

# Modèles Estimés sur Données de Panel

## Introduction

Il est fréquent en économétrie qu'on ait à composer avec des données à deux dimensions :

- une dimension chronologique
- une dimension spatiale

Par exemple, on s'intéresse à la relation qui existe entre le ROA d'une banque et un certain nombre de ratios représentatifs soit de la structure de son bilan, soit de son activité au cours de l'exercice écoulé. On a rassemblé à cet effet un certain nombre d'informations relatives aux valeurs prises par ces ratios pour un échantillon de 24 banques européennes (UK, France, Allemagne) sur la période 1995 – 2000 (fichier ROA.xls).

Cet échantillon est composé de 24 banques : c'est la dimension « spatiale » de l'échantillon. Ces banques étant de tailles et de nationalités différentes on comprend que la question de l'homogénéité / hétérogénéité des données se pose avec acuité.

Pour chaque banque on dispose d'observations dans le temps (de 1995 à 2000) : c'est la dimension chronologique.

Quel intérêt peut-il y avoir à traiter de l'estimation d'une relation avec ce genre de données ? Ce qui est intéressant avec les données de panel, c'est qu'elles vont nous permettre de répondre à certaines questions qu'il eut été impossible d'élucider avec un simple échantillon chronologique ou un simple échantillon en « cross section ». Je m'explique :

1. Imaginons que la relation Ratios de structure  $\rightarrow$  ROA ait été estimée sur un échantillon de banques en coupe instantanée (par exemple, les observations relatives à la seule année 1995) et qu'on ait pu mettre en évidence le caractère effectif de cette relation. La question qu'on est en droit de se poser, à l'issue de cet ajustement est : « qu'est-ce

qui nous garantit que cette relation, effective pour l'année 1995, le sera aussi pour les années ultérieures ? ». Nous n'avons, à ce stade là, aucun élément qui nous permette de nous prononcer sur la validité de cette relation dans le temps.

2. De façon un peu différente, supposons à présent que cette même relation ait été estimée sur un échantillon de nature chronologique... mais pour une seule banque. Qu'est-ce qui nous autorise à extrapoler cette relation, propre à la banque A, à une autre banque B ? En d'autres termes : la banque A étudiée ne présente-t-elle pas des spécificités telles qu'il devient difficile de généraliser les résultats.

Seul un échantillon en données de panel est susceptible de nous aider à nous prononcer quant à la stabilité de la relation dans sa double dimension spatiale et temporelle.

**Illustration** : *FATALITY.WF1, FATALITY-1.DOC*

L'exemple traité ici est emprunté à J.H. STOCK et M.W.WATSON<sup>1</sup>. Il est destiné à illustrer justement tout l'intérêt qu'on peut trouver à exploiter un échantillon en données de panel.

La problématique est la suivante : tous les ans, aux USA on dénombre approximativement 40000 décès par accidents de la route. Dans 1/3 des cas ces accidents impliquent un conducteur en état d'ébriété. Une enquête a même montré qu'entre 1 h et 3 h du matin, un quart des automobilistes dépasse le taux autorisé d'alcoolémie. Des mesures de lutte contre l'alcool au volant ont été adoptées ici et là dans différents états des USA. C'est sur l'efficacité de ces mesures au regard du nombre des décès par accident de la route qu'on s'interroge ici.

L'échantillon : pour 48 états des USA on dispose d'observations annuelles sur la période 1982 – 1988 pour ce qui concerne un certain nombre de variables et, notamment :

- le taux de mortalité par accident de la route  $MRALL_{it}$  (nombre de décès par accident de la route pour 10000 habitants dans l'état  $i$  au cours de

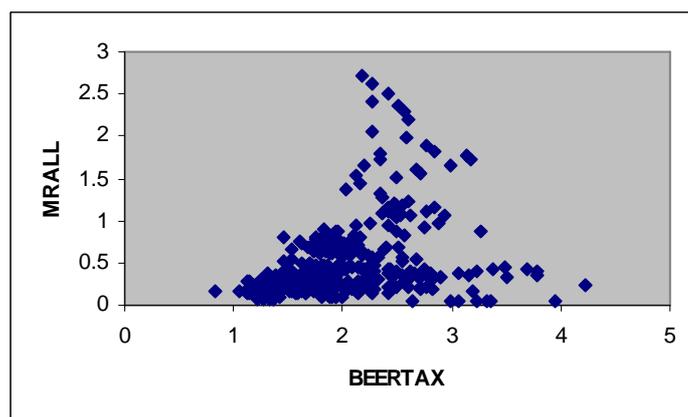
---

<sup>1</sup> J.H. STOCK et M.W.WATSON : "Introduction to Econometrics" – Addison Wesley 2003.

l'année t)

- le taux de taxe sur les alcools ajusté de l'inflation  $BEERTAX_{it}$

On a reproduit ci-dessous la relation Taxe sur les Alcools → Taux de Mortalité en utilisant, pour commencer, l'ensemble de toutes les observations dans le temps et dans l'espace, soit, au total,  $48 \times 7 = 336$  couples observés :



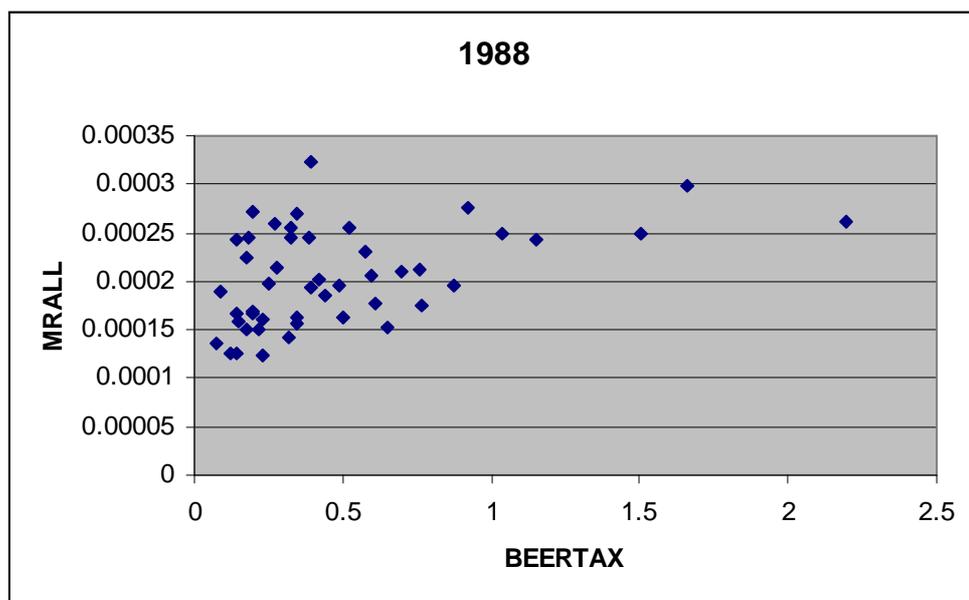
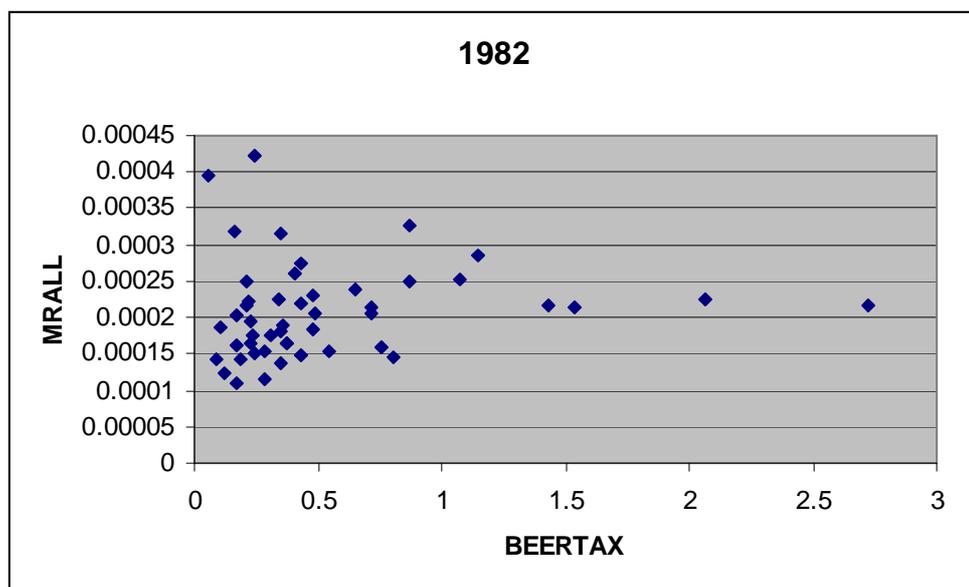
Ce graphique suggère, paradoxalement, une relation positive (!) entre la taxe et le taux de mortalité, ce que confirme l'ajustement réalisé sur données empilées :

Dependent Variable: MRALL\_?\*10000  
 Method: Pooled Least Squares  
 Date: 14/05/04 Time: 09:04  
 Sample: 1982 1988  
 Included observations: 7  
 Cross-sections included: 48  
 Total pool (balanced) observations: 336

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.853308	0.043567	42.53913	0.0000
BEERTAX_?	0.364605	0.062170	5.864668	0.0000
R-squared	0.093363	Mean dependent var		2.040444
Adjusted R-squared	0.090648	S.D. dependent var		0.570194
S.E. of regression	0.543736	Akaike info criterion		1.625230
Sum squared resid	98.74685	Schwarz criterion		1.647951
Log likelihood	-271.0387	F-statistic		34.39433
Durbin-Watson stat	0.132716	Prob(F-statistic)		0.000000

On retrouve ce même résultat paradoxal quand on se limite à un simple échantillon en coupe instantanée pour l'année 1982 (on a, dans ce cas, 48 observations qui

correspondent, chacune, à un des états) ou encore pour l'année 1988 :



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.859073	0.105989	17.54030	0.0000
BEERTAX_?	0.438755	0.164454	2.667950	0.0105
R-squared	0.134003	Mean dependent var	2.069594	

#### Ajustement 1988

Sur la base de ces résultats peut-on conclure à une relation positive ou, pour le moins,

à l'absence de relation entre le taux de taxe et le taux de mortalité ? On peut nourrir quelque soupçon quant au sens de cette relation... A ce stade de nos investigations on doit se souvenir que la relation, telle qu'elle est ici spécifiée (le taux de mortalité ne dépendrait **que** du taux de taxe) occulte un certain nombre de variables qui exercent, vraisemblablement, une certaine influence sur le taux de mortalité :

- l'état des routes
- l'état du parc automobile
- les habitudes socio-culturelles par rapport à la consommation d'alcool
- la densité urbaine de l'état considéré

autant de variables qui sont susceptibles d'être corrélées à la fois avec la variable expliquée et avec la variable explicative et dont l'omission induit nécessairement un **biais** sur la valeur estimée du coefficient attaché à la taxe sur les alcools. Certaines de ces variables dont on sait bien qu'elles exercent une influence sur le taux de mortalité sont particulièrement difficiles à mesurer. Comment mesurer en particulier le niveau de tolérance culturelle dans un état de l'union par rapport à la consommation d'alcool ? Faut-il pour autant renoncer à la prise en compte de ces variables en acceptant du même coup les conséquences de leur omission ? Pas nécessairement si certaines conditions sont vérifiées.

Imaginons que les valeurs prises par ces variables inobservables ou difficilement mesurables puissent être considérées comme constantes à travers le temps pour un individu (état) donné. Tel est par exemple le cas des variables socio-culturelles dont les caractéristiques ne se déforment que très lentement dans le temps. Le modèle peut alors être ainsi spécifié :

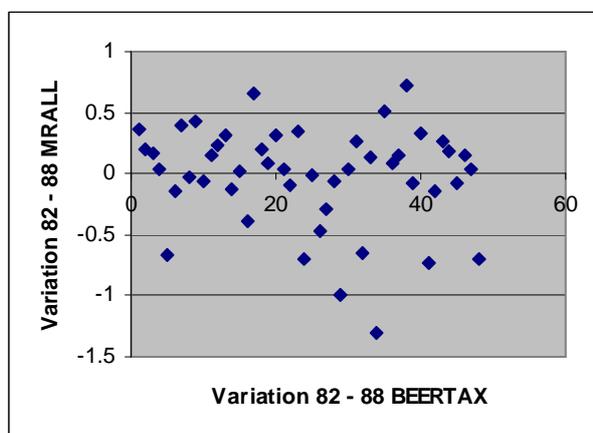
$$MRALL_{it} = \alpha + \beta BEERTAX_{it} + \gamma Z_i + \varepsilon_{it}$$

En exprimant les variables en différences premières on voit qu'il est possible de se débarrasser des variables inobservables et d'obtenir ainsi une estimation sans biais du coefficient attaché à la variable explicative :

$$\Delta MRALL_{it} = \beta \Delta BEERTAX_{it} + \varepsilon_{it} - \varepsilon_{i,t-1}$$

Le graphique et le tableau ci-dessous reproduisent la relation qui unit la variation de la taxe sur les alcools entre 1982 et 1988 à la variation du taux de mortalité sur cette

même période. On voit par conséquent que, lorsqu'on prend la précaution, par ce truchement, de composer avec la présence de spécificités individuelles, la relation est conforme à ce que dictait l'intuition : il y a bien un lien négatif entre l'importance de la taxe et le taux de mortalité.



	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	-0.0720371	0.06064401	-1.18786832	0.24098262
Variable X 1	-1.04097257	0.41722785	-2.49497383	0.01624848

Avec cet exemple, on perçoit déjà un peu les bénéfices qu'on peut retirer de l'utilisation de données de panel :

- l'estimation de la relation  $X - Y$  sur données chronologiques pour un individu n'est pas toujours possible. Tel est ici le cas : on ne dispose que de 7 observations dans le temps, ce qui est évidemment trop peu pour qu'on puisse estimer efficacement le coefficient  $\beta$ . Quand bien même disposerait-on d'un nombre d'observations dans le temps suffisamment grand se poserait ensuite la question du choix de l'individu "représentatif" et du caractère généralisable des résultats obtenus pour cet individu
- l'estimation en coupe instantanée (envisageable ici puisque le nombre  $N$  des individus est suffisamment grand) ne permet pas de composer avec la présence de spécificités individuelles... ce qui induit, dans ce contexte, un biais sur les coefficients estimés

Conscients de tout l'intérêt qu'il peut y avoir à travailler avec des données de panel, on

étudie à présent les techniques qui peuvent être mises en oeuvre pour exploiter le plus efficacement possible toute l'information contenue dans ce type d'échantillon.

## **Section I : structure générale et typologie des modèles**

### **Paragraphe 1 : structure générale du modèle**

Quand on travaille sur des données de panel, la structure générale du modèle peut être exprimée, pour l'individu  $i$  et pour la date  $t$ , sous la forme :

$$Y_{it} = X_{it} \beta + Z_i \alpha + \varepsilon_{it}$$

où  $Y_{it}$  est l'observation relative au  $i^e$  individu à la date  $t$ .

$\varepsilon_{it}$  est l'erreur du modèle relative à l'individu  $i$  et à la date  $t$ . On fait sur ce terme d'erreur un certain nombre d'hypothèses sur le bien fondé desquelles on pourra par la suite s'interroger. En attendant, on suppose que :

$$E \varepsilon_{it} = E \varepsilon_{it} | X = 0 \quad \forall i, t$$

$$E \varepsilon_{it}^2 = E \varepsilon_{it}^2 | X = \sigma_\varepsilon^2 \quad \forall i, t$$

Le modèle ci-dessus peut être récrit sous forme matricielle relative à l'individu  $i$  :

$$Y_i = X_i \beta + Z_i \alpha + \varepsilon_i$$

$X_i$  est une matrice  $T \times K$  (où  $K$  est le nombre de variables explicatives)<sup>2</sup>

$$X_i = \begin{bmatrix} X_{1i1} & X_{2i1} & X_{Ki1} \\ X_{1i2} & X_{2i2} & X_{Ki2} \\ \vdots & \vdots & \vdots \\ X_{1iT} & X_{2iT} & X_{KiT} \end{bmatrix}$$

On notera  $X_{k,it}$  la réalisation de la  $k^e$  variable pour l'individu  $i$  et la date  $t$ .

$\beta$  est, lui, un vecteur colonne (vecteur des coefficients de pente).

Dans tout ce qui suit, on supposera que le jeu des coefficients  $\beta_k$  est intangible dans le temps et dans l'espace quoique cette hypothèse puisse être levée... au prix d'un degré de complexité bien plus élevé.

---

<sup>2</sup> Attention : les réalisations de  $X_{1i}$  ne sont pas égales à 1.

L'élément  $Z_i \alpha$  est incontestablement le plus intéressant de ce modèle en cela qu'il va permettre à l'utilisateur, en fonction de sa composition, de spécifier l'homogénéité ou l'hétérogénéité de la relation dans sa dimension spatiale.  $Z_i$  est une matrice dont la première colonne  $Z_{1i}$  est une constante (supposée égale à 1) qui sera la même d'un individu à l'autre et quelle que soit la date considérée. Les autres colonnes, s'il y en a, sont les réalisations d'un certain nombre de variables,

1. susceptibles de différer d'un individu à l'autre
2. mais réputées constantes à travers le temps pour un même individu.

Le modèle ci-dessus peut donc être réécrit sous une forme plus développée :

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix} = \begin{bmatrix} X_{1i1} & X_{2i1} & \cdots & X_{Ki1} \\ X_{1i2} & X_{2i2} & \cdots & X_{Ki2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1iT} & X_{2iT} & \cdots & X_{KiT} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} 1 & Z_{2i} & \cdots & Z_{Hi} \\ 1 & Z_{2i} & \cdots & Z_{Hi} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_{2i} & \cdots & Z_{Hi} \end{bmatrix} \times \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_H \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{iT} \end{bmatrix}$$

Dans le cas de l'explication du ROA de la banque  $i$  à la date  $t$  on retrouve au rang des composantes de  $Z_i$  :

- . une variable indicatrice de l'activité (spécialisée ou universelle) de cette banque
- . une variable indicatrice de sa nationalité
- . une variable indicatrice de la cotation de cette banque et c...

## Paragraphe 2 : typologie des modèles

On peut imaginer trois grands types de modèles :

- . le modèle des MCO
- . le modèle à effet fixes connu aussi sous le nom de LSDV (Least Squares Dummy Variables)
- . le modèle à effet aléatoire connu aussi sous le nom de modèle à composantes d'erreur

- . le modèle à coefficients aléatoires
- . le modèle à structure de covariance

### **A - Le modèle MCO**

C'est le plus simple. Il se fonde sur le postulat que les individus qui composent l'échantillon sont rigoureusement homogènes c'est à dire ne se démarquent les uns des autres par aucune caractéristique spécifique. Dès lors, il n'y a qu'une composante au vecteur  $Z_i$  : la constante, commune à tous les individus. Le modèle est ainsi spécifié :

$$Y_{it} = \alpha + X_{it} \beta + \varepsilon_{it} = \alpha + \beta_1 X_{1,it} + \dots + \beta_K X_{K,it} + \varepsilon_{it}$$

où les coefficients sont estimés sur la base d'un échantillon à l'intérieur duquel les données sont « empilées » sans égard par rapport aux individus non plus que par rapport aux dates.

En termes matriciels et pour l'individu  $i$  on a encore :

$$Y_i = X_i \beta + Z_i \alpha + \varepsilon_i$$

mais avec une matrice  $Z_i$  à une seule colonne dont tous les éléments sont égaux à 1 de telle sorte que la constante  $\alpha$  est la même pour tous les individus.

A la condition que le postulat d'homogénéité soit fondé d'une part et que, d'autre part, les propriétés relatives à  $\varepsilon_{it}$  soient vérifiées (en particulier l'absence d'autocorrélation) l'estimateur MCO de  $\beta$  est sans biais, convergent et de variance minimum.

### **B - Le modèle à effets fixes (LSDV)**

Imaginons que chaque individu présente des caractéristiques propres susceptibles d'affecter la relation étudiée. Dans ce contexte d'hétérogénéité des individus une spécification MCO sur données « empilées » qui postule une même structure  $X \rightarrow Y$  quel que soit l'individu étudié induit un biais d'omission : l'estimateur MCO des  $\beta_k$  est biaisé et non convergent (cf. cours Econométrie III).

