^2}\left(2Y'\Omega\left(\lambda\right)X - 2X'\Omega\left(\lambda\right)X\beta\right) = 0$$

$$\hat{\beta} = \left(X'\Omega\left(\lambda\right)X\right)^{-1}X'\Omega\left(\lambda\right)Y \, .$$

$$\frac{\partial \ln\left(L\right)}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\left(Y - X\beta\right)'\Omega\left(\lambda\right)\left(Y - X\beta\right) = 0.$$

$$\hat{\sigma}^2 = \frac{\left(Y - X\hat{\beta}\right)'\Omega\left(\lambda\right)\left(Y - X\hat{\beta}\right)}{N}.$$

En remplaçant ces estimateurs dans la fonction logvraisemblance, on obtient une fonction qui ne dépend que de λ .

$$Ln\left(L_c\right) = -\frac{N}{2}\ln\left(2\pi\right) - \frac{N}{2} - \frac{N}{2}\ln\left(\left(Y - X\hat{\beta}\right)'\Omega\left(\lambda\right)\left(Y - X\hat{\beta}\right)/N\right) + \ln\left(\left|I-\lambda w\right|\right).$$

A partir des conditions du premier ordre et ces définitions, en posant $A=\left|I-\lambda w\right|$, on a :

$$\frac{\partial \ln\left(\left|A\right|\right)}{\partial \lambda}=Tr\left(A^{-1}\frac{\partial A}{\partial \lambda}\right).$$

$$\frac{\partial A}{\partial \lambda}=-w\,.$$

Alors $\frac{\partial \ln\left(\left|A\right|\right)}{\partial \lambda}=-Tr\left(\left(I-\rho w\right)^{-1}w\right)$.

$$\frac{\partial \ln\left(L_c\right)}{\partial \lambda}=\frac{\hat{\varepsilon}_0'\left(w+w'\right)\hat{\varepsilon}_0-2\lambda\left(w\hat{\varepsilon}_0\right)'\left(w\hat{\varepsilon}_0\right)}{2\hat{\sigma}^2}-Tr\left(\left(I-\rho w\right)^{-1}w\right)=0\,.$$

Où $\hat{\varepsilon}_0=Y-X\hat{\beta}$.

L'estimateur de λ est la valeur qui annule la fonction précédente, il est estimé par des méthodes numériques. Les éléments de la matrice de variance covariance asymptotique pour ces estimateurs sont obtenus par les dérivées partielles secondes de logvraisemblance (Anselin, 1988), tout en respectant les éléments du vecteur des paramètres.

Pour les éléments de la diagonale :

$$\frac{\partial^2(\ln L\left(Y/\theta\right))}{\partial\beta\partial\beta'}=-\frac{X'\Omega\left(\lambda\right)X}{\sigma^2}\,.$$

$$\frac{\partial^2(\ln L\left(Y/\theta\right))}{\partial\sigma^2\partial\sigma^2}=\frac{N}{2\sigma^4}-\frac{\left(Y-X\beta\right)'\Omega\left(\lambda\right)\left(Y-X\beta\right)}{2\sigma^6}\,.$$

$$\frac{\partial^2(\ln L_c\left(Y/\theta\right))}{\partial\lambda^2}=-\frac{\left(w\hat{\varepsilon}_0\right)'\left(w\hat{\varepsilon}_0\right)}{\hat{\sigma}^2}-Tr\left(\left(I-\lambda w\right)^{-1}w\left(I-\lambda w\right)^{-1}w\right).$$

Pour les autres termes :

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\beta\partial\lambda}=\frac{\partial^2 L_c\left(Y/\theta\right)}{\partial\lambda\partial\beta}=-\frac{Y'\left(2\lambda ww'-w-w'\right)X-X'\left(2\lambda ww'-w-w'\right)X\beta}{\sigma^2}.$$

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\beta\partial\sigma^2}=\frac{\partial^2\left(\ln L\left(Y/\theta\right)\right)}{\partial\sigma^2\partial\beta}=-\frac{1}{\sigma^4}\left(-Y'\Omega(\lambda)X-X'\Omega(\lambda)X\beta\right).$$

$$\frac{\partial^2(\ln L(Y/\theta))}{\partial\sigma^2\partial\lambda}=\frac{\partial^2\left(\ln L_c\left(Y/\theta\right)\right)}{\partial\lambda\partial\sigma^2}=-\frac{\left(Y-X\beta\right)'\left(2\lambda ww'-\left(w+w'\right)\right)\left(Y+X\beta\right)}{2\sigma^4}.$$

Résultats :

D’après le modèle considéré (3-2-2-1), nous avons

$$\varepsilon=\left(I-\lambda w\right)^{-1}\mu$$

$$\mathrm{E}\left(\varepsilon\right)=0.$$

$$\mathrm{E}\left(\varepsilon\varepsilon'\right)=\left(I-\lambda w\right)^{-1}\sigma^2\left(I-\lambda w\right)^{-1'}.$$

$$\mathrm{E}\left(Y\right)=X\beta.$$

$$\mathrm{E}\left(YY'\right)=\left(I-\lambda w\right)^{-1}\sigma^2\left(\left(I-\lambda w\right)^{-1}\right)'+\left(X\beta\right)\left(X\beta\right)'$$

$$V\left(Y\right)=\left(I-\lambda w\right)^{-1}\sigma^2\left(\left(I-\lambda w\right)^{-1}\right)'.$$

En utilisant Théorème 1 et ces résultats, on trouve :

$$\left(I(\theta)\right)^{-1}=\begin{bmatrix}\frac{X'\Omega(\lambda)X}{\sigma^2} & 0 & 0\\ 0 & \frac{1-N}{2\sigma^2} & \frac{Tr\left(w_B\right)}{\sigma^2}\\ 0 & \frac{Tr\left(w_B\right)}{\sigma^2} & Tr\left(w_B'w_B\right)+Tr\left(\left(w_B\right)^2\right)\end{bmatrix}^{-1}$$

Avec $w_B = (I - \lambda w)^{-1} w$

III-3-4. Autres méthodes d'estimation

L'estimation par l'approche du maximum de vraisemblance est la méthodologie la plus connue pour les modèles spatiaux. C'est la seule technique considérée et appliquée. Cependant, plusieurs alternatives on été suggérées pour éviter quelques problèmes associés à l'estimation par MV, spécialement la complexité numérique de l'optimisation non linéaire et le domaine de définition des paramètres. Deux alternatives sont considérées et nous suggérons un aperçu sur l'application de ces méthodes sur quelques modèles:

- L'estimation par les variables instrumentales.
- L'approche Bayesienne.
-

III-3-4-1. Estimation par les variables instrumentales

Malgré le fait que les propriétés asymptotiques de cette méthode soient similaires à la méthode de MV et la facilité de l'exécution numérique cette méthode est peu appliquée pour les modèles spatiales. Elle est considérée comme une technique d'estimation pour une spécification d'un modèle autorégressif spatial pur du premier ordre (Anselin, 1988; Le gallo, 2000). Si on revoit le début de ce chapitre, on remarque que la méthode de MCO est affaiblie pour le modèle autorégressif pur, cela est dû à la corrélation existante entre les variables spatiales et le terme d'erreur.

De là, l'approche par les variables instrumentales est basée sur l'existence d'un ensemble d'instrument «Q», qui est fortement corrélé avec les variables d'origines (ie [wY ,X]) mais asymptotiquement non corrélé avec les termes d'erreurs :

$\mathrm{P}\lim\left(1/N\right)Q'\varepsilon = 0$.

$\mathrm{P}\lim\left(1/N\right)Q'Z = M_{QZ}$ une matrice finie non singulière.

Alors, pour un modèle autorégressif on a :

$$Y = \left[wY, X\right]\theta + \varepsilon = Z\theta + \varepsilon \, .$$

En multipliant cette équation par $(1/N)Q'$ et en utilisant les propriétés de la limite en probabilité supposées, on obtient :

$$\mathrm{P}\lim(1/N)Q'Y = \mathrm{P}\lim(1/N)Q'Z\theta + 0\,.$$

Cette expression peut nous donner $\hat{\theta}$ l'estimateur des variables instrumentales

$$\hat{\theta} = [Q'Z]^{-1} Q'Y \quad \text{P.P}$$

Pour vu que $Q'Z$ soit inversible, Q et Z ont le même nombre de colonnes.

Dans le cas où on a plus d'instruments que de variables d'origines, il n'existe pas une solution exacte. Par conséquent, on utilise une approche similaire à l'estimation par MCO. Le problème peut être formulé comme une minimisation de la distance quadratique par rapport à la norme S :

$$\min \Phi(\theta) = \min\Big((Y - Z\theta)' QSQ'(Y - Z\theta) \Big).$$

En choisissant $S = (Q'Q)^{-1}$ la solution de cette optimisation est donnée par :

$$\hat{\theta} = \left(Z \mathrm{P}_Q Z \right)^{-1} Z' \mathrm{P}_Q Y$$

Avec $\mathrm{P}_Q = Q(Q'Q)^{-1}Q'$ une matrice de projection symétrique idempotente.

La covariance asymptotique des paramètres estimés est :

$$\mathrm{cov}(\hat{\theta}) = \overset{2}{\sigma}(Z'P_Q Z)^{-1}$$

où la variance des erreurs est estimée par :

$$\overset{\hat{2}}{\sigma} = (Y - Z\hat{\theta})'(Y - Z\hat{\theta})/N$$

Sous des hypothèses raisonnables satisfaites lorsque les poids sont basés sur la contiguïté, les estimateurs des variables instruments sont convergents et asymptotiquement normaux (Anselin, 1988).

D'après le gallo (2000), Kelejian et Procha ont développé une approche par la méthode des moments généralisés (GMM), pour le coefficient spatial dans le modèle à erreurs autocorrélées, car la méthode des variables instrumentales ne fournit pas des estimateurs convergents. Ils développent ensemble des conditions sur les moments pour les paramètres dans le modèle à erreurs autocorrélées, soit :

$$Y = X\beta + \varepsilon$$

$$\varepsilon = \lambda w\varepsilon + \mu$$

Si $\mu \sim iid\,\mathrm{N}\left(0, \sigma^2 \mathrm{I}\right)$ les trois conditions sont les suivantes :

$$E\left(\mu'\mu/N\right) = \sigma^2 ; \qquad E\left(\left(w\mu\right)'\left(w\mu\right)/N\right) = \sigma^2\left(\frac{1}{N}Tr\left(w'w\right)\right) ; \qquad E = \left(\mu' w\mu/N\right) = 0 .$$

En remplaçant μ par $\hat{\varepsilon} - \lambda w\hat{\varepsilon}$, où $\hat{\varepsilon}$ est le vecteur des résidus des MCO. On obtient un système des trois équations pour les paramètres $\lambda, \lambda^2, \sigma^2$. Les deux méthodes d'estimations peuvent être combinées pour estimer les paramètres d'un modèle général.

III-3-4-2. L'approche Bayesienne pour l'estimation des modèles spatiaux

L'application des techniques Bayé siennes pour l'estimation des modèles spatiaux ont été suggérés par Hepple et Anselin, pour les modèles autorégressifs de premier ordre et pour les modèles de régression avec une autocorrélation spatiale des erreurs (Anselin, 1988).

La statistique d'inférence dans les travaux bayésiens est basée sur la combinaison de l'information a priori concernant les paramètres du modèle et l'information contenue dans l'ensemble des données. La loi a priori prise doit être donnée ou non informative s'il n'y a pas d'hypothèse précise. L'information contenue dans l'ensemble des données est représentée par la fonction de vraisemblance. L'information contenue dans le modèle doit être prise comme une distribution conjointe en fonction de la densité des variables et paramètres $h\left(\theta, y\right)$.

En terme de probabilité conditionnelle, c'est équivalent à :

$$h\left(\theta, y\right) = g\left(\theta/y\right) \cdot f\left(y\right) = f\left(y/\theta\right) \cdot g\left(\theta\right) .$$

où *f* appartient à la distribution des données et *g* la distribution des paramètres. Dans la terminologie bayesienne $g\left(\theta/y\right)$ est la densité a posteriori. C'est-à-dire l'information concernant des paramètres après avoir observé les données.

$g\left(\theta\right)$ est la densité a priori des paramètres, c'est-à-dire l'information concernant les paramètres avant d'observer les données.

La densité a postériori peut être exprimée comme suit:

$$g(\theta/y) = f(y/\theta)\cdot g(\theta)/f(y).$$

Cette formule est appelée formule de bayes.

