Econométrie Spatiale

Plan

- Processus stochastiques spatiaux
  - Spécification de la covariance spatiale
  - SAR et SMA

- Modèles économétriques avec dépendance spatiale
  - Modèle à variable endogène spatialement décalée
  - Modèle à variables exogènes spatialement décalées
  - Modèle de Durbin spatial
  - Modèle à erreurs spatialement autocorrélées
  - Modèles d'ordre supérieur
  - Externalités spatiales

Plan (suite)

- **Hétérogénéité spatiale**
  - **Le problème**
  - **Hétérogénéité discrète**
  - **Hétérogénéité continue**

- **Estimation : maximum de vraisemblance**
  - **Propriétés des estimateurs des MCO**
  - **Analyse asymptotique spatiale**
  - **Méthodes d'estimation des modèles spatiaux**
    - **Principes généraux du Maximum de Vraisemblance**
    - **Modèle avec variable endogène spatialement décalée (SAR)**
    - **Modèle avec erreurs spatialement autocorrélées (SEM)**

Plan (suite)

- **Tests de spécification**
  - **Principes fondamentaux**
  - **Tests sur les résidus estimés par les MCO**
  - **Tests du multiplicateur de Lagrange contre l'hypothèse d'erreurs spatialement corrélées : LM_{ERR}**
  - **Tests du multiplicateur de Lagrange contre l'hypothèse d'une variable endogène spatialement décalé : LM_{LAG}**
  - **Tests robustes**
  - **Recherche de spécifications**

Processus stochastiques spatiaux

Spécification de la covariance spatiale

Processus stochastiques spatiaux

- Processus stochastiques spatiaux
  - $\{Z(i) : i \in D\}$
  - $i \in R^d$ localisation génériques des observations (vecteur de coordonnées)
  - $D \subset R^d$ ensemble des indices (sous-ensemble des localisations potentielles)

$\Rightarrow$ *données ponctuelles : D* est un processus ponctuel, *i* est aléatoire

$\Rightarrow$ *données géostatistiques : D* est un sous ensemble fixe de R^d (continu)

$\Rightarrow$ ***données régionales :*** *D* est un ensemble fixe de points dénombrables dans R^d

unités spatiales discrètes et finies / les observations sont localisées en des points fixes ou des lieux discrets (par exemple centroïde / PIB par tête régional)

- *Z* est une variable aléatoire, *z*(*i*) en est une réalisation en *i*

- Exemples
  - *i* représente les coordonnées *x, y* des transactions immobilières, *Z* représente le prix de vente en *i*
  - *i* représente les unités spatiales (comtés ou régions), *Z* est le taux de criminalité en *i*

Structure de dépendance spatiale

- Trois méthodologies pour modéliser l'autocorrélation spatiale

- Représentation directe
  - covariance entre deux observations comme fonction directe de la distance
    - $\operatorname{cov}(y_i, y_j) = f(\theta, d_{ij})$
    - covariance symétrique, définie positive

- **Modèles de processus spatiaux**
  - processus stochastique spatial $\{Z(i), i \in D\}$
  - la spécification du processus détermine la structure de la covariance

- Non-paramétrique
  - structure en blocs avec covariances communes
    - $\operatorname{cov}(y_i, y_j) = \sigma_h$ pour tout *i* et *j* à l'intérieur d'une bande de distance *h*

Représentation directe (1)

- Exemple : 5 points sur une droite : $x = 0, 1, 2, 3, 4$
- Matrice symétrique de distance de terme général d_{ij}

0	1	2	3	4
1	0	1	2	3
2	1	0	1	2
3	2	1	0	1
4	3	2	1	0

Représentation directe (2)

- Covariance spatiale : exponentielle négative
  - $\mathrm{cov}_{ij} = \exp(-0,3 \times d_{ij})$
  - vérifier que la matrice est définie positive

1	0,74082	0,54881	0,40657	0,30119
0,74082	1	0,74082	0,54881	0,40657
0,54881	0,74082	1	0,74082	0,54881
0,40657	0,54881	0,74082	1	0,74082
0,30119	0,40657	0,40657	0,74082	1

det = 0,0414

Covariances déterminées par le processus temporel AR(1)

- Expression du processus AR(1)
  - $y_t = \rho y_{t-1} + u_t \qquad u_t \sim i.i.d(0, \sigma_u^2)$
  - par substitution, on obtient :

$$y_t = \rho(\rho y_{t-2} + u_{t-1}) + u_t = \rho^2 y_{t-2} + \rho u_{t-1} + u_t$$

…

…

$$y_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots$$

$\Rightarrow$ série infinie en u_t

Structure de la covariance processus temporel AR(1)

