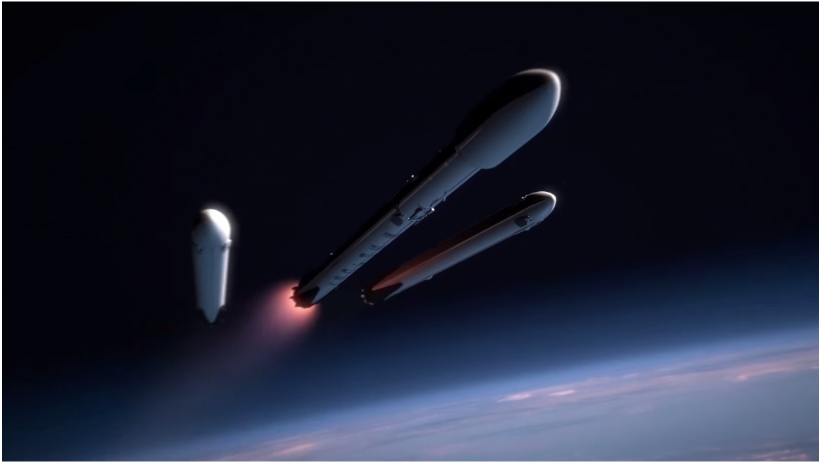
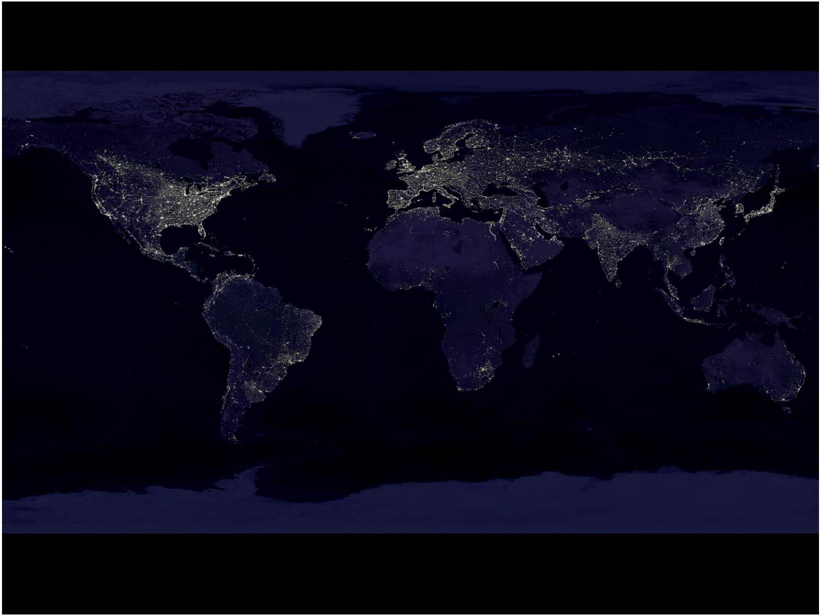
Prendre en compte la dépendance spatiale avec l'économétrie spatiale

Lionel Védrine
CESAER, AgroSup Dijon, INRA, UBFC

17 mai 2018
Séminaire DMCSI

Econométrie spatiale?





Dépendance spatiale?

Causes économiques

Au-delà des caractéristiques géographiques, la configuration spatiale des unités n'est pas neutre:

- externalités de diffusion,
- mobilité/échanges,
- interactions stratégiques.

Définition

La dépendance spatiale résulte des ces différents mécanismes et implique:

- les unités ne sont pas indépendantes les unes des autres,
- l'intensité des interactions est décroissante avec l'éloignement.

Dépendance spatiale?

Causes économiques

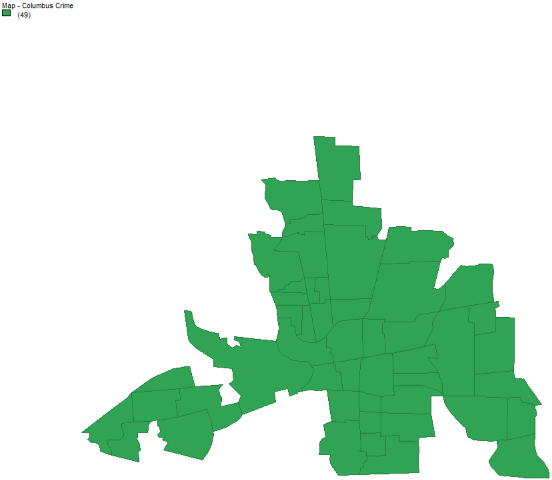
Au-delà des caractéristiques géographiques, la configuration spatiale des unités n'est pas neutre:

- externalités de diffusion,
- mobilité/échanges,
- interactions stratégiques.

