

Économétrie spatiale et données empilées dans le temps : proposition d'une modélisation adaptée

Jean Dubé ¹ Diègo Legros ²

¹Université Laval, Québec, Canada, Email : jean.dube@esad.ulaval.ca

²Université de Bourgogne, France, Email : diego.legros@u-bourgogne.fr

ACFAS
Rimouski, Québec, Canada
25-29 mai 2015

Modélisation hédonique des prix immobiliers

- La théorie hédonique de Rosen (1974) suggère que le prix d'un bien immobilier est expliqué par :
 - Caractéristiques intrinsèques
 - Surface, nombre de pièces, type de biens, étage, date de construction etc
 - Caractéristiques extrinsèques :
 - Proximité avec les écoles, nombre de commerces, taux de criminalité dans le voisinage, pollution etc.
- Variables temporelles
 - Dummies trimestrielles, annuelles
 - Capter l'évolution nominale des prix
- La forme fonctionnelle la plus souvent retenue est linéaire dans les paramètres

$$y_{it} = \alpha + X_{it}\beta + D_{it}\delta + \varepsilon_{it} \quad (1)$$

- y_{it} est souvent mis en logarithme

Modélisation hédonique des prix immobiliers

- Depuis la fin des années 80, un ensemble de travaux (Dubin et Sung, 1987 ; Can, 1992 ; Dubin, 1998) intègrent la dimension spatiale dans la détermination du prix d'un bien immobilier
- Intuition : les prix des biens immobiliers peuvent être :
 - corrélés dans l'espace ...
 - et liés aux prix des biens voisins
- Les méthodologies développées en économétrie spatiale permettent
 - de tester la présence d'une éventuelle autocorrélation spatiale
 - et de modéliser cette autocorrélation

Modélisation hédonique des prix immobiliers

- Il existe, en analyse immobilière, deux principales approches pour introduire l'autocorrélation spatiale
 - L'introduire dans les perturbations : modèle à erreurs autocorrélées (SEM)

$$\begin{aligned}
 y_{it} &= \alpha + X_{it}\beta + D_{it}\delta + \mu_{it} \\
 \mu_{it} &= \lambda W \mu_{it} + \varepsilon_{it}
 \end{aligned}
 \tag{2}$$

- L'introduire via une spécification autorégressive : le prix d'un bien est expliqué par le prix de biens spatialement décalés (SAR).

$$y_{it} = \alpha + \rho W y_{it} + X_{it}\beta + D_{it}\delta + \varepsilon_{it}
 \tag{3}$$

- Cette approche est très populaire en modélisation hédonique.
- Ces approches requièrent la construction d'une matrice de pondérations spatiales, W , informant du lien entre chaque observation.

Construction de la matrice de pondérations spatiales W

- Matrice de pondérations spatiales est une matrice carrée de taille $N \times N$.

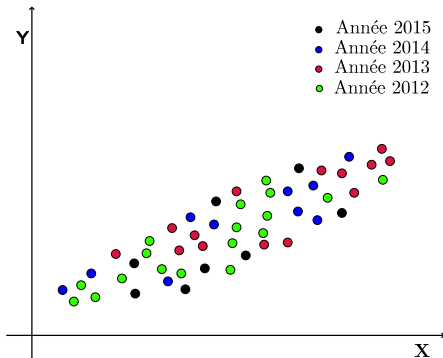
$$W = \begin{pmatrix} 0 & w_{12} & w_{13} & \cdots & w_{1N} \\ w_{21} & 0 & w_{23} & \cdots & w_{2N} \\ w_{31} & w_{32} & 0 & \cdots & w_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & w_{N3} & \cdots & 0 \end{pmatrix} \quad (4)$$

- w_{ij} mesure l'intensité du lien entre l'observation i et j
 - w_{ij} non stochastique, non négatif et fini.
 - w_{ij} exogène, défini *a-priori* par le chercheur

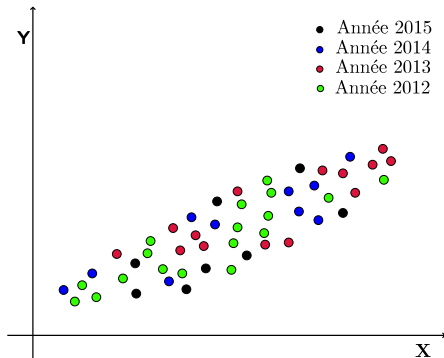
Données immobilières = données spatio-temporelles

- Les données de transactions immobilières sont :
 - spatialement localisées
 - apparaissent à une date précise
 - sont rarement répétées dans le temps
- Habituellement, la dimension temporelle est absente dans les modèles d'économétrie spatiale
 - si ce n'est par l'introduction de dummies temporelles dans les régresseurs.
 - La matrice de pondérations spatiales « standard » W est a-temporelle
 - Impose implicitement une symétrie temporelle : la distance temps est neutre (passé = présent = futur)
 - Des individus proches spatialement peuvent avoir été observés à des dates très éloignées l'une de l'autre.

