



Cursus intégré

Année 2006

AIDE-MEMOIRE
D'ECONOMETRIE

Alain Trognon et Jean-Marc Fournier

Chapitre 1 : LE MODELE LINEAIRE

1. Définitions :

1.1 Définition intrinsèque :

Définition 1 : On appelle modèle linéaire, le modèle statistique $(R^N, \mathcal{R}^N, \mathcal{P})$ où \mathcal{P} est une famille de lois de probabilité telles que : $\exists L$, sous espace vectoriel de $R^N / \forall P \in \mathcal{P}, \exists(m, \sigma^2) \in L \times (R^+ - \{0\})$:

i) $E_P Y = m, m \in L,$

ii) $V_P Y = \sigma^2 I_N,$

iii) l'application qui à toute $P \in \mathcal{P}$ associe (m, σ^2) est une application surjective de \mathcal{P} sur $L \times (R^+ - \{0\})$.

Remarques : a) modèle est semi-paramétrique

b) si e , le vecteur « unitaire » dont toutes les composantes sont égales à 1 sur la base canonique, appartient à L , on dit que le modèle comporte un terme constant.

c) si le modèle comporte un terme constant et si $\dim L = 2$ le modèle est un modèle linéaire simple.

1.2 Définition classique :

Définition 1' : Le modèle $(R^N, \mathcal{R}^N, \mathcal{P})$ est un modèle linéaire s'il existe une matrice $X(N \times K)$ de plein rang, telle que : $\forall P \in \mathcal{P}, \exists(b, \sigma^2) \in R^K \times (R^+ - \{0\})$:

i) $E_P Y = Xb$

ii) $V_P Y = \sigma^2 I_N$

iii) l'application qui à toute $P \in \mathcal{P}$ associe (b, σ^2) est une application surjective de \mathcal{P} sur $R^K \times (R^+ - \{0\})$.

Ecriture matricielle :

$$\begin{cases} y = Xb + u, \text{rg}X = K \\ Eu = 0, Vu = \sigma^2 I_N \end{cases}$$

Multicolinéarité : $\text{rg}X < K$

2. Modèles linéaires, modèles conditionnels :

2.1 La régression probabiliste :

Soit le modèle d'échantillonnage gaussien sur R^{K+1} :

$$\left(R^{K+1}, R^{K+1}, N \left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right) \right)^{\otimes N}$$
$$\mu \in R, \nu \in R^K$$

L'échantillon est constitué de vecteurs d'observations indépendantes $(y_n, x_n) \in R^{K+1}$. La loi de $Y_n / X_n = x_n$ est une loi normale d'espérance :

$$E(Y_n / X_n = x_n) = \mu + \Sigma_{yx} \Sigma_{xx}^{-1} (x_n - \nu)$$

et de variance :

$$V(Y_n / X_n = x_n) = \sigma_y^2 - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}$$

2.2 Le modèle linéaire conditionnel :

Un modèle pour lequel il existe a, b_1, b_2, \dots, b_K tels que :

$$E(Y_n / X_1 = x_1, X_2 = x_2, \dots, X_N = x_N) = a + \sum_{k=1}^K x_{kn} b_k$$

$$V(Y_n / X_1 = x_1, X_2 = x_2, \dots, X_N = x_N) = \sigma^2$$

$$\text{cov}(Y_n, Y_{n'} / X_1 = x_1, X_2 = x_2, \dots, X_N = x_N) = 0, \forall n \neq n'$$

est un modèle linéaire conditionnel.

3.Représentations géométriques :

3.1 L'espace des variables : R^{K+1}

3.2 L'espace des observations : R^N

4. La méthode des moindres carrés ordinaires :

4.1 Cas de la définition intrinsèque :

Le modèle postule que $E_p Y = m \in L$, si y est le vecteur des observations la méthode des MCO propose d'estimer m par \hat{m} défini par :

$$\arg \min_{m \in L} (\|y - m\|^2)$$

en conséquence : $\hat{m} = P_L y$, où P_L est le projecteur orthogonal L .

4.2 Cas de la définition classique :

Le modèle postule que $E_p Y = Xb$. L est alors l'espace linéaire engendré par les vecteurs colonnes de X ; $L = L(X)$.

L'estimateur des MCO \hat{b} de b réalise le minimum de :

$$Q(b) = \|y - Xb\|^2$$

4.3 Les équations normales :

Lorsque X est de plein rang la fonction $Q(b)$ est strictement concave et \hat{b} est défini par les conditions du premier ordre :

$$\frac{\partial Q}{\partial b}(\hat{b}) = -2X'(y - X\hat{b}) = 0$$

Ce système porte le nom d'équations normales écrit sous la forme :

$$X'(y - X\hat{b}) = 0$$

Puisque X est de plein rang, $X'X$ est inversible et $\hat{b} = (X'X)^{-1} X'y$.

4.4 Terminologie :

- mettre en œuvre la méthode des MCO se dit très souvent : « faire les MCO de y sur X » ou encore « régresser y sur X ».
- $\hat{y} = \hat{m} = X\hat{b}$ est le vecteur des « valeurs ajustées », ou « prédites ».
- $\hat{u} = y - \hat{y}$ est le vecteur des « résidus ».
- $\hat{m} = \hat{y} = P_L y$, $P_L = X(X'X)^{-1}X'$, projecteur orthogonal sur $L = L(X)$.
- $\hat{u} = y - \hat{y} = (I - P_L)y = (I - X(X'X)^{-1}X')y = M_L y$, $M_L = I - X(X'X)^{-1}X'$ est le projecteur orthogonal sur L^\perp .
- l'équation $y = \sum_{k=1}^K x_k \hat{b}_k$ est l'équation de l'hyperplan de régression dans R^{K+1} .
- $Q(\hat{b}) = \|y - \hat{m}\|^2 = \|y - \hat{y}\|^2 = \|y - X\hat{b}\|^2 = \|\hat{u}\|^2 = \sum_{n=1}^N \hat{u}_n^2 = SCR$, « somme des carrés des résidus ».
- $N - \dim L = N - K =$ nombre de degrés de liberté.

5. Propriétés algébriques de la méthode des MCO.

5.1 Modèle avec terme constant : $e \in L$

- la somme des résidus est nulle : $e' \hat{u} = \sum_{n=1}^N \hat{u}_n = 0$
- $\bar{y} = \frac{1}{N} \sum_{n=1}^N y_n = \bar{\hat{y}} = \frac{1}{N} \sum_{n=1}^N \hat{y}_n$
- Equation d'analyse de la variance :

$$\|y - \bar{y}e\|^2 = \|\hat{y} - \bar{\hat{y}}e\|^2 + \|\hat{u}\|^2$$

$$\frac{1}{N} \sum_{n=1}^N (y_n - \bar{y})^2 = \frac{1}{N} \sum_{n=1}^N (\hat{y}_n - \bar{\hat{y}})^2 + \frac{1}{N} \sum_{n=1}^N \hat{u}_n^2$$

« variance totale = variance expliquée + variance résiduelle »

- Coefficient de détermination :

$$R^2 = \frac{\|\hat{y} - \bar{\hat{y}}e\|^2}{\|y - \bar{y}e\|^2} = 1 - \frac{\|\hat{u}\|^2}{\|y - \bar{y}e\|^2} \in [0,1]$$

Rapport de la variance expliquée et de la variance totale.

5.2 Modèle sans terme constant : $e \notin L$

La somme des résidus n'est pas nulle. Beaucoup de relations précédentes ne sont plus valables.

5.3 Théorème de Frisch et Waugh

Partition des variables explicatives en deux groupes : $X = [W : Z]$, W de dimension $N \times K_1$, Z de dimension $N \times K_2$ avec $K = K_1 + K_2$.

La régression de y sur X a donné :

$$y = W\hat{c} + Z\hat{d} + \hat{u}$$

Théorème : \hat{c} est le résultat de la régression de $M_Z y$ sur $M_Z W$ où $M_Z = I - Z(Z'Z)^{-1}Z'$ est le projecteur orthogonal sur l'orthogonal de l'espace linéaire image de Z ; noté $L(Z)$.

6. Propriétés statistiques de la méthode des MCO

Le modèle :

$$\begin{cases} y = Xb + u, X(N \times K), \text{rg} X = K \\ Eu = 0, Vu = \sigma^2 I_N \end{cases}$$

Les MCO :

$$\hat{b} = (X'X)^{-1} X'y$$

6.1 Moments des MCO :

$E\hat{b} = b$, l'estimateur des MCO de b est sans biais.

$$V(\hat{b}) = \sigma^2 (X'X)^{-1}$$

6.2 Estimateur de σ^2 :

$$SCR = \|\hat{u}\|^2$$

$$E(SCR) = (N - K)\sigma^2$$

On prend $\hat{\sigma}^2$ comme estimateur sans biais de σ^2 :

$$\hat{\sigma}^2 = \frac{SCR}{N - K} = \frac{\|\hat{u}\|^2}{N - K}$$

on en déduit un estimateur sans biais de la matrice des variances covariances de \hat{b} :

$$\hat{V}(\hat{b}) = \hat{\sigma}^2(X'X)^{-1}$$

6.3 Le théorème de Gauss-Markov

Théorème : Soit : $C = \{b^+ / \exists A(K \times N), b^+ = Ay, Eb^+ = b, \forall b \in R^K\}$, la classe des estimateurs linéaires et sans biais de b :

- i) $\hat{b} \in C$
- ii) $V(b^+) - V(\hat{b})$ est une matrice de type positif (semi-définie positive), $\forall b^+ \in C$.
- iii) $V(b^+) = V(\hat{b}) \Leftrightarrow b^+ = \hat{b}$.

\hat{b} est donc le meilleur estimateur linéaire et sans biais au sens de la minimisation de la variance (*BLUE : Best Linear Unbiased Estimator*).

7. Le modèle linéaire normal

7.1 Définitions

$$\left\{ R^N, \mathcal{R}^N, N(m, \sigma^2 I_N), (m, \sigma^2) \in L \times (R^+ - \{0\}) \right\}$$

$$\left\{ R^N, \mathcal{R}^N, N(Xb, \sigma^2 I_N), (b, \sigma^2) \in R^K \times (R^+ - \{0\}) \right\}$$

Écritures standardisées :

$$y = m + u, u \approx N(0, \sigma^2 I_N), m \in L$$

$$y = Xb + u, u \approx N(0, \sigma^2 I_N), X(N \times K), \text{rg}X = K$$

7.2 Vraisemblance, log-vraisemblance, équations de vraisemblance

$$\left\{ \begin{aligned} L_N(b, \sigma^2) &= (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} Q(b)\right] \\ \log L(b, \sigma^2) &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} Q(b) \end{aligned} \right.$$

$$Q(b) = \|y - Xb\|^2$$

Les équations de vraisemblance :

$$\begin{cases} \frac{\partial \log L}{\partial b} = -\frac{1}{2\sigma^2} \frac{\partial Q}{\partial b} = 0 \Leftrightarrow \tilde{b} = \hat{b} \\ \frac{\partial \log L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} Q = 0 \Leftrightarrow \tilde{\sigma}^2 = \frac{N-K}{N} \hat{\sigma}^2 \end{cases}$$

$(\tilde{b}, \tilde{\sigma}^2)$ estimateur du MV et $(\hat{b}, \hat{\sigma}^2)$ estimateur des MCO de (b, σ^2) .

Propriétés :

- i) Le modèle linéaire normal est exponentiel.
- ii) $(\hat{b}, \hat{\sigma}^2)$ est une statistique exhaustive minimale et totale.
- iii) $(\hat{b}, \hat{\sigma}^2)$ est l'estimateur optimal de (b, σ^2) .

Loi du couple $(\hat{b}, \hat{\sigma}^2)$

- i) $\hat{b} \approx N(b, \sigma^2(X'X)^{-1})$
- ii) $(N-K) \frac{\hat{\sigma}^2}{\sigma^2} \approx \chi^2(N-K)$
- iii) \hat{b} et $\hat{\sigma}^2$ sont des variables aléatoires indépendantes.

Fonctions pivotales :

- i) $\mathcal{T}_k = \frac{\hat{b}_k - b_k}{\sqrt{\hat{\sigma}^2 h_{kk}}} \approx T(N-K)$ où b_k est le k ème élément de b et $h_{kk} = [(X'X)^{-1}]_{kk}$ est le k ème élément de la diagonale de $(X'X)^{-1}$.
- ii) $\mathcal{F} = \frac{1}{K\hat{\sigma}^2} (\hat{b} - b)' X' X (\hat{b} - b) \approx F(K, N-K)$
- iii) $\mathcal{R} = (N-K) \frac{\hat{\sigma}^2}{\sigma^2} \approx \chi^2(N-K)$

8. Intervalle de confiance et test sur un coefficient

8.1 Intervalle de confiance

La fonction pivotale \mathcal{T}_k permet de produire un intervalle de confiance pour le coefficient b_k . Au niveau $1-\alpha$ l'intervalle d'étendue minimale est donné par :

$$I_{1-\alpha}^k = \left[\hat{b}_k - t_{1-\frac{\alpha}{2}}(N-K)\sqrt{\hat{\sigma}^2 h_{kk}}, \hat{b}_k + t_{1-\frac{\alpha}{2}}(N-K)\sqrt{\hat{\sigma}^2 h_{kk}} \right]$$

où $t_z(q)$ est le quantile d'ordre z de la loi de Student à q degrés de liberté.

La fonction pivotale \mathcal{F} permet de construire des ellipsoïdes de confiance.

8.2 Test bilatéral sur un coefficient

Le problème de test est $H_0 : b_k = b_k^0$, $H_a : b_k \neq b_k^0$ ou b_k^0 est une valeur privilégiée par l'économètre. Lorsque cette valeur est nulle on parle de test de significativité.

Région critique du test de « Student » de niveau $1-\alpha$:

$$\begin{aligned} C &= \left\{ \hat{b}_k / b_k^0 \notin I_{1-\alpha}^k \right\} \\ &= \left\{ \hat{b}_k / \left| \frac{\hat{b}_k - b_k^0}{\sqrt{\hat{\sigma}^2 h_{kk}}} \right| > t_{1-\frac{\alpha}{2}}(N-K) \right\} \\ &= \left\{ \hat{b}_k / \left| 0 \right| > t_{1-\frac{\alpha}{2}}(N-K) \right\} \end{aligned}$$

$$\text{avec } \mathcal{T}_k^0 = \frac{\hat{b}_k - b_k^0}{\sqrt{\text{var}(\hat{b}_k)}}$$

La mise en œuvre pratique est la suivante : si \mathcal{T}_k^0 est en valeur absolue supérieur à $t_{1-\frac{\alpha}{2}}(N-K)$ ($\cong 2$ si $\alpha = 0,05$ et $N-K > 20$) on rejette l'hypothèse que $b_k = b_k^0$ sinon on ne réfute pas cette hypothèse.

Ce test a d'excellentes propriétés théoriques puisqu'il est uniformément le plus puissant dans la classe des tests sans biais.

8.2 Test unilatéral sur un coefficient

Le problème de test est : $H_0': b_k \leq b_k^0, H_a': b_k > b_k^0$. Dans le cas où $b_k^0 = 0$ c'est un test de signe.

La région critique est pour ce test :

$$C' = \left\{ \hat{b}_k / \mathcal{T}_k^0 \geq t_{1-\alpha}(N-K) \right\}$$

Ce test unilatéral possède d'aussi bonne propriétés théoriques que le précédent.

8.3 Tests sur la variance.

On peut être amené à considérer deux types de problèmes de tests dans ce cas :

- (1) $H_0 : \sigma^2 \leq \sigma_0^2, H_a : \sigma^2 > \sigma_0^2$
- (2) $H_0' : \sigma^2 = \sigma_0^2, H_a' : \sigma^2 \neq \sigma_0^2$

Les régions critiques sont respectivement pour un niveau $1 - \alpha$:

$$(1) \mathcal{W} = \left\{ (N-K) \frac{\hat{\sigma}^2}{\sigma_0^2} \geq \chi_{1-\alpha}^2(N-K) \right\}$$

$$(2) \mathcal{W}' = \left\{ (N-K) \frac{\hat{\sigma}^2}{\sigma_0^2} \geq c_2, (N-K) \frac{\hat{\sigma}^2}{\sigma_0^2} \leq c_1 \right\}$$

$$P(c_1 \leq \chi^2(N-K) \leq c_2) = 1 - \alpha$$

$$\int_{c_1}^{c_2} x f(x) dx = (N-K)(1-\alpha) \quad \text{où } f(x) \text{ est la densité de } \chi^2(N-K).$$

Chapitre 2 : LE MODELE LINEAIRE SOUS CONTRAINTES

1. Définitions

Définition 1 : définition intrinsèque

$(R^N, \mathcal{R}^N, \mathcal{P})$ est un modèle linéaire sous contraintes s'il existe un sous espace affine A de R^N tel que :

- i) l'extrémité de $E_p Y \in A$
- ii) $V_p Y = \sigma^2 I_N$

Définition 2 : définition matricielle :

$$\begin{cases} y = Xb + u, b \in R^K, y \in R^N, \text{rg}X = K \\ Eu = 0, Vu = \sigma^2 I_N \\ Rb = r, R(p \times K), \text{rg}R = p < K, r(p \times 1) \end{cases}$$

Propriété : les définitions 1 et 2 sont équivalentes.

2. Moindres carrés ordinaires sous contraintes

2.1 Contraintes résolues

$$Rb = r \Leftrightarrow b = b_0 + S\beta, \beta \in R^{K-p}$$

b_0 est une solution particulière et S est formé par des vecteurs de base du noyau de R .

L'estimateur contraint \hat{b}_c de b , est solution du programme :

$$\min_{\beta \in R^{K-p}} \|y - Xb_0 - XS\beta\|^2$$

$$\Rightarrow \hat{\beta} = (S'X'XS)^{-1}S'X'(y - Xb_0) \text{ et } \hat{b}_c = b_0 + S\hat{\beta}$$

2.2 Minimisation directe

$$\min_{b \in \mathbb{R}^K / Rb=r} \|y - Xb\|^2$$

$$\Rightarrow \hat{b}_c = \hat{b} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[r - R\hat{b}]$$

3. Le test de l'hypothèse linéaire générale : Le test de Fisher

Le modèle est linéaire normal ; deux approches : intrinsèque, matricielle :

Intrinsèque :

$$\begin{cases} y \approx N(m, \sigma^2 I_N) \\ m \in L, \dim L = K \end{cases}$$

Le problème de test :

$$H_0 : m \in L_0, H_a : m \notin L_0, m \in L, L_0 \text{ sev de } L \subset \mathbb{R}^N.$$

Matricielle :

$$\begin{cases} y \approx N(Xb, \sigma^2 I_N) \\ X(N \times K), \text{rg}X = K \end{cases}$$

Le problème de test :

$$H_0 : Rb = r, H_a : Rb \neq r, R(p \times K), \text{rg}R = p$$

3.1 Présentation géométrique :

Notations :

$$L_1 = L \cap L_0^\perp$$

P projecteur orthogonal sur L , $\dim L = K$

P_0 projecteur orthogonal sur L_0 , $\dim L_0 = K_0$

P_1 projecteur orthogonal sur L_1 , $\dim L_1 = K_1 = K - K_0$

$$P = P_0 + P_1$$

Lemme : Soit $Z \approx N(\theta, \sigma^2 I_N)$ un vecteur normal.

- i) Si P est un projecteur orthogonal alors $\left\| \frac{PZ}{\sigma} \right\|^2 \approx \chi^2(\text{rg}P, \frac{\|P\theta\|^2}{\sigma^2})$
- ii) Si P et Q sont deux projecteurs orthogonaux tels que $PQ = 0$ (c'est-à-dire $\text{Im}P$ et $\text{Im}Q$ sous-espaces orthogonaux) alors $\left\| \frac{PZ}{\sigma} \right\|^2$ et $\left\| \frac{QZ}{\sigma} \right\|^2$ sont des variables indépendantes.

Remarque : Si $X \approx N(\mu, I_n)$ par définition $\|X\|^2 \approx \chi^2(n, \|\mu\|^2)$, Chi-deux décentré et $\chi^2(n, 0) = \chi^2(n)$ est le Chi-deux centré.

Application: on pose :

$\hat{y} = Py$, estimateur de m sans contrainte

$\hat{u} = y - \hat{y} = (I_N - P)y$, résidu de l'estimation sans contrainte

$\hat{y}_0 = P_0 y$, estimateur de m sous contraintes

$\hat{u}_0 = y - \hat{y}_0 = (I_N - P_0)y$, résidu de l'estimation sous contraintes

on a donc $\hat{u} - \hat{u}_0 = \hat{y} - \hat{y}_0 = (P - P_0)y = P_1 y$.

D'après le lemme :

$$\left\| \frac{\hat{u} - \hat{u}_0}{\sigma} \right\|^2 = \left\| \frac{\hat{y} - \hat{y}_0}{\sigma} \right\|^2 \approx \chi^2(\text{rg}P_1, \frac{\|P_1 m\|^2}{\sigma^2})$$

$$\left\| \frac{\hat{u}}{\sigma} \right\|^2 \approx \chi^2(\text{rg}(I_N - P))$$

et ces deux variables sont indépendantes.

Statistique de Fisher :

$$\mathcal{F} = \frac{\frac{\|\hat{u} - \hat{u}_0\|^2}{\text{rg}P_1}}{\frac{\|\hat{u}\|^2}{\text{rg}(I_N - P)}} = \frac{\frac{\|\hat{y} - \hat{y}_0\|^2}{\text{rg}P_1}}{\frac{\|\hat{y}\|^2}{\text{rg}(I_N - P)}} = \frac{N - K}{K - K_0} \frac{\|\hat{u}_0\|^2 - \|\hat{u}\|^2}{\|\hat{u}\|^2}$$

Lorsque $m \in L_0, P_1 m = 0$. Ceci correspond à l'hypothèse nulle du problème de test et dans ce cas :

$$\mathcal{F} \approx \frac{\frac{\chi^2(\text{rg}P_1)}{\text{rg}P_1}}{\frac{\chi^2(\text{rg}(I_N - P))}{\text{rg}(I_N - P)}} = F(\text{rg}P_1, \text{rg}(I_N - P))$$

loi de Fisher à $\text{rg}P_1 = K_1 = K - K_0$ et $\text{rg}(I_N - P) = N - K$ degrés de liberté.

La région critique du test associé à cette statistique pour un seuil α est :

$$C = \{\mathcal{F} \geq F_{1-\alpha}(K - K_0, N - K)\}$$

3.2 Présentation matricielle

L'hypothèse nulle est $Rb = r, R(p \times K), \text{rg}R = p$. Ce système représente p contraintes linéaires indépendantes. Les solutions de ce système de p équations sont : $b = b_0 + S\beta, S(K \times (K - p))$, où b_0 est une solution particulière et où S est formée d'une base de $\text{Ker}R$.

Analyse de l'hypothèse nulle :

$$b = b_0 + S\beta$$

$$y = Xb_0 + XS\beta + u$$

$$y_c = y - Xb_0 = XS\beta + u$$

$$L_0 = L(XS), \dim L_0 = \text{rg}(XS) = K_0 = K - p$$

Analyse de l'hypothèse générale :

$$y = Xb + u = Xb_0 + X(b - b_0) + u$$

$$\delta = b - b_0$$

$$y_c = y - Xb_0 = X\delta + u$$

$$L = L(X), \dim L = \text{rg}(X) = K$$

Estimateurs contraint et non-contraint :

$$\hat{\delta} = \hat{b} - b_0$$

$$\hat{b}_c = b_0 + XS\hat{\beta}$$

$$\hat{y}_c = X\hat{\delta} = X(\hat{b} - b_0)$$

$$\hat{y}_{c0} = XS\hat{\beta} = X(\hat{b}_c - b_0)$$

Construction de la statistique de Fisher :

$$\|\hat{y}_c - \hat{y}_{c0}\|^2 = \|X\hat{b} - X\hat{b}_c\|^2$$

$$\|y_c - \hat{y}_c\|^2 = \|y - X\hat{b}\|^2$$

Statistique de Fisher :

$$\mathcal{F} = \frac{N-K}{p} \times \frac{\|X\hat{b} - X\hat{b}_c\|^2}{\|y - X\hat{b}\|^2}$$

Sous $H_0 : Rb = r$, \mathcal{F} est distribuée selon la loi de Fisher à $p, N-K$ degrés de liberté.

