

Avant-propos

Introduction

Avant même d’aborder le sujet de front, il nous semble important de préciser la portée que nous souhaitons donner à ce livre. Le titre en lui-même est assez évocateur : il s’agit d’une introduction aux outils de base de l’économétrie spatiale. L’accent est porté sur les microdonnées : des observations qui consistent en des points sur une projection géographique plutôt qu’à des formes géométriques décrivant les limites (quelles qu’elles soient) d’une zone géographique. Nous proposons donc de couvrir les méthodes de détection et d’analyse spatiale descriptive, et la modélisation spatiale et spatio-temporelle.

En aucun cas nous souhaitons substituer cet ouvrage à des références importantes dans le domaine telles que Anselin (1988) [ANS 88], Anselin et Florax (1995) [ANS 95], LeSage (1999) [LES 99], ou encore la plus récente référence dans le domaine : LeSage et Pace (2009) [LES 09]. Nous considérons ces références comme des incontournables pour quiconque veut s’investir dans le domaine.

L’objectif du livre est de faire le lien entre les approches quantitatives existantes (analyse de corrélation, analyse bivariée, régression linéaire) et la façon dont on peut généraliser ces approches à des cas où les données disponibles à l’analyse possède une dimension spatiale. Bien que plusieurs équations soient présentées, notre approche est largement basée sur la description de l’intuition derrière chacune des équations. Le langage mathématique est un incontournable en analyse statistique et quantitative. Or, pour plusieurs, l’acquisition des connaissances nécessaires à la bonne lecture et à la bonne compréhension des équations est souvent décourageante. Pour cette raison, nous tâcherons de faire les liens entre l’intuition des équations et la formalisation mathématique à proprement parler. Trop peu d’ouvrages, à notre avis, mettent l’accent sur

cette structuration qui est pourtant la pierre angulaire de l'analyse quantitative. Après tout, le but de l'approche quantitative est de fournir un ensemble d'outils puissants permettant d'isoler certains effets que l'on cherche à identifier. Or, l'amplitude de ces effets dépendent justement du type d'outil utilisé pour le mesurer.

L'originalité de l'approche est, à notre avis, quadruple. Premièrement, le livre présente des exemples fictifs simples. Ces exemples permettent aux lecteurs de suivre, pour de petits échantillons, le détail des calculs à la main pour chacune des étapes de la construction des matrices de pondérations et des statistiques descriptives. Le lecteur pourra répliquer les calculs sur un logiciel de base tel qu'Excel pour s'assurer qu'il comprend bien chacune des étapes. Selon nous, cette démarche permet au lecteur non spécialiste d'intégrer les particularités des équations, des calculs et des données spatiales.

Deuxièmement, le livre propose de faire le lien entre l'écriture en sommation (voir double sommation) des statistiques (ou des modèles) et l'écriture matricielle. Plusieurs personnes peuvent rencontrer des difficultés pour faire le passage de l'un à l'autre. Nous présentons, dans cet ouvrage, pour quelques indices spatiaux, les deux écritures en insistant sur le passage d'une écriture à l'autre. La compréhension de l'écriture matricielle est importante puisqu'elle est plus compacte que l'écriture en sommation et rend plus lisible les expressions mathématiques comportant des doubles sommations tels que les indices de détection des schémas d'autocorrélation spatiale ; ceci est particulièrement utile dans les détections des schémas locaux. Le recours au calcul matriciel et à des exemples simples permettront aux lecteurs de généraliser les calculs à des jeux de données plus importants afin de faciliter la compréhension de l'économétrie spatiale. La forme matricielle rend également les calculs directement transposables dans un logiciel spécialisé (tel que MatLab, Mata (Stata), etc.) permettant de faire les calculs sans passer par des programmes préalablement écrits, du moins pour la construction des matrices de pondérations spatiales et pour le calcul des indices de concentration. La présentation des calculs matriciels pas à pas permet de bien saisir les étapes de calculs.

Troisièmement, le livre propose, en annexes, des programmes qui permettent de simuler des microdonnées spatiales et spatio-temporelles. Ces programmes permettent ensuite de transposer les présentations des chapitres sur des cas où la réalité est connue à l'avance. Cette approche, de type Monte Carlo, pourra bénéficier à certains lecteurs qui souhaiteraient examiner le comportement des statistiques de test ainsi que celui des estimateurs dans certains contextes bien précis. Les avantages de cette approche par simulations sont multiples :

- elle permet d'établir, de manière intuitive, les propriétés d'outils statistiques plutôt qu'une démonstration mathématique formelle ;

- elle permet de mieux connaître les processus générateurs de données (PGD) et d'établir des liens avec l'application des modèles statistiques ;
- elle offre la possibilité de tester l'impact de l'omission d'une dimension particulière (spatiale ou temporelle) sur les estimations et les résultats ;
- elle donne l'occasion au lecteur de mettre en œuvre ses propres expériences, moyennant quelques modifications mineures.