III-3-4-2-1. L'approche Bayesienne pour le modèle autorégressif du premier ordre

Pour utiliser cette approche, il est nécessaire de supposer une distribution fondamentale qui nous donne une bonne vraisemblance, et une densité à priorité des paramètres. Pour le modèle auto régressif du premier ordre :

$$y = \rho wy + e$$

L'approche usuelle est de supposer une loi normale pour le terme d'erreur. Par conséquent, la fonction de vraisemblance est comme suit :

$$f(y/\rho\sigma) = L(y/\rho\sigma) \propto |I-\rho w|\sigma^{-N}\exp\left\{-\left(2\sigma^2\right)^{-1} y'\left(I-\rho w\right)'\left(I-\rho w\right)y\right\}$$

Les deux paramètres dans le modèle sont des coefficients autorégressif ρ et de variance de l'erreur σ^2, suivant l'approche standard en économétries, les densités a priori de ces paramètres sont données par :

$$p(\sigma) \propto \sigma^{-1} \qquad 0 \langle \sigma \langle +\infty$$

$$p(\rho) \propto constante \qquad -1 \langle \rho \langle +1$$

On suppose de plus que ρ et σ sont a priori indépendantes, la densité conjointe a priori a la forme suivante :

$$p(\rho,\sigma) \propto \sigma^{-1}$$

La loi a posteriori conjointe de (ρ,σ) est donnée par l'application directe de la formule de Bayes.

$$g(\rho,\sigma/y) = p(\rho,\sigma/y) = L(y/\rho,\sigma)\cdot p(\rho,\sigma) \Big/ \int_{\Theta} L(y/\rho,\sigma)\cdot p(\rho,\sigma)\, d\rho d\sigma$$

$$\propto \quad |I-\rho w|\sigma^{-(N+1)}\exp\left\{-\left(2\sigma^2\right)^{-1} y'\left(I-\rho w\right)'\left(I-\rho w\right)y\right\}$$

La loi a posteriori marginal de ρ est obtenue par l'intégration par rapport à σ de la loi a posteriori conjointe. Alors :

$$p(\rho/y) = \int_0^{+\infty} p(\rho,\sigma/y)d\sigma = \int_0^{+\infty} |I-\rho w|\sigma^{-(N+1)} \exp\left\{-\left(2\sigma^2\right)^{-1} y'(I-\rho w)'(I-\rho w)y\right\}d\sigma$$

On a :

$$\Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx \qquad t\rangle 0 .$$

On pose :

$$x = -\left(2\sigma^2\right)^{-1} y'(I-\rho w)'(I-\rho w)y .$$

$$dx = \sigma^{-3} y'(I-\rho w)'(I-\rho w)y \;\; d\sigma .$$

$$p(\rho/y) = \int_0^{+\infty} p(\rho,\sigma/y)d\sigma = |I-\rho w|\left(y'(I-\rho w)'(I-\rho w)y\right)^{\frac{-N}{2}} \Gamma\left(\frac{N}{2}\right)\cdot 2^{\left(\frac{N}{2}-1\right)}$$

.

$$\propto \qquad |I-\rho w|\left(y'(I-\rho w)'(I-\rho w)y\right)^{\frac{-N}{2}} .$$

L'estimateur de Bayes minimise l'espérance à postériori de la fonction de perte proposée. Supposant que la fonction de perte est quadratique, c'est-à-dire

$$L(\rho,\hat{\rho}) = (\rho-\hat{\rho})^2$$

Avec $\hat{\rho}$ est l'estimateur de ρ .

L'estimateur de Bayes pour ρ est obtenu en minimisant :

$$\mathrm{E}_{\rho/y}\left(L(\rho,\hat{\rho})\right) = \int_{-1}^{1} (\rho-\hat{\rho})^2 \cdot p(\rho/y)\, d\rho$$

$$\frac{\partial}{\partial\hat{\rho}} \mathrm{E}_{\rho/y}\left(L(\rho,\hat{\rho})\right) = \frac{\partial}{\partial\hat{\rho}}\left[\int_{-1}^{1} (\rho-\hat{\rho})^2 \cdot p(\rho/y)\, d\rho\right]$$

$$\frac{\partial}{\partial\hat{\rho}}\mathrm{E}_{\rho/y}\left(L\left(\rho,\hat{\rho}\right)\right)=-2\int_{-1}^{1}\rho\cdot p\left(\rho/y\right)d\rho+2\hat{\rho}\int_{-1}^{1}p\left(\rho/y\right)d\rho$$

$$\frac{\partial}{\partial\hat{\rho}}\mathrm{E}_{\rho/y}\left(L\left(\rho,\hat{\rho}\right)\right)=0\Leftrightarrow\hat{\rho}=\int_{-1}^{1}\rho\cdot p\left(\rho/y\right)d\rho$$

Donc l’estimateur de Bayes est :

$$\hat{\rho}=\mathrm{E}\left(\rho/y\right)$$

CHAPITRE IV

LES TESTS

IV 1. Introduction

La modélisation des données spatialisées peut s'effectuer de différentes façon: on peut inclure des variables décalées (endogènes ou exogènes), une auto corrélation spatiale des erreurs ou estimer les modèles avec différentes matrices de poids. Le choix entre ces différentes alternatives passe par la mise en œuvre des tests de spécification.

Dans ce chapitre, nous verrons différents types de test, les trois grands principaux tests en économétrie standard: le test du multiplicateur de Lagrange (test du score de Rao), le test du ratio de vraisemblance et le test de Wald. Au début, la littérature en économétrie spatiale a été dominée par les deux derniers types de tests (Cliff et Ord 1981), ils contribuent toujours dans la recherche de la spécification du modèle bien qu'ils nécessitent l'estimation du modèle non-contraint qui doit être estimé par des méthodes non-linéaires. Au contraire, le test du score est basé uniquement sur le résultat du modèle. Sous l'hypothèse nulle et il s'agit la plupart du temps du modèle linéaire standard estimé par les MCO (Anselin, 1988). Ainsi que le test de Moran qui est le plus ancien et le plus utilisé pour test l'autocorrélation spatiale.

IV-2. Les tests d'hypothèses basées sur le principe du maximum de vraisemblance

Ces types de tests de spécification. Ils sont basés sur les propriétés optimales de l'estimateur de MV. Ces propriétés donnent normalité asymptotique pour ces estimateurs. Formellement, différence entre le paramètre estimé et la population converge en distribution vers une loi normale avec une moyenne de « 0 » et une variance correspondante à la matrice d'information inverse (Anselin, 1988, 1999).

C'est-à-dire :

$$N^{-1/2}(h-\theta) \to N\left(0, \lim_{N \to +\infty} \left(I(\theta)/N\right)^{-1}\right)$$

où θ est le paramètre estimé et h l'estimateur.

Ces tests d'hypothèses peuvent être formulés comme un test sur une fonction du paramètre du modèle.

C'est-à-dire : $\mathrm{H}_0 : g(\theta) = 0 vs \mathrm{H}_1 : g(\theta) \neq 0$

où : g est une fonction linéaire q-dimensionnelle ou une fonction matricielle non linéaire dont les éléments sont les vecteurs du paramètre θ.

En terme générale, les tests asymptotiques sont basés sur différentes mesures de distance entre les paramètres estimés sans restriction et les paramètres qui satisfont les contraintes impliquées par l'hypothèse nulle.

Les trois tests de spéciations qui détectent la présence des variables exogènes spatialement décalées sont le test de Wald, le test du rapport de vraisemblance et test de multiplicateur de Lagrange.

IV-2-1. Test de Wald

L'expression générale du test pour un modèle général est donnée par (Le gallo, 2000) :

$$W = g'\left[G'VG\right]^{-1} g$$

avec : g comme un vecteur de dimension q des valeurs des paramètres estimés par MV où q est le nombre de contraintes.

G matrice (3+k, q) de dériver partielle $\partial g'(\theta)/\partial\theta$ évaluée pour les paramètres estimés.

V est la matrice asymptotique estimée de dimension 3+k.

Le test de WALD est distribué selon une loi de χ^2 à q degrés de liberté, où q est le nombre de paramètres.

IV-2-2. Test du rapport de vraisemblance

Le test du rapport de vraisemblance est basé sur la différence entre log vraisemblance pour le modèle sans restriction (θ) et le modèle avec restriction (θ_R).

$$L_R = 2\left[L_{(\theta)} - L\left(\theta_R\right)\right].$$

L correspond à log vraisemblance. Le test de L_R est aussi asymptotiquement distribué par une χ^2 avec q degrés de liberté (Anselin, 1988).

IV-2-3. Le test du multiplicateur de Lagrange

Le test de LM est connu sous le nom du test du score, il est basé sur la première condition d'optimisation pour la fonction de Lagrange en log vraisemblance.

$$f = L(\theta) + \eta' g(\theta)$$

où f est le Lagrangien

L est le log vraisemblance

η est le vecteur du multiplicateur de Lagrange correspondant à q restrictions g(θ).

Ce test est exprimé seulement sur les coefficients restreints, il est basé sur l'estimation d'un modèle simple. Il est donné par :

$$L_M = d_{R'} \left(I_{\theta_R} \right)^{-1} d_R$$

Où

d_R le vecteur score qui est $\partial L / \partial \theta$ évalué sous l'hypothèse nulle.

$I(\theta_R)$ L'estimateur consistant pour la matrice d'information, sous l'hypothèse nulle.

La statistique du test LM est aussi asymptotiquement distribué par une χ^2 avec q degrés de liberté (Anselin, 1988, 2001).

IV-3. Autres tests

Dans le paragraphe précédent nous avons donnés des tests de spécification permettant de détecter une présence de variables exogènes spatialement décalées. D'autres tests permettent de détecter une omission de l'autocorrélation spatiale et la forme prise par cette dernière dans le modèle. Ce type de tests peut se diviser en plusieurs catégories. Premièrement, le test de Moran qui vise à tester l'autocorrélation spatiale résidu lorsque les erreurs suivent un processus autorégressifs ou moyen mobile. Deuxièmement, les tests du multiplicateur de Lagrange qui ont été développés et qui peuvent être soit unidirectionnels lorsqu'une hypothèse simple est testée, en supposant une spécification correcte pour le reste du modèle, soit multidirectionnels lorsque plusieurs types de dépendance spatiale sont testées.

IV-3-1. Test de Moran I

Le test I a été développé par Moran adopté aux résidus d'une régression et se présente formellement de la façon suivante en notation matricielle (Anselin et Florax, 1995).

$$I = N\left(\hat{\varepsilon}' w\hat{\varepsilon}\right) / \delta_0 \, \hat{\varepsilon}'\hat{\varepsilon}$$

Où :

$\hat{\varepsilon} = Y - X\hat{\beta}$ est le vecteur des résidus estimés par MCO.

$\hat{\beta} = \left((X'X)^{-1} X'Y \right)$ et δ_0 est un facteur de standardisation égale à la somme de tous les éléments de w. Cette statistique se simplifie pour une matrice standardisée $\delta_0 = N$.

Sous l'hypothèse nulle d'indépendance spatiale Cliff et Ord ont dérivé les deux premiers moments de I.

$$E(I) = tr(Mw) / N - K$$

et

$$V(I) = \frac{\left(tr\left(MwMw' \right) + tr(Mw)^2 + \{tr(Mw)\}^2 \right)}{(N-K)(N-K+2)} - \left(\mathrm{E}(I)\right)^2$$

où

$$M = I - X\left(X'X\right)^{-1} X' .$$

Le test se base sur la statistique de Moran centrée et réduite.

$$Z(I) = I - E(I) / V(I)$$

Pour des résidus normalement distribués et une matrice de poids « bien élevée », Z(I) suit asymptotiquement une loi normale centrée et réduite.

IV-3-2. Tests unidirectionnels

Ce test a été suggéré par Burridge (Anselin, 2001). Il considère tout d'abord le cas où les erreurs suivent un processus spatial autorégressif : c'est-à-dire $\varepsilon = \lambda w\varepsilon + \mu$ et sur lequel il test $H_0 : \lambda = 0$, sous l'hypothèse nulle, il retrouve le modèle linéaire classique.

La statistique du multiplicateur de Lagrange se calcule de la façon suivante :

$$LM_{ERR} = d'\left(\hat{\theta}\right) \cdot I\left(\hat{\theta}\right)^{-1} d\left(\hat{\theta}\right). \ .$$

Où :

$d(\theta)$, $I(\theta)$ le vecteur score et la matrice de l'information évolués sous l'hypothèse nulle.

Pour le modèle à erreurs auto régressive $\theta = \left(\beta', \sigma^2, \lambda\right)'$ la fonction log-vraisemblance :

$$\ln\left(l\left(y/\theta\right)\right) = -\frac{N}{2}\ln\left(2\Pi\right) - \frac{N}{2}\log(\sigma^{2}) + \ln\left|A\right| - \frac{1}{2\sigma^2}\varepsilon'\left(\mathrm{A'}A\right)^{-1}\varepsilon .$$

où : $\mathrm{A} = \mathrm{I} - \lambda w$.

$\varepsilon = Y - X\beta$.