3.3 Ecritures équivalentes :

a) Forme de Wald :

$$X(\hat{b} - \hat{b}_c) = X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{b} - r]$$

$$\Rightarrow \|X(\hat{b} - \hat{b}_c)\|^2 = [R\hat{b} - r][R(X'X)^{-1}R']^{-1}[R\hat{b} - r]$$

$$\hat{\sigma}^2 = \|y - X\hat{b}\|^2 / (N-K)$$

et la statistique de Fisher associée :

$$\mathcal{F} = \frac{[R\hat{b} - r][R(X'X)^{-1}R']^{-1}[R\hat{b} - r]}{p\hat{\sigma}^2}$$

Puisque $[R\hat{b} - r] \approx N([Rb - r], \sigma^2 R(X'X)^{-1}R')$, et que \hat{b} est un estimateur du maximum de vraisemblance non-contraint, \mathcal{F} est une statistique de type Wald. On voit d'ailleurs que \mathcal{F} converge en loi vers la loi $\chi^2(p)$ lorsque $N \rightarrow \infty$.

b) Ecriture mnémotechnique

Sommes des carrés

$$SCR = \|\hat{u}\|^2, SCR_0 = \|\hat{u}_0\|^2$$

$$SCR_0 - SCR = \|\hat{u} - \hat{u}_0\|^2 = \|X\hat{b} - X\hat{b}_c\|^2$$

Degrés de liberté

$$dl = N - K$$

$$dl_0 = N - (K - p) = N - K + p$$

$$dl_0 - dl = p$$

Statistique de Fisher :

$$\mathcal{F} = \frac{SCR_0 - SCR}{SCR} \times \frac{dl}{dl_0 - dl} \underset{H_0}{\approx} F(dl_0 - dl, dl)$$

3.4 Exemples d'application

- a) Test de significativité de l'ensemble des coefficients
- b) Tests de stabilité, d'homogénéité... dits de Chow

Chapitre 3 : PROPRIETE ASYMPTOTIQUES DES MOINDRES CARRES ORDINAIRES

1.Introduction et notation

Le modèle linéaire sous forme détaillée :

$$y_n = \sum_{k=1}^K x_{kn} b_k + u_n = x_n' b + u_n, n = 1, 2, \dots, N$$

$$Eu_n = 0, Vu_n = \sigma^2, \text{cov}(u_n, u_{n'}) = 0, n \neq n'$$

et, sous forme matricielle :

$$Y_N = X_N b + U_N \quad \text{avec} \quad Y_N = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix}, X_N = \begin{pmatrix} x_{11} & \dots & x_{K1} \\ \dots & \dots & \dots \\ x_{1N} & \dots & x_{KN} \end{pmatrix}$$

$$EU_N = 0, VU_N = \sigma^2 I_N$$

Les estimateurs des MCO :

$$\hat{b}(N) = (X_N' X_N)^{-1} X_N' Y_N$$

$$\hat{\sigma}^2(N) = \frac{1}{N - K} \sum_{n=1}^N \hat{u}_{Nn}^2$$

$$\hat{U}_N = \begin{pmatrix} \hat{u}_{N1} \\ \cdot \\ \cdot \\ \cdot \\ \hat{u}_{NN} \end{pmatrix} = Y_N - X_N \hat{b}(N)$$

Ce chapitre a pour objet l'étude asymptotique de ces estimateurs, convergence et lois limites.

Notations : b_k et $\hat{b}_k(N)$, k ème élément de b et son estimateur des MCO.

$$h_{kk}(N) = \left[(X_N' X_N)^{-1} \right]_{kk}, \text{ } k\text{ème élément de la diagonale.}$$

2. Convergence des estimateurs des MCO

Propriété : Si $(X_N' X_N)^{-1} \rightarrow 0$ alors :

$$\text{i) } \hat{b}(N) \xrightarrow{P, L^2} b$$

et si de plus $\exists \delta > 1 / (\text{Log} N)^2 (\text{Log}(\text{Log} N))^\delta h_{kk}(N) \xrightarrow{N \rightarrow \infty} 0$

$$\text{ii) } \hat{b}_k(N) \xrightarrow{p.s.} b_k$$

iii) Si les perturbations sont indépendantes et équidistribuées :
 $\hat{b}(N) \xrightarrow{p.s.} b$

Remarque : Si $\frac{X_N' X_N}{N} \rightarrow A, A$ inversible : $\hat{b}(N) \xrightarrow{p.s.} b$

3. Loi asymptotique des estimateurs des MCO

Structure de $\hat{b}(N)$:

$$\begin{aligned} \hat{b}(N) &= (X_N' X_N)^{-1} X_N' Y_N \\ &= b + (X_N' X_N)^{-1} X_N' U_N \\ &= b + \sum_{n=1}^N (X_N' X_N)^{-1} x_n u_n \quad \text{avec } x_n' = (x_{1n}, x_{2n}, \dots, x_{Kn}), B_{Nn} = (X_n' X_N)^{-1} x_n \\ &= b + \sum_{n=1}^N B_{Nn} u_n \end{aligned}$$

Lemme : Soit $\{u_n, n=1,2,\dots\}$ une suite de variables aléatoires indépendantes d'espérance nulle et de variances constantes égale à σ^2 , et une suite triangulaire de vecteurs de taille p , $\{A_{Nn}, N=1,2,\dots, n=1,2,\dots, N\}$, tels que :

$$\sum_{n=1}^N A_{Nn} A_{Nn}' = I_p \quad \text{alors :}$$

$$\max_{n \leq N} \|A_{Nn}\|^2 \xrightarrow{N \rightarrow \infty} 0 \Leftrightarrow \sum_{n=1}^N A_{Nn} u_n \xrightarrow{L} N(0, \sigma^2 I_p)$$

Théorème : Soit le modèle de l'introduction où les perturbations sont indépendantes et équidistribuées, alors :

$$\max_{n \leq N} x_n' (X_N' X_N)^{-1} x_n \xrightarrow{N \rightarrow \infty} 0 \Leftrightarrow (X_N' X_N)^{-1/2} (\hat{b}(N) - b) \xrightarrow{L} N(0, \sigma^2 I_K)$$

4) Convergence et loi asymptotique de $\hat{\sigma}^2(N)$

Théorème : Si les perturbations sont indépendantes et équidistribuées, alors :

i) $\hat{\sigma}^2(N) \xrightarrow{p} \sigma^2$

ii) Si $E u_n^4 < \infty$, alors : $\hat{\sigma}^2(N) \xrightarrow{p.s.} \sigma^2$ et $\sqrt{N}(\hat{\sigma}^2(N) - \sigma^2) \xrightarrow{L} N(0, \text{var}(u_n^2))$

Remarque : Lorsque les perturbations sont indépendantes et équidistribuées et lorsque $\frac{X_N' X_N}{f(N)} \rightarrow A$ où $f(N) \rightarrow \infty$ et A est inversible, alors

$$\sqrt{f(N)}(\hat{b}(N) - b) \xrightarrow{L} N(0, \sigma^2 A^{-1})$$

$$\sqrt{N}(\hat{\sigma}^2(N) - \sigma^2) \xrightarrow{L} N(0, \text{var}(u^2))$$

En première approximation on considère que :

$$\hat{b}(N) \approx N(b, \sigma^2 (X_N' X_N)^{-1})$$

$$\hat{\sigma}^2(N) \approx N(\sigma^2, N^{-1} \text{var}(u^2))$$

Chapitre 4 : LE MODELE LINEAIRE GENERALISE

1. Introduction et définitions

Un modèle linéaire généralisé est un modèle linéaire au premier ordre dont la matrice des variances-covariances est proportionnelle à une matrice connue qui n'est pas l'identité.

$$\text{Définition détaillée : } \begin{cases} y_n = \sum_{k=1}^K x_{kn} b_k + u_n, n = 1, 2, \dots, n \\ Eu_n = 0 \\ \text{var } u_n = \sigma_n^2, \text{cov}(u_n, u_m) = \sigma_{nm}, n \neq m \end{cases}$$

$$\text{Définition matricielle : } \begin{cases} y = Xb + u, X(N \times K), \text{rg}X = K \\ Eu = 0, Vu = \Sigma \end{cases}$$

Remarques : Σ , matrice carrée $N \times N$, est symétrique, définie positive. On rencontre des problèmes réels où Σ n'est pas inversible. On écarte cependant ce cas particulier dans ce chapitre.

2. Etude du modèle linéaire généralisé

On suppose que Σ est connue à un facteur de proportionnalité près : $\Sigma = \sigma^2 \Omega$ où σ^2 est un paramètre et où Ω est une matrice connue.

Pour conserver une certaine analogie avec la modèle linéaire où $trI_N = N$ on utilise parfois une règle dite de normalisation en imposant : $tr\Omega = N$.

2.1 Sphéricisation du modèle

Puisque Ω est symétrique de type positif elle est diagonalisable dans le groupe orthogonal :

$$\exists H / H' = H^{-1}, \Omega = HDH', D = \text{diag}(d_1^2, d_2^2, \dots, d_N^2), d_i > 0 \forall i$$

on construit une matrice $\Omega^{-1/2} = HD^{-1/2}H'$, avec $D^{-1/2} = \text{diag}(d_1^{-1}, d_2^{-1}, \dots, d_N^{-1})$ et on transforme les variables en posant :

$$y^* = \Omega^{-1/2}y, X^* = \Omega^{-1/2}X, u^* = \Omega^{-1/2}u$$

on obtient le **modèle sphéricisé** qui est un modèle linéaire :

$$\begin{cases} y^* = X^*b + u^* \\ Eu^* = 0, Vu^* = \sigma^2 I_N \end{cases}$$

Le modèle linéaire généralisé possède toutes les propriétés algébriques et statistiques du modèle sphéricisé.

2.2 Méthode d'estimation du modèle linéaire généralisé : MCG

Théorème de Gauss-Markov : Le meilleur estimateur linéaire et sans biais (*BLUE*) de b est l'estimateur des MCO sur le modèle sphéricisé :

$$\begin{aligned} \tilde{b} &= (X^{*'}X^*)^{-1}X^{*'}y^* \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \end{aligned}$$

Remarques : i) Cet estimateur est appelé estimateur des **moindres carrés généralisés, MCG**.

ii) Il est sans biais et sa variance est :
 $V\tilde{b} = \sigma^2(X^{*'}X^*)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$.

iii) On définit un estimateur sans biais de σ^2 par :

$$\tilde{\sigma}^2 = \frac{\|y^* - X^* \tilde{b}\|^2}{N - K} = \frac{(y - X\tilde{b})' \Omega^{-1} (y - X\tilde{b})}{N - K}$$

iv) Si le modèle est normal : $u \approx N(0, \sigma^2 \Omega)$, on a :

$$\begin{cases} \tilde{b} \approx N(b, \sigma^2 (X' \Omega^{-1} X)^{-1}) \\ (N - K) \frac{\tilde{\sigma}^2}{\sigma^2} \approx \chi^2(N - K) \end{cases} \text{ et ces deux variables sont indépendantes.}$$

Dans ce cadre \tilde{b} et $\frac{N - K}{N} \tilde{\sigma}^2$ sont les estimateurs du maximum de vraisemblance.

Chapitre 5 : LE MODELE QUASI-LINEAIRE

1. Introduction

Le modèle linéaire généralisé suppose une connaissance presque parfaite de la structure du deuxième ordre du modèle :

$$\begin{cases} y = Xb + u, X(N \times K), \text{rg}X = K \\ Eu = 0, Vu = \sigma^2\Omega \end{cases}$$

où Ω est une matrice connue. On ne peut en pratique laisser libres tous les paramètres de cette matrice symétrique. On recourt ainsi à l'hypothèse suivante : $\exists S : \Theta \rightarrow \mathcal{E}$ où \mathcal{E} est l'espace des matrices symétriques définies positives et $\Theta \subset R^m$, avec $m=1,2$ ou 3 en pratique, tel que $\exists \theta \in \Theta, S(\theta) = \Omega$. On suppose S injective (identifiabilité de θ).

Définition : On appelle modèle **quasi-linéaire** :

$$\begin{cases} y = Xb + u, X(N \times K), \text{rg}X = K \\ Eu = 0, Vu = \sigma^2 S(\theta), \theta \in \Theta \subset R^m \end{cases}$$

on ne détaille pas à ce stade les propriétés de régularité de S .

Exemple : $u_n \approx MA(1)$, $u_n = \varepsilon_n + \theta\varepsilon_{n-1}$, $(\varepsilon_n) \approx BB(\sigma^2)$.

2. Propriétés des estimateurs usuels

2.1 MCO, MCG

MCO : $\hat{b} = (X'X)^{-1}X'y$, MCG : $\tilde{b} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$.

\hat{b} est véritablement un estimateur au sens où il est calculable ; \tilde{b} est une référence idéale car Ω est inconnue !

\hat{b} et \tilde{b} sont sans biais.

Les programmes informatiques calculent une estimation de la variance de \hat{b} par :

$$\hat{V}\hat{b} = \hat{\sigma}^2(X'X)^{-1} \text{ avec } \hat{\sigma}^2 = \frac{\|y - X\hat{b}\|^2}{N - K}$$

alors que :

$$V\hat{b} = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Théorème de Greenwald : Soit $\omega_{nn'}$ les éléments de Ω et x_{kn} les observations sur les variables exogènes :

$$y_n = \sum_{k=1}^K x_{kn}b_k + u_n, \text{cov}(u_n, u_{n'}) = \sigma^2 \omega_{nn'}$$

Si la covariance empirique $\text{cov}_e(\omega_{nn'}, x_{kn}x_{kn'}) > 0$ alors $E \text{vâr } \hat{b}_k < \text{var } \hat{b}_k$.

Remarque : L'estimateur de la variance produit par la méthode des moindres carrés est un estimateur biaisé et il existe de nombreuses circonstances où il sous estime les véritables variances, biaisant ainsi fortement les tests de significativité entre autres.

2.2 MCQG

Soit $\hat{\theta}$ un estimateur de θ . On peut proposer un estimateur de Ω en calculant : $\hat{\Omega} = S(\hat{\theta})$.

Définition : $b^* = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$ est appelé **estimateur des moindres carrés quasi généralisés, MCQG**.

Propriétés des MCQG : - b^* n'a pas nécessairement d'espérance ; s'il en a une il est en général biaisé... mais cf plus loin...

- même si la loi de la perturbation est connue, le calcul de la loi de b^* n'est pas possible en général. On recourt alors à des expériences de simulation ou aux approximations asymptotiques.

Voici deux théorèmes un peu ad hoc !

Théorème : Si $P\lim \frac{X'\hat{\Omega}^{-1}X}{N}$ existe et est inversible et si $P\lim \frac{X'\hat{\Omega}^{-1}u}{N} = 0$ alors $P\lim b^* = b$. $P\lim$ désigne la limite en probabilité.

Théorème : Si

$$- \frac{X'\Omega^{-1}X}{N} \rightarrow Q \text{ où } Q \text{ est une matrice inversible,}$$

$$- P\lim \frac{1}{N} X'(\hat{\Omega}^{-1} - \Omega^{-1})X = 0,$$

$$- P\lim \frac{1}{\sqrt{N}} X'(\hat{\Omega}^{-1} - \Omega^{-1})u = 0,$$

- $\Omega^{-1/2}u$ est un vecteur de variables indépendantes et identiquement distribuées,

$$\text{alors : } P\lim \sqrt{N}(b^* - \tilde{b}) = 0.$$

Remarques : Autrement dit les estimateurs des MCG et des MCQG ont même loi asymptotique.

$\sqrt{N}(b^* - b) \xrightarrow{L} N(0, \sigma^2 Q^{-1})$ puisque $P\lim \frac{X'\hat{\Omega}^{-1}X}{N} = Q$ on « approxime » la loi de b^* par la loi normale $N(b, \sigma^2(X'\hat{\Omega}^{-1}X)^{-1})$.

Les expériences de simulation montrent dans de nombreuses circonstances que cette approximation n'est pas bonne pour des échantillons de taille modérée : de quelques dizaines à quelques centaines...

Disposer d'un estimateur de θ convergent ne garantit pas que $P\lim(b^* - \tilde{b}) = 0$.

Si les perturbations sont normales la dernière hypothèse du théorème est vérifiée. Si les autres hypothèses le sont aussi alors l'estimateur des MCQG est asymptotiquement efficace.

2.3 Modèle quasi linéaire normal: Maximum de vraisemblance.

$$\begin{cases} y = Xb + u, X(N \times K), \text{rg}X = K \\ u \approx N(0, \sigma^2 S(\theta)), \theta' = (\theta_1, \theta_2, \dots, \theta_m) \end{cases} \quad \theta \rightarrow S(\theta) \text{ injective.}$$

Log-vraisemblance :

$$L(b, \sigma^2, \theta) = -\frac{N}{2} \text{Log} 2\pi - \frac{N}{2} \text{Log} \sigma^2 - \text{Log} \det S(\theta) - \frac{1}{2\sigma^2} u' S(\theta)^{-1} u \text{ avec } u = y - Xb.$$

Equations de vraisemblance :

$$\begin{cases} \frac{\partial}{\partial b} L = X' S(\theta)^{-1} u = 0 \\ \frac{\partial}{\partial \sigma^2} L = \text{tr}(\sigma^2 S(\theta) - uu') S(\theta)^{-1} = 0 \\ \frac{\partial}{\partial \theta_k} L = \text{tr}(\sigma^2 S(\theta) - uu') \sigma^{-2} \frac{\partial}{\partial \theta_k} S(\theta)^{-1} = 0, k = 1, 2, \dots, m \end{cases}$$

On en déduit :

$$b^* = (X' S(\theta^*)^{-1} X)^{-1} X' S(\theta^*)^{-1} y$$

$$\sigma^{2*} = \frac{u^{*'} S(\theta^*)^{-1} u^*}{N}$$

$$u^* = y - Xb^*$$

où b^*, σ^{2*} sont les estimateurs du MV de b, σ^2 respectivement et où θ^* est celui de θ défini implicitement par le système :

$$(*) \left\{ \sigma^{2*} \text{tr} S(\theta^*) \frac{\partial}{\partial \theta^k} S(\theta^*)^{-1} = u^{*'} S(\theta^*)^{-1} u^*, k = 1, 2, \dots, m. \right.$$

« Estimateur en une étape » et calcul itératif du MV :

i) Il existe en général une valeur θ_0 de θ telle que $S(\theta_0) = I_N$.

ii) Calculer $b^*(\theta_0) = (X' S(\theta_0)^{-1} X)^{-1} X' S(\theta_0)^{-1} y = (X' X)^{-1} X' y = \hat{b}$,

$$u^*(\theta_0) = y - Xb^*(\theta_0) = y - X\hat{b} = \hat{u} \text{ et } \sigma^{2*}(\theta_0) = \frac{u^*(\theta_0)' S(\theta_0)^{-1} u^*(\theta_0)}{N} = \frac{\hat{u}' \hat{u}}{N} = \hat{\sigma}^2.$$

- iii) En déduire $\theta_1^* = \theta^*(u^*(\theta_0))$ solution du système (*) dans lequel $u^* = u^*(\theta_0) = \hat{u}$ et $\sigma^{2*} = \sigma^{2*}(\theta_0) = \hat{\sigma}^2$.
- iv) Calculer $b^*(\theta_1^*) = (X'S(\theta_1^*)^{-1}X)^{-1}X'S(\theta_1^*)^{-1}y \dots$

Propriétés :

- i) Les estimateurs « en une étape » et du MV sont des estimateurs des MCQG.
- ii) Ces estimateurs sont distribués symétriquement autour de b , la vraie valeur du paramètre.
- iii) Si les espérances de ces deux estimateurs existent (!), ils sont sans biais.

Information de Fisher :

Elle se partitionne de la manière suivante :

$$I_f = \begin{pmatrix} I_{bb} & I_{\sigma^2 b} & I_{\theta b} \\ I_{b\sigma^2} & I_{\sigma^2\sigma^2} & I_{\theta\sigma^2} \\ I_{b\theta} & I_{\sigma^2\theta} & I_{\theta\theta} \end{pmatrix} \text{ avec } \begin{cases} I_{bb} = \sigma^{-2} X' \Omega^{-1} X, I_{\sigma^2 b} = 0, I_{\theta b} = 0 \\ I_{\sigma^2\sigma^2} = \frac{N}{2\sigma^4}, I_{\theta_k\sigma^2} = -\frac{1}{2\sigma^2} \text{tr} \frac{\partial S^{-1}}{\partial \theta_k} S, k = 1, 2, \dots, m \\ I_{\theta_k\theta_{k'}} = \frac{1}{2} \text{tr} \frac{\partial S^{-1}}{\partial \theta_k} S \frac{\partial S^{-1}}{\partial \theta_{k'}} S \end{cases}$$

Les estimateurs du MV des premier et second ordres sont donc asymptotiquement indépendants.

3. Test sur la variance du modèle : un exemple.

On considère le modèle :

$$\begin{cases} y = Xb + u, X(N \times K), \text{rg}X = K \\ u \approx N(0, \sigma^2 S(\theta)), \theta \in R, m = 1 \end{cases}$$

On suppose de plus que pour $\theta = 0, S(0) = I_N$ et qu'on s'intéresse à $H_0 : \theta = 0, H_a : \theta \neq 0$. Sous H_0 le modèle est linéaire ; il relève de la méthode des moindres carrés ordinaires. On utilise pour ce problème test, le test du score qui repose uniquement sur les estimateurs (du MV) sous l'hypothèse nulle.

Résultat : Pour le problème de test ci-dessus la statistique de test du score est :

$$\xi_S = \frac{(\hat{u}' A \hat{u})^2}{2\hat{\sigma}^2 \text{tr} A^2}$$

avec $A = \frac{\partial}{\partial \theta} S(0)$. Sous H_0 , ξ_S est asymptotiquement distribué selon la loi du χ^2 à un degré de liberté, et tend vers $+\infty$ sous l'alternative. Pour un niveau égal à $1-\alpha$, la région critique est :

$$\{\xi_S \geq \chi^2_{1-\alpha}(1)\}$$

Exemple : $(u_n) \approx MA(1)$ ou $(u_n) \approx AR(1)$: $\xi_S = \frac{N^2}{N-1} \frac{(\sum_{n=2}^N (\hat{u}_n \hat{u}_{n-1}))^2}{\sum_{n=1}^N \hat{u}_n^2} \cong (\sqrt{N} \hat{\rho})^2$ avec $\hat{\rho}$

résultat de la régression de \hat{u}_n sur \hat{u}_{n-1} : $\hat{u}_n = \hat{\rho} \hat{u}_{n-1} + \hat{\varepsilon}_n$.

Remarque : Le test ne tient compte que de la forme de $S(\theta)$ au voisinage de $\theta = 0$: $S(\theta) = I_N + \theta A + \dots$.

3. Propriétés de l'estimateur des MCO

3.1 Propriétés statistiques de l'estimateur des MCO

Sans tenir compte de la structure de variance du modèle linéaire généralisé l'estimateur des MCO est une alternative simple à l'estimateur des MCG.

