

ÉCONOMÉTRIE APPLIQUÉE DES DONNÉES DE PANEL

Emmanuel DUGUET

Version 9, 2014

Table des matières

1 Le modèle linéaire	3
1.1 Ecriture du modèle	3
1.2 Le modèle à erreur simple	5
1.2.1 Exemple : la fonction de production (1)	5
1.2.2 Définition	6
1.2.3 Estimation du paramètre	7
1.2.4 Estimation de la variance de l'estimateur	8
1.2.5 Estimation séparée du terme constant	12
1.3 Programmation	13
1.3.1 Avec la procédure REG	13
1.3.2 Avec la procédure IML	17
2 Le modèle à effet individuel	23
2.1 Le modèle à erreur composée	23
2.1.1 Exemple : la fonction de production (2)	24
2.1.2 Présentation	24
2.1.3 Propriétés de la perturbation	25
2.1.4 Propriétés des moindres carrés ordinaires	28
2.1.5 Covariance robuste	29
2.1.6 Estimation par les moindres carrés généralisés	30
2.2 Programmation	31
2.3 Le modèle à effet fixe	31
2.3.1 Exemple : la fonction de production (3)	32
2.3.2 Présentation	33
2.3.3 Estimation par les moindres carrés ordinaires	34
3 Estimation inter-individuelle	37
3.1 Ecriture du modèle	37
3.2 Estimateur Between	40

3.3	Questions complémentaires	42
3.3.1	Expression séparée du terme constant	42
3.3.2	Les variables qui ne varient que dans une dimension	43
3.4	Programmation	44
3.4.1	Avec la procédure REG	44
3.4.2	Avec la procédure IML	46
4	Estimation intra-individuelle	49
4.1	Ecriture du modèle	49
4.2	Traitement du terme constant	52
4.3	Estimateur Within	53
4.4	Questions complémentaires	54
4.4.1	Optimalité asymptotique	54
4.4.2	Les variables qui ne varient que dans une dimension	55
4.5	Programmation	55
4.5.1	Avec la procédure REG	56
4.5.2	Avec la procédure IML	58
5	Estimation optimale	63
5.1	Les moindres carrés généralisés	63
5.1.1	Estimation	63
5.1.2	Expression séparée du terme constant	66
5.1.3	Cas extrêmes : MCO et Within	67
5.1.4	Ecriture comme une moyenne pondérée	68
5.1.5	Estimateur avec terme constant séparé	70
5.2	Les moindres carrés quasi-généralisés	71
5.2.1	Variance inter-individuelle (between)	72
5.2.2	Variance intra-individuelle (within)	74
5.2.3	Estimation des composantes de la variance	76
5.2.4	Estimation du paramètre	77
5.2.5	Application des moindres carrés quasi-généralisés	78
5.3	Programmation sous IML	78
6	Les tests de spécification	83
6.1	Le test de Fisher	83
6.2	Le test de Hausman	84
6.3	Le test de Mundlak	88
6.3.1	Test direct	88
6.3.2	Test par régression auxiliaire	89
6.4	Equivalence entre les tests de Hausman et Mundlak	91

6.5	Un autre test	92
6.6	Programmation sous IML	94
7	Effet individuel et prévision	97
7.1	Estimateur brut de l'effet individuel	97
7.2	Les variables constantes dans le temps	100
7.3	Estimateur centré de l'effet individuel	101
7.4	Prévision et résidu du modèle	102
7.5	Programmation sous IML	102
8	Modèles estimés en différences	109
8.1	Estimation	109
8.2	Application	113
9	Modèle général à erreur composée	117
9.1	Les transformations	119
9.2	Les estimateurs	123
9.2.1	Le between individuel	123
9.2.2	Le between temporel	124
9.2.3	Le double within	124
9.2.4	Les moindres carrés généralisés	125
9.3	Estimation des composantes de la variance	127
9.3.1	Estimation de σ_ε^2	127
9.3.2	Trace des matrices de projection	128
9.3.3	Estimation de σ_α^2	130
9.3.4	Estimation de σ_β^2	131
9.4	Tests de spécification	133
9.4.1	Test de Hausman	133
9.4.2	Tests à la Mundlak	133
9.5	Programmation sous IML	135
10	Estimation sur panels non cylindrés	143
10.1	Estimation inter-individuelle (between)	144
10.2	Estimation intra-individuelle (within)	147
10.3	Estimation des composantes de la variance	149
10.3.1	Variance inter-individuelle (between)	149
10.3.2	Variance intra-individuelle (within)	151
10.4	Estimation par les moindres carrés généralisés	153
10.5	Estimation par les moindres carrés quasi-généralisés	155
10.6	Tests de spécification	156
10.7	Programmation	156

11 Le modèle apparemment linéaire	173
11.1 Le modèle avec erreurs de mesure	173
11.2 Le modèle avec simultanéité	178
11.3 Le modèle autorégressif	179
11.3.1 Modèle avec effet fixe	179
11.3.2 Modèle à erreur composée	180
11.3.3 Modèle à perturbations autocorrélées	181
12 Estimation par variables instrumentales	183
12.1 Le modèle	183
12.2 Les variables instrumentales	185
12.3 L'estimateur à variables instrumentales	186
12.4 Cas particuliers	191
12.5 Comment trouve t-on une variable instrumentale?	191
12.6 Les tests de spécification	193
12.6.1 Le test de Wu-Hausman	194
12.6.2 Le test de Hausman-Taylor	201
12.7 Programmation 1 : test de Wu-Hausman	206
13 Le modèle non sphérique	211
13.1 Motivation	211
13.2 L'estimateur des moments généralisés	213
13.2.1 Covariance robuste	213
13.2.2 L'estimateur des moments généralisés	214
13.3 Les tests de spécification	215
13.3.1 Le test de Sargan-Hansen	215
13.3.2 Les conditions d'orthogonalité	216
13.3.3 La statistique de test	217
14 Le choix des instruments	219
14.1 Modèles estimés en niveaux	219
14.2 Modèles estimés en différences	222
15 Le modèle VAR	225
15.1 Cas général	225
15.1.1 Ecriture et estimation de la forme réduite	226
15.1.2 Estimation de la forme structurelle	228

Introduction

Les progrès continus de l'informatique ont permis de constituer des bases de données de plus en plus riches. Après une période essentiellement consacrée à l'économétrie des séries temporelles, puis des données individuelles, les praticiens purent enfin exploiter les données de panel. Les données de panel sont des données portant sur un ensemble d'individus observés à plusieurs dates. On peut donc les voir aussi bien comme un regroupement des séries temporelles d'un grand nombre d'individus que comme le regroupement de données individuelles prises à plusieurs dates. Les données de panel proviennent généralement du regroupement d'enquêtes individuelles annuelles. Ainsi, les données de panel, par la plus grande quantité d'information qu'elles contiennent, devraient permettre de mieux estimer les relations structurelles rencontrées en économie. Contrairement aux séries temporelles toutefois, les données de panel sont rarement utilisées pour la prévision ou la simulation mais pour révéler les variables significatives dans l'explication d'un comportement (production, consommation etc.).

