

Econométrie spatiale : une introduction pratique

Jean-Michel Floch, Ronan Le Saout

JMS, 2 Avril 2015

Motivations

- Un emploi plus fréquent ;
- Des notions complexes (matrice de voisinage, autocorrélation spatiale) ;
- De multiples modèles en concurrence ;
- Nécessité d'un guide détaillant pas à pas la mise en place de ces modèles.

Quelques utilisations au sein de l'INSEE

- Le chômage et le marché de l'emploi : Blanc et Hild (2009), puis le PSAR de Toulouse (2014) ;
- La non-réponse à l'enquête emploi : Loonis (2012) ;
- La relation entre les prix immobiliers et les risques industriels : Grislain-Létrémy et Katosky (2013) ;
- Les migrations résidentielles : Guymarc (2015) ;
- Une prise en compte de l'espace mais des modèles différents. Pourquoi ?

Plan de la présentation

- Quelques justifications à la mise en place de ces modèles
- Notions de voisinage et d'autocorrélation spatiale
- Choix de modèle et interprétation des résultats
- Illustration à partir des taux de chômage locaux

Quelques justifications à la mise en place de ces modèles

- L'interaction spatiale, organisationnelle ou sociale des agents économiques est classique en économie : effets de voisinage, de pair, interactions stratégiques, copie par mimétisme ou par les normes sociales...
- Deux théories générales (Anselin 2002) :
 - La décision d'un agent économique (une entreprise par exemple) dépend de la décision des autres agents (ses concurrents) ;
 - La décision d'un agent économique dépend d'une ressource rare ;
 - Dans les 2 cas, la variable dépendante y_i s'explique par les autres agents y_{-i} et des variables explicatives x_i .
- Problème : ces deux théories amènent à implémenter un même modèle spatial ou d'interaction. Elles sont équivalentes d'un point de vue observationnel.

Quelques raisons économétriques (LeSage et Pace 2009)

- Corriger un biais de variable omise ;
- Améliorer le calcul de la précision ;
- Confronter plusieurs modèles d'économétrie spatiale pour discuter l'incertitude du processus générateur des données ;
- Justifier et généraliser certains choix effectués en économétrie "classique" (variables explicatives du type distance ou indicateurs agrégés par zone géographique).

Notions de voisinage et d'autocorrélation spatiale

- Si nous observons N régions, il y a $N \cdot (N - 1) / 2$ couples différents de régions

⇒ Il est nécessaire de définir *a priori* les relations de voisinage entre les agents ou les zones géographiques pour caractériser l'autocorrélation spatiale ;

- De nombreux choix possibles pour la matrice de voisinage W : contiguïté, inverse de la distance, k plus proches voisins ;
- Quand W est normée par ligne, Wy s'interprète simplement comme la moyenne du voisinage pour la variable y .
- Supposée exogène, i.e. indépendante des variables explicatives X .
- Choix empirique de définir plusieurs matrices pour tester la robustesse des analyses.

Matrices de voisinage



2 plus proches



5 plus proches



10 plus proches

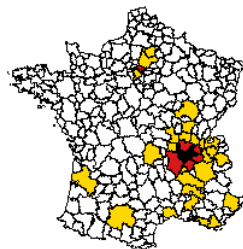
Matrices de voisinage



Contiguité



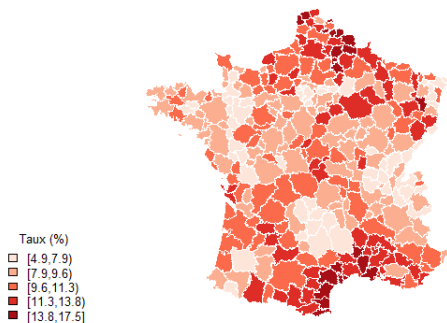
Distance Inverse



Flux domicile-travail

Illustrer l'autocorrélation ou l'hétérogénéité spatiale, à l'aide de cartes

Taux de chômage (2013)



Caractériser l'autocorrélation, à l'aide de tests

- Test de Moran

- Indicateur de Moran $I = \frac{N}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$;
- Une valeur positive sera synonyme d'autocorrélation spatiale positive ;
- Sous l'hypothèse H_0 d'absence d'autocorrélation spatiale ($I = 0$), la statistique $I^* = \frac{I - \mathbb{E}(I)}{\sqrt{\mathbb{V}(I)}}$ suit asymptotiquement une loi normale $\mathcal{N}(0, 1)$.

- Affiner l'analyse à l'aide d'indicateurs locaux (Floch 2012).

Choix de modèle et interprétation des résultats

3 types d'interaction spatiale :

- Une interaction endogène, i.e. que la décision économique d'un agent va dépendre de la décision de ses voisins ;
- Une interaction exogène, i.e. que la décision économique d'un agent va dépendre des caractéristiques observables de ses voisins ;
- Une corrélation spatiale des effets liée à de mêmes caractéristiques inobservées.