Propriétés : i) $\hat{b} = (X'X)^{-1} X'y$ est un estimateur linéaire et sans biais.

ii) Sa variance est : $V\hat{b} = \sigma^2 (X'X)^{-1} (X'\Omega X) (X'X)^{-1}$ et $V\hat{b} - V\tilde{b}$ est une matrice semi définie positive.

iii) Si \hat{b} est un estimateur linéaire sans biais l'estimateur des MCO

$\hat{\sigma}^2 = \frac{\|y - X\hat{b}\|^2}{N-K}$ de σ^2 est un estimateur biaisé : $E\hat{\sigma}^2 = \sigma^2 \frac{N - \text{tr}(X'X)^{-1} X'\Omega X}{N-K}$.

3.3 Cas d'égalité des MCO et des MCG

Théorème de Kruskal : $\{\tilde{b} = \hat{b}\} \Leftrightarrow \{\exists V / \Omega X = XV\}$

Remarques : i) Ce théorème est valable lorsque Ω n'est pas inversible.

ii) La condition ci-dessus se traduit aussi par : $\Omega L(X) \subset L(X)$; soit si Ω est inversible par $L(X)$ est invariant par Ω .

Chapitre 6 : HETEROSCEDASTICITE

1. Généralités

-Nuage en entonnoir

-Un exemple : Le modèle linéaire simple avec hétéroscédasticité proportionnelle.

$$\begin{cases} y_n = a + bx_n + u_n, n = 1, 2, \dots, N \\ Eu_n = 0, \text{var } u_n = \sigma_n^2, \text{cov}(u_n, u_{n'}) = 0, n \neq n' \end{cases}$$

Spécification proportionnelle à une variable exogène : $\sigma_n^2 = s^2 z_n$ où z_n est une variable positive observable et s^2 un paramètre.

Sphéricisation :

$$\begin{aligned} \frac{y_n}{\sqrt{z_n}} &= a \frac{1}{\sqrt{z_n}} + b \frac{x_n}{\sqrt{z_n}} + \varepsilon_n, \varepsilon_n = \frac{u_n}{\sqrt{z_n}}, n = 1, 2, \dots, N \\ E\varepsilon_n &= 0, \text{var } \varepsilon_n = s^2, \text{cov}(\varepsilon_n, \varepsilon_{n'}) = 0, n \neq n' \end{aligned}$$

L'estimation de ce modèle sphéricisé relève de la méthode des moindres carrés ordinaires.

-Modèles à coefficients aléatoires :

$$\begin{cases} y_n = a_n + b_n x_n + u_n, n = 1, 2, \dots, N \\ E \begin{pmatrix} a_n \\ b_n \\ u_n \end{pmatrix} = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}, V \begin{pmatrix} a_n \\ b_n \\ u_n \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} & 0 \\ \sigma_{ba} & \sigma_b^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \end{cases}$$

Les vecteurs (a_n, b_n, u_n) $n = 1, 2, \dots, N$ ne sont pas corrélés. Ce modèle s'écrit de manière équivalente :

$$\begin{cases} y_n = a + bx_n + \varepsilon_n, \varepsilon_n = (a_n - a) + (b_n - b)x_n + u_n, n = 1, 2, \dots, N \\ E\varepsilon_n = 0, \text{var } \varepsilon_n = \sigma_a^2 + \sigma^2 + 2\sigma_{ab}x_n + \sigma_b^2x_n^2, \text{cov}(\varepsilon_n, \varepsilon_{n'}) = 0, n \neq n' \end{cases}$$

2. Méthodes d'estimations : exemple de l'hétéroscédasticité multiplicative :

De manière générale $\sigma_n^2 = f(z_n, \theta)$ où $f(\dots)$ est une fonction connue, z_n des observations sur un vecteur de variables exogènes, et θ est un vecteur de paramètres. On est dans le cas d'un modèle à hétéroscédasticité multiplicative si : $f(z_n, \theta) = \exp(z_n' \theta)$.

Pour un modèle :

$$\begin{cases} y_n = x_n' b + u_n, n = 1, 2, \dots, N, x_n \in R^K \\ Eu_n = 0, \text{var } u_n = \exp(z_n' \theta), \text{cov}(u_n, u_{n'}) = 0, n \neq n', z_n \in R^p \end{cases}$$

l'estimateur « idéal » des MCG est :

$$\tilde{b} = \left(\sum_{n=1}^N \exp(-z_n' \theta) x_n x_n' \right)^{-1} \sum_{n=1}^N \exp(-z_n' \theta) y_n x_n$$

2.1 Le maximum de vraisemblance :

On impose de plus que le modèle est normal :

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_N \end{pmatrix} \approx N(0, D)$$

avec $D = \text{diag}(\dots, \exp(z_n' \theta), \dots)$.

Les équations de vraisemblance du chapitre 5 produisent le système suivant où b^* et θ^* sont les estimateurs du MV de b et θ respectivement :

$$\begin{aligned} b^* &= \left(\sum_{n=1}^N \exp(-z_n' \theta^*) x_n x_n' \right)^{-1} \sum_{n=1}^N \exp(-z_n' \theta^*) y_n x_n \\ \sum_{n=1}^N z_n (u_n^{*2} \exp(-z_n' \theta^*) - 1) &= 0, u_n^* = y_n - x_n' b^* \end{aligned}$$

Ce système peut conduire à un estimateur en une étape ou à l'estimateur du MV calculé par une méthode itérative comme il est suggéré au chapitre 5.

2.2 Méthode des régressions auxiliaires : MCQG

On remarque : $\text{Log}(\sigma_n^2) = z_n' \theta$ donc si on pose : $v_n = \text{Log}\left(\frac{u_n^2}{\exp(z_n' \theta)}\right) = \text{Log}u_n^2 - z_n' \theta$ on voit que v_n est distribué selon une loi $\text{Log}(\chi^2(1))$, or :

$$E\text{Log}(\chi^2(1)) \cong -1,2704$$

$$V\text{Log}(\chi^2(1)) \cong 4,9348$$

Ainsi :

$$\begin{cases} \text{Log}u_n^2 = z_n' \theta + v_n, n = 1, 2, \dots, N \\ E v_n \cong -1,2704, \text{var } v_n \cong 4,9348 \end{cases}$$

les perturbations v_n sont indépendantes. On voit apparaître un modèle linéaire auxiliaire en remplaçant dans le modèle ci-dessus u_n par \hat{u}_n le résidu des MCO. On obtient un estimateur $\hat{\theta}$ de θ en « régressant » $\text{Log}\hat{u}_n^2 + 1,2704$ sur z_n . On calcule ensuite un estimateur des MCQG de b par :

$$\tilde{b} = \left(\sum_{n=1}^N \exp(-z_n' \hat{\theta}) x_n x_n' \right)^{-1} \sum_{n=1}^N \exp(-z_n' \hat{\theta}) y_n x_n$$

Cet estimateur est convergent et asymptotiquement normal et équivalent à l'estimateur b^* , ce qui n'est pas le cas de $\hat{\theta}$ vis à vis de θ^* .

3. Test de l'hypothèse d'homoscédasticité contre celle d'hétéroscédasticité

Le problème de test repose sur l'hypothèse nulle d'homoscédasticité $H_0 : \sigma_n^2 = \sigma^2, \forall n$; l'alternative d'hétéroscédasticité variant selon les modèles.

3.1 Test historique de Goldfeld et Quandt

Le modèle est :

$$\begin{cases} y_n = x_n' b + u_n, x_n \in R^K \\ u_n \approx N(0, \sigma_n^2), n = 1, 2, \dots, N \end{cases}$$

les perturbations sont indépendantes et leurs variances sont liées à une variable réelle positive z_n .

La méthode de Goldfeld et Quandt est la suivante :

- classer les observations en fonction de z_n croissante,
- déterminer la médiane des $z_n, n = 1, 2, \dots, N$ et éliminer c observations de part et d'autre de cette médiane : on élimine donc $2c$ observations.
- on obtient deux sous échantillons E_1 et E_2 dont le nombre d'observations est, si N est pair : $(N/2) - c$.
- Les MCO sur chacun des sous échantillons produisent SCR_1, SCR_2 telles que :

$$\frac{SCR_1}{\sigma^2} \approx \chi^2((N/2) - c - K), \frac{SCR_2}{\sigma^2} \approx \chi^2((N/2) - c - K)$$

sous l'hypothèse d'homoscédasticité. Calculées sur deux échantillons disjoints, ces deux fonctions pivotales sont de plus indépendantes. Ainsi, sous H_0 ,

$$\frac{SCR_1}{SCR_2} \approx F((N/2) - c - K, (N/2) - c - K)$$

La région critique peut prendre l'une des trois formes suivantes selon que sous l'alternative on sait que :

- 1) la variance est corrélée positivement à z
- 2) corrélée négativement à z
- 3) sans idée à priori sur le signe de cette corrélation

$$1) : C_1 = \left\{ \frac{SCR_1}{SCR_2} \leq F_{\alpha}((N/2) - c - K, (N/2) - c - K) \right\}$$

$$2) : C_2 = \left\{ \frac{SCR_1}{SCR_2} \geq F_{1-\alpha}((N/2) - c - K, (N/2) - c - K) \right\}$$

3) :

$$C_3 = \left\{ \frac{SCR_1}{SCR_2} \geq F_{1-\alpha/2} \left(\frac{N}{2} - c - K, \frac{N}{2} - c - K \right), \frac{SCR_1}{SCR_2} \leq F_{\alpha/2} \left(\frac{N}{2} - c - K, \frac{N}{2} - c - K \right) \right\}$$

3. Le test de Breush et Pagan

Le modèle est normal et la forme de la variance est : $\sigma_n^2 = h(\theta_1 + \sum_{j=2}^p \theta_j z_{jn})$ où h est un fonction quelconque. Le problème de test prend néanmoins une forme paramétrique :

$$H_0 : \theta_2 = \theta_3 = \dots = \theta_p = 0, H_a : \exists j, 2 \leq j \leq p / \theta_j \neq 0$$

C'est un test du score. Sa mise en œuvre repose donc sur les estimateurs sous l'hypothèse nulle, c'est-à-dire sur les estimateurs des moindres carrés. Si \hat{u}_n sont les résidus des MCO, la mise en œuvre pratique repose sur deux régressions auxiliaires :

- \hat{u}_n^2 sur $cte, z_{2n}, z_{3n}, \dots, z_{pn}$ qui produit SCR_2

- \hat{u}_n^2 sur cte qui produit SCR_1

La statistique de test est :

$$\xi_{BP} = N \frac{SCR_1 - SCR_2}{SCR_1}$$

Sous H_0 , $\xi_{BP} = N \frac{SCR_1 - SCR_2}{SCR_1} \xrightarrow{L} \chi^2(p-1)$ et $\xi_{BP} \rightarrow +\infty$ sous l'alternative.

La région critique du test asymptotique pour un niveau $1-\alpha$ est donc :

$$C = \left\{ \xi_{BP} \geq \chi_{1-\alpha}^2(p-1) \right\}$$

3.2 Test de Szroeter

Ce test suppose que l'on sait ordonner les variances : $Vu_n = \sigma_n^2 \Rightarrow \sigma_{n-1}^2 < \sigma_n^2$. Le problème de test est alors :

$$H_0 : \sigma_{n-1}^2 = \sigma_n^2, \forall n, H_a : \sigma_{n-1}^2 < \sigma_n^2, \forall n$$

Szroeter propose de calculer :

$$\tilde{h} = \frac{\sum_{n=1}^N n \hat{u}_n^2}{\sum_{n=1}^N \hat{u}_n^2}$$

$$Q = \sqrt{\frac{6N}{N^2-1}} \left(\tilde{h} - \frac{N+1}{2} \right)$$

Sous $H_0 : Q \xrightarrow{L} N(0,1)$ et sous $H_a : Q \rightarrow +\infty$. Il propose donc la région critique suivante :

$$C = \{Q \geq u_{1-\alpha}\}$$

où $u_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite quand le test est construit pour un niveau $1-\alpha$.

Remarques : ce test serait plus puissant que les deux précédents.

4. Approche de White

4.1 Le modèle

$(y_n, x_n), n = 1, 2, \dots$ est une suite de vecteurs aléatoires iid de taille $K+1$ et de carrés intégrable ($\in L^2$). On suppose que : $E(y_n / x_n) = x_n' b, V(y_n / x_n) = \sigma^2(x_n)$.

Pour un échantillon de taille N on obtient un modèle linéaire hétéroscédastique conditionnel :

$$\begin{cases} y = Xb + u \\ E(u / X) = 0, V(u / X) = D = \text{diag}(\dots, \sigma^2(x_n), \dots) \end{cases}$$

4.2 Loi asymptotique des MCO

$$\hat{b} = (X'X)^{-1} X'y = \left(\sum_{n=1}^N x_n x_n' \right)^{-1} \sum_{n=1}^N y_n x_n = \left(\frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \frac{1}{N} \sum_{n=1}^N y_n x_n \xrightarrow{p.s.} (Exx')^{-1} Eyx = b$$

$$\sqrt{N}(\hat{b} - b) = \left(\frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \frac{1}{\sqrt{N}} \left(\sum_{n=1}^N u_n x_n \right)$$

$$\frac{1}{N} \sum_{n=1}^N x_n x_n' \xrightarrow{p.s.} E x x', LFGN$$

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N u_n x_n \xrightarrow{L} N(0, E u^2 x x'), TCL$$

On en déduit que :

$$\sqrt{N}(\hat{b} - b) \xrightarrow{L} N(0, \Omega)$$

$$\Omega = (E x x')^{-1} E(u^2 x x') (E x x')^{-1}$$

$$E(u^2 x x') = E E(u^2 x x' / x) = E(E(u^2 / x) x x') = E \sigma^2(x) x x'$$

Remarques : Si le modèle est homoscédastique $\sigma^2(x) = \sigma^2$ et $\Omega = \sigma^2 (E x x')^{-1}$ qui est estimable de façon convergente par $\hat{\sigma}^2 (X' X / N)^{-1}$.

Si $\sigma^2(x)$ n'est pas corrélé (indépendant) avec (de) $x x'$ alors $E \sigma^2(x) x x' = E \sigma^2(x) E x x'$. Dans ce cas, $\Omega = E \sigma^2(x) (E x x')^{-1}$ a la structure de la variance des MCO pour un modèle homoscédastique. On montre que $\hat{\sigma}^2 (X' X / N)^{-1}$ est un estimateur convergent de Ω . Le modèle de White montre ainsi que l'hétéroscédasticité n'est un problème que lorsque la variance de la perturbation est corrélée avec $x_n x_n'$.

4.3 Estimation robuste de White

Les remarque précédentes montrent que la variance asymptotique de l'estimateur des MCO peut être estimée, dans le cas idéal où les perturbations sont connues, par :

$$\Omega^* = \left(\frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \frac{1}{N} \sum_{n=1}^N u_n^2 x_n x_n' \left(\frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1}$$

On peut traduire matriciellement ce résultat :

$$\Omega^* = \left(\frac{1}{N} X' X \right)^{-1} \frac{1}{N} X' D^* X \left(\frac{1}{N} X' X \right)^{-1}$$

$$D^* = \text{diag}(\dots, u_n^2, \dots)$$

White démontre que remplacer u_n^2 par \hat{u}_n^2 ne modifie pas les propriétés de convergence ; il propose d'estimer Ω par :

$$\hat{\Omega}^* = \left(\frac{1}{N} X' X\right)^{-1} \frac{1}{N} X' \hat{D}^* X \left(\frac{1}{N} X' X\right)^{-1}$$

$$\hat{D}^* = \text{diag}(\dots, \hat{u}_n^2, \dots), \hat{u}_n = y_n - x_n' \hat{b}$$

On en déduit une estimation dite robuste de la matrice de variance de l'estimateur des MCO, parce qu'elle est valable (convergente) avec ou sans hétéroscédasticité :

$$\hat{V}^*(\hat{b}) = (X' X)^{-1} X' \hat{D}^* X (X' X)^{-1}$$

4.4 Test d'hétéroscédasticité de White

Les considérations précédentes montrent que la détection de l'hétéroscédasticité est importante lorsque les variances des perturbations sont corrélées avec les éléments de $X'X$. White propose un test de type Breush et Pagan fondé sur la régression de \hat{u}_n^2 sur *cte*, $x_{1n}, x_{2n}, \dots, x_{kn}, x_{1n}^2, x_{2n}^2, \dots, x_{kn}^2, \dots, x_{kn}x_{k'n}, \dots; k \neq k'$. Ainsi lorsque le modèle est simple on calcule SCR_2 à partir de la régression du carré de la perturbation sur *cte*, x_n, x_n^2 .

Chapitre 7 : Régression multivariée, modèle de régressions empilées

1. Définition

Le modèle de régressions empilées concerne le traitement simultané de plusieurs variables endogènes reliées à un groupe de variables exogènes par des modèles linéaires.

Exemples : modèles de consommation
modèles de production
modèles VAR

2. Formalisme

G variables endogènes : $y_{gn}, g = 1, 2, \dots, G, n = 1, 2, \dots, N$

G modèles linéaires :

$$y_{gn} = \sum_{k=1}^{K_g} x_{gkn} b_{gk} + u_{gn}, g = 1, 2, \dots, G, n = 1, 2, \dots, N$$

« empilement » :

$$\left\{ \begin{array}{l} y_{-1} = X_1 b_1 + u_{-1} \\ \dots \\ y_{-g} = X_g b_g + u_{-g} \\ \dots \\ y_{-G} = X_G b_G + u_{-G} \end{array} \right.$$

avec : $y'_{-g} = (y_{g1}, y_{g2}, \dots, y_{gN}), g = 1, 2, \dots, G$.

Ce modèle est souvent qualifié de « modèle SUR », seemingly unrelated regression.

Hypothèses sur les perturbations :

- bonne spécification : $Eu_{gn} = 0, g = 1, 2, \dots, G, n = 1, 2, \dots, N$
- corrélation pour une même observation : $\text{cov}(u_{gn}, u_{jn}) = \sigma_{gj}, \forall g, \forall j, \forall n$
- pas de corrélation entre deux observations différentes : $\text{cov}(u_{gn}, u_{jm}) = 0, \forall g, \forall j, \forall n \neq m$

Notations :

$$\Sigma = (\sigma_{gj}), u_{-n}' = (u_{1n}, u_{2n}, \dots, u_{Gn})$$

$$Eu_{-n} u_{-n}' = \Sigma, Eu_{-n} u_{-m}' = 0, n \neq m$$

3. Utilisation du produit de Kronecker

$$Y = \begin{bmatrix} y \\ y_{-1} \\ \dots \\ y \\ -G \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & \dots & 0 \\ 0 & X_2 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & X_{G-1} & 0 \\ 0 & \dots & \dots & 0 & X_G \end{bmatrix} \begin{bmatrix} b_1 \\ \dots \\ b_G \end{bmatrix} + \begin{bmatrix} u_{-1} \\ \dots \\ u_{-G} \end{bmatrix}$$

avec : $u_{-g}' = (u_{g1}, u_{g2}, \dots, u_{gN}), g = 1, 2, \dots, G$.

$$\begin{cases} Y = Zb + U \\ EU = 0, VU = \begin{pmatrix} \sigma_{11}I_N & \dots & \dots & \sigma_{1G}I_N \\ \dots & \dots & \dots & \dots \\ \dots & \sigma_{g1}I_N & \dots & \dots \\ \sigma_{G1}I_N & \dots & \dots & \sigma_{GG}I_N \end{pmatrix} = \Sigma \otimes I_N \end{cases}$$

où $Z(NG \times \sum_{i=1}^G K_i), b(\sum_{i=1}^G K_i \times 1)$

4. Les cas d'équivalence des MCO et des MCG

4.1 MCO

$$\hat{b} = (Z'Z)^{-1}Z'Y = \begin{bmatrix} X_1'X_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & X_G'X_G \end{bmatrix}^{-1} \begin{bmatrix} X_1'y_{-1} \\ \dots \\ X_G'y_{-G} \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y_{-1} \\ \dots \\ (X_G'X_G)^{-1}X_G'y_{-G} \end{bmatrix} = \begin{bmatrix} \hat{b}_1 \\ \dots \\ \hat{b}_G \end{bmatrix}$$

autrement dit « MCO sur l'ensemble = MCO équation par équation ».

Mais, attention, $V\hat{b}$ n'est pas une matrice bloc-diagonale.

4.2 MCG

$$\tilde{b} = (Z'\Sigma^{-1} \otimes I_N Z)^{-1} Z'\Sigma^{-1} \otimes I_N Y$$

Théorème de Zellner (formulation de Monfort) :

Si $\forall (i, j) \in \{1, 2, \dots, G\}^2 : L(X_i) = L(X_j)$ ou $\sigma_{ij} = 0$ alors : $\tilde{b} = \hat{b}$.

Remarque : cette condition est suffisante mais pas nécessaire.

Démonstration : application du théorème de Kruskal en montrant que $L(\Sigma \otimes I_N Z) \subset L(Z)$ dans les conditions du théorème.

Cas particulier : la **régression multivariée** ou modèle linéaire multivarié.

Dans ce cas il n'y a pas de contrainte sur Σ , mais $X_1 = X_2 = \dots = X_G = X$.

$$Y = I_G \otimes Xb + U \Rightarrow MCO = MCG$$

5. Estimation des MCOG

Cette procédure ne présente d'intérêt qu'en dehors des cas d'application du théorème de Zellner.

5.1 Description de la méthode

a) Estimations de Σ .

La mise en œuvre des MCO sur chaque équation produit un estimateur sans biais \hat{b}_g de b_g et un vecteur de résidus $\hat{u}_{-g} = y_{-g} - X_g \hat{b}_g$. On remarque alors que

$E \hat{u}_{-i} \hat{u}_{-j}' = \sigma_{ij} \text{tr} M_i M_j$ avec $M_g = I_N - X_g (X_g' X_g)^{-1} X_g'$, $g = 1, 2, \dots, G$. On propose

donc un estimateur sans biais de σ_{ij} et de Σ :

$$\hat{\sigma}_{ij} = \frac{\hat{u}_{-i}' \hat{u}_{-j}}{\text{tr} M_i M_j}, \hat{\Sigma} = (\hat{\sigma}_{ij})$$

Remarque : $\hat{\Sigma}$ est un estimateur sans biais, mais cette matrice n'est pas nécessairement définie positive. On lui préfère en général l'estimateur suivant :

$$\tilde{\sigma}_{ij} = \frac{\hat{u}_{-i} \hat{u}_{-j}}{\sqrt{(N-K_i)(N-K_j)}}, \tilde{\Sigma} = (\tilde{\sigma}_{ij})$$

Par construction $\tilde{\Sigma}$ est définie positive mais n'est pas nécessairement sans biais...sauf pour les éléments de la diagonale !

b) Estimateur de b

C'est l'estimateur des MCQG : $b^* = (Z' \dot{\Sigma}^{-1} \otimes I_N Z)^{-1} Z' \dot{\Sigma}^{-1} \otimes I_N Y$ avec $\dot{\Sigma} = \hat{\Sigma}$ ou $\dot{\Sigma} = \tilde{\Sigma}$.

Propriétés :

- si les perturbations sont normales, b^* est sans biais (voir les propriétés du MV et de l'estimateur en une étape au chapitre 5).
- b^* est un estimateur convergent et asymptotiquement normal (voir les théorèmes du chapitre 5 ; ici les conditions de ces théorèmes sont faciles à vérifier car Σ est de dimension fixe).
- les expériences de simulation montrent que b^* est meilleur au sens de l'erreur quadratique moyenne (variance) que \hat{b} .
- b^* est d'autant meilleur que les variables explicatives des diverses régressions sont orthogonales...
- les MCQG sont programmés dans SAS : proc syslin.