Par rapport aux données en coupe (i.e., individuelles ou sur un intervalle unique de temps) elle permettent de traiter au moins partiellement l'effet des variables explicatives manquantes. En effet, une variable constante dans le temps ne peut expliquer les variations de comportement d'un individu donné. En conséquence, il doit être possible de neutraliser l'effet de ces variables individuelles en utilisant la dimension temporelle des données. Cette propriété est intéressante car elle permet de neutraliser l'effet des variables individuelles inobservables. Ainsi, on devrait parvenir à des estimations plus fiables dès lors que l'on dispose de données de panel. Deux modèles principaux sont utilisés pour traiter les variables individuelles inobservables : le modèle à effet individuel aléatoire et le modèle à effet fixe. Dans leur version de base, le premier modèle postule une absence de corrélation entre les variables individuelles inobservables et les variables explicatives du modèle alors que le second modèle autorise cette corrélation. Le choix entre ces deux modèles se fait à l'aide des tests de spécification.

Le premier chapitre contient les rappels nécessaires sur le modèle linéaire standard et une présentation des modèles de panel. Il permet de relier l'économétrie des panels aux méthodes habituelles. Le second chapitre porte sur la décomposition des données dans les dimensions inter-individuelle, dite *between*, et intra-individuelle, dite *within*, et sur les méthodes d'estimation qui y sont associées. L'estimateur optimal du modèle à erreur composée est ensuite relié à ces décompositions dans le troisième chapitre, qui discute brièvement l'estimation du modèle à effet fixe. Le chapitre 4 est consacré aux tests usuels de spécification, qui permettent de choisir la bonne méthode d'estimation. Les méthodes d'estimation précédentes sont prolongées au cas des panels non cylindrés de le chapitre 5.

Les chapitres qui suivent sont consacrés aux modèles dans lesquels, d'une part, la perturbation est corrélée aux variables explicatives du modèle et, d'autre part, où il n'est pas possible de régler ce problème par une transformation des données de départ. Les cas de base sont présentés dans le chapitre 6 qui présente les modèles avec erreur de mesure et autorégressif. Dès lors que l'on rencontre ces cas, on doit procéder à une estimation par variable instrumentale, présentée dans le chapitre 7. Cette estimation n'est toutefois pas optimale dans de nombreux cas, en raison de l'existence de problèmes d'autocorrélation ou d'hétéroscédasticité. La solution à ce problème est fournie dans le chapitre 8, consacré à la méthode des moments généralisés. Le chapitre 9 présente les tests de spécification associés au modèle apparemment linéaire, qui visent à tester l'admissibilité des instruments utilisés lors de l'estimation. L'ouvrage se conclut sur l'estimation d'un modèle vectoriel autorégressif dans le dernier chapitre.

Chapitre 1

Le modèle linéaire

Les modèles que nous estimons dans cet ouvrage sont tous linéaires. La méthode d'estimation de base associée aux modèles linéaires est celle des moindres carrés ordinaires. Dans le cas des données de panel, cette méthode est très importante car beaucoup des estimateurs que nous allons voir peuvent s'interpréter comme des moindres carrés ordinaires appliqués soit à une transformation des données de départ (*between*, *within*, MCG, MQCG) soit à un modèle auxiliaire défini de manière *ad hoc* (variables instrumentales, GMM). Il est donc important de bien réviser ce modèle avant de commencer à pratiquer l'économétrie des panels.

1.1 Ecriture du modèle

Le modèle théorique que l'on souhaite estimer est donné par :

$$y = f(X_1, \dots, X_p) + u,$$

où u représente l'erreur que l'on fait en expliquant la variable y par la théorie précédente et (X_1, \dots, X_p) les p variables qui expliquent les valeurs que prend y . Ici, on suppose que la forme de la fonction f est linéaire par rapport aux paramètres :¹

$$f(X_1, \dots, X_p) = b_1 X_1 + \dots + b_p X_p.$$

1. Les variables X peuvent être des transformations des grandeurs économiques étudiées, comme leur logarithme. Ainsi, la fonction que l'on estime n'est généralement linéaire que par rapport aux paramètres du modèle. On peut également raisonner sur des transformations connues des paramètres de départ, à condition d'utiliser le théorème de Slutsky qui permet d'obtenir des estimateurs convergents.

Dans l'ensemble, on peut donc écrire que :

$$y = (X_1, \dots, X_p) \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} + u = Xb + u$$

où X désigne le vecteur $(1, p)$ des variables explicatives et b le paramètre $(p, 1)$ que nous cherchons à estimer.

On possède un échantillon comportant des informations sur un ensemble d'individus indicés par $i = 1, \dots, N$ que l'on suppose présents pendant toute la période d'étude.² Les dates auxquelles les données sont observées sont indicées par $t = 1, \dots, T$. Pour chaque individu, on peut écrire que :

$$\begin{aligned} y_{it} &= X_{it}b + u_{it}, \\ i &= 1, \dots, N \\ t &= 1, \dots, T \end{aligned}$$

Ainsi les variables explicatives pour l'individu i à la date t sont regroupées dans un vecteur ligne :

$$X_{it} = (X_{it1}, \dots, X_{pit}).$$

La première opération que nous allons effectuer est d'écrire le modèle individu par individu, puis nous écrirons le modèle qui empile toutes les données. Pour l'individu i les données se présentent comme une simple série temporelle. On a :

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} \quad \text{et} \quad X_i = \begin{pmatrix} X_{i1} \\ (1,p) \\ \vdots \\ X_{iT} \\ (1,p) \end{pmatrix}.$$

Le modèle pour un individu s'écrit donc :

$$y_i = X_i b + u_i, \quad i = 1, \dots, N.$$

La seule différence par rapport au cas usuel est la présence d'un indice d'individu pour chaque série temporelle, puisqu'il y a N séries temporelles dans notre panel. Comme le paramètre b est le même pour tous les individus, on

2. On parle d'échantillon *cylindré*.

peut se ramener à un modèle empilé qui utilisera toutes les observations. On a :

$$\begin{cases} y_1 = X_1 b + u_1 \\ \vdots \\ y_N = X_N b + u_N \end{cases}$$

ce qui se réécrit :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} b + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}$$

On pose donc :

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} \quad \text{et} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix},$$

ce qui donne le modèle linéaire :

$$y = Xb + u.$$

Concrètement, ce rangement des données revient à trier d'abord par individu puis, pour chaque individu, à trier par date. Selon les propriétés de la perturbation u on sera amené à choisir différents estimateurs. Nous commençons par le cas le plus simple : celui d'un bruit blanc.

