

Chapitre 3

Détecter la dépendance spatiale

Jaouad Madkour

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

Plan du chapitre

1. L'indice de Moran
2. Le C de Geary
3. Les indicateurs locaux (LISA)
4. Tester la significativité

1 L'indice de Moran

1.1 Un coefficient de corrélation, version spatiale

Indice de Moran

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

L'indice I mesure ainsi la corrélation existant entre la valeur observée d'une unité et celle de ses voisines, pondérées selon W . Il se lit exactement comme un coefficient de corrélation usuel : positif si les valeurs semblables se regroupent effectivement dans l'espace, négatif si elles alternent au contraire, et proche de zéro en l'absence de toute structure spatiale identifiable.

Ce que l'espérance sous indépendance révèle

Sous l'hypothèse nulle d'absence totale d'autocorrélation spatiale, on a $E[I] = -1/(n-1)$, une valeur très proche de zéro pour n suffisamment grand — et non de zéro exactement, à la différence notable d'un coefficient de corrélation ordinaire. Un I observé nettement supérieur à cette référence signale ainsi une dépendance positive ; nettement inférieur, une dépendance négative.

1.2 Le diagramme de Moran

Diagramme de Moran

Un nuage de points portant en abscisse la variable centrée-réduite z_i , et en ordonnée son décalage spatial $(Wz)_i$. La pente de la droite de régression de Wz sur z est précisément égale à I lui-même. Le nuage obtenu se lit en quatre quadrants distincts : Haut-Haut et Bas-Bas signalent des groupements de valeurs semblables — autocorrélation positive —, Haut-Bas et Bas-Haut signalent au contraire des unités qui contrastent nettement avec leur voisinage — autocorrélation négative, valeurs atypiques spatiales.

Reprise des quatre régions du chapitre 2

Avec $y_A = 7$, $y_B = 8$, $y_C = 10$, $y_D = 6$ et la contiguïté reine standardisée en ligne déjà construite au chapitre précédent, chaque région étant voisine des trois autres, I se calcule directement à partir des covariances entre chaque y_i et son décalage spatial $(Wy)_i$ correspondant. Un I positif et statistiquement significatif confirmerait ici que les taux de chômage élevés (observés en C) et faibles (observés en D) ne se distribuent nullement au hasard dans l'espace considéré.

2 Le C de Geary

2.1 Une mesure complémentaire, sensible au contraste local

Indice de Geary

$$C = \frac{(n-1)}{2 \sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2}$$

C varie typiquement entre zéro et deux; la valeur un correspond précisément à l'absence de dépendance spatiale, $C < 1$ signale une autocorrélation positive, $C > 1$ une autocorrélation négative — soit une échelle rigoureusement inversée par rapport à celle de l'indice I.

Pourquoi utiliser conjointement les deux indices

L'indice de Moran constitue une mesure globale de covariance : elle peut manquer une dépendance pourtant forte mais très localisée, si celle-ci se dilue dans l'ensemble de l'échantillon considéré. L'indice de Geary, construit sur des différences au carré terme à terme, se révèle plus sensible aux dissemblances proprement locales entre voisins immédiats. Les deux indices concordent le plus souvent en pratique, mais un désaccord persistant entre eux constitue lui-même une information précieuse.

3 Les indicateurs locaux

3.1 Passer du diagnostic global au diagnostic unité par unité

LISA, indicateurs locaux d'association spatiale

Le Moran local associé à l'unité i décompose directement l'indice global :

$$I_i = z_i \sum_j w_{ij} z_j, \quad \sum_i I_i \propto I$$

Chaque I_i est alors testé individuellement, ce qui permet de cartographier précisément où se situent réellement les groupements significatifs, plutôt que de se contenter d'un unique verdict global portant sur l'ensemble de l'échantillon disponible.

Une lecture cartographique immédiate et parlante

Une carte LISA colore chaque unité selon son quadrant propre du diagramme de Moran, uniquement lorsque le test local associé se révèle significatif : Haut-Haut en rouge pour un foyer de valeurs fortes, Bas-Bas en bleu pour un foyer de valeurs faibles, Haut-Bas et Bas-Haut en couleurs contrastées pour les valeurs atypiques repérées. C'est l'outil de base de toute étude sérieuse de la pauvreté régionale, de la criminalité urbaine ou de la diffusion effective d'une épidémie.

4 Tester la significativité

4.1 Approximation asymptotique et test par permutation

Deux façons de calculer une valeur p

Une approximation normale asymptotique, valable pour n suffisamment grand, fournit une statistique standardisée $z = (I - E[I]) / \sqrt{\text{Var}(I)}$. Une alternative solide, le test par permutation, redistribue aléatoirement les valeurs de y entre les unités un très grand nombre de fois, recalcule I à chaque tirage effectué, et situe ensuite la valeur observée au sein de cette distribution simulée sous l'hypothèse d'indépendance.

Pourquoi préférer parfois la permutation

L'approximation normale suppose une distribution sous-jacente bien précise — normalité, ou tout du moins un moment d'ordre quatre fini — que les petits échantillons régionaux vérifient assez rarement en pratique. Le test par permutation ne suppose en revanche rien de tel : il construit sa propre référence directement à partir des données disponibles, au prix d'un calcul plus lourd — largement compensé aujourd'hui par la puissance de calcul couramment disponible.

5 Ce qu'il faut retenir

Cinq points

L'indice de Moran généralise directement la notion de corrélation à la dimension spatiale du problème. Le diagramme de Moran rend cette dépendance lisible au moyen de quatre quadrants bien distincts. Le C de Geary offre une mesure complémentaire, plus sensible aux contrastes proprement locaux entre voisins. Les indicateurs LISA localisent enfin le diagnostic unité par unité, plutôt qu'un unique verdict global. Détecter une dépendance ne dit cependant pas encore où elle loge précisément dans le modèle économétrique — dans y lui-même, ou plutôt dans le terme d'erreur. Les chapitres 4 et 5 répondent directement à cette question centrale.

Pour aller plus loin

- [1] Patrick A.P. Moran, « Notes on Continuous Stochastic Phenomena », *Biometrika*, 1950.
- [2] Robert C. Geary, « The Contiguity Ratio and Statistical Mapping », *The Incorporated Statistician*, 1954.
- [3] Luc Anselin, « Local Indicators of Spatial Association — LISA », *Geographical Analysis*, 1995.
- [4] Luc Anselin, *Spatial Econometrics : Methods and Models*, Kluwer Academic Publishers, 1988.
- [5] Andrew Cliff et J. Keith Ord, *Spatial Processes : Models and Applications*, Pion, 1981.