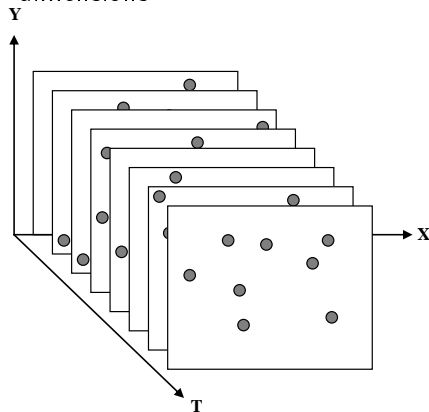
Les données immobilières sont envisagées en 2 dimensions



Les données immobilières sont envisagées en 2 dimensions



alors qu'elles possèdent 3 dimensions



Construction d'une matrice de pondérations spatio-temporelles

- Proposition : construire une matrice de pondérations intégrant les deux dimensions : **Matrice de pondérations spatio-temporelles**
- Pour cela : s'appuyer sur deux matrices
 - Matrice de pondérations spatiales notée S
 - L'élément s_{ij} mesure la distance spatiale entre deux observations ($s_{ij} = s_{ji}$)
 - Matrice de pondérations temporelles notée T
 - L'élément t_{ij} mesure la distance temporelle entre deux observations ($t_{ij} \neq t_{ji}$)

Construction d'une matrice de pondérations spatio-temporelles

- La matrice de pondérations spatio-temporelles, de dimension $N \times N$ est obtenue à l'aide des matrices S et T :

$$W = S \odot T \quad (5)$$

- de terme général $w_{ij} = s_{ij} \times t_{ij}$
- De façon générale, la matrice spatio-temporelle, W , s'écrit :

$$W = S \odot T = \begin{pmatrix} 0 & s_{12}t_{12} & s_{13}t_{13} & \dots & s_{1N}t_{1N} \\ s_{21}t_{21} & 0 & s_{23}t_{23} & \dots & s_{2N}t_{2N} \\ s_{31}t_{31} & s_{32}t_{32} & 0 & \dots & s_{3N}t_{3N} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ s_{N1}t_{N1} & s_{N2}t_{N2} & s_{N3}t_{N3} & \dots & 0 \end{pmatrix}$$

Mobilisation de la matrice de pondérations spatio-temporelles

- Vérifier la robustesse des mesures d'autocorrélation spatiale
 - I de Moran
 - Coefficients λ et ρ des modèles SAR, SEM.
- Possibilité de créer de nouvelles variables explicatives en fixant des seuils sur les distances spatiale et temporelles
 - Effets dynamiques de prix spatialement localisés

$$y_{s,t-l} = Wy$$

- \Rightarrow prix moyen des biens vendus dans un voisinage et période fixée (1 trimestre avant, 2 trimestres avant etc.)

Mobilisation de la matrice de pondérations spatio-temporelles

- Décomposition de l'effet spatial en un effet multidirectionnel et un effet unidirectionnel

$$y_{it} = \alpha + \rho \underline{S} y_{it} - \psi \underline{W} y_{it-1} + X_{it} \beta + D_{it} \delta + \varepsilon_{it} \quad (6)$$

- \underline{S} matrice spatiale bloc diagonale (effets multidirectionnels dans une même période)
- \underline{W} matrice spatio-temporelle triangulaire inférieure avec seuil (les effets unidirectionnels des observations de la période précédente)

Surévaluation I de Moran

Application sur 20 246 données québécoises 1986-1996

- Introduction de la matrice W dans un modèle hédonique
 - Comparaison de la valeur des I de Moran selon la matrice de pondérations retenue
- Modélisation hédonique du prix de transactions immobilières
 - Caractéristiques intrinsèques
 - Caractéristiques extrinsèques
 - Dummies temporelles
- Détection autocorrélation spatiale dans les résidus (I de Moran)
- Matrices de pondérations spatiales utilisées
 - Matrice de contiguïté
 - Matrice de distance 500 mètres
 - Matrice de distance 1 000 mètres
- Des différences notables entre les I de Moran calculés S et W

Surévaluation I de Moran

Application sur des données québécoises 1986-1996

	Matrice contiguïté		Matrice distance 500m	
	I Moran	V(I)	I Moran	V(I)
Matrice spatiale S	0.13139	0.00904	0.19399	0.01213
$W = S \odot T$ (Passé : 2-3 trim av.)	0.04841	0.00817	0.10148	0.01086
$W = S \odot T$ (Présent : 1 trim)	0.02748	0.00745	0.05933	0.01199