Deux cas de figure peuvent être envisagés :

- . toutes les caractéristiques spécifiques sont observables et quantifiables

. certaines ne le sont pas... quoiqu'on sache qu'elles existent

### 1°) Toutes les caractéristiques spécifiques sont observables et quantifiables

En admettant que les spécificités individuelles puissent être mesurées, de façon exhaustive, à l'aune des réalisations de deux variables  $Z_2$  et  $Z_3$  (variables dans l'espace mais constantes dans le temps pour un même individu)<sup>3</sup>, le modèle peut alors être réécrit sous la forme :

$$Y_i = X_i \beta + Z_i \alpha + \varepsilon_{it}$$

$$\text{avec : } Z_i = [i_T, Z_{2i}, Z_{3i}] = \begin{bmatrix} 1 & Z_{2i} & Z_{3i} \\ 1 & Z_{2i} & Z_{3i} \\ \vdots & \vdots & \vdots \\ 1 & Z_{2i} & Z_{3i} \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

où  $i_T$  est un vecteur de dimension T dont toutes les composantes sont égales à 1.

Sous forme matricielle, on peut récrire ce modèle comme :

$$Y = X \beta + Z \alpha + \varepsilon$$

$$Y = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ \vdots \\ Y_{N1} \\ \vdots \\ Y_{NT} \end{bmatrix} \quad X = \begin{bmatrix} X_{1,11} & \cdots & X_{K,11} \\ \vdots & & \vdots \\ X_{1,1T} & \cdots & X_{K,1T} \\ \vdots & & \vdots \\ X_{1,N1} & \cdots & X_{K,N1} \\ \vdots & & \vdots \\ X_{1,NT} & \cdots & X_{K,NT} \end{bmatrix} \quad Z = \begin{bmatrix} 1 & Z_{2,1} & Z_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & Z_{2,1} & Z_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & Z_{2,N} & Z_{3,N} \\ \vdots & \vdots & \vdots \\ 1 & Z_{2,N} & Z_{3,N} \end{bmatrix}$$

Tel qu'il est spécifié, ce modèle peut, en principe, être estimé par les MCO. Il est peu vraisemblable cependant que les caractéristiques qui déterminent les spécificités individuelles puissent être recensées et mesurées de manière exhaustive et, dans ce

---

<sup>3</sup> Par exemple, dans le cas de l'explication du ROA de la banque  $i$  à la date  $t$  on retrouve au rang des composantes de  $Z_i$  :

- . une variable indicatrice de l'activité (spécialisée ou universelle) de cette banque
- . une variable indicatrice de sa nationalité
- . une variable indicatrice de la cotation de cette banque et c...

cas, le risque d'un biais d'omission demeure... sauf à envisager une autre spécification du modèle.

## 2°) Certaines des caractéristiques ne sont pas observables.

Plus simplement, un moyen commode de composer avec la présence de spécificités non observables consiste à introduire dans la matrice  $Z$  autant de variables dummy qu'il y a d'individus. Par exemple, dans le cas  $N = 2$  et  $T = 3$  on a :

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Les coefficients  $\alpha_1$ ,  $\alpha_2$  représentent alors une mesure synthétique de l'ensemble de toutes les caractéristiques spécifiques –observables ou non observables- susceptibles d'influencer la relation étudiée. Le risque de biais d'omission est ainsi considérablement réduit... sous réserve que l'hypothèse de constance dans le temps des spécificités individuelles ait un réel fondement. Bien sûr, comme il s'agit désormais d'une mesure synthétique, il devient *ipso facto* impossible d'isoler au sein de l'effet spécifique ce qui peut être attribué à telle ou telle caractéristique en particulier.

On verra plus loin que l'estimation de ce modèle LSDV peut poser quelques problèmes de faisabilité (liées à la puissance des moyens de calcul) lorsque le nombre d'individus (et donc de variables dummy) est important. Dans ce contexte on verra que la mise en oeuvre de l'opérateur *within* peut être une alternative intéressante pour l'estimation des coefficients du modèle.

### C - Le modèle à effets aléatoires (modèle à composantes d'erreur)

Une autre manière d'aborder la question de l'hétérogénéité des individus à l'intérieur d'un échantillon en données de panel consiste à interpréter le terme d'erreur comme étant la somme de deux composantes (d'où la terminologie utilisée de modèle à composantes d'erreur) :

- une première composante  $\varepsilon_{it}$  similaire à celle qui apparaissait déjà dans les modèles précités

- une seconde, plus originale, postule que chaque individu se démarque des autres par la réalisation d'une variable aléatoire  $u_i$  dont les caractéristiques (en particulier, moyenne et variance) sont identiques d'un individu à l'autre.

Ce type de modèle est ainsi spécifié :

$$Y_{it} = \alpha + X_{it} \beta + u_i + \varepsilon_{it}$$

avec  $u_i \sim \text{IID}(0, \sigma_u)$ .

Attention, contrairement à ce qui se passe dans le cadre du modèle à effet fixe pour lequel les individus se démarquent les uns des autres par un élément constant, la composante  $u_i$  qui apparaît ici n'est pas une constante mais bien la réalisation d'une variable aléatoire.

Bien sûr, comme on sait que la présence d'une corrélation entre terme d'erreur et variables explicatives engendre des problèmes de biais dans l'estimation des coefficients du modèle, l'hypothèse sous-jacente à l'usage d'un modèle à composante d'erreur est que la composante aléatoire spécifique  $u_i$  n'est pas corrélée avec les variables explicatives du modèle. Cette hypothèse sous-jacente peut être suspecte dans certaines circonstances. Imaginons par exemple qu'il s'agisse d'expliquer le salaire d'un individu. Au rang des variables explicatives on retrouvera sans surprise des variables représentatives de l'ancienneté dans l'entreprise mais aussi, et de manière plus intéressante, le nombre d'années d'études ou encore le niveau de diplôme. L'élément spécifique aléatoire est réputé rendre compte des influences exercées par toutes les variables omises ou non observables... au rang desquelles figurent vraisemblablement les qualités innées de l'individu, dont on peut penser qu'elles ne sont sans doute pas totalement indépendantes du nombre d'années d'études ou du niveau de diplôme. Il est alors difficile de soutenir que cette composante aléatoire est sans corrélation avec les variables explicatives retenues.

#### **D - Le modèle à coefficients aléatoires.**

A vrai dire, le modèle à composante d'erreur se démarque du modèle MCO rapidement présenté au A en cela que ce qui est nécessairement constant dans le modèle MCO :

$$Y_{it} = \alpha + X_{it} \beta + \varepsilon_{it}$$

devient aléatoire dans le modèle à composante d'erreur :

$$Y_{it} = (\alpha + u_i) + X_{it} \beta + \varepsilon_{it}$$

On peut généraliser ce traitement réservé jusqu'ici à l'élément constant  $\alpha$  à l'ensemble de tous les coefficients du modèle qui devient alors :

$$Y_{it} = (\alpha + u_i) + X_{it} (\beta + h_i) + \varepsilon_{it}$$

L'hypothèse qui est faite dans ce cas est que les valeurs des coefficients peuvent différer aléatoirement d'un individu à l'autre quoiqu'en espérance ces coefficients soient identiques.

Les modèles hiérarchiques complètent cette panoplie de modèles à coefficients aléatoires en ménageant la possibilité de coefficients aléatoires dont les espérances pourraient varier d'un individu à l'autre. Le modèle est alors ainsi spécifié :

$$Y_{it} = (\alpha + u_i) + X_{it} (\beta + h_i + \mu z_i) + \varepsilon_{it}$$

où  $z_i$  est un vecteur de variables dont les réalisations sont spécifiques à l'individu  $i$  (p. ex. le niveau de diplôme pour reprendre l'exemple précédent) et  $\mu$  un vecteur de coefficients. On voit que, dans ce cas, le coefficient attaché à chacune des variables représentées dans la matrice  $X$  est susceptible de varier d'un individu à l'autre à la fois en niveau et en espérance.

### **E - Le modèle à structure de covariance**

Une autre façon (assez peu exploitée jusqu'ici) de modéliser les spécificités individuelle consiste à introduire l'élément d'hétérogénéité non pas au niveau des coefficients du modèle mais plutôt à celui de la matrice des variances covariances du terme d'erreur  $\varepsilon_{it}$  en posant que le terme d'erreur présente une propriété d'hétéroscédasticité dans la dimension des individus :

$$\text{Var } \varepsilon_{it} = \sigma_{\varepsilon_i}^2$$

Exemple : prendre le fichier « AIRLINE\_COST\_FUNCTION.WF1 » (source Greene). Sous Eviews 5, il sera bon de structurer le fichier en fonction des individus

et des dates. On s'intéresse à la modélisation de la fonction de coût à laquelle serait confrontée une compagnie aérienne. A cet effet, on a collecté un certain nombre d'informations relatives à 6 compagnies aériennes sur une période de 15 ans pour ce qui concerne :

- le coût total de production C apprécié en milliers de US \$
- le niveau de la production Q (nombre de miles par passager transporté)
- le prix du fuel P
- le taux d'utilisation des capacités de production LF

La fonction de coût est spécifiée sous la forme :

$$\text{Log } C = \alpha + \beta_1 \text{ Log } Q + \beta_2 \text{ Log } P + \beta_3 \text{ LF} + \varepsilon$$

L'estimation par les MCO sur données « empilées » (pooled data) a abouti au résultat suivant :

Dependent Variable: LOGC					
Method: Least Squares					
Sample: 1 90					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
LOGQ	0.883	0.013	66.599	0.000	
LOGP	0.454	0.020	22.359	0.000	
LF	-1.628	0.345	-4.713	0.000	
C	9.517	0.229	41.514	0.000	
R-squared	0.988	Mean dependent var			13.366
Adjusted R-squared	0.988	S.D. dependent var			1.132
S.E. of regression	0.125	Akaike info criterion			-1.284
Sum squared resid	1.335	Schwarz criterion			-1.173
Log likelihood	61.770	F-statistic			2419.341
Durbin-Watson stat	0.383	Prob(F-statistic)			0.000

On notera que les coefficients qui résultent de cet ajustement sont significatifs et ont le signe attendu.

### **Paragraphe 3 : Homogénéité des données et homogénéité des comportements**

#### **A - Intérêt de la question : conséquences d'une hétérogénéité ou d'une homogénéité mal spécifiée**

La toute première question à laquelle on devrait normalement apporter une réponse avant même de songer à l'estimation des coefficients du modèle c'est celle qui consiste à se demander : les individus qui composent l'échantillon présentent-ils ou non des spécificités individuelles susceptibles d'induire des comportements différents pour ce qui concerne la relation étudiée. La question est d'importance car elle va déterminer le choix entre deux grandes catégories de modèles.

- si on peut montrer (par un test approprié) que les individus ont des comportements homogènes, il n'y a pas lieu de tenir compte de spécificités individuelles... sauf à accepter une perte d'efficacité dans l'estimation des coefficients attachés aux différentes variables explicatives. On privilégie dans ce cas une estimation par les MCO sur un échantillon pour lequel les données dans le temps et dans l'espace sont finalement « empilées » les unes sur les autres.
- si, en revanche, on peut montrer qu'il y a bien hétérogénéité des comportements, l'estimation MCO sur données empilées n'est plus appropriée puisqu'elle ignore les spécificités individuelles. On doit alors redouter l'existence d'un biais probable d'omission.

Exemple : Hetero\_Non\_Specifiee.prg. On montre ici qu'une hétérogénéité des comportements non prise en compte se traduit généralement <sup>4</sup> par un estimateur biaisé des coefficients attachés aux différentes variables explicatives.

### **B - Les tests d'homogénéité des comportements**

Il s'agit ici de répondre à la question de l'opportunité de la prise en compte d'une hétérogénéité des comportements. On s'intéresse uniquement à la possibilité d'une hétérogénéité interindividuelle <sup>5</sup>. En partant de l'hypothèse nulle la plus contraignante –celle d'une homogénéité complète des comportements- et en progressant vers des hypothèses alternatives de moins en moins contraignantes pour s'acheminer vers celle d'une hétérogénéité complète, on va construire pour chaque couple d'hypothèse ( $H_0$  versus  $H_A$ ) une statistique de test de Fisher. C'est sur la base de ces statistiques que peut être établi, *in fine*, le type d'hétérogénéité auquel on est confronté.

Cette stratégie de test peut être présentée conformément au schéma de la page suivante.

La statistique de test est une statistique de Fisher construite sur la base des sommes

---

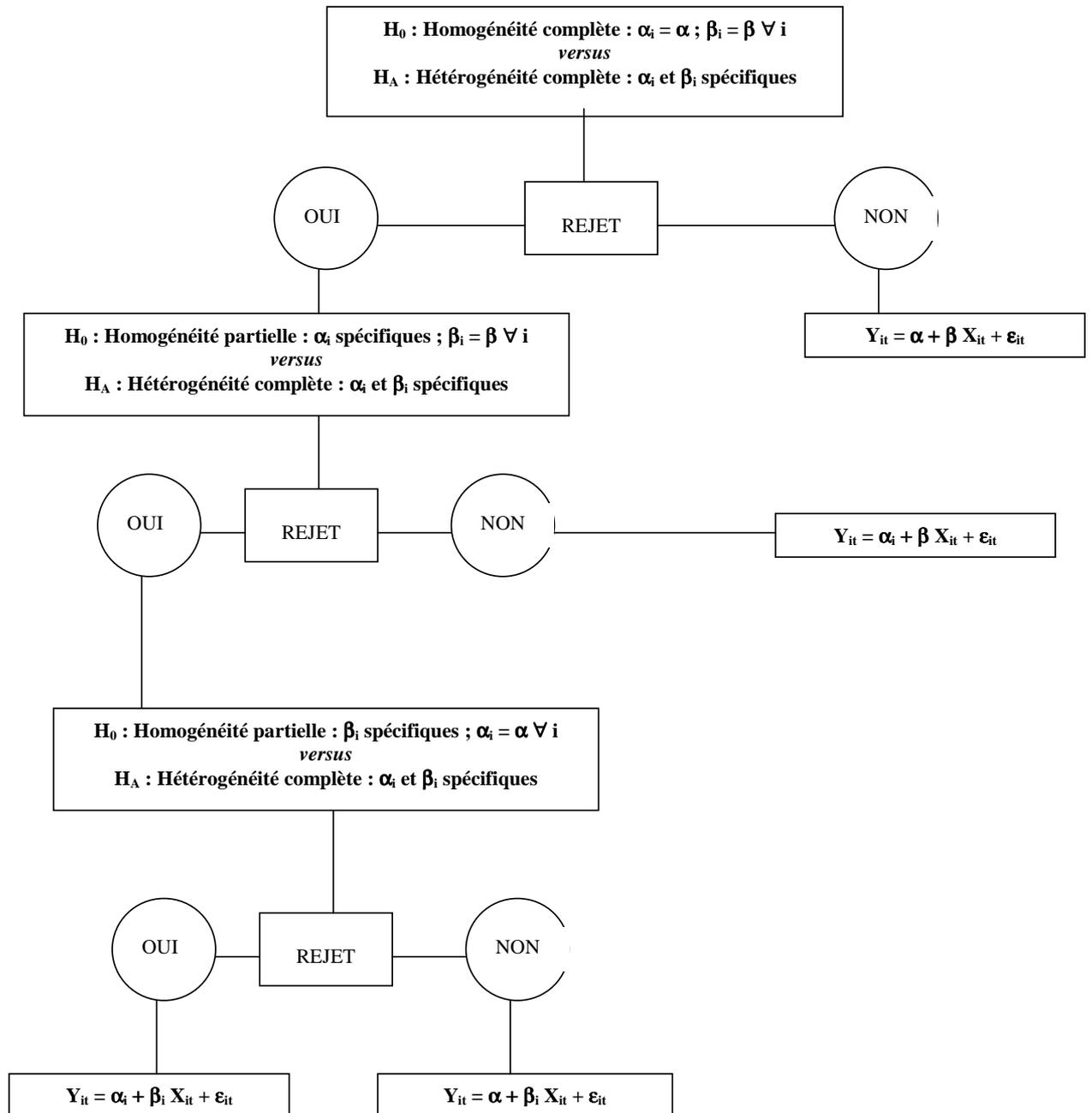
<sup>4</sup> Dans le cas pour lequel la (les) variable(s) explicative n'entretient aucun lien avec l'importance de l'effet fixe, l'estimateur MCO de  $\beta$  obtenu sur données « empilées » est simplement inefficace quoique sans biais et convergent.

<sup>5</sup> On peut en effet concevoir que l'hétérogénéité des comportements puisse s'exprimer non seulement dans la dimension individuelle mais aussi dans la dimension temporelle. Cette hétérogénéité se traduit alors par un modèle pour lequel les coefficients sont susceptibles de prendre des valeurs différentes en fonction de la date à laquelle la relation est appréciée.

des carrés des résidus induites par l'estimation des modèles sous-jacents à l'hypothèse nulle et à l'hypothèse alternative faisant l'objet du test. C'est ainsi, par exemple, que, dans le cas du test d'homogénéité complète *versus* hétérogénéité total, la statistique de test est ainsi calculée :

$$F = \frac{\frac{ESS(H_0) - ESS(H_A)}{(k+1)(N-1)}}{\frac{ESS(H_A)}{NT - N(k+1)}}$$

où N est le nombre d'individus, T le nombre d'observations dans le temps, k le nombre de variables explicatives (hors constante).



Remarque : pour la réalisation de ces différents tests d'hypothèse on utilise le plus souvent, pour des raisons de commodité, la méthode LSDV... ce qui revient à admettre que, si spécificités il y a, celles-ci seraient de type fixe. Comprenons bien que ces tests de spécification ont pour vocation de déceler la présence d'une possible hétérogénéité des comportements mais qu'ils ne nous renseignent en aucune façon sur la nature –fixe ou aléatoire- de ces effets spécifiques.

Note : le plus souvent, on pourra se contenter d'une procédure de test simplifiée qui consiste à confronter les sommes des carrés des résidus de deux modèles simplement :

- le modèle qui postule l'homogénéité complète des comportements d'un individu à l'autre (hypothèse nulle  $H_0$ ) :

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$$

- celui qui intègre une hétérogénéité partielle des comportements, cette hétérogénéité prenant la forme de constantes spécifiques :

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

### **C - Homogénéité des données : la décomposition de la variance**

Au delà de la question de l'homogénéité des comportements, se pose aussi celle de l'homogénéité des données elles mêmes. On dira que, pour une variable donnée (qu'il s'agisse au demeurant de la variable expliquée  $Y$  ou de l'une des  $K$  variables explicatives  $X_j$ ), il y a homogénéité des données si la distribution de cette variable dans le temps est très proche d'un individu à l'autre. Sur la base de cette définition, l'une des conditions de l'homogénéité des données est que, pour la variable étudiée, les moyennes individuelles soient identiques.

Un bon indicateur du degré d'homogénéité des données peut être construit à partir de la décomposition de la variabilité totale TSS de la variable étudiée. On montre ici que la variabilité totale d'une variable (par exemple  $Y$ ) peut être décomposée comme la somme

- de sa variabilité inter-individuelle (variabilité *between* BSS)
- de sa variabilité intra-individuelle (variabilité *within* WSS)

La variabilité totale de Y est donnée par :

$$TSS_{YY} = \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y})^2$$

où  $\bar{Y}$  est la moyenne générale de Y telle qu'elle peut être calculée sur l'ensemble de l'échantillon :

$$\bar{Y} = \frac{1}{NT} \sum_i \sum_t Y_{it}$$

On montre que :

$$TSS_{YY} = WSS_{YY} + BSS_{YY}$$

avec :

$$WSS_{YY} = \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})^2 \quad BSS_{YY} = T \sum_{i=1}^N (\bar{Y}_{io} - \bar{Y})^2$$

$$\text{avec } \bar{Y}_{io} = \frac{1}{T} \sum_t Y_{it}$$

En effet, on a :

$$\begin{aligned} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y})^2 &= \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io} + \bar{Y}_{io} - \bar{Y})^2 \\ &= \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})^2 + \sum_{i=1}^N \sum_{t=1}^T (\bar{Y}_{io} - \bar{Y})^2 + 2 \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})(\bar{Y}_{io} - \bar{Y}) \end{aligned}$$

Or,  $\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})(\bar{Y}_{io} - \bar{Y}) = 0$  puisque :

$$\begin{aligned} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})(\bar{Y}_{io} - \bar{Y}) &= \sum_{i=1}^N \left[ (\bar{Y}_{io} - \bar{Y}) \sum_{t=1}^T (Y_{it} - \bar{Y}_{io}) \right] \\ &= \sum_{i=1}^N \left[ (\bar{Y}_{io} - \bar{Y}) \cdot N(\bar{Y}_{io} - \bar{Y}_{io}) \right] = 0 \end{aligned}$$

$$\Rightarrow \quad TSS_{YY} = WSS_{YY} + BSS_{YY} \quad \text{CQFD}$$

La variabilité totale de Y peut donc être interprétée comme la somme de la variabilité

entre les moyennes individuelles et de la variabilité dans le temps des Y par rapport à leurs moyennes individuelles.

