

Résumé économétrie spatiale

Vincent Robitaille

1. Modèle de régression linéaire classique

Sous la forme vectorielle:

$$y = X\beta + \varepsilon, \quad (\varepsilon, X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n)$$

L'hypothèse que les résidus conditionnels à X soient indépendants et identiquement distribués selon cette loi normale avec une variance constante a deux implications:

1. Exogénéité, c'est à dire que les chocs sont indépendants des variables explicatives. $E[\varepsilon|X] = 0$
2. Sphéricité des termes d'erreur, c'est à dire l'absence d'autocorrélation et l'absence d'hétéroscédasticité dans les termes d'erreur. Cela implique que les chocs sont parfaitement indépendant entre eux et que leur variance n'est pas influencée par les valeurs de X .

Lorsque ces conditions sont respectées, on considère alors l'estimateur des moindres carrées β_{OLS} comme étant le meilleur estimateur linéaire non biaisé (BLUE). L'estimation par OLS consiste à trouver les valeurs des paramètres qui minimisent la somme des termes d'erreur au carré (la différence entre la valeur prédite et la valeur réelle pour l'ensemble des observations). La somme des erreurs au carré est définie comme $S(\beta) = e'e$, où $e = y - X\hat{\beta}$. La solution de ce problème d'optimisation nous amène alors à la solution

$$\hat{\beta} = (X'X)^{-1}X'y$$

On peut vérifier que l'estimateur est bel et bien sans biais puisque $E[\hat{\beta}|X] = \beta$. De plus, sa variance est $Var[\hat{\beta}|X] = (X'X)^{-1}\sigma_\varepsilon^2$. Cela revient donc à dire que la distribution de l'estimateur des OLS suit une normale iid d'espérance et de variances précédemment définies, soit

$$(\hat{\beta}|X) \stackrel{iid}{\sim} N(\beta, (X'X)^{-1}\sigma_\varepsilon^2)$$

L'estimation par OLS coïncide aussi dans ce cas (loi normale) avec l'estimateur du maximum de vraisemblance (MLE). La vraisemblance est une fonction qui calcul la densité de l'observation des données (y_1, \dots, y_n) selon un vecteur de paramètres θ . Celle-ci est obtenue à partir de la distribution conjointe de l'ensemble des observations de l'échantillon disponible. Pour l'optimisation par MLE, le vecteur de paramètres qui maximise cette fonction sert d'estimation ponctuelle pour les paramètres d'intérêt.

Dans le cas où les résidus suivent une loi normale iid en respectant les hypothèses énoncées, l'expression de la fonction de vraisemblance est la suivante:

$$L(\beta, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \varepsilon_i^2 \right\}$$

Encore une fois, la résolution du problème d'optimisation nous permet d'obtenir une solution fermée pour les estimateurs du MLE, ainsi que la matrice d'information de Fisher soient:

$$\begin{aligned} \hat{\beta}_{MLE} &= (X'X)^{-1}X'y \\ \hat{\sigma}_{MLE}^2 &= \frac{e'e}{n} \\ I(\beta, \sigma_\varepsilon^2) &= \begin{bmatrix} \frac{1}{\hat{\sigma}_\varepsilon^2} X'X & 0 \\ 0 & \frac{n}{2\hat{\sigma}_\varepsilon^4} \end{bmatrix} \end{aligned}$$

De plus, l'estimation par la méthodes moments permet également d'obtenir la même expression pour l'estimateur de β .

$$\begin{aligned} \frac{1}{n} X'e &= E[X'\varepsilon] = 0 \\ \Rightarrow \frac{1}{n} X'(y - X\beta) &= 0 \\ \Rightarrow \frac{1}{n} X'y - \frac{1}{n} X'X\beta &= 0 \\ \Rightarrow \hat{\beta}_{MM} &= (X'X)^{-1}X'y \end{aligned}$$

2. Introduction aux données spatiales

S'il y a présence d'autocorrélation dans les termes d'erreur, certains éléments de hors diagonale de la matrice de variance-covariance se seront pas nuls. Les propriétés optimales des MCO se seraient alors pas valides et l'estimation par GLS serait uniquement possible si on pouvait spécifier exactement la forme d'autocorrélation. Soit les régions $1, \dots, n$ avec une autocorrélation spatiale, nous pourrions nous attendre à observer au moins l'un de ses deux phénomènes:

- Une autocorrélation spatiale positive (négative) apparaît lorsque les individus *proches* les uns des autres sont plus (moins) semblables que ceux éloignés.
- L'hétérogénéité spatiale apparaît lorsque certaines régions présentent plus de variabilité que d'autres. Cela peut être représenté à l'aide de la matrice de variance-covariance.

Afin de travailler avec des données spatiales, nous devons posséder deux types d'information à propos des individus :

1. Valeurs observées de variables individuelles (social, économique, etc.)
2. La localisation de l'observation de ces variables et les liens de proximité entre les observations spatiales.