- Structure de la covariance de y
  - **Hypothèses**
  - processus stationnaire en variance dans le temps :

$$V(u_t) = V(u_{t-h}) = \sigma_u^2 \text{ pour tout } h$$

  - $|\rho| < 1$ processus non explosif

$$V(y_t) = V(u_t) + \rho^2 V(u_{t-1}) + \rho^4 V(u_{t-2}) + \ldots \qquad \Rightarrow V(y_t) = \frac{\sigma_u^2}{1-\rho^2}$$

  - $\operatorname{cov}(y_t, y_{t-1}) = E(y_t y_{t-1}) = E[y_{t-1}(\rho y_{t-1} + u_t)]$

$$\operatorname{cov}(y_t, y_{t-1}) = \rho E(y_{t-1}^2) + E(u_t)$$

$$\operatorname{cov}(y_t, y_{t-1}) = \frac{\rho \sigma_u^2}{1-\rho^2} \quad \text{et} \quad \operatorname{cov}(y_t, y_{t-h}) = \frac{\rho^h \sigma_u^2}{1-\rho^2}$$

- Le processus AR(1) détermine complètement la structure de la covariance

Matrice de variances-covariances

- Exemple : $y_t = 0{,}3y_{t-1} + u_t$

1,0989	0,32967	0,0989	0,02967	0,0089
0,32967	1,0989	0,32967	0,0989	0,02967
0,0989	0,32967	1,0989	0,32967	0,0989
0,02967	0,0989	0,32967	1,0989	0,32967
0,0089	0,02967	0,0989	0,32967	1,0989

Covariances déterminées par le processus spatial AR(1) sur une droite

- Localisations sur une droite

i-2	***i-1***	***i***	***i+1***	***i+2***

- Les processus spatiaux sont plus complexes :
  - $y_i = \rho(y_{i-1} + y_{i+1}) + u_i$
  - par substitution, on obtient :
    - *i-1* dépend de *i-2* et de *i*, *i+1* dépend de *i* et de *i+2*
    - notons la présence de y_i (ce qui n'est pas le cas dans le domaine temporel)

$$y_i = \rho\left\{\left[\rho\left(y_{i-2} + y_i\right) + u_{i-1}\right] + \left[\rho\left(y_i + y_{i+2}\right) + u_{i+1}\right]\right\} + u_i$$

$$y_i = 2\rho^2 y_i + \rho^2 y_{i-2} + \rho^2 y_{i+2} + \rho u_{i-1} + \rho u_{i+1} + u_i$$

Structure de la covariance processus spatial AR(1) sur une droite

- Structure de la covariance de y
  - **Hypothèses**
    - processus stationnaire en variance dans l'espace :

$$V(y_{i+h}) = V(y_{i-h}) \text{ pour } h = 1,2$$

    - $|\rho| < 0.577$ processus non explosif
    - l'espace du paramètre dépend du nombre de voisins

$$V(y_i) = \rho^2 V(y_{i-1} + y_{i+1}) + \sigma_u^2 + 2\,\text{cov}\left[\left(y_{i-1} + y_{i+1}\right), u_i\right]$$

$$V(y_i) = \rho^2\left[V(y_{i-1}) + V(y_{i+1}) + 2\,\text{cov}(y_{i-1}, y_{i+1})\right] + \sigma_u^2 + 2\,\text{cov}\left[\left(y_{i-1} + y_{i+1}\right), u_i\right]$$

- $\text{cov}(y_{i-1}, y_{i+1}) \neq 0 \qquad \text{cov}(y_{i-1}, u_i) \neq 0 \qquad \text{cov}(y_{i+1}, u_i) \neq 0$
  - y_{i-1} et y_{i+1} contiennent tous deux y_i, par conséquent $E(y_{i-1}y_{i+1}) \neq 0, E(y_{i\pm1}u_i) \neq 0$
  - différent du cas temporel : pas de forme récursive simple
  - mais la structure de la covariance est entièrement déterminée par le processus AR(1) spatial

Processus spatial AR(1) sur une droite

- Modélisation avec une matrice de pondération spatiale $\boldsymbol{W}$:

$$y = \rho W y + u, \text{ ou } y = (I - \rho W)^{-1} u$$

avec $\boldsymbol{W}$ une matrice de contiguïté de dimension (5×5)