6. Le maximum de vraisemblance

On empile pour cela les G équations observation par observation :

$$y_{-n} = B_{(G \times K)} x_{-n} + u_{-n}, n = 1, 2, \dots, N$$

K est le nombre de variables exogènes différentes dans l'ensemble des régressions. La g ième ligne de la n ième observation est :

$$y_{gn} = B_g x_{-n} + u_{ng}$$

où B_g est le vecteur ligne de dimension K dont les éléments sont les coefficients $b_{g1}, b_{g2}, \dots, b_{gK_g}$ correspondant aux variables exogènes présentes dans l'équation g et qui sont égaux à 0 pour les autres variables du modèle global. Ainsi le modèle linéaire multivarié correspond à une matrice B « pleine » c'est-à-dire sans élément égal à zéro.

Dans cette écriture et avec l'hypothèse de normalité des perturbations les vecteurs aléatoires $u_{-n}, n = 1, 2, \dots, N$ sont $Niid(0, \Sigma)$.

La log-vraisemblance :

$$\begin{aligned} L(B, \Sigma) &= -\frac{NG}{2} \log(2\pi) - \frac{N}{2} \log(\det \Sigma) - \frac{1}{2} \sum_{n=1}^N (y_{-n} - Bx_{-n})' \Sigma^{-1} (y_{-n} - Bx_{-n}) \\ &= -\frac{NG}{2} \log(2\pi) - \frac{N}{2} \log(\det \Sigma) - \frac{1}{2} \text{tr} \Sigma^{-1} H_N \end{aligned}$$

avec $H_N = \sum_{n=1}^N u_{-n} u_{-n}'$ et $u_{-n} = y_{-n} - Bx_{-n}$.

Equations de vraisemblance :

Comme tout modèle linéaire généralisé les équations de vraisemblance des paramètres b se résolvent selon un estimateur des MCG :

$$b^* = (Z' \Sigma^{*-1} \otimes I_N Z)^{-1} Z' \Sigma^{*-1} \otimes I_N Y$$

Celles relatives aux paramètres de Σ s'écrivent sous forme matricielle :

$$\frac{\partial L}{\partial \Sigma} = \left(\frac{\partial L}{\partial \sigma_{ij}} \right) = -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} H_N \Sigma^{-1} = 0$$

équations qui produisent :

$$\Sigma^* = \frac{1}{N} H_N^* \text{ soit } \sigma_{ij}^* = \frac{1}{N} \sum_{n=1}^N u_i^* u_j^* \text{ avec } u_g^* = y_{gn} - x_{-gn}' b_g^*.$$

Remarque : - l'estimateur du MV est un estimateur des MCQG itéré.

- Σ^* est symétrique semi-définie positive.

7. Un exemple de test asymptotique

Soit le modèle :

$$\begin{aligned} Y &= Zb + U \\ U &\approx N(0, \Sigma \otimes I_N) \end{aligned}$$

et le problème de test : $H_0 : Rb = r, H_a : Rb \neq r$.

La forme des équations de vraisemblance conduit à un maximum de la vraisemblance de la forme suivante :

$$L(B^*, \Sigma^*) = -\frac{GN}{2} \log(2\pi) - \frac{N}{2} \log(\det \Sigma^*) - \frac{GN}{2}$$

ce maximum ne dépend que de Σ^* ; ceci est aussi vrai pour l'estimation sous contraintes sur b . Le test du rapport des vraisemblances a donc la forme suivante si l'on pose :

Estimation sans contrainte : B^*, Σ^*

Estimation sous contrainte (H_0) : B_0^*, Σ_0^*

Statistique du rapport des vraisemblances :

$$\xi_{RV} = -2\{L(B_0^*, \Sigma_0^*) - L(B^*, \Sigma^*)\} = N \log\left(\frac{\det \Sigma_0^*}{\det \Sigma^*}\right)$$

Sous H_0 , $\xi_{RV} \xrightarrow{L} \chi^2(\text{rg}R)$, sous H_a , $\xi_{RV} \rightarrow +\infty$.

Et la région critique asymptotique est pour un test de niveau $1-\alpha$:

$$C = \{\xi_{RV} \geq \chi_{1-\alpha}^2(\text{rg}R)\} = \{\det \Sigma_0^* \geq e^{N^{-1}\chi_{1-\alpha}^2(\text{rg}R)} \det \Sigma^*\}$$

Chapitre 8 : Autocorrélation des perturbations

1. Introduction et définitions

Modèles pour séries temporelles à temps discret (données trimestrielles, données annuelles).

Modélisation de phénomènes de persistance, de rémanence dans les modèles économétriques . Modélisation des constats fréquents d'enroulement des séries observées autour du modèle linéaire estimé.

Utilisation des modèles de séries temporelles stationnaires. Modélisation des perturbations par un AR(1) :

$$u_t = \rho u_{t-1} + \varepsilon_t, (\varepsilon_t) \approx BB(\sigma^2)$$

Lorsque $|\rho| < 1$, la solution stationnaire au sens large de l'équation de récurrence précédente est : $u_t = \sum_{\tau=1}^{\infty} \rho^\tau \varepsilon_{t-\tau}$. On en déduit :

$$Eu_t = 0, Vu_t = \frac{\sigma^2}{(1-\rho^2)}, \text{cov}(u_t, u_{t-n}) = \rho^n \frac{\sigma^2}{(1-\rho^2)} .$$

Si $u' = (u_1, u_2, \dots, u_T)$, $Vu = \sigma_u^2 S(\rho)$ avec :

$$\sigma_u^2 = \frac{\sigma^2}{(1-\rho^2)}, S(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & \dots & 1 \end{pmatrix}$$

Définition : Le modèle quasi-linéaire à **autocorrélation des perturbations** est :

$$\begin{cases} y_t = x_t' b + u_t, t = 1, 2, \dots, T \\ u_t = \rho u_{t-1} + \varepsilon_t, |\rho| < 1, (\varepsilon_t) \approx BB(\sigma^2) \end{cases}$$

ou

$$\begin{cases} y = Xb + u, X(T \times K), \text{rg}X = K \\ Eu = 0, Vu = \sigma_u^2 S(\rho) \end{cases}$$

2. Les estimateurs usuels

2.1 Les MCO

$$\hat{b} = (X'X)^{-1} X'y, E\hat{b} = b, V\hat{b} = \sigma_u^2 (X'X)^{-1} X'S(\rho)X(X'X)^{-1}$$

$$\hat{V}\hat{b} = \hat{\sigma}^2 (X'X)^{-1}, \hat{\sigma}^2 = \frac{\|y - X\hat{b}\|^2}{T - K}$$

exemple : $y_t = bx_t + u_t, \rho > 0, r_j = \frac{\sum_{t=j+1}^T x_t x_{t-j}}{\sum_{t=1}^T x_t^2} > 0, \forall j$ alors :

$$E\hat{\sigma}^2 \leq \sigma_u^2, E \text{vâr} \hat{b} \leq \frac{\sigma_u^2}{\sum_{t=1}^T x_t^2} \leq \text{var}(\hat{b})$$

2.2 Les MCG

Modèle quasi-différencié :

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'b + u_t - \rho u_{t-1} = (x_t - \rho x_{t-1})'b + \varepsilon_t, t = 2, 3, \dots, T$$

Si ρ est connu, on pose : $y_t^* = y_t - \rho y_{t-1}, x_t^* = x_t - \rho x_{t-1}$:

$$\begin{cases} y_t^* = x_t^* b + \varepsilon_t, t = 2, 3, \dots, T \\ E\varepsilon_t = 0, \text{var}(\varepsilon_t) = \sigma^2, \text{cov}(\varepsilon_t, \varepsilon_{t'}) = 0, \forall t \neq t' \end{cases}$$

C'est un modèle linéaire dit « modèle quasi-différencié » qui n'est pas équivalent au modèle avec autocorrélation.

Modèle sphérisé :

Ajout d'une première observation au modèle quasi-différencié :

$y_1 = x_1' b + u_1$ est d'espérance nulle, sans corrélation avec $\varepsilon_t, t = 2, 3, \dots, T$ mais sa variance est différente de celle commune aux ε_t . En revanche :

$$\sqrt{1 - \rho^2} y_1 = \sqrt{1 - \rho^2} x_1' b + \sqrt{1 - \rho^2} u_1$$

produit un perturbation dont la variance est égale à la variance des ε_t :

$$\begin{cases} \sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}x_1'b + \varepsilon_1', \varepsilon_1' = \sqrt{1-\rho^2}u_1 \\ y_t^* = x_t'b + \varepsilon_t, t = 2,3,\dots,T \\ E\varepsilon_1' = 0, \text{var}(\varepsilon_1') = \sigma^2, E\varepsilon_t = 0, \text{var}(\varepsilon_t) = \sigma^2, t = 2,3,\dots,T \\ \text{cov}(\varepsilon_1', \varepsilon_t) = 0, \text{cov}(\varepsilon_t, \varepsilon_{t'}) = 0, t > 1, t \neq t' \end{cases}$$

Les estimateurs des MCG de b, σ^2 s'obtiennent à partir des MCO sur ce modèle sphéricisé.

3. Les procédures d'estimations : ρ inconnu.

3.1 Procédures de type M-estimation

Le modèle quasi-différencié conduit à la procédures de moindres carrés non-linéaires fondées sur la fonctions objectif :

$$f(b, \rho) = \sum_{t=2}^T (y_t - \rho y_{t-1} - (x_t - \rho x_{t-1})'b)^2$$

La minimisation de cette fonction ne pose pas de problème numérique, mais on peut encore trouver dans des anciens articles d'économétrie des références à des méthodes d'estimation itératives (Cochrane-Orcutt) ou par balayage (Hildreth-Lu).

3.2 MCQG : méthode de Durbin

Pour appliquer la méthode des MCQG, il convient de disposer d'un estimation préalable de ρ . Durbin a proposer d'estimer par les MCO la version autorégressive du modèle quasi-différencié sans tenir compte des contraintes non-linéaire sur les coefficients des variables.

$$y_t = \rho y_{t-1} + x_t'b - \rho x_{t-1}'b + \varepsilon_t, t = 2,3,\dots,T$$

qu'on écrit avec de nouveaux paramètres :

$$y_t = \alpha y_{t-1} + x_t'\beta + x_{t-1}'\gamma + \varepsilon_t, t = 2,3,\dots,T$$

Les MCO (sans tenir compte de la contrainte $\gamma + \rho\beta = 0$) produisent un estimateur $\hat{\rho}$ de ρ égal à $\hat{\alpha}$. On ne retient pas $\hat{\beta}$ comme estimateur de b en raison des risques de multicollinéarité entre x_t et x_{t-1} .

On calcule ensuite l'estimateur des MCQG avec $\hat{\rho} = \hat{a}$, ce qui peut être réalisé en utilisant le modèle sphéricisé.

3.3 Le maximum de vraisemblance : les perturbations sont normales

Pour construire la vraisemblance on s'appuie sur le modèle sphéricisé où le vecteur des perturbation est :

$$(\sqrt{1-\rho^2}u_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_T) \approx N(0, \sigma^2 I_T)$$

en tenant compte du Jacobien on obtient la log-vraisemblance :

$$L(b, \rho, \sigma^2) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 + \frac{1}{2} \log(1-\rho^2) - \frac{1}{2\sigma^2} \left[(1-\rho^2)u_1^2 + \sum_{t=2}^T (u_t - \rho u_{t-1})^2 \right]$$

$$u_t = y_t - x_t' b$$

Les équations de vraisemblances produisent les estimateurs de b, σ^2 sous la forme d'estimateurs des MCQG :

$$b^* = (X' \Omega(\rho^*)^{-1} X)^{-1} X' \Omega(\rho^*)^{-1} y$$

$$\sigma^{2*} = \frac{1}{T} \left[(1-\rho^{*2})u_1^{*2} + \sum_{t=2}^T (u_t^* - \rho^* u_{t-1}^*)^2 \right]$$

$$u_t^* = y_t - x_t' b^*$$

L'équation de vraisemblance relative à ρ s'écrit :

$$\frac{\partial}{\partial \rho} L = \frac{G(\rho)}{\sigma^2(1-\rho^2)}$$

$$G(\rho) = (1-\rho^2) \left(\sum_{t=2}^T u_t u_{t-1} - \rho \sum_{t=2}^{T-1} u_t^2 \right) - \rho \sigma^2$$

$$d^{\circ}G = 3, G(-1) = \sigma^2, G(0) = \sum_{t=2}^T u_t u_{t-1}, G(1) = -\sigma^2$$

Il existe donc une seule solution de cette équation en ρ comprise entre -1 et $+1$. La procédure de maximisation qui respecte cette contrainte porte le nom de procédure de Beach et McKinnon.

4. Détection de l'autocorrélation

Le problème de test : $H_0 : \rho = 0, H_a : \rho \neq 0$

4.1 Test du score

L'équation de vraisemblance relative à ρ , évaluée aux estimateurs du MV sous l'hypothèse nulle est égale à :

$$\left. \frac{\partial}{\partial \rho} L \right|_{\hat{b}, \hat{\sigma}^2, 0} = \frac{G(0)}{\hat{\sigma}^2} = \frac{\sum_{t=1}^T \hat{u}_t \hat{u}_{t-1}}{\hat{\sigma}^2} = T \frac{\sum_{t=1}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2} = T\hat{\rho} \quad \text{où } \hat{\rho} \text{ est le coefficient}$$

d'autocorrélation du premier ordre des résidus. Par ailleurs on constate que la matrice d'information de Fisher est diagonale sous l'hypothèse nulle et que :

$-E \frac{\partial}{\partial \rho^2} L(b, \sigma^2, 0) = T$. Ainsi, la statistique du test du Score est :

$$\xi_S = \frac{(T\hat{\rho})^2}{T} = T\hat{\rho}^2$$

Sous H_0 , $\xi_S \xrightarrow{L} \chi^2(1)$ et sous H_a , $\xi_S \rightarrow +\infty$. Pour un test de niveau $1-\alpha$, la région critique asymptotique est : $C = \{\xi_S \geq \chi_{1-\alpha}^2(1)\}$.

4.2 Test de Durbin et Watson

Il existe un test exact, avec une région d'indécision, qui repose sur la statistique de test de Durbin et Watson :

$$d = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}$$

Cette statistique est comprise entre 0 et 4 et est proche de $2(1-\hat{\rho})$. On doit donc s'attendre à ce que si $\rho > 0$, d soit inférieur à 2, et que si $\rho < 0$, d soit supérieur à 2, la valeur 2 correspondant à l'absence d'autocorrélation.

La loi de d n'est pas fixe sous l'hypothèse nulle ; elle dépend des valeurs des variables exogènes. Mais on peut trouver deux variables d_L et d_U dont les lois sont indépendantes des valeurs des exogènes sous l'hypothèse nulle et telles que :

$$d_L \leq d \leq d_U$$

les quantiles des lois de ces deux variables « encadrantes » ont été tabulées initialement lorsque le modèle comporte un terme constant et plus tard pour tous les types de modèles.

Mise en œuvre

Problème de test : $H_0 : \rho = 0, H_a : \rho > 0$; seuil : $1 - \alpha$; d_L^*, d_U^* les quantiles d'ordre α des lois de d_L et de d_U sous l'hypothèse nulle, on propose la règle de décision à trois niveaux :

$$\begin{cases} d \leq d_L^* \Rightarrow \text{rejet de } H_0 \\ d \geq d_U^* \Rightarrow H_0 \text{ n'est pas rejetée} \\ d_L^* < d < d_U^* \Rightarrow ? \end{cases}$$

Remarques : 1) par précaution on rejettera l'hypothèse nulle dans la région d'incertitude. Des travaux ont établi le positionnement de la véritable valeur critique dans l'intervalle d'incertitude en fonction de la périodicité des variables exogènes...

2) pour le problème de test $H_0 : \rho = 0, H_a : \rho < 0$ on utilise les mêmes seuils que précédemment avec pour statistique de test $4 - d$.

3) La présence d'un phénomène d'autocorrélation est souvent le signe d'un problème de spécification : omission d'une variable croissante avec le temps, mauvaise modélisation d'une courbure de la forme fonctionnelle.

5. Prédiction en présence d'autocorrélation

5.1 Prédiction dans un modèle linéaire généralisé

Soit le modèle :

$$\begin{cases} y = Xb + u, X(T \times K) \\ Eu = 0, Vu = \Omega \end{cases}$$

on cherche à prédire $y_\tau, \tau > T$, sachant le vecteur des variables exogènes observé en $\tau > T$, x_τ , et $\text{cov}(u_\tau, u') = Eu_\tau u' = \omega'$.

Si $y_\tau^* = E^*(y_\tau / y)$ est la prédiction linéaire de y_τ , on a :

$$y_\tau^* = x_\tau' b + \omega' \Omega^{-1} (y - Xb)$$

5.2 Application

Dans le cas d'un modèle avec autocorrélation et où $\tau = T + s$ on a :

$\omega' = \sigma^2 \rho^s (\rho^{(T-1)}, \rho^{(T-2)}, \dots, \rho, 1)$ on voit donc que : $\omega' \Omega^{-1} = \rho^s (0, 0, \dots, 1)$. On en déduit :

$$y_{T+s}^* = x'_{T+s} b + \rho^s (0, 0, \dots, 0, 1)(y - Xb) = x'_{T+s} b + \rho^s (y_T - x'_T b) = x'_{T+s} b + \rho^s u_T.$$

En pratique on utilise \hat{y}_{T+s}^* comme prédicteur de y_{T+s} :

$$\hat{y}_{T+s}^* = x'_{T+s} \hat{b} + \hat{\rho}^s \hat{u}_T !$$

Chapitre 9 : Théorie des variables instrumentales

1. Généralités

Exemple : le modèle Keynésien élémentaire. Notions de variables endogènes et exogènes. Quelques éléments sur l'exogénéité.

$$\begin{cases} C_t = \alpha + \beta Y_t + u_t \\ Y_t = C_t + I_t, I_t \text{ exogène} \end{cases}$$

Hypothèse : $\text{cov}(I_t, u_t) = 0$

Conséquence : $\text{cov}(Y_t, u_t) \neq 0$

2. Le cadre d'analyse

y est une variable endogène et x_1, x_2, \dots, x_K sont des variables « explicatives » d'une équation :

$$y = \sum_{k=1}^K x_k b_k + u = x'b + u, b \in R^K$$

$$Eu = 0, \exists k / Ex_k u \neq 0$$

On suppose pour simplifier que les observations sur les diverses variables sont issues de tirages indépendants : $(u_n, x_{1n}, \dots, x_{Kn})$ sont des vecteurs aléatoires iid du second ordre ($\in L^2$).

3. Non-convergence des MCO

$$\begin{aligned} \hat{b} &= (\sum_n x_n x_n')^{-1} \sum_n y_n x_n \\ &= b + (\sum_n x_n x_n')^{-1} \sum_n u_n x_n \end{aligned}$$

avec les hypothèses ci-dessus on voit que :

$$\frac{1}{N} \sum_n x_n x_n' \rightarrow Exx', \frac{1}{N} \sum_n u_n x_n \rightarrow Eux, \hat{b} \rightarrow b + (Exx')^{-1} Eux \neq b$$

4. Variables instrumentales

Soit H variables z_1, z_2, \dots, z_H telles que :

$$i) Euz_h = 0, \forall h = 1, 2, \dots, H,$$

$$ii) z_1, z_2, \dots, z_H \text{ est un système indépendant de } L^2, (Ezz')^{-1} \text{ existe,}$$

$$iii) rgEzx' = \min(K, H), \text{ cette matrice est de rang maximum.}$$

Les variables z_1, z_2, \dots, z_H sont des **variables instrumentales**.

Remarque : pour sortir du cadre de l'échantillonnage (iid) on recourt à des hypothèses moins fortes sur les limites en probabilité :

$$P \lim N^{-1} \sum_n u_n z_n = 0, P \lim N^{-1} \sum_n z_n z_n' = Q / Q^{-1} \text{ existe, } P \lim N^{-1} \sum_n z_n x_n' = R / rgR = \min(H, K).$$

Exemple 1 : $y_t = \alpha y_{t-1} + \beta x_t + u_t$, u_t autocorrélé, x_t et x_{t-1} sont des variables instrumentales si x est une variable exogène.

Exemple 2 : $y_n = \beta x_n + u_n$, $E u_n = 0$, $E x_n \neq 0$, $E u_n x_n \neq 0$ alors $z_n = 1, \forall n$ est une variable instrumentale.

Exemple 3 : modèle à équations simultanées.

5. Méthode des variables instrumentales

Hypothèse : $(u_n, x_{1n}, \dots, x_{Kn}, z_{1n}, \dots, z_{Hn})$ sont des vecteurs aléatoires iid de carrés intégrables ($\in L^2$) .

La première caractéristique de variables instrumentales implique :

$$Ez_h y = Ez_h x' b, h = 1, 2, \dots, H$$

Le paramètre b vérifie ces **conditions de moments** dont les contreparties empiriques sont :

$$\frac{1}{N} \sum_n y_n z_{hn} = \frac{1}{N} \sum_n z_{hn} x_n' b, h = 1, 2, \dots, H$$

Soit matriciellement avec des notations standards :

$$Z' y = Z' X b$$

- i) Si $H < K$ le système ci-dessus a une infinité de solution,
- ii) SI $H = K$ le système ci-dessus a une seule solution : $\tilde{b} = (Z'X)^{-1}Z'y$
- iii) Si $H > K$, le système n'est en général pas compatible ; le problème est dit **suridentifié**.

Dans le cas iii), on se ramène au cas ii) en réduisant le nombre d'instruments : $z_A = Az$ où A est une matrice de taille $K \times H$. Puisque $Z_A = ZA'$ on estime b par :

$$\tilde{b}(A) = (AZ'X)^{-1}AZ'y$$

Ces estimateurs sont appelés **estimateurs des variables instrumentales**.

Propriété : Par construction : $\tilde{b}(A) \xrightarrow{N \rightarrow \infty} b$. La convergence est presque sûre ou en probabilité selon les hypothèses retenues.

6. Loi asymptotique des estimateurs des variables instrumentales.

Il faut décrire le modèle jusqu'au second ordre et on le fait ici conditionnellement aux variables explicatives et instrumentales :

$$\begin{cases} y = x'b + u \\ E(u/x) = 0, E(u/z) = 0, V(u/z) = \sigma^2 \end{cases}$$

On se place dans le cas iid.

a) Cas $H = K$:

Propriété : $\sqrt{N}(\tilde{b} - b) \xrightarrow{L} N(0, \Sigma), \Sigma = \sigma^2(Ezx')^{-1}Ezz'(Exz')^{-1}$

Remarque : la matrice Σ peut être estimée par

$\tilde{\Sigma} = \tilde{\sigma}^2 \left(\frac{Z'X}{N} \right)^{-1} \frac{Z'Z}{N} \left(\frac{X'Z}{N} \right)^{-1} = N \tilde{\sigma}^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}$ où $\tilde{\sigma}^2 = \frac{\|y - X\tilde{b}\|^2}{N}$ et une

« approximation » de la variance de \tilde{b} est $\hat{V}\tilde{b} = \tilde{\sigma}^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}$.

b) Cas $H > K$: $z_A = Az$

Propriété : $\sqrt{N}(\tilde{b}(A) - b) \xrightarrow{L} N(0, \Sigma(A)), \Sigma(A) = \sigma^2 (AEzx')^{-1} AEzz' A' (Exz' A')^{-1}$

Lorsqu'il y a suridentification le choix de A peut être résolu en minimisant $\Sigma(A)$.

Estimateur des doubles moindres carrés.

Propriété : Si $A^* = Exz'(Ezz')^{-1}$ alors $\Sigma(A) - \Sigma(A^*)$ est une matrice semi-définie positive $\forall A$, de dimension $K \times H$.

L'estimateur des variables instrumentales optimal est donc fondé sur $z_{A^*} = Exz'(Ezz')^{-1}z$. En pratique on utilise $\hat{A}^* = X'Z(Z'Z)^{-1}$ qui conduit à l'estimateur des **doubles moindres carrés** :

$$\hat{b}_{2MC} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$$

Remarques : 1) Remplacer A^* par \hat{A}^* ne modifie pas la loi asymptotique.