Plusieurs critères de proximité peuvent être utilisés. Parmi ceux-ci, on retrouve notamment, en référence aux échecs : - *Rook's criterion*: les unités sont proches si elles partagent un côté. - *Queen's criterion*: les unités sont proches si elles partagent un côté ou un coin. - En réalité, les espaces géographiques sont assez complexes (régions administratives, irrégularité géométriques, etc) et d'autres types de critères peuvent être utilisés. On retrouve notamment la distance entre le centre des régions, les plus proches voisins, de même que des mesures de distance non géographiques telles que des interactions sociales dans un réseau et autres mesures de distance et d'association non géographiques.

2.1 Matrice de pondération spatiale W et lag spatial

La matrice de pondération spatiale W permet de décrire les associations entre les différentes observations. Dans sa forme la plus simple, les éléments w_{ij} ($i \neq j$) de celle-ci prennent la valeur 1 si l'individu i est voisin de l'individu j (si $j \in N(i)$) et 0 sinon. L'ensemble $N(i)$ représente les voisins de la localisation i selon différents critères tels que :

- Territoires adjacents
- Distance maximale ($j \in N(i)$ si $d_{ij} < d_{max}$)
- Critère du plus proche voisin
- Fonction négative de distances géographiques, économiques ou sociales entre des régions au lieu de données dichotomiques.

$$W = \begin{bmatrix} w_{11} & \cdots & w_{n1} \\ \vdots & w_{ij} & \vdots \\ w_{1n} & \cdots & w_{nn} \end{bmatrix}$$

La matrice standardisée W^* , dans laquelle les éléments sont $w^*_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}}$ a pour utilité principale le calcul du lag spatial d'un vecteur d'intérêt. Le lag spatial revient à prendre la

moyenne pondérée valeurs observés chez les voisins d'un individu et est défini comme:

$$L(y) = W * y, \quad L(y_i) = \sum_{j=1}^n w_{ij} y_j$$

3. Modèles de régression linéaire spatiaux

3.1 Forme générale

De manière similaires aux modèles de séries chronologiques qui relâchent l'hypothèse d'absence d'autocorrélation temporelle, nous allons maintenant relâcher l'hypothèse d'indépendance spatiale dans les chocs stochastiques. Posons d'abord que W est la matrice standardisée. Sous sa forme la plus générale, nous allons considérer le modèle autorégressif spatial avec erreurs autorégressives SARAR(1,1) suivant:

$$\begin{aligned} y &= \lambda W y + X \beta_{(1)} + W X \beta_{(2)} + u, & |\lambda| < 1 \\ \Leftrightarrow y &= \lambda W y + Z \beta + u, & Z = [X, W X], \quad \beta = [\beta_{(1)}, \beta_{(2)}] \\ u &= \rho W u + \varepsilon, & |\rho| < 1, \quad (\varepsilon|X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n) \end{aligned}$$

Théorème de Gershgorin

Soit

$$\begin{aligned} (I - \lambda W)y &= X \beta_{(1)} + W X \beta_{(2)} + u \\ \Leftrightarrow y &= (I - \lambda W)^{-1} [X \beta_{(1)} + W X \beta_{(2)} + u] \\ u &= (I - \rho W)^{-1} \varepsilon \end{aligned}$$

Sous la condition de l'existence des deux matrices inverses, alors lorsque W est standardisée, les deux matrices inverses existent si $|\rho| < 1$ et $|\lambda| < 1$.

De ce modèle, six cas particuliers de modèles sont présentés.

3.1.1 Modèle spatial autorégressif pur

Nous avons ici $\beta = 0$ et soit $\lambda = 0$ ou $\rho = 0$. De plus, W est non stochastique et $(\varepsilon|X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n)$. Le modèle devient alors une simple autorégression spatiale qui peut être estimée par MLE:

$$\begin{aligned} \rho = 0 &\Rightarrow y = \lambda W y + \varepsilon \\ \lambda = 0 &\Rightarrow y = \lambda W y + \varepsilon, \quad \text{où } y = u \end{aligned}$$

3.1.2 Lag spatial de X (SLX)

Soit $\lambda = \rho = 0$, $(\varepsilon|X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n)$, W et X non stochastiques et Z de plein rang, alors le modèle ne contient qu'un lag spatial de certaines variables indépendantes. Cela ne cause aucun problème pour l'estimation qui peut être faite par OLS.

$$y = WX\gamma + X\beta + u$$

3.1.3 Modèle d'erreurs spatiales (SEM)

Nous avons $\lambda = 0$ et $\rho \neq 0$.

$$\begin{aligned} y &= Z\beta + u \\ u &= \rho Wu + \varepsilon \end{aligned}$$

De plus, si $(\varepsilon|X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n)$, alors

$$\begin{aligned} &\Rightarrow u = (I - \rho W)^{-1} \varepsilon \\ &\Rightarrow E[u] = 0 \\ &E[uu'] = \sigma_\varepsilon^2 (I - \rho W)^{-1} (I - \rho W)'^{-1} = \sigma_\varepsilon^2 \Omega \\ &\quad \rightarrow \text{prend en compte à la fois l'hétéroscédasticité et l'autocorrélation des } u \\ &\quad \rightarrow \text{l'estimation GLS peut être utilisée si } \rho \text{ est connu a priori} \\ (I - \rho W)y &= (I - \rho W)Z\beta + (I - \rho W)u \\ &\Rightarrow y = \rho Wy + Z\beta - WZ\rho\beta + \varepsilon \\ &\Rightarrow y = \rho Wy + Z\beta - WZ\gamma + \varepsilon, \quad \gamma = \rho\beta \end{aligned}$$