1.2 Le modèle à erreur simple

1.2.1 Exemple : la fonction de production (1)

Nous allons prendre l'exemple d'une fonction de production dans l'ensemble de l'ouvrage. Ceci nous permettra de mieux voir les différences d'hypothèses entre les différents modèles de panel sur un exemple simple. On considère une fonction de production de type Cobb-Douglas :

$$Y_{it} = A_{it} C_{it}^{\gamma_C} L_{it}^{\gamma_L}, \quad (1.1)$$

où Y_{it} est la valeur ajoutée, C_{it} le capital physique et L_{it} l'emploi. Le terme A_{it} est appelé *productivité globale des facteurs* et représente le potentiel productif de l'entreprise i pour la raison suivante. La productivité globale des facteurs est reliée à l'efficacité productive de la technologie de production employée. Si $A_{it} > A_{jt}$, l'entreprise i peut produire plus que l'entreprise j avec les *mêmes* quantités d'inputs à la même date. Pour obtenir un modèle linéaire, on prend la relation (1.1) en logarithmes :

$$\ln Y_{it} = \ln A_{it} + \gamma_C \ln C_{it} + \gamma_L \ln L_{it} \Leftrightarrow y_{it} = a_{it} + \gamma_C c_{it} + \gamma_L \ell_{it},$$

avec $y_{it} = \ln Y_{it}$, $c_{it} = \ln C_{it}$, $\ell_{it} = \ln L_{it}$ et $a_{it} = \ln A_{it}$. Pour faire apparaître un terme d'erreur dans ce modèle, il suffit de supposer que le logarithme de la productivité globale des facteurs se décompose entre une partie fixe, noté a , qui représente sa valeur moyenne sur l'échantillon, et un bruit blanc, noté ε_{it} , qui résume l'hétérogénéité des productivités des entreprises. On pose donc :

$$a_{it} = a + \varepsilon_{it}, \quad (1.2)$$

ce qui permet de réécrire la fonction de production sous la forme :

$$\begin{aligned} y_{it} &= a + \varepsilon_{it} + \gamma_C c_{it} + \gamma_L \ell_{it} \\ &= X_{it}b + u_{it}, \end{aligned}$$

avec :

$$X_{it} = (1, c_{it}, \ell_{it}), \quad b = \begin{pmatrix} a \\ \gamma_C \\ \gamma_L \end{pmatrix} \text{ et } u_{it} = \varepsilon_{it}$$

ε_{it} non corrélé avec (c_{it}, ℓ_{it}) .

1.2.2 Définition

Les différents modèles de panel se distinguent par les hypothèses qu'il font sur le terme d'erreur. Dans cette section, nous supposons simplement que le terme d'erreur vérifie toutes les hypothèses habituelles. Afin d'assurer la cohérence des notations avec les chapitres qui suivent, nous poserons :

$$u_{it} = \varepsilon_{it}, \quad \varepsilon_{it} \rightsquigarrow BB(0, \sigma_\varepsilon^2),$$

où la notation BB désigne un bruit blanc. Plus précisément, un bruit blanc vérifie les hypothèses suivantes :

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

sans perte de généralité tant que les variables explicatives contiennent un terme constant. Les variances sont données par :

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

et les covariances sont nulles aussi bien entre les individus à toute date que dans le temps pour un même individu :³

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0 \quad \forall i \neq j \forall t,$$

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

Sous ces hypothèses on peut écrire le modèle sous la forme :

$$y = Xb + u,$$

avec :

$$E(u|X) = 0, \quad V(u|X) = \sigma_{\varepsilon}^2 I_{NT}.$$

1.2.3 Estimation du paramètre

La forme particulière de la matrice de covariance, scalaire, est due à l'absence de corrélation entre les individus à toute date, et entre les différentes dates pour tous les individus. D'après le théorème de Gauss-Markov, le meilleur estimateur sans biais linéaire en y est donné par la méthode des moindres carrés ordinaires. Il est défini par :

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

En remplaçant y par son expression dans \hat{b} , on voit que :

$$\begin{aligned} \hat{b} &= (X'X)^{-1} X'(Xb + u) \\ &= (X'X)^{-1} X'Xb + (X'X)^{-1} X'u \\ &= b + (X'X)^{-1} X'u, \end{aligned}$$

tout l'aléa qui porte sur l'estimation provient donc du terme d'erreur. On déduit de l'expression précédente que l'estimateur des moindres carrés ordinaires est sans biais :

$$\begin{aligned} E(\hat{b} - b | X) &= E\left((X'X)^{-1} X'u \mid X\right) \\ &= (X'X)^{-1} X'E(u|X) \\ &= 0. \end{aligned}$$

3. Nous relâcherons cette hypothèse dans les chapitres ultérieurs.

La variance de cet estimateur est donnée par :

$$\begin{aligned}
 V(\hat{b}|X) &= V(\hat{b} - b|X) \\
 &= V\left((X'X)^{-1} X'u \mid X\right) \\
 &= (X'X)^{-1} X'V(u|X) X (X'X)^{-1} \\
 &= \sigma_\varepsilon^2 (X'X)^{-1} \underbrace{X'I_{NT}X}_{X'X} (X'X)^{-1} \\
 &= \sigma_\varepsilon^2 (X'X)^{-1}.
 \end{aligned}$$

Dans la pratique, on ne peut connaître la valeur exacte de σ_ε^2 donc il faut l'estimer. On se base sur la somme des carrés des résidus du modèle. La prévision du modèle est donnée par $\hat{y} = X\hat{b}$. Le résidu de la régression est défini par l'écart entre la vraie valeur de la variable expliquée y et sa prévision par les moindres carrés : $\hat{u} = y - \hat{y}$.

1.2.4 Estimation de la variance de l'estimateur

La prévision est obtenue par une projection orthogonale de y sur le sous-espace vectoriel engendré par les variables explicatives X , que nous noterons $\text{Im}(X)$. En effet, la prévision est égale à :

$$\begin{aligned}
 \hat{y} &= X\hat{b} \\
 &= X(X'X)^{-1} X'y \\
 &= P_X Y,
 \end{aligned}$$

et la matrice P_X représente une projection orthogonale. On vérifie en effet que P_X est à la fois idempotente et symétrique.⁴ Plus précisément, la matrice est idempotente car :

$$\begin{aligned}
 P_X^2 &= P_X P_X \\
 &= X(X'X)^{-1} X'X \underbrace{(X'X)^{-1} X'}_{I_p} \\
 &= X(X'X)^{-1} X' \\
 &= P_X
 \end{aligned}$$

4. RAPPEL. Une matrice A est idempotente si $A = A^2$ et elle est symétrique si $A = A'$.

et elle est symétrique car :

$$\begin{aligned} P'_X &= \left[X (X'X)^{-1} X' \right]' \\ &= X' (X'X)^{-1} X \\ &= P_X. \end{aligned}$$

Ceci nous permet de définir le résidu comme :

$$\begin{aligned} \hat{u} &= y - \hat{y} \\ &= y - P_X y \\ &= (I_{NT} - P_X) y. \end{aligned}$$

Cette relation nous permet de définir une deuxième matrice de projection orthogonale sur l'espace orthogonal à l'espace de prévision. On note cet espace $\text{Im}^\perp(X)$. La matrice de projection orthogonale sur ce sous-espace vectoriel est donnée par :

$$M_X = I_{NT} - P_X.$$

On vérifie qu'elle est idempotente car :

$$\begin{aligned} M_X^2 &= M_X M_X \\ &= (I_{NT} - P_X) (I_{NT} - P_X) \\ &= I_{NT}^2 - I_{NT} P_X - P_X I_{NT} + P_X^2 \\ &= I_{NT} - P_X - P_X + P_X \\ &= I_{NT} - P_X \\ &= M_X. \end{aligned}$$

et elle est symétrique puisque :⁵

$$\begin{aligned} M'_X &= (I_{NT} - P_X)' \\ &= I'_{NT} - P'_X \\ &= I_{NT} - P_X \\ &= M_X. \end{aligned}$$