Le modèle de Manski

- Le modèle de Manski (1993) intègre ces 3 effets :

$$Y = \rho \cdot WY + X \cdot \beta + WX \cdot \theta + u$$

$$u = \lambda \cdot Wu + \varepsilon$$

Avec ρ l'interaction endogène, θ l'interaction exogène et λ l'interaction résiduelle.

- Problème : ce modèle n'est pas identifiable dans le cas général.
- Contraintes d'identification : contraindre un ou deux paramètres à 0.

La galaxie des modèles d'économétrie spatiale

- Deux modèles classiques :
 - Le modèle SAR (Spatial AutoRegression)
 $Y = \rho \cdot WY + X \cdot \beta + \varepsilon$, i.e. $\lambda = \theta = 0$
 - Le modèle SEM (Spatial Error Model) $Y = X \cdot \beta + u$ et
 $u = \lambda \cdot Wu + \varepsilon$, i.e. $\rho = \theta = 0$
- Le modèle actuellement plébiscité (LeSage et Pace 2009), le Spatial Durbin Model (SDM), $Y = \rho \cdot WY + X \cdot \beta + WX \cdot \theta + \varepsilon$ i.e. $\lambda = 0$
 - Intègre les cas particuliers des modèles SAR et SEM ;
 - Robuste à la présence d'autocorrélation résiduelle.
- De nombreux autres modèles.
- Estimation par maximum de vraisemblance des modèles dans le cas de variables exogènes.

Une approche itérative de choix de modèle (Elhorst 2010)

- Pas de règles intangibles ;
- A partir du modèle non spatial (MCO), tester la présence d'autocorrélation spatiale ;
- En cas d'autocorrélation, estimer le modèle SDM et tester la significativité des paramètres d'interactions ;
- En cas d'absence d'autocorrélation, estimer le modèle à interactions exogènes ($\theta \neq 0$) et tester la significativité des paramètres θ ;
- Estimer différents modèles permet d'étudier les différents types d'interactions.

L'interprétation des résultats : attention aux rétroactions

- Le modèle SEM s'interprète comme le modèle MCO ;
- Pour le modèle SAR $Y = \rho \cdot WY + X\beta + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, le modèle peut également s'écrire

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

- La valeur prédite est donc $\hat{y} = (1 - \hat{\rho} W)^{-1} X\hat{\beta}$ et non $X\hat{\beta}$ comme dans un modèle MCO classique.
- L'effet marginal (pour une variable quantitative) d'une modification de la variable X_{ir} (pour l'individu i et la variable r) n'est pas β_r mais la valeur diagonale de rang i de la matrice $(1 - \rho W)^{-1} \beta_r$.
- Une modification de mon territoire impacte mes voisins, ce qui m'impacte en retour.
- L'effet direct moyen (leSage et Pace 2009) correspond à la moyenne des termes diagonaux de la matrice $(1 - \rho W)^{-1} \beta_r$. C'est cet indicateur qui est le plus proche de l'interprétation des coefficients β calculés par MCO.

Pour aller plus loin

- Les modèles supposent une observation exhaustive de la population, ce qui est peu adapté aux données d'enquête ;
- Il peut exister de l'hétérogénéité spatiale, i.e. des déterminants différents par zone plus qu'un phénomène de diffusion ;
- Le découpage administratif ne correspond pas forcément à la réalité économique des relations entre agents (MAUP "Modifiable Areal Unit problem") ;
- L'interprétation des résultats n'est valable que pour le découpage géographique choisi (risque de régression dite écologique).

Illustration : Taux de chômage locaux

- Modélisant le taux de chômage localisé (par zone d'emplois) à l'aide de caractéristiques de
 - la population active (% des peu diplômés et des moins de 30 ans dans la population active) ;
 - la structure économique (% des emplois dans le secteur industriel) ;
 - du marché du travail (taux d'activité) ;
 - du dynamisme du marché du travail mesuré par la variation d'emploi.
- Modèle descriptif et illustratif ;
- Aucune conclusion causale possible ;
- Description des commandes du logiciel R.

Résultats

- Les critères statistiques conduiraient à retenir un modèle SDM ;
- Pour des raisons de parcimonie, le choix d'un modèle SEM pourrait être envisagé ;
- Le choix d'un modèle SAR serait ici déconseillé. Un test montre qu'une autocorrélation spatiale résiduelle reste présente. L'interprétation des résultats peut alors être erronée ;

Conclusion et perspectives

- Les modèles d'économétrie spatiale définissent un cadre cohérent (et paramétrique) pour modéliser tout type d'interactions entre agents économiques, pas uniquement géographiques ;
- La force de ces modèles est de mettre en avant si un problème "spatial" se pose et sous quelle forme. Tenir compte de ces interactions, c'est toujours mieux que de ne rien faire ;
- Le raffinement méthodologique doit être mis en regard de la complexité des nouveaux modèles, en matière d'interprétation.