Le vecteur score sous H_0

Les deux éléments du vecteur score sont :

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2} X'\left(A'A\right)^{-1}\varepsilon \ .$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{-N}{\sigma^2} + \frac{1}{2\sigma^4}\varepsilon'\left(A'A\right)^{-1}\varepsilon \ .$$

Et l'élément du vecteur score correspondant au paramètre λ est :

$$\frac{\partial l}{\partial \lambda} = -\frac{1}{\sigma^2}\varepsilon' A' w\left(A'A\right)^{-1}\varepsilon - Tr\left(A'w\right) \ .$$

Sous l'hypothèse nulle A=I, les deux premiers éléments produisent les estimateurs de MCO pour $\beta et \sigma^2$; l'élément score pour les paramètres autorégressifs spatial est :

$$\left.\frac{\partial l}{\partial \lambda}\right|_{H_0} = -\frac{1}{\hat{\sigma}^2} e' w e \ .$$

La matrice d'information sous H_0

La matrice d'information obtenue dans le chapitre précédent est donnée comme suit :

$$\left(I(\theta)\right)=\begin{bmatrix} \dfrac{X'\left(A'A\right)X}{\sigma^2} & 0 & 0 \\ 0 & \dfrac{N}{2\sigma^4} & \dfrac{Tr\left(w_A\right)}{\sigma^2} \\ 0 & \dfrac{Tr\left(w_A\right)}{\sigma^2} & Tr\left(w_A{}'w_A\right)+Tr\left(\left(w_A\right)^2\right) \end{bmatrix}$$

où

$$w_A = A^{-1}w$$

Sous l'hypothèse nulle, cette expression sera :

$$I\left(\hat{\theta}\right)\Big|_{H_0} = \begin{bmatrix} \dfrac{X'X}{\hat{\sigma}^2} & 0 & 0 \\ 0 & \dfrac{N}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & Tr\left(w\ 'w\ \right)+Tr\left(w^2\right) \end{bmatrix}$$

La statistique du multiplicateur de Lagrange est :

$$LM_{err} = d'\left(\hat{\theta}\right)\left(I\left(\hat{\theta}\right)\right)^{-1}d\left(\hat{\theta}\right).$$

$$LM_{err} = \left(0,.0,\frac{e'we}{\hat{\sigma}^2}\right)\begin{pmatrix} \dfrac{X'X}{\hat{\sigma}^2} & 0 & 0 \\ 0 & \dfrac{N}{2\hat{\sigma}^4} & 0 \\ 0 & 0 & Trw'w+Tr\left(w^2\right) \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 0 \\ \dfrac{e'we}{\hat{\sigma}^2} \end{pmatrix} = \left(\frac{e'we}{\hat{\sigma}^2}\right)^2 .\frac{1}{T}$$

où :

T = Tr (w'w) + Tr(w)² et $\hat{\sigma}^2$ sont les estimateurs obtenues sous l'hypothèse nulle puisqu'il n'y a qu'une seul constante. $LM_{err} \xrightarrow{D} \chi_1^2$ (Converge en distribution).

La statistique de test est la même si on spécifie comme hypothèse alternative le processus moyenne mobile et comme test $\delta = 0$. LM$_{err}$ est localement optimale pour les deux alternatives (autorégressive et moyenne mobile) et lorsque l'hypothèse nulle est rejetée. Le test ne donne pas d'indications quand à la nature du processus des erreurs (Anselin, 2001; Anselin et florax, 1995).

IV-3-2-1. Test d'une variable endogène décalée

Il a été proposé par Anselin (1988). Soit l'hypothèse nulle, $H_0 : \rho = 0$ dans le modèle $Y = \rho wY + X\beta + \varepsilon$ avec $\varepsilon \to N\left(0, I\sigma^2\right)$, en utilisant la fonction de vraisemblance du chapitre précédent.

$$Ln\left(L\left(Y/\theta\right)\right) = -\frac{N}{2}\log\left(2\pi\right) + \log\left|B\right| - \frac{N}{2}\log\left(\sigma^2\right)\frac{\left(Y - \rho wY - X\beta\right)'\left(Y - \rho wY - X\beta\right)}{2\sigma^2}$$

où :

$B = \left(I - \rho w\right)$

Le vecteur score sous H_0

Les éléments du vecteur score sont :

$$\frac{\partial Ln\left(L\right)}{\partial\beta} = \frac{1}{\sigma^2}X'\left(Y - \rho wY - X\beta\right)$$

$$\frac{\partial Ln\left(L\right)}{\partial\sigma^2} = -\frac{N}{\sigma^2} - \frac{1}{2\sigma^4}\left(Y - \rho wY - X\beta\right)'\left(Y - \rho wY - X\beta\right)$$

et

$$\frac{\partial Ln\left(L\right)}{\partial\rho} = \frac{\left(Y - \rho wY - X\beta\right)' Yw}{\sigma^2} - T_r\left(I - \rho w\right)^{-1} w.$$

Sous l'hypothèse nulle B=I, les deux premiers éléments produisent des estimateurs de MCO pour le dernier :

$$\left.\frac{\partial Ln(L)}{\partial \rho}\right|_{H_0} = \frac{e'Yw}{\hat{\sigma}^2}.$$

La matrice d'information sous H_0

La matrice d'information obtenue dans ce cas est :

$$(I(\theta)) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 & \frac{X'w_B X\beta}{\sigma^2} \\ 0 & \frac{N}{2\sigma^4} & \frac{Tr(w_B)}{\sigma^2} \\ \frac{(X'w_B X\beta)'}{\sigma^2} & \frac{Tr(w_B)}{\sigma^2} & Tr(w_B w_B') + Tr\left((w_B)^2\right) + \frac{(w_B X\beta)'(w_B X\beta)}{\sigma^2} \end{bmatrix}.$$

Sous l'hypothèse nulle :

$$\left.\left(I(\hat{\theta})\right)\right|_{H_0} = \begin{bmatrix} \frac{X'X}{\hat{\sigma}^2} & 0 & \frac{X'wX\beta}{\hat{\sigma}^2} \\ 0 & \frac{N}{2\hat{\sigma}^4} & 0 \\ \frac{(X'wX\beta)'}{\hat{\sigma}^2} & 0 & Tr(ww') + Tr\left((w)^2\right) + \frac{(wX\beta)'(wX\beta)}{\hat{\sigma}^2} \end{bmatrix}.$$

La statistique du multiplicateur de Lagrange est :

$$LM_{LAG} = \left(0, 0, \frac{e'wY}{\hat{\sigma}^2}\right) \begin{pmatrix} \frac{X'X}{\hat{\sigma}^2} & 0 & \frac{X'wX\beta}{\hat{\sigma}^2} \\ 0 & \frac{N}{2\hat{\sigma}^4} & 0 \\ \frac{(X'wX\beta)'}{\hat{\sigma}^2} & 0 & Trw'w + Tr\left(w^2\right) + (wX\beta)'(wX\beta)/\hat{\sigma}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ \frac{e'wY}{\hat{\sigma}^2} \end{pmatrix}.$$

On pose : $T = Tr(ww') + Tr(w)^2$.

$$\det\left(I\left(\hat{\theta}\right)\right) = \frac{N}{2\hat{\sigma}^8}\left[X'X\left(wX\hat{\beta}\right)'\left(wX\hat{\beta}\right) - \left(X'wX\hat{\beta}\right)'\left(X'wX\hat{\beta}\right) + X'XT\hat{\sigma}^2\right]$$

$$\left(I\left(\hat{\theta}\right)\right)^{-1} = \frac{1}{\det\left(I\left(\hat{\theta}\right)\right)} \times$$

$$\begin{bmatrix} \frac{N}{2\hat{\sigma}^6}\left[\left(wX\hat{\beta}\right)'\left(wX\hat{\beta}\right) + T\hat{\sigma}^2\right] & 0 & -\frac{N(X'wX\hat{\beta})}{2\hat{\sigma}^6} \\ 0 & \frac{X'X}{\hat{\sigma}^2}\left[\left(wX\hat{\beta}\right)'\left(wX\hat{\beta}\right)/\hat{\sigma}^2 - \left(X'wX\hat{\beta}\right)'\left(X'wX\hat{\beta}\right)/\hat{\sigma}^4 + T\right] & 0 \\ \frac{N\left(X'wX\hat{\beta}\right)'}{2\hat{\sigma}^6} & 0 & \left(X'X\right)/2\hat{\sigma}^6 \end{bmatrix}$$

$$LM_{LAG} = \left(e'wY/\hat{\sigma}^2\right)^2 / \hat{T}_1 .$$

où : $\hat{T}_1 = \dfrac{\left[\left(wX\hat{\beta}\right)'\left(I - X\left(X'X\right)^{-1}X'\right)\left(wX\hat{\beta}\right) + T\hat{\sigma}^2\right]}{\hat{\sigma}^2}$

Sous H_0 : $LM_{LAG} \to \chi_1^2$.

IV-3-3. Tests multidirectionnels

IV-3-3-1. Test en présence d'une autocorrélation des erreurs et d'une variable décalée

Anselin et Bera remarquent que LM_{ERR} est la statistique de test correspondant à $H_0 : \lambda = 0$, en supposant que $\rho = 0$. En revanche, si $\rho \neq 0$; c'est-à-dire que le modèle correct contient à la fois une autocorrélation des erreurs et une variable autorégressive, ce test n'est plus valide même asymptotiquement et il n'est plus distribué selon une χ^2 à un degré de liberté. Pour une inférence statistique valide, il est nécessaire de prendre en compte la possible variable endogène décalée lorsqu'on test l'autocorrélation spatiale des erreurs et vice versa. Plusieurs stratégies sont possibles pour ce problème. On peut effectuer un test joint de présence d'une variable décalée et d'une autocorrélation des erreurs, mais si l'hypothèse nulle est rejetée, on ne connaît pas la nature exacte de la dépendance spatiale. Une autre solution consiste à estimer un modèle avec une variable endogène décalée et tester ensuite s'il y a encore une autocorrélation des erreurs et vice versa. Dans ce cas, il faudrait estimer les modèles par le maximum de vraisemblance (Le gallo, 2001).

Anselin et al ont finalement proposé des tests basés sur les résidus des MCO dans le modèle simple mais qui sont capables de prendre en compte une autocorrélation des erreurs lorsqu'on test la présence d'une variable endogène décalée et vice versa (Le gallo, 2001).

Première approche : elle consiste à tester l'hypothèse nulle jointe $H_0 : \lambda = \rho = 0$, dans le modèle générale grâce au principe du multiplicateur de Lagrange. Ainsi le test peut être effectué à partir des résidus des MCO dans le modèle simple.

La fonction de vraisemblance pour le modèle générale est :

$$L(Y/\theta) = -\frac{n}{2}Ln(2\pi) - \frac{n}{2}Ln\sigma^2 + Ln|\beta| + Ln|A| - \frac{1}{2\sigma^2}(AY - X\beta)'(B'B)(AY - X\beta)$$

où :

$A = I - \rho w_1$. et $B = I - \lambda w_2$.

Le vecteur score sous H_0

Les deux premiers éléments du vecteur score $d'\theta|_{H_0} = \frac{\partial L}{\partial \theta}$ sont :

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\varepsilon'(B'B)X \ .$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon'(B'B)\varepsilon$$

où :

$$\varepsilon = (AY - X\beta) \ .$$

Les éléments du vecteur score correspondant aux paramètres $\rho et \lambda$ sont :

$$\frac{\partial L}{\partial \rho} = \frac{1}{\sigma^2} \varepsilon' \left(B'B\right) w_1 Y - tr\left(A^{-1} w_1\right).$$

$$\frac{\partial L}{\partial \lambda} = -Tr\left(B^{-1} w_2\right) + \frac{1}{\sigma^2} \varepsilon' B' w_2 \varepsilon .$$

Sous l'hypothèse nulle A=B=I.

Les deux premières éléments du vecteur score nous produisent des estimateurs de MCO pour β et σ^2. Pour les paramètres autorégressifs spatiaux, on a :

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = \frac{1}{\hat{\sigma}^2}\left(e' w_1 Y\right), \text{ car } Tr w_1 = 0 .$$

$$\left.\frac{\partial L}{\partial \lambda}\right|_{H_0} = \frac{1}{\hat{\sigma}^2} e' w_1 e , \ \text{ car } Tr w_1 = 0 .$$

La partie la plus compliquée est la matrice d'information $I_\theta = -E\left[\frac{\partial^2 L}{\partial \theta \, \partial \theta'}\right]$.

Comme l'on a déjà vu dans le chapitre précédent, on a besoin des dérivées secondes de certaines définitions et du théorème 1.

Rappel

Soit :

θ : Vecteur des paramètres.

w : Matrice de poids.