0	1	0	0	0
1	0	1	0	0
0	1	0	1	0
0	0	1	0	1
0	0	0	1	0

Matrice de variances-covariances

- Matrice de variances-covariances de y

$$V(y) = E(yy') = \left(I - \rho W\right)^{-1} E(uu')\left(I - \rho W'\right)^{-1}$$

$$E(uu') = \sigma_u^2 I \quad \text{avec } \sigma_u^2 = 1; \quad \rho = 0,3 \quad \Rightarrow \quad V(y) = (I - 0,3W)^{-1}(I - 0,3W')^{-1}$$

1,3887	0,92523	0,46106	0,20068	0,072398
0,92523	1,8497	1,1259	0,53346	0,20068
0,46106	1,1259	1,9221	1,1259	0,46106
0,20068	0,53346	1,1259	1,8497	0,92523
0,072398	0,20068	0,46106	0,92523	1,3887

1	0,3	0,09	0,027	0,0081
0,3	1	0,3	0,09	0,027
0,09	0,3	1	0,3	0,09
0,027	0,09	0,3	1	0,3
0,0081	0,027	0,09	0,3	1

Matrice de corrélation pour le processus temporel AR(1)

$$\sigma_u^2 = 1 \text{ et } \rho = 0,3$$

$$\operatorname{cor}(y_t, y_{t-h}) = \frac{\operatorname{cov}(y_t, y_{t-h})}{\sqrt{\operatorname{var}(y_t)\operatorname{var}(y_{t-h})}} = \rho^h$$

$$\sigma_u^2 = 1 \text{ et } \rho = 0,3$$

Matrice de corrélation pour le processus spatial AR(1)

$$\operatorname{cor}(y_i, y_j) = \frac{\operatorname{cov}(y_i, y_j)}{\sqrt{\operatorname{var}(y_i)\operatorname{var}(y_j)}}$$

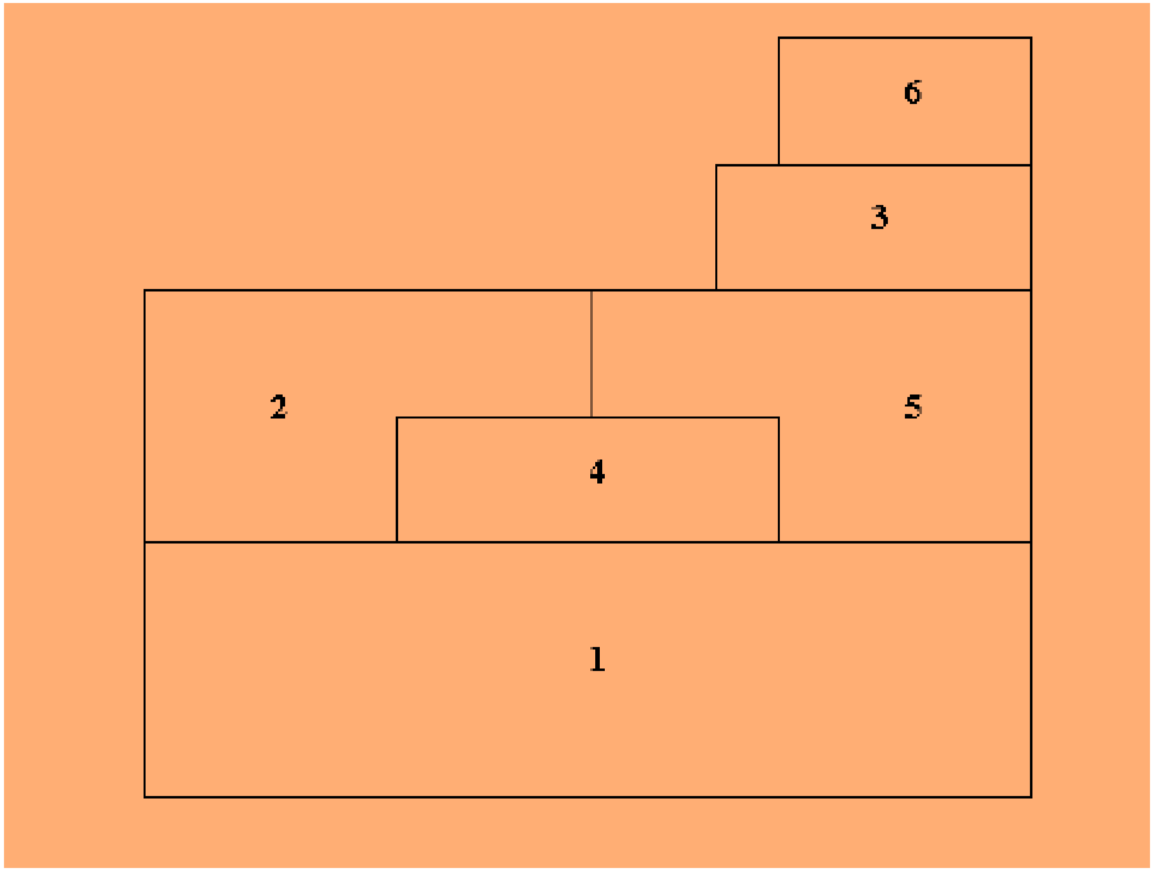
1	0,5773	0,28221	0,12522	0,052135
0,5773	1	0,59712	0,2884	0,12522
0,28221	0,59712	1	0,59712	0,28221
0,12522	0,2884	0,59712	1	0,5773
0,052135	0,12522	0,28221	0,5773	1