2) L'estimateur des doubles moindres carrés peut être calculé en deux étapes : on projette orthogonalement le modèle sur l'espace linéaire engendré par Z ($P_Z = Z(Z'Z)^{-1}Z'$) et dans une seconde étape on régresse $P_Z y$ sur $P_Z X$.

7. Application : modèle à erreurs sur les variables

Présentation sur le modèle linéaire simple.

7.1 Définition

$$y_n = a + bx_n + u_n, n = 1, 2, \dots, N$$

Les observations sont iid et x est indépendant de u ; x est exogène. Mais on ne mesure pas correctement x . x est mesuré par x^* telle que : $x^* = x + \varepsilon$ où ε est une perturbation. Il vient donc :

$$y_n = a + bx_n^* + v_n, v_n = u_n - b\varepsilon_n, n = 1, 2, \dots, N$$

$$Ev_n = 0, E\varepsilon_n = 0, V(\varepsilon_n) = \sigma_\varepsilon^2$$

$$\text{cov}(x_n^*, v_n) = -b\sigma_\varepsilon^2$$

7.2 Biais asymptotique des MCO

$$\hat{b} = \frac{\sum_n (x_n^* - \bar{x}^*)(y_n - \bar{y})}{\sum_n (x_n^* - \bar{x}^*)^2} \rightarrow \frac{\text{cov}(x^*, y)}{\text{var}(x^*)} = b + \frac{\text{cov}(x^*, v)}{\text{var}(x^*)} = b - \frac{b\sigma_\varepsilon^2}{\text{var}(x^*)}$$

remarques : 1) on dit que l'estimateur des MCO est biaisé vers zéro.

2) ce biais ne dépend pas de l'erreur sur l'équation : u_n .

7.3 Problèmes d'identification : modèles gaussiens

Le modèle : $y = a + bx, x \approx N(Ex, Vx) \Rightarrow Ey = a + bEx, Vy = b^2Vx$

Les variables observables : $y^* = y + u, u \approx N(0, \sigma_u^2)$
 $x^* = x + \varepsilon, \varepsilon \approx N(0, \sigma_\varepsilon^2)$

6 paramètres libres à estimer : $a, b, Ex, Vx, \sigma_u^2, \sigma_\varepsilon^2$. Mais la loi des variables observables ne dépend que de 5 paramètres libres : $Ey^*, Ex^*, Vy^*, \text{cov}(y^*, x^*), Vx^*$ puisque :

$$\begin{pmatrix} y^* \\ x^* \end{pmatrix} \approx N \left(\begin{pmatrix} Ey^* \\ Ex^* \end{pmatrix}, \begin{pmatrix} Vy^* & \text{cov}(x^*, y^*) \\ \text{cov}(x^*, y^*) & Vx^* \end{pmatrix} \right)$$

Informations identifiantes :

a) $a = 0 \Rightarrow b = \frac{Ey^*}{Ex^*}$, estimable par $b^* = \frac{\bar{y}^*}{\bar{x}^*}$

b) $\sigma_\varepsilon^2 = k\sigma_u^2, k$ connu : lorsque $k=1$: régression orthogonale.

c) variable instrumentale : $\text{cov}(z, \varepsilon) = \text{cov}(z, u) = 0$ et $\text{cov}(z, x) \neq 0$: $b = \frac{\text{cov}(z, y^*)}{\text{cov}(z, x^*)}$

Exemple : le revenu permanent.

1. Méthodes des moments généralisés

Le modèle de variable instrumentale ne définit l'identification des paramètres (du premier ordre) que par les conditions dites d'orthogonalités :

$$Ez_{hn}(y_n - x_n'b) = 0, h = 1, 2, \dots, H$$

ou en notation vectorielle $E(y_n - x_n'b)z_n = 0$

Ce cadre « linéaire » a été généralisé à des fonctions quelconques pour l'identification d'un paramètre θ estimable à partir de variables observables w_n telles que il existe un ensemble de fonctions $h_j(w_n, \theta), j = 1, 2, \dots, H$ vérifiant :

$$Eh_j(w_n, \theta) = 0, j = 1, 2, \dots, H$$

ou $Eh(w_n, \theta) = 0, h(\cdot, \cdot)' = (h_1(\cdot, \cdot), \dots, h_H(\cdot, \cdot))$

Ces H conditions sont supposées conduire à l'identification du paramètre d'intérêt θ . Lorsque la densité de la loi de w_n est $f(w, \theta_0)$ cette condition d'identification se traduit en :

$$Eh_j(w_n, \theta) = \int h_j(w, \theta) f(w, \theta_0) dw = 0, j = 1, 2, \dots, H \Leftrightarrow \theta = \theta_0$$

Dans le cas du modèle de variables instrumentales où $\theta = b$ et $h(w_n, \theta) = (y_n - x_n' b) z_n$ cette condition impose que : $H \geq K$ et $Ez_n x_n'$ soit de rang maximum.

Pour estimer le paramètre θ on distingue deux cas : le cas juste identifié ($H = K$) et le cas suridentifié ($H > K$).

a) $H = K$: on estime θ par la méthode des moments en résolvant le système (en général non-linéaire) :

$$\sum_{n=1}^N h_j(w_n, \theta) = 0, j = 1, 2, \dots, H$$

Ces équations ont le plus souvent une seule solution $\hat{\theta}_N$.

Exemple : $\sum_{n=1}^N (y_n - x_n' b) z_n = Z'(y - Xb) = 0 \Leftrightarrow \hat{b}_N = (Z'X)^{-1} Z'y$.

b) $H > K$: le système précédent est en général incompatible ; on estime θ par la méthode des **moments généralisés** qui consiste à minimiser le critère suivant :

$$\hat{\theta}(S) = \arg \min_{\theta} (\sum_n h(w_n, \theta))' S (\sum_n h(w_n, \theta))$$

où S est une matrice symétrique définie positive.

Remarque : la méthode des moments généralisés se réduit à la méthode des moments lorsque $H = K$.

Propriétés asymptotiques : 1) $\hat{\theta}(S)$ est un estimateur convergent.

$$2) \sqrt{N}(\hat{\theta}(S) - \theta) \xrightarrow{L} N(0, \Sigma(S))$$

$$\text{avec } \Sigma(S) = \left[E \frac{\partial h'}{\partial \theta} SE \frac{\partial h}{\partial \theta'} \right]^{-1} E \frac{\partial h'}{\partial \theta} SVhSE \frac{\partial h}{\partial \theta'} \left[E \frac{\partial h}{\partial \theta'} SE \frac{\partial h'}{\partial \theta} \right]^{-1}$$

Remarque : la deuxième partie de ce théorème se démontre très facilement dans le cas iid. Lorsqu'on a affaire à des processus ergodiques ces résultats restent valables.

Exemple : $\sum_n h_n = Z'(y - Xb), S = (Z'Z)^{-1}$ la fonction objectif est $(y - Xb)'Z(Z'Z)^{-1}Z'(y - Xb)$ qui conduit à $\hat{b}((Z'Z)^{-1}) = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$ c'est-à-dire à l'estimateur des doubles moindres carrés.

Propriété d'optimalité asymptotique : Si $S^* = Vh^{-1}$ alors $\Sigma(S) - \Sigma(S^*)$ est semi-définie positive $\forall S$ symétrique définie positive.

Estimation en deux étapes : L'estimateur optimal fondé sur S^* n'est pas utilisable en pratique. On utilise donc \hat{S}^* fondé sur une estimation préalable de la variance Vh . Ainsi on calcule par exemple dans un premier temps $\hat{\theta}(I)$ l'estimateur des moments généralisés avec $S = I$ et on en déduit :

$$\hat{V}h = \frac{1}{N} \sum_n h(w_n, \hat{\theta}(I))h(w_n, \hat{\theta}(I))', \hat{S}^* = \hat{V}h^{-1} \text{ et } \hat{\theta}(\hat{S}^*).$$

Cet estimateur de θ possède les mêmes propriétés asymptotiques que celui fondé sur S^* ; il atteint donc la borne de la variance asymptotique des estimateurs des « MMG » (GMM en anglais).

Exemple : dans le cas des variables instrumentales : $\hat{b}(I) = (X'ZZ'X)^{-1}X'ZZ'y, \hat{u}_n = y_n - x_n'\hat{b}(I), h(w_n, \hat{b}(I)) = \sum_n \hat{u}_n z_n$.

On obtient une estimation préalable de Vh :

$$\hat{V}h = \frac{1}{N} \sum_n \hat{u}_n^2 z_n z_n' = Z' \hat{D}Z \text{ avec } \hat{D} = \text{diag}(\hat{u}_1^2, \hat{u}_2^2, \dots, \hat{u}_N^2).$$

L'estimateur en deux étapes de b est finalement égal à :

$$\hat{b}(\hat{S}^*) = (X'Z(Z'\hat{D}Z)^{-1}Z'X)^{-1}X'Z(Z'\hat{D}Z)^{-1}Z'y$$

C'est l'estimateur optimal des MMG ; il surclasse l'estimateur des doubles moindres carrés lorsqu'il y a hétéroscédasticité : Vh_n dépend de n . Cf. le chapitre 6 et l'estimation robuste de White pour la variance des MCO...

Chapitre 10 : Séries temporelles

1. Introduction

Les séries temporelles modélisent des données mesurées à des intervalles de temps réguliers (par exemple des données macroéconomiques, des données financières,...). On peut considérer une série temporelle univariée ou multivariée (processus vectoriel). On fera des études en temps discret donc on indicera les séries de façon dénombrable par l'entier relatif t .

2. Formalisme, définitions, premières propriétés

On considère les observations, sur des dates $1, \dots, T$, x_1, \dots, x_T réalisations des variables aléatoires $X_1, \dots, X_T : (\Omega, \mathcal{A}, P) \rightarrow R$, $\omega \in \Omega$ est un état de la nature tel que $x_t = X_t(\omega)$.

Dans la suite, on supposera $\forall t \in Z, X_t \in L^2(\Omega, \mathcal{A}, P)$

On appelle $(X_t)_{t \in Z}$ **processus stochastique** et $(x_t)_{t \in Z}$ est une **trajectoire du processus**.

2.1 Stationnarité

Définition $(X_t)_{t \in Z}$ est un **processus stationnaire au sens strict** (ou **fortement stationnaire**) si et seulement si :

$$\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n), \forall h \in Z, \text{ la loi de } (X_{t_1}, \dots, X_{t_n}) \text{ est identique à la loi de } (X_{t_1+h}, \dots, X_{t_n+h})$$

Définition $(X_t)_{t \in Z}$ est un **processus stationnaire du second ordre** (ou **faiblement stationnaire**) si et seulement si :

- (i) $\forall t \in Z, E(X_t) = m$
- (ii) $\forall t \in Z, \forall h \in Z, Cov(X_t, X_{t+h}) = \gamma(h)$

$\gamma(h)$ est l'**auto-covariance** d'ordre h du processus.

A partir de (ii), on déduit, pour $h=0$: $\forall t \in Z, V(X_t) = \sigma^2 = \gamma(0)$

Si un processus est stationnaire au sens strict alors il est faiblement stationnaire.
Si le processus est gaussien alors il est stationnaire au sens strict si et seulement s'il est faiblement stationnaire.

Lorsque $m=0$ et, pour $h \neq 0$, $\gamma(h)=0$, le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un **bruit blanc faible** noté $BB(0, \sigma^2)$. Si on suppose également que les ε_t sont *i.i.d.*, alors le processus est un **bruit blanc fort**.

Remarque Une marche aléatoire et une tendance déterministe ne sont pas stationnaires.

Définition On appelle **fonction d'autocorrélation** d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ la fonction à valeurs dans \mathbb{C} définie par :

$$\forall h \in \mathbb{Z}, \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \text{Corr}(X_t, X_{t+h})$$

Sa représentation graphique est appelée **auto-corrélogramme théorique**.

2.2 Polynômes retard et avance

Définition L'**opérateur retard** L (lag) ou B (backward) est défini sur la classe des processus stationnaires comme étant :

$$L : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = X_{t-1}, \text{ noté } LX_t = X_{t-1}$$

De même :

$$F : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = X_{t+1}, \text{ noté } FX_t = X_{t+1}$$

Définition A tout polynôme P de degré p , on associe un polynôme retard $P(L)$ défini par :

$$P(L)X_t = \left(\sum_{k=0}^p a_k L^k \right) X_t = \sum_{k=0}^p a_k X_{t-k}$$

On définit de façon similaire le polynôme avance à partir de l'opérateur F .

Remarque On élargit cette définition à des séries entières en L ou en F (considérées comme des polynômes de degré infini).

Définition $A(L)$ est **inversible** si et seulement s'il existe $B(L)$ série entière telle que : $A(L) \circ B(L) = Id$

Remarque Tout polynôme étant scindé dans \mathbb{C} , l'inversion du polynôme se ramène à l'inversion de monômes de degré 1 de la forme $(1-\lambda L)$ dont on montre qu'ils sont inversibles lorsque $|\lambda| \neq 1$.

Proposition

- (i) ϕ est inversible si et seulement si ses racines sont de modules distincts de 1.
- (ii) Si toutes les racines sont de module strictement supérieur à 1, alors l'inverse de ϕ est une série entière de rayon de convergence supérieur à 1.

2.3 Transformation d'un processus stationnaire par une moyenne mobile infinie

Définition-Proposition Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $(a_j)_{j \in \mathbb{Z}}$ une suite de réels tels que $\sum_j |a_j| < +\infty$. Alors $Y_t = \sum_j a_j X_{t-j}$ est défini (p.s.) pour tout t . On a les propriétés suivantes :

- (i) $\forall t \in \mathbb{Z}, Y_t \in L^2(\Omega, \mathcal{A}, \mathbb{P})$
- (ii) $(Y_t)_{t \in \mathbb{Z}}$ est un processus stationnaire tel que

$$EY_t = m_Y = \left(\sum_{j \in \mathbb{Z}} a_j \right) m_X$$

$$\forall h \in \mathbb{Z}, \gamma_Y(h) = \sum_{j, k \in \mathbb{Z}} a_j a_k \gamma(h + k - j)$$

On dit que $(Y_t)_{t \in \mathbb{Z}}$ est la **transformée de $(X_t)_{t \in \mathbb{Z}}$ par la moyenne mobile infinie associée aux $(a_j)_{j \in \mathbb{Z}}$** .

Définition Si $X_t = \varepsilon_t \rightarrow BB(0, \sigma^2)$ alors $Y_t \rightarrow MA(\infty)$.

2.4 Autocorrélation partielle

Définition On appelle **autocorrélation partielle** d'ordre p

$$\begin{aligned} r(p) &= \text{Corr}(X_t - EL(X_t | X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p} | X_{t-1}, \dots, X_{t-p+1})) \\ &= \frac{\text{Cov}(X_t - EL(X_t | X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p} | X_{t-1}, \dots, X_{t-p+1}))}{\sqrt{V(X_t - EL(X_t | X_{t-1}, \dots, X_{t-p+1}))V(X_{t-p} - EL(X_{t-p} | X_{t-1}, \dots, X_{t-p+1}))}} \end{aligned}$$

On montre que $r(p) = a_p$ coefficient de X_{t-p} dans $EL(X_t | X_{t-1}, \dots, X_{t-p})$.

Pour la démonstration, on utilise le théorème de Frisch-Waugh.

On appelle **auto-corrélogramme partiel** la représentation graphique associée.

2.5 Innovations

Définition La **régression affine théorique** de X_t sur $X_{t-1}, \dots, X_{t-p}, \dots$ est la projection orthogonale dans $L^2(\Omega, A, P)$ de X_t sur $H = \text{vect}(1, X_{t-1}, \dots, X_{t-p}, \dots)$. On note $\underline{X}_{t-1} = X_{t-1}, \dots, X_{t-p}, \dots$ et $\bar{L}(\underline{X}_{t-1})$ l'espace H .

Théorème Soient $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire, $X_t^* = EL(X_t | \underline{X}_{t-1})$ la régression affine de X_t sur $\bar{L}(\underline{X}_{t-1})$ et $\varepsilon_t = X_t - X_t^*$, alors

- (i) $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible
- (ii) $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ pour tout $k > 0$

Définition (Processus des innovations) Avec les notations ci-dessus :

- (i) $(\varepsilon_t)_{t \in \mathbb{Z}}$ est le **processus des innovations** de $(X_t)_{t \in \mathbb{Z}}$
- (ii) ε_t est l'**innovation** de X_t
- (iii) X_t^* est la **prévision optimale** de X_t à la date $t-1$.

Théorème de Wold Soient $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $(\varepsilon_t)_{t \in \mathbb{Z}}$ le processus des innovations correspondant.

Alors

$$\exists (a_k)_{k \in \mathbb{Z}} \left| \sum_k |a_k| < +\infty \right. \text{ et } X_t = m + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k}$$

2.6 Densité spectrale et auto-corrélations inverses

Proposition-définition (densité spectrale) Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de la forme :

$$X_t = m + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k} \text{ où } (\varepsilon_t)_{t \in \mathbb{Z}} \rightarrow BB \text{ et } \sum_k |a_k| < +\infty$$

alors :

- (i) $\sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$
- (ii) $\forall \omega \in [-\pi, \pi], f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{i\omega h}$

f_X est la **densité spectrale** de $(X_t)_{t \in \mathbb{Z}}$

Proposition Sous les hypothèses précédentes, avec les mêmes notations,

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \cos(\omega h)$$

Proposition (injectivité) Sous les hypothèses précédentes, avec les mêmes notations,

$$\gamma_X(h) = \int_{-\pi}^{\pi} f_X(\omega) e^{-i\omega h} d\omega = \int_{-\pi}^{\pi} f_X(\omega) \cos(\omega h) d\omega$$

On en déduit que $(X_t)_{t \in \mathbb{Z}}$ est entièrement caractérisé par sa densité spectrale f_X .

Définition (Auto-corrélations inverses)

Soit $(X_t)_{t \in \mathbb{Z}}$ tel que $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ où $(\varepsilon_t)_{t \in \mathbb{Z}} \rightarrow BB$ et $\sum_{j \in \mathbb{Z}} |a_j| < +\infty$

On suppose que $\omega \mapsto \frac{1}{f_X(\omega)} e^{-i\omega h}$ est intégrable sur $[-\pi, \pi]$.

On appelle **auto-covariance inverse** d'ordre h de $(X_t)_{t \in \mathbb{Z}}$

$$\gamma_X^i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} e^{-i\omega h} d\omega$$

L'**auto-corrélation inverse** d'ordre h est alors définie comme

$$\rho_X^i(h) = \frac{\gamma_X^i(h)}{\gamma_X^i(0)}$$

Le graphe de l'auto-corrélation inverse est appelé **auto-corrélogramme inverse**.

2.7 Estimations

On considère un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ tel que pour tout $t \in \mathbb{Z}$, $E(X_t) = m$.
On estime les grandeurs $\gamma(h)$, $\rho(h)$, $r(h) = a_h^h$, $f(\omega)$, $\gamma^i(h)$, $\rho^i(h)$ et ceci sachant qu'on observe X_1, \dots, X_T (on assimile ici le processus et ses observations dans les notations).

On prend comme estimateurs :

$$\hat{m} = \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_T \text{ moyenne empirique,}$$

$$\hat{\gamma}_X(h) = \frac{1}{T-h-1} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T) \text{ auto-covariance empirique d'ordre } h$$

(estimation acceptable si h n'est pas trop grand).

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$$

$$\hat{r}(h) = \hat{a}_h^h \text{ dans la régression empirique (m.c.o.) de } x_t \text{ sur } 1, x_1, \dots, x_{t-h}$$

$\hat{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-H}^H \hat{\gamma}_X(h) e^{i\omega h}$, le problème ici étant qu'on voudrait un H suffisamment grand mais prendre un H trop grand est risqué pour l'estimation de $\hat{\gamma}_X(h)$. On prend alors un estimateur corrigé :

$$\hat{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-H}^H \underbrace{\left(1 - \frac{|h|}{H+1}\right)}_{\text{coefficient de Newey-West}} \hat{\gamma}_X(h) e^{i\omega h}$$

Proposition Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire alors tous les estimateurs présentés ci-dessus sont convergents.

Démonstration Application de la loi des grands nombres.

Proposition Si $X_t = m + \sum_{j \geq 0} a_j \varepsilon_{t-j}$ où $E(\varepsilon_t^4) = \eta < +\infty$, alors tous ces estimateurs ont des lois jointes asymptotiquement gaussiennes dont les matrices de variance-covariance sont calculables. On peut en déduire des intervalles de confiance empiriques.

3. Processus ARMA et ARIMA : propriétés, estimation et prévision

3.1 Processus ARMA(p,q)

Définition Un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une représentation ARMA (p,q) canonique minimale s'il vérifie une équation :

$$\phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

où

- (i) $(\varepsilon_t)_{t \in \mathbb{Z}} \rightarrow BB$
- (ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, avec $\varphi_p \neq 0$
- (iii) $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, avec $\theta_q \neq 0$
- (iv) ϕ et θ ont toutes leurs racines de module strictement supérieur à 1 (représentation canonique).
- (v) ϕ et θ n'ont pas de racine commune (représentation minimale).

Remarques

- Il existe des solutions non stationnaires (qui ne sont pas des ARMA).

- Si $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, alors les racines de ϕ sont de module distinct de 1. On pourrait considérer le cas où θ a des racines de module 1 (c'est compatible avec la stationnarité).
- Si on suppose que ϕ et θ ont toutes leurs racines de module distincts de 1, on peut toujours se ramener à la représentation canonique avec des racines de module >1 .

Proposition

(i) $EX_t = \frac{\mu}{\phi(1)} = m$

(ii) $\phi(L)(X_t - m) = \theta(L)\varepsilon_t$

Remarque par centrage, on peut donc se ramener au cas où $\mu = 0$

Proposition (auto-covariance et auto-corrélation)

(i) Pour $h > q$, les $\gamma(h)$ et les $\rho(h)$ vérifient les équations de récurrence d'ordre p :

$$\gamma(h) - \phi_1\gamma(h-1) - \dots - \phi_p\gamma(h-p) = 0$$

$$\rho(h) - \phi_1\rho(h-1) - \dots - \phi_p\rho(h-p) = 0$$

(ii) Elles décroissent donc vers 0 exponentiellement avec h , pour $h > q$.

Equations de Yule-Walker L'équation précédente pour $h = q+1, \dots, q+p$ donne :

$$\begin{pmatrix} \rho(q) & \cdots & \rho(q+p-1) \\ \vdots & \ddots & \vdots \\ \rho(q+p-1) & \cdots & \rho(q) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(p+1) \\ \vdots \\ \rho(p+q) \end{pmatrix}$$

Quand ρ est connu ou estimé, on peut alors calculer les ϕ_j

Inversement, quand les ϕ_j sont connus, on calcule $\rho(p+1), \dots, \rho(p+q)$ qui seront les conditions initiales pour le calcul de $\rho(h)$ tel que $h > q$

Définition Une représentation ARMA(p,0) est appelée représentation **AR(p)**. Une représentation ARMA(0,q) est appelée représentation **MA(q)**. On définit également les représentations $AR(\infty)$ et $MA(\infty)$ de manière évidente.