$A = I - \theta w$: une matrice.

$$\frac{\partial \theta w}{\partial \theta} = w . \qquad \frac{\partial Ln\left|A\right|}{\partial \theta} = trA^{-1}\frac{\partial A}{\partial \theta} = tr\left(A^{-1}\right)\left(-w\right) .$$

$$\frac{\partial A}{\partial \theta} = \frac{\partial I}{\partial \theta} - \frac{\partial \theta w}{\partial \theta} = -w . \qquad \frac{\partial A^{-1}}{\partial \theta} = -A^{-1}\left(\frac{\partial A}{\partial \theta}\right)A^{-1} = A^{-1} w A^{-1} .$$

$$\frac{\partial tr\left(A^{-1} w\right)}{\partial \theta} = tr\left(\frac{\partial A^{-1} w}{\partial \theta}\right) = tr\left(A^{-1} w A^{-1} w\right) .$$

Les dérivées secondes

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2} X'\left(B'B\right)X \quad .$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma} = -\frac{1}{\sigma^4} X'\left(B'B\right)\left(AY - XB\right) = \left[\frac{\partial^2 L}{\partial\sigma\partial\beta'}\right] \quad .$$

$$\frac{\partial^2 L}{\partial\beta\partial\rho} = -\frac{1}{\sigma^2} X'\left(B'B\right)w_1 Y = \left(\frac{\partial^2 L}{\partial\rho\partial\beta'}\right) \quad .$$

$$\frac{\partial^2 L}{\partial\beta\partial\lambda} = \left(\frac{\partial^2 L}{\partial\lambda\partial\beta^t}\right) = -\frac{1}{\sigma^2} X'B'w_2\varepsilon - \frac{1}{\sigma^2}\varepsilon'B'w_2 X \ .$$

$$\frac{\partial^2 L}{\partial^2\partial} = +\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon'\left(B'B\right)\varepsilon \quad .$$

$$\frac{\partial^2 L}{\partial\rho\partial\sigma^2} = \left(\frac{\partial^2 L}{\partial\sigma^2\partial\rho}\right) = -\frac{1}{\sigma^4}\varepsilon'\left(B'B\right)w_1 Y \ .$$

$$\frac{\partial^2 L}{\partial\lambda\partial\sigma^2} = \left(\frac{\partial^2 L}{\partial\sigma^2\partial\lambda}\right) = -\frac{1}{\sigma^4}\varepsilon'B'w_2\varepsilon \ .$$

$$\frac{\partial^2 L}{\partial^2\rho} = -\frac{1}{\sigma^2}\left(w_1 Y\right)'\left(B'B\right)w_1 Y - tr\left(A^{-1}w_1 A^{-1}w_1\right).$$

$$\frac{\partial^2 L}{\partial\rho\partial\lambda} = -\frac{1}{\sigma^2}\varepsilon'w_2'Bw_1 Y - \frac{1}{\sigma^2}\varepsilon'B'w_2 w_1 Y \ .$$

$$\frac{\partial^2 L}{\partial^2\lambda} = -\frac{1}{\sigma^2}\varepsilon'w_2'w_2\varepsilon - tr\left(B^{-1}w_2 B^{-1}w_2\right).$$

Résultats

Soit

$Y = A^{-1}X\beta + A^{-1}\varepsilon$

$\varepsilon = B^{-1}\mu$ où $E\left(\mu\right) = 0, V\left(\mu\right) = I\sigma^2$

Donc :

$$E(\varepsilon)=0$$

$$E(\varepsilon\varepsilon')=E\left(B^{-1}\mu\mu'\left(B^{-1}\right)'\right)=B^{-1}B'^{-1}E(\mu\mu')=\sigma^2(B'B)^{-1}$$

$$E(Y)=A^{-1}X\beta$$

$$E(YY')=\left(A^{-1}X\beta\right)\left(A^{-1}X\beta\right)'+\sigma^2A^{-1}(B'B)^{-1}(A')^{-1}$$

$$V(Y)=E(YY')-E(Y)\left(E(Y)\right)'=\sigma^2A^{-1}(B'B)^{-1}(A')^{-1}$$

La matrice d'information

Les éléments de la matrice d'information sont donnés par :

$$I(\theta)=-E\left[\frac{\partial^2 L}{\partial\theta\,\partial\theta'}\right],\qquad I_{\beta\beta'}=\frac{1}{\sigma^2}X'(B'B)X\ .$$

$$I_{\beta\sigma}=\frac{1}{\sigma^4}X'(B'B)E(AY-XB)=0\ Car\ E(\varepsilon)=0\ ,\qquad I_{\beta\rho}\frac{1}{\sigma^2}X'(B'B)w_1A^{-1}X\beta\ .$$

$$I_{\beta\lambda}=0\ Car\ E(\varepsilon)=0\,.$$

$$I_{\sigma\sigma}=-\frac{N}{2\sigma^4}+\frac{1}{\sigma^6}E(\varepsilon'(B'B)\varepsilon)=-\frac{N}{2\sigma^4}+\frac{1}{\sigma^6}Tr(B'B)E(\varepsilon\varepsilon')+\left(E(\varepsilon)\right)'B'BE(\varepsilon)=\frac{N}{2\sigma^4}\ .$$

$$I_{\sigma\rho}=\frac{1}{\sigma^4}E\left(\varepsilon'(B'B)w_1Y\right)=\frac{1}{\sigma^2}Tr\left(A^{-1}w_1\right).$$

$$I_{\sigma\lambda}=\frac{1}{\sigma^4}E\left(\varepsilon'B'w_2\varepsilon\right)=\frac{1}{\sigma^4}Tr\left(B'w_2E(\varepsilon\varepsilon')\right)+(E(\varepsilon))'B'w_2E(\varepsilon)=\frac{1}{\sigma^4}Tr\left(B^{-1}w_2\right).$$

$$I_{\rho\rho}=\frac{1}{\sigma^2}E\left(\left(w_1Y\right)'(B'B)w_1Y\right)+tr\left(A^{-1}w_1A^{-1}w_1\right)=$$

$$\frac{1}{\sigma^2}\left(\left(Bw_1A^{-1}X\beta\right)'\left(Bw_1A^{-1}X\beta\right)\right)+Tr\left(\left(Bw_1A^{-1}B^{-1}\right)'\left(Bw_1A^{-1}B^{-1}\right)\right)+tr\left(A^{-1}w_1A^{-1}w_1\right).$$

$$I_{\rho\lambda}=\frac{1}{\sigma^2}E\left(\varepsilon'w_2'Bw_1Y\right)+\frac{1}{\sigma^2}E\left(\varepsilon'B'w_2w_1Y\right)=Tr\left(\left(w_2B^{-1}\right)Bw_1A^{-1}B^{-1}\right)+Tr\left(w_2w_1A^{-1}B^{-1}\right)$$

$$I_{\lambda\lambda}=\frac{1}{\sigma^2}E\left(\varepsilon'w_2'w_2\varepsilon\right)+Tr\left(B^{-1}w_2B^{-1}w_2\right)=Tr\left(B^{-1}w_2B^{-1}w_2\right)+Tr\left(\left(B^{-1}w_2\right)'B^{-1}w_2\right).$$

Sous H_0 ces expressions seront :

$$I_{\hat{\beta}\hat{\beta}'}\Big|_{H_0} = \frac{1}{\sigma^2} X'X \; , \qquad I_{\hat{\beta}\hat{\sigma}}\Big|_{H_0} = 0 \; , \qquad I_{\hat{\beta}\hat{\rho}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2} X' w_1 X \hat{\beta} \; , \qquad I_{\hat{\beta}\hat{\lambda}}\Big|_{H_0} = 0 .$$

$$I_{\hat{\sigma}\hat{\sigma}}\Big|_{H_0} = \frac{N}{2\sigma^4} \; , \qquad I_{\hat{\sigma}\hat{\lambda}}\Big|_{H_0} = \frac{1}{\sigma^2} Tr\left(w_2\right) = 0 , \qquad I_{\hat{\sigma}\hat{\rho}}\Big|_{H_0} = \frac{1}{\sigma^2} Tr\left(w_1\right) = 0 .$$

$$I_{\hat{\rho}\hat{\rho}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2}\left(\left(w_1 X\hat{\beta}\right)'\left(w_1 X\hat{\beta}\right)\right) + Tr\left(\left(w_1\right)'\left(w_1\right)\right) + tr\left(w_1 w_1\right)$$

$$I_{\hat{\rho}\hat{\lambda}}\Big|_{H_0} = Tr\left(\left(w_2\right)' w_1\right) + Tr\left(w_2 w_1\right), \qquad I_{\hat{\lambda}\hat{\lambda}}\Big|_{H_0} = Tr\left(w_2 w_2\right) + Tr\left(\left(w_2\right)' w_2\right).$$

Après avoir évalué tous les éléments du vecteur score et la matrice d'informations sous l'hypothèse nulle $H_0 : \lambda = \rho = 0$, la statistique qui en résulte est la suivante:

$$SARMA = d'\left(\hat{\theta}\right) I\left(\hat{\theta}\right) d\left(\hat{\theta}\right)$$

où

$$\hat{\theta} = \left(\hat{\beta}', \hat{\sigma}^2, \hat{\lambda}, \hat{\rho}\right).$$

$SARMA =$

$$\begin{pmatrix} 0 & 0 & e' w_1 e / \hat{\sigma}^2 & e' w_1 Y / \hat{\sigma}^2 \end{pmatrix} \begin{pmatrix} \dfrac{X'X}{\hat{\sigma}^2} & 0 & 0 & \dfrac{\left(X' w_1 X\hat{\beta}\right)}{\hat{\sigma}^2} \\ 0 & \dfrac{N}{2\hat{\sigma}^4} & 0 & 0 \\ 0 & 0 & T_{22} & T_{21} \\ \dfrac{\left(X' w_1 X\hat{\beta}\right)'}{\hat{\sigma}^2} & 0 & T_{21} & T_{11} + \dfrac{\left(w_1 X\hat{\beta}\right)'\left(w_1 X\hat{\beta}\right)}{\hat{\sigma}^2} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ e' w_1 e / \hat{\sigma}^2 \\ e' w_1 Y / \hat{\sigma}^2 \end{pmatrix}$$

où :

$$T_{ij} = Tr\left(w_i w_j\right) + Tr\left(w_i' w_j\right)$$

$$\left(I\left(\hat{\theta}\right)\right)^{-1} = \frac{1}{\det\left(I\left(\hat{\theta}\right)\right)} \bullet$$

$$\begin{pmatrix} \frac{N}{2\hat{\sigma}^4}\left[T_{22}\left(T_{11}+\frac{\left(w_1X\hat{\beta}\right)'\left(w_1X\hat{\beta}\right)}{\hat{\sigma}^2}\right)-T_{21}{}^2\right] & 0 & -\frac{N\left(X'w_1X\hat{\beta}\right)T_{21}}{2\hat{\sigma}^6} & -\frac{N\left(X'w_1X\hat{\beta}\right)T_{22}}{2\hat{\sigma}^6} \\ 0 & \begin{matrix}\frac{X'X}{2\hat{\sigma}^2}\left[T_{22}\left(T_{11}+\frac{\left(w_1X\hat{\beta}\right)'\left(w_1X\hat{\beta}\right)}{\hat{\sigma}^2}\right)-T_{21}{}^2\right] \\ -\frac{T_{22}\left(X'w_1X\hat{\beta}\right)'\left(X'wX\hat{\beta}\right)}{2\hat{\sigma}^4}\end{matrix} & 0 & 0 \\ \frac{-N\left(X'w_1X\hat{\beta}\right)'T_{21}}{2\hat{\sigma}^6} & 0 & \begin{matrix}\frac{NX'X}{2\hat{\sigma}^6}\left[T_{11}+\frac{\left(w_1X\hat{\beta}\right)'\left(w_1X\hat{\beta}\right)}{\hat{\sigma}^2}\right] \\ -\frac{N\left(X'w_1X\hat{\beta}\right)'\left(X'wX\hat{\beta}\right)}{2\hat{\sigma}^8}\end{matrix} & -\frac{NX'X}{2\hat{\sigma}^6}T_{21} \\ -\frac{N\left(X'w_1X\hat{\beta}\right)'T_{22}}{2\hat{\sigma}^6} & 0 & -\frac{NX'X}{2\hat{\sigma}^6}T_{21} & \frac{NX'X}{2\hat{\sigma}^6}T_{22} \end{pmatrix}$$

$$\det\left(I\left(\hat{\theta}\right)\right)=\frac{NX'X}{2\hat{\sigma}^8}\left[\left(\left(w_1X\hat{\beta}\right)'M\left(w_1X\hat{\beta}\right)+\hat{\sigma}^2T_{11}\right)T_{22}-\hat{\sigma}^2\left(T_{21}\right)^2\right].$$

On pose :

$$\hat{D}_1=\left(w_1X\hat{\beta}\right)'M\left(w_1X\hat{\beta}\right)+\hat{\sigma}^2T_{11} \text{ et } \hat{D}=\hat{D}_1T_{22}-\hat{\sigma}^2\left(T_{21}\right)^2/\hat{\sigma}^2 .$$

où :

$$M=I-X\left(X'X\right)^{-1}X'.$$

Alors

$$SARMA=\begin{pmatrix} \frac{e'w_1e\left(X'w_1X\hat{\beta}\right)'\left(T_{21}+T_{22}\right)}{\hat{\sigma}^2X'X\hat{D}} & 0 & \frac{e'w_1e*\hat{D}_1/\hat{\sigma}^2{}^2}{\hat{\sigma}^2\hat{D}}-\frac{e'w_1YT_{21}}{\hat{\sigma}^2\hat{D}} & \frac{-e'w_1eT_{21}}{\hat{\sigma}^2\hat{D}}+\frac{e'w_1YT_{22}}{\hat{\sigma}^2\hat{D}} \end{pmatrix}*\begin{pmatrix} 0 \\ 0 \\ e'w_1e/\hat{\sigma}^2 \\ e'w_1Y/\hat{\sigma}^2 \end{pmatrix}$$

$$=\left[\left(\frac{e'w_1e}{\hat{\sigma}^2}\right)^2\frac{\hat{D}_1}{\hat{\sigma}^2}+\left(\frac{e'w_1Y}{\hat{\sigma}^2}\right)^2T_{22}-2\left(\frac{e'w_1e}{\hat{\sigma}^2}\right)\left(\frac{e'w_1Y}{\hat{\sigma}^2}\right)T_{21}\right]\bullet\frac{1}{\hat{D}}$$

Si $w_1=w_2=w$ alors $T_{11}=T_{21}=T_{22}=T=Tr\left(ww\right)+Tr\left(w'w\right)$.

$$SARMA=\left[\left(\frac{e'w_1e}{\hat{\sigma}^2}\right)^2\Big/T+\left(\left(\frac{e'w_1e}{\hat{\sigma}^2}\right)-\left(\frac{e'w_1Y}{\hat{\sigma}^2}\right)\right)^2\Big/\sigma^{-2}\left(D_1-T\sigma^2\right)\right]$$

Sous $H_0:\lambda=\rho=0$, SARMA converge vers un χ^2 à deux degrés de libertés.

