



ÉCONOMÉTRIE SPATIALE

Master 2 APE

Chapitre 4

Gilles de Truchis

gilles.detruchis@gmail.com

Site : www.varennnes-ecofin.com/



LES CHAPITRES DU COURS

1. Modèles spatiaux avec hétérogénéité

1.1 Introduction à l'hétérogénéité

1.2 Traitement de l'hétérogénéité

2. Annexe sur les techniques d'estimation

2.1 Rappels sur les estimateurs usuels

2.2 Propriétés des estimateurs

3. Conclusion

MODÈLES SPATIAUX AVEC HÉTÉROGÉNÉITÉ

CONTENTS

1. Modèles spatiaux avec hétérogénéité

1.1 Introduction à l'hétérogénéité

1.2 Traitement de l'hétérogénéité

2. Annexe sur les techniques d'estimation

2.1 Rappels sur les estimateurs usuels

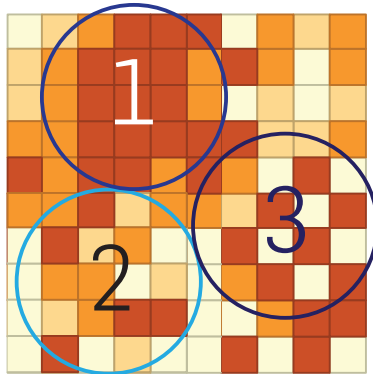
2.2 Propriétés des estimateurs

3. Conclusion

LE CONCEPT D'HÉTÉROGÉNÉITÉ

- ▶ Dans l'espace, des schémas différents peuvent apparaître
- ⇒ e.g. l'autocorrélation peut être uniquement locale et non globale
- ▶ L'hétérogénéité va transparaître à travers différents symptômes
- ⇒ Instabilité structurelle
 - ▶ les coefficients du modèles sont différents selon la localisation
- ⇒ Hétéroscédasticité
 - ▶ les variances des termes d'erreurs sont différentes selon la localisation (modèle mal spécifié)

INTUITION DE L'HÉTÉROGÉNÉITÉ



- Dans l'espace, des schémas différents apparaissent
 - Dans la zone 1 : autocorrélation positive
 - Dans la zone 2 : autocorrélation nulle
 - Dans la zone 3 : autocorrélation négative

PROBLÈMES LIÉS À L'HÉTÉROGÉNÉITÉ

- ▶ L'estimation d'un schéma global est plus simple
 - ▶ modèle unique impliquant une fonction des paramètres stable dans l'espace
 - ▶ L'estimation de schémas locaux est plus complexe
 - ▶ modèles multiples impliquant non-linéarité ou non-stationnarité au niveau global
- ⇒ Potentiellement, il faudrait estimer n modèles :

$$y_i = \sum_{j=1}^k x_{ij} \beta_{ij} + \varepsilon_i$$

- ▶ Problème : on dispose de n observation pour estimer $n \times k$ coefficients β

SOLUTION POUR L'HÉTÉROGÉNÉITÉ

► Imposer des variations discrètes :

- dans la moyenne : approche ANOVA (analysis of variance)
- dans les paramètres : régimes spatiaux

► Imposer des variations continues :

- dans la moyenne : analyse de la tendance de surface (ATS)
- plus générale : méthode d'expansion spatiale de Casetti
- dans les paramètres : paramètres variants spatialement (geographically weighted regression : GWR)

CONTENTS

1. Modèles spatiaux avec hétérogénéité

1.1 Introduction à l'hétérogénéité

1.2 Traitement de l'hétérogénéité

2. Annexe sur les techniques d'estimation

2.1 Rappels sur les estimateurs usuels

2.2 Propriétés des estimateurs

3. Conclusion

ANOVA SPATIALE

- Soit $Y = \{y_i, i \in I\}$ un échantillon scindé arbitrairement en deux sous-échantillons G_1 et G_2
- On suppose G_1 et G_2 stationnaire et

$$\mathbb{E}(y_i) = \mu_1, \quad i \in G_1$$

$$\mathbb{E}(y_i) = \mu_2, \quad i \in G_2$$

- On peut alors tester la présence d'hétérogénéité dans la moyenne
 - ⇒ Sous l'hypothèse nulle que $\mu_1 = \mu_2$ il y a homogénéité spatiale
 - ⇒ Sous l'hypothèse alternative que $\mu_1 \neq \mu_2$ il y a hétérogénéité spatiale
- Approche alternative : on utilise un modèle de régression avec indicatrices

$$y_i = \mu_1 d_{1i} + \mu_2 d_{2i} + \varepsilon_i$$

$$d_{1i} = 1, \quad \forall i \in G_1, \quad d_{1i} = 0, \quad \forall i \notin G_1$$

$$d_{2i} = 1, \quad \forall i \in G_2, \quad d_{2i} = 0, \quad \forall i \notin G_2$$