Ce qui est dit ici de la variabilité de Y peut l'être tout aussi bien de n'importe quelle autre variable ou même de la covariabilité de X et Y.

On pourrait montrer de la même façon que la variabilité totale de Y peut aussi être interprétée comme la somme de :

$$\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y})^2 = \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{ot})^2 + \sum_{i=1}^N \sum_{t=1}^T (\bar{Y}_{ot} - \bar{Y})^2$$

où  $\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{ot})^2$  est la variabilité des Y par rapport leurs moyennes temporelles et

$\sum_{i=1}^N \sum_{t=1}^T (\bar{Y}_{ot} - \bar{Y})^2$  est la variabilité des moyennes temporelles par rapport à la moyenne générale.

Deux cas particuliers méritent d'être signalés :

1. S'il y a parfaite homogénéité des données, les moyennes individuelles de Y (resp. de X) sont identiques entre elles et identiques à la moyenne générale de Y (resp. de X). Dans ce contexte, la variabilité inter-individuelle  $BSS_{YY}$  de Y est nulle et la variabilité totale est égale à la variabilité intra-individuelle  $WSS_{YY}$ .

$$\text{Homogénéité parfaite des données} \rightarrow TSS_{YY} = WSS_{YY}$$

2. Si, en revanche, pour un même individu (et ceci quel que soit l'individu considéré), la valeur de Y ne s'écarte jamais de sa moyenne individuelle (pas de variabilité dans le temps), la variabilité intra-individuelle est nulle et la variabilité totale est alors égale à la variabilité inter-individuelle  $BSS_{YY}$

Stabilité dans le temps de la valeur de Y pour un même individu  $\rightarrow TSS_{YY} = BSS_{YY}$

Intérêt de la décomposition de la variabilité des X et des Y : si on peut montrer que la variabilité des X et des Y à travers le temps (variabilité intra-individuelle) est très faible pour un même individu (quel que soit celui-ci), il n'est pas très utile de chercher à exploiter la dimension temporelle du panel : dans ce cas on peut très bien se contenter de travailler simplement sur les moyennes individuelles des X et des Y pour

estimer le modèle... en coupe instantanée (cft *infra* l'estimateur Between).

Exercice : airline\_cost\_function.wf1

On se pose ici la question du degré d'homogénéité des observations relatives aux coûts de production de 6 compagnies aériennes. Appréciez, sous Eviews 5, ce degré d'homogénéité en décomposant la variabilité totale de LOGC.

- Sélectionnez la variable LOGC puis cliquez sur « VIEW » - « Tests for Descriptive Statistics » - « Equality Test by Classification »
- Choisissez la série « i » comme variable classifiante et cochez le test d'égalité des moyennes (individuelles)

Test for Equality of Means of LOGC  
Categorized by values of I  
Date: 14/05/04 Time: 19:18  
Sample: 1 15  
Included observations: 90

Method	df	Value	Probability
Anova F-statistic	(5, 84)	31.87474	0.0000

Analysis of Variance			
Source of Variation	df	Sum of Sq.	Mean Sq.
Between	5	74.67988	14.93598
Within	84	39.36101	0.468583
Total	89	114.0409	1.281358

Category Statistics				
I	Count	Mean	Std. Dev.	Std. Err. of Mean
1	15	14.67563	0.494617	0.127710
2	15	14.37247	0.680536	0.175714
3	15	13.37231	0.522066	0.134797
4	15	13.13580	0.727387	0.187810
5	15	12.36304	0.711945	0.183823
6	15	12.27441	0.891749	0.230249
All	90	13.36561	1.131971	0.119320

On rejette fortement l'hypothèse nulle d'égalité des moyennes individuelles. La part de la variabilité inter-individuelle est forte au sein de la variabilité totale. Il y a hétérogénéité des données relatives à cette variable.

Attention : l'hétérogénéité des données et, en particulier le fait que les moyennes individuelles d'une même variable soient significativement différentes les unes des

autres ne permet pas de préjuger de l'hétérogénéité des comportements. Les moyennes individuelles des variables du modèles peuvent être très différentes d'un individu à l'autre sans que, pour autant ; les comportements soient fondamentalement différents (les paramètres de la fonction de consommation peuvent très bien être identiques d'un individu à l'autre même si ces individus ont des niveaux de revenu et des niveaux de consommation peuvent être très différents.

## **Section II : Le modèle sur données empilées : estimateur MCO et estimateur Between**

Ce type de modélisation n'est approprié que dans le cas pour lequel l'homogénéité des comportements a pu être établi. On passera assez rapidement sur cette première classe de modèle qui ne diffère pas des modèles traditionnels.

En admettant que l'homogénéité des comportements ait donc pu être établie, rien ne s'oppose à ce que la relation  $X \rightarrow Y$  puisse être spécifiée sous la forme :

$$Y = i_{NT} \alpha + X_{NT,K} \beta + \varepsilon$$

et estimée directement par les MCO. Sous réserve que les bonnes propriétés habituelles de l'erreur soient ici encore vérifiées et que l'hypothèse d'homogénéité des comportements soit fondée, l'estimateur MCO de  $\beta$  est sans biais, convergent et efficace.

Dans le cas particulier pour lequel la variabilité intra-individuelle (*within*) de  $Y$  est négligeable (en termes relatifs), on peut, sans grande perte d'information, utiliser l'estimateur Between ou estimateur sur les moyennes individuelles. Le principe en est relativement simple puisqu'il consiste, comme l'indique son nom, à conduire un ajustement MCO non plus sur les données brutes mais sur les moyennes individuelles :

$$\bar{Y}_{i_0} = \alpha + \bar{X}_{i_0} \beta + \bar{\varepsilon}_{i_0}$$

Il est clair que ce type de modèle ne peut être envisagé lorsque les observations sur  $X_k$

et  $Y$  présentent une forte variabilité dans le temps (*within*) : dans ce cas, le fait de ne travailler que sur les moyennes individuelles aboutit à une perte importante d'information (puisque, à vrai dire, on remplace chaque  $X_{k,it}$  par sa moyenne individuelle  $\bar{X}_{k,i0}$ ) qui se paie en termes d'efficacité de l'estimateur.

### Section III : Le modèle à effets fixes

#### Paragraphe 1 : Spécification et estimation. Estimateur LSDV et estimateur Within.

Avec le modèle à effets fixes, on prend en compte l'hétérogénéité des comportements des individus qui composent l'échantillon en considérant que les équations qui régissent les relations  $X \rightarrow Y$  se démarquent, d'un individu à l'autre, par un simple élément constant  $\alpha_i$  de telle sorte que ce modèle a pour expression :

$$/2.1/ \quad Y_{it} = \alpha_i + \beta_1 X_{1,it} + \dots + \beta_K X_{K,it} + \varepsilon_{it}$$

ou encore, sous forme matricielle et pour un individu  $i$  :

$$Y_i = i_T \alpha_i + X_i \beta + \varepsilon_{it}$$

avec :

$$i_T = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ de dimension } T \times 1, \quad X_i = \begin{bmatrix} X_{1,i1} & X_{2,i1} & \dots & X_{K,i1} \\ X_{1,i2} & X_{2,i2} & \dots & X_{K,i2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1,iT} & X_{2,iT} & \dots & X_{K,iT} \end{bmatrix} \text{ de dimension } T \times K$$

$$\text{et } \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \text{ de dimension } K \times 1$$

A un niveau plus agrégé encore (celui de l'ensemble des individus) le modèle est récrit :

$$/2.2/ \quad Y = X \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + D \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \varepsilon$$

$$\text{avec } X = \begin{bmatrix} X_{1,11} & \dots & X_{K,11} \\ \vdots & & \vdots \\ X_{1,1T} & \dots & X_{K,1T} \\ \vdots & & \vdots \\ X_{1,N1} & \dots & X_{K,N1} \\ \vdots & & \vdots \\ X_{1,NT} & \dots & X_{K,NT} \end{bmatrix} \text{ de dimension } NT \times K$$

$$D = \begin{bmatrix} i_T & 0 & \cdots & 0 \\ 0 & i_T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & i_T \end{bmatrix} = [d_1, d_2, \dots, d_N] \text{ de dimension } NT \times N$$

et  $\beta$  de dimension  $K \times 1$

Dans un souci de concision on peut aussi concaténer :

- les matrices X et D horizontalement :

$$Z = [D : X]$$

- les vecteurs de coefficients  $\alpha$  et  $\beta$  verticalement :

$$\delta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad \text{dimension } K+N \times 1$$

et le modèle peut être récrit plus simplement encore :

$$Y = Z \delta + \varepsilon$$

Comme on l'a déjà dit, rien ne s'oppose, théoriquement à ce que les coefficients du modèle (les K coefficients du vecteur  $\beta$  et les N coefficients du vecteur  $\alpha$ ) soient estimés par les MCO :

$$\hat{\delta} = (Z' Z)^{-1} Z' Y$$

Les estimateurs ainsi obtenus directement s'appellent les estimateurs LSDV (Least Square Dummy Variables) en cela que les estimations des constantes spécifiques sont obtenues directement comme on le voit en régressant, en particulier, la variable Y sur un ensemble de variables dummy qui permettent d'identifier les différents individus (c'est la matrice D). Toutefois, dans le cas pour lequel N est grand, on peut se heurter à un problème de capacité de calcul. Dès lors, on doit trouver un artifice pour rendre possible le calcul de ces coefficients.

Les estimateurs  $\hat{\alpha}$  et  $\hat{\beta}$  des coefficients du modèle /2.2/ sont nécessairement tels qu'ils minimisent la somme des carrés des résidus :

$$ESS = [Y - X \hat{\beta} - D \hat{\alpha}]' [Y - X \hat{\beta} - D \hat{\alpha}]$$

Les conditions du premier ordre s'énoncent ainsi :

$$\text{condition C1 : } X' Y = X' X \hat{\beta} + X' D \hat{\alpha}$$

$$\text{condition C2 : } D' Y = D' X \hat{\beta} + D' D \hat{\alpha}$$

Ces conditions du premier ordre constituent un système de  $K + N$  équations à  $K + N$  inconnues dont la résolution fournit les estimations des coefficients.

Ce système – qui peut être très volumineux – peut être considérablement simplifié dans le cas pour lequel  $N$  est grand. On note en effet que de la condition 2 on peut déduire :

$$\hat{\alpha} = (D'D)^{-1} [D' Y - D' X \hat{\beta}]$$

En reportant cette expression de  $\hat{\alpha}$  dans la première équation il est possible de donner l'expression de la matrice des estimateurs des seuls coefficients  $\beta_k$  :

$$\begin{aligned} X' Y &= X' X \hat{\beta} + X' D (D'D)^{-1} [D' Y - D' X \hat{\beta}] \\ &= [X' X - X' D (D'D)^{-1} D' X] \hat{\beta} + X' D (D'D)^{-1} D' Y \end{aligned}$$

soit encore :

$$X' [I - D (D'D)^{-1} D'] Y = X' [I - D (D'D)^{-1} D'] X \hat{\beta}$$

ce qui peut s'écrire plus simplement :

$$X' W Y = X' W X \hat{\beta}$$

en posant  $W = I - D (D'D)^{-1} D'$ . On étudiera par la suite les propriétés – intéressantes – de la matrice  $B = D (D'D)^{-1} D'$  (opérateur « *Between* ») et de la matrice  $W = I - B$  (opérateur « *Within* »).

On en déduit l'expression de l'estimateur *within*  $\hat{\beta}_{with}$  :

$$\hat{\beta}_{with} = (X' W X)^{-1} X' W Y$$

Munis des estimations des coefficients  $\beta_j$  il est ensuite possible, dans un deuxième temps, d'en déduire celles des coefficients  $\alpha_i$ , c'est à dire des effets fixes. On sait en

effet, compte tenu de la deuxième condition du premier ordre que  $\hat{\alpha}$  et  $\hat{\beta}$  sont nécessairement tels que :

$$D'Y = D'X \hat{\beta} + D'D \hat{\alpha}$$

On en déduit que  $\hat{\alpha}_{with} = (D'D)^{-1} D' (Y - X \hat{\beta}_{with})$

## Paragraphe 2 : les opérateurs *between* et *within*

On étudie les propriétés

. de la matrice  $B = D (D'D)^{-1} D'$  opérateur *Between*

. de la matrice  $W = I - D (D'D)^{-1} D'$  opérateur *Within*

et on montre que l'estimateur  $\hat{\beta}$  qui vient d'être défini est l'estimateur des MCO d'un modèle

$$Y_{it} - \bar{Y}_{i \cdot} = \beta_1 (X_{1,it} - \bar{X}_{1,i \cdot}) + \dots + \beta_K (X_{K,it} - \bar{X}_{K,i \cdot}) + (\varepsilon_{it} - \bar{\varepsilon}_{i \cdot})$$

$$\text{avec } \bar{Y}_{i \cdot} = \frac{1}{T} \sum_{t=1}^T Y_{it} \quad \bar{X}_{k,i \cdot} = \frac{1}{T} \sum_{t=1}^T X_{k,it}$$

où les variables, comme on le voit, sont exprimées en écart par rapport à leurs moyennes individuelles.

La matrice  $W = I_{NT,NT} - D (D'D)^{-1} D'$  est de dimension  $NT \times NT$ . Elle a une structure et des propriétés très particulières.

Rappelons que  $D = \begin{bmatrix} i_T & 0 & \dots & 0 \\ 0 & i_T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & i_T \end{bmatrix}$  de telle sorte que cette matrice peut aussi être

réécrite comme le produit de Kronecker :

$$D_{NT \times N} = I_N \otimes i_T$$

de telle sorte que :

$$(\mathbf{D}' \mathbf{D})_{N,N} = \begin{bmatrix} T & 0 & \dots & 0 \\ 0 & T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T \end{bmatrix} \quad \text{et } (\mathbf{D}' \mathbf{D})^{-1}_{N,N} = \begin{bmatrix} \frac{1}{T} & 0 & \dots & 0 \\ 0 & \frac{1}{T} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{T} \end{bmatrix}$$

$$(\mathbf{D} (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}')_{NT,NT} = \begin{bmatrix} \begin{bmatrix} \frac{1}{T} & \dots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \dots & \frac{1}{T} \end{bmatrix} & \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & \begin{bmatrix} \frac{1}{T} & \dots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \dots & \frac{1}{T} \end{bmatrix} & \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & \begin{bmatrix} \frac{1}{T} & \dots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \dots & \frac{1}{T} \end{bmatrix} \end{bmatrix}$$

Au sein de cette matrice  $\mathbf{D} (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}'$  :

- il y a en ligne comme en colonne  $N$  sous blocs de matrices
- où chaque bloc de matrice est lui-même de dimension  $T \times T$

Par exemple, dans le cas pour lequel  $N = 2$  et  $T = 3$  on a :

$$\mathbf{D} (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}' = \begin{bmatrix} 0.33 & 0.33 & 0.33 & 0.00 & 0.00 & 0.00 \\ 0.33 & 0.33 & 0.33 & 0.00 & 0.00 & 0.00 \\ 0.33 & 0.33 & 0.33 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \end{bmatrix}$$

Qu'est-ce qui se passe par conséquent quand on multiplie, par exemple, cette matrice  $\mathbf{D} (\mathbf{D}' \mathbf{D})^{-1} \mathbf{D}'$  par la matrice  $\mathbf{Y}_{NT,NT}$  ? On obtient alors une matrice dont les

composantes correspondent aux moyennes individuelles de  $Y$  :

$$D (D'D)^{-1} D' \times Y = B Y = \begin{bmatrix} \bar{Y}_1 \\ \vdots \\ \bar{Y}_1 \\ \vdots \\ \bar{Y}_N \\ \vdots \\ \bar{Y}_N \end{bmatrix}$$

De la même façon on a :

$$D (D'D)^{-1} D' \times X = BX = \begin{bmatrix} \bar{X}_{1,1} & \dots & \bar{X}_{K,1} \\ \bar{X}_{1,1} & \dots & \bar{X}_{K,1} \\ \vdots & \ddots & \vdots \\ \bar{X}_{1,N} & \dots & \bar{X}_{K,N} \\ \bar{X}_{1,N} & \dots & \bar{X}_{K,N} \end{bmatrix} \text{ qui est de dimension } NT \times K.$$

On voit par conséquent que la matrice  $B = D (D'D)^{-1} D'$  est, en quelque sorte, un **opérateur de calcul des moyennes individuelles**. On l'appelle plus communément l'opérateur « Between ». C'est ainsi qu'on l'appellera désormais.

De ce qui vient d'être dit il découle que la matrice  $W = I_{NT,NT} - B$  autorise, elle, le calcul des écarts d'une variable (qu'il s'agisse de  $Y$  ou des composantes de  $X$ ) par rapport à ses moyennes individuelles :

$$W Y = \begin{bmatrix} Y_{11} - \bar{Y}_{1o} \\ \vdots \\ Y_{1T} - \bar{Y}_{1o} \\ \vdots \\ Y_{N,1} - \bar{Y}_{No} \\ \vdots \\ Y_{N,T} - \bar{Y}_{No} \end{bmatrix} \quad W X = \begin{bmatrix} X_{1,1,1} - \bar{X}_{1,1o} & \dots & X_{K,1,1} - \bar{X}_{K,1o} \\ X_{1,1,T} - \bar{X}_{1,1o} & & X_{K,1,T} - \bar{X}_{K,1o} \\ \vdots & \ddots & \vdots \\ X_{1,N,1} - \bar{X}_{1,No} & \dots & X_{K,N,1} - \bar{X}_{K,No} \\ X_{1,N,T} - \bar{X}_{1,No} & & X_{K,N,T} - \bar{X}_{K,No} \end{bmatrix}$$

L'opérateur *Within*  $W$  présente deux propriétés importantes (tout comme l'opérateur *Between* d'ailleurs) :

- 1/ la matrice  $W$  est symétrique
- 2/ la matrice  $W$  est idempotente

De ces deux propriétés il découle que :

$$W \times W' = W' \times W = W$$

En résumé, on peut écrire que :

$$B = D (D'D)^{-1} D' = \frac{1}{T} I_N \quad J_T$$

$$W = I_{NT} - B = I_N \quad I_T - \frac{1}{T} I_N \quad J_T$$

où  $J_T$  est une matrice carrée de dimension  $T$  dont toutes les composantes sont égales à 1,  $I_N$  et  $I_T$  sont des matrices identité de dimensions  $N$  et  $T$  respectivement.

Exercice : `proprietes_mat_D.prg` : ce programme autorise le calcul de  $D$ ,  $B$ ,  $W$  et d'expérimenter les différentes propriétés de l'opérateur *Within*.

### Paragraphe 3 : variance de l'erreur et variance de l'estimateur *within* dans le modèle à effets fixes

#### A - L'estimateur *within* de la variance de l'erreur

On vient de voir qu'il est possible de simplifier la question de l'estimation des coefficients  $\beta_j$  dans le modèle à effets fixes  $Y_{it} = \alpha_i + \beta_1 X_{1,it} + \dots + \beta_K X_{K,it} + \varepsilon_{it}$  en travaillant sur des variables centrées par rapport à leurs moyennes individuelles. On aboutit ainsi à l'expression de l'estimateur *within* des  $K$  coefficients  $\beta_j$

$$\hat{\beta}_{with} = (X' W X)^{-1} X' W Y$$

et à celle des N effets fixes  $\alpha_i$  :

$$\hat{\alpha}_{with} = (D'D)^{-1} D' (Y - X \hat{\beta}_{with})$$

de telle sorte que, pour chaque  $\alpha_i$  on a :

$$\hat{\alpha}_i = \bar{Y}_{i0} - \hat{\beta}_1 \bar{X}_{1,i0} - \dots - \hat{\beta}_K \bar{X}_{K,i0}$$

Munis de ces estimations des pentes et des effets fixes, il est possible de calculer les résidus de l'ajustement :

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_1 X_{1,it} - \dots - \hat{\beta}_K X_{K,it}$$

ainsi que la somme des carrés de ces résidus et la variance estimée  $s^2_\varepsilon$  de l'erreur du modèle (supposée homoscedastique). Cette variance estimée de l'erreur est donnée par :

$$s^2_\varepsilon = \frac{1}{NT - N - K} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \hat{\alpha}_i - \hat{\beta}_1 X_{1,it} - \dots - \hat{\beta}_K X_{K,it})^2$$

Attention : on a vu que l'estimateur *within* des  $\beta_j$  est l'estimateur des MCO appliqué à des variables centrées sur leurs moyennes individuelles. Si, après avoir centré les variables, on utilise sans précautions la méthode des MCO on va calculer (selon la procédure automatique) la variance de l'erreur comme

$$s^2_\varepsilon = \frac{1}{NT - K} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \hat{\alpha}_i - \hat{\beta}_1 X_{1,it} - \dots - \hat{\beta}_K X_{K,it})^2$$

en oubliant, à tort, que les moyennes individuelles ont du, elles-mêmes, faire l'objet d'une estimation. On sous-estime alors la variance de l'erreur... et donc celles des estimateurs des coefficients.