Cette digression va nous permettre de simplifier bien des calculs, à la fois dans ce chapitre et dans les suivants. On voit immédiatement que le résidu est

5. On vérifie que la matrice identité est elle-même une matrice de projection orthogonale, sur le sous-espace vectoriel engendré par la base canonique de \mathbb{R}^{NT} .

une fonction linéaire de la perturbation du modèle :

$$\begin{aligned}\hat{u} &= M_X y \\ &= M_X (Xb + u) \\ &= M_X Xb + M_X u,\end{aligned}$$

or la projection de X sur l'espace orthogonal au sous-espace engendré par X est forcément nulle :

$$\begin{aligned}M_X X &= (I_{NT} - P_X) X \\ &= X - X \underbrace{(X'X)^{-1} X'X}_{I_p} \\ &= 0,\end{aligned}$$

en conséquence :⁶

$$\hat{u} = M_X u,$$

ce qui nous permet d'écrire la somme des carrés des résidus du modèle comme :

$$\begin{aligned}S &= \begin{matrix} \hat{u}' & \hat{u} \\ (1, NT) & (NT, 1) \end{matrix} \\ &= u' M_X' M_X u \\ &= u' M_X u.\end{aligned}$$

Le calcul de l'espérance de cette somme des carrés fournit la solution à notre problème d'estimation :⁷

$$E(S|X) = E(\text{tr}(S)|X) \quad \text{car } S \text{ est un nombre réel,}$$

donc :

$$\begin{aligned}E(S|X) &= E(\text{tr}(u' M_X u) | X) \\ &= E(\text{tr}(u u' M_X) | X) \\ &= E(\text{tr}(M_X u u') | X) \\ &= \text{tr}(M_X E(u u' | X)),\end{aligned}$$

6. Cette relation montre bien la différence qui existe entre la perturbation du modèle u et le résidu \hat{u} .

7. RAPPELS. Soit A une matrice carrée :

- $\text{tr}(A) =$ somme des éléments diagonaux de $A = \sum_i a_{ii}$.
- $E(\text{tr}A) = \text{tr}E(A)$.
- Soient trois matrices A, B, C on a $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$ sous réserve que les formats des matrices coïncident.

or

$$E(uu'|X) = V(u|X) = \sigma_\varepsilon^2 I_{NT},$$

ce qui implique :

$$E(S|X) = \sigma_\varepsilon^2 \text{tr}(M_X).$$

Il nous reste donc à trouver la trace de cette matrice de projection orthogonale pour avoir enfin un estimateur de σ_ε^2 . La solution est facile :

$$\begin{aligned} \text{tr}(M_X) &= \text{tr}\left(I_{NT} - X(X'X)^{-1}X'\right) \\ &= \text{tr}(I_{NT}) - \text{tr}\left(X(X'X)^{-1}X'\right) \\ &= NT - \text{tr}\left(X'X(X'X)^{-1}\right) \\ &= NT - \text{tr}(I_p) \\ &= NT - p. \end{aligned}$$

Nous pouvons donc écrire que :

$$E(S|X) = \sigma_\varepsilon^2 (NT - p),$$

ce qui est équivalent à dire que :

$$E\left(\frac{S}{NT - p} \middle| X\right) = \sigma_\varepsilon^2,$$

d'où l'estimateur sans biais de la variance de la perturbation :

$$\widehat{\sigma}_\varepsilon^2 = \frac{S}{NT - p} = \frac{\hat{u}'\hat{u}}{NT - p}.$$

Cet estimateur s'écrit également :

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{NT - p} \sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2.$$

Cette double sommation ne doit pas surprendre car bien que la somme se fasse sur toutes les observations, celles-ci sont repérées par deux indices au lieu d'un seul. L'estimateur de la variance de \hat{b} est donc :

$$\widehat{V}(\hat{b}|X) = \widehat{\sigma}_\varepsilon^2 (X'X)^{-1}.$$

1.2.5 Estimation séparée du terme constant

Pour les comparaisons que nous effectuerons plus loin, nous aurons besoin d'une expression séparée des coefficients des variables autres que le terme constant. Pour cela, nous séparons les variables de la manière suivante :

$$X = (e_{NT}, \underline{X}),$$

ainsi que les coefficients :

$$b = \begin{pmatrix} a \\ \underline{b} \end{pmatrix},$$

où a désigne le terme constant du modèle et \underline{b} le vecteur des coefficients des autres variables explicatives. L'estimateur des MCO est défini par la solution du système d'équations suivant :

$$X'X \hat{b} = X'y,$$

ce qui est équivalent à :

$$\begin{pmatrix} e'_{NT} \\ \underline{X}' \end{pmatrix} (e_{NT}, \underline{X}) \begin{pmatrix} \hat{a} \\ \underline{\hat{b}} \end{pmatrix} = \begin{pmatrix} e'_{NT} \\ \underline{X}' \end{pmatrix} y,$$

et donne le système :

$$\begin{cases} e'_{NT} e_{NT} \hat{a} + e'_{NT} \underline{X} \underline{\hat{b}} = e'_{NT} y \\ \underline{X}' e_{NT} \hat{a} + \underline{X}' \underline{X} \underline{\hat{b}} = \underline{X}' y \end{cases}$$

La première équation permet d'écrire :

$$\hat{a} = (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \underline{\hat{b}}),$$

expression que l'on reporte dans la seconde équation :

$$\underline{X}' e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \underline{\hat{b}}) + \underline{X}' \underline{X} \underline{\hat{b}} = \underline{X}' y,$$

on remarque que la matrice :

$$B_{NT} \triangleq e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT},$$

est symétrique et idempotente. Il s'agit de la matrice de projection orthogonale sur $\text{Im}(e_{NT})$ qui calcule la moyenne sur toutes les observations et la réplique NT fois dans un vecteur colonne. On obtient finalement :

$$\begin{aligned} \underline{\hat{b}} &= (\underline{X}' \underline{X} - \underline{X}' B_{NT} \underline{X})^{-1} (\underline{X}' y - \underline{X}' B_{NT} y) \\ &= (\underline{X}' (\mathbb{I}_{NT} - B_{NT}) \underline{X})^{-1} \underline{X}' (\mathbb{I}_{NT} - B_{NT}) y. \end{aligned}$$

L'estimateur obtenu consiste donc à centrer les variables globalement, par la matrice de projection orthogonale $I_{NT} - B_{NT}$ puis à appliquer les MCO. Intuitivement, le centrage des données enlève le terme constant du modèle, de sorte qu'il suffit de faire des MCO sur les données centrées pour obtenir directement les coefficients des variables explicatives différentes du terme constant.

1.3 Programmation

Pour obtenir l'estimation par les moindres carrés ordinaires, les programmes usuels d'estimation ne posent pas de problème particulier car les degrés de liberté sont calculés de manière correcte. Toutefois, pour effectuer des estimations plus avancées, nous aurons besoin d'un langage de programmation adapté aux matrices. Dans cet ouvrage, nous rappellerons l'utilisation de la procédure standard de MCO sous SAS (REG) avant de programmer l'estimateur et les statistiques de test sous SAS-IML.