RÉGIMES SPATIAUX

- On considère deux régimes pour les paramètres structurels du modèle

$$y_i = \sum_{j=1}^k x_{ij}\beta_{1j} + \varepsilon_i, \quad i \in G_1$$

$$y_i = \sum_{j=1}^k x_{ij}\beta_{2j} + \varepsilon_i, \quad i \in G_2$$

- On peut alors tester la présence d'hétérogénéité
 - ⇒ Sous l'hypothèse nulle que $\beta_{1,.} = \beta_{2,.}$ il y a homogénéité spatiale
 - ⇒ Sous l'hypothèse alternative que $\beta_{1,.} \neq \beta_{2,.}$ il y a hétérogénéité spatiale
- Si ε_i i. i. d. $(0, \sigma_2)$, cela revient à effectuer un test de Chow

ANALYSE DE LA TENDANCE DE SURFACE (ATS)

► Soit $Z = \{z_i, i \in I\}$ et $I \in \mathbb{R}^2$

⇒ les coordonnées de z_i sont des couples (x_i, y_i)

► L'ATS consiste à régresser Z sur une expansion polynomiale de ces coordonnées

$$z_i = \alpha + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 x_i y_i + \beta_5 y_i^2 + \dots + \varepsilon$$

► Ce modèle permet de révéler les grandes tendances (e.g. Nord-Sud, Est-Ouest) de la surface de régression

► Si les coefficients sont significatifs, ils révèlent une **instabilité sur les paramètres**

► Limites :

⇒ La méthode est sujette à la multi-colinéarité

⇒ Le choix de l'ordre du polynôme est délicat

⇒ Capture uniquement des grandes tendances et non les variations abruptes locales

MÉTHODE DE CASETTI

- ▶ Cette méthode généralise la méthode ATS

- ▶ Repartons d'un modèle du type

$$y_i = x_i\beta_i + \varepsilon_i$$

- ▶ Il n'est pas possible d'estimer les n coefficients β_i , Casetti propose une alternative
⇒ β_i sera déterminé par une fonction de variables auxiliaires

$$\beta_i = f(z_i, \gamma)$$

- ▶ Si z est un vecteur de 2 variables, une spécification simple est

$$\beta_i = \gamma_0 + \gamma_1 z_{1,i} + \gamma_2 z_{2,i}$$

- ▶ Souvent, les variables z seront les coordonnées des observations (méthode ATS)
⇒ Capture uniquement des grandes tendances et non les variations abruptes locales
⇒ Le choix de l'expansion est un problème

MÉTHODE DE CASETTI

- ▶ La méthode de Casetti peut être généralisée au cas d'une expansion aléatoire
- ▶ La spécification de β comprend donc un terme d'erreur

$$\beta_i = \gamma_0 + \gamma_1 z_{1,i} + \gamma_2 z_{2,i} + u_i$$

- ▶ Une fois intégré dans le modèle initial, cela donne le modèle final suivant

$$y_i = \gamma_0 x_i + \gamma_1 (z_i x_i) + \gamma_1 (z_i x_i) + v_i$$

$$v_i = \varepsilon_i + u_i x_i$$

- ▶ En découle l'hétéroscédasticité du processus car

$$\mathbb{V}(v_i) = \mathbb{V}(\varepsilon_i) + x_i^2 \mathbb{V}(u_i)$$

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

- ▶ La GWR permet de dépasser les limites des approches précédentes
 - ⇒ Approche non-paramétrique
 - ⇒ Capture pour chaque y_i les variations des coefficients dans l'espace
- ▶ Repartons du modèle hétérogène initiale :

$$y_i = \sum_{j=1}^k x_{ij} \beta_{ij} + \varepsilon_i$$

- ▶ On suppose que les données autour de i ont plus d'influence dans l'estimation de β_{ij}
 - ▶ On intègre dans l'estimation une pondération V_i pour en tenir compte
 - ⇒ On obtient des MCO pondérés

$$\hat{\beta}_i = (X' V_i X)^{-1} X' V_i Y$$

$$V_i = \text{diag}(v_{i1}, \dots, v_{in})$$

- ▶ v_{ik} : poids de l'observation k sur l'estimation du modèle autour de i

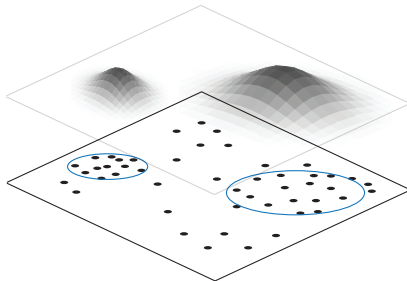
GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