La deuxième approche consiste à faire un test du multiplicateur de Lagrange pour une forme de dépendance spatiale, lorsque l'autre forme n'est pas constante. Cela consiste à tester l'hypothèse nulle : $H_0 : \lambda = 0$ en présence de ρ.

C'est-à-dire sous l'hypothèse nulle, on retrouve le module autorégressif. Alors que, sous l'hypothèse alternative, on retrouve le modèle général.

Le test est alors basé sur les résidus de l'estimation par le maximum de vraisemblance pour le modèle autorégressif.

On garde la même fonction de vraisemblance pour le modèle général mais on a une seule contrainte qui est λ.

Donc les premiers éléments du vecteur score $\partial\theta = \frac{\partial L}{\partial \theta}$ sont :

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\varepsilon' B' B X$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon' B' B \varepsilon$$

$$\frac{\partial L}{\partial \rho} = \frac{1}{\sigma^2}\varepsilon'\left(B'B\right) w_1 Y - tr\left(A^{-1} w_1\right)$$

où :

$$\varepsilon = AY - XB$$

L'élément du vecteur score correspondant au paramètre λ est:

$$\frac{\partial L}{\partial \lambda} = \frac{1}{\sigma^2}\varepsilon' B' w_2 \varepsilon - tr\left(B^{-1} w_2\right).$$

Sous l'hypothèse nulle, B = I les trois premiers éléments du vecteur score produisent les estimateurs de MV pour $\beta, \hat{\sigma}, \rho$, avec e comme résidus de l'estimation par MV.

Pour le paramètre λ

$$\left.\frac{\partial L}{\partial \lambda}\right|_{H_0} = \frac{1}{\hat{\sigma}^2} e' w_2 e \,.$$

Pour la matrice d'information, on garde la même matrice que l'approche précédente sauf qu'il ne faut pas oublier que la contrainte est λ. Sous l'hypothèse nulle on trouve.

$$I_{\hat{\beta}\hat{\beta}'}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2} X'X$$

$$I_{\hat{\beta}\hat{\sigma}}\Big|_{H_0} = 0$$

$$I_{\hat{\beta}\hat{\rho}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2} X' w_1 \hat{A}^{-1} X \hat{\beta}$$

$$I_{\hat{\beta}\hat{\lambda}}\Big|_{H_0} = 0$$

$$I_{\hat{\sigma}\hat{\sigma}}\Big|_{H_0} = \frac{N}{2\hat{\sigma}^4}$$

$$I_{\hat{\sigma}\hat{\lambda}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2} Tr\left(w_2\right) = 0$$

$$I_{\hat{\sigma}\hat{\rho}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2} Tr\left(\hat{A}^{-1} w_1\right) = 0$$

$$I_{\hat{\rho}\hat{\rho}}\Big|_{H_0} = \frac{1}{\hat{\sigma}^2}\left(\left(w_1\hat{A}^{-1}X\hat{\beta}\right)^t\left(w_1\hat{A}^{-1}X\hat{\beta}\right)\right) + Tr\left(\left(w_1\hat{A}^{-1}\right)^t\left(w_1\hat{A}^{-1}\right)\right) + tr\left(\hat{A}^{-1}w_1\right)^2$$

$$I_{\hat{\rho}\hat{\lambda}}\Big|_{H_0} = Tr\left(w_2{}^t w_1 \hat{A}^{-1}\right) + Tr\left(w_2 w_1 \hat{A}^{-1}\right)$$

$$I_{\hat{\lambda}\hat{\lambda}}\Big|_{H_0} = Tr\left(w_2\right)^2 + Tr\left(w_2{}' w_2\right)$$

où $\hat{\rho}$ est l'estimateur de ρ évalué par MV dans le modèle autorégressif spatial obtenu par optimisation non linéaire. La statistique est

$$LM^*_{err} = d'\left(\hat{\theta}\right) I\left(\hat{\theta}\right) d\left(\hat{\theta}\right)$$

$$où$$

$$\hat{\theta} = \left(\hat{\beta}', \hat{\sigma}^2, \hat{\rho}, \hat{\lambda}\right).$$

LM^*_{err} =

$$\begin{pmatrix}0 & 0 & 0 & e'w_2e/\hat{\sigma}^2\end{pmatrix}\begin{pmatrix}\frac{X'X}{\hat{\sigma}^2} & 0 & \frac{\left(X'w_1\hat{A}^{-1}X\hat{\beta}\right)}{\hat{\sigma}^2} & 0\\ 0 & \frac{N}{2\hat{\sigma}^4} & \frac{1}{\hat{\sigma}^2}Tr\left(\hat{A}^{-1}w_1\right) & 0\\ \frac{\left(X'w_1\hat{A}^{-1}X\hat{\beta}\right)'}{\hat{\sigma}^2} & \frac{1}{\hat{\sigma}^2}Tr\left(\hat{A}^{-1}w_1\right) & T_{11A}+\frac{\left(w_2\hat{A}^{-1}X\hat{\beta}\right)'\left(w_2\hat{A}^{-1}X\hat{\beta}\right)}{\hat{\sigma}^2} & T_{21A}\\ 0 & 0 & T_{21A} & T_{22}\end{pmatrix}^{-1}\begin{pmatrix}0\\0\\0\\e'w_2e/\hat{\sigma}^2\end{pmatrix}$$

$$\det\left(I\left(\hat{\theta}\right)\right)=\frac{NX'X}{2\hat{\sigma}^8}\left[\hat{D}_1T_{22}-\frac{2\hat{\sigma}^2}{N}\left(Tr\left(\hat{A}^{-1}\right)\right)^2T_{22}-\hat{\sigma}^2\left(T_{21A}\right)^2\right]$$

où:

$$\hat{D}_1=\left(w_1\hat{A}^{-1}X\hat{\beta}\right)'M\left(w_1\hat{A}^{-1}X\hat{\beta}\right)+\hat{\sigma}^2T_{11}\quad \text{et}\quad M=I-X\left(X'X\right)^{-1}X'.$$

Alors

$LM^*_{err}=$

$$\begin{pmatrix}\frac{Ne'w_2e\left(X'w_1\hat{A}^{-1}X\hat{\beta}\right)'}{2\hat{\sigma}^8\det\left(I\left(\hat{\theta}\right)\right)} & \frac{e'w_2eT_{21A}Tr\left(\hat{A}^{-1}w_1\right)X'X}{\hat{\sigma}^6\det\left(I\left(\hat{\theta}\right)\right)} & \frac{e'w_2eT_{21A}NX'X}{\hat{\sigma}^8} & \frac{X'XN\left(\hat{D}_1-2\hat{\sigma}^2\left(Tr\hat{A}^{-1}w_1\right)^2/N\right)}{\hat{\sigma}^6\det\left(I\left(\hat{\theta}\right)\right)}\end{pmatrix}*$$

$$\begin{pmatrix}0\\0\\0\\e'w_2e/\hat{\sigma}^2\end{pmatrix}$$

$$=\left(\frac{e'w_1e}{\hat{\sigma}^2}\right)^2\frac{1}{T_{22}-\left(T_{21A}\right)^2F}$$

où :

$$F=\frac{\hat{\sigma}^2}{\hat{D}_1-2\hat{\sigma}^2\left(Tr\left(\hat{A}^{-1}w_1\right)\right)^2/N}.$$

Sous $H_0:\lambda=0$, LM^*_{err} converge en distribution vers un χ^2 à un degré de liberté.

On peut également tester l'hypothèse nulle $H_0:\rho=0$. En présence de λ, même principe et mêmes étapes, c'est-à-dire le test est basé sur les résidus de l'estimation par le MV dans le modèle avec une autocorrélation des erreurs, la statistique est :

$$LM^*_{LAG}=\frac{\left[e'\hat{B}'\hat{B}w_1Y/\hat{\sigma}^2\right]^2}{H_\rho-H_{\theta\rho}\hat{V}\left(\hat{\theta}\right)H'_{\theta\rho}}$$

où :

e est le vecteur des résidus estimés par le maximum de vraisemblance dans le modèle avec erreur autorégressif.

$$\hat{B} = I - \hat{\lambda} w_2$$

$$H_\rho = Tr\left(w_1\right)^2 + Tr\left(\hat{B} w_1 \hat{B}^{-1}\right)' \left(\hat{B} w_1 \hat{B}^{-1}\right) + \frac{\left(\hat{B} w_1 X \hat{\beta}\right)' \left(\hat{B} w_1 X \hat{\beta}\right)}{\hat{\sigma}^2}$$

$$H'_{\theta\rho} = \begin{pmatrix} \dfrac{\left(\hat{B} X\right)' \hat{B} w_1 X \hat{\beta}}{\hat{\sigma}^2} \\ Tr\left(w_2 \hat{B}^{-1}\right)' \hat{B} w_1 \hat{B}^{-1} + Tr w_2 w_1 \hat{B}^{-1} \\ 0 \\ \\ \end{pmatrix}$$

$\hat{V}\left(\hat{\theta}\right)$ est la matrice des variances covariances estimée par de $\hat{\theta}$ dans le modèle avec erreur autorégressive. Sous $H_0 : LM^*_{LAG}$ converge, en distribution vers un χ^2 à un degré de liberté.

Une dernière approche de ces tests nécessite uniquement une estimation du modèle simple classique par les MCO. Cette approche est celle de Bera, Escudero et Yoon qui a été reprise par Anselin et al (Bera, Escudero et Yoon, 2000). Elle consiste à utiliser des tests robustes pour une mauvaise spécification locale. Par exemple, on ajuste LM_{err} pour que sa distribution asymptotique reste un χ^2 central, même en présence locale de ρ.

La statistique modifiée pour, le test de $H_0 : \lambda = 0$ est:

$$RLM_{err} = \left(\frac{e' w_1 e}{\hat{\sigma}^2} - T_{12} \hat{\sigma}^2 \hat{D}_1^{-1} \left(\frac{e' w_1 Y}{\hat{\sigma}^2}\right)\right)^2 \Big/ T_{22} - \left(T_{12}\right)^2 \hat{\sigma}^2 \hat{D}_1^{-1} .$$

Si $w_1 = w_2 = w$ alors $T_{11} = T_{21} = T_{22} = T = Tr\left(ww\right) + Tr\left(w' w\right)$.

$$RLM_{err} = \left(\frac{e' w_1 e}{\hat{\sigma}^2} - T \hat{\sigma}^2 \hat{D}_1^{-1} \left(\frac{e' w_1 Y}{\hat{\sigma}^2}\right)\right)^2 \Big/ T - T^2 \hat{\sigma}^2 \hat{D}_1^{-1}$$

Même chose pour le test LM_{LAG} en pr ésence locale de λ ; la statistique de ce test est :

$$RLM_{LAG} = \left(\left(\frac{e'w_1Y}{\hat{\sigma}^2} \right) - T_{12}T_{22}^{-1} \frac{e'w_1e}{\hat{\sigma}^2} \right)^2 \Big/ \hat{\sigma}^2 \hat{D}_1 - T_{22}^{-1} (T_{12})^2 .$$

Si $w_1 = w_2 = w$ alors $T_{11} = T_{21} = T_{22} = T = Tr(ww) + Tr(w'w)$.

$$RLM_{LAG} = \left(\frac{e'w_1Y}{\hat{\sigma}^2} - \frac{e'w_1e}{\hat{\sigma}^2} \right)^2 \Big/ \hat{\sigma}^{-2} \hat{D}_1 - T$$

IV-4. Puissance et robustesse des tests

Des simulations récentes de Monte Carlo effectuées en particulier par Anselin et Rey (1991), Anselin et Florax (1995) fournissent quelques indications quant aux performances de ces tests asymptotiques en échantillon fini.