1.3.1 Avec la procédure REG

Les données que nous utilisons sont tirées d'une étude de Hall et Mairesse (1995). Nous avons extrait de ces données un panel cylindré de $N = 671$ entreprises couvrant les années 1981 à 1989, soit $T = 9$ années. Le nombre total d'observations de l'échantillon est donc égal à $NT = 6039$. Les variables utilisées sont les suivantes :

- logarithme des ventes (variables expliquée) : y (nom : LogS) ;
- logarithme de l'emploi (variable explicative) : ℓ (nom : LogL) ;
- logarithme du capital physique (variable explicative) : c (nom : LogC) ;
- Logarithme du capital de recherche et développement (variable explicative) : k (nom : LogK).

Le modèle que nous voulons estimer est donc :

$$y_{it} = a + \gamma_L \ell_{it} + \gamma_C c_{it} + \gamma_K k_{it} + u_{it}.$$

avec $u_{it} = \varepsilon_{it}$ pour une erreur simple. Les données sont rangées dans un tableau SAS `tab` situé dans le répertoire de travail ("Work"), de sorte que l'on entre les commandes :

Programme 1.1.

```
proc reg data=tab;
model LogS=LogL LogC LogK;
run;
```

ce qui fournit le résultat :

Sortie 1.1.

The REG Procedure
Model: MODEL1
Dependent Variable: LogS

Number of Observations Read 6039
Number of Observations Used 6039

Analyse de variance

Source	DF	Somme des carrés	Carré moyen	Valeur F
Model	3	22726	7575.37098	43226.8
Error	6035	1057.61515	0.17525	
Corrected Total	6038	23784		

Analyse de variance

Source	Pr > F
Model	<.0001
Error	
Corrected Total	

Root MSE	0.41863	R-Square	0.9555
Dependent Mean	5.67592	Adj R-Sq	0.9555
Coeff Var	7.37546		

Résultats estimés des paramètres

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t
Intercept	1	3.27944	0.02134	153.68
LogL	1	0.60255	0.00833	72.34
LogC	1	0.35557	0.00629	56.57
LogK	1	0.03663	0.00367	9.97

Résultats estimés des paramètres

Variable	DF	Pr > t
Intercept	1	<.0001
LogL	1	<.0001
LogC	1	<.0001
LogK	1	<.0001

On peut également sortir les résultats dans un tableau SAS nommé B, en entrant les instructions :

Programme 1.2.

```
proc reg data=tab outest=b tableout;
model LogS=LogL LogC LogK;
proc print data=b;
run;
```

ce qui fournit le tableau :

Sortie 1.2.

					I				
					n				
					t				
					e				
					r				
					c	L	L	L	L
					e	o	o	o	o
					p	g	g	g	g
					t	L	C	K	S
1	MODEL1	PARMS	LogS	0.41863	3.279	0.6026	0.3556	0.03663	-1
2	MODEL1	STDERR	LogS	0.41863	0.021	0.0083	0.0063	0.00367	.
3	MODEL1	T	LogS	0.41863	153.679	72.3357	56.5700	9.97187	.
4	MODEL1	PVALUE	LogS	0.41863	0.000	0.0000	0.0000	0.00000	.
5	MODEL1	L95B	LogS	0.41863	3.238	0.5862	0.3433	0.02943	.
6	MODEL1	U95B	LogS	0.41863	3.321	0.6189	0.3679	0.04383	.

Ce format n'est toutefois pas utilisé par les économètres pour présenter les résultats, ce qui justifie que nous transposions ce tableau en utilisant les noms des statistiques comme nouveaux noms de colonne :

Programme 1.3.

```
proc transpose data=b out=b;
id _type_;
proc print;
run;
```

Sortie 1.3.

Obs	_NAME_	_LABEL_	PARMS		
1	_RMSE_	Root mean squared error	0.41863		
2	Intercept	Intercept	3.27944		
3	LogL		0.60255		
4	LogC		0.35557		
5	LogK		0.03663		
6	LogS		-1.00000		

Obs	STDERR	T	PVALUE	L95B	U95B
1	0.41863	0.419	0.41863	0.41863	0.41863
2	0.02134	153.679	0.00000	3.23760	3.32127
3	0.00833	72.336	0.00000	0.58622	0.61888
4	0.00629	56.570	0.00000	0.34325	0.36790
5	0.00367	9.972	0.00000	0.02943	0.04383
6

L'expression "RMSE" est l'abréviation de "Root Mean Squared Error" et donne l'estimation de $\hat{\sigma}_\varepsilon$, "Intercept" signifie "Constante" et le coefficient -1 pour y indique qu'il s'agit de la variable expliquée. La colonne PARMS donne les estimations des paramètres du modèle, STDERR les estimations des écarts types des paramètres estimés, T le t de Student, PVALUE la probabilité critique des estimations des paramètres (si elle est inférieure ou égale à α , on rejette l'hypothèse que le paramètre est nul au seuil α), L95B est la borne inférieure de l'intervalle de confiance de niveau 95% et U95B la borne supérieure correspondante.

On peut ensuite exporter ces résultats dans une feuille de calcul afin de les mettre en forme pour les inclure dans un rapport. Par exemple, si l'on prend une feuille Excel :

Programme 1.4.

```
proc export data=b outfile="c:\resultats.XLS" replace
dbms=excel5; run;
```

1.3.2 Avec la procédure IML

La majorité des estimateurs que nous verrons dans cet ouvrage ne sont pas disponibles par des procédures standard, il nous faudra donc les programmer nous mêmes. Pour cela nous utiliserons le langage matriciel de SAS disponible dans la procédure IML ("Interactive Matrix Language"). Un autre avantage de cette procédure est qu'elle permet de faire du "sur mesure" aussi bien pour les sorties que pour les statistiques de test. Comme dans la section précédente, nous supposerons que les données sont rangées dans un tableau TAB situé sous le répertoire de travail. Voici le programme d'estimation commenté :

Programme 1.5.

1. `proc iml;`
Début de la procédure IML.
2. `ny={LogS}; nx={LogL LogC LogK}; nx=t(nx);`
Création de deux vecteurs de caractères, `ny` contient la chaîne de caractère `LogS`, et `nx` est un vecteur ligne à deux éléments contenant respectivement les chaînes de caractères `LogL`, `LogC` et `LogK`. On transpose ensuite la liste des variables explicatives de manière à obtenir un vecteur colonne. La fonction IML `t(.)` transpose un vecteur ou une matrice.
3. `use tab;`
On ouvre le tableau SAS `tab`;
4. `read all var (ny) into y;`
Dans le tableau SAS `tab`, on lit le nom de la variable expliquée contenue dans le vecteur de caractères `ny` et l'on place cette colonne de données dans un vecteur IML de nom `y`. Les noms IML sont gérés indépendamment des noms des variables de la table SAS.