Proposition Sous les hypothèses précédentes,

(i) $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $AR(\infty)$,

$$\sum_{k=0}^{+\infty} a_k X_{t-k} = \mu + \varepsilon_t \quad \text{où } a_0 = 1 \quad \text{et} \quad \sum_{k=0}^{+\infty} |a_k| < +\infty$$

(ii) $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $MA(\infty)$,

$$X_t = m + \sum_{k=0}^{+\infty} b_k \varepsilon_{t-k} \quad \text{où } b_0 = 1 \text{ et } \sum_{k=0}^{+\infty} |b_k| < +\infty$$

(iii) $\bar{L}(X_t) = \bar{L}(\varepsilon_t)$

(iv) $(\varepsilon_t)_{t \in \mathbb{Z}}$ est l'innovation de $(X_t)_{t \in \mathbb{Z}}$

Remarque

En vertu du théorème de Wold, $X_t = m + B(L)\varepsilon_t$ où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est le processus des innovations, si de plus $X_t \rightarrow ARMA(p, q)$ alors $B(L) = \frac{\theta(L)}{\phi(L)}$

Propriétés On peut résumer dans le tableau suivant les propriétés des modèles **AR(p)** et **MA(q)**

	AR(p)	MA(q)
$\rho(h)$	décroît exponentiellement vers 0 avec h	0 si $ h > q$ et non nul si $h = q$
$r(h)$	0 si $ h > p$ et non nul si $h = p$	-
$\rho^i(h)$	0 si $ h > p$ et non nul si $h = p$	décroît exponentiellement vers 0 avec h

Remarques

- $ARMA(p, q) \equiv AR(\infty) \approx AR(P)$ si P grand
- $ARMA(p, q) \equiv MA(\infty) \approx MA(Q)$ si Q grand

Souvent, l'un des paramètres (p ou q) est petit alors que l'autre est grand. Avec l'approximation précédente on a alors moins de paramètres à estimer.

3.2. Processus ARIMA(p,d,q)

Définition (représentation canonique minimale) $(X_t)_{t \geq -pd}$ est un processus ARIMA(p,d,q) en représentation canonique minimale s'il vérifie une équation du type :

$$\forall t \geq 0, (1-L)^d \phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

où

- (i) $(\varepsilon_t)_{t \in \mathbb{Z}} \rightarrow BB(0, \sigma^2)$
- (ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, avec $\varphi_p \neq 0$
- (iii) $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, avec $\theta_q \neq 0$

- (iv) ϕ et θ ont toutes leurs racines de module strictement supérieur à 1 (représentation canonique).
- (v) ϕ et θ n'ont pas de racine commune (représentation minimale).
- (vi) conditions initiales

$$Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$$

telles que $\text{cov}(Z, \varepsilon_t) = 0$

Remarque Dès que $d \geq 1$, ces processus sont non stationnaires.

Exemple la marche aléatoire est un processus ARIMA(0,1,0).

Proposition $(1-L)^d \phi(L)X_t = \phi(L)(1-L)^d X_t$, on pose $Y_t = (1-L)^d X_t$. Y_t suit alors le processus :

$$\phi(L)Y_t = \mu + \theta(L)\varepsilon_t$$

Sous les hypothèses précédentes, $Y_t = (1-L)^d X_t$ est alors asymptotiquement équivalent à un processus ARMA(p,q).

Proposition (Approximation auto-régressive d'un ARIMA(p,d,q)) :

Avec les notations précédentes,

$$\left. \begin{array}{l} \exists A_t(L), A_t(L) = \sum_{j=0}^t a_j^t L^j \text{ et } a_0^t = 1 \\ \exists \mu_0 \\ \exists h(t) \in R^{p+d+q} \text{ et } \lim_{t \rightarrow +\infty} h(t) = 0 \end{array} \right\} \text{tels que } A_t(L)X_t = \mu_0 + \varepsilon_t + h(t)'Z \Leftrightarrow X_t = -\sum_{j=1}^t a_j^t X_{t-j} + \varepsilon_t + h(t)'Z$$

Proposition (Approximation moyenne mobile d'un ARIMA(p,d,q)) :

Avec les notations précédentes, sous les mêmes hypothèses,

$$\left. \begin{array}{l} \exists B_t(L), B_t(L) = \sum_{j=0}^t b_j^t L^j \text{ et } b_0^t = 1 \\ \exists \mu_1 \\ \exists \tilde{h}(t) \in R^{p+d+q} \text{ et } \lim_{t \rightarrow +\infty} \tilde{h}(t) = 0 \end{array} \right\} \text{tels que } X_t = \mu_1 + B_t(L)\varepsilon_t + \tilde{h}(t)'Z$$

Corollaire $\forall t, \varepsilon_t \in L(X_0, \dots, X_t, 1, Z)$

$X_t \in L(\varepsilon_0, \dots, \varepsilon_t, 1, Z) \Rightarrow \varepsilon_t = X_t - EL(X_0 | X_0, \dots, X_t, 1, Z)$ est le processus des innovations.

Proposition (Calcul de $E(X_t)$) Si l'on note $m_t = E(X_t)$ alors m_t vérifie $\psi(L)m_t = \mu$. On obtient ainsi :

Une équation de récurrence dont le polynôme caractéristique est $z^{p+d+1}\psi\left(\frac{1}{z}\right)$,

Une forme générale de la solution (pour $\mu = 0$ et $\mu \neq 0$).

3.3 Estimation d'un processus ARIMA

On dispose d'observations x_1, \dots, x_T de X_1, \dots, X_T qu'on souhaite modéliser par un ARMA ou un ARIMA. La démarche pour l'identification est la suivante :

- (i) Choix de d
- (ii) Choix de (p, q)
- (iii) Estimation des coefficients (ce qui peut se faire par le maximum de vraisemblance sous l'hypothèse que les sont i.i.d.).
- (iv) Phase de vérification

3.3.a Choix de d

Si $X_t \rightarrow ARMA(p, q)$, $\rho(h)$ décroît exponentiellement vers 0 avec h (pour $h > q$). Si $(X_t)_{t \in \mathbb{Z}}$ admet une racine unité, ce n'est plus vrai, il y a **persistance des chocs**. En pratique, lorsque les $\hat{\rho}_X(h)$ restent proches de 1 ou décroissent linéairement avec h alors le processus est sans doute non stationnaire. Dans ce cas, on considère la série des différences premières dont on estime à nouveau l'auto-corrélogramme. On réitère la procédure jusqu'à l'obtention d'un auto-corrélogramme présentant une décroissance exponentielle.

On peut tester l'existence (ou l'absence) de racine unité. Soulignons 4 tests de racines unités usuels. Dans les trois premiers (Dickey-Fuller, Phillips-Perron, Schmidt-Phillips), l'hypothèse nulle est l'hypothèse de non-stationnarité de la série. Dans le quatrième (test KPSS), l'hypothèse nulle est celle de la stationnarité. Ces tests sont peu puissants.

3.3.b Choix de p et q

On suppose d déterminé et on travaille sur la série stationnaire $Y_t = (1-L)^d X_t$ suivant un processus $ARMA(p, q)$.

Choix de p On montre que, pour un $AR(p) : \sqrt{T}(\hat{r}_T(h) - r(h)) \xrightarrow{loi} N(0,1)$
 On teste $H_0 : r(h) = 0$ contre $H_a : r(h) \neq 0$. On refuse H_0 au seuil de 5% si $\sqrt{T}|\hat{r}_T(h)| > 1,96$.

Sur l'auto-corrélogramme partiel, c'est dernier rang significatif.

Choix de q On montre que, pour un $MA(q)$, pour $h > q$:

$$\sqrt{T} \frac{\hat{\rho}_T(h)}{\sqrt{\sum_1^q \hat{\rho}^2(k)}} \xrightarrow{loi} Student$$

On teste $H_0 : \rho(h) = 0$ contre $H_a : \rho(h) \neq 0$ avec un test de Student.

Sur l'auto-corrélogramme, c'est dernier rang significatif.

3.3.c Estimation des coefficients des polynômes

On a choisi d et divers couples (p, q) compatibles avec les données. Le modèle s'écrit :

$$\forall t \geq 0, (1-L)^d \phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

On suppose que $\varepsilon_t i.i.d. \xrightarrow{loi} N(0, \sigma_\varepsilon^2)$

Le paramètre à estimer est $\omega = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$.

On calcule l'estimateur du maximum de vraisemblance (EMV). On a recours pour cela à une procédure de maximisation numérique (on utilise notamment les équations de Yule-Walker pour l'initialisation).

3.3.d Tests sur les résidus

On vérifie que $(\varepsilon_t)_{t \in \mathbb{Z}}$ est bien un bruit blanc avec le **test du porte-manteau** :

$$\hat{\rho}_\varepsilon(k) = \frac{1}{T-1} \sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}$$

L'auto-corrélation empirique d'ordre k de $\hat{\varepsilon}_t$ est alors $Q_K = T \sum_{k=1}^K \hat{\rho}_\varepsilon^2(k)$

Pour K assez grand ($K \geq 12$), on montre que si $(\varepsilon_t)_{t \in \mathbb{Z}} \rightarrow BB(0, \sigma^2)$ alors

$$Q_K \xrightarrow{L} \chi^2(K - p - q)$$

Remarque Q_K peut éventuellement être remplacé par

$$Q_K' = T(T+2) \sum_{k=1}^K \frac{1}{T-k} \hat{\rho}_{\varepsilon}^2(k)$$

3.4 Prédiction

Définition La **prédiction optimale** dans un $ARMA(p,q)$ est définie par :

$${}_t X_{t+j}^* = EL(X_{t+j} | X_t)$$

On trouve alors, dans un $AR(p)$:

$${}_t X_{t+j}^* = \mu + \sum_{j=1}^{h-1} \varphi_j X_{t+h-j}^* + \sum_{j=h}^p \varphi_j X_{t+h-j}$$

dans un $MA(q)$:

$$\text{Si } h > q ; \quad {}_t X_{t+j}^* = m$$

$$\text{Si } h \leq q ; \quad {}_t X_{t+j}^* = m - \sum_{j=h}^q \theta_j \varepsilon_{t+h-j}$$

Cette forme est exacte mais n'est pas utilisable en pratique car les ε_{t+h-j} ne sont pas observés. On peut les calculer à partir des observations en utilisant la forme $AR(\infty)$:

$$\theta(L)^{-1}(X_t - m) = \varepsilon_t \Leftrightarrow \theta(L)^{-1} X_t = \mu + \varepsilon_t \text{ où } \mu = \frac{m}{\theta(1)}$$

On en déduit la prédiction optimale suivante :

$${}_t X_{t+j}^* = \mu - \sum_{j=1}^{h-1} a_j X_{t+h-j}^* - \sum_{j=h}^{\infty} a_j X_{t+h-j}$$

En pratique, on doit négliger les termes contenant X_t pour $t < 0$. On n'a qu'une prédiction approchée en tronquant.

Dans un $ARMA(p,q)$, on se ramène, de même, à une forme $AR(\infty)$

Dans un $ARIMA(p,d,q)$, on se ramène à un $ARMA(p,q)$ en différenciant d fois la série.

4 Processus multivariés, modèles VAR

4.1 Introduction

Définition (processus vectoriel) $(X_t)_{t \in \mathbb{Z}}$ est un processus à valeurs dans R^n si

$$X_t = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{n,t} \end{pmatrix}$$

avec $x_{i,t}$ processus à valeurs dans R .

On généralise ici des définitions et des propriétés vues pour les processus univariés.

Définition (Processus vectoriel du second ordre) $(X_t)_{t \in \mathbb{Z}}$ est un processus du second ordre si et seulement si :

$$\begin{aligned} & \forall t \in \mathbb{Z}, X_t \in L^2_{R^n}(\Omega, A, P) \\ \Leftrightarrow & \quad \forall i, \forall t \in \mathbb{Z}, x_{it} \in L^2(\Omega, A, P) \\ \Leftrightarrow & \quad \forall t \in \mathbb{Z}, X_t' X_t = \|X_t\|_2^2 \in L^1(\Omega, A, P) \end{aligned}$$

Définition (Processus vectoriel stationnaire au second ordre) Soit $(X_t)_{t \in \mathbb{Z}}$ est un processus du second ordre. On dit que $(X_t)_{t \in \mathbb{Z}}$ est stationnaire au second ordre si et seulement si :

- (i) $E(X_t) = m$
- (ii) $E(X_t - m)(X_{t+h} - m)' = \Gamma_X(h)$

Remarques

(ii) $\Rightarrow V(X_t) = \Sigma = \Gamma_X(0)$

$(X_t)_{t \in \mathbb{Z}}$ stationnaire $\Rightarrow \forall i, (x_{it})_{t \in \mathbb{Z}}$ stationnaire. La réciproque est fautive en général.

Proposition $(X_t)_{t \in \mathbb{Z}}$ stationnaire $\Rightarrow \forall h, \Gamma(-h) = \Gamma(h)'$

Proposition Soit $(X_t)_{t \in \mathbb{Z}}$ un processus vectoriel stationnaire. Soit $(A_j)_{j \in \mathbb{Z}}$ une suite de matrices telle que $\sum_{j \in \mathbb{Z}} \|A_j\| < +\infty$. Alors

- (i) $\forall t \in \mathbb{Z}, Y_t = \sum_j A_j X_{t-j} \in L^2_{R^n}(\Omega, A, P)$

(ii) $(Y_t)_{t \in \mathbb{Z}}$ est stationnaire avec :

$$E(Y_t) = m_Y = \left(\sum_j A_j \right) m_X$$

$$\Gamma_Y(h) = \sum_{j,k} A_j \Gamma_X(h+k-j) A_k'$$

Cas particulier Si $X_t = \varepsilon_t \rightarrow BB$, alors $Y_t = \sum_j A_j \varepsilon_{t-j}$ est un processus $VMA(\infty)$

Définition-Proposition (Densité spectrale) Si $X_t = \sum_j A_j \varepsilon_{t-j}$ où $\varepsilon_t \rightarrow BB$ et

$$\sum_{j \in \mathbb{Z}} \|A_j\| < +\infty \text{ alors :}$$

(i) $\sum_{h \in \mathbb{Z}} \|\Gamma_X(h)\| < +\infty$

(ii) $S_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \Gamma_X(h) e^{i\omega h}$ est la matrice de densité spectrale de $(X_t)_{t \in \mathbb{Z}}$

Proposition

(i) Si $\varepsilon_t \rightarrow BB(0, \Omega)$ alors $S_\varepsilon(\omega) = \frac{1}{2\pi} \Omega$

(ii) Soit $X_t = \sum_j A_j \varepsilon_{t-j} = A(L)\varepsilon_t$ tel que $\sum_{j \in \mathbb{Z}} \|A_j\| < +\infty$

et $Y_t = \sum_j B_j X_{t-j} = B(L)X_t$ tel que $\sum_{j \in \mathbb{Z}} \|B_j\| < +\infty$

alors $Y_t = \sum_j C_j \varepsilon_{t-j} = C(L)\varepsilon_t$ et $S_Y(\omega) = B(e^{i\omega}) S_X(\omega) B(e^{-i\omega})'$

Théorème (injectivité) On considère le modèle $X_t = A(L)\varepsilon_t$ où $\varepsilon_t \rightarrow BB(0, \Omega)$ alors :

$$\forall h \in \mathbb{Z}, \Gamma_X(h) = \int_{-\pi}^{\pi} S_X(\omega) e^{-i\omega h} d\omega$$

4.2 Innovation et prévision d'un processus vectoriel

Définition (Prévision) On appelle $X_t^* = EL(X_t | \underline{x}_{t-1})$ **prévision linéaire optimale** de X_t à la date $t-1$

Proposition Soit X_t^* la prévision linéaire optimale de X_t à la date $t-1$. Si \hat{X}_t est une autre prévision linéaire de X_t à la date $t-1$, alors :

$$V(X_t - X_t^*) \leq V(X_t - \hat{X}_t) \text{ (au sens des matrices symétriques positives)}$$

On a égalité si et seulement si $X_t^* = \hat{X}_t$.

Proposition (processus des innovations) Soit $(X_t)_{t \in \mathbb{Z}}$ un processus vectoriel stationnaire et $X_t^* = EL(X_t | x_{t-1})$. L'**innovation** de $(X_t)_{t \in \mathbb{Z}}$ est le processus :

$$\varepsilon_t = X_t - X_t^*$$

On montre que $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.

Remarques

Ce qui a été présenté est valable lorsque $E(X_t) = 0$. Sinon, on définit :

$$X_t^* = EL(X_t | 1, x_{t-1})$$

$(\varepsilon_{it})_{t \in \mathbb{Z}}$ n'est pas l'innovation du processus univarié $(x_{it})_{t \in \mathbb{Z}}$.

Théorème de Wold Soit $(X_t)_{t \in \mathbb{Z}}$ un processus vectoriel stationnaire et $(\varepsilon_t)_{t \in \mathbb{Z}}$ le processus des innovations. On a alors :

$$\exists (A_k)_{k \in \mathbb{Z}} \mid \sum_k \|A_k\| < +\infty \text{ et } X_t = m_t + \sum_{k=0}^{+\infty} A_k \varepsilon_{t-k} \text{ avec } A_0 = I$$

On définit les estimateurs (moments empiriques) suivants :

$$\hat{m} = \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_T \text{ moyenne empirique,}$$

$$\hat{\Gamma}_X(h) = \frac{1}{T-h-1} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)$$

Ces estimateurs sont convergents et asymptotiquement normaux.

4.3 Processus VAR stationnaires

Définition Soit $\mu \in R^n, (\phi_j)_{1 \leq j \leq p} \in (M_n(R))^p, \phi_p \neq 0, \varepsilon_t \rightarrow BB(0, \Omega), \Omega \in M_n(R)$ tel que

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

$$\text{Soit } \phi(L)X_t = \mu + \varepsilon_t \text{ où } \phi(L) = I_n - \phi_1 L + \dots + \phi_p L^p$$

Alors on dit que $(X_t)_{t \in \mathbb{Z}}$ suit un **processus vectoriel auto-régressif** noté $X_t \rightarrow VAR(p)$

Proposition Si $\det \phi(Z)$ a toutes ses racines de modules strictement supérieurs que 1, alors :

- (i) $\phi(L)$ est inversible

(ii) $\phi(L)^{-1} = \sum_{k=0}^{\infty} A_k L^k$ tel que $A_0 = I$ et $\sum_k \|A_k\| < +\infty$.

Remarque (racine unitaire) Si $\det \phi(1) = 0$ alors $(X_t)_{t \in \mathbb{Z}}$ est non stationnaire.

Proposition Si $\phi(L)X_t = \mu + \varepsilon_t$ et si les racines de $\det \phi(Z)$ sont de modules strictement supérieurs à 1, alors

(i) $X_t = m + \sum_{k=0}^{\infty} A_k \varepsilon_{t-k}$ avec $A_0 = I$ et $\sum_k \|A_k\| < +\infty$

(ii) $(\varepsilon_t)_{t \in \mathbb{Z}}$ le processus des innovations de $(X_t)_{t \in \mathbb{Z}}$

Estimation On montre (théorème de Zellner) que l'estimation d'un modèle VAR par les MCQG peut se ramener à l'estimation des MCO processus par processus. Sous l'hypothèse de normalité des résidus, cette estimation coïncide également avec l'EMV.

4.4 Définition de la cointégration

Une combinaison linéaire de séries temporelles non stationnaires peut être stationnaire. Dans ce cas, on dit que les séries sont cointégrées.

Définition Soit $(X_t)_{t \in \mathbb{Z}}$ un intégré d'ordre d . $(X_t)_{t \in \mathbb{Z}}$ est dit **cointégré** si et seulement s'il existe un vecteur de cointégration α tel que $(\alpha'X_t)_{t \in \mathbb{Z}}$ est intégré d'ordre $k < d$.

Chapitre 11: Modèles pour données en panel

1. Introduction

On dispose d'informations sur des individus $i, i = 1, 2, \dots, N$ (ménages, personnes, entreprises, ..., régions, pays, ...) observés au cours du temps $t = 1, 2, \dots, T_i$. On dit que l'échantillon est cylindré lorsque : $T_i = T, \forall i$. La double dimension « individu×temps » renvoie par analogie à un « panneau » d'où le terme emprunté à l'anglais « panel ».

Intérêt : cette double dimension permet d'utiliser des hypothèses d'indépendance dans la dimension individuelle est d'éviter des hypothèses trop restrictives pour les séries temporelles. Ceci ne peut être fait qu'en contrôlant la variabilité des comportements individuels ; c'est-à-dire en contrôlant l'hétérogénéité inobservée.

Exemple : assurance automobile et risques individuels.

2. Le modèle de la covariance

2.1 Définition

Le modèle statistique classique de la covariance offre un cadre intéressant pour introduire une certaine hétérogénéité individuelle en respectant le principe de parcimonie :

$$\begin{cases} y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T \\ E\varepsilon_{it} = 0, V\varepsilon_{it} = \sigma_\varepsilon^2, \text{cov}(\varepsilon_{it}, \varepsilon_{i't'}) = 0, i \neq i' \text{ ou } t \neq t' \end{cases}$$

Dans ce modèle les paramètres sont $(\beta, \alpha_1, \alpha_2, \dots, \alpha_N)$ et σ_ε^2 . Ce modèle respecte un comportement commun au travers de sa partie βx_{it} mais introduit des différences au travers des paramètres α_i encore appelé *effets fixes*. Par ce biais on tient compte très simplement d'une *hétérogénéité individuelle*.

2.2 Ecriture matricielle

Pour chaque individu on pose

$$\underline{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \cdot \\ \cdot \\ y_{iT} \end{pmatrix}, \underline{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ x_{iT} \end{pmatrix}, \underline{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdot \\ \cdot \\ \varepsilon_{iT} \end{pmatrix}$$

et le modèle est classé par individu:

$$\underline{y}_i = \alpha_i \underline{t}_T + \beta \underline{x}_i + \underline{\varepsilon}_i, i = 1, 2, \dots, N, \underline{t}_T' = (1, 1, \dots, 1), \dim \underline{t}_T = T$$

$$E \underline{\varepsilon}_i = 0, V \underline{\varepsilon}_i = \sigma_\varepsilon^2 I_T, \text{cov}(\underline{\varepsilon}_i, \underline{\varepsilon}_{i'}) = 0, i \neq i'$$

puis sur le temps :

$$\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \underline{y}_N \end{pmatrix} = I_N \otimes \underline{t}_T \alpha + \beta \underline{x} + \underline{\varepsilon}, \underline{x} = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \cdot \\ \cdot \\ \underline{x}_N \end{pmatrix}, \underline{\varepsilon} = \begin{pmatrix} \underline{\varepsilon}_1 \\ \underline{\varepsilon}_2 \\ \cdot \\ \cdot \\ \underline{\varepsilon}_N \end{pmatrix}, E \underline{\varepsilon} = 0, V \underline{\varepsilon} = \sigma_\varepsilon^2 I_{NT}$$

On rassemble sur la totalité du panel en un modèle linéaire :

$$\underline{y} = X \gamma + \underline{\varepsilon}, \quad X_{(NT \times (N+1))} = (I_N \otimes \underline{t}_T, \underline{x})$$

$$E \underline{\varepsilon} = 0, V \underline{\varepsilon} = \sigma_\varepsilon^2 I_{NT}$$

On élimine les effets fixes α_i en application du théorème de Frisch et Waugh, c'est-à-dire en appliquant le projecteur :

$$M = I_{NT} - I_N \otimes \underline{t}_T (I_N \otimes \underline{t}_T' I_N \otimes \underline{t}_T)^{-1} I_N \otimes \underline{t}_T' = I_{NT} - \frac{I_N \otimes \underline{t}_T \underline{t}_T'}{T}$$

Ce projecteur opère sur les variables du panel en les transformant en écart à leur moyenne individuelle, exemple :

$$M \underline{y} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ y_{it} - y_{i.} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, y_{i.} = \frac{1}{T} \sum_{t=1}^T y_{it}$$

Pour obtenir l'estimateur des MCO de β il convient donc de régresser $y_{it} - y_{i.}$ sur $x_{it} - x_{i.}$.

Remarques : L'estimateur ainsi obtenu est appelé dans la littérature statistique *estimateur de la covariance* ; en économétrie des panels cet estimateur est appelé *estimateur Within* ; on trouve encore dans des articles en français : *estimateur intra (individuel)*.