► La construction des V_i peut se faire sur la base de divers méthodes

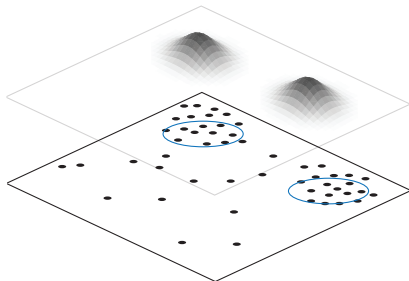
► Noyau fixe Gaussien : $V^{(GFK)} = \exp(-\frac{1}{2}(\frac{d_{ij}}{b})^2)$

► b représente une bande de distance

► Noyau fixe bi-square : $V^{(BFK)} = (1 - (\frac{d_{ij}}{b})^2)^2$ si $d_{ij} < b$ et $V = 0$ sinon



Gaussian Adaptive Kernel

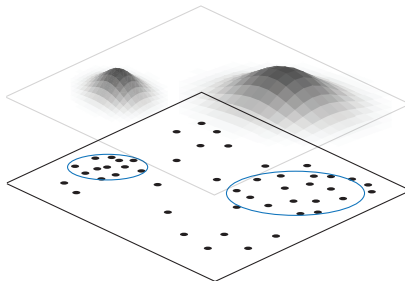


Gaussian Fixed Kernel

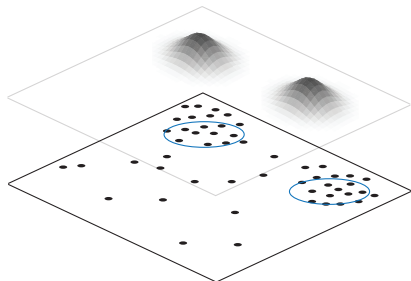
GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

► Les méthodes fixed kernel ne sont pas toujours adaptées \Rightarrow méthodes adaptatives

- Noyau adaptatif Gaussien : $V^{(GAK)} = \exp(-R_{ij}/b)$
 - R_{ij} est le classement du point j en terme de distance par rapport à i
- DNN : $V^{(DNN)} = 1$ si j est un des k plus proches voisins et $V = 0$ sinon
- CNN : $V^{(CNN)} = V^{(BFK)}$ si j est un des k plus proches voisins et $V = 0$ sinon



Gaussian Adaptive Kernel



Gaussian Fixed Kernel

ANNEXE SUR LES TECHNIQUES D'ESTIMATION

CONTENTS

1. Modèles spatiaux avec hétérogénéité

1.1 Introduction à l'hétérogénéité

1.2 Traitement de l'hétérogénéité

2. Annexe sur les techniques d'estimation

2.1 Rappels sur les estimateurs usuels

2.2 Propriétés des estimateurs

3. Conclusion

PRINCIPE DU MAXIMUM DE VRAISEMBLANCE

- ▶ Partons d'un exemple : soit un échantillon $X_t = X_1, \dots, X_n \sim P(\theta)$
 - ▶ $P(\theta)$ dénote la distribution de Poisson dont la fonction de masse est

$$\Pr(X_i = x) = \frac{\exp(-\theta)\theta^x}{x!}, \quad \theta > 0, \quad \forall x \in \mathbb{N}$$

- ▶ Soit une réalisation de l'échantillon $x_t = x_1, \dots, x_n$

- ▶ La probabilité d'observer cette réalisation est

$$\Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n))$$

- ▶ L'indépendance des tirages donne l'équivalence avec le produit des probabilités marginales

$$\Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n)) = \prod_{i=1}^n \Pr(X_i = x_i)$$

L'ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

- En remplaçant par la fonction de masse de la loi de Poisson on obtient

$$\Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n)) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = e^{-n\theta} \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

- Il s'agit donc d'une fonction dépendant de x_1, \dots, x_n et de θ
- θ est un paramètre inconnu mais on observe x_1, \dots, x_n
- Par la suite on notera :

$$L_n(\theta; x_1, \dots, x_n) = \Pr((X_1 = x_1) \cap \dots \cap (X_n = x_n))$$

- Le principe du **maximum de vraisemblance** est le suivant :
 - Trouver le θ qui maximise la probabilité d'apparition de x_1, \dots, x_n