Anselin et Rey (1991) ainsi que Anselin et Florax (1995) ont comparés les performances des tests de Moran, LMerr, LMlag , RLMerr et RLMlag pour différentes matrices de poids (sur zonage régulier ou non), différentes distributions des erreurs et différentes tailles d'échantillon.

Plusieurs résultats ressortent :

- D'une manière générale, les puissances des tests diminuent dans les petits échantillons et augmentent avec des valeurs plus fortes des paramètres spatiaux, LMlag étant le test le plus puissant et robuste à la non-normalité des erreurs.
- Le test I de Moran apparaît puissant pour les deux alternatives, variable endogène décalée ou autocorrélation des erreurs. Ce test devient ainsi un indicateur général d'une mauvaise spécification du modèle, quelque soit la forme de la dépendance spatiale omise.
- Les tests du score LMerr et LMlag ont les plus grades puissances, pour leur alternative respective.
- Les tests ajustés RLMerr et RLMlag ont également de bonnes performances en termes de puissance et de taille empiriques.

IV-5. Les regèle de décisions

Les diffèrent tests de spécification peuvent être combinées à fin de choisir la meilleur spécification du modèle, Anselin et Florax proposent cette règle de décision afin de spécifier le modèle (Le gallo, 2001).

1/ La première étape consiste à estimer le modèle simple par les MCO et effectuer le test de Moan et le test de SAMRA (test joint d'une présence de présence d'autocorrélation des erreurs et d'une variable autorégressive). Le rejet de l'hypothèse nulle indique une mauvaise spécification du modèle et une omission de l'autocorrélation spatiale. Les tests LM_{ERR}, LM_{LAG} et leurs versions robustes permettent de spécifier la forme de l'autocorrélation spatiale.

2/ Si les résultats des tests indiquent une présence de dépendance spatiale, il est souvent utile de commencer par inclure dans le modèle, si possible, des variables supplémentaires. Il peut s'agir de variables exogènes supplémentaires qui sont susceptibles d'éliminer la dépendance spatiale (si cette dernière provient d'une mauvaise spécification) ou des variables exogènes décalées spatialement.

3/ si l'ajout des variables exogènes supplémentaires n'a pas éliminé l'autocorrélation spatiale, il faut alors estimer un modèle incorporant une variable autorégressive ou une autocorrélation des erreurs. Le choix entre ces deux formes de la dépendance spatiale en comparant les niveaux de significativité des tests du multiplicateur de Lagrange s'effectue selon les valeurs relatives des tests du multiplicateur de Lagrange LM_{ERR}, LM_{LAG} et leurs versions robustes.

* si l'on considère que LM_{LAG} et LM_{ERR}, Anselin et Rey (1991) proposent de choisir l'une ou l'autre forme fonctionnelle en appliquant la règle de décision simple suivantes :
- Si le test du modèle autorégressif n'aboutit pas au rejet de l'hypothèse nulle ou si les deux tests aboutissent au rejet de l'hypothèse nulle et que le deuxième test et plus significatif que le premier, on choisit le modèle avec autocorrélation des erreurs.
- Si le test du modèle autorégressif aboutit au rejet de l'hypothèse nulle ou si les deux tests aboutissent au rejet de l'hypothèse nulle et que le premier test et plus significatif que le test de l'autocorrélation des erreurs, on choisit le modèle autorégressif.
*Anselin et Florax (1995) affinent cette règle de décision en la complétant par l'utilisation des tests robustes : si LM_{LAG} est plus significatif que LM_{ERR} et RLM_{LAG} le sont mais RLM_{ERR} ne l'est pas, on inclut une variable endogène décalée. Une façon similaire, la présence d'une autocorrélation des erreurs peut être identifié à travers RLM_{ERR}. Dans ce dernier, le test du

facteur commun doit encore être effectué pour vérifier que le modèle avec autocorrélation et le meilleur.

4/ une fois que le modèle spatiale adéquat ait été estimé, trois test supplémentaires peuvent être mobilisés.

* Pour un modèle autorégressif, le test LM^*_{ERR} permet de savoir si une corrélation spatiale des erreurs et encore nécessaire.

* Pour un modèle avec autocorrélation des erreurs, le test LM^*_{LAG} permet de savoir si une variable endogène décalé est encore nécessaire. Le test du facteur commun indique si la restriction $\delta+\lambda\beta=0$ peut être ajoutée ou non. Si elle ne l'est pas, le modèle se réduit au modèle avec corrélation des erreurs.

5/ si plusieurs restent encore en compétition, le test de J sert à comparer les modèles spatiaux comportant des matrices de poids différentes. Le choix entre les modèles peut aussi se faire avec les critères traditionnels tels que les critères d'information:

$$\mathrm{INF}=-2\ln L+q(K)$$

$\ln L$ est la valeur de la fonction de logvraisemblance à l'optimum, K est le nombre de paramètres inconnus et q un facteur de correction qui varie selon les formulations : $q=2K$ pour le critère d'Akaïke ou $q=\log NK$ pour le critère de Schwartz. Lorsqu'on compare deux modèles selon leurs critères d'information, on choisit celui qui minimise ce coefficient.

CHAPITRE V

SIMULATION

V-1. Introduction

Nous avons vu dans les chapitres précédents, que l'autocorrélation spatiale se modélise grâce aux matrices de poids. A travers les règles de décision, les tests de spécification permettent de déterminer la forme prise par l'autocorrélation spatiale et son interprétation. Comme nous l'avons déjà souligné, l'une des raisons à l'attention portée sur les effets spatiaux est l'apparition des logiciels qui proposent l'estimation des principaux modèles spatiaux, tels que SpaceStat (Anselin, 1999), GeoDa (Anselin, 2006), logiciels R (Bivand, 2002) ou la librairie pour Matlab (Le sage, 2005). Vue la disponibilité de la librairie Matlab, nous l'avons choisi pour simuler des données à partir d'un modèle spatiale, dont nous effectuons des tests de l'autocorrélation spatiale, à fin de détecter la dépendance spatiale et spécifier la forme du modèle, puis estimer ces coefficients en utilisant les toolbox proposer par le sage (2005), nous concluons par l'interprétation des résultats.

Le Matlab est constitué d'un noyau relativement réduit, capable d'interpréter puis d'évaluer les expressions numériques matricielles qui lui sont adressées. Ce noyau est complété par une bibliothèque de fonctions prédéfinies, très souvent sous forme de fichiers m-files et regroupés en paquetages ou toolbox. A coté des toolbox requises local, il est possible d'ajouter des toolbox spécifiques à tel ou tel problème mathématique.

V-2. Simulation des données spatiale

Nous avons développé un programme sous le Matlab pour une simulation des données suivant un modèle autorégressif spatial, donné par :

$$Y = \left(I - \rho w\right)^{-1} X\beta + \left(I - \rho w\right)^{-1} \mathrm{E} \;\ldots\ldots\ldots\ldots(1)$$

où X est une matrice de (N,k) qui représente les observations des variables exogènes, avec N la taille de l'échantillon et k le nombre de paramètres. Ces variables explicatives peuvent être par exemple dans le domaine météorologique la température, pression ou précipitation des pluies, ou dans le domaine économique peut être la consommation, investissement, etc… . Ces variables peuvent être générées par des lois de probabilité, ou par des modèles déjà connus, dans notre cas nous avons choisi le modèle suivant, où i représente le $i^{ième}$ site et j représente la $j^{ième}$ variable :

$$X\left(i, j\right) = X\left(i-1, j\right) A(j) + \mathrm{E1}(j) \;\; ; \; i = 1,\ldots., N \;\text{ et }\; j = 1,\ldots., k$$

Tel que $X(i-1, j)$, $E1(j)$ suivent la loi normale (0,1) et $A(j)$ est générée par une loi uniforme [-1,1]. Si nous déroulons juste cette partie du programme, nous aurons une matrice dont les lignes représentent les sites et les colonnes représentent les variables.

Par exemple : $X(2,3)$ est la valeur de la troisième variable par rapport au deuxième site.

Donc $X(i, j)$ représente la valeur de la $j^{ème}$ variable par rapport au $i^{ème}$ site.

Revenons au modèle initial :

E : est le vecteur des erreurs qui suit la loi normale (0,1).

β : est le vecteur des paramètres, qui est générée par la loi uniforme [-1,1].

Apres avoir déterminé et généré les parties classiques du modèle (1); nous entamons la partie spatiale par le paramétrer ρ, qui est le paramètre spatial autorégressif indiquant l'ampleur de l'interaction existant entre les données de Y. Nous avons vu dans le chapitre III que les conditions de régularité de logvraisemblance sont prises sur les valeurs des coefficients spatiaux; dans notre cas il faut que $\rho \in \left]1/\omega_{\min}, 1/\omega_{\max}\right[$, tel que $\omega_{\min}$ est la valeur propre négative la plus grande en valeur absolue de w (matrice de séparation spatiale) et $\omega_{\max}$ est la valeur propre positive maximale. Pour cela nous avons programmé cette condition où ρ est généré par la loi uniforme [-1,1].

V-2-1. Simulation de la matrice de poids spatiale

Comme nous remarquons dans le modèle (1), w qui est la matrice de poids spatiale permet de modéliser la structure spatiale des données. Pour cela, nous avons déroulé un programme pour générer des matrices de séparation spatiale à différent ordre; en se basant sur la méthode du pointeur (vue en chapitre II). Dans notre exemple, nous avons supposés que cette matrice est définie à partir des dispositions géographiques, où chaque site i est reconnu par deux coordonnées $x(i)$ *et* $y(i)$ que nous pourrons considérer comme par exemple une altitude et longitude et qu'elles suivent une loi uniforme [100,3000].quoique cette matrice en réalité peut être définie à partir des dispositions géographiques et les relations existantes entre observations. La simulation de ces coordonnées va nous permettre de calculer les différentes distances existantes entre les différents sites. Cette distance est donnée par :

$$d(i, j) = \sqrt{(x(i) - x(j))^2 - (y(i) - y(j))^2}$$

où i représente le $i^{ème}$ site et j représente le $j^{ème}$ site.

En exécutant ce programme, nous obtenons une matrice de distance. D'après la méthode du pointeur, pour calculer les paramètres de séparations spatiale qui construisent les différents voisinages autour de chaque point, nous devons ordonner ces distances suivant les lignes et retirer le min «strictement positif » de chaque ligne (la plus petite distance de chaque ligne), puis retrouver le max de ces min. Ce max est le premier paramètre de séparation spatiale α_1.

Pour avoir le deuxième paramètre de séparation, nous retranchons de la matrice de distance le α_1 et nous aurons une autre matrice avec des valeurs négatives, nulles ou positives. Le même principe qui se répète, nous ordonnons la matrice suivant les lignes et nous retirons pour chaque ligne la plus petite valeur, toujours strictement positive, puis nous cherchons le max de ces valeurs. Ce max sera le deuxième paramètre de séparation spatiale.

Nous répétons ce procédé jusqu'à ce que nous trouvons un entier *m* tel que :

$$\sum_{i=1}^{m} \alpha_i = \max_{1\le i\le n}\left[\max_{1\le j\le n}(d_{ij})\right]$$

De cette manière, nous retrouvons tous les paramètres de séparation spatiale. A partir de ces paramètres, nous programmons le voisinage spatial de chaque site par :

$$a_0 = 0, \qquad a_h = \sum_{i=1}^{h} \alpha_i, 1 \le h \le m$$

et

$$V_i^{(s)} = \begin{cases} \{i\}\ si\ s = 0 (pour\ tout\ \ i = 1,2,......) \\ \{ j \in G\ telque\ \ a_{s-1} < d_{ij} \le a_s \}\ \ si\ \ (i \ne j), 1 \le s \le \eta_i \le m. \\ \phi\ \ si\ \ non \end{cases}$$

où s est l'ordre de séparation spatiale. Puis nous avons utilisé ces voisinages pour générer les matrices de séparation spatiale à différentes ordres, en se basant sur :

$$W_{ij}^{(s)} = \begin{cases} \left|v_i^{(s)}\right|^{-1} & j \in v_i^{(s)} \\ 0\ \ si\ non \end{cases}$$

Cette formule nous donne plusieurs matrices à différent ordre de séparation. Nous remarquons que les matrices du premier ordre fournissent plus d'information concernant les interactions existantes entre les sites, d'où nous avons choisis une combinaison pour les trois premières matrices de séparation spatiale, qui est donnée par :

$$W = \beta_1 W^{(1)} + \beta_2 W^{(2)} + \beta_3 W^{(3)}$$

tel que $\beta_1 = h_1/(h_1 + h_2 + h_3)$

$\beta_2 = h_2/(h_1 + h_2 + h_3)$

$\beta_3 = h_3/(h_1 + h_2 + h_3)$

et chacune des h_1, h_2, h_3 suivent la loi normale centrée et réduite.