5. `read all var (nx) into x;`

Dans le tableau SAS `tab`, on lit les noms des variables explicatives contenues dans le vecteur de caractères `nx` et l'on place les colonnes de données dans une matrice IML de nom `x`.

6. `nt=nrow(y);`

On définit une variable IML de nom `nt`, égale au nombre de lignes du vecteur IML `y`, c'est-à-dire au nombre total d'observations. Son nom vient du fait que le nombre d'observations est égal à $N \times T$. Plus généralement, `nrow(.)` donne le nombre de lignes d'une matrice.

7. `x=J(nt,1,1)||x;`

On ajoute un terme constant à la matrice des variables explicatives. La fonction IML `J(a,b,c)` crée une matrice constante d'éléments tous égaux à `c`, dont les dimensions sont de `a` lignes et `b` colonnes. L'expression `J(nt,1,1)` crée donc un vecteur colonne de `nt` lignes, dont tous les éléments sont égaux à 1. Pour ajouter ce terme constant à la matrice des variables explicatives, on utilise une concaténation horizontale, effectuée par l'instruction `"||"` (touches `AltGr+6`). L'instruction `A||B` crée une matrice égale à (A, B) . Il faut que les matrices `A` et `B` aient le même nombre de lignes. Ici l'instruction `J(nt,1,1)||x` ajoute une colonne unité à la matrice des variables explicatives. On réaffecte le résultat à la matrice `x`, qui a donc gagné une colonne.

8. `nx="Constante"//nx;`

Comme nous venons d'ajouter un terme constant à la matrice `x`, il faut modifier la liste des noms des variables afin que `nx` corresponde toujours à `x`. On choisit ici le nom "Constante" pour désigner le terme constant. Toutefois, comme `nx` est un vecteur colonne, il faut effectuer une concaténation verticale, que l'on obtient par l'instruction `"//"`. L'instruction `A//B` donne la matrice :

$$\begin{pmatrix} A \\ B \end{pmatrix},$$

Il faut que les matrices `A` et `B` aient le même nombre de colonnes. Ici l'instruction `nx="Constante"//nx` ajoute la chaîne "Constante" au vecteur des noms des variables explicatives.

9. `p=ncol(x);`

On affecte le nombre de colonnes de la matrice des variables explicatives à la variable IML `p`. Ce nombre servira à calculer les degrés de liberté plus loin.

10. `ixx=inv(t(x)*x);`

On calcule $(X'X)^{-1}$ et que l'on affecte à la variable IML `ixx`. On la stocke

ainsi parce que la matrice $(X'X)^{-1}$ est utilisée deux fois dans le programme et qu'une inversion matricielle est coûteuse en temps de calcul. La fonction IML `inv(.)` inverse une matrice carrée et l'étoile `*` représente le produit matriciel.

11. `b=ixx*t(x)*y;`

Calcul de l'estimation des MCO, $(X'X)^{-1} X'y$, que l'on affecte au vecteur IML `b`.

12. `u=y-x*b;`

Calcul des résidus $\hat{u} = y - X\hat{b}$ que l'on affecte au vecteur IML `u`.

13. `su2=ssq(u)/(nt-p);`

Estimation sans biais de la variance de la perturbation du modèle

$$\hat{\sigma}_u^2 = \frac{1}{NT-p} \sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2$$

la fonction IML `ssq` prend la somme des carrés des éléments de la matrice IML `u`.

14. `vb=su2#ixx;`

Estimation de la variance de l'estimateur :

$$\hat{\sigma}_u^2 (X'X)^{-1},$$

on remarque que le produit d'une matrice par un nombre réel est effectué par l'instruction `#`, et non par l'instruction `*`.

15. `sb=sqrt(vecdiag(vb));`

Estimation des écarts types des estimateurs. L'instruction IML `vecdiag` prend la diagonale d'une matrice carrée et la range dans un vecteur colonne. Elle effectue donc la transformation de :

$$\begin{pmatrix} \widehat{V}(\hat{b}_1) & \widehat{Cov}(\hat{b}_1, \hat{b}_2) & \cdots & \widehat{Cov}(\hat{b}_1, \hat{b}_p) \\ \widehat{Cov}(\hat{b}_1, \hat{b}_2) & \widehat{V}(\hat{b}_2) & \cdots & \widehat{Cov}(\hat{b}_2, \hat{b}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{Cov}(\hat{b}_1, \hat{b}_p) & \widehat{Cov}(\hat{b}_2, \hat{b}_p) & \cdots & \widehat{V}(\hat{b}_p) \end{pmatrix}$$

en

$$\begin{pmatrix} \widehat{V}(\hat{b}_1) \\ \widehat{V}(\hat{b}_2) \\ \vdots \\ \widehat{V}(\hat{b}_p) \end{pmatrix},$$

la fonction IML `sqrt(.)` prend la racine carrée élément par élément dans le vecteur précédent, de sorte que l'on obtient les écarts types des estimations des coefficients du modèle.

16. `tb=abs(b)/sb;`
Calcul des t de Student, qui sont mis dans le vecteur IML `tb`.
17. `print "Estimation dans la dimension totale",`
Début de l'impression des résultats. La virgule représente un saut de ligne. Cette instruction `print` ne finira que lorsqu'elle rencontrera un point-virgule.
18. `"Variable expliquée :"` `ny,`
Imprime le nom de la variable expliquée.
19. `b [rowname=nx] sb tb;`
Imprime les estimations du paramètre, rangées dans `b`, avec en commentaire le vecteur `nx`, les estimations des écarts types et les t de Student.
20. `res=b||sb||tb;`
Crée une matrice IML `res`, qui contient le tableau des résultats d'estimation. On ne peut mettre que des variables de même type dans une matrice (tout en numérique ou tout en caractères).
21. `create nx from nx[colname="Variables"];`
`append from nx;`
Crée un tableau SAS de nom `nx` (indiqué par `create`) à partir du vecteur IML `nx` (indiqué par `from`). La colonne du tableau SAS qui vient d'être créée s'appellera "Variables". Le tableau SAS est alimenté par le vecteur IML `nx` (indiqué par `append`).
22. `create res from res[colname=("PAR"// "STD"// "T")];`
`append from res;`
Crée un tableau SAS de nom `res` (indiqué par `create`) à partir du vecteur IML `res` (indiqué par `from`). La colonne du tableau SAS qui vient d'être créée s'appellera "Variables". Le tableau SAS est alimenté par le vecteur indiqué directement dans l'instruction `colname` en respectant la syntaxe des vecteurs colonnes. Notons que l'on aurait pu utiliser la syntaxe des vecteurs ligne en entrant `colname=("PAR"||"STD"||"T")`. Le tableau SAS `res` est alimenté par le vecteur IML `res` (indiqué par `append`).
23. `quit; run;`
Ferme la procédure IML. On retourne donc en langage SAS standard.
24. `data res; merge nx res;`
On fusionne les tableaux SAS `nx` et `res` ligne à ligne (instruction `merge`) et l'on place le résultat dans le tableau `res` (instruction `data`);

25. proc print; run;

On imprime le tableau SAS.

