

Chapitre 6

Modèles généraux et choix de spécification

Jaouad Madkour

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

Plan du chapitre

1. Réunir SAR et SEM
2. L'hypothèse du facteur commun
3. La stratégie des multiplicateurs de Lagrange
4. Ce qu'il faut retenir

1 Réunir SAR et SEM

1.1 Le modèle de Durbin spatial

Spatial Durbin Model

$$y = \rho Wy + X\beta + WX\theta + \varepsilon$$

Le SDM ajoute au modèle SAR les régresseurs eux-mêmes décalés spatialement, WX , assortis de leur propre vecteur de coefficients θ . Il contient ainsi le SAR comme cas particulier lorsque $\theta = 0$, et une version restreinte du SEM comme autre cas particulier possible.

Pourquoi ce modèle mérite d'être le point de départ naturel

Le SDM répond directement à une objection tout à fait naturelle : pourquoi les régions voisines influenceraient-elles y_i uniquement via Wy , et jamais via leurs propres caractéristiques observées X_j ? En incluant explicitement WX , le SDM laisse alors les données elles-mêmes trancher si cette influence directe existe réellement, plutôt que de l'exclure a priori par simple hypothèse de départ.

1.2 Le modèle combiné SAC ou SARAR

Modèle SAC / SARAR

$$y = \rho Wy + X\beta + u, \quad u = \lambda Wu + \varepsilon$$

Ce modèle superpose ainsi un décalage spatial dans y et une erreur autorégressive spatiale, avec potentiellement deux matrices de pondération distinctes retenues — W_1 pour le décalage, W_2 pour l'erreur elle-même.

Le prix réel de cette généralité accrue

Plus le modèle retenu est général, plus il devient difficile à estimer correctement — identification plus fragile, calcul nettement plus lourd — et plus les résultats obtenus deviennent sensibles à la spécification exacte de W_1 et de W_2 retenue. Cette généralité accrue n'est donc jamais gratuite : elle s'achète toujours au prix d'une puissance statistique et d'une robustesse d'estimation moindres.

2 L'hypothèse du facteur commun

2.1 Un test qui relie formellement SDM et SEM

Restriction du facteur commun

Le modèle SEM peut en réalité se réécrire comme un SDM particulier, sous la contrainte non linéaire $\theta = -\rho\beta$. Un test du rapport de vraisemblance portant sur cette restriction précise indique alors si les données recueillies sont compatibles avec un simple SEM, ou si elles exigent au contraire la généralité complète du SDM.

Ce que rejeter cette restriction signifie concrètement

Rejeter la restriction du facteur commun signifie que les régions voisines comptent pour des raisons qui ne se réduisent jamais à une simple corrélation d'erreurs fortuite : leurs caractéristiques propres X_j exercent un effet réel sur y_i , bien distinct de tout artefact purement statistique. C'est là un argument solide en faveur du SDM plutôt que du simple SEM.

3 La stratégie des multiplicateurs de Lagrange

3.1 Choisir sans avoir à estimer tous les modèles

Tests LM-lag et LM-error d'Anselin

Deux tests du multiplicateur de Lagrange, calculables directement à partir des seuls résidus MCO du modèle sans dépendance spatiale : LM_ρ teste l'hypothèse $H_0 : \rho = 0$ contre l'alternative d'un SAR, LM_λ teste $H_0 : \lambda = 0$ contre l'alternative d'un SEM. Aucun des deux tests n'exige d'estimer préalablement le modèle spatial complet correspondant.

Le piège bien réel des tests non robustes

Un modèle SAR mal spécifié comme SEM, ou l'inverse, contamine mutuellement les deux tests précédents : LM_ρ peut alors rejeter à tort simplement parce qu'une dépendance de type SEM est en réalité présente, et réciproquement pour LM_λ . D'où l'existence de versions robustes, notées LM_ρ^* et LM_λ^* , corrigées pour rester valides même en présence de l'autre type de dépendance concurrente.

3.2 L'arbre de décision usuellement retenu

La procédure d'Anselin et Florax

Estimer d'abord le modèle par MCO simples, calculer ensuite LM_ρ et LM_λ . Si aucun des deux n'est significatif, les MCO suffisent amplement. Si un seul des deux l'est, retenir alors le modèle correspondant, SAR ou SEM selon le cas. Si les deux se révèlent significatifs, examiner les versions robustes LM_ρ^* et LM_λ^* : celle qui demeure significative désigne le modèle à privilégier finalement ; si les deux versions robustes restent significatives, le modèle SAC ou SARAR s'impose alors de lui-même.

Une philosophie concurrente : partir directement du modèle général

Elhorst recommande à l'inverse une démarche descendante : partir directement du SDM le plus général possible, puis tester s'il se réduit effectivement à un simple SAR (lorsque $\theta = 0$) ou à un simple SEM (via la restriction du facteur commun). Cette approche du général vers le particulier évite le risque, propre à la démarche ascendante d'Anselin, de rester durablement bloqué sur un modèle trop restreint, faute d'avoir testé au préalable l'alternative la plus large possible.

Ce que la pratique retient effectivement des deux approches

Aucune des deux démarches proposées ne s'impose unanimement comme supérieure à l'autre : la stratégie ascendante se révèle plus simple et suffit souvent en présence de peu de régresseurs disponibles ; la stratégie descendante se révèle plus prudente, mais nettement plus coûteuse en calcul et en degrés de liberté disponibles. La plupart des travaux appliqués sérieux rapportent aujourd'hui les deux démarches à la fois, et vérifient soigneusement leur convergence effective.

4 Ce qu'il faut retenir

Cinq points

Le SDM et le SAC/SARAR généralisent tous deux SAR et SEM, en autorisant plusieurs mécanismes de dépendance spatiale à la fois. La restriction du facteur commun relie formellement le SDM et le SEM entre eux. Les tests du multiplicateur de Lagrange, en versions simples puis robustes, permettent de choisir une spécification appropriée sans avoir à estimer tous les modèles candidats disponibles. La démarche ascendante d'Anselin et la démarche descendante d'Elhorst constituent deux philosophies concurrentes et également légitimes. En pratique sérieuse, les deux démarches sont généralement rapportées ensemble, pour vérifier leur convergence mutuelle. Le chapitre 7 s'attaque à présent à une question restée en suspens depuis le chapitre 4 : comment interpréter correctement un coefficient, une fois le modèle spatial finalement choisi.

Pour aller plus loin

- [1] Luc Anselin, *Spatial Econometrics : Methods and Models*, Kluwer Academic Publishers, 1988.
- [2] Luc Anselin et Raymond Florax, *New Directions in Spatial Econometrics*, Springer, 1995.
- [3] J. Paul Elhorst, « Applied Spatial Econometrics : Raising the Bar », *Spatial Economic Analysis*, 2010.
- [4] James LeSage et Robert Kelley Pace, *Introduction to Spatial Econometrics*, CRC Press, 2009.
- [5] Harry Kelejian et Dennis Robinson, « A Suggested Method of Estimation for Spatial Interdependent Models », *Papers in Regional Science*, 1993.