Finalement, pour la première partie de ce programme qui est simulation des données, il suffit de donner la taille de l'échantillon N et le nombre de paramètre k pour avoir les données.

V-3.Exemple

Nous avons déroulé le programme pour un exemple de N=12, k= 3, où nous obtenons 12 sites et 3 variables explicatives. Ces sites sont reconnus par leurs coordonnées qui sont données dans tableau 1 :

	x(i)	y(i)
1	2855	2773
2	770	2241
3	1860	611
4	1509	1277
5	2685	2813
6	2310	2759
7	1424	1290
8	154	2692
9	2482	268
10	1390	1123
11	1885	2458
12	2397	129

Tableau 1

La matrice de distance associés à ces sites et donnée dans l'annexe A. Et comme nous l'avons déjà dit ces matrices nous permettent de retrouver les paramètres de séparation spatiale, dans notre cas il y a « 6 » qui sont données comme suit :

$\alpha_1 =$ 763.4507, $\alpha_2 =$983.2940, $\alpha_3 =$637.8695, $\alpha_4 =$306.2958, $\alpha_5 =$669.9470, $\alpha_6 =$45.0229

Maintenant, nous pourrons calculer les différentes matrices de voisinages à différent ordre de séparation, dans ce cas nous obtenons « 6 »matrices, et nous avons données dans l'annexe B deux matrices de voisinage à l'ordre 2 et 4. Ces matrices sont notées par V(:,:,2), V(:,:,4), ces matrices contiennent que des 0 et 1, tel que le 1 veut dire qu'il y a voisinage et 0 il n'y a pas de voisinage. Si nous lisons dans l'annexe B, le site numéro 6 est voisin avec le site numéro (2 ,4,7) mais à d'ordre 2; par contre le site 4 est voisin d'ordre 2 à{2,6,9,11,12}et il n'est voisin à aucun site à ordre 4 (tableau 3).

Une fois l'ordre de séparation est choisi, nous obtenons notre matrice de poids ou matrice de séparation spatiale. Nous pourrons avoir 6 matrices de séparation spatiale, nous utilisons la combinaison déjà cité en haut et nous obtenons une seule matrice de poids spatiale, qui est données en annexe C.

Une fois la matrice de poids est choisi, nous calculons d'abord les bornes de l'intervalle où le paramétrer ρ doit être choisis, pour cette exemple, l'intervalle cherché et $\left]-0.0855,1.0000\right[$, et $\rho = 0.7$. Et comme nous avons choisis trois variables explicatives, nous pouvons obtenir les données sur lesquelles nous exécutons le prochain programme. Les données sont présentées dans l'annexe D.

V-4.Programmation des tests de l'autocorrélation spatiale

Maintenant que nous avons les données, nous passons à la programmation des tests de l'autocorrélation spatiale .suivant les règles de décisions suggérées par Ainslin et Florax (1995) nous commençons par estimer le modèle simple par MCO. D'où, nous obtenons les différents estimateurs que nous utilisons pour tests l'autocorrélation spatiale. Nous avons programmé le test de I Moran et de SARMA sur les résidus du modèle de régression qui sont données par leurs formes matricielles :

I=(e'*w*e)/(e'*e)

Tel que

e sont résidus estimés de MCO et w est la matrice de séparation spatiale. Le test se base sur la statistique de Moran centrée et réduite :

Z(I)=(I-E(I))/V(I)

E(I) et V(I) sont les deux premiers moments de I.

Pour SARMA :

SARMA=(dlan^2/T1)+(drou-dlan)^2/(sige^(-2))*(D1-T1*sige^2).

Tel que :

T1=trace($w*w$)+trace($w'*w$)

D1=((w*x*b)'*M*(w*x*b))+((sige^2)*T1)

dlan =(e'*w*e)/(sige^2)

drou=(e'*w*y)/(sige^2)

Où :sige est la variance estimés du modèle simple et b l'estimateur de B .

Ces deux tests sont très puissants contre toutes formes de dépendance spatiale mais ne permets pas de les discriminer. Pour cela, nous utilisons les deux tests du multiplicateur de Lagrange :

LMERR=dlan^2/T1.

Et

LMLAG=drou^2/D

Où :

D=(((w*x*b)'*M*(w*x*b))/sige^2)+T1

Et $M = I - X\left(X'X\right)^{-1}X'$

V-5.Exemple

Nous exécutons ce programme, pour l'exemple précédant, et nous obtenons les résultats suivant :

Les tests	Les résultats
I Moran	23.5512
SARMA	1.5639e+026
LMERR	1.5131
LMLAG	6.2167
RLMERR	0.1847
RLMLAG	4.8884

Tableau 4

Nous avons effectués six tests pour détecter l'autocorrélation spatiale. Premièrement le test de I Moran et de SARMA sur les résidus de régression (Cliff et Ord, 1980), ces test sont très puissant contre toutes les formes de dépendance spatiale mais ne permet pas de les discriminer. Pour cela, nous utilisons les deux tests du multiplicateur de Lagrange ainsi que leurs versions robustes $LM_{LAG}, LM_{ERR}, RLM_{LAG} et RLM_{ERR}$ (chapitre précèdent); LM_{ERR} permet de tester une autocorrélation spatial de erreurs et LM_{LAG} permet de tester une variable endogène décalée. Leurs versions robustes $RLM_{LAG} et RLM_{ERR}$ ont une bonne puissance contre leur alternative spécifique.

V-5-1.Interprétation des résultats

Les résultats des tests sont présentés dans le tableau 4. Au seuil $\alpha = 5\%$ nous représentons les valeurs tabulés de toutes les statistiques utilisées :

	La loi normale	La khi deux à 2 degrés de liberté	La khi deux à 1 degrés de liberté
$\alpha = 5\%$	1.96	5.99	3.84

En comparent les résultats trouvés avec les valeurs tabulés, nous avons $ZI = 23.5512 > 1.96$; C'est à dire le test de I Moran rejette l'hypothèse nulle au seuil $\alpha = 5\%$ ce qui indique la présence d'une certaine forme d'autocorrélation spatiale; ainsi que le test de SARMA, nous remarquons aussi que sa valeur est très grande ce qui veut dire quelque soit le seuil le test est significatif. Nous remarquons aussi, $LM_{LAG} = 6.2167 > 3.84$ et $LM_{ERR} = 1.5131 < 3.84$, c'est à dire le test LM_{LAG} rejet l'hypothèse nulle par contre le test LM_{ERR} ne le rejet pas. D'après Les règles de décisions suggérés par (Ainslin et Florax, 1995b) et présentés dans le chapitre IV page nous avons la présence d'une variable endogène spatialement décalé plutôt que la présence d'une autocorrélation spatiale des erreurs. Donc, la spécification la plus appropriés pour nos données est les modèles autorégressifs; car le test LM_{LAG} a la plus grande puissance pour son alternative. Dans notre cas, les tests robustes n'affinent pas les résultats car il n'y a pas de comparaison entre les tests, du au faite que le test LM_{ERR} n'est pas significatif. Ces résultats étaient prévisibles car les données ont été simulées suivant un modèle autorégressif.

V-6 Estimation du paramètre spatial

Dans ce paragraphe, nous déroulons un programme en utilisant les toolbox proposés par Le sage (2005) pour estimer notre paramètre spatial. En exécutant ce programme, nous obtenons:

1. La matrice de poids.
2. M1= la valeur propre négative la plus grande en valeur absolu
3. M2= la valeur propre positive la plus grande.
4. nous choisissons une valeur de rou à partir de M1 et M2, rou est paramètre spatial.
5. Les données Y.
6. Les trois variables explicatives.
7. Les tests de spécification.
8. L'estimateur de rou.

Pour N=12 comme nombre de sites et k=3 comme nombre de variables explicatives nous obtenons:

Rho=0.7020 qui est l'estimateur de la vraie valeur rou= 0.7000. donc notre paramètre a était bien estimé, aussi Le modèle a était bien détecté par les tests, où nous avons ZI = 2.8500 >

1.96 ce qui indique une présence d'autocorrélation spatiale. Nous remarquons aussi que LMERR = 2.8995 < 3.84 mais LMLAG = 8.2123 > 3.84 ce qui indique une présence d'une variable endogène spatialement décalé. Nous pourrons voir les résultats de ce programme dans l'annexe E

Pour vérifier si notre programme estime bien notre paramètre, nous l'exécutons pour un pas d'itération égal 100 jusqu'à 1000, N=12, et nous fixons rou= 0.9955**,** nous obtenons ce tableau :

N / résultats	100	200	300	400	500	600	700	800	900	1000
sigRho	0.1509	0.1375	0.1227	0.0789	0.0513	0.0497	0.0487	0.0438	0.0361	0.0349
Rho	0.9559	0.9866	0.9897	0.9835	0.9819	0.9848	0.9869	0.9857	0.9875	0.9859

Où: sigRho est l'écart-type des estimateurs et Rho est la moyenne des estimateurs.
A travers le tableau résumant les résultats obtenus lors des simulations. Nous constatons qu'autant le nombre d'itération augmente l'écart-type diminue. Nous remarquons aussi que pour chaque itération la moyenne des estimateurs ce rapproche de notre paramètre spatial « 0.9955 ». De là nous concluons que l'estimateur converge vers la vraie valeur.

Conclusion

Lors de ce travail, nous avons donné des définitions concernent l'un des deux effets spatiaux ; la dépendance spatiale (l'autocorrélation spatiale) considérée comme une relation fonctionnelle entre ce qui arrive en un point de l’espace et ce qui arrive ailleurs. Nous avons vu son introduction dans les modèles économétrique spatiaux tels que les modèles de régression spatiale. Elle se modélise grâce aux matrices de poids. Dont nous avons défini et apporté quelles méthodes aidant à leurs construction tel que la méthode du pointeur proposé par Boukhebouze et Dreosbeke. La dépendance spatiale est multidimensionnelle et multidirectionnelle ; ce qui entraîne la perte des propriétés des estimateurs obtenus par la méthode des moindres carrés ordinaires. Par conséquent, nous avons développé quelques méthodes d'estimation, particulièrement celle du maximum de vraisemblance, ainsi que les tests de spécification de. Une simulation des données a été faite, afin d'utiliser des tests de spécification pour la validation du type du modèle étudié, ainsi qu'un programme basé sur la méthode du pointeur proposée par Boukhebouze et Dreosbeke a été développé pour construire une matrice de poids, aussi la méthode du maximum de vraisemblance estime bien notre paramètre spatial.

En revanche, il reste beaucoup à développer dans ce récent domaine. Un volet de perspective ouverte pour la construction des matrices de poids. Car, à partir d'un champ de travail, on peut avoir une variété de matrice de poids spatiale et d'agrégation spatiale. Cette variété crée des problèmes méthodologiques en économétrie spatiale. L'un des problèmes important est : l'unité superficielle modifiable et le manque de critère pour le choix de la matrice de poids spatiale. Egalement, d'autres méthodes d'estimation restent à développer telle la méthode des moments et l’approche bayesien. Aussi, l'introduction de la dépendance spatiale dans les séries temporelles mérite d'être étudiée.

Une application de cette technique sur des données réelles est nécessaire, qui était notre objective. Mais vu les manques de donnée dans toutes les domaines, ne nous a pas aidé à développer une matrice de poids adéquate à notre problème, qui était la vente des carburants.