### **B - L'estimateur *within* de la variance des estimateurs des coefficients**

La matrice des variances covariances estimées des estimateurs des coefficients se calcule comme dans le cas du modèle de régression multiple à ceci près qu'on travaille désormais sur des variables centrées sur leurs moyennes individuelles.

La matrice des variances covariances est donc déterminée à partir des matrices  $Y^*$  et  $X^*$  qui sont les matrices des valeurs centrées de  $Y$  et des  $X_j$  sur leurs moyennes

individuelles ( $Y^* = W Y$  et  $X^* = W X$ ). On la détermine donc comme :

$$\begin{aligned}\hat{V}(\hat{\beta}_{with}) &= s^2_{\varepsilon} (X^{*'} X^*)^{-1} = s^2_{\varepsilon} (X' W' W X)^{-1} \\ &= s^2_{\varepsilon} (X' W X)^{-1} \text{ puisque l'opérateur } within \text{ } W \text{ est symétrique et idempotent.}\end{aligned}$$

Connaissant les variances estimées des estimateurs des coefficients  $\beta_j$  il est ensuite facile de retrouver celles des effets fixes en se servant du fait que les estimateurs des effets fixes sont des combinaisons linéaires des  $\hat{\beta}_j$  et de  $Y$  puisque :

$$\hat{\alpha}_{with} = (D'D)^{-1} D' (Y - X \hat{\beta}_{with}) = \begin{bmatrix} \bar{Y}_{1o} \\ \bar{Y}_{2o} \\ \vdots \\ \bar{Y}_{No} \end{bmatrix} - \begin{bmatrix} \bar{X}_{1,1o} & \bar{X}_{2,1o} & \cdots & \bar{X}_{K,1o} \\ \bar{X}_{1,2o} & \bar{X}_{2,2o} & \cdots & \bar{X}_{K,2o} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{X}_{1,No} & \bar{X}_{2,No} & \cdots & \bar{X}_{K,No} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

Il est alors possible de montrer que l'estimateur de la variance de  $\hat{\alpha}_{i \text{ within}}$  est donné par :

$$\begin{aligned}\hat{V}(\hat{\alpha}_{i \text{ within}}) &= \frac{s^2_{\varepsilon}}{T} + \bar{x}_{io} \hat{V}(\hat{\beta}) \bar{x}_{io}' \\ \text{avec } \bar{x}_{io} &= [\bar{x}_{1,io} \quad \bar{x}_{2,io} \quad \cdots \quad \bar{x}_{K,io}]\end{aligned}$$

où, de façon plus générale :

$$\hat{V}(\hat{\alpha}_{wit}) = \frac{s^2_{\varepsilon}}{T} + \bar{X} \hat{V}(\hat{\beta}) \bar{X}' \quad \text{avec } \bar{X}_{N,K} = \begin{bmatrix} \bar{X}_{1,1o} & \bar{X}_{2,1o} & \cdots & \bar{X}_{K,1o} \\ \bar{X}_{1,2o} & \bar{X}_{2,2o} & \cdots & \bar{X}_{K,2o} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{X}_{1,No} & \bar{X}_{2,No} & \cdots & \bar{X}_{K,No} \end{bmatrix}$$

Il est intéressant de noter que la transformation des variables (et donc du modèle) par le truchement de l'opérateur *within* crée automatiquement un phénomène d'autocorrélation des erreurs là où, dans le modèle séminal, il n'y en avait pas. En effet, on passe du modèle :

$$Y_{it} = \alpha_i + \beta_1 X_{1,it} + \cdots + \beta_K X_{K,it} + \varepsilon_{it}$$

à un modèle en variables d'écart aux moyennes individuelles :

$$Y_{it} - \bar{Y}_{io} = \beta_1 (X_{1,it} - \bar{X}_{1,io}) + \cdots + \beta_K (X_{K,it} - \bar{X}_{K,io}) + (\varepsilon_{it} - \bar{\varepsilon}_{io})$$

Il est clair, dans ces conditions, que le nouveau terme d'erreur ( $\varepsilon_{it} - \bar{\varepsilon}_{io}$ ) est autocorrélé puisque ( $\varepsilon_{it} - \bar{\varepsilon}_{io}$ ) et ( $\varepsilon_{is} - \bar{\varepsilon}_{io}$ ) ont un élément commun  $\bar{\varepsilon}_{io}$ . On note toutefois que cette autocorrélation ne peut jamais être observée qu'à l'intérieur du groupe d'observations relatives à un même individu. En dépit de cette autocorrélation artificielle, il est possible de montrer (sur la base du théorème de Kruskal qui s'applique ici et sous réserve que par ailleurs le terme d'erreur  $\varepsilon_{it}$  présente toutes les bonnes propriétés énoncées plus haut) que l'estimateur des MCO sur données en écarts aux moyennes individuelles est rigoureusement équivalent à l'estimateur MCG appliqué à ces mêmes données<sup>6</sup>.

Exercice : *airline\_cost\_function.wf1* et *between\_estimator.prg* : donnez les estimateurs *between* du modèle :

$$\text{LOGC}_{it} = \alpha + \beta_1 \text{LOGQ}_{it} + \beta_2 \text{LOGP}_{it} + \beta_3 \text{LF}_{it} + \varepsilon_{it}$$

Calculez aussi les t de student et le R2.

## **Section IV : Le modèle à effets aléatoires**

### **Paragraphe 1 : Le modèle**

#### **A - La problématique**

Un des inconvénients du modèle à effets fixes tient à ce que le coefficient  $\alpha_i$  est vraiment propre à l'individu  $i$  et uniquement à cet individu. Ceci signifie que les prévisions construites à partir du modèle estimé ne peuvent jamais porter que sur les individus faisant partie de l'échantillon d'estimation. Autrement dit, les résultats mis en évidence pour les individus faisant partie de cet échantillon ne peuvent être extrapolés à l'ensemble de la population dont cet échantillon a été extrait. Cette propriété n'est guère gênante si l'objet d'étude est un groupe de pays clairement identifiés et dont il s'agit d'identifier les spécificités. Il peut être très intéressant par exemple, s'agissant de l'étude des fonctions de production macroéconomiques des pays qui composent la zone euro de mettre en évidence, d'essayer de voir s'il y a des

---

<sup>6</sup> Cf SEVESTRE pp. 27.

spécificités pays et, le cas échéant, d'identifier les origines possibles de ces spécificités. En revanche, imaginons à présent qu'on s'intéresse aux habitudes de consommation des français en général ; on a constitué à cet effet un échantillon de 10000 individus représentatifs de la population. Après avoir mis en évidence l'opportunité de modéliser l'hétérogénéité des individus on a estimé un modèle à effets fixes qui attribue à chaque individu de l'échantillon une constante spécifique. La question est : quel intérêt y a-t-il à savoir que Mr X présente une constante spécifique supérieure à celle de Mlle Y ? Dans ce cas, le modèle, tel qu'il a pu être estimé sur la base de cet échantillon, est de peu d'intérêt dès lors qu'il s'agit de généraliser les résultats obtenus à l'ensemble de la population. Dans un tel contexte, on doit privilégier l'usage d'un modèle à effets aléatoires.

### B - Spécification du modèle et propriétés supposées des termes d'erreur

On pose désormais que :

$$Y_{it} = \alpha + X_{it} \beta + u_i + \varepsilon_{it} = \alpha + X_{it} \beta + \eta_{it}$$

avec  $\eta_{it} = u_i + \varepsilon_{it}$

ou, en termes matriciels :

- au niveau individuel  $i$  :

$$Y_i = \mathbf{i}_T \alpha + X_i \beta + u_i + \varepsilon_i$$

$\begin{matrix} T \times 1 & T \times 1 & T \times K & K \times 1 & T \times 1 & T \times 1 \end{matrix}$

avec :

$$u_i = \begin{bmatrix} u_i \\ u_i \\ \vdots \\ u_i \end{bmatrix} \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

- au niveau de l'ensemble de tous les individus :

$$Y = \mathbf{i}_{NT} \alpha + X \beta + u + \varepsilon$$

$\begin{matrix} NT \times 1 & NT \times 1 & NT \times K & K \times 1 & NT \times 1 & NT \times 1 \end{matrix}$

On suppose désormais que la spécificité individuelle  $u_i$  dont la valeur, pour un même individu, est stable dans le temps, est la réalisation d'une variable aléatoire dont la loi de distribution est la même pour tous les individus. A partir de là, et en admettant que les coefficients et paramètres du modèle aient déjà été estimés, il devient possible de

construire une prévision pour n'importe quel individu tiré de la population mère quand bien même celui-ci n'aurait pas fait partie de l'échantillon utilisé pour l'estimation... exercice impossible avec le modèle à effet fixe.

On fait sur les erreurs  $u_i$  et  $\varepsilon_{it}$  les hypothèses suivantes (en espérant qu'elles soient vérifiées !):

- $E(\varepsilon_{it} | X) = E(u_i | X) = 0 \forall i, t$  : l'information contenue dans  $X$  n'est d'aucune utilité pour prévoir la valeur de  $\varepsilon_{it}$  ou celle de  $u_i$
- $E(\varepsilon_{it}^2 | X) = \sigma_\varepsilon^2 \forall i, t$  : l'erreur  $\varepsilon$  est homoscédastique
- $E(u_i^2 | X) = \sigma_u^2 \forall i$  : l'éventail des valeurs possibles de  $u_i$  est le même (et avec les mêmes probabilités) d'un individu à l'autre
- $E(u_j \varepsilon_{it} | X) = 0 \forall i, j, t$  : les deux composantes de l'erreur ne sont pas corrélées entre elles, qu'il s'agisse d'un même individu ou non
- $E(\varepsilon_{it} \varepsilon_{js} | X) = 0 \forall i, j, t, s$ , pour  $i \neq j$  ou  $t \neq s$  : les erreurs  $\varepsilon$  ne sont ni autocorrélées ni corrélées entre des individus différents
- $E(u_i u_j | X) = 0 \forall i, j, i \neq j$ : les erreurs spécifiques ne sont pas corrélées d'un individu à l'autre.

**C - Matrice des variances covariances de l'erreur du MCE**

On note que la matrice des variances – covariances des erreurs  $\eta_{it}$  du modèle présente une structure particulière... qui justifiera l'usage d'un estimateur adapté. En effet, en notant  $\eta$  la matrice des erreurs avec  $\eta_{it} = u_i + \varepsilon_{it}$  on a :

$$\eta = \begin{bmatrix} \eta_{11} \\ \vdots \\ \eta_{1T} \\ \vdots \\ \eta_{N1} \\ \vdots \\ \eta_{NT} \end{bmatrix} \text{ et donc } \Omega_{NT \times NT} = E \left[ \begin{array}{ccc|ccc|ccc} \boxed{\begin{matrix} \eta_{11}^2 & \dots & \eta_{11}\eta_{1T} \\ \vdots & \ddots & \vdots \\ \eta_{11}\eta_{1T} & \dots & \eta_{1T}^2 \end{matrix}} & \dots & \boxed{\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}} \\ \vdots & \ddots & \vdots \\ \boxed{\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}} & \dots & \boxed{\begin{matrix} \eta_{N1}^2 & \dots & \eta_{N1}\eta_{NT} \\ \vdots & \ddots & \vdots \\ \eta_{N1}\eta_{NT} & \dots & \eta_{NT}^2 \end{matrix}} \end{array} \right] = I_N \Sigma$$

ou, encore : 
$$\Omega = \begin{bmatrix} \Sigma & 0 & 0 & \dots & 0 \\ 0 & \Sigma & 0 & \dots & 0 \\ 0 & 0 & \Sigma & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Sigma \end{bmatrix}$$

avec  $\Sigma_{T \times T} = E \begin{bmatrix} \eta_{i1}^2 & \dots & \eta_{i1}\eta_{iT} \\ \vdots & \ddots & \vdots \\ \eta_{i1}\eta_{iT} & \dots & \eta_{iT}^2 \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \dots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix}$

$$= \sigma_\varepsilon^2 I_T + \sigma_u^2 i_T i_T' = \sigma_\varepsilon^2 I_T + \sigma_u^2 J_T$$

où  $i_T$  est un vecteur colonne unitaire de dimension  $T$  et  $J_T = i_T i_T'$  une matrice carrée unitaire.

On en déduit que  $\Omega = \sigma_\varepsilon^2 [I_N \otimes I_T] + \sigma_u^2 [I_N \otimes J_T]$ .

Il est clair, compte tenu des propriétés des erreurs énoncées ci-dessus, que :

- les covariances entre les erreurs de deux individus différents sont nulles
- les covariances entre les erreurs d'un même individu mais à des dates différentes sont égales à  $\sigma_u^2$  (puisque  $u_i$  est alors la seule composante commune de ces deux erreurs)
- la variance de l'erreur relative à l'individu  $i$  à la date  $t$  est égale à  $\sigma_u^2 + \sigma_\varepsilon^2$ .

On montre, pour en terminer avec l'étude des propriétés de la matrice  $\Omega$ , que celle-ci peut être exprimée comme une combinaison particulière des **opérateurs** *within* et *between*.

En effet on a :

$$\begin{aligned} \Omega &= \sigma_\varepsilon^2 [I_N \otimes I_T] + \sigma_u^2 [I_N \otimes J_T] \\ &= \sigma_\varepsilon^2 I_N \otimes [I_T + \frac{\sigma_u^2}{\sigma_\varepsilon^2} J_T] \end{aligned}$$

$$\begin{aligned}
&= \sigma_\varepsilon^2 \mathbf{I}_N \otimes \left[ \mathbf{I}_T - \frac{\mathbf{J}_T}{T} + \frac{\mathbf{J}_T}{T} + \frac{\sigma_u^2}{\sigma_\varepsilon^2} \mathbf{J}_T \right] \\
&= \sigma_\varepsilon^2 \mathbf{I}_N \otimes \left[ \left( \mathbf{I}_T - \frac{\mathbf{J}_T}{T} \right) + \frac{\mathbf{J}_T}{T} \left( 1 + T \frac{\sigma_u^2}{\sigma_\varepsilon^2} \right) \right] \\
&= \sigma_\varepsilon^2 \mathbf{I}_N \otimes \left[ \left( \mathbf{I}_T - \frac{\mathbf{J}_T}{T} \right) + \frac{\mathbf{J}_T}{T} \cdot \frac{\sigma_\varepsilon^2 + T \sigma_u^2}{\sigma_\varepsilon^2} \right] \\
&= \sigma_\varepsilon^2 \left[ \mathbf{I}_N \otimes \left( \mathbf{I}_T - \frac{\mathbf{J}_T}{T} \right) + \mathbf{I}_N \otimes \frac{\mathbf{J}_T}{T} \cdot \frac{\sigma_\varepsilon^2 + T \sigma_u^2}{\sigma_\varepsilon^2} \right] \\
&= \sigma_\varepsilon^2 \left[ \mathbf{W} + \frac{\sigma_\varepsilon^2 + T \sigma_u^2}{\sigma_\varepsilon^2} \mathbf{B} \right] \\
&= \sigma_\varepsilon^2 \mathbf{W} + (\sigma_\varepsilon^2 + T \sigma_u^2) \mathbf{B}
\end{aligned}$$

On voit donc bien que la matrice des variances covariances est, en effet, une combinaison linéaire des **opérateurs** *within* et *between*.

On peut facilement vérifier (résultat qu'on aura à réutiliser) que :

$$\Omega^{-1} = \frac{1}{\sigma^2} [\mathbf{W} + \psi \mathbf{B}] \quad \text{avec } \psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T \sigma_u^2}$$

## **Paragraphe 2 : Propriétés des estimateurs *Between* et *Within* appliqués au modèle à composantes d'erreur.**

En admettant que le modèle  $X \rightarrow Y$  prenne la forme d'un modèle à composantes d'erreur :

$$Y = \mathbf{i} \alpha + \mathbf{X} \beta + \mathbf{u} + \varepsilon$$

on peut envisager plusieurs classes d'estimateurs « élémentaires » pour les coefficients du modèle.

L'estimateur des MCO :  $\hat{\beta}_{mco} = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{Y}$

en prenant la précaution de définir  $\mathbf{X}_1$  comme étant la matrice  $\mathbf{X}$  augmentée d'une première colonne unitaire

L'estimateur *Within* :  $\hat{\beta}_{within} = (X' W X)^{-1} X' W Y$  <sup>7</sup>

L'estimateur *Between* :  $\hat{\beta}_{Between} = (X_1' B X_1)^{-1} X_1' B Y$

On note au passage que le vecteur  $\hat{\beta}_{between}$  (tout comme le vecteur  $\hat{\beta}_{mco}$  d'ailleurs) présente une composante de plus (l'estimateur du coefficient  $\alpha$ ) que le vecteur  $\hat{\beta}_{within}$ .

Quoiqu'il puisse être démontré que, dans le contexte défini par les hypothèses relatives aux termes d'erreur, ces estimateurs demeurent sans biais <sup>8</sup> et convergents, ils ne sont pas, en revanche, de variance minimum (cf. cours de licence). On doit leur préférer l'estimateur des MCG.

### Paragraphe 3 : L'estimateur des MCG

On sait d'ores et déjà (démonstration *supra*) que la matrice  $\Omega$  des variances – covariances des erreurs du modèle MCE

1. n'est pas diagonale
2. est une combinaison linéaire des **opérateurs** *Within* et *Between*.

On sait également (c'est un résultat connu de l'algèbre des matrices) que, si la matrice  $\Omega$  est définie positive (ce qui est en effet le cas), alors il existe nécessairement une matrice  $H$ , de même dimension que la matrice  $\Omega$  telle que :

$$H \Omega H' = I$$

On déduit de ce résultat que :

---

<sup>7</sup> Attention : l'estimateur *within* ne peut être défini comme  $(X_1' W X_1)^{-1} X_1' W Y$  puisque, dans ce cas, la matrice  $(X_1' W X_1)$  n'est pas inversible. On note en effet que toutes les composantes de la première colonne de la matrice  $X_1$  sont égales à 1. On en déduit que la première colonne de  $X_1' W$  est nulle... et donc  $(X_1' W X_1)$  n'est pas inversible.

<sup>8</sup> Ceci est particulièrement simple à mettre en évidence dans le cas de l'estimateur *within* puisque, en travaillant sur les écarts aux moyennes individuelles, on gomme la totalité des effets aléatoires spécifiques  $u_i$  :

$$Y_{it} = \alpha + \beta X_{it} + u_i + \varepsilon_{it} \quad \bar{Y}_i = \alpha + \beta \bar{X}_i + \bar{u}_i + \bar{\varepsilon}_i \quad Y_{it} - \bar{Y}_i = \beta (X_{it} - \bar{X}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

$$\Omega = (H' H)^{-1} \quad 9$$

soit encore :

$$H' H = \Omega^{-1}$$

On peut montrer, compte tenu de la structure désormais familière de la matrice  $\Omega$ , que cette matrice  $H$  a pour expression :

$$H = I_N \otimes H_i$$

avec :

$$H_i = \frac{1}{\sigma_\varepsilon} \left[ I_T - \frac{\theta}{T} i_T i_T' \right]$$

$$\text{où } \theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_u^2}} \quad 10$$

Quel intérêt y a-t-il à définir cette matrice  $H$  ?

La réponse à cette question passe par l'étude des propriétés du terme d'erreur dans le modèle ainsi transformé :

$$HY = HX \beta + H(\varepsilon + u) \quad 11$$

On note en effet que, si dans le modèle primitif ( $Y = X_1 \beta + \varepsilon + u$ ) le terme d'erreur  $\varepsilon + u$  est auto-corrélé il en va différemment dans le modèle transformé puisque :

$$E [H(\varepsilon + u)(\varepsilon + u)' H'] = H \Omega H' = I_{NT}$$

---

<sup>9</sup> En effet on a :  $H \Omega H' = I \Rightarrow H' H \Omega H' = H' \Rightarrow \Omega H' = (H' H)^{-1} H' \Rightarrow \Omega H' H = I \Rightarrow \Omega = (H' H)^{-1}$

<sup>10</sup> Le programme `omega_hprimh.prg` met en évidence l'équivalence entre ces deux écritures de  $\Omega$  :

$$\Omega = \sigma_\varepsilon^2 [I_N \otimes I_T] + \sigma_u^2 [I_N \otimes J_T]$$

$$\Omega = (H' H)^{-1}$$

<sup>11</sup> Rappelons que la matrice  $X$  est ici augmentée d'une première colonne unitaire.

de telle sorte que, au final, l'erreur  $H(\varepsilon + u)$  du nouveau modèle est homoscédastique... et le modèle peut (pourrait) alors être estimé par les MCO.