On obtient la sortie suivante par la procédure IML :

Sortie 1.4.

```

Estimation dans la dimension totale
                                NY
                                Variable expliquée : LOGS
B                                SB                                TB
Constante 3.2794352 0.0213396 153.67858
LOGL      0.6025503 0.0083299 72.335715
LOGC      0.3555747 0.0062856 56.570018
LOGK      0.036631 0.0036734 9.9718733

```

et la sortie suivante par la procédure *print* :

Sortie 1.5.

Obs	Variabes	PAR	STD	T
1	Constante	3.27944	0.021340	153.679
2	LOGL	0.60255	0.008330	72.336
3	LOGC	0.35557	0.006286	56.570
4	LOGK	0.03663	0.003673	9.972

ce dernier tableau peut alors être exporté sous une feuille de calcul comme nous l'avons vu plus haut.

Chapitre 2

Le modèle à effet individuel

Les hypothèses faites pour le modèle à erreur simple sont difficiles à soutenir dans de nombreux cas. En effet, peu de modèles peuvent résumer la totalité de l'hétérogénéité des comportements des individus au moyen des seules variables explicatives. On parle d'hétérogénéité inobservée, soit parce que les données individuelles qui la déterminent ne sont pas présentes dans l'enquête soit parce que ces données ne sont pas mesurables. Pour les entreprises, on peut penser à la qualité de la gestion, pour les ménages, à l'influence du milieu familial. L'avantage des données de panel réside dans leur capacité à fournir des estimateurs convergents même lorsque ces informations sont manquantes.

Il existe deux modèles principaux qui prennent en compte l'hétérogénéité des individus. Le modèle à erreur composée suppose que l'effet individuel est sans corrélation avec les variables explicatives. On parle de modèle à effet non corrélé. Le modèle à effet fixe, que nous verrons plus loin suppose que l'hétérogénéité est corrélée avec les variables explicatives. On parle de modèle à effet corrélé. Ce dernier modèle est considéré comme le plus pertinent par une majorité de spécialistes.

2.1 Le modèle à erreur composée

Le modèle à erreur composée a été introduit en 1966 par Balestra et Nerlove. Il consiste à introduire un effet individuel dans la perturbation du modèle. Ceci implique que l'estimation par les moindres carrés ordinaires n'est pas optimale. De plus, les statistiques de test déduites des formules habituelles ne sont plus valables de sorte qu'il faut modifier le calcul de la matrice de covariance de l'estimateur. La solution réside dans l'application des moindres carrés généralisés. Toutefois, pour pouvoir mettre en oeuvre cet estimateur, il faut

dra d'abord estimer la matrice de covariance des perturbations du modèle à erreur composée, ce qui implique d'introduire deux autres estimateurs : l'estimateur inter-individuel (*between*) et l'estimateur intra-individuel (*within*).

2.1.1 Exemple : la fonction de production (2)

On considère la fonction de production (1.1). Pour faire apparaître un effet individuel dans ce modèle, il suffit de modifier l'hypothèse (1.2). Cette fois ci, on décompose la productivité globale des facteurs en trois composantes. La première composante est la valeur moyenne de la productivité globale des facteurs sur l'échantillon, notée a , la seconde ε_{it} est un différentiel global de productivité et la troisième est un différentiel *individuel* de productivité pris en écart à la productivité moyenne, donc d'espérance nulle, noté α_i . On généralise donc la relation (1.2) de la manière suivante :

$$\begin{aligned} a_{it} &= a + \varepsilon_{it} + \alpha_i, \\ E(\alpha_i|X) &= 0, V(\alpha_i|X) = \sigma_\alpha^2, \\ \text{Cov}(\alpha_i, \alpha_j|X) &= 0 \quad \forall i \neq j, \text{Cov}(\alpha_i, \varepsilon_{jt}|X) = 0 \quad \forall (i, j, t) \end{aligned}$$

Le modèle à erreur composée suppose que a_i et ε_{it} sont des variables aléatoires indépendantes entre elles et indépendantes des inputs de la fonction de production (c_{it}, ℓ_{it}). On peut alors réécrire la fonction de production sous la forme :

$$\begin{aligned} y_{it} &= a + \varepsilon_{it} + \alpha_i + \gamma_C c_{it} + \gamma_L \ell_{it} \\ &= X_{it}b + u_{it}, \end{aligned}$$

avec :

$$X_{it} = (1, c_{it}, \ell_{it}), b = \begin{pmatrix} a \\ \gamma_C \\ \gamma_L \end{pmatrix} \text{ et } u_{it} = \varepsilon_{it} + \alpha_i,$$

ε_{it} et α_i ni corrélés entre eux ni avec (c_{it}, ℓ_{it}) .

2.1.2 Présentation

Le modèle à erreur composée est défini de la manière suivante :

$$\begin{aligned} y_{it} &= X_{it}b + u_{it} \\ u_{it} &= \varepsilon_{it} + \alpha_i \\ i &= 1, \dots, N \\ t &= 1, \dots, T \end{aligned}$$

L'erreur est composée de deux termes, d'où le nom du modèle. Le premier terme ε_{it} est un bruit blanc identique à celui du modèle à erreur simple. Le second terme α_i ne dépend que de l'individu i et ne varie pas dans le temps. Ainsi, il résume les différences permanentes de comportement entre individus qui ne sont pas prises en compte par les variables explicatives X et qui ont pourtant une influence sur la variable expliquée y . La définition de ce modèle serait incomplète sans les propriétés des perturbations. La première composante vérifie :

$$E(\varepsilon_{it}|X) = 0, V(\varepsilon_{it}|X) = \sigma_\varepsilon^2 \quad \forall (i, t),$$

ainsi que :

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}|X) = 0 \quad \forall i \neq j \forall t,$$

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{is}|X) = 0 \quad \forall t \neq s \forall i.$$

La seconde composante du modèle, l'effet aléatoire individuel, vérifie :

$$E(\alpha_i|X) = 0, V(\alpha_i|X) = \sigma_\alpha^2 \quad \forall i.$$

avec :

$$\text{Cov}(\alpha_i, \alpha_j|X) = 0 \quad \forall i \neq j,$$

$$\text{Cov}(\alpha_i, \varepsilon_{jt}|X) = 0 \quad \forall (i, j, t),$$

$$\text{Cov}(X_{it}, \alpha_i|X) = 0 \quad \forall (i, t).$$

Ceci revient à supposer qu'il existe une distribution des effets individuels dans la population totale, dont la moyenne est constante dans le temps et de variance σ_α^2 constante.¹ Cet effet individuel n'est pas corrélé avec l'erreur ε , ce qui correspond juste à une hypothèse visant à bien séparer les effets individuels des autres sources d'hétérogénéité. Enfin, les effets aléatoires individuels ne sont pas corrélés entre eux, ce qui revient à dire que toutes les corrélations entre les comportements des individus passent par les variables explicatives observables X du modèle.