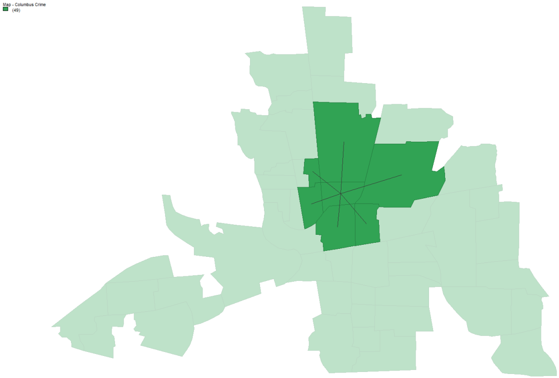
Autocorrélation spatiale

corollaire statistique:
$cov(y_i, y_j) \neq 0$ pour $i \neq j$

De l'arrangement spatial au terme spatial



De l'arrangement spatial au terme spatial



De l'arrangement spatial au terme spatial

$$C = \begin{array}{c} \\ R1 \\ R2 \\ R3 \\ R3 \\ R5 \\ R6 \\ R7 \end{array} \begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

De l'arrangement spatial au terme spatial

$$Wy = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}$$

$$= \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix}$$

De l'arrangement spatial au terme spatial

$$Wy = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}$$

$$= \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix}$$

L'opérateur de décalage spatial

En coupe:

$$WY = W_N.Y_N = \sum_{i \neq j} w_{ij} y_j$$

De l'arrangement spatial au terme spatial

L'opérateur de décalage spatial

En panel, sous l'hypothèse de constance de la matrice de pondération spatiale:

$WY = (I_T \otimes W_N)Y_N = \sum_{i \neq j} w_{ij} y_{j,t}$

$$\begin{pmatrix} W_N & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & W_N \end{pmatrix}$$

Les effets spatiaux dans un modèle linéaire

Modèle sur données empilées

$$y_{it} = x_{it}\beta + \alpha + \epsilon_{it} \quad (1)$$

où ϵ_{it} est $iid(0, \sigma^2)$

Les effets spatiaux dans un modèle linéaire

On introduit un premier terme spatial:

$$y_{it} = x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt} \theta + \alpha + \epsilon_{it} \quad (1)$$

un effet d'interaction exogène, θ : la valeur de Y_i dépend des caractéristiques de son voisinage (contrôle de la corrélation de l'environnement des unités).
Utiliser pour modéliser les effets externes de politiques publiques.

Les effets spatiaux dans un modèle linéaire

On introduit un terme autorégressif spatial:

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \alpha + \epsilon_{it} \quad (1)$$

un effet d'interaction endogène (ρ) : la valeur de Y_i dépend directement de la valeur prise dans son voisinage WY => Modèle représentant le résultat d'équilibre d'un processus faisant interagir les décisions des individus

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$$

$$y = \underbrace{(I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots)^{-1} X\beta}_{\textit{effet multiplicateur}} + \underbrace{(I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots)^{-1} \varepsilon}_{\textit{effet diffusion}}$$

L'omission de Wy implique:

- biais dans l'estimation de β
- biais dans l'estimation de σ^2

Les effets spatiaux dans un modèle linéaire

On introduit un terme autorégressif spatial dans les erreurs:

$$\begin{aligned} y_{it} &= x_{it}\beta + \alpha + u_{it} && (1) \\ u_{it} &= \lambda \sum_{i \neq j} w_{ij} u_{jt} + \epsilon_{it} && (2) \end{aligned}$$

un effet d'interaction des inobservables, (λ): les effets inobservables qui affectent la valeur de Y_i sont autocorrélés avec ceux de son voisinage.

$$y = X\beta + \underbrace{(I\lambda W)^{-1}\epsilon}_{\textit{Effet diffusion}}$$

L'omission de Wu implique:

- biais dans l'estimation de σ^2

Les effets spatiaux dans un modèle linéaire

Modèle "générateur des données" non identifiable (Manski, 1993):

$$\begin{aligned} y_{it} &= \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha + u_{it} \qquad (1) \\ u_{it} &= \lambda \sum_{i \neq j} w_{ij} u_{jt} + \epsilon_{it} \end{aligned}$$