- L'**estimateur du maximum de vraisemblance** est donc :

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^+} L_n(\theta; x_1, \dots, x_n)$$

L'ESTIMATEUR DU MAXIMUM DE LA LOG-VRAISEMBLANCE

- Dans l'exemple reposant sur la loi de Poisson on a

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^+} e^{-n\theta} \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

- La formule est complexe et la présence d'un produit n'arrange rien
- Simplifions le programme de maximisation en considérant la **log-vraisemblance**

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^+} \ln L_n(\theta; x_1, \dots, x_n)$$

- Dans le cadre de notre exemple la **log-vraisemblance** est

$$\ln L_n(\theta; x_1, \dots, x_n) = -n\theta + \ln(\theta) \sum_{i=1}^n x_i - \ln \left(\prod_{i=1}^n x_i! \right)$$

MODÈLE DE RÉGRESSION ET MLE

- Reprenons le modèle de régression multiple

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \Leftrightarrow \varepsilon_i = y_i - \beta_0 - \sum_{k=1}^p \beta_k x_{ik}$$

- Le MLE requiert une hypothèse sur la loi des erreurs : $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$
- La vraisemblance associée à la densité de la loi normale est donc :

$$\begin{aligned} L_n(\theta; \varepsilon_1, \dots, \varepsilon_n) &= \prod_{i=1}^n (\sigma_\varepsilon \sqrt{2\pi})^{-1} e^{-\varepsilon_i^2 / (2\sigma_\varepsilon^2)} \\ &= (2\pi\sigma_\varepsilon^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n \varepsilon_i^2}{2\sigma_\varepsilon^2}\right) \\ \ln L_n(\theta; \varepsilon_1, \dots, \varepsilon_n) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \varepsilon_i^2 \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\Sigma) - \frac{1}{2} \varepsilon' \Sigma^{-1} \varepsilon, \quad \Sigma = \sigma_\varepsilon^2 I \end{aligned}$$

- Le MLE du modèle est donc $\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^+} \ln L_n(\theta; \varepsilon_1, \dots, \varepsilon_n)$

CONTENTS

1. Modèles spatiaux avec hétérogénéité

1.1 Introduction à l'hétérogénéité

1.2 Traitement de l'hétérogénéité

2. Annexe sur les techniques d'estimation

2.1 Rappels sur les estimateurs usuels

2.2 Propriétés des estimateurs

3. Conclusion

LES MCO DANS LES MODÈLES DE TYPE SAR

- Rappelons que les modèles de type SAR sont de la forme :

$$\begin{aligned} Y &= \rho WY + \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I) \\ &= (I - \rho W)^{-1} \beta X + (I - \rho W)^{-1} \varepsilon \end{aligned}$$

- Observons que WY est corrélée avec ε même si $\varepsilon \sim \text{i.i.d.}$ car

$$\begin{aligned} \mathbb{E}(WY\varepsilon') &= \mathbb{E}(W(I - \rho W)^{-1}(X\beta + \varepsilon)\varepsilon') \\ &= W(I - \rho W)^{-1} \mathbb{E}((X\beta + \varepsilon)\varepsilon') = \sigma_\varepsilon^2 W(I - \rho W)^{-1} \end{aligned}$$

- On comprend alors pourquoi l'estimateur **MCO est inconsistant**
 - résultat différent des séries temporelles car le décalage spatial introduit un biais de simultanéité
 - si ρ connu, β peut s'estimer par MCO

LES MCO DANS LES MODÈLES DE TYPE SEM

- Rappelons que les modèles de type SEM sont de la forme :

$$Y = \beta X + u,$$
$$u = \lambda W u + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$$

- Observons que dans le modèle SEM, les erreurs ne sont plus sphériques

$$\mathbb{E}(uu') = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1} = \Omega$$

- Comme Ω n'est pas diagonale, l'estimateur **MCO est non-biaisé mais inefficace**
⇒ l'estimateur n'atteint pas la borne de Cramer-Rao
- si λ connu, β peut s'estimer par MCO

LE MLE DANS LES MODÈLES DE TYPE SAR

- On supposera la normalité des erreurs dans le SAR :

$$Y = \rho WY + \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$$

- ε étant inobservé, comment construire $L_n(\theta; \varepsilon)$?