On estime les *effets individuels* par $\hat{\alpha}_i = y_{i.} - \hat{\beta}_W x_{i.}$ et la variance de la régression par $\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)-1} \sum_{i=1}^N \sum_{t=1}^T [(y_{it} - y_{i.}) - \hat{\beta}_W (x_{it} - x_{i.})]^2 \dots$ tous ces estimateurs sont sans biais et relèvent du théorème de Gauss-Markov pour ceux du premier ordre...

Ce modèle à effets fixes a été présenté pour un modèle avec une seule variable explicative il se généralise à un nombre quelconque sans aucune difficulté.

3. Le modèle à erreurs composées

3.1 Définition

Le modèle à erreurs composées traite les effets individuels comme des variables aléatoires ; c'est un modèle linéaire généralisé :

$$\begin{cases} y_{it} = x_{it}' \beta + \alpha_i + \varepsilon_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T \\ E\alpha_i = 0, V\alpha_i = \sigma_\alpha^2, \text{cov}(\alpha_i, \alpha_{i'}) = 0, i \neq i' \\ E\varepsilon_{it} = 0, V\varepsilon_{it} = \sigma_\varepsilon^2, \text{cov}(\varepsilon_{it}, \varepsilon_{i't'}) = 0, i \neq i' \text{ ou } t \neq t' \\ \text{cov}(\alpha_i, \varepsilon_{i't'}) = 0, \forall i, i', t \end{cases}$$

avec $x_{it}' = (x_{1it}, x_{2it}, \dots, x_{Kit})$. Ces variables explicatives seront traitées comme fixes dans la présentation ; on abordera le cas où elles sont aléatoires et leur possible corrélation avec les effets individuels et les perturbations dans des sections suivantes. Quoiqu'il en soit le modèle à erreurs composées peut être un modèle conditionnel ; pour cela il suffit de considérer que les espérances et variances de la définitions sont conditionnelles aux variables explicatives... ce qui en fait des variables strictement exogènes, non corrélées avec les effets aléatoires.

3.2 Ecriture matricielle

Si \underline{y} est le vecteur des observations sur la variable endogène comme il est défini en 2.2 et si $X, (N \times K)$ est la matrice des observations sur les variables exogènes rangées en conformité avec le rangement de \underline{y} on a :

$$\begin{cases} \underline{y} = X\beta + u, u = I_N \otimes \iota_T \alpha + \varepsilon \\ Eu = 0, Vu = \Omega = \sigma_\alpha^2 I_N \otimes \iota_T \iota_T' + \sigma_\varepsilon^2 I_{NT} \end{cases}$$

On posera : $J_T = \iota_T \iota_T'$; cette matrice est la matrice carrée de taille T composée uniquement de 1 .

On voit que :

$$\begin{aligned} \Omega &= \sigma_\varepsilon^2 (I_{NT} - \frac{I_N \otimes J_T}{T}) + (\sigma_\varepsilon^2 + T\sigma_\alpha^2) \frac{I_N \otimes J_T}{T} \\ \Omega^{-1} &= \frac{1}{\sigma_\varepsilon^2} [(I_{NT} - \frac{I_N \otimes J_T}{T}) + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} \frac{I_N \otimes J_T}{T}] \\ \sigma_\varepsilon \Omega^{-1/2} &= (I_{NT} - \frac{I_N \otimes J_T}{T}) + \theta \frac{I_N \otimes J_T}{T}, \theta = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}} \end{aligned}$$

3.3 Les différents estimateurs

Estimateur des MCG

De manière générale il s'écrit :

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \underline{y}$$

Il s'obtient, à condition de connaître $\sigma_a^2, \sigma_\varepsilon^2$, en estimant le *modèle sphéricisé* par les moindres carrés ordinaires :

$$\begin{cases} y_{it} - (1-\theta)y_i = (x_{it} - (1-\theta)x_i)' \beta + v_{it} \\ v_{it} = u_{it} - (1-\theta)u_i = \theta\alpha_i + \varepsilon_{it} - (1-\theta)\varepsilon_i, & i = 1, 2, \dots, N, t = 1, 2, \dots, T \\ Ev_{it} = 0, Vv_{it} = \sigma_\varepsilon^2, \text{cov}(v_{it}, v_{i't'}) = 0, i \neq i' \text{ ou } t \neq t' \end{cases}$$

Modèle Within, estimateur Within

La structure de la matrice des variance covariance du modèle à erreurs composées met en évidence le projecteur $W = (I_{NT} - \frac{I_N \otimes J_T}{T})$ qui est apparu dans la transformation de Frisch et Waugh. Le modèle transformé par W est appelé *modèle Within* :

$$W \underline{y} = WX\beta + W \underline{\varepsilon}$$

dont l'écriture détaillée est :

$$y_{it} - y_i = (x_{it} - x_i)' \beta + \varepsilon_{it} - \varepsilon_i.$$

L'estimateur des MCO produit l'*estimateur Within* :

$$\hat{\beta}_W = (X'WX)^{-1} X'W \underline{y}$$

qui n'est rien d'autre que l'estimateur de la covariance.

Remarques : Les variables constantes dans la dimension temporelle, c'est-à-dire indépendantes du temps sont éliminées par la transformation W . Dans le calcul de l'estimateur on se restreint aux autres variables...

L'estimateur de σ_ε^2 sur ce modèle *Within* est l'estimateur correspondant présenté au modèle de la covariance.

Modèle Between, estimateur Between

L'autre projecteur de la matrice Ω est $B = \frac{I_N \otimes J_T}{T}$. Le modèle transformé par B est appelé *modèle Between* :

$$B \underline{y} = BX\beta + Bu$$

Ce projecteur calcule les moyennes individuelles des variables. Ce modèle *Between* est :

$$y_{i.} = x_{i.}'\beta + u_{i.}, u_{i.} = \alpha_i + \varepsilon_i, i = 1, 2, \dots, N$$

La régression de $y_{i.}$ sur $x_{i.}$ donne l'estimateur *Between* $\hat{\beta}_B$ et l'estimateur de la variance de $u_{i.}$, $Vu_{i.} = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{T}$ est défini par :

$$\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}_\varepsilon^2}{T} = \frac{1}{N-K} \sum_{i=1}^N (y_{i.} - x_{i.}'\hat{\beta}_B)^2$$

Estimateur des MCQG

Avec les modèles *Within* et *Between* on obtient des estimateurs sans biais des variances de l'erreur composée. On peut donc calculer un estimateur $\tilde{\beta}$ des MCQG pour le paramètre β ; cet estimateur s'obtient par la régression du modèle sphérisé dans lequel on remplace θ par $\hat{\theta} = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T\hat{\sigma}_\alpha^2}$.

On montre que $\tilde{\beta} = \Delta\hat{\beta}_B + (I_K - \Delta)\hat{\beta}_W$ avec $\Delta = [V(\hat{\beta}_B)^{-1} + V(\hat{\beta}_W)^{-1}]^{-1}V(\hat{\beta}_B)^{-1}$. Les estimateurs des MCG et des MCQG sont donc des moyennes pondérées vectorielles des estimateurs *Within* et *Between*. Pour les MCQG on utilise $\hat{\Delta}$ calculé en fonction des estimation des variance des deux estimateurs. Ces estimateurs *Within* et *Between* ne sont pas corrélés puisqu'ils sont issus de deux modèles qui résultent de projections sur deux sous espaces orthogonaux.

Tous les estimateurs *Between*, *Within*, MCG sont sans biais ; on montre aussi que sous l'hypothèse de normalité des perturbations et des effets, l'estimateur des MCQG est sans biais, mais il faut pour que l'espérance existe que $N \geq K + 5, T \geq 2$.

Remarques : 1) L'estimateur *Between* exploite les différences entre les individus du panel. Il est convergent lorsque $N \rightarrow \infty$ à condition que $Ex_i.\alpha_i = 0$ et $Ex_i.\varepsilon_i = 0$. Ceci est possible lorsque les variables explicatives sont strictement exogènes et sans corrélation avec les effets individuels.

2) L'estimateur *Within* exploite les différence à l'intérieur (intra-within) des séries individuelles. C'est un estimateur convergent en $T \rightarrow \infty$ ou $N \rightarrow \infty$ à condition que $E(x_{it} - x_i.)\varepsilon_{it} = 0$. Cette condition est vérifiée lorsque les variables

explicatives sont strictement exogènes. Mais pour cet estimateur la convergence ne dépend pas des corrélations entre ces variables explicatives et les effets individuels puisque la transformation Within élimine ces effets.

3) L'estimateur des MCG et celui des MCQG sont convergent en N ou $T \rightarrow \infty$ sous les conditions des remarques 1 et 2.

4) Pour mémoire l'estimateur des MCO est convergent en N ou $T \rightarrow \infty$ à condition que $Ex_{it}(\alpha_i + \varepsilon_{it}) = 0$. Ceci est obtenu s'il y a absence de corrélation entre variables explicatives et effets aléatoires et plus simplement qu'en 2 puisqu'il suffit de l'absence de corrélation « contemporaine » avec la perturbation et non pas sur toutes les périodes

4. Test de spécification

4.1 généralités

La littérature économétrique a été animée par le sujet de la spécification que l'on doit retenir pour modéliser l'hétérogénéité. Le modèle de la covariance, à effets fixes, et le modèle à erreurs composées, à effets aléatoires, sont concurrents. Un premier guide pour le choix entre fixe et aléatoire tient au caractère fermé ou ouvert de l'échantillon des individus : si i se rapporte à des pays, des régions, de Etats... la population est « close ». L'indice n'est pas un simple identifiant d'une population plus large et ouverte, il a un sens connu et reconnu. Dans ce cas le recours à un modèle à effet fixe semble se justifier. En revanche dans le cas des ménages ou des personnes, le panel est généralement issu d'un échantillon ; on considère que l'utilisation des effets aléatoires approche mieux l'hétérogénéité de la population ; cette hétérogénéité et donc les effets ne peut être abordée que sous la forme d'une distribution. D'où le recours à l'aléatoire.

Des considération plus statistiques renvoient aux questions de convergence des estimateurs. Si les effets individuels sont des paramètres, c'est-à-dire des effets fixes, on aborde leur loi de manière non paramétrique et puisqu'on ne leur impose aucune contrainte, la corrélation entre les variables explicatives et les effets est abordée de manière implicite. Ce modèle à effet fixe supporte donc l'hypothèse de corrélation entre explicatives et effets. Aujourd'hui les économètres considèrent que le choix entre effets aléatoires et effets fixes relève du choix de la spécification du modèle : accepte-t-on de spécifier la loi conjointe des variables explicatives et des effets, et de leur possible corrélation ? Ou la laisse-t-on libre. Dans le premier cas les effets sont aléatoires, dans le second ils sont fixes.

Le problème de la spécification se concentre donc sur la question de la corrélation effets-variables explicatives. On va donc aborder ces problèmes de spécification en considérant le problème de test : peut-on oui ou non ne pas spécifier des corrélations entre « explicatives » et « effets ».

4.2 Le test d'Hausman

Dans toute cette section les variables explicatives sont strictement exogènes : $Ex_{it}\varepsilon_{is} = 0, \forall t, s$. Et on retient comme problème de test :

$$H_0 : \text{Bonne spécification} : Ex_{it}\alpha_i = 0, \forall i, \forall t$$

$$H_a : Ex_{it}\alpha_i \neq 0$$

Sous H_0 , $\hat{\beta}_B$ et $\hat{\beta}_W$ sont convergents.

Sous H_a , $\hat{\beta}_B$ n'est pas convergent, $\hat{\beta}_W$ est convergent.

En suivant les considérations d'Hausman on voit que la différence entre ces deux estimateurs est une indication du rejet de H_0 : si cette différence est « grande » on rejettera H_0 . Sous H_0 les estimateurs Between et Within ne sont pas corrélés. Sous cette hypothèse on a donc $V(\hat{\beta}_B - \hat{\beta}_W) = V(\hat{\beta}_B) + V(\hat{\beta}_W)$. On note $\xi_H = (\hat{\beta}_B - \hat{\beta}_W)' [V(\hat{\beta}_B) + V(\hat{\beta}_W)]^{-1} (\hat{\beta}_B - \hat{\beta}_W)$ et $\hat{\xi}_H$ la statistique issue de ξ_H où les variances sont remplacées par des estimateurs convergents.

Propriété : Sous l'hypothèse de bonne spécification (H_0) $\xi_H \xrightarrow{L} \chi^2(K)$ lorsque $N \rightarrow \infty$. Mais sous H_a $\xi_H \rightarrow +\infty$.

Statistique d'Hausman : $\hat{\xi}_H = (\hat{\beta}_B - \hat{\beta}_W)' [\hat{V}(\hat{\beta}_B) + \hat{V}(\hat{\beta}_W)]^{-1} (\hat{\beta}_B - \hat{\beta}_W)$ où les variances estimées sont issues des modèles Between et Within.

Test d'Hausman : pour un test asymptotique de seuil 5% la région critique est :

$$C = \left\{ \hat{\xi}_H \geq \chi_{95\%}^2(K) \right\}$$

Remarques : Dans son article d'Econometrica en 1978 Hausman a retenu une autre formulation pour son test. Sa formulation est plus « universelle » que celle proposée ici puisque pour Hausman un test de spécification est fondé sur la différence entre un estimateur convergent asymptotiquement efficace sous H_0 mais non convergent sous H_a et un estimateur convergent sous l'hypothèse de bonne spécification et son alternative. A partir des trois estimateurs Between,

Within et MC(Q)G on produit deux tests au sens strict d'Hausman et celui présenté plus haut. Dans le cas présent ces trois tests sont numériquement égaux.

Il convient de prendre garde au fait que pour le test présenté ci-dessus les formules peuvent ne pas être opératoire. Exemple si dans les variables explicatives certaines ne dépendent pas du temps l'opérateur W les élimine. On se restreint alors aux autres variables¹ pour élaborer le test.

Approche de Mundlak : Si comme nous l'avons supposé les variables explicatives sont strictement exogènes, les biais asymptotiques des estimateurs proviennent de $Ex_i \alpha_i \neq 0$. Pour tenir compte de cette corrélation, on peut utiliser la décomposition de la régression probabiliste $\alpha_i = x_i' \theta + \tilde{\alpha}_i$ où $Ex_i \tilde{\alpha}_i = 0$. On introduit un nouvel effet individuel non corrélé aux explicatives. Y. Mundlak propose donc de considérer le modèle de régression étendu :

$$y_{it} = x_{it}' \beta + x_{it}' \theta + \tilde{\alpha}_i + \varepsilon_{it}$$

On montre sans peine que l'estimateur des MCG de θ est proportionnel à $(\hat{\beta}_B - \hat{\beta}_W)$. Le test de significativité de θ revient au test d'Hausman à condition de calculer convenablement la variance de l'estimateur des MCQG puisque le modèle de Mundlak est bien un modèle à erreurs composées.

5 Les modèles dynamiques de panels

5.1 Introduction

Les problèmes d'estimation des modèles dynamiques de panels ont été observés par M. Nerlove et P. Balestra en 1967 dans une étude sur le gaz naturel aux USA modélisé par un modèle autoregressif à retards échelonnés :

$$y_{it} = \beta y_{it-1} + \gamma_1 x_{it} + \gamma_2 x_{it-1} + \alpha_i + \varepsilon_{it}$$

Ils ont constaté que tous les estimateurs classiques (MCO, MCQG,...) sont biaisés même lorsque la variable explicative est fortement exogène. Ils ont aussi butté sur la mise en œuvre du maximum de vraisemblance qui se réduit très souvent à la méthode des MCO !

¹ Voir l'article: J.A. Hausman, W.E. Taylor "Panel Data and Unobservable Individual Effects", *Econometrica*, vol. 49, n°6, nov. 1981.

5.2 Etude du modèle AR(1) de panel

Définition

$$\begin{cases} y_{it} = \beta y_{it-1} + \alpha_i + \varepsilon_{it}, |\beta| < 1 \\ \alpha_i \approx iid(0, \sigma_\alpha^2), \varepsilon_{it} \approx iid(0, \sigma_\varepsilon^2), \text{cov}(\alpha_i, \varepsilon_{js}) = 0, \forall i, \forall j, \forall s \\ \text{données initiales : } y_{i0}, i = 1, 2, \dots, N \end{cases}$$

Les estimateurs classiques et leurs limites

$$MCO : \hat{\beta}_{MCO} = \frac{\sum_i \sum_t y_{it} y_{it-1}}{\sum_i \sum_t y_{it-1}^2}, \lim_{N \rightarrow \infty} \hat{\beta}_{MCO} = \beta + \frac{\sigma_\alpha^2 (1 - \beta^2)}{\sigma_\alpha^2 (1 + \beta) + \sigma_\varepsilon^2 (1 - \beta)}$$

$$Within : \hat{\beta}_W = \frac{\sum_i \sum_t (y_{it} - y_{i.})(y_{it-1} - y_{i.(-1)})}{\sum_i \sum_t (y_{it-1} - y_{i.(-1)})^2},$$

$$l \lim_{N \rightarrow \infty} \hat{\beta}_W = \beta - \frac{1 + \beta}{T - 1} \left\{ 1 - \frac{1}{T} \frac{1 - \beta^T}{1 - \beta} \right\} \left\{ 1 - \frac{2\beta}{(1 - \beta)^{T-1}} \left[1 - \frac{1}{T} \frac{1 - \beta^T}{1 - \beta} \right]^{-1} \right\}^{-1} \approx \beta - \frac{1 + \beta}{T - 1}$$

$$Between : \hat{\beta}_B = \frac{\sum_i y_{i.} y_{i.(-1)}}{\sum_i y_{i.(-1)}^2},$$

$$\lim_{N \rightarrow \infty} \hat{\beta}_B = \beta + (1 - \beta) \frac{1 - \frac{1 - \beta^T}{1 + \beta} \frac{\theta}{T}}{1 - \frac{2\beta(1 - \beta^T)}{1 + \beta^2} \frac{\theta}{T}} \approx 1 - \frac{\theta}{T} (1 - \beta^T) \frac{1 - \beta}{1 + \beta}, \theta = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$$

On montre que ces estimateurs s'échelonnent de la façon suivante ; leurs limites en $N \rightarrow \infty$ aussi :

$$\hat{\beta}_W \leq \hat{\beta}_{MCOG} \leq \beta \leq \hat{\beta}_{MCG} \leq \hat{\beta}_{MCO} \leq \hat{\beta}_B$$

Si de plus $T \rightarrow \infty$:

$$\beta = \lim \hat{\beta}_W = \lim \hat{\beta}_{MCOG} = \lim \hat{\beta}_{MCG} < \lim \hat{\beta}_{MCO} < \lim \hat{\beta}_B = 1$$

Remarque : lorsque le modèle comporte des variables explicatives strictement exogènes les biais sont réduits mais ne s'annulent pas pour T fixe. L'ordre des estimateurs n'est d'ailleurs pas modifié.

Le maximum de vraisemblance

Lorsque les perturbations et les effets sont gaussiens, on peut utiliser la méthode du MV. Deux vraisemblances sont envisageables : la vraisemblance conditionnelle aux premières observations (*MV conditionnel*) et la vraisemblance qui intègre la loi des premières observations (*MV non conditionnel*).

MV conditionnel

Dans ce cas on montre, par des approximations asymptotiques, que dans le cadre du modèle AR(1) de panel si $\beta \geq \frac{(T-3)^2-8}{(T+1)^2}$ alors l'estimateur du MV est égal à l'estimateur des MCO. L'introduction de variables explicatives fortement exogènes réduit la surface des régions où MV=MCO.

MV non conditionnel

Il faut dans ce cas tenir compte des lois des valeurs initiales. Si on observe (y_{i0}, x_{i0}) dans le modèle :

$$y_{it} = \beta y_{it-1} + \gamma x_{it} + \delta z_i + \alpha_i + \varepsilon_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

on décompose y_{i0} sur les valeurs initiales des explicatives :

$$y_{i0} = \varphi z_i + \vartheta x_{i0} + v_{i0}$$

et on décompose toujours dans L^2 l'effet individuel sur la perturbation v_{i0} :
 $\alpha_i = \psi v_{i0} + \mu_i$.

Si $(v_{i0}, \mu_i, \varepsilon_{i1}, \dots, \varepsilon_{iT})$ sont des vecteurs gaussiens $Niid(0, \text{diag}(\sigma_v^2, \sigma_\mu^2, \sigma_\varepsilon^2 I_T))$, la log vraisemblance se calcule sans peine. La maximisation de cette vraisemblance non conditionnelle produit des estimateurs qui ne présentent pas les défauts de ceux issus du MV conditionnel.

Les méthodes des moments

Sans entrer dans les détails de cours plus approfondis sur les panels on donne ici des premières indication sur les méthodes des moments pour ces modèles dynamiques. On se place dans le cas où les séries $y_i^T = (y_{i1}, y_{i2}, \dots, y_{iT})$ sont iid. Pour ces vecteurs du second ordre on pose $E y_i^T y_i^T = \Omega$, matrice dont les éléments sont notés $\omega_{ts}, t = 1, 2, \dots, T, s = 1, 2, \dots, T$.

Si le modèle est :

$$\begin{cases} y_{it} = \beta y_{it-1} + \alpha_i + \varepsilon_{it} \\ E(\varepsilon_{it} / y_i^{t-1}) = 0, y_i^{t-1} = (y_{i1}, y_{i2}, \dots, y_{it-1}), t = 2, 3, \dots, T \\ E\alpha_i = \gamma, V\alpha_i = \sigma_\alpha^2, E\varepsilon_{it}^2 = \sigma_t^2 \end{cases}$$

on en déduit que $Ey_i^{t-2}(\Delta y_{it} - \beta \Delta y_{it-1}) = 0, t = 3, \dots, T$; on dispose ainsi de $\frac{(T-2)(T-1)}{2}$ conditions d'orthogonalité pour estimer le paramètre β . Ces conditions se traduisent sur les paramètres ω_{ts} par :

$$\omega_{ts} = \beta \omega_{t-1s} + c_s, t = 2, 3, \dots, T, s = 1, 2, \dots, t-1$$

avec $c_s = Ey_{is}\alpha_i, s = 1, 2, \dots, T$. On constate que pour ce modèle autorégressif on prend en charge que très peu d'hypothèses garantissant ainsi une bonne flexibilité au modèle. La matrice Ω dépend de $2T$ paramètres : $\alpha, c_1, c_2, \dots, c_{t-1}, \omega_{11}, \omega_{22}, \dots, \omega_{TT}$. Les paramètres ω s'estiment par leurs contreparties empiriques, les autres à partir de la méthode des moments ; par exemple si $T = 3$, le modèle est juste identifié, α, c_1, c_2 sont déterminés par les 3 conditions :

$$\begin{aligned} \omega_{32} &= \alpha \omega_{22} + c_2 \\ \omega_{31} &= \alpha \omega_{21} + c_1 \\ \omega_{21} &= \alpha \omega_{11} + c_1 \end{aligned}$$

en estimant ω_{ts} par $\hat{\omega}_{ts} = \frac{1}{N} \sum_i y_{it} y_{is}$. Dans ce cas particulier on a donc :

$$\hat{\alpha} = \frac{\sum_i (y_{i3} - y_{i2}) y_{i1}}{\sum_i (y_{i2} - y_{i1}) y_{i1}}. \text{ Dans les cas plus généraux c'est à la théorie de la méthode}$$

des moments généralisés ou des moindres carrés asymptotiques qu'il convient de recourir.