(2)

Les effets spatiaux dans un modèle linéaire

Modèle "générateur des données" non identifiable (Manski, 1993):

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha + u_{it} \quad (1)$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \epsilon_{it}$$

(2)

Faire une restriction sur certains paramètres:
SARAR

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \alpha + u_{it} \quad (3)$$

$$u_{it} = \lambda \sum w_{ij} u_{jt} + \epsilon_{it} \quad (4)$$

Les effets spatiaux dans un modèle linéaire

Modèle "générateur des données" non identifiable (Manski, 1993):

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha + u_{it} \quad (1)$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \epsilon_{it}$$

(2)

Faire une restriction sur certains paramètres:
SDM

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha + \epsilon_{it} \quad (3)$$

(4)

Les effets spatiaux dans un modèle linéaire

Modèle "générateur des données" non identifiable (Manski, 1993):

$$\begin{aligned} y_{it} &= \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt} \theta + \alpha + u_{it} \quad (1) \\ u_{it} &= \lambda \sum_{i \neq j} w_{ij} u_{jt} + \epsilon_{it} \end{aligned}$$

(2)

Faire une restriction sur certains paramètres:
SDEM

$$\begin{aligned} y_{it} &= x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt} \theta + \alpha + u_{it} \quad (3) \\ u_{it} &= \lambda \sum w_{ij} u_{jt} + \epsilon_{it} \quad (4) \end{aligned}$$

Interprétations des coefficients en présence d'effet multiplicateur spatial

- Avec la présence d'un terme autorégressif spatial (SAR):
  $\frac{\partial E(y_i)}{\partial X_k} = I\beta + \rho W\beta + \rho^2 W^2\beta + ... + \rho^n W^n\beta$
  qui se décompose en:
  - effet direct du changement de $x_i$sur le résultat de i:
    $\frac{\partial E(y_i)}{\partial X_{ik}} = (I - \widehat{\rho}W)^{-1}_{ii}(I\widehat{\beta}_k)$
    - inclut l'effet de feedback où i affecte j et j affecte aussi i en retour,
  - effet indirect du changement de x_j sur le résultat de i:
    $\frac{\partial E(y_i)}{\partial X_{jk}} = (I - \widehat{\rho}W)^{-1}_{ij}(I\widehat{\beta}_k)$
    - où $(I - \widehat{\rho}W)^{-1}_{ij}$ représente le ij eme élément de la matrice de transformation spatiale inverse $(I - \widehat{\rho}W)^{-1}$
    - représente l'effet du changement de la valeur prise par une observation j de la variable X_j et affectant (à travers l'effet multiplicateur spatial) le résultat de l'observation i (Y_i).

Exemple: le rendement de l'éducation

Model	Direct Effect	Indirect Effect	Total Effect
OLS	0.2891	/	/
OLS Fixed Effects	0.2653	/	/
SDM	0.2829	1.1231	1.4061
SDM Fixed Effects	0.2663	0.0753	0.342
SDM: regime 1	0.2864	1.1521	1.4386
SDM: regime 2	0.2202	0.8855	1.1057
SDM Fixed Effects: regime 1	0.2994	0.0866	0.386
SDM Fixed Effects: regime 2	0.2388	0.0691	0.308

Seconde relation de Verdoorn

Cette loi relie, de manière linéaire, les taux de croissance de la productivité du travail (p) à ceux de l'output (q) dans le secteur manufacturier pour un ensemble d'économies. La spécification de base est donnée par:

$$p_{it} = b_0 + b_1 q_{it} + b_2 G_{it} + b_3 u_{it} + b_4 d_{it} + \epsilon_{it} \tag{5}$$

$$p_{it} = b_0 + \rho \sum_{i \neq j} w_{ij} p_{jt} + b_1 q_{it} + b_2 G_{it} + b_3 u_{it} + b_4 d_{it} + \epsilon_{it} \tag{6}$$

$$p_{it} = b_0 + b_1 q_{it} + b_2 G_{it} + b_3 u_{it} + b_4 d_{it} + \epsilon_{it} \tag{7}$$