- On utilise la transformation $f(Y) = \varepsilon = (I - \rho W)Y - X\beta$
- Pour tenir compte de la contribution à $L_n(\theta; \varepsilon)$ apportée par l'observation t dans la transformation, on inclura le déterminant du jacobien de la transformation

$$\det \left(\frac{\partial f(Y)}{\partial Y} \right) = \det(I - \rho W) = \det(J)$$

- La log-vraisemblance normale est donc

$$\ln L_n(\theta; \varepsilon) = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \ln \det(J) - \frac{1}{2\sigma_\varepsilon^2} \left((I - \rho W)Y - X\beta \right)' \left((I - \rho W)Y - X\beta \right)$$

- On a $\ln \det(J) = \ln \sum_i (1 - \rho \omega_i)$ avec ω_i les valeurs propres de W

- Espace du paramètre $\rho : \omega_{\min}^{-1} < \rho < \omega_{\max}^{-1}, \quad \omega_{\max} = 1 \text{ si } W^{(s)}$

LE MLE DANS LES MODÈLES DE TYPE SEM

- On supposera la normalité des erreurs dans le SEM :

$$Y = \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \Omega(\varphi))$$

- En effet, on sait que $\mathbb{E}(\varepsilon \varepsilon') = \sigma_\varepsilon^2 \Omega(\varphi)$
 - φ représente des paramètres de nuisance
 - On suppose φ indépendant de β et que $\Omega(\varphi)$ est bloqué diagonale

- La log-vraisemblance normale est donc

$$\ln L_n(\theta; \varepsilon) = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \ln \det(\Omega(\varphi)) - \frac{1}{2\sigma_\varepsilon^2} (Y - X\beta)' (Y - X\beta)$$

- Pour φ connu, le MLE est équivalent aux MCG

$$\hat{\beta}(\varphi) = (X' \Omega(\varphi)^{-1} X)^{-1} X' \Omega(\varphi)^{-1} Y$$

- Pour φ inconnu, il nous faut un estimateur convergent

LE MLE DANS LES MODÈLES DE TYPE SEM

- Si le modèle a des erreurs de type AR spatial

$$\varepsilon = \lambda W\varepsilon + u$$

- On en déduit $\mathbb{E}(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \left((I - \lambda W)'(I - \lambda W) \right)^{-1}$

⇒ nécessite un estimateur de λ car $\Omega(\lambda)^{-1} = (I - \lambda W)'(I - \lambda W)$

- Si le modèle a des erreurs de type MA spatial

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

- On en déduit $\mathbb{E}(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \left((I + \lambda W)'(I + \lambda W) \right)$

⇒ nécessite une inversion donc pas de forme analytique pour
 $\Omega(\lambda)^{-1} = \left(I + \lambda(W + W') + \lambda^2 WW' \right)^{-1}$

LE MLE DANS LES MODÈLES DE TYPE SEM

- Supposons que le modèle a des erreurs de type AR spatial est donc

$$Y = \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \left((I - \lambda W)'(I - \lambda W) \right)^{-1})$$

- ε étant inobservé :

- On utilise la transformation $f(Y) = \varepsilon_\lambda = (I - \lambda W)(Y - X\beta)$
- Le déterminant du jacobien de la transformation est alors

$$\det \left(\frac{\partial f(Y)}{Y} \right) = \det(I - \lambda W) = \det(J)$$

- La log-vraisemblance normale est donc

$$\ln L_n(\theta; \varepsilon) = -\frac{n}{2} \ln(2\pi\sigma_\varepsilon^2) - \ln \det(J) - \frac{1}{2\sigma_\varepsilon^2} \varepsilon'_\lambda \varepsilon_\lambda$$

- On a $\ln \det(J) = \ln \sum_i (1 - \lambda \omega_i)$ avec ω_i les valeurs propres de W
- Espace du paramètre $\lambda : \omega_{\min}^{-1} < \lambda < \omega_{\max}^{-1}$

MESURE DU POUVOIR EXPLICATIF DU MLE

- ▶ Le coefficient de détermination R^2 n'est pas exploitable
 - ▶ Il repose sur la somme des carrés des erreurs non pondérée
 - ▶ Il ne tient pas compte de la perte d'information due à la dépendance
- ▶ On peut utiliser les critères d'information
 - ▶ AIC (Akaike Information Criterion) : avec K le nombre de paramètres du modèle
$$AIC = -2 \ln(L_n) + 2K$$
 - ▶ BIC (Bayesian Information Criterion) : avec K le nombre de paramètres du modèle
$$BIC = -2 \ln(L_n) + \ln(n)K$$
 - ▶ HQS (Hannan, Quinn, Schwarz) : avec K le nombre de paramètres du modèle
$$HQS = -2L_n + q(K), \quad q = \ln(nK) \text{ or } q = 2 \ln(\ln(nK))$$

CONCLUSION

ET MAINTENANT...

- ▶ On passe de la théorie à la pratique