

Chapitre 2

La matrice de pondération spatiale

Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger
Polycopié de cours · Économétrie spatiale (Master)

Plan du chapitre

1. Formaliser le voisinage
2. Trois familles de critères
3. La standardisation en ligne
4. Un choix qui engage toute la suite

1 Formaliser le voisinage

1.1 L'outil central de tout ce cours

Matrice de pondération spatiale

La matrice de pondération spatiale W est une matrice carrée de dimension $n \times n$, où n désigne le nombre d'unités géographiques effectivement observées. Son élément w_{ij} mesure précisément l'intensité du lien spatial existant entre l'unité i et l'unité j . Par convention constante, $w_{ii} = 0$: une unité donnée n'est jamais considérée comme sa propre voisine.

Ce que W accomplit concrètement

W transforme ainsi une notion intuitive mais floue — celle d'« être proche » — en un objet parfaitement manipulable algébriquement. Une fois W fixée une bonne fois pour toutes, toute la suite de ce cours en découle directement : le test de dépendance spatiale du chapitre 3, le modèle SAR du chapitre 4, le modèle SEM du chapitre 5 s'écrivent tous, sans exception, à partir de cette même matrice W .

2 Trois familles de critères

2.1 La contiguïté

Contiguïté de type tour et de type reine

La contiguïté de type tour relie deux unités entre elles si elles partagent une frontière commune non réduite à un simple point. La contiguïté de type reine ajoute en outre les unités qui ne se touchent, elles, qu'en un seul coin. On pose alors $w_{ij} = 1$ si i et j sont effectivement contiguës selon le critère retenu, et 0 dans le cas contraire.

Quatre régions disposées en carré

Quatre régions disposées en carré : A et B côte à côte au nord, C et D côte à côte au sud, A située au-dessus de C , B au-dessus de D . En contiguïté tour, A est voisine de B et de C , mais pas de D , puisqu'elles ne partagent qu'un simple coin commun. En contiguïté reine en revanche, A devient voisine des trois autres régions à la fois.

2.2 La distance

Pondération fondée sur la distance

$$w_{ij} = 1/d_{ij}^\alpha$$

où d_{ij} désigne la distance séparant les centroïdes respectifs de i et de j , et où $\alpha > 0$ règle précisément la vitesse de décroissance de l'influence considérée. Une variante à seuil fixe pose alternativement $w_{ij} = 1$ si d_{ij} reste inférieure à un seuil donné, et 0 au-delà de ce seuil.

Un choix qui n'a en réalité rien d'anodin

Une valeur $\alpha = 1$ donne une décroissance relativement lente de l'influence : des régions très éloignées conservent alors un poids non négligeable dans le calcul. Une valeur $\alpha = 2$, en décroissance de type carré inverse — à l'image même de la loi de gravitation — éteint en revanche beaucoup plus vite l'influence des régions lointaines. Aucun de ces choix n'est jamais réellement neutre : il engage à chaque fois une véritable hypothèse économique sur la portée effective du phénomène étudié.

2.3 Les k plus proches voisins

Critère des k plus proches voisins

$w_{ij} = 1$ si j fait effectivement partie des k unités les plus proches de i considérée, et 0 sinon. Contrairement aux deux critères précédents, cette matrice n'est pas nécessairement symétrique : j peut parfaitement être un proche voisin de i sans que la réciproque soit vraie, dans un semis de points géographiquement irrégulier.

Pourquoi ce critère se révèle précieux en pratique

Il garantit à chaque unité observée un nombre fixe de voisins retenus, ce qui évite complètement le problème des « îles » isolées — des unités qui n'auraient sinon aucun voisin sous un critère de distance trop strict, ou sous un critère de contiguïté si elles ne touchent effectivement aucune autre unité du jeu de données. C'est donc le choix par défaut le plus naturel pour des points précis — villes, entreprises — plutôt que pour des polygones administratifs entiers.

3 La standardisation en ligne

3.1 Transformer W pour l'interpréter comme une moyenne

Standardisation en ligne

$$w_{ij}^* = \frac{w_{ij}}{\sum_k w_{ik}}$$

Chaque ligne de la matrice W^* ainsi obtenue somme alors exactement à un. Le produit W^*y donne dès lors, pour chaque unité considérée, la moyenne pondérée de la variable y observée chez l'ensemble de ses voisines — une quantité directement interprétable en pratique.

Calcul complet sur les quatre régions

Avec la contiguïté reine retenue, A possède trois voisines directes (B, C, D), chacune pondérée à $1/3$ après standardisation en ligne. Si les taux de chômage observés valent $y_B = 8$, $y_C = 10$, $y_D = 6$, le décalage spatial associé à A vaut alors exactement

$$(Wy)_A = \frac{1}{3} \times 8 + \frac{1}{3} \times 10 + \frac{1}{3} \times 6 = 8.$$

Une lecture immédiate et directe

$(Wy)_A = 8$ signifie très précisément : « le taux de chômage moyen chez les voisines directes de A s'élève à huit pour cent ». Cette quantité précise, notée systématiquement Wy , constitue l'ingrédient qui apparaîtra explicitement au membre de droite du modèle SAR étudié au chapitre 4 qui suit.

4 Un choix qui engage toute la suite

4.1 W n'est jamais neutre

La mise en garde formulée par Anselin

W doit impérativement être fixée avant même l'estimation du modèle, à partir d'une information réellement exogène — géographie physique, distance mesurée —, et jamais à partir d'une variable économique elle-même corrélée avec y , sous peine de confondre artificiellement une dépendance spatiale réelle avec une simple corrélation entre régresseurs. Choisir W à partir des échanges commerciaux effectivement observés entre régions pour expliquer ensuite leur PIB respectif, par exemple, mélange dangereusement la cause et la conséquence du phénomène étudié.

Que faire lorsque W demeure incertaine

En pratique courante, on estime alors souvent le même modèle avec plusieurs matrices W également plausibles — contiguïté, distance à différents seuils, k plus proches voisins pour plusieurs valeurs de k successives — et l'on vérifie ensuite que les conclusions qualitatives obtenues résistent bien au changement de spécification retenu. Une conclusion qui ne survit finalement qu'à un seul choix particulier de W doit être considérée comme fragile.

5 Ce qu'il faut retenir

Cinq points

W formalise rigoureusement le voisinage à partir de trois grandes familles de critères possibles : contiguïté, distance, k plus proches voisins. Sa version standardisée en ligne transforme le produit Wy en une moyenne directement interprétable chez les voisins de chaque unité. Ce choix de W , effectué en amont et sur une base résolument exogène, conditionne ensuite toute l'analyse qui suit dans ce cours. La robustesse d'une conclusion se vérifie en la confrontant à plusieurs matrices W plausibles et concurrentes. Le chapitre 3 se sert précisément de cet outil pour construire le tout premier outil de diagnostic disponible : l'indice de Moran.

Pour aller plus loin

- [1] Luc Anselin, *Spatial Econometrics : Methods and Models*, Kluwer Academic Publishers, 1988.
- [2] James LeSage et Robert Kelley Pace, *Introduction to Spatial Econometrics*, CRC Press, 2009.
- [3] Luc Anselin, « Under the Hood : Issues in the Specification and Interpretation of Spatial Regression Models », *Agricultural Economics*, 2002.
- [4] J. Paul Elhorst, *Spatial Econometrics : From Cross-Sectional Data to Spatial Panels*, Springer, 2014.
- [5] Harry Kelejian et Ingmar Prucha, « A Generalized Spatial Two-Stage Least Squares Procedure », *Journal of Real Estate Finance and Economics*, 1998.