L'estimateur MCG est donc l'estimateur MCO du modèle transformé :

$$\hat{\beta}_{mcg} = (\mathbf{X}'\mathbf{H}'\mathbf{H}\mathbf{X})^{-1} \mathbf{X}'\mathbf{H}'\mathbf{H}\mathbf{Y} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}$$

On peut montrer que l'estimateur MCG est une moyenne pondérée des estimateurs *within* et *between*. En effet, on sait que :

$$\boldsymbol{\Omega} = \sigma_{\varepsilon}^2 [\mathbf{I}_N \otimes \mathbf{I}_T] + \sigma_u^2 [\mathbf{I}_N \otimes \mathbf{J}_T]$$

et donc :

$$\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma_{\varepsilon}^2} \left[ \mathbf{I}_N \otimes \mathbf{I}_T - \frac{\sigma_u^2}{\sigma_{\varepsilon}^2 + T\sigma_u^2} \mathbf{I}_N \otimes \mathbf{J}_T \right]$$

où, rappelons-le,  $\mathbf{J}_T$  est une matrice carrée de dimension  $T$  et dont tous les éléments sont égaux à 1.

En partant de cette dernière expression de  $\boldsymbol{\Omega}^{-1}$  on peut écrire que :

$$\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma_{\varepsilon}^2} \left[ \mathbf{I}_N \otimes \mathbf{I}_T - \frac{1}{T} \mathbf{I}_N \otimes \mathbf{J}_T + \frac{1}{T} \mathbf{I}_N \otimes \mathbf{J}_T - \frac{\sigma_u^2}{\sigma_{\varepsilon}^2 + T\sigma_u^2} \mathbf{I}_N \otimes \mathbf{J}_T \right]$$

Or, on se souvient que l'opérateur *between*  $\mathbf{B}$  et l'opérateur *within*  $\mathbf{W}$  sont tels que :

$$\mathbf{B} = \mathbf{D} (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}' = \frac{1}{T} \mathbf{I}_N \quad \mathbf{J}_T$$

$$\mathbf{W} = \mathbf{I}_{NT} - \mathbf{B} = \mathbf{I}_N \quad \mathbf{I}_T - \frac{1}{T} \mathbf{I}_N \quad \mathbf{J}_T$$

de telle sorte que :

$$\begin{aligned} \boldsymbol{\Omega}^{-1} &= \frac{1}{\sigma_{\varepsilon}^2} \left[ \mathbf{W} + \frac{1}{T} \mathbf{I}_N \otimes \mathbf{J}_T - \frac{\sigma_u^2}{\sigma_{\varepsilon}^2 + T\sigma_u^2} \mathbf{I}_N \otimes \mathbf{J}_T \right] \\ &= \frac{1}{\sigma_{\varepsilon}^2} \left[ \mathbf{W} + \frac{1}{T} \left[ \mathbf{I}_N \otimes \mathbf{J}_T - \frac{T\sigma_u^2}{\sigma_{\varepsilon}^2 + T\sigma_u^2} \mathbf{I}_N \otimes \mathbf{J}_T \right] \right] = \frac{1}{\sigma_{\varepsilon}^2} \left[ \mathbf{W} + \frac{1}{T} \left[ \frac{\sigma_{\varepsilon}^2}{\sigma_{\varepsilon}^2 + T\sigma_u^2} \mathbf{I}_N \otimes \mathbf{J}_T \right] \right] \end{aligned}$$

$$= \frac{1}{\sigma_\varepsilon^2} \left[ W + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} B \right] = \frac{1}{\sigma_\varepsilon^2} [W + \psi B]$$

en posant que :

$$\psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}$$

Dès lors, l'expression de l'estimateur  $\hat{\beta}_{mcg}$  devient :

$$\begin{aligned} \hat{\beta}_{mcg} &= [X' (W + \psi B) X]^{-1} X' (W + \psi B) Y \\ &= [X' (W + \psi B) X]^{-1} X' W Y + \psi [X' (W + \psi B) X]^{-1} X' B Y \end{aligned}$$

Posons à présent :

$$F^W = [X' (W + \psi B) X]^{-1} X' W X$$

$$F^B = \psi [X' (W + \psi B) X]^{-1} X' B X$$

On voit tout de suite que  $F^W + F^B = 1$  et que, donc :  $F^B = I - F^W$ .

$$\text{Or } F^W \hat{\beta}_{with} = [X' (W + \psi B) X]^{-1} X' W X (X' W X)^{-1} X' W Y$$

$$= [X' (W + \psi B) X]^{-1} X' W Y \text{ correspond au premier élément de } \hat{\beta}_{mcg}$$

De la même façon, on a :

$$F^B \hat{\beta}_{betw} = (I - F^W) \hat{\beta}_{betw} = \psi [X' (W + \psi B) X]^{-1} X' B X (X' B X)^{-1} X' B Y$$

$$= \psi [X' (W + \psi B) X]^{-1} X' B Y \text{ correspond, lui, au second élément de } \hat{\beta}_{mcg}.$$

On en déduit que :

$$\hat{\beta}_{mcg} = F^W \hat{\beta}_{within} + (I - F^W) \hat{\beta}_{between} \quad cqfd$$

On notera que :

1. dans le cas pour lequel  $\psi = 0$  (ce qui se produit, compte tenu de l'expression même de  $\psi$ , lorsque le nombre  $T$  des observations dans le temps tend vers l'infini), tous les éléments de la matrice  $F^B = I - F^W$  sont nuls ; l'estimateur des MCG converge donc vers l'estimateur *within* quand  $T$  tend vers l'infini.

2. dans le cas opposé pour lequel  $\psi = 1$  (ce qui se produit lorsque la variance de  $u_i$  tend vers 0, c'est à dire lorsqu'il n'y a pas d'effets spécifiques puisque les  $u_i$  sont alors identiques), l'estimateur des MCG converge vers l'estimateur MCO du modèle en pooling (aucune spécificité individuelle) puisque, en effet, dans ce cas on a :

$$\Omega = \sigma_\varepsilon^2 [I_N \otimes I_T] + \sigma_u^2 [I_N \otimes J_T] = \sigma_\varepsilon^2 I_{NT}$$

On voit tout de suite qu'on retrouve alors une matrice des variances-covariances qui est diagonale (pas d'autocorrélation) et homoscedastique : l'estimateur des MCG est alors analogue à celui des MCO.

Dans le cas général pour lequel  $\psi$  est compris strictement entre 0 et 1, l'estimateur des MCG doit par conséquent être interprété comme une combinaison linéaire de l'estimateur *within* (hétérogénéité forte) et de l'estimateur MCO sur données empilées (absence d'hétérogénéité).

La matrice des variances covariances de l'estimateur des MCG est donnée (résultat connu) par :

$$V(\hat{\beta}_{mcg}) = [X' \Omega^{-1} X]^{-1} = \sigma_\varepsilon^2 [X' (W + \psi B) X]^{-1}$$

On peut montrer que les éléments diagonaux de cette matrice (c'est à dire les variances des estimateurs MCG des coefficients) sont inférieurs à ce qu'eussent été ces mêmes éléments, dans le même contexte marqué par la présence d'effets aléatoires, si on avait privilégié l'usage (alors inapproprié) de l'estimateur à effets fixes *within*. Dans ce contexte, l'estimateur des MCG est l'estimateur BLUE (quoique, rappelons-le, l'estimateur *within*, en particulier, demeure un estimateur sans biais et convergent).

#### **Paragraphe 4 : L'estimateur des MCQG**

Un des problèmes qui se pose avec le calcul de l'estimateur  $\hat{\beta}_{mcg}$  (et non des moindres !) c'est qu'il requiert nécessairement que soit connue la matrice des variances covariances  $\Omega$ , ou, ce qui est équivalent, que soit connue la valeur du paramètre  $\psi$ , c'est à dire encore les variances  $\sigma_\varepsilon^2$  et  $\sigma_u^2$ . Or, généralement, ces

paramètres sont inconnus : il faut les estimer. On peut imaginer pour cela plusieurs procédures. L'une des méthodes <sup>12</sup> les plus couramment utilisées (sans doute en raison de sa simplicité) consiste à exploiter les variances des résidus issus de l'estimation du modèle à composantes d'erreur par les méthodes *within* d'une part et *between* d'autre part.

### La variance résiduelle de l'estimateur *within* du modèle à effets aléatoires

Rappelons que, en partant de l'expression générale du modèle à erreurs composées

$$Y_{it} = \alpha + \sum_{k=1}^K \beta_k X_{k,it} + \eta_{it} \quad \text{avec } \eta_{it} = u_i + \varepsilon_{it}$$

on peut éliminer les spécificités individuelles en travaillant sur des variables centrées autour de leurs moyennes individuelles :

$$Y_{it} - Y_{io} = \sum_{k=1}^K \beta_k (X_{k,it} - X_{k,io}) + \varepsilon_{it} - \varepsilon_{io}$$

soit matriciellement :

$$W Y = W X \beta + W \varepsilon$$

L'estimateur *within* du modèle à erreurs composées est donné par :

$$\hat{\beta}_{with} = (X' W' W X)^{-1} X' W' W Y$$

et, comme la matrice  $W$  (opérateur *within*) est symétrique et idempotente

$$\hat{\beta}_{with} = (X' W X)^{-1} X' W Y$$

La matrice des résidus issus de l'estimation du modèle à effets aléatoires sur données centrées sur les moyennes individuelles est donnée par :

$$\hat{\varepsilon}_{with} = W Y - W X \hat{\beta}_{with}$$

On peut montrer<sup>13</sup> que la somme des carrés des résidus pondérée par le nombre de degrés de liberté de ce modèle (NT observations – N moyennes individuelles – K coefficients)

---

<sup>12</sup> Méthode de Swamy et Arora (1972).

<sup>13</sup> Voir P. SEVESTRE p.60-62

$$s_w^2(\varepsilon) = \frac{\hat{\varepsilon}_{\text{with}}' \hat{\varepsilon}_{\text{with}}}{N(T-1) - K}$$

est un estimateur sans biais de la variance théorique de la composante non spécifique de l'erreur dans le modèle MCE :

$$E s_w^2(\varepsilon) = \sigma_\varepsilon^2$$

### La variance résiduelle de l'estimateur *between* du modèle à effets aléatoires

En appliquant l'opérateur *between* au modèle à composantes d'erreur il vient :

$$Y_{io} = \beta_0 + \sum_{j=1}^k \beta_j X_{j,io} + \eta_{io} \quad \text{avec } \eta_{io} = u_i + \varepsilon_{io}$$

soit, matriciellement :

$$B Y = B X_1 \beta + B \eta$$

en prenant, on le rappelle, la précaution de définir  $X_1$  comme la matrice  $X$  augmentée d'une première colonne dont tous les éléments sont égaux à l'unité.

L'estimateur *between* du modèle ainsi transformé est :

$$\hat{\beta}_{betw} = (X_1' B X_1)^{-1} X_1' B Y$$

et le vecteur des résidus <sup>14</sup> qui en résulte :

$$\hat{\eta}_{betw} = B Y - B X_1 \hat{\beta}_{betw}$$

La variance des résidus issus de ce modèle :

$$s_B^2(\eta) = \frac{1}{(N - k_b) T} \sum_{i=1}^N \sum_{t=1}^T \hat{\eta}_{it}^2 = \frac{\hat{\eta}'_{betw} \hat{\eta}_{betw}}{(N - k_b) T}$$

est, elle, un estimateur sans biais de la variance théorique de  $u_i + \varepsilon_{io}$ , c'est à dire de

---

<sup>14</sup> Attention : il convient de noter tout particulièrement que ce vecteur est un vecteur à  $NT$  composantes et non pas à  $N$  composantes individuelles seulement. Pour un même individu il y a  $T$  résidus identiques. Cette remarque a son importance dans le calcul de la somme des carrés de ces résidus !

$$\sigma_u^2 + \frac{\sigma_\varepsilon^2}{T} \quad 15.$$

Ainsi munis des estimateurs de  $\sigma_u^2$  d'une part et de  $\sigma_u^2 + \frac{\sigma_\varepsilon^2}{T}$  d'autre part il est alors

possible de fournir un estimateur sans biais du paramètre  $\psi = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}$  :

$$\hat{\psi} = \frac{s_w^2}{T s_B^2}$$

L'estimateur des MCQG est alors fourni par :

$$\hat{\beta}_{mcqg} = [X' (W + \hat{\psi} B) X]^{-1} X' (W + \hat{\psi} B) Y$$

L'estimateur MCQG de la constante  $\beta_0$  dans le modèle à composante d'erreur est donné par :

$$\hat{\beta}_{0 mcqg} = Y_{00} - \sum_{j=1}^K \hat{\beta}_{jmcqg} X_{j,00}$$

Enfin, la matrice des variances covariances théoriques des estimateurs des coefficients :

$$V(\hat{\beta}_{mcg}) = [X' \Omega^{-1} X]^{-1} = \sigma_\varepsilon^2 [X' (W + \psi B) X]^{-1}$$

où  $\sigma_\varepsilon^2$  est la variance théorique de la composante non spécifique de l'erreur peut être approchée par  $s_w^2(\varepsilon)$  puisque  $E s_w^2(\varepsilon) = \sigma_\varepsilon^2$ . Dès lors, la matrice des variances covariances des estimateurs des coefficients peut être estimée par :

$$\hat{V}(\hat{\beta}_{mcg}) = s_w^2(\varepsilon) [X' (W + \hat{\psi} B) X]^{-1}$$

### **Paragraphe 5 : Effets fixes ou effets aléatoires : le test d'Hausman.**

En admettant que nous ayons déjà la certitude de la présence de spécificités individuelles, se pose maintenant la question de la nature de ces spécificités : doit-on retenir une spécification de type « effets fixes » ou, au contraire, de type « effets aléatoire ». La question n'est pas sans importance.

On se souvient du résultat selon lequel l'estimateur des MCG est l'estimateur « optimal » du modèle à effets aléatoires... **à la condition que l'erreur du modèle (et, notamment, sa composante spécifique) ne soit pas suspecte de corrélation avec les différentes variables explicatives.** Si, malheureusement, tel n'était pas le cas, l'estimateur des MCG est alors biaisé... ce qui est évidemment très gênant. Or, si on peut être disposé à admettre l'absence de corrélation entre la composante non spécifique de l'erreur et les variables explicatives, on éprouve quelque réticence à admettre ce même postulat d'absence de corrélation entre la composante spécifique et les variables explicatives qui sont elles-mêmes sensées rendre compte, justement, des caractéristiques propres de chaque individu. C'est **la critique de Mundlack.**

On commence par expérimenter

1. le caractère **biaisé** de l'estimateur des MCQG et le caractère **non biaisé** de l'estimateur *within* si l'erreur spécifique est corrélée avec l'une au moins des variables explicatives. Dans ce cas, il existe une différence significative entre les estimations engendrées par ces deux estimateurs
2. le caractère **non biaisé** des estimateurs *within et MCQG* si l'erreur spécifique n'est pas corrélée avec les variables explicatives. Dans ce cas, il n'y a pas de différence significative entre les estimations engendrées par ces deux estimateurs. On privilégiera toutefois, dans ce contexte, l'usage de l'estimateur des MCQG dont on vérifie qu'il est alors plus efficace que l'estimateur *within*.

On définit ensuite la statistique de test de Hausman pour l'hypothèse nulle d'absence de corrélation entre erreur spécifique et variables explicatives. Du résultat de ce test dépend le choix de la méthode d'estimation qui sera finalement retenue.

### **A - Corrélation des erreurs spécifiques avec les variables explicatives : quelles propriétés pour les estimateurs *within* et *mcqg* ?**

On expérimente les propriétés des deux estimateurs (*within* et *mcqg*) en présence et en l'absence de corrélation entre erreur spécifique et variables explicatives.

---

<sup>15</sup> On doit noter en effet que  $\text{Var}(\eta_{io}) = \text{Var}(u_i + \varepsilon_{io}) = \text{Var}(u_i) + \text{Var}(\varepsilon_{io}) = \text{Var}(u_i) + \text{Var}(\varepsilon_{it}) / T$

Programme : WITH\_OU\_MCQG\_EFF\_ALEAT.PRG

A 100 reprises, on crée un échantillon de données de panel pour les variables X et Y. On note N le nombre des individus et T la taille de l'échantillon dans la dimension temporelle. Pour la simplicité on suppose

- . que le panel est « cylindré »
- . que le nombre des variables explicatives est égal à 1 (ce qui ne nuit pas à la généralité du résultat).

On a supposé ici que  $N = 20$  et  $T = 20$ .

A chaque itération et pour chaque individu

1/ on produit une chronique d'erreurs non spécifiques  $\varepsilon_{it}$ . On suppose ces erreurs non autocorrélées dans le temps et non corrélées de manière interindividuelle. On suppose que les  $\varepsilon_{it}$  sont IIDN(0,1).

2/ on produit une erreur spécifique u telle que la réalisation de cette erreur pour l'individu i est tirée aléatoirement d'une distribution Normale centrée et d'écart type  $\sigma_u = 5$ . On se souvient que la valeur prise par cette erreur, spécifique à l'individu i, est la même quel que soit t.

3/ On crée une chronique pour la variable explicative X en ayant soin, selon le cas, d'introduire ou, au contraire, de ne pas introduire un élément de corrélation entre l'erreur spécifique et les réalisations de X

- s'il n'y a pas de corrélation, les réalisations de X sont tirées aléatoirement d'une distribution du Chi-Deux à 1 DDL
- s'il y a corrélation, on augmente la valeur de  $X_{it}$  d'un élément  $\rho u_i$  en posant ici  $\rho = 0.5$

4/ Munis des réalisations de X, de U et de  $\varepsilon$  on crée ensuite la chronique de Y sur la base d'une relation du type :

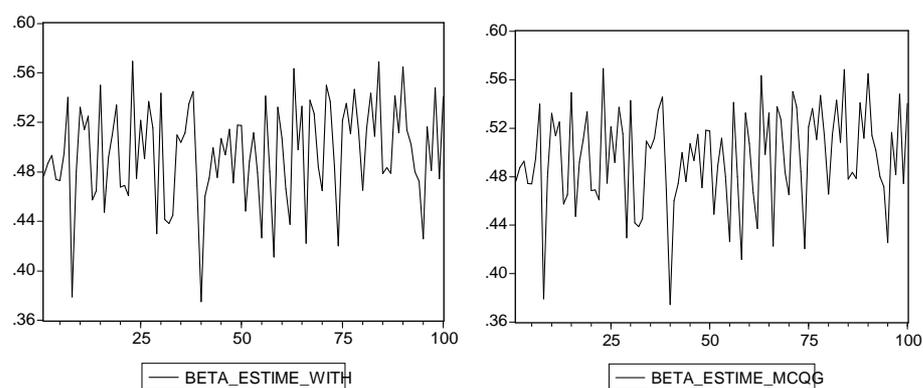
$$Y_{it} = 10 + 0.5 X_{it} + \eta_{it}$$

$$\text{avec } \eta_{it} = u_i + \varepsilon_{it}$$

5/ On utilise successivement les trois estimateurs : OLS sur données empilées,

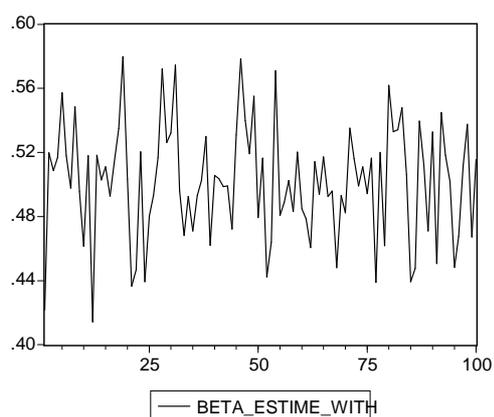
### Within et MCQG.