ANNEXE A

La matrice de distance

La matrice de distance associée aux coordonnées du tableau 1 est :

$d(i,j) =$ 1.0e+003 *

Columns 1 through 7

0	2.1518	2.3800	2.0124	0.1746	0.5452	2.0608
2.1518	0	1.9609	1.2147	1.9986	1.6248	1.1542
2.3800	1.9609	0	0.7528	2.3515	2.1946	0.8069
2.0124	1.2147	0.7528	0	1.9345	1.6846	0.0860
0.1746	1.9986	2.3515	1.9345	0	0.3789	1.9773
0.5452	1.6248	2.1946	1.6846	0.3789	0	1.7155
2.0608	1.1542	0.8069	0.0860	1.9773	1.7155	0
2.7022	0.7635	2.6909	1.9591	2.5339	2.1570	1.8917
2.5326	2.6122	0.7103	1.4017	2.5531	2.4969	1.4710
2.2065	1.2784	0.6950	0.1946	2.1291	1.8769	0.1704
1.0199	1.1359	1.8472	1.2394	0.8752	0.5208	1.2557
2.6834	2.6660	0.7216	1.4514	2.6994	2.6314	1.5148

Columns 8 through 12

2.7022	2.5326	2.2065	1.0199	2.6834
0.7635	2.6122	1.2784	1.1359	2.6660
2.6909	0.7103	0.6950	1.8472	0.7216
1.9591	1.4017	0.1946	1.2394	1.4514
2.5339	2.5531	2.1291	0.8752	2.6994
2.1570	2.4969	1.8769	0.5208	2.6314
1.8917	1.4710	0.1704	1.2557	1.5148
0	3.3609	1.9974	1.7467	3.4059
3.3609	0	1.3869	2.2699	0.1629
1.9974	1.3869	0	1.4238	1.4150
1.7467	2.2699	1.4238	0	2.3846
3.4059	0.1629	1.4150	2.3846	0

ANNEXE B

Les tableaux de voisinages

V(:,:,2)	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	1	0
2	0	0	0	1	0	1	1	0	0	1	1	0
3	0	0	0	0	0	0	1	0	0	0	0	0
4	0	1	0	0	0	1	0	0	1	0	1	1
5	0	0	0	0	0	0	0	0	0	0	1	0
6	0	1	0	1	0	0	1	0	0	0	0	0
7	0	1	1	0	0	1	0	0	1	0	1	1
8	0	0	0	0	0	0	0	0	0	0	1	0
9	0	0	0	1	0	0	1	0	0	1	0	0
10	0	1	0	0	0	0	0	0	1	0	1	1
11	1	1	0	1	1	0	1	1	0	1	0	0
12	0	0	0	1	0	0	1	0	0	1	0	0

Tableau 2

V(:,:,4)	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	1	0	0	1
2	0	0	0	0	0	0	0	0	1	0	0	1
3	0	0	0	0	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	1	1	0	0	0
6	0	0	0	0	0	0	0	0	1	0	0	1
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	1	0	1	0	0	0	0	0	0	0
9	1	1	0	0	1	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	0	0	0	1	0	0	0	0	0	0

Tableau 3

ANNEXE C

La matrice de poids spatiale

$W =$

Columns 1 through 7

0	0.0735	0.0735	0.0735	0.1285	0.1285	0.0735
0.1226	0	0.1226	0.0750	0.1226	0.0750	0.0750
0.0735	0.0735	0	0.0643	0.0735	0.0735	0.3752
0.1226	0.0750	0.0857	0	0.1226	0.0750	0.0857
0.1285	0.0735	0.0735	0.0735	0	0.1285	0.0735
0.0857	0.1251	0.1226	0.1251	0.0857	0	0.1251
0.1226	0.0625	0.0625	0.1285	0.1226	0.0625	0
0	0.2570	0	0.0919	0	0.0919	0.0919
0	0	0.1285	0.1251	0	0	0.1251
0.0919	0.0938	0.0857	0.0857	0.0919	0.0919	0.0857
0.0536	0.0536	0.1226	0.0536	0.0536	0.2570	0.0536
0	0	0.1285	0.1251	0	0	0.1251

Columns 8 through 12

0	0	0.0735	0.3752	0
0.2570	0	0.0750	0.0750	0
0	0.0643	0.0643	0.0735	0.0643
0.1226	0.0750	0.0857	0.0750	0.0750
0	0	0.0735	0.3752	0
0.1226	0	0.1226	0.0857	0
0.1226	0.0625	0.1285	0.0625	0.0625
0	0	0.0919	0.3752	0
0	0	0.1251	0.3677	0.1285
0.0919	0.0938	0	0.0938	0.0938
0.0536	0.1226	0.0536	0	0.1226
0	0.1285	0.1251	0.3677	0

ANNEXE D

Les données de la simulation

Les données de la simulation sont
Y=1.0e+006 *
1.1178
0.8597
1.0704
1.1202
1.1187
0.8749
1.0926
1.1311
1.6807
1.5420
2.4587
5.7136

ANNEX E

Les données de la simulation

Y=
-2.9518
-4.9961
-2.3541
-2.3512
-2.9525
-2.5847
-2.5478
-3.5502
-0.1479
-5.6319
-3.1301
-2.6082

Les valeurs des variables explicatives

X =

0.1370	1.8626	0.2631
1.5556	-0.2426	1.5750
0.7826	1.6174	1.1792
0.5859	-0.7223	0.7370
-0.3781	0.5630	-0.2468
0.6631	-0.0406	0.5496
-1.0340	-0.7406	-0.8498
0.6049	0.9460	0.3166
-1.0122	-1.7212	-0.8209
1.4689	2.7926	1.1854
0.0225	-2.1913	0.4347
0.0669	2.2525	0.0211

W =

0	0.1346	0.1401	0.0089	0.1346	0.1401	0.1346	0.1401	0.0089	0.0089	0.0089	0.1401
0.0577	0	0.0577	0	0.0357	0.0577	0.5605	0.0577	0	0.0577	0.0577	0.0577
0.2802	0.0808	0	0	0. 0808	0.0808	0.0808	0.2802	0	0	0.0357	0.0808
0.0059	0	0	0	0.0059	0.0059	0	0.0059	0.5605	0.4038	0.0059	0.0059
0.1010	0.0059	0.1010	0.0059	0.1010	0.0059	0.1010	0.0059	0.0059	0.0059	0.5605	
0.1868	0.0673	0.0673	0.0178	0.0673	0	0.0673	0.1868	0.0178	0.0673	0.0673	0.1868
0.0808	0.5605	0.0808	0	0.0119	0.0808	0	0.0808	0	0.0119	0.0119	0.0808
0.1401	0.1346	0.1401	0.0119	0.1346	0.1401	0.1346	0	0	0.0119	0.0119	0.1401
0.0071	0	0	0.5605	0.0071	0.0071	0	0	0	0.4038	0.0071	0.0071
0.0089	0.0808	0	0.0808	0.0089	0.0808	0.0089	0.0089	0.0808	0	0.5605	0.0808
0.0045	0.2019	0.0045	0.0045	0.0045	0.2019	0.0045	0.0045	0.0045	0.5605	0	0.0045
0.1401	0.1010	0.1010	0.0119	0.1401	0.1401	0.1010	0.1401	0.0119	0.1010	0.0119	0

Les tests

ZI = 2.8500
LMERR = 2.8995
LMLAG = 8.2123

Référence

1. Alia. L, Lebretonb. M, 2007, *The ERM breakdown: a spatial econometric approach*, Applied Economics Letters, 14, 197–201
2. Anselin. L, 1988, *Spatial Econometrics: Methods and Models*, Springer
3. Anselin. L, 1999, *Spatial Data Analysis with SpaceStat and ArcView* , spacestat_workbook. Pdf.
4. Anselin. L, 2001, *Rao's Score Test in Spatial Econometrics*, *Journal of Statistical Planning and Inference* 97 (1), 113-139.
5. Anselin. L, Cohen. J, Cook. D, Gorr. W, Tita. G, 2000, *Spatial Analyses of Crime*, Criminal Justice, vol 4, pp.213-260
6. Anselin. L, Florax .R. J. G. M, Rey. S. J, 2OO4, *Advances in Spatial Econometrics Methodology, Tools and Applications*, Springer-Verlag.
7. Anselin. L, Florax. R. J. G. M, 1995, *Spatial Effects in Linear Regression Models*, in «New Directions in Spatial Econometrics», Springer, Berlin.
8. Anselin. L, Syabri. I, Kho. Y, 2006, *GeoDa: une introduction à l'analyse de données spatiales*, Geographical Analysis, Vol. 38, pp. 5-22.
9. Anselin.L, 1999, *SPATIAL ECONOMETRIC*, www.csiss.org/learning_resources/content/papers/baltchap.pdf
10. Anselin.L, Rey.S, 1991, «*Properties of Tests for Spatial Dependence in Linear Regression Models*», Geographical Analysis, 23, pp. 112-131.
11. Baumont.C, Ertur. C, Le Gallo.J, 2000, «*Geographic Spillover and Growwth A Spatial Econometric Analysis for European Regions*»,http://www.u-bourgogne.fr/LEG/ertur/e2000-07.pdf
12. Bavaud. F, 1998, *Models for Spatial Weights: A Systematic Look*, Geographical Analysis, vol 30, pp. 154-170.
13. Bavaud. F,2003, *Une approche markovienne générale à l'analyse spatiale*, in "Actes des XXXVèmes Journées de Statistique", Société Française de Statistique (éditeur), pp. 169-172, Lyon
14. Bera. A. K, Escudero. W. S, 2002, *GMM Besed Tests for Locally Misspecified Models*, www.depeco.econo.unlp.edu.ar/doctrab/doc22.pdf
15. Bera. A. K, Escudero. W. S, Yoon. M, 2000, «*Tests for the Error Component Model in the Presence of Local Misspecification*», www.depeco.econo.unlp.edu.ar/doctrab/doc22.pdf

16. Bivand. R, 2002, *Spatial econometrics functions in R: Classes and methods*, J Geograph Syst (2002) 4:405–421
17. Blackman. A, Albers. H. J, Avalos-sartorio. B, Murphy. C. L, 2008, *LAND COVER IN A MANAGED FOREST ECOSYSTEM: MEXICAN SHADE COFFEE*, American Journal of Agricultural Economics, Vol. 90, Issue 1, pp. 216-231
18. Boukhebouze. A, 1999, « *A propos du concept de matrice de séparation spatiale* », Methodologica, N°7, 83-110
19. Boukhabouze. A, 1999, Some particular divisions on spatial data analysis, Pré-publication Institut de Mathématiques, Université des Sciences et de la Technologie Houari Boumédiène Bab-Ezzouar Alger, Algérie.
20. Boukhabouze. A, 2000, A propos du concept de matrice de séparatrion spatiale, Methodologica, (soumis pour publication).
21. Boukhabouze. A, Droesbeke J. J, 2000, Neighbourhood structure and independent classes in spatial data analysis, soumis pour publication.
22. Cliff. A.D, Ord. J.K, 1973, *Spatial Autocorrelation*, London, Pion.
23. Cliff. A. D, Ord. J. K, 1981, *Spatial Processes: Models and Applications*, London, Pion .
24. Ertur. C, Koch. W, 2OO4, *Analyse spatiale des disparités régionales dans l'Europe élargie*, ungaro.u-bourgogne.fr/pages/docs-travail.htm - 101k
25. Fingleton B. López-Bazo. E, 2006, *Empirical growth models with spatial effects,* Papers in Regional Science, Volume 85, Number 2, pp. 177-198(22).
26. Fischer M. M, Stirböck. C, 2006,*Pan-European regional income growth and club-convergence: Insights from a spatial econometric perspective*, Ann Reg Sci 40:693–721.
27. Fourgeaud, Fuchs. A, 1967, *Statistique*, Dunod, Paris.
28. Griffith D. A, 1980, *Towards a Theory of Statistics*, Geographical Analysis, 12, pp. 325-338
29. Griffith. D. A, Paelinck. J. H. P, 2007, *An equation by any other name is still the same: on spatial econometrics and spatial statistics*, Ann Reg Sci , 41, 209–227.
30. Guillaume. M, 1971, *Modèles Economique*, presses universitaires de France.

    in "Actes des XXXVèmes Journées de Statistique", Société Française de Statistique (éditeur), pp. 169-172, Lyon
31. Jayet. H, 2001, *Econométrie des données spatiales. Une introduction à la pratique*, Cahiers d'Economie et de Sociologie Rurales, 58-59, 105-129

32. Kissling W. D, Gudrun. C, 2007, *Spatial autocorrelation and the selection of simultaneous autoregressive models*, Global Ecology and Biogeography, 17, 59-71
33. Le Gallo. J, 2000, « *Econométrie spatiale 1 –Autocorrélation spatiale* », http://www.infotheque.org/ressource/4181.html
34. Le Gallo. J, 2004, *Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire*, www.cairn.info/load_pdf.php?ID_ARTICLE=ECOP_155_0139.
35. Le sage. J. P, 1998, *Spatial Econometrics* , www.spatialeconometrics.com/html/wbook.pdf
36. Le sage. J. P, 2004, *Bayesian estimation of spatial regression models*, www4.fe.uc.pt/spatial/doc/lecture3_slide.pdf
37. Le sage. J. P, 2004, *Maximum likelihood estimation of spatial regression models*, www4.fe.uc.pt/spatial/doc/lecture1.pdf
38. Le Sage. J. P, 2005, *Econometrics Toolbox for MATLAB*, http://www.spatial-econometrics.com
39. LICHSTEIn. W. J, SIMONS. R. T, SHRINER. S. A, FRANZREB. K. E, 2002, *SPATIAL AUTOCORRELATION AND AUTOREGRESSIVE MODELS IN ECOLOGY*, Ecological Monographs, 72(3), 2002, pp. 445–463.
40. Lundrg. J, 2006, *Using Spatial Econometrics to Analyse Local Growth in Sweden*, Regional Studies, Vol. 40.3, pp. 303–316.
41. Mulkay. B, Mars 2007, « *Estimation et test des modèles spatiaux* », www.crest.fr/semecopanels/Econométrie%20spatiale%20(4).pdf.
42. Pinkse. J, 2004, *Moran-Flavoured Tests with Nuisance Parameters: Examples*, forthcoming in: \New Advances in Spatial Econometrics,"Luc Anselin and Raymond J.G.M. Florax, eds., Springer (New York).