Finalement, la plus grande particularité de ce livre est certainement l'accent mis sur l'utilisation des microdonnées spatiales. La plupart des ouvrages et des applications en économétrie spatiale reposent sur des données spatiales agrégées. Cette représentation suppose alors que chacune des observations prend la forme d'un polygone (forme géométrique quelconque) représentant une délimitation fixe de frontières géographiques décrivant, par exemple, un pays, une région, une ville ou un quartier. Les données représentent alors une statistique agrégée d'observations individuelles (moyenne, médiane, proportion) plutôt que le détail de chacune des observations. A notre avis, les applications reposant sur les microdonnées sont une voie d'avenir pour la mise en pratique des méthodes économétriques spatiales, mais également pour une meilleure compréhension de plusieurs phénomènes. Les microdonnées spatiales permettent de contourner le problème classique de l'erreur écologique ² [ROB 50] en plus de répondre directement à plusieurs critiques disant que les données agrégées spatiales ne permettent pas de capter certains détails qui ne sont observables qu'à microéchelle. De plus, sans être exemptes du problème des unités spatiales modifiables (MAUP) ³ [ARB 01, OPE 79], elles ont au moins l'avantage de permettre explicitement la possibilité de tester l'effet de l'agrégation spatiale sur les résultats d'analyses.

Ainsi, le livre se veut un intermédiaire qui permet une transition, pour des non-économètres et des non-statisticiens, vers les livres de référence en économétrie spatiale. Le livre ne se veut donc pas être un ouvrage économétrique théorique axé sur des démonstrations mathématiques formelles ⁴, mais

2. Le problème de l'erreur écologique vient de la transposition des conclusions tirées sur des unités spatiales agrégées à des unités spatiales individuelles composant l'agrégation spatiale.

3. Le concept de MAUP (ou *Modifiable Area Unit Problem* en anglais) a été proposé par Openshaw et Taylor en 1979 pour désigner l'influence du découpage spatial (effets d'échelle et effets de zonage) sur les résultats de traitements statistiques ou de modélisation.

4. Le lecteur intéressé par une présentation plus formalisée de l'économétrie spatiale est invité à consulter un ouvrage récent de LeSage et Pace (2009) [LES 09] qui est considéré par certains chercheurs comme une référence marquant un « grand pas en avant » dans « l'état de l'art de l'économétrie spatiale » [ELH 10, p. 9].

plutôt un document d'introduction à l'économétrie spatiale appliquée des microdonnées.

A qui s'adresse l'ouvrage

La lecture de ce livre suppose néanmoins un minimum de connaissances en statistique et en économétrie. Il ne nécessite pas de connaissances particulières des systèmes d'informations géographiques (SIG). Même si l'ouvrage présente des programmes permettant la simulation de données à la fin de chacun des chapitres, il ne requiert aucune expérience particulière ou aptitude à la programmation.

Plus particulièrement, le livre s'adresse surtout aux étudiants de deuxième et troisième cycles dans les domaines liés aux sciences régionales. Comme le domaine des sciences régionales est plutôt large et multidisciplinaire, nous voulons établir une certaine mise en contexte pour les personnes qui désireraient se lancer dans l'analyse quantitative spatiale et pousser l'aventure plus loin. Selon nous, l'application des statistiques et des modèles statistiques ne peuvent plus se faire en ignorant la réalité spatiale des observations. L'aspect spatial apporte une richesse d'information qu'il est pertinent de considérer lors des analyses empiriques quantitatives.

Le livre s'adresse également aux étudiants des premier et deuxième cycles en économie qui désirent introduire la dimension spatiale dans leurs analyses. Nous croyons que ce livre permettra une excellente mise en contexte avant d'aborder formellement les aspects théoriques de l'économétrie visant à développer les estimateurs, à démontrer les preuves de convergences ainsi qu'à développer les tests de détection selon les approches classiques (tests du ratio de vraisemblance (LR), du multiplicateur de Lagrange et de Wald).

Nous souhaitons aussi atteindre les chercheurs qui ne sont pas des économètres ou des statisticiens, mais qui souhaitent en apprendre un peu plus sur la logique et les méthodes permettant de détecter la présence d'autocorrélation spatiale ainsi que sur les méthodes de correction d'éventuels problèmes survenant lors de la présence d'autocorrélation.