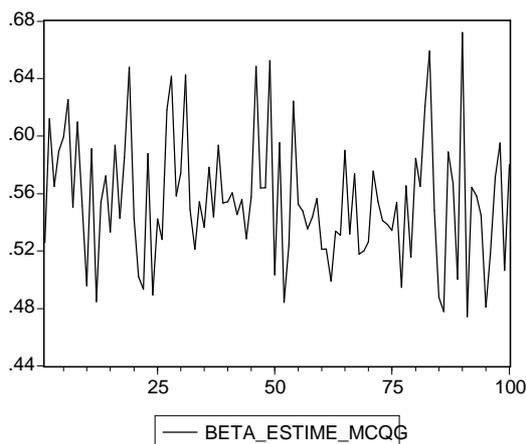
Dans le cas, pour commencer, pour lequel il n'y a pas de corrélation entre  $X_{it}$  et l'erreur spécifique  $u_i$  les estimations *within* et *mcqg* du coefficient attaché à X obtenues sur la base de ces 100 simulations présentent les caractéristiques suivantes :



On note que, dans ce cas, les deux estimateurs sont sans biais et fournissent des estimations assez voisines.

Dans le cas, à présent, pour lequel il existerait un lien entre la variables explicative et l'erreur spécifique (cas de figure qui n'a rien d'exceptionnel) les 100 estimations engendrées par l'usage des estimateurs *within* et *mcqg* présentent les caractéristiques suivantes :





Il est important de noter ici que l'estimateur des MCQG est désormais biaisé (il est aussi non convergent) de telle sorte qu'il existe une différence sensible entre les estimations issues de ces deux estimateurs.

Le test de Hausman est, précisément, fondé sur ces propriétés comparées des estimateurs *within* et *mcqg* en présence ou, au contraire, en l'absence d'un phénomène de corrélation  $X - u$ .

### **B - La statistique de test de Hausman**

On veut éprouver l'hypothèse nulle d'exogénéité des variables explicatives par rapport à l'erreur spécifique du modèle... étant bien entendu que, du résultat de ce test, dépend la mise en oeuvre d'un estimateur ou d'un autre. L'intuition du test est que :

. s'il y a une différence significative entre les estimations *within* et *mcqg* c'est que, vraisemblablement, l'une au moins des variables explicatives est corrélée avec l'erreur spécifique. Dans ce cas il convient de privilégier l'usage de l'estimateur *within* qui garantit des estimations sans biais des coefficients  $\beta_j$

. s'il n'y a pas de différence significative entre les estimations *within* et *mcqg* c'est que les variables sont exogènes par rapport à l'erreur spécifique ; dans ce cas on privilégie l'usage de l'estimateur *mcqg* qui est alors sans biais et plus efficace que l'estimateur *within*.

Il s'agit donc de tester l'hypothèse nulle :

$$H_0 : \hat{\beta}_{j \text{ within}} - \hat{\beta}_{j \text{ mcqg}} = 0 \quad \forall j$$

$$\text{versus } H_A : \exists j \mid \hat{\beta}_{j \text{ within}} - \hat{\beta}_{j \text{ mcqg}} \neq 0$$

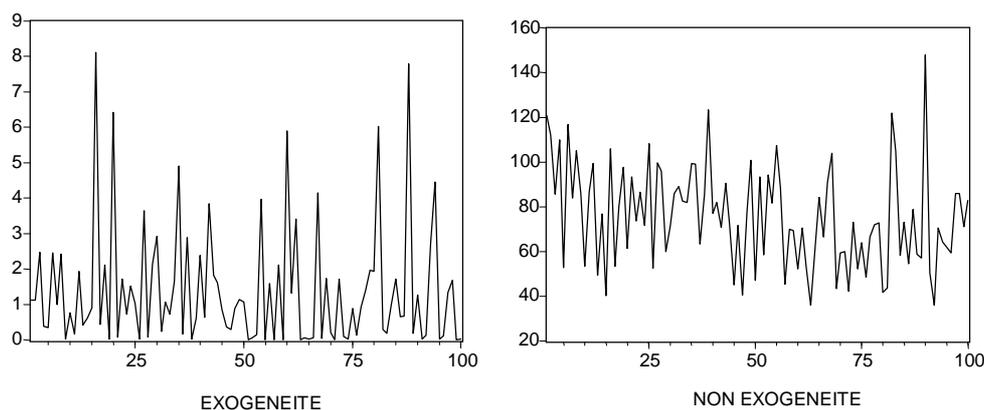
La statistique de test est :

$$\xi_H = (\hat{\beta}_{within} - \hat{\beta}_{mcqg}) [\hat{V}(\hat{\beta}_{within} - \hat{\beta}_{mcqg})]^{-1} (\hat{\beta}_{within} - \hat{\beta}_{mcqg})$$

où  $\hat{V}(\hat{\beta}_{within} - \hat{\beta}_{mcqg})$  est la matrice des variances – covariances des différences entre les deux estimateurs. Sous  $H_0$ , cette statistique est distribuée selon une loi du Chi-deux à  $k$  degrés de liberté (où  $k$  est le nombre de composantes du vecteur  $\beta$  des coefficients). Il est possible de montrer que, sous  $H_0$ , cette matrice peut être réécrite comme :

$$\hat{V}(\hat{\beta}_{within} - \hat{\beta}_{mcqg}) = \hat{V}(\hat{\beta}_{within}) - \hat{V}(\hat{\beta}_{mcqg})$$

Le graphique ci-dessous reproduit les valeurs qui ont été obtenues pour la statistique de Hausman dans chaque cas de figure (absence / présence d'un phénomène de corrélation en  $X$  et  $u$ ) :



La valeur critique pour un seuil de 5 % est, ici, égale à 3.841, de telle sorte que, dans le premier cas (exogénéité) on n'est amené à rejeter l' $H_0$  que dans, approximativement, 5 % des cas.

### **Section V : L'appréciation de l'aptitude du modèle à reproduire les volatilités totale, intra et inter individuelles de Y**

A l'issue de l'estimation d'un modèle en données de panel on est en droit de se poser quelques questions sur la qualité elle-même de l'ajustement. Dans le cadre traditionnel du modèle de régression estimé soit en coupe instantanée soit sur séries chronologiques on apprécie la « performance » d'ensemble du modèle au regard de

son aptitude à reproduire la variabilité de Y. Dans le cadre plus complexe d'un modèle estimé sur données de panel, la performance du modèle peut être appréciée à l'aune de plusieurs indicateurs.

On se souvient <sup>16</sup> en effet que la variabilité totale TSS de Y (section I, paragraphe 3, B), c'est à dire la somme des carrés des écarts des  $Y_{it}$  par rapport à sa moyenne d'ensemble  $\bar{Y}$  peut se décomposer comme la somme de la variabilité interindividuelle BSS et de la variabilité intra individuelle WSS :

$$TSS_{YY} = WSS_{YY} + BSS_{YY}$$

avec :

$$TSS_{YY} = \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y})^2$$

$$WSS_{YY} = \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{io})^2$$

$$BSS_{YY} = T \sum_{i=1}^N (\bar{Y}_{io} - \bar{Y})^2$$

La performance du modèle étudié (quel qu'il soit au demeurant) peut être appréciée par rapport à son aptitude à expliquer chacune de ces trois variabilités ; il est alors possible de définir, pour chaque modèle, trois  $R^2$  :

. le  $R^2_{Total}$  traditionnel qui nous renseigne sur la part expliquée de la variabilité totale de Y

. le  $R^2_{within}$  qui nous renseigne sur la part expliquée de la variabilité intra-individuelle de Y

. le  $R^2_{between}$  qui nous renseigne sur la part expliquée de la variabilité inter-individuelle.

Bien sûr, chaque modèle (MCO-Pooled, Within et Between) est plus particulièrement dédié à l'explication de l'une de ces trois variabilités :

. le modèle MCO sur données empilées, par construction, est « calibré » pour

---

<sup>16</sup> section I, paragraphe 3, B

minimiser la somme des carrés des résidus entre Y observés et Y calculés. C'est le modèle qui expliquera le mieux la variabilité totale de Y ; c'est donc avec cette méthode que le  $R^2$  traditionnel sera le plus élevé

. quand on utilise le modèle WITHIN sur données centrées autour des moyennes individuelles, on cherche à expliquer la variabilité intra-individuelle de Y ; c'est donc avec cette méthode que le  $R^2_{within}$  sera le plus élevé

. quand on utilise le modèle BETWEEN sur données centrées autour des moyennes individuelles, on cherche à expliquer la variabilité intra-individuelle de Y ; c'est donc avec cette méthode que le  $R^2_{between}$  sera le plus élevé

De façon très pratique, comment doit-on s'y prendre pour calculer ces différents  $R^2$  ? On prendra comme exemple le cas du modèle sur données « empilées » (modèle *Pooled*)

- le calcul du  $R^2_{total}$  est tout à fait standard ; il n'y a donc pas lieu de s'y attarder
- le calcul du  $R^2_{within}$  se fait ainsi : on commence par calculer les écarts observés  $Y_{it} - \bar{Y}_{io}$  des  $Y_{it}$  à leurs moyennes individuelles ; on construit ensuite les écarts  $\hat{Y}_{it} - \bar{\hat{Y}}_{io}$  des Y calculés à leurs moyennes individuelles avec :

$$\hat{Y}_{it} = \sum_{j=1}^K \hat{\beta}_{j,pool} X_{jit} \quad \text{et} \quad \bar{\hat{Y}}_{io} = \frac{1}{T} \sum_{t=1}^T \hat{Y}_{it}$$

Le  $R^2_{within}$ , pour ce modèle, est alors calculé comme le carré du coefficient de corrélation entre les  $Y_{it} - \bar{Y}_{io}$  et les  $\hat{Y}_{it} - \bar{\hat{Y}}_{io}$ .

$$R^2_{within} = \left[ \frac{\text{Cov}\left((Y_{it} - \bar{Y}_{io}), (\hat{Y}_{it} - \bar{\hat{Y}}_{io})\right)}{s(Y_{it} - \bar{Y}_{io}) \times s(\hat{Y}_{it} - \bar{\hat{Y}}_{io})} \right]^2$$

- le calcul du  $R^2_{between}$  se fait ainsi : on commence par calculer les moyennes individuelles des Y observés et calculés. Le  $R^2_{between}$  est alors calculé comme le carré du coefficient de corrélation entre ces moyennes individuelles :

$$R^2_{between} = \left[ \frac{\text{Cov}(\bar{Y}_{io}, \hat{\bar{Y}}_{io})}{s(\bar{Y}_{io}) \times s(\hat{\bar{Y}}_{io})} \right]^2$$

En partant à présent du modèle *Within*, une précaution doit être prise pour le calcul des  $\hat{Y}_{it}$  qui serviront au calcul des trois  $R^2$ . Il convient en effet de ne pas introduire les effets fixes dans le calcul de ces  $\hat{Y}$  ... sauf à se retrouver avec un  $R^2_{between}$  égal à 1 puisque les effets fixes sont justement calculés de telle sorte que les moyennes individuelles observées et calculées soient identiques<sup>17</sup> ; dans ces conditions, il y aurait adéquation parfaite entre moyennes individuelles observées et moyennes individuelles calculées.

Exemple : On dispose de données de panel relatives au nombre de grèves dans le secteur industriel<sup>18</sup>. On note  $GREVE_{it}$  le nombre de jours de grève pour 1000 salariés du secteur industriel dans le pays  $i$  au cours de l'année  $t$ . On souhaite relier cette variable au taux de chômage ( $CHOM_{it}$ ) et au taux d'inflation ( $INFL_{it}$ ) du pays  $i$  tels qu'ils ont pu être observés au cours de l'année  $t$ . Le panel est constitué de 17 individus-pays et de 35 observations dans le temps échelonnées de 1951 à 1985. On estime les 5 modèles étudiés jusqu'ici (*Pool*, *LSDV*, *Within*, *Between* et *MCQG*) et, pour chaque modèle, on évalue sa performance au regard de la reproduction des différentes variabilités.

Les données et le programme utilisés sont respectivement :

- hurlin\_greves\_emp.wfl
- estimation\_panel\_data.prg (programme EViews)

<sup>17</sup> On peut s'en assurer sous Eviews 5.0 (page structurée) en comparant les propriétés statistiques des valeurs observées et calculées (*within*) de  $Y$ .

<sup>18</sup> Les données sont issues de <http://lib.stat.cmu.edu/datasets/>. L'exemple est issu de Ch. HURLIN : « L'économétrie des données de panel : modèles linéaires simples ». Document de Travail - Ecole Doctorale Edocif – Séminaire Méthodologique (2000).

MCO - Données Empilées				
Variable	Coefficient	Ect.type	T Stat	Prob
CHOM	27.437903	7.5399679	3.6389947	0.0002976
INFL	18.613561	4.9865259	3.7327713	0.0002077
Const	95.078706			
ESS	183837589	SE	557.25799	
R2 Total =	0.0529	R2 Intra =	0.0026	R2 Inter = 0.5929

Estimateur LSDV				
Variable	Coefficient	Ect.type	T Stat	Prob
CHOM	-21.596840	9.1915834	-2.3496321	0.0191286
INFL	16.272887	4.7565830	3.4211297	0.0006678
ESS	146958281	SE	505.10979	
R2 Total =	0.0004	R2 Intra =	0.0252	R2 Inter = 0.1969

Estimateur WITHIN				
Variable	Coefficient	Ect.type	T Stat	Prob
CHOM	-21.596840	9.1915834	-2.3496321	0.0191286
INFL	16.272887	4.7565830	3.4211297	0.0006678
ESS	146958281	SE	505.10979	
R2 Total =	0.0004	R2 Intra =	0.0252	R2 Inter = 0.1969

Estimateur BETWEEN				
Variable	Coefficient	Ect.type	T Stat	Prob
CHOM	80.954240	23.165030	3.4946745	0.0035728
INFL	59.388224	32.606827	1.8213433	0.0899879
Const	-341.54624			
ESS	503607.36	SE	189.66274	
R2 Total =	0.0528	R2 Intra =	0.0032	R2 Inter = 0.5933

Estimateur MCQG				
Variable	Coefficient	Ect.type	T Stat	Prob
CHOM	-12.081434	8.7862316	-1.3750417	0.1696541
INFL	16.424731	4.7266967	3.4748858	0.0005496
Const	248.62161			
ESS	192970787	SE	579.31112	
R2 Total =	0.0081	R2 Intra =	0.0234	R2 Inter = 0.0385

On notera que, conformément à ce qui a été dit précédemment, des différents modèles,

1. celui qui maximise la valeur du  $R^2_{total}$ , c'est le modèle MCO sur données empilées

2. celui qui maximise la valeur du  $R^2_{within}$ , c'est le modèle Within sur données en différences aux moyennes individuelles
3. celui qui maximise la valeur du  $R^2_{between}$ , c'est le modèle Between sur les moyennes individuelles.

Accessoirement, le programme fournit également deux autres informations :

- un test pour l'hypothèse d'homogénéité ( $H_0$  : « Homogénéité complète » *versus*  $H_A$  : « Présence d'effets fixes »)
- un test de Hausman pour l'hypothèse nulle d'exogénéité des variables explicatives par rapport aux erreurs spécifiques (le résultat du test détermine le choix d'un modèle à effets fixes ou à composantes d'erreur).

Test d'homogénéité (absence d'effets fixes)						
H0 : Homogénéité <i>versus</i>						
Ha : Présence d'effets fixes						
F calc	DDL	F* 5%	Conclusion	F* 1%	Conclusion	
9.03	16, 576	1.66	Rejet de H0	2.03	Rejet de H0	

Le résultat du test permet de rejeter l'hypothèse d'homogénéité interindividuelle de la relation (chômage, inflation) → grèves. Il faut donc privilégier un modèle tenant compte des spécificités individuelles.

Test d'Hausman						
H0 : EXOGENEITE DES Xj						
H calc	DDL	H* 5%	Conclusion	H* 1%	Conclusion	
13.92	2	5.99	Rejet de H0	9.21	Rejet de H0	

Le résultat du test est donc que l'hypothèse nulle d'exogénéité des variables explicatives doit être rejetée. Dans ce contexte, on sait que les estimateurs des MCQG sont biaisés. On privilégiera donc l'usage d'un estimateur de type Within qui a le mérite de demeurer sans biais dans un tel environnement.

## **Section VI : Hétéroscédasticité et autocorrélation : détection et correction**

On a supposé jusqu'ici que les erreurs du modèle étudié présentaient quelques bonnes propriétés. Notamment :

- 1/ elles sont supposées homoscédastiques dans la dimension interindividuelle
- 2/ elles sont supposées non corrélées dans la dimension chronologique

Plus formellement, ceci peut être exprimé comme :

- $E(\varepsilon_{it}^2 | X) = \sigma_\varepsilon^2 \quad \forall i, t$  : l'erreur  $\varepsilon$  est homoscédastique
- $E(\varepsilon_{it} \varepsilon_{js} | X) = 0 \quad \forall t, s, i, j, i \neq j$  : les erreurs  $\varepsilon$  ne sont ni autocorrélées ni corrélées entre des individus différents

On pose ici deux questions :

- celle de la détection d'une possible violation de l'une ou l'autre de ces deux hypothèses
- celle de la prise en compte d'un phénomène d'autocorrélation ou d'hétéroscédasticité.

On traitera de cette double question dans le cadre théorique d'un modèle à effets fixes. Cette question n'est pas vaine : comme on le sait déjà (cft. cours de Licence) une autocorrélation ou une hétéroscédasticité non prises en compte (utilisation par exemple de l'estimateur des MCO) ont pour conséquence le caractère biaisé et non convergent des estimateurs de la variance de l'erreur et, par voie de conséquence, des estimateurs des variances et covariances des estimateurs des différents coefficients.

Illustration de ce résultat connu : *GENR\_PANEL\_AUTOCORRELATION.PRG*

A 100 reprises on génère un panel de données pour lesquelles le terme d'erreur qui affecte la relation  $X - Y$  est autocorrélé :

$$\varepsilon_{it} = \rho \varepsilon_{i, t-1} + v_t$$

où on suppose que les  $v_{it}$  sont IIDN(0,1). La variance théorique des  $\varepsilon_{it}$ , dans ce contexte, est donc telle que :

$$\sigma_\varepsilon^2 = \rho^2 \sigma_\varepsilon^2 + \sigma_v^2 = \frac{1}{1-\rho^2}$$

On expérimente le fait que, quoique l'estimateur *within* du coefficient associé à  $X$  soit sans biais, celui de la variance de l'erreur  $\varepsilon_{it}$  est, lui, biaisé (on pourrait éventuellement montrer qu'il est aussi non convergent). On comparera notamment la moyenne des valeurs estimées de  $\sigma_\varepsilon$  à sa valeur théorique (*ectyp\_theorique*).

**Paragraphe 1 : violation des hypothèses : la détection**

La détection des phénomènes d'autocorrélation ou d'hétéroscédasticité procède, à quelques difficultés supplémentaires près, de la même méthodologie que celle qui est mise en oeuvre dans le cas d'échantillons plus « conventionnels ».

### A - Le test d'absence d'autocorrélation

Soit le modèle  $Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$ . On veut éprouver le bien fondé de l'hypothèse nulle d'absence d'autocorrélation. Dire qu'il n'y a, dans le modèle étudié, aucun phénomène d'autocorrélation, c'est reconnaître que :

$$E \varepsilon_{is} \varepsilon_{jt} = 0 \quad \forall i, j, s, t, s \neq t$$

A dire vrai, le test d'hypothèse traditionnellement mis en oeuvre se borne à éprouver l'hypothèse nulle d'absence d'autocorrélation d'ordre 1 pour quelque individu  $i$  que ce soit. En d'autres termes, il s'agit de tester :

$$H_0 : E \varepsilon_{it} \varepsilon_{i,t-1} = 0 \quad \forall i, t$$

La statistique de test utilisée est elle-même une forme étendue de la statistique de Durbin Watson, proposée par Bhargava, Franzini et Narendranathan (1982)<sup>19</sup>. La statistique DWBFN est calculée comme une statistique de DW usuelle, à ceci près que les résidus utilisés sont les résidus issus de l'estimation du modèle à effets fixes et centrés autour de leurs moyennes individuelles  $\tilde{\varepsilon}_{it} = \hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_{i0}$  où, il faut le préciser, les  $\hat{\varepsilon}_{it}$  sont les résidus issus de l'estimation du modèle :

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_1 X_{1it} - \dots - \hat{\beta}_k X_{kit}$$

$$DWBFN = \frac{\sum_{i=1}^N \sum_{t=2}^T (\tilde{\varepsilon}_{it} - \tilde{\varepsilon}_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \tilde{\varepsilon}_{it}^2}$$

La statistique ainsi calculée est généralement produite par la plupart des logiciels qui rendent possible l'estimation des modèles sur la base de données de panel.

---

<sup>19</sup> Bhargava, Franzini et Narendranathan : « Serial correlation and fixed effect model ». Review of economic studies, 49, 1982, pp.533-549.