AR(1) spatial versus AR(1) temporel

- Espace
  - Variances non constantes et plus grandes
  - Covariances plus grandes
  - Corrélations plus grandes
  - Covariances liées au décalage spatial, mais ne prennent pas la même valeur pour tous les voisins de même ordre
- Temps
  - Variance constante
  - Covariances/corrélations identiques pour le même décalage temporel

Processus stochastiques spatiaux

SAR et SMA



exemple pour N = 6 unités spatiales

contiguïté représentée sur un plan ou une carte

contiguïté = frontière commune

Matrice de contiguïté du premier ordre

0	1	0	1	1	0
1	0	0	1	1	0
0	0	0	0	1	1
1	1	0	0	1	0
1	1	1	1	0	0
0	0	1	0	0	0

Modèle autorégressif spatial - SAR

- Spécification
  - supposons sans perte de généralité que $E(y)=0$
  - $y=\rho Wy+u$ et $u\sim(0,\sigma_u^2 I)$
  - $(I-\rho W)y=u \quad \Rightarrow \quad y=(I-\rho W)^{-1}u$

- Structure de la covariance
  - $E(yy')=\left(I-\rho W\right)^{-1}E(uu')\left(I-\rho W'\right)^{-1}$

$$E(yy')=\sigma_u^2\left(I-\rho W\right)^{-1}\left(I-\rho W'\right)^{-1}$$

$$E(yy')=\sigma_u^2\left[\left(I-\rho W\right)'\left(I-\rho W\right)\right]^{-1}$$

Covariance - SAR

$$\rho = 0,2 \; ; \; \sigma^2 = 1$$

2,1436	1,4492	0,41829	1,4492	1,5316	0,10197
1,4492	2,1436	0,41829	1,4492	1,5316	0,10197
0,41829	0,41829	1,3565	0,41829	0,79701	0,49108
1,4492	1,4492	0,41829	2,1436	1,5316	0,10197
1,5316	1,5316	0,79701	1,5316	2,3971	0,21435
0,10197	0,10197	0,49108	0,10197	0,21435	1,1422

Corrélation - SAR

$\rho = 0,2 \; ; \; \sigma^2 = 1$

1	0,67604	0,2453	0,67604	0,67567	0,06517
0,67604	1	0,2453	0,67604	0,67567	0,06517
0,2453	0,2453	1	0,2453	0,44199	0,39453
0,67604	0,67604	0,2453	1	0,67567	0,06517
0,67567	0,67567	0,44199	0,67567	1	0,12954
0,06517	0,06517	0,39453	0,06517	0,12954	1

Modèle moyenne mobile spatial - SMA

- Spécification
  - $y = \rho W u + u$
  - $y = (I + \rho W) u$

- Structure de la covariance
  - $$E(yy') = \left(I + \rho W\right) E(uu') \left(I + \rho W\right)'$$
    $$E(yy') = \sigma_u^2 \left(I + \rho W\right)\left(I + \rho W'\right)$$
    $$E(yy') = \sigma_u^2 \left[I + \rho(W + W') + \rho^2 W W'\right]$$

Covariance - SMA

$$\rho = 0,2\ ;\ \sigma^2 = 1$$

1,12	0,48	0,04	0,48	0,48	0
0,48	1,12	0,04	0,48	0,48	0
0,04	0,04	1,08	0,04	0,4	0,4
0,48	0,48	0,04	1,12	0,48	0
0,48	0,48	0,4	0,48	1,16	0,04
0	0	0,4	0	0,04	1,04

Corrélation - SMA

$$\rho = 0,2 \; ; \; \sigma^2 = 1$$

1	0,42857	0,03637	0,42857	0,42112	0
0,42857	1	0,03637	0,42857	0,42112	0
0,03637	0,03637	1	0,03637	0,35737	0,37743
0,42857	0,42857	0,03637	1	0,42112	0
0,42112	0,42112	0,35737	0,42112	1	0,036418
0	0	0,37743	0	0,036418	1