Conditions d'orthogonalité et moments généralisés

On peut approcher l'estimation des modèles dynamiques par diverses méthodes fondées sur diverses conditions d'orthogonalité. La littérature économétrique

retient aujourd'hui deux méthodes GMM dites DIF² et SYS³ qu'on peut présenter en guise d'introduction et d'illustration sur le modèle suivant :

$$y_{it} = \beta y_{it-1} + \delta x_{it} + \alpha_i + \varepsilon_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

On retient pour la variable explicative x les hypothèses suivantes :

$$\begin{aligned} E\alpha_i x_{it} &\neq 0, t = 1, 2, \dots, T \\ E\varepsilon_{it} x_{it} &\neq 0, t = 1, 2, \dots, T, r = t, \dots, T \\ E\varepsilon_{it} x_{is} &= 0, t = 1, 2, \dots, T, s = 1, 2, \dots, t-1 \end{aligned}$$

La variable explicative est donc une variable « prédéterminée, d'endogénéité contemporaine, corrélée aux effets individuels ». On en déduit un premier système classique de conditions d'orthogonalité par la suppression de l'effet individuel en différence :

$$\begin{aligned} Ey_{is} \Delta u_{it} &= 0, t = 2, 3, \dots, T, s = 0, 1, \dots, t-2 \\ Ex_{is} \Delta u_{it} &= 0, t = 2, 3, \dots, T, s = 1, 2, \dots, t-1 \\ \text{avec } u_{it} &= y_{it} - \beta y_{it-1} - \delta x_{it} \end{aligned}$$

La méthode des GMM appliquée à ce système porte le nom de *GMM DIF*.

Si on ajoute des conditions de stationnarité sur les variables on complète ce système par :

$$\begin{aligned} Eu_{it} \Delta y_{it-1} &= 0, t = 1, 2, \dots, T \\ Eu_{it} \Delta x_{it-1} &= 0, t = 2, \dots, T \end{aligned}$$

Ajoutées aux conditions précédentes, la méthode des GMM appliquée à ce « système » conduit à la méthode *GMM SYS*.

Remarque : Les études empiriques et quelques études théoriques ont montré que les méthodes de type GMM portent des défauts de biais et de manque de précision sur des échantillons finis, malgré leurs très bonnes propriétés asymptotiques. Des travaux sont publiés sur la question du comportement des GMM lorsqu'il y a abondance de conditions d'orthogonalité ce qui est le cas ici lorsque la dimension temporelle est grande. Ces travaux se rapportent aussi au problème des instruments faibles.

² Arellano, Bond, Journal of Econometrics, 1991

³ Arellano, Bover, Journal of Econometrics, 1995 ; Blundell, Bond, Journal of Econometrics, 1998.

Chapitre 12 : Modèles à variables latentes : modèles de variables qualitatives et dépendantes limitées

1. Exemples :

1.1 Possession d'un bien durable ou d'une caractéristique:

Accès à l'Internet :
$$\begin{cases} y_i = 1 \text{ si le ménage } i \text{ a accès} \\ y_i = 0 \text{ sinon} \end{cases}$$

$x_i, i = 1, 2, \dots, n$ désigne le vecteur des caractéristiques des ménages qui influence l'accès à l'Internet.

Modèle de choix :

$$\begin{cases} \text{Si accès : } U(1, x_i) \\ \text{Sinon : } U(0, x_i) \\ y_i = 1 \Leftrightarrow U(1, x_i) > U(0, x_i) \\ y_i = 0 \Leftrightarrow U(0, x_i) > U(1, x_i) \end{cases}$$

Variable latente :

$$y_i^* = U(1, x_i) - U(0, x_i) := x_i' b + u_i$$

Modèle dichotomique :

$$\begin{cases} y_i = 1 \Leftrightarrow y_i^* > 0 \\ y_i = 0 \Leftrightarrow y_i^* < 0 \end{cases}$$

Lien entre la variable latente et la variable observable ; processus d'observation :

$$y_i = 1[y_i^* > 0]$$

1.2 Généralisation : Modèles de choix discrets, modèles polytomiques :

Enquête loisir : « Si vous sortez dimanche après-midi, préférez-vous aller :

1. au théâtre
2. au cinéma
3. au musée »

. $y_i \in \{1,2,3\}$, trois niveaux d'utilité : $U(j, x_i)$, $j = 1,2,3$.

. variables latentes : $y_{ij}^* = U(j, x_i) := x_i' \beta_j + u_{ji}$, $j = 1,2,3, i = 1,2, \dots, n$

. processus d'observation :

$$y_i = \sum_{j=1}^3 j 1[y_{ji}^* = \max(y_{1i}^*, y_{2i}^*, y_{3i}^*)]$$

1.3 Modèle polytomique ordonné :

Exemple : collecte du salaire dans un système de tranches :

« Dans quelle tranche se situe votre salaire ? »

$0 \leq . < 1000$	$y = 1$
$1000 \leq . < 1500$	$y = 2$
$1500 \leq . < 2250$	$y = 3$
$2250 \leq . < 3000$	$y = 4$
$3000 \leq . < 5000$	$y = 5$
$5000 \leq .$	$y = 6$

Etude du salaire en fonction du niveau d'étude mesuré par le nombre d'années qualifiante, notée : x_i .

$$s_i = a + bx_i + u_i$$

s_i est une variable latente, seule la variable y_i est observable au travers du processus d'observation de recueil des informations sur le salaire par le système de tranches ci-dessus :

$$y_i = 1[0 \leq s_i < 1000] + 2[1000 \leq s_i < 1500] + 3[1500 \leq s_i < 2250] + 4[2250 \leq s_i < 3000] + 5[3000 \leq s_i < 5000] + 6[5000 \leq s_i]$$

On en déduit par exemple que :

$$P(y_i = 2/x_i) = P(1000 - a - bx_i \leq u_i < 1500 - a - bx_i) = F(1500 - a - bx_i) - F(1000 - a - bx_i)$$

où $F(\cdot)$ est la fonction de répartition de u_i .

1.4 Modèle de Poisson (modèle de comptage) :

La variable dépendante $Y_i \in \mathbb{N}$ correspond au nombre d'occurrences d'un événement (nombre d'accidents, de dépôts de brevet,...). On suppose dans le modèle de Poisson que

$$P(Y_i = y_i | x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \text{ avec } \lambda_i = e^{x_i'b}$$

Ce modèle est paramétrique et peut être estimé par l'EMV.

1.5 Modèles de variable dépendante limitée :

Exemple 1 : le salaire minimum.

. s_i salaire potentiel : $s_i = x_i'b + u_i$

. dispositif de salaire minimum : s_0 ; le salaire effectivement payé est

$$\begin{cases} s_i, s_i \geq s_0 \\ s_0, s_i < s_0 \end{cases}$$

. on pose $y_i^* = s_i - s_0 = x_i'b + u_i$

$$\begin{cases} y_i = y_i^*, y_i^* \geq 0 \\ y_i = 0, y_i^* < 0 \end{cases}$$

soit

$$y_i = y_i^* 1[y_i^* \geq 0]$$

y est la variable observable et y^* est la variable latente.

Exemple 2 : le modèle de déséquilibre.

Equation d'offre : $S_t = z_t' b_1 + u_{1t}$

Equation de demande : $D_t = x_t' b_1 + u_{2t}$

Quantités échangées : $Q_t = \min(S_t, D_t)$

- . Q_t est la variable observable.
- . D_t, S_t sont des variables latentes.

2. Formalisation : modèles latents, modèles observables.

a) Variables scalaires :

y_i^* variable latente (réelle)

y_i variable observable (réelle)

x_i vecteur des variables exogènes (dimension K)

Modèle latent : $y_i^* = x_i' b + u_i$, u_i perturbation.

Processus d'observation :

$$\exists g : R \rightarrow R$$

g fonction connue non injective telle que :

$$y_i = g(y_i^*)$$

Exemples :

. $g(z) = 1[z \geq 0]$: modèle dichotomique,

. $g(z) = z1[z \geq 0]$: modèle TOBIT,

. $g(z) = \sum_{j=0}^p c_j 1[a_j \leq z < a_{j+1}]$, $a_0 = -\infty < a_1 < \dots < a_p < a_{p+1} = +\infty$

b) Variables vectorielles :

y_i^* variable latente (dimension q)

y_i variable observable (dimension p)

Modèle latent :

$$y_i^* = \prod_{(qxK)} x_i + v_i$$

où v_i est le vecteur des perturbations.

Processus d'observation : $g : R^q \rightarrow R^p$.

- . $q = 2, p = 1 : g(z_1, z_2) = \min(z_1, z_2)$, déséquilibre
- . $q = 2, p = 1 : g(z_1, z_2) = z_1 1[z_2 \geq 0]$, TOBIT généralisé
- . $p = 1, g(z_1, z_2, \dots, z_q) = \sum_{j=1}^q j 1[z_j = \max(z_1, z_2, \dots, z_q)]$, modèle polytomique.

3. Modèles dichotomiques : PROBIT, LOGIT

3.1 Définitions :

. Modèle latent : $y_i^* = x_i' b + u_i, -u_i \approx F$, F fonction de répartition d'une loi connue d'espérance nulle.

. Modèle observable : $y_i = 1[y_i^* \geq 0]$

. Les observations sont indépendantes.

Conséquence : les variables y_i sont des variables de Bernoulli indépendantes :

$$y_i \approx B(1, p_i), p_i = P(y_i = 1) = P(x_i' b + u_i \geq 0) = F(x_i' b)$$

Cas particuliers :

i) $F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt := \Phi(z)$, modèle PROBIT

ii) $F(z) = \frac{1}{1 + \exp(-z)} := L(z)$, modèle LOGIT

3.2 Méthode d'estimation : le maximum de vraisemblance.

. $y_i \approx B(1, p_i)$, densité : $p_i^{y_i} (1 - p_i)^{(1-y_i)}$.

. densité de y_1, y_2, \dots, y_n : $\prod_{i=1}^n p_i^{y_i} (1 - p_i)^{(1-y_i)}$

. Log-vraisemblance : $L_n(b) = \sum_{i=1}^n \{y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\}$

$$L_n(b) = \sum_{i=1}^n \{y_i \log(F(x_i' b)) + (1 - y_i) \log(1 - F(x_i' b))\}$$

. Estimateur : $\hat{b}_n = \arg \max L_n(b)$

Remarques :

. si $\log(F(\cdot)), \log(1 - F(\cdot))$ sont des fonction strictement concaves, $L_n(b)$ est strictement concave.

. $\log(\Phi), \log(1 - \Phi), \log(L), \log(1 - L)$ sont des fonctions strictement concaves.

Equations de vraisemblance : Cas « linéaire simple » $x_i' b = \alpha + \beta z_i$:

$$\begin{cases} \frac{\partial L_n}{\partial \alpha} = \sum_{i=1}^n \frac{f(\alpha + \beta z_i)}{F(\alpha + \beta z_i)(1 - F(\alpha + \beta z_i))} (y_i - F(\alpha + \beta z_i)) = 0 \\ \frac{\partial L_n}{\partial \beta} = \sum_{i=1}^n \frac{f(\alpha + \beta z_i)}{F(\alpha + \beta z_i)(1 - F(\alpha + \beta z_i))} z_i (y_i - F(\alpha + \beta z_i)) = 0 \\ f(z) = \frac{\partial F}{\partial z}(z) \end{cases}$$

Les estimateurs $\hat{\alpha}, \hat{\beta}$ de α, β sont solutions de ce système non linéaire.

Remarque : Le logiciel SAS permet d'estimer très facilement les modèles LOGIT et PROBIT par les PROC LOGISTIC et PROBIT...

Cas favorable du Modèle LOGIT :

$$F(z) = L(z) = \frac{1}{1 + \exp(-z)}, f(z) = F(z)(1 - F(z))$$

Conséquence les équations de vraisemblance ont une forme analogue aux équations normales d'un modèle linéaire : si $\hat{\varepsilon}_i = y_i - F(\hat{\alpha} + \hat{\beta}z_i)$ est le résidu les équations de vraisemblances sont :

$$\begin{cases} \sum_{i=1}^n \hat{\varepsilon}_i = 0 \\ \sum_{i=1}^n z_i \hat{\varepsilon}_i = 0 \end{cases}$$

Remarque : Le modèle Logit est un modèle GLM (generalised linear model) au sens de Nelder et McCullagh.

Applications :

- . z_i variable indicatrice,
- . le scoring.

3.3 Variables explicatives endogènes

- i) Dans le cas d'un modèle apparemment linéaire dans lequel une variable explicative dichotomique est endogène :

$$y_i = b_1 x_{1i} + b_2 x_{2i} + u_i \text{ avec } x_{1i} \in \{0,1\} \text{ et } E(u_i x_{1i}) \neq 0$$

On peut appliquer la première étape des 2MC : cette première régression n'est pas structurelle. Cette méthode est conseillée pour produire l'estimateur des moments à partir de la contrainte identifiante $E(u_i z_i) = 0$.

On peut également utiliser un modèle dichotomique si l'on est prêt à faire une hypothèse structurelle sur la variable dichotomique endogène.

On construit alors la valeur prédite $F(z_i' \hat{c})$ (estimateur de $E(x_{1i} | z_i)$). On peut alors faire la régression augmentée

$$y_i = b_1 x_{1i} + b_2 x_{2i} + d \underbrace{(x_{1i} - F(z_i' \hat{c}))}_{\hat{v}_i} + u_i$$

Les hypothèses assurant la convergence sont plus fortes que dans le cas de 2MC car on doit spécifier $E(x_{1i} | z_i)$ et imposer $E(u_i | z_i) = 0$

ii) Modèle à variable dépendante dichotomique et variable explicative endogène

On suppose que dans le modèle latent $y_i^* = b_1 x_{1i} + b_2 x_{2i} + u_i$, on a $E(x_{1i} u_i) \neq 0$

Sous l'hypothèse $\begin{cases} u_i = \alpha z_i + v_i \\ u_i = \gamma v_i + \varepsilon_i \end{cases}$ où ε_i est indépendante de u_i et de v_i , la technique de la régression augmentée est valide.

L'estimation en deux étapes oblige à une correction des écarts-types des estimateurs.

4. Modèles Tobit

4.1 Modèle TOBIT simple

a) Définition

Modèle latent : $y_i^* = a + b x_i + u_i, u_i \approx \text{Niid}(0, \sigma^2), i = 1, 2, \dots, n$

Modèle observable : $\begin{cases} y_i = y_i^*, y_i^* \geq 0 \\ y_i = 0, y_i^* < 0 \end{cases}$

Processus d'observation : $y_i = y_i^* 1[y_i^* \geq 0]$

b) Inadaptation des MCO

. Intuition graphique

. Calcul :

Echantillon observable : variable non tronquée : $i / y_i > 0$

$$\begin{aligned} E(y_i / y_i > 0) &= E(a + b x_i + u_i / a + b x_i + u_i > 0) \\ &= a + b x_i + E(u_i / u_i > -a - b x_i) \end{aligned}$$

$u_i \approx N(0,1)$:

$$E(u_i / u_i > -a - bx_i) = \sigma \frac{\varphi(\frac{a + bx_i}{\sigma})}{\Phi(\frac{a + bx_i}{\sigma})}$$

avec $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$, $\Phi(t) = \int_{-\infty}^t \varphi(x)dx$.

On appelle Ratio de Mill la fonction : $\lambda(t) = \frac{\varphi(t)}{\Phi(t)}$.

Pour les observations $i / y_i > 0$ on a donc :

$$y_i = a + bx_i + \sigma z_i + \varepsilon_i$$

$$E\varepsilon_i = 0, z_i = \lambda\left(\frac{a + bx_i}{\sigma}\right)$$

Si on estime cette équation par les MCO en omettant la variable z_i on obtient des estimateurs de a, b qui portent un biais de variable omise.

Echantillon complet : $i = 1, 2, \dots, n$

$$E(y_i) = E(y_i / y_i > 0) = \Phi\left(\frac{a + bx_i}{\sigma}\right)(a + bx_i) + \sigma\varphi\left(\frac{a + bx_i}{\sigma}\right)$$

Le modèle est non linéaire, l'estimateur des MCO n'est pas convergent.

c) Méthodes d'estimation

i) Méthode d'Heckman

Modèle PROBIT associé au modèle TOBIT :

On pose : $d_i = 1[y_i > 0]$, $d_i, i = 1, 2, \dots, n$ est observable.

Puisque $P(d_i = 1) = \Phi\left(\frac{a + bx_i}{\sigma}\right)$; on peut estimer les paramètres $a^* = \frac{a}{\sigma}, b^* = \frac{b}{\sigma}$ par le maximum de vraisemblance. Soit \hat{a}^*, \hat{b}^* ces estimateurs. On peut ainsi estimer le ratio de Mill z_i par $\hat{z}_i = \lambda(\hat{a}^* + \hat{b}^* x_i)$.

Pour estimer les paramètres a, b, σ , Heckman propose d'appliquer les moindres carrés ordinaires (MCO) à la régression étendue sur l'échantillon observable $i / y_i > 0$:

$$y_i = a + bx_i + \sigma \hat{z}_i + \varepsilon_i^*$$

Remarques : C'est une méthode simple qui ne garantit pas cependant que $\hat{\sigma}$ soit toujours positif. De plus, il ne faut pas utiliser les variances estimées issues des MCO directement.

ii) Méthode du maximum de vraisemblance :

La loi de y_i a une masse en 0 :

$$P(y_i = 0) = P(y_i^* \leq 0) = P(u_i \leq -a - bx_i) = 1 - \Phi\left(\frac{a + bx_i}{\sigma}\right)$$

La densité de y_i se décompose en :

$$y_i > 0 : \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - a - bx_i)^2\right)$$

$$y_i = 0 : 1 - \Phi\left(\frac{a + bx_i}{\sigma}\right)$$

Avec l'hypothèse d'indépendance des observations la vraisemblance est :

$$L_n(a, b, \sigma) = \prod_{i/y_i=0} \left[1 - \Phi\left(\frac{a + bx_i}{\sigma}\right)\right] \prod_{i/y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - a - bx_i)^2\right)$$

Les estimateurs de a, b, σ s'obtiennent en maximisant cette vraisemblance ou son logarithme.

Remarques : Par une reparamétrisation adaptée on obtient un fonction objectif

concave : si on pose $a^* = \frac{a}{\sigma}, b^* = \frac{b}{\sigma}, h = \frac{1}{\sigma}$ et $L_n^*(a^*, b^*, h) = L_n\left(\frac{a^*}{h}, \frac{b^*}{h}, \frac{1}{h}\right)$, cette

fonction est strictement concave et garantit l'unicité du maximum de vraisemblance. On peut utiliser le logiciel SAS pour estimer un modèle TOBIT simple par la PROC LIFEREG.

4.2 Le modèle TOBIT généralisé (modèle de sélection)

a) Définition

Modèle latent :

$$\begin{cases} y_{1i}^* = x_{1i}'b_1 + u_{1i}, \begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \approx Niid(0, \Sigma) \\ y_{2i}^* = x_{2i}'b_2 + u_{2i}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix}, i = 1, 2, \dots, n \end{cases}$$

Processus d'observation, modèle observable : $y_i = y_{1i}^* 1[y_{2i}^* \geq 0]$.

b) Inadaptation des MCO :

Les variables y_{1i}^* sont observables lorsque $y_{2i}^* \geq 0$, sinon l'observation est égale à 0. Sur l'échantillon des observations pour lesquelles y_{1i}^* est observable on a :

$$\begin{aligned} E(y_i / y_i \neq 0) &= E(x_{1i}'b_1 + u_{1i} / x_{2i}'b_2 + u_{2i} \geq 0) \\ &= x_{1i}'b_1 + E(u_{1i} / u_{2i} \geq -x_{2i}'b_2) \\ &= x_{1i}'b_1 + \rho\sigma_1\lambda(x_{2i}'b_2) \end{aligned}$$

Remarque : lorsque $\rho = 0$ la « sélection des observation est ignorable ». C'est-à-dire que la censure doit être indépendante de la variable d'intérêt pour que l'estimateur des MCO sur l'échnatillon des observations non censurées soit convergent.

c) Méthodes d'estimation

1) Méthode d'Heckman

On considère le modèle Probit associé : $d_i = 1$ si y_{1i}^* est observable, $d_i = 0$ sinon:

$$P(d_i = 1) = P(y_{2i}^* \geq 0) = \Phi(x_{2i}'b_2)$$

Les variables d_i sont régies par un modèle PROBIT qui peut être estimé par le maximum de vraisemblance : \hat{b}_2 .

Cette estimateur permet de calculer une estimation du ratio de Mill :
 $\hat{z}_i = \lambda(x_{2i}'\hat{b}_2)$. Sur l'échantillon des observables on construit comme dans le cas du modèle TOBIT simple une régression augmentées :

$$y_i = x_{1i}'b_1 + \rho\sigma_1\hat{z}_i + \varepsilon_i^*$$

La méthode des MCO sur cette dernière conduit à des estimateurs convergents de $b_1, \sigma_{12} = \rho\sigma_1$. Mais comme précédemment il convient de ne pas utiliser directement les estimateurs des MCO des variances de ces estimateurs.

2) Maximum de vraisemblance

Comme pour le modèle TOBIT simple, le modèle TOBIT généralisé comporte une masse de probabilité en 0.

Si $y_i = 0, P(y_i = 0) = P(y_{2i}^* < 0) = 1 - \Phi(x_{2i}'b_2) := l_{1i}(b_2)$

Si $y_i \neq 0$, la densité est $\frac{1}{\sigma_1} \varphi\left(\frac{y_{1i} - x_{1i}'b_1}{\sigma_1}\right) \Phi\left(\frac{x_{2i}'b_2 + \rho\left(\frac{y_{1i} - x_{1i}'b_1}{\sigma_1}\right)}{\sqrt{1-\rho^2}}\right) := l_{2i}(b_1, b_2, \rho, \sigma_1)$

La vraisemblance de la totalité des observations $y_i, i = 1, 2, \dots, n$ est alors :

$$L_n(b_1, b_2, \rho, \sigma_1) = \prod_{i/y_i=0} l_{1i}(b_1) \prod_{i/y_i \neq 0} l_{2i}(b_1, b_2, \rho, \sigma_1)$$

Remarque : A ρ fixé cette vraisemblance est strictement concave en $\frac{b_1}{\sigma_1}, b_2, \frac{1}{\sigma_1}$.

5.1. Modèle polytomique ordonné : EMV

On note α_j les seuils des tranches. On a :

$$P(Y_i = j | X_i = x_i) = F\left(\frac{\alpha_{j+1} - bx_i}{\sigma}\right) - F\left(\frac{\alpha_j - bx_i}{\sigma}\right)$$

On en déduit la log-vraisemblance :

$$L_n(a, b, \sigma) = \sum_{i=1}^N \ln \left[F\left(\frac{\alpha_{y_i+1} - bx_i}{\sigma}\right) - F\left(\frac{\alpha_{y_i} - bx_i}{\sigma}\right) \right]$$

En supposant F connu, on peut utiliser l'estimateur du maximum de vraisemblance.

Lorsque les seuils sont connus (avec au moins trois tranches), on identifie b et σ . Lorsque les seuils sont inconnus, ce n'est plus le cas. Dans ce dernier cas, on peut fixer $\sigma = 1$ et la constante du modèle linéaire latent à 0.

5.2. Modèle polytomique non ordonné : EMV

On se place dans le cadre du modèle de choix discret présenté en début de chapitre. On suppose les perturbations indépendantes des variables explicatives.

On utilise également l'estimateur du maximum de vraisemblance.

On a : $P(Y_i = j | X_i = x_i) = \int \prod_{k \neq j} F[-x_i'(b_k - b_j) + u] f(u) du$

Dans le cas du modèle logit multinomial, on suppose : $F(u) = \exp(-\exp(-u))$

Alors on obtient :

$$P(Y_i = j | X_i = x_i) = \frac{e^{b_j x_i}}{\sum_{k=1}^m e^{b_k x_i}}$$

On identifie les coefficients b à une constante près.

Pour interpréter les paramètres, on peut utiliser les ratios de chances (odds ratio) :

$$\frac{P(Y_i = j | X_i = x_i)}{P(Y_i = k | X_i = x_i)} = \frac{e^{b_j x_i}}{e^{b_k x_i}} = \exp((b_j - b_k) x_i)$$