On reproduit ci-dessous le tableau des valeurs critiques telles qu'elles ont été produites par BFN (1982) pour un nombre d'individus relativement important :

		N = 100		N = 500		N = 1000	
		$d_{\min}$	$d_{\max}$	$d_{\min}$	$d_{\max}$	$d_{\min}$	$d_{\max}$
T = 6	K = 3	1.859	1.880	1.939	1.943	1.957	1.959
	K = 9	1.839	1.902	1.935	1.947	1.954	1.961
T = 10	K = 3	1.891	1.904	1.952	1.954	1.967	1.968
	K = 9	1.878	1.916	1.949	1.957	1.965	1.970

Pour un nombre d'observations très important  $N = 1000$ , on voit qu'il est possible de rejeter l'hypothèse nulle d'absence d'autocorrélation si la valeur calculée de la statistique est inférieure à 1.96.

### B - Un test d'homoscédasticité

Ici encore le test de l'hypothèse nulle d'homoscédasticité est une extension au cas des données de panel du test bien connu de White. Il suffit simplement, à l'issue de l'estimation, sur données de panel, du modèle principal :

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \varepsilon_{it}$$

de procéder à un ajustement auxiliaire, toujours sur données de panel, du type :

$$\hat{\varepsilon}_{it}^2 = \lambda_0 + \lambda_1 X_{1it}^2 + \dots + \lambda_k X_{kit}^2 + v_{it} \quad ^{20}$$

où, évidemment, les  $\hat{\varepsilon}_{it}$  sont les résidus issus de l'ajustement principal.

Sous l'hypothèse nulle d'homoscédasticité, la statistique  $W = N.T. R_{\text{auxiliaire}}^2$  est réputée être distribuée selon une loi du Chi Deux à  $k$  degrés de liberté.

Exemple : strikes\_hurlin.wf1

On a repris les données produites par Ch. Hurlin, relatives à l'explication du nombre de jours de grèves en fonction du taux de chômage et du taux d'inflation dans un pays  $i$  à une date  $t$ . Rappelons que l'estimation est conduite sur la base d'un panel de 17 pays pour lesquels on dispose de 35 observations dans le temps (le panel est cylindré).

<sup>20</sup> Le cas échéant, on peut introduire, de façon complémentaire, les variables en niveaux et les produits croisés de celles-ci.

L'estimation *within* du modèle est reproduite ci-dessous (on a négligé de faire apparaître les effets fixes) :

Dependent Variable: GREVES?  
 Method: Pooled Least Squares  
 Date: 02/16/04 Time: 15:37  
 Sample: 1951 1985  
 Included observations: 35  
 Cross-sections included: 17  
 Total pool (balanced) observations: 595

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INFL?	16.27289	4.756583	3.421130	0.0007
CHOM?	-21.59684	9.191583	-2.349632	0.0191

Munis des résidus de cet ajustement principal on a ensuite procédé à l'ajustement auxiliaire :

Dependent Variable: RESID?^2  
 Sample: 1951 1985  
 Included observations: 35  
 Cross-sections included: 17  
 Total pool (balanced) observations: 595

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INFL?^2	4.581650	1111.859	0.004121	0.9967
CHOM?^2	-1781.656	2780.008	-0.640882	0.5218
C	286068.0	128204.0	2.231350	0.0260

R-squared 0.000701 Mean dependent var 246988.7

La statistique W est :

$$W = 595 * 0.000701 = 0.417$$

La valeur prise par la statistique permet de conclure à l'impossibilité de rejeter  $H_0$ .

## Paragraphe 2 : La violation des hypothèses : la prise en compte

On suppose à présent que l'une ou l'autre des deux hypothèses d'absence d'autocorrélation et d'homoscédasticité a été rejetée. Comment peut on en tenir compte.

### A - La correction de l'autocorrélation

Le test de Durbin a mis en évidence, pour le modèle étudié, la réalité d'un phénomène d'autocorrélation des résidus. On peut donner deux interprétations de ce résultat qui débouchent sur un traitement spécifique de cette autocorrélation.

- la mise en évidence d'un phénomène d'autocorrélation des résidus doit nécessairement amener l'utilisateur à s'interroger, avant toute chose, sur l'éventuelle omission d'une variable essentielle qui serait elle-même autocorrélée. Une telle omission, au delà de l'autocorrélation des résidus qu'elle suscite, est susceptible d'engendrer aussi un biais sur l'estimation des coefficients associés aux différentes variables explicatives
- si l'éventualité d'une autocorrélation des résidus induite par l'omission d'une variable essentielle peut être écartée, il faut alors songer à un traitement spécifique de la probable autocorrélation des erreurs. Il est alors toujours possible de mettre en oeuvre une méthode de correction du type Cochrane – Orcutt qui consiste à préférer au modèle  $Y_{it} = \alpha_i + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \varepsilon_{it}$  pour lequel l'erreur  $\varepsilon_{it}$  est, intraindividuellement, autocorrélée une spécification du type :

$$Y_{it} = \alpha_i (1-\rho) + \beta_1 (X_{1it} - \rho X_{1i,t-1}) + \dots + \beta_k (X_{kit} - \rho X_{ki,t-1}) + \varepsilon_{it} - \rho \varepsilon_{i,t-1}$$

pour laquelle le terme d'erreur  $\varepsilon_{it} - \rho \varepsilon_{i,t-1}$  ne serait pas autocorrélé (sous réserve que le paramètre  $\rho$  ait été correctement estimé).

Exercice : `genr_panel_autocorrelation.prg`.

*Ce programme fabrique à 100 reprises un panel de données pour lequel l'erreur du modèle est autocorrélée. On montre que :*

- *l'estimateur within demeure un estimateur sans biais de beta (à la condition, bien sûr, que l'erreur ne soit pas aussi corrélée avec la variable explicative X !)*
- *que, dans ce contexte, la correction à la Cochrane Orcutt induit un estimateur de beta qui est, lui, plus efficace.*

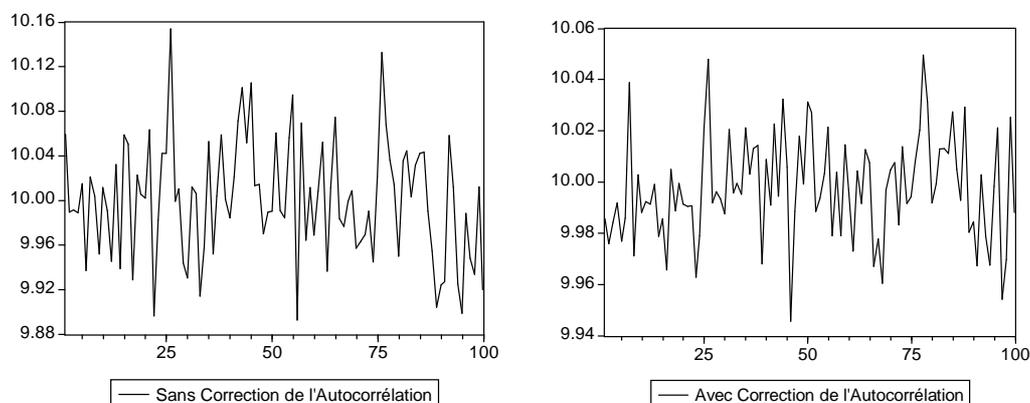
La relation X – Y est du type :

$$Y_{it} = 10 + 10 X_{it} + \varepsilon_{it}$$

avec  $\varepsilon_{it} = 0.95 \varepsilon_{i,t-1} + v_t$  où les  $v_t$  sont iid  $N(0,1)$ .

Le graphique ci-dessous reproduit les estimations qui ont été obtenues du coefficient  $\beta$  associé à la variable X pour chaque panel de données en utilisant d'abord

l'estimateur *within* sans tenir compte du caractère autocorrélé des erreurs puis en prenant en compte cette autocorrélation. On expérimente que les estimations induites par la méthode Cochrane Orcutt présentent un degré d'efficacité supérieur.



## B - La correction de l'hétéroscédasticité

Dans le cas pour lequel la statistique de White révélerait la présence d'un phénomène d'hétéroscédasticité interindividuelle on met traditionnellement en oeuvre une méthode de correction inspirée de la méthode des MC Pondérés. On admet, par hypothèse, que la variance de l'erreur  $\varepsilon_{it}$  est homoscédastique pour un même individu (homoscédasticité intraindividuelle) mais hétéroscédastique d'un individu à l'autre (hétéroscédasticité interindividuelle) :

$$\text{Var } \varepsilon_{it} = \sigma_i^2 \quad \forall i, t$$

Dans ce contexte on peut essayer de corriger l'hétéroscédasticité en pondérant toutes les observations relatives à un même individu par un coefficient spécifique à cet individu. Il reste, évidemment, à déterminer la valeur de ces coefficients de pondération spécifiques. Un moyen simple consiste à estimer, dans un premier temps, la relation  $X - Y$  sur données empilées et à calculer les variances intra individuelles des résidus issus de cet ajustement préliminaire. Les écarts type intra individuels ainsi calculés seront ensuite utilisés pour pondérer les observations relatives aux différents individus.

Exemple : `genr_panel_heteroscedasticite.prg`. On montre que, en présence d'une hétéroscédasticité interindividuelle des erreurs, la méthode des MCP induit des estimateurs plus efficaces.

A 100 reprises on construit un panel de données pour 20 individus (numérotés  $i =$

1, ..., 20) et 30 observations dans le temps avec :

- effets fixes : on suppose que l'effet fixe  $\alpha_i$  relatif au  $i^{\text{e}}$  individu est égal à  $i$ .
- hétéroscédasticité des erreurs : on suppose que l'écart type intra individuel de l'erreur associée au  $i^{\text{e}}$  individu est lié à l'ordre de grandeur de  $X$ , lui-même variable d'un individu à l'autre

$$\varepsilon_{it} = 10 \bar{X}_{i0} \times v_{it} \quad \text{avec } v_{it} \text{ iid } N(0,1)$$

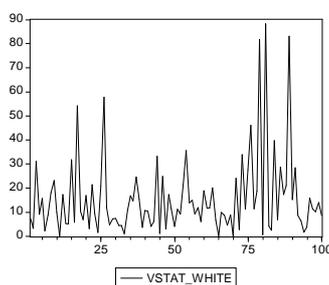
- une relation  $X - Y$  du type :

$$Y_{it} = \alpha_i + 10 X_{it} + \varepsilon_{it}$$

A chaque itération,

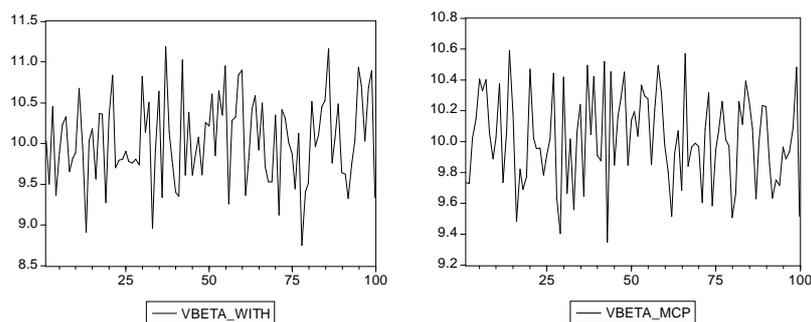
1. on estime pour commencer un modèle à effet fixe sans correction de l'hétéroscédasticité
2. on calcule la statistique de White pour l'hypothèse nulle d'homoscédasticité
3. on estime un modèle à effets fixes en ayant soin de pondérer les variables du modèle par les écarts type intra individuels tels qu'ils peuvent être calculés.

On expérimente que la statistique de White, dans ce cas d'une hétéroscédasticité liée à l'ordre de grandeur de  $X$ , détecte correctement le phénomène <sup>21</sup> d'une part et que, d'autre part, la pondération des variables induit des estimateurs de  $\beta$  plus efficaces.




---

<sup>21</sup>  $\chi^2_{5\%}(1) = 3.85$



Dans le cas pour lequel le phénomène d'hétéroscédasticité s'exprime dans les deux dimensions, c'est à dire le cas pour lequel :

$$\text{Var } \varepsilon_{it} = \sigma_{it}^2$$

il est possible de mettre en oeuvre une autre technique pour l'estimation des écarts type  $\sigma_{it}$ .

1/ On commence par estimer un modèle à effets fixes sans pondération des variables utilisées

2/ Munis des coefficients estimés, on calcule les résidus de cette estimation :

$$\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta} X_{it}$$

3/ Sous l'hypothèse d'une hétéroscédasticité liée à l'ordre de grandeur de l'une des variables du modèle (par exemple X) selon la relation fonctionnelle :

$$\sigma_{it}^2 = f(\lambda_0 + \lambda_1 X_{it} + \lambda_2 X_{it}^2)$$

on obtient une estimation de l'élément  $\lambda_0 + \lambda_1 X_{it} + \lambda_2 X_{it}^2$  en régressant  $f^{-1}(\hat{\varepsilon}_{it}^2)$  sur X et son carré. On obtient par ce truchement une estimation de  $\sigma_{it}^2$  qu'on peut ensuite utiliser pour pondérer les variables du modèle.

## Section VII : Les modèles dynamiques

On s'intéresse ici à une classe particulière de modèles de données de panel. Ce sont les modèles qui font dépendre la valeur de  $Y$  pour le  $i^{\text{e}}$  individu à la date  $t$  non seulement des valeurs prises par les variables  $X_j$  pour ce même individu mais aussi – et c'est là que réside la source des difficultés associées à l'estimation de ces modèles – des valeurs retardées de la variable expliquée elle-même. Le modèle étudié ici peut donc être spécifié sous la forme :

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Y_{i,t-1} + \varepsilon_{it}$$

si on s'en tient à une seule variable exogène.

On pourrait considérer la valeur retardée de  $Y$  comme une variable explicative tout à fait ordinaire et procéder par conséquent à l'estimation des coefficients du modèle en mettant en oeuvre les techniques qui ont été présentées jusqu'ici (par exemple en utilisant l'estimateur *within*). Malheureusement, on va montrer que, précisément,  $Y_{i,t-1}$  n'est pas une variable explicative « comme une autre » et que l'estimateur (*within* notamment) du coefficient qui lui est attaché est nécessairement biaisé et non convergent. Dès lors, d'autres techniques d'estimation doivent être envisagées qui font appel à l'usage de variables instrumentales.

### Paragraphe 1 : Les estimateurs traditionnels sont biaisés et non convergents

On met ici en évidence le fait que les estimateurs usuels développés jusqu'ici sont biaisés et non convergents dès lors qu'ils sont utilisés dans le cadre d'un modèle dynamique. Pour ne pas alourdir inutilement l'exposé, on se contentera de mettre en évidence cette propriété fâcheuse dans le cas de l'estimateur *within*.

On part du modèle :

$$/1/ \quad Y_{it} = \alpha_i + \beta X_{it} + \gamma Y_{i,t-1} + \varepsilon_{it}$$

Il implique :

$$/2/ \quad \bar{Y}_{io} = \alpha_i + \beta \bar{X}_{io} + \gamma \bar{Y}_{io(-1)} + \bar{\varepsilon}_{io}$$

où  $\bar{Y}_{io(-1)}$  est la moyenne intraindividuelle des  $Y_{it}$  calculée sur les  $T-1$  premières observations relatives à l'individu  $i$ .

L'estimateur *within* consiste, on le rappelle, à appliquer l'estimateur des MCO au modèle /3/ obtenu en retranchant le contenu de l'équation /2/ à celui de l'équation /1/ :

$$/3/ \quad (Y_{it} - \bar{Y}_{io}) = \beta (X_{it} - \bar{X}_{io}) + \gamma (Y_{i,t-1} - \bar{Y}_{io(-1)}) + (\varepsilon_{it} - \bar{\varepsilon}_{io})$$

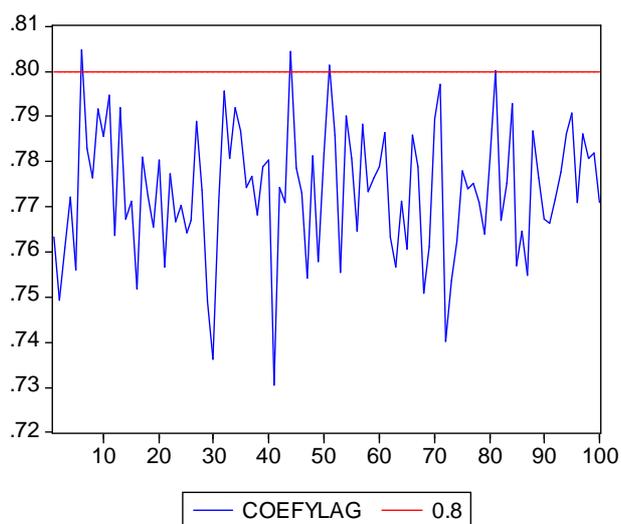
Le problème majeur qui se pose avec cette nouvelle équation est que sa spécification introduit *de facto* un phénomène de corrélation entre la nouvelle erreur du modèle  $(\varepsilon_{it} - \bar{\varepsilon}_{io})$  et la variable  $(Y_{i,t-1} - \bar{Y}_{io(-1)})$ . L'explication d'un tel phénomène est simple à comprendre : le calcul de  $\bar{Y}_{io(-1)}$  fait intervenir implicitement les T-1 premières erreurs  $\varepsilon_{it}$ ... de telle sorte que, par construction,  $(Y_{i,t-1} - \bar{Y}_{io(-1)})$  et  $(\varepsilon_{it} - \bar{\varepsilon}_{io})$  sont nécessairement corrélées. Les conséquences d'un tel phénomène sont bien connues : l'estimateur  $\hat{\gamma}$  du coefficient associé à la variable endogène retardée est biaisé et non convergent (cf. cours Econométrie III). On expérimente cette propriété à l'aide d'une simulation.

Illustration : `genr_panel_dynamique.prg` (alternativement on peut aussi utiliser `panel_dynamique_biais.prg`)

A 100 reprises on a créé un panel de données de 24 individus et 32 observations dans le temps. On introduit une relation dynamique du type :

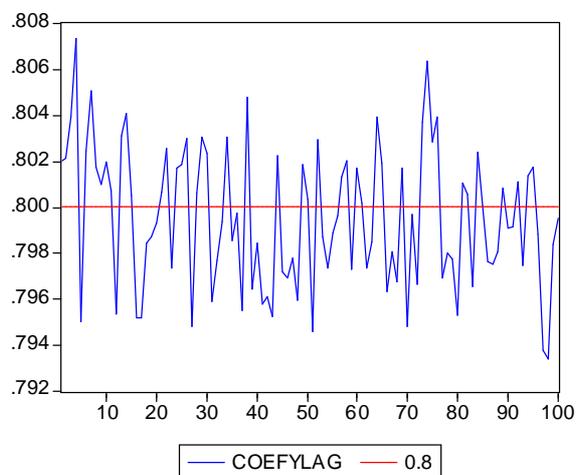
$$Y_{it} = \alpha_i + X_{it} + 0.8 Y_{i,t-1} + \varepsilon_{it}$$

Les réalisations du terme d'erreur  $\varepsilon_{it}$  sont tirées de manière purement aléatoire et indépendante d'une distribution normale standard. Les effets fixes relatifs aux différents individus sont tirés d'une distribution uniforme dont les modalités sont des entiers compris entre 0 et 100. On expérimente le fait que l'estimateur *within* est biaisé :



En faisant varier la taille de l'échantillon dans l'une puis dans l'autre des deux dimensions on peut vérifier que cet estimateur biaisé est

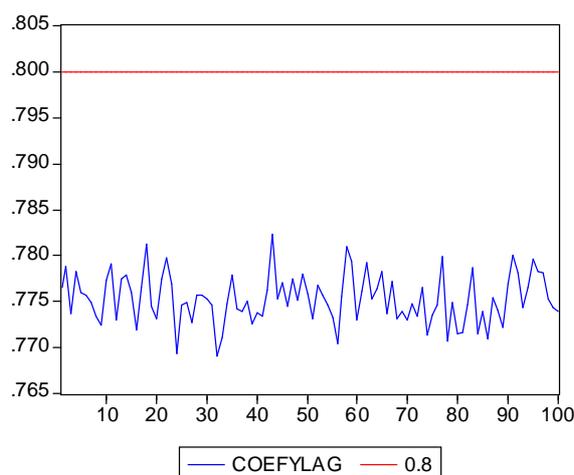
- . convergent dans la dimension temporelle
- . non convergent dans la dimension individuelle<sup>22</sup> :



$N = 24 ; T = 1000$

---

<sup>22</sup> Sur ce point, on pourra consulter M. Verbeek, *op.cit.* pp. 327-329



$N = 1000 ; T = 32$

On montrerait sur la base du même argument que l'estimateur  $\hat{\gamma}_{between}$  et l'estimateur  $\hat{\gamma}_{DI}$  du modèle récrit en différence première sont, tous les deux, biaisés. On note en effet que, pour le modèle en différence première :

$$\Delta Y_{it} = \gamma \Delta Y_{i,t-1} + \beta \Delta X_{it} + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

Le terme  $\Delta Y_{i,t-1} = Y_{i,t-1} - Y_{i,t-2}$  et le terme d'erreur  $(\varepsilon_{it} - \varepsilon_{i,t-1})$  ont un élément commun,  $\varepsilon_{i,t-1}$  en l'occurrence.

