

Chapitre 3

Le modèle autorégressif spatial (SAR)

Feuille d'exercices

Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger
Exercices et corrigés · Économétrie spatiale (Master)

1 Exercices du chapitre

Exercice 1 — Pourquoi les MCO échouent

On considère le modèle SAR $y = \rho Wy + X\beta + \varepsilon$.

- (a) Pourquoi Wy est-il corrélé avec le terme d'erreur ε , alors que W elle-même est fixée de façon purement exogène (chapitre 1) ?
- (b) En quoi ce problème diffère-t-il d'une simple omission de variable, où l'on pourrait au moins invoquer une correction ex post des écarts-types ?
- (c) Un étudiant propose de simplement régresser y sur Wy et X par MCO, en considérant Wy comme un régresseur ordinaire. Pourquoi cette stratégie produit-elle un $\hat{\rho}$ biaisé, quelle que soit la taille de l'échantillon ?
- (d) Citer les deux méthodes d'estimation convergentes présentées dans ce chapitre, et pour chacune, le principe qui la rend valide malgré l'endogénéité de Wy .

Exercice 2 — Le multiplicateur spatial, terme à terme

On reprend la contiguïté **reine** des quatre régions (A, B, C, D , chacune voisine des trois autres, poids $1/3$) et $\rho = 0,3$.

- (a) Écrire les trois premiers termes du développement en série de Neumann : $I, \rho W, \rho^2 W^2$.
- (b) Un choc de $+1$ sur X_C affecte y_C directement à hauteur de 1 (terme I). Calculer l'effet de ce même choc sur y_A (ou y_B , ou y_D , par symétrie) au premier tour de propagation (terme ρW).
- (c) Calculer l'effet de ce choc, au deuxième tour de propagation (terme $\rho^2 W^2$), sur y_C elle-même (effet de « retour ») et sur y_A .
- (d) En sommant les trois premiers termes, donner une approximation de l'effet total du choc sur y_C et sur y_A . Le calcul exact de $(I - \rho W)^{-1}$ donne 0,1299 pour l'effet sur y_A (colonne C) : l'approximation à l'ordre 2 est-elle déjà proche du résultat exact ?

Exercice 3 — La condition de stabilité

Avec une matrice W standardisée en ligne, la plus grande valeur propre de W vaut exactement 1, et la série de Neumann ne converge que si $|\rho| < 1$.

- (a) Pourquoi cette condition est-elle qualifiée, dans ce chapitre, d'« analogue » à la condition de stationnarité $|\phi| < 1$ d'un processus AR(1) ?
- (b) Un chercheur estime $\hat{\rho} = 1,15$ sur des données régionales. Que peut-on en conclure sur la validité de son modèle SAR, indépendamment même de la significativité statistique de ce coefficient ?
- (c) Un second chercheur trouve $\hat{\rho} = -0,6$, statistiquement significatif. Ce résultat est-il compatible avec la condition de stabilité ? Que signifierait, économiquement, un ρ négatif (par opposition à un ρ positif) ?

2 Exercice cumulatif (chapitre 3, chapitre 2, chapitre 1 et introduction)

Exercice 4 — De la détection à l'estimation

Une chercheuse dispose de données sur le prix de l'immobilier dans les quatre régions A, B, C, D du chapitre 1 (contiguïté reine). Elle calcule d'abord l'indice de Moran (chapitre 2) sur les résidus d'une régression MCO du prix sur des caractéristiques locales, et trouve $I = 0,38$, significativement supérieur à $E[I] = -1/3$.

- (a) D'après l'introduction du cours, ce seul résultat permet-il de savoir si la dépendance loge dans y (justifiant un modèle SAR) ou dans le terme d'erreur (justifiant un modèle SEM, chapitre 4) ?
- (b) La chercheuse décide malgré tout d'estimer un modèle SAR, en argumentant que les prix

immobiliers se déterminent réellement par comparaison directe entre régions voisines (un mécanisme de « dépendance substantielle » au sens de l'introduction). Pourquoi cet argument économique, plutôt que le seul résultat du test de Moran, est-il ce qui justifie réellement le choix du modèle SAR ?

- (c) Elle obtient $\hat{\rho} = 0,25$. En utilisant la matrice W construite au chapitre 1 (poids 1/3), décrire qualitativement comment un choc positif sur le prix d'une région se propagerait aux trois autres, en trois étapes successives.
- (d) Pourquoi ne peut-elle pas simplement réutiliser son estimation MCO initiale (celle qui a servi à calculer les résidus pour le test de Moran) comme point de départ fiable pour interpréter $\hat{\rho}$?

3 Applications en sciences économiques

Exercice 5 — Spillovers de prix immobiliers

Une étude estime, sur 200 quartiers d'une métropole, un modèle SAR du prix au m² avec $\hat{\rho}_{ML} = 0,45$ (maximum de vraisemblance) et $\hat{\rho}_{IV} = 0,41$ (variables instrumentales, WX et W^2X comme instruments), tous deux significatifs au seuil de 1 %.

- (a) Pourquoi la proximité des deux estimations (0,45 et 0,41) renforce-t-elle la confiance dans le résultat, plutôt que de se contenter de l'une des deux méthodes ?
- (b) Un promoteur immobilier veut savoir de combien le prix moyen de la métropole augmenterait si une politique d'aménagement faisait monter durablement le prix d'un seul quartier central. Pourquoi le multiplicateur spatial $(I - \rho W)^{-1}$, et non le seul coefficient $\hat{\rho}$, est-il l'outil pertinent pour répondre à cette question ?
- (c) Expliquer pourquoi une hausse artificielle et isolée du prix affiché dans un seul quartier (sans changement réel de ses caractéristiques) ne produirait, en réalité, aucun des effets de propagation prédits par le modèle SAR estimé sur données observationnelles.

Exercice 6 — Diffusion d'une innovation agricole, reprise

On reprend le scénario de l'introduction : l'adoption d'une nouvelle semence par des coopératives agricoles voisines. Un chercheur formalise maintenant ce mécanisme par un modèle SAR du taux d'adoption, avec $\hat{\rho} = 0,55$.

- (a) Quelle est l'interprétation économique directe de $\hat{\rho} = 0,55$ dans ce contexte précis (par opposition à une simple corrélation) ?
- (b) Une politique publique subventionne la première année d'adoption dans un petit nombre de coopératives « pilotes », dans l'espoir d'un effet d'entraînement. D'après le multiplicateur spatial, cette stratégie a-t-elle un fondement économétrique solide dans ce modèle, contrairement au cas de la corrélation nuisible évoqué dans l'introduction ?
- (c) Le chercheur choisit d'estimer ρ par IV/GMM plutôt que par ML, car ses données comportent $n = 3\,500$ coopératives. Pourquoi ce choix est-il cohérent avec les recommandations pratiques du cours ?