Structure du livre

Le livre est divisé en six chapitres qui suivent une logique bien précise. Le chapitre 1 propose une introduction aux analyses spatiales portant sur les données désagrégées ou individuelles (microdonnées spatiales). Une attention particulière est mise sur la structure des bases de données spatiales ainsi qu'à

leurs particularités. Il montre pourquoi il est essentiel de s'intéresser à l'économétrie spatiale si le chercheur possède des données qui sont géolocalisées et il présente un bref historique du développement de la branche de l'économétrie spatiale depuis sa formalisation.

Le chapitre 2 est certainement la pièce centrale de l'ouvrage. Il sert d'intrant aux autres chapitres qui ont recours, dans le calcul, aux matrices de pondérations. Il s'avère donc crucial et c'est la raison pour laquelle une emphase particulière y est apportée avec de nombreux exemples. Un exemple fictif est développé et repris au chapitre 3 pour démontrer le calcul des indices de détection des patrons d'autocorrélation spatiale.

Le chapitre 3 présente les mesures les plus utilisées pour détecter la présence de patrons spatiaux dans la distribution d'une variable donnée. Ces mesures s'avèrent particulièrement cruciales pour vérifier l'hypothèse d'absence de relation spatiale entre les résidus d'un modèle de régression. La présence d'une autocorrélation spatiale viole l'une des hypothèses qui assure l'efficacité de l'estimateur des moindres carrés ordinaires et peut modifier les conclusions issues du modèle statistique. La détection d'un tel patron spatial nécessite de corriger le modèle de régression et de recourir aux modèles de régression spatiale et spatio-temporelle. Evidemment, les indices de détection peuvent aussi être utilisés comme outils descriptifs et c'est en bonne partie sur cette assise que repose le chapitre.

Les chapitres 4 et 5 présentent les modèles standards utilisés en économétrie spatiale. Les modèles spatiaux (chapitre 4) peuvent facilement être transposés aux applications spatio-temporelles (chapitre 5) en développant une matrice de pondération adaptée à la réalité analysée. Ils présentent notamment l'intuition derrière le recours à un type de modèle plutôt qu'à un autre : c'est l'idée fondamentale derrière le processus générateur des données. En fonction du modèle supposé, les conséquences de la relation spatiale détectée entre les résidus du modèle de régression peuvent être plus ou moins importantes, allant d'une imprécision dans le calcul de la variance à un biais des estimateurs obtenus. Les annexes liées aux chapitres 4 (modélisation spatiale) et 5 (modélisation spatio-temporelle) sont basées sur la simulation de données et l'estimation des modèles autorégressifs à partir de matrices de pondération construites au préalable.

La conclusion souligne le rôle central de la construction de la matrice de pondérations spatiales en économétrie spatiale et les différentes voies possibles permettant de transposer les techniques et méthodes existantes à différentes définitions de la « distance ».

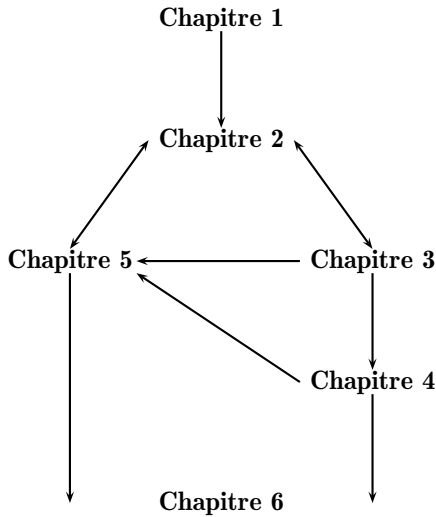


Figure 1. *Liens entre les chapitres*

En espérant que ce survol des fondements de l'économétrie spatiale donnera le goût à certains étudiants et chercheurs d'utiliser la modélisation spatiale économétrique dans le but de tirer le maximum possible de leurs bases de données et inspirer certains d'entre eux à proposer de nouvelles approches originales qui viendront compléter les méthodes développées à ce jour. Après tout, le développement des méthodes spatiales permet notamment d'intégrer les notions de proximités spatiales (et autres). Cet aspect est particulièrement crucial pour certains courants théoriques liés à la science régionale et à la nouvelle géographie économique (NGE), largement inspirés par les travaux de Paul Krugman [FUJ 04, KRU 91a, KRU 91b, KRU 98], récipiendaire du prix nobel en économie en 2008 [BEH 09], mais aussi pour d'autres champs reposant sur la formalisation de liens entre observations.

Bonne lecture.