## Paragraphe 2 : L'estimateur des Variables Instrumentales (Anderson et Hsiao - 1981)

On vient de voir que l'estimateur *within* et l'estimateur des MCO appliqué aux variables en différences premières sont biaisés. Récrivons justement le modèle en différences premières (on supposera, pour la simplicité, que le modèle est purement autorégressif) :

$$Y_{it} - Y_{i,t-1} = \gamma (Y_{i,t-1} - Y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

On vient de voir que l'estimateur des MCO appliqué à ce modèle est biaisé par construction puisque la variable explicative  $\Delta Y_{t-1}$  est corrélée avec le nouveau terme d'erreur  $\Delta \varepsilon_{it}$  ( $\varepsilon_{i,t-1}$  est une des composantes de  $Y_{i,t-1}$ ). Pour autant,  $Y_{i,t-2}$ , lui, n'est théoriquement pas corrélé avec  $\Delta \varepsilon_{it}$  tout en demeurant fortement corrélée avec

$\Delta Y_{t-1}$ <sup>23</sup>.  $Y_{t-2}$  présente donc à ce titre toutes les qualités attendues d'un bon instrument. Dans ce contexte, la logique veut qu'on mette en oeuvre la méthode des variables instrumentales. Rappelons que, si  $Z_t$  est un instrument pour la variable explicative  $X_t$  corrélée avec le terme d'erreur, l'estimateur VI des variables instrumentales, pour le modèle  $Y_t = \alpha + \beta X_t + \varepsilon_t$  est donné par :

$$\hat{\beta}_{VI} = \frac{\sum_{t=1}^T Z_t y_t}{\sum_{t=1}^T Z_t x_t}$$

Pour le modèle dynamique étudié, l'estimateur des variables instrumentales (Anderson et Hsiao 1981) a pour expression :

$$\hat{\gamma}_{VI} = \frac{\sum_{i=1}^N \sum_{t=2}^T y_{it-2} (Y_{it} - Y_{it-1})}{\sum_{i=1}^N \sum_{t=2}^T y_{it-2} (Y_{it-1} - Y_{it-2})}$$

où  $y_{it-2}$  est la valeur centrée de  $Y_{it-2}$  (on admet que  $Y_{it} - Y_{it-1}$  est d'espérance nulle).

Alternativement, Anderson et Hsiao ont aussi proposé d'utiliser :

$$\hat{\gamma}_{VI} = \frac{\sum_{i=1}^N \sum_{t=3}^T (Y_{it-2} - Y_{it-3})(Y_{it} - Y_{it-1})}{\sum_{i=1}^N \sum_{t=3}^T (Y_{it-2} - Y_{it-3})(Y_{it-1} - Y_{it-2})}$$

qui présente toutefois l'inconvénient de conduire à la perte d'un degré de liberté.

**Illustration** : `genr_panel_dynamique_stacked.prg`

A 100 reprises, on construit un panel de données et une relation :

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Y_{it-1} + \varepsilon_{it}$$

avec  $\alpha_i$  tiré aléatoirement, pour chaque individu, d'une distribution uniforme [0, 100].

$$\beta = 1 \quad \gamma = 0.8$$

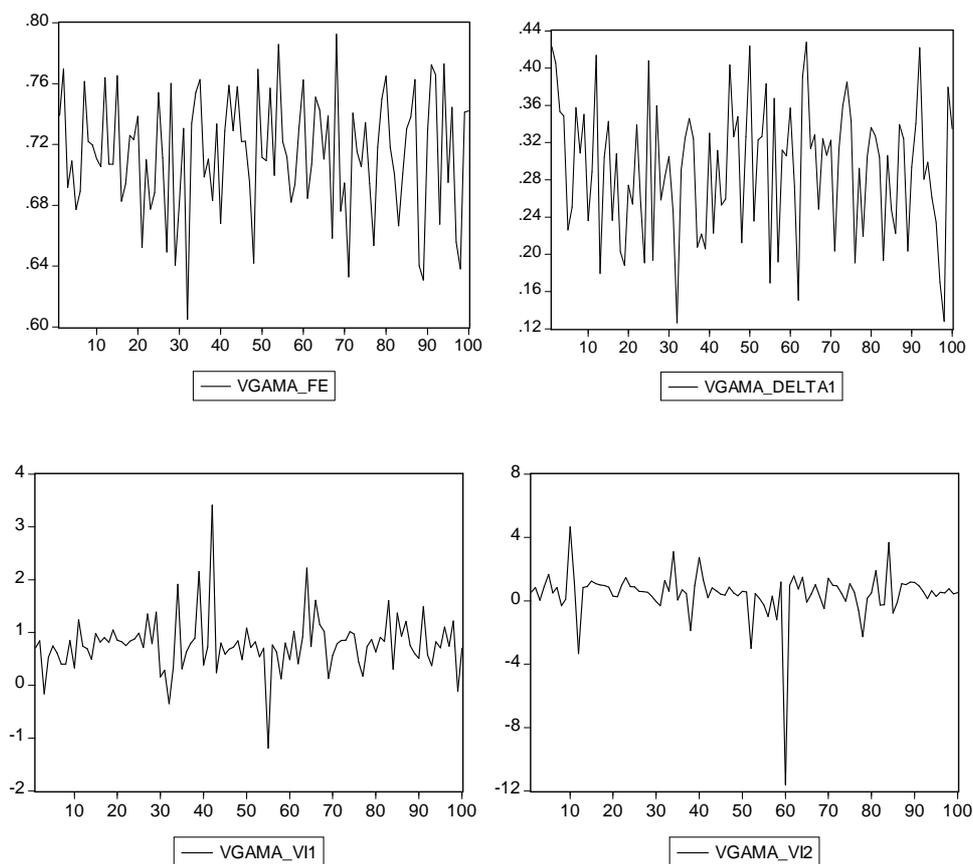
---

<sup>23</sup> ... à la condition évidemment que l'erreur  $\varepsilon_{it}$  ne soit pas elle-même autocorrélée.

On cherchera, à chaque itération, à retrouver les valeurs de ces coefficients en estimant les coefficients d'un modèle dynamique  $Y_{it} = \alpha_i + \beta X_{it} + \gamma Y_{i,t-1} + \varepsilon_{it}$ . A cet effet, on utilisera

- . un simple modèle à effets fixes
- . un modèle en différences premières
- . la méthode des variables instrumentales (appliquée au modèle en différence) avec  $Y_{i,t-2}$  comme instrument
- . la méthode des variables instrumentales avec  $\Delta Y_{i,t-2}$  comme instrument.

Les résultats obtenus peuvent être appréciés au travers des graphes représentatifs des estimations qui ont été obtenues du coefficient  $\gamma$  attaché à  $Y_{i,t-1}$  :



	Moyenne des $\hat{\gamma}$
Effet Fixe	0.71

Différences	0.29
Variables instrumentales 1	0.77
Variables instrumentales 2	0.43

### Paragraphe 3 : l'estimateur des moments généralisés (Arellano et Bond – 1991)

On vient de voir, en nous plaçant dans le cadre analytique simplifié d'un modèle purement autorégressif, que la méthode des variables instrumentales consiste à retenir  $Y_{i,t-2}$  comme instrument à la place de  $\Delta Y_{i,t-1}$  dans le modèle récrit en différence première :

$$Y_{it} - Y_{i,t-1} = \gamma(Y_{i,t-1} - Y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

$$\Delta Y_{it} = \gamma \Delta Y_{i,t-1} + \Delta \varepsilon_{it}$$

On sait que, à la condition bien sûr que  $Y_{i,t-2}$  soit effectivement non corrélé avec l'erreur  $\Delta \varepsilon_{it}$ , l'estimateur des variables instrumentales est convergent. Est-ce pour autant qu'il est efficient ? Arellano et Bond (1991) montrent précisément qu'un gain d'efficience peut être obtenu en exploitant, pour chaque observation, tous les instruments disponibles... opportunité qui a été négligée jusqu'ici et, notamment, comme on va le voir dans un instant, par la méthode des variables instrumentales qui vient d'être présentée. On se limitera ici à une présentation simplifiée de la méthode de Arellano et Bond dans le cadre du modèle autorégressif pur<sup>24</sup>

#### A - L'estimateur des moments généralisés : le cas d'un modèle purement autorégressif

Soit donc ici le modèle :

$$/1/ \quad Y_{it} = \gamma Y_{i,t-1} + \alpha_i + \varepsilon_{it}$$

qu'on récrit, comme on l'a déjà fait, en vue d'éliminer les effets fixes, en différence première :

---

<sup>24</sup> Pour une généralisation au cas pour lequel le modèle serait augmenté d'un certain nombre de variables exogènes on pourra consulter B.H.BALTAGI pp. 134 – 136.

$$/2/ \quad \Delta Y_{it} = \gamma \Delta Y_{i,t-1} + \Delta \varepsilon_{it}$$

On suppose qu'on dispose d'un échantillon  $E_T$  de  $T$  observations indicées  $t = 1, 2, \dots, T$ . Dans ce contexte, la première réalisation observable du modèle /2/ spécifié en différence première est :

$$Y_{i3} - Y_{i2} = \gamma (Y_{i2} - Y_{i1}) + \varepsilon_{i3} - \varepsilon_{i2}$$

La dernière réalisation observable (la  $T-2^e$ ) est :

$$Y_{iT} - Y_{i,T-1} = \gamma (Y_{i,T-1} - Y_{i,T-2}) + \varepsilon_{iT} - \varepsilon_{i,T-1}$$

On fait sur les erreurs  $\varepsilon_{it}$  les hypothèses habituelles :

$$E(\varepsilon_{it}) = 0 \quad \forall i, t$$

$$E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 \quad \forall i, t$$

$$E(\varepsilon_{is} \varepsilon_{it}) = 0 \quad \forall i, s, t, s \neq t$$

Les erreurs  $\Delta \varepsilon_{it}$  du modèle en différence première sont donc telles que :

$$E(\Delta \varepsilon_{it}) = 0 \quad \forall i, t$$

$$E(\Delta \varepsilon_{it}^2) = 2 \sigma_\varepsilon^2 \quad \forall i, t$$

$$E(\Delta \varepsilon_{it} \Delta \varepsilon_{i,t-1}) = -\sigma_\varepsilon^2 \quad \forall i, t$$

$$E(\Delta \varepsilon_{it} \Delta \varepsilon_{i,t-k}) = 0 \quad \forall i, t, k > 1$$

La matrice des variances – covariances des erreurs  $\Delta \varepsilon_{it}$  du modèle en différences premières pour le  $i^e$  individu est donc :

$$E(\Delta \varepsilon_i \Delta \varepsilon_i') = \sigma_\varepsilon^2 G$$

où la matrice  $G$ , de dimension  $T-2 \times T-2$ , est telle que :

$$G = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$

Si on écrit le modèle sous forme matricielle :

$$/3/ \quad \Delta Y = \gamma \Delta Y_{-1} + \Delta \varepsilon$$

avec  $\Delta \varepsilon' = \Delta \varepsilon_{13}, \Delta \varepsilon_{14}, \dots, \Delta \varepsilon_{1T}, \dots, \Delta \varepsilon_{N3}, \Delta \varepsilon_{N4}, \dots, \Delta \varepsilon_{NT}$

la matrice des variances – covariances des erreurs pour l'ensemble des individus est de la forme :

$$E(\Delta \varepsilon \Delta \varepsilon') = \sigma_\varepsilon^2 I_N \otimes G = \sigma_\varepsilon^2 \Omega_{N(T-2) \times N(T-2)}$$

où  $\otimes$ , on le rappelle, est le produit de Kronecker.

Lorsqu'il s'agit, pour la première réalisation du modèle en différence,

$$Y_{i3} - Y_{i2} = \gamma(Y_{i2} - Y_{i1}) + \varepsilon_{i3} - \varepsilon_{i2}$$

de définir un instrument pouvant se substituer à la variable explicative  $(Y_{i2} - Y_{i1})$ , corrélée avec le terme d'erreur  $(\varepsilon_{i3} - \varepsilon_{i2})$ , à vrai dire, le seul instrument sur lequel on puisse s'appuyer est  $Y_{i1}$ . C'est cet instrument qu'on pourrait qualifier d'instrument de premier plan qui avait été utilisé implicitement par l'estimateur des variables instrumentales. Dès lors qu'il s'agit de la deuxième réalisation disponible

$$Y_{i4} - Y_{i3} = \gamma(Y_{i3} - Y_{i2}) + \varepsilon_{i3} - \varepsilon_{i2}$$

deux instruments sont offerts qui peuvent se substituer à  $(Y_{i3} - Y_{i2})$  : l'instrument de premier plan  $Y_{i2}$  (déjà utilisé implicitement avec la méthode des variables instrumentales) mais aussi un instrument de second plan,  $Y_{i1}$ , qui avait été délaissé jusqu'ici. *A fortiori*, lorsqu'il s'agit de définir les instruments utilisables pour la T-2<sup>e</sup> et dernière réalisation du modèle :

$$Y_{iT} - Y_{iT-1} = \gamma(Y_{iT-1} - Y_{iT-2}) + \varepsilon_{iT} - \varepsilon_{iT-1}$$

on peut alors compter sur

- un instrument de premier plan :  $Y_{iT-2}$
- T-3 instruments de second plan :  $Y_{iT-3}, Y_{iT-4}, \dots, Y_{i1}$

soit, au total, T-2 instruments.

On comprend qu'il peut être dommageable, en termes d'efficience, de ne pas tenir

compte de l'information – délaissée jusqu'ici - contenue dans les instruments de second plan. La prise en considération de cette information passe par la construction d'une matrice  $Z_i$  d'instruments de premier et de second plans<sup>25</sup> pour l'individu  $i$  :

$$Z_i = \begin{bmatrix} [Y_{i1}] & 0 & \dots & 0 \\ 0 & [Y_{i1}, Y_{i2}] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & [Y_{i1}, Y_{i2}, \dots, Y_{iT-2}] \end{bmatrix}$$

de dimension  $T-2 \times (1 + 2 + \dots + T-2)$ . Au niveau agrégé de l'ensemble des individus, la matrice  $Z$  des instruments est donnée par :

$$Z = \begin{bmatrix} [Z_1] \\ [Z_2] \\ \vdots \\ [Z_N] \end{bmatrix} \text{ de dimension } N(T-2) \times (1 + 2 + \dots + T-2)$$

On doit bien noter que les composantes de cette matrice  $Z$  doivent, par nature, vérifier les conditions suivantes dites « conditions sur les moments » :

$$E(Z' \Delta \varepsilon) = 0$$

puisque ces composantes sont réputées être des « instruments » pour les variables qu'elles représentent.

Rien n'interdit de prémultiplier chacun des éléments du modèle /3/ par cette matrice des instruments de telle sorte que :

$$/4/ \quad Z' \Delta Y = \gamma Z' \Delta Y_{-1} + Z' \Delta \varepsilon$$

La matrice des variances – covariances du modèle ainsi récrit a pour contenu :

---

<sup>25</sup> On peut rapprocher cette matrice de celle qui avait été implicitement utilisée avec la méthode des variables instrumentales :

$$Z_i = \begin{bmatrix} Y_{i1} & 0 & \dots & 0 \\ 0 & Y_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Y_{iT-2} \end{bmatrix}$$

$$E(Z' \Delta \varepsilon \Delta \varepsilon' Z) = \sigma_\varepsilon^2 Z' \Omega Z$$

Quel bénéfice peut-on retirer d'une telle transformation du modèle sérial /3/ ? Ce qu'il faut bien comprendre (et ce qui a déjà été mis en évidence) c'est que l'estimateur MCO  $\hat{\gamma}$  du modèle en différence première est, par construction, biaisé pour N grand et T fini. En revanche, l'estimateur MCO  $\hat{\gamma}$  du modèle /4/ pour lequel les différentes composantes du modèle ont été prémultipliées par  $Z'$  présente une propriété intéressante. On note en effet que cet estimateur défini comme :

$$\begin{aligned} \hat{\gamma}_{mco} &= (\Delta Y_{-1}' Z Z' \Delta Y_{-1})^{-1} \Delta Y_{-1}' Z Z' \Delta Y \\ &= (\Delta Y_{-1}' Z Z' \Delta Y_{-1})^{-1} \Delta Y_{-1}' Z Z' (\delta \Delta Y_{-1} + \Delta \varepsilon) \\ &= \gamma + (\Delta Y_{-1}' Z Z' \Delta Y_{-1})^{-1} \Delta Y_{-1}' Z Z' \Delta \varepsilon \end{aligned}$$

Si les conditions sur les moments sont effectivement vérifiées alors la limite en probabilité de  $\hat{\gamma}_{mco}$  est égale à  $\gamma$ . L'estimateur est donc asymptotiquement sans biais.

Malheureusement, et parce que les erreurs  $\Delta \varepsilon_{it}$  sont autocorrélées, cet estimateur n'est pas efficient. On doit lui préférer l'estimateur des MCG qui tient compte explicitement du caractère autocorrélé des erreurs. Il est calculé comme (c'est l'estimateur « *one step* » de Arellano et Bond :

$$\hat{\gamma}_{ABI} = (\Delta Y_{-1}' Z (Z' \Omega Z)^{-1} Z' \Delta Y_{-1})^{-1} \Delta Y_{-1}' Z (Z' \Omega Z)^{-1} Z' \Delta Y$$

Muni de l'estimation  $\hat{\gamma}_{ABI}$  du coefficient  $\gamma$ , il est possible de calculer le vecteur  $\widehat{\Delta \varepsilon}$  des résidus du modèle en différence première. Il est alors possible de construire l'estimateur de Arellano et Bond « *two step* » qui consiste à substituer à la matrice  $Z' \Omega Z$  une estimation plus fine de la matrice des variances – covariances :

$$V = Z' \widehat{\Delta \varepsilon} \widehat{\Delta \varepsilon}' Z$$

Au final, l'estimateur des moments généralisés à la Arellano – Bond est donné par :

$$\hat{\gamma}_{AB2} = (\Delta Y_{-1}' Z V^{-1} Z' \Delta Y_{-1})^{-1} \Delta Y_{-1}' Z V^{-1} Z' \Delta Y$$

et l'estimateur de la variance de l'estimateur de  $\gamma$  est :

$$s^2(\hat{\gamma}_{AB2}) = (\Delta Y_{-1}' Z V^{-1} Z' \Delta Y_{-1})^{-1}$$

Application :

Dependent Variable: GREVES  
 Method: Panel Generalized Method of Moments  
 Transformation: First Differences  
 Date: 02/25/04 Time: 22:14  
 Sample (adjusted): 1952 1985  
 Cross-sections included: 17  
 Total panel (balanced) observations: 561  
 White period instrument weighting matrix  
 White period standard errors & covariance (d.f. corrected)  
 Instrument list: @DYN(GREVES,-2,-2) INFL CHOM

---



---

Variable	Coefficien t	Std. Error	t-Statistic	Prob.
GREVES(-1)	0.023364	0.000104	224.0578	0.0000
INFL	17.50172	0.190061	92.08494	0.0000
CHOM	-23.15917	0.955292	-24.24302	0.0000

---



---

#### Effects Specification

Cross-section fixed (first differences)

---



---

R-squared	-0.019134	Mean dependent var	-11.31142
Adjusted R-squared	-0.022787	S.D. dependent var	661.9903
S.E. of regression	669.4901	Sum squared resid	2.50E+08
Durbin-Watson stat	2.906874	J-statistic	15.87603
Instrument rank	17.00000		

---



---

**Introduction****Section I : structure générale et typologie des modèles****Section II : Le modèle sur données empilées : estimateur MCO et estimateur Between****Section III : Le modèle à effets fixes****Section IV : Le modèle à effets aléatoires****Section V : L'appréciation de l'aptitude du modèle à reproduire les volatilités totale, intra et inter individuelles de Y****Section VI : Hétéroscédasticité et autocorrélation : détection et correction****Section VII : Les modèles dynamiques**