Cette dernière expression montre une surparamétrisation dû à $\gamma = \rho\beta$ et Wy est corrélée avec les termes d'erreur, ce qui est source d'endogénéité.

$$\begin{aligned} (I - \lambda W)y &= Z\beta - WZ\gamma + \varepsilon \\ \Rightarrow y &= (I - \lambda W)^{-1}(Z\beta - WZ\gamma) + (I - \lambda W)^{-1}\varepsilon \end{aligned}$$

De manière à ce que la covariance entre Wy et le terme d'erreur puisse être exprimé comme:

$$\begin{aligned} E[(Wy)\varepsilon'] &= E[W(I - \lambda W)^{-1}(Z\beta - WZ\gamma) + W(I - \lambda W)^{-1}\varepsilon\varepsilon'] \\ &= W(I - \lambda W)^{-1}(Z\beta - WZ\gamma)E[\varepsilon'] + W(I - \lambda W)^{-1}E[\varepsilon\varepsilon'] \\ &= \sigma_\varepsilon^2 W(I - \lambda W)^{-1}I \neq 0 \end{aligned}$$

On peut donc voir que le terme d'erreur est corrélé avec Wy , ce qui rend OLS non optimal pour l'estimation. Il est également impossible d'utiliser des instruments, car la procédure serait inconsistante dû à l'impossibilité d'identifier des instruments pour lesquels Wy est linéairement dépendante des deux autres régresseurs Z et WZ .

3.1.4 Modèle de lag spatial (SLM)

Nous avons $\lambda \neq 0$ et $\rho = 0$.

$$y = \lambda W y + Z \beta + u$$

Un problème d'endogénéité apparaît puisque y est corrélé avec les chocs. Nous pouvons voir que:

$$\begin{aligned} (I - \lambda W)y &= Z\beta + u \\ y &= (I - \lambda W)^{-1}Z\beta + (I - \lambda W)^{-1}u \end{aligned}$$

La corrélation entre Wy et u :

$$\begin{aligned} E[(Wy)u'] &= E[W(I - \lambda W)^{-1}Z\beta + (I - \lambda W)^{-1}uu'] \\ &= W(I - \lambda W)^{-1}Z\beta E[u'] + (I - \lambda W)^{-1}E[uu'] \\ &= \sigma_\varepsilon^2 W(I - \lambda W)^{-1}, \quad I \neq 0 \end{aligned}$$

Modèle spatial Durbin (SDM)

Il s'agit essentiellement d'un modèle à lag spatial (SLM) auquel a été ajouté un lag de variables indépendantes.

$$y = \lambda W y + W X \gamma + X \beta + u$$

Ce type de modèle est souvent présenté comme préférable dans la littérature pour les raisons suivantes:

1. Choisir un SDM revient à estimer un SEM sous contrainte.

$$\begin{aligned} y &= X\beta + u, \quad u = \rho W u + \varepsilon \\ (I - \rho W)y &= (I - \rho W)X\beta + (I - \rho W)u \\ \text{en utilisant } \varepsilon &= (I - \rho W)u \\ y &= \rho W y + X\beta - \rho W X\beta + \varepsilon \\ \text{en prenant } \gamma &= -\rho\beta, \text{ cela coïncide avec le SDM} \end{aligned}$$

2. Le biais par omission de variables est commun dans la modélisation spatiale où il est presque impossible d'observer tous les prédicteurs pertinents. De plus ceux qui sont inobservables risquent fortement d'être reliés à la variable dépendante. L'utilisation d'un SDM réduit ce type de biais puisque le lag des variables exogènes aide à expliquer les effets des variables omises.
3. Peut aider à mitiger les effets provenant d'une hétérogénéité inobservée.
4. Un SDM produit des estimations non biaisées même lorsqu'on échoue à spécifier le bon processus de génération des données, qui pourrait être un SLM ou un SEM. Au contraire, si on commet une erreur de spécification en n'identifiant pas le SDM comme le vrai processus, toutes les spécifications alternatives vont souffrir d'un biais par omission de variables.

Modèle général SARAR(1,1) - Modèle spatial autorégressif avec erreurs autorégressives

$$y = Z\beta + \lambda Wy + u$$

$$u = \rho Wu + \varepsilon$$

$$(\varepsilon|X) \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2 I_n)$$

note: le modèle peut seulement exister si $\beta \neq 0$

Le modèle présente deux difficultés concernant l'estimation:

1. Problème d'endogénéité associée à la présence du lag de y (Wy), similaire au SLM.
2. Présence d'autorégression dans les chocs stochastiques fait en sorte qu'on ne peut utiliser le GLS à moins de connaître ρ a priori.