Comparaison OLS, SAR, STAR

Application sur 294 768 transactions en Ile de France (1990-2003)

- Test de performance de 3 modélisations en termes de prévisions hors-échantillon
- Les modèles testés

$$y_{it} = \alpha + D_{it}\delta + X_{it}\beta + \varepsilon_{it} \quad (7)$$

$$y_{it} = \alpha + D_{it}\delta + W y_{it} + X_{it}\beta + \varepsilon_{it} \quad (8)$$

$$y_{it} = \alpha + D_{it}\delta + \underline{S} y_{it} + \underline{W} y_{it} + X_{it}\beta + \varepsilon_{it} \quad (9)$$

- Deux échantillons de 10 000 extraits aléatoirement
 - Estimation des paramètres sur un échantillon
 - Prévisions effectuées sur le deuxième échantillon, calcul de la corrélation entre prévisions et variable endogène

Comparaison OLS, SAR, STAR

Application sur 294 768 transactions en Ile de France (1990-2003)

	ρ		MSE	
OLS specification	0.8157		0.1424	
Distance seuil (km)	Matrice S		Matrice W	
	ρ	MSE	ρ	MSE
0.5	0.8155	0.1428	0.8248	0.1361
1	0.8096	0.1470	0.8401	0.1252
1.5	0.8229	0.1373	0.8410	0.1245
2	0.7065	0.2659	0.8272	0.1474
2.5	0.5517	0.7596	0.8088	0.1791
3	0.4971	1.9313	0.7841	0.2250

Effets dynamiques

Application sur 25 357 transactions du comté de Lucas, OHIO, 1993-1998

- Modélisation hédonique prix biens immobiliers
- 3 effets dynamiques

$$y_{it-1} = y_{it} \times (S \odot T_1) \quad (10)$$

$$y_{it-2} = y_{it} \times (S \odot T_2) \quad (11)$$

$$y_{it-3} = y_{it} \times (S \odot T_3) \quad (12)$$

- S : matrice spatiale 1km
- T_1 : 1-er trimestre
- T_2 : 2-ème trimestre
- T_3 : 3-ème trimestre

$$y_{it} = \alpha + y_{it-r}\rho + D_{it}\delta + X_{it}\beta_k + \varepsilon_{it} \quad (13)$$

- Estimation d'un modèle à erreur spatialement autocorrélées (SEM)

Effets dynamiques

Application sur 25 357 transactions du comté de Lucas, OHIO, 1993-1998

Table : Résultats Estimation OLS - log(prix)

Variables	Model HPM		Model CHPM	
	Coefficients	t-statistic	Coefficients	t-statistic
Y_{it-1}	-	-	0.0180***	10.27
Y_{it-2}	-	-	0.0170***	10.09
Y_{it-3}	-	-	0.0103***	6.81
R^2	0.7311	-	0.7387	-
\bar{R}^2	0.7310	-	0.7385	-
Moran I with W	0.2234***	77.50	0.2180***	75.70
Moran I with S	0.3240***	337.43	0.3083***	321.56

Effets dynamiques

Application sur 25 357 transactions du comté de Lucas, OHIO, 1993-1998

Table : Résultats Estimation SEM - matrice (S) - log(prix)

	Model HPM		Model CHPM	
Variables	Coefficients	t-statistic	Coefficients	t-statistic
y_{it-1}	-	-	0.000 7	0.54
y_{it-2}	-	-	0.003 1**	2.49
y_{it-3}	-	-	0.001 5	1.39
λ	0.990 0***	132.80	0.990 0***	133.29

Effets dynamiques

Application sur 25 357 transactions du comté de Lucas, OHIO, 1993-1998

Table : Résultats Estimation SEM - matrice (W) - log(prix)

Variables	Model HPM		Model CHPM	
	Coefficients	t-statistic	Coefficients	t-statistic
y_{it-1}	-	-	0.008 7***	4.54
y_{it-2}	-	-	0.015 5***	8.04
y_{it-3}	-	-	0.010 5***	5.51
λ	0.580 9***	197.78	0.684 0***	174.60

Conclusion

- Négliger la dimension temporelle dans les données spatio-temporelles conduit à une surestimation le coefficient d'autocorrélation spatiale
- Impacte la qualité de l'estimation des paramètres notamment le paramètre d'autocorrélation spatiale et donc de la qualité de la mesure des **effets marginaux**.
 - OLS : Effets marginaux : β
 - SAR : Effets marginaux : $(I - \lambda W)\beta$
- La décomposition de l'effet spatial en un effet multidirectionnel et un effet unidirectionnel fournit de meilleures performances en termes de prévision hors échantillon.
- Possibilité de créer :
 - Des variables dynamiques locales
 - Ces nouvelles variables peuvent constituer des instruments lors de problème d'endogénéité (surface)