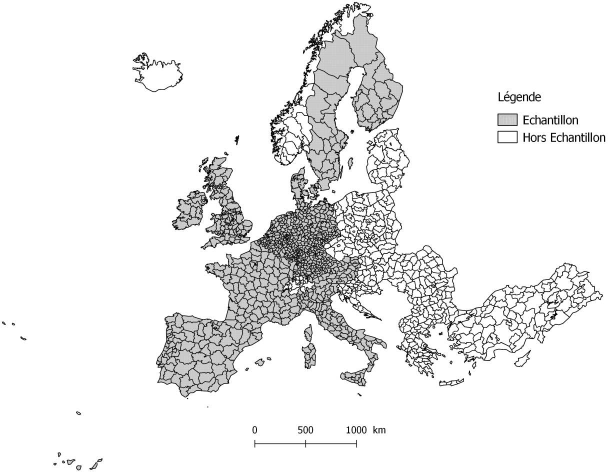
$$\epsilon_{it} = \alpha_i + \lambda \sum_{i \neq j} w_{ij} \epsilon_{jt} + v_{it} \tag{8}$$

ou:

$$\epsilon_{it} = \lambda \sum w_{ij} \epsilon_{jt} + v_{it} \tag{9}$$

Données

1032 régions (NUTS3) pour 1991-2008 (moyennes de 3 ans)



Résultats

Modèle:	données empilées	effets fixes (within)	effets aléatoires (MCG)
	(1)	(2)	(3)
q	0.692***	0.640***	0.701***
	(0.009)	(0.010)	(0.010)
u	0.0001***	−0.0002	0.0001***
	(0.00001)	(0.0002)	(0.00001)
G	0.0001	0.002***	0.0003***
	(0.0001)	(0.0001)	(0.0001)
d	−0.008***	−0.182***	−0.033***
	(0.003)	(0.005)	(0.003)
Constante	0.146***		0.228***
	(0.012)		(0.014)
Observations	5,160	5,160	5,160
Adjusted R^2	0.523	0.587	0.552
Note:			*p<0.1; **p<0.05; ***p<0.01

Table: Estimations sans prise en compte de l'autocorrélation spatiale

Tests de spécification

```
chisq = 1040.8, df = 4, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent

Test d'Hausman robuste à l'autocorrélation spatiale (SEM)

Hausman test for spatial models
chisq = 1263.8, df = 4, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent

Test d'Hausman robuste à l'autocorrélation spatiale (SAR)
Hausman test for spatial models

chisq = 1504, df = 4, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Tests de spécification

```
LM test for spatial lag dependence

LM = 326.41, df = 1, p-value < 2.2e-16
alternative hypothesis: spatial lag dependence


LM test for spatial error dependence

LM = 1115.5, df = 1, p-value < 2.2e-16
alternative hypothesis: spatial error dependence


Locally robust LM test for spatial lag dependence sub spatial
error

LM = 0.0025551, df = 1, p-value = 0.9597
alternative hypothesis: spatial lag dependence
```

Résultats

		p		
Modèle:	données empilées	effets fixes (MV)		effets fixes (MMG)
		erreur Baltagi	erreur KKP	
	(1)	(2)	(3)	(4)
q	0.716***	0.650***	0.650***	0.836***
	(0.017)	(0.008)	(0.008)	(0.009)
u	0.0001***	0.0001	0.0001	0.0001
	(0.00001)	(0.0002)	(0.0002)	(0.0002)
G	-0.0004***	0.001***	0.001***	0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
d	-1.70***	-0.163***	-0.163***	-0.164***
	(0.003)	(0.0005)	(0.0005)	(0.005)
Constante	0.2***			
	(0.02)			
λ		0.566***	0.566***	0.513***
		(0.02)	(0.02)	(0.02)
Observations	5,160	5,160	5,160	5,160
Note:				*p<0.1; **p<0.05; ***p<0.01

Table: Estimations du modèle sur données empilées et du modèle à effets fixes avec autocorrélation spatiale des erreurs

Limites et extensions

- La dépendance spatiale permet de modéliser des configurations de données plus complexes, permettant de:
  - étudier les interactions stratégiques entre individus,
  - contrôler de variables inobservables.
- Développement récent de l'économétrie spatiale sur données de panel améliore la précision et la crédibilité des résultats pour l'identification d'interactions spatiales:
  - panel statique avec matrice fixe dans le temps,
  - panel dynamique spatiaux,
  - matrices spatio-temporelles.
- Développement simultané et synergie avec:
  - les modèles multi-niveaux,
  - les modèles à facteurs communs.

Merci de votre attention

Contact

lionel.vedrine@inra.fr
http://lionelvedrine.wordpress.com
lionelvedrine