2.1.3 Propriétés de la perturbation

Pour déterminer la méthode d'estimation optimale, il faut commencer par étudier la variance de la perturbation du modèle empilé. Pour un individu i

1. L'hypothèse de nullité de l'espérance de l'effet individuel se fait sans perte de généralité tant que le modèle comporte un terme constant.

donné, on a :

$$\begin{cases} u_{i1} = \varepsilon_{i1} + \alpha_i \\ \vdots \\ u_{iT} = \varepsilon_{iT} + \alpha_i \end{cases}$$

ce qui peut s'écrire sous forme vectorielle :

$$\begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}_{(T,1)} = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}_{(T,1)} + \begin{pmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{pmatrix}_{(T,1)} = \varepsilon_i + \alpha_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{(T,1)},$$

on introduit donc le vecteur unité d'ordre T que l'on note :

$$e_T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{(T,1)},$$

d'où l'erreur du modèle pour un individu :

$$u_i = \varepsilon_i + \alpha_i e_T, \quad i = 1, \dots, N$$

Cette erreur est d'espérance nulle :

$$E(u_i|X) = E(\varepsilon_i|X) + E(\alpha_i|X) e_T = \mathbf{0}_{(T,1)}, \quad i = 1, \dots, N$$

en empilant les erreurs de tous les individus, on trouve donc que :

$$E(\mathbf{u}|X) = E \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} \Big| X = \begin{pmatrix} E(u_1|X) \\ \vdots \\ E(u_N|X) \end{pmatrix} = \mathbf{0}_{(NT,1)}.$$

La variance du modèle individuel est égale à :

$$\begin{aligned} V(u_i|X) &= E(u_i u_i' | X) - E(u_i | X) E(u_i | X)' \\ &= E((\varepsilon_i + \alpha_i e_T) (\varepsilon_i' + \alpha_i e_T') | X) \\ &= E(\varepsilon_i \varepsilon_i' + (\alpha_i \varepsilon_i) e_T' + e_T (\alpha_i \varepsilon_i') + \alpha_i^2 e_T e_T' | X) \\ &= E(\varepsilon_i \varepsilon_i' | X) + E(\alpha_i \varepsilon_i | X) e_T' + e_T E(\alpha_i \varepsilon_i' | X) + E(\alpha_i^2 | X) e_T e_T' \\ &= E(\varepsilon_i \varepsilon_i' | X) + E(\alpha_i^2 | X) e_T e_T', \end{aligned}$$

car l'absence de corrélation entre les deux composantes du terme d'erreur implique que :

$$E(\alpha_i \varepsilon_i | X) = E(\alpha_i \varepsilon_i' | X)' = 0_{(T,1)}.$$

La première partie de la variance est identique à celle du modèle à erreur simple :

$$E(\varepsilon_i \varepsilon_i' | X) = \sigma_\varepsilon^2 I_T.$$

La seconde partie de la variance est égale à :

$$E(\alpha_i^2 | X) e_T e_T' = \sigma_\alpha^2 \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

car $e_T e_T'$ n'est autre que la matrice unité carrée d'ordre T . Dans l'ensemble, on obtient :

$$\begin{aligned} V(u_i | X) &= \sigma_\varepsilon^2 \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} + \sigma_\alpha^2 \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\alpha^2 \end{pmatrix}. \end{aligned}$$

Cette forme particulière de matrice de covariance est intéressante car elle montre bien que l'existence d'un effet aléatoire individuel crée de l'autocorrélation au sein de la série temporelle propre à chaque individu. Plus précisément, cette autocorrélation est constante dans le temps et égale à :

$$\rho_u = \frac{\text{Cov}(u_{it}, u_{it-h} | X)}{V(u_{it} | X)} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} \in [0, 1], \quad \forall h \neq 0.$$

L'origine de cette autocorrélation des observations d'un même individu est claire : le même effet individuel influence les comportements d'un individu à *toutes* les dates.

La conséquence immédiate de cette forme d'autocorrélation est de rendre inefficace une estimation par les moindres carrés ordinaires. La matrice de covariance des perturbations est la même pour tous les individus. On la note Σ :

$$\Sigma = V(u_i | X) = \sigma_\varepsilon^2 I_T + \sigma_\alpha^2 e_T e_T', \quad i = 1, \dots, N$$

Comme les perturbations des différents individus ne sont pas corrélées, la variance du modèle empilé est égale à :

$$\begin{aligned} V(u|X) &= V \left(\begin{array}{c|c} u_1 & \\ \vdots & \\ u_N & \end{array} \middle| X \right) \\ &= \begin{pmatrix} V(u_1|X) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & V(u_N|X) \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{(T,T)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma_{(T,T)} \end{pmatrix}. \end{aligned}$$

La réécriture avec le produit de Kronecker permet de simplifier cette expression :

$$\Omega_{(NT,NT)} = V(u|X) = I_{(N,N)} \otimes \Sigma_{(T,T)}.$$

Cette matrice de covariance n'est pas scalaire, en conséquence le théorème de Gauss-Markov implique que l'utilisation des moindres carrés ordinaires n'est plus l'estimateur optimal. Il faut appliquer le théorème d'Aitken, qui établit que l'estimateur des moindres carrés généralisés est l'estimateur optimal.

2.1.4 Propriétés des moindres carrés ordinaires

Le modèle empilé se met sous la forme :

$$y = Xb + u$$

avec

$$E(u|X) = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}_{(NT,1)} \quad \text{et} \quad V(u|X) = \Omega_{(NT,NT)}.$$

Les moindres carrés ordinaires sont toujours sans biais, mais leur matrice de covariance ne prend plus la forme habituelle. Pour l'espérance, on a :

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

donc

$$E(\hat{b} - b | X) = (X'X)^{-1} X'E(u|X) = \mathbf{0}.$$

Pour la variance, on a :

$$\begin{aligned} V(\hat{b}|X) &= E\left((\hat{b} - b)(\hat{b} - b)' \middle| X\right) \\ &= E\left((X'X)^{-1} X' u u' X (X'X)^{-1} \middle| X\right) \\ &= (X'X)^{-1} X' E(u u' | X) X (X'X)^{-1} \\ &= (X'X)^{-1} X' \Omega X (X'X)^{-1}, \end{aligned}$$

et aucune simplification ne permet de se ramener à la forme du modèle à erreur simple. La propriété précédente implique que l'on ne peut plus se fier aux statistiques calculées par les programmes de moindres carrés ordinaires et qu'il faut recourir à une estimation par les moindres carrés généralisés.

2.1.5 Covariance robuste

La matrice de covariance de l'estimateur des moindres carrés ordinaires n'est plus scalaire dans un modèle à effet individuel, ce qui empêche d'utiliser la formule usuelle de l'estimateur de la covariance de l'estimateur des moindres carrés ordinaires. Il existe toutefois une méthode, que nous réutiliserons plus loin, qui permet d'obtenir un estimateur de la bonne matrice de covariance dans le cas général. On a :

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

et tout le problème d'estimation vient du terme en $\Omega_{(NT,NT)}$. En fait, cette matrice est diagonale et peut se réécrire, dans le cas général, sous la forme :

$$\Omega = \text{diag}(\Omega_i),$$

toutefois, il n'est généralement pas possible d'estimer cette matrice sans faire d'hypothèse. En effet, chaque bloc Ω_i contient jusqu'à $T(T+1)/2$ éléments, de sorte que Ω peut contenir jusqu'à $NT(T+1)/2$ paramètres, nombre qui tend vers l'infini lorsque N ou T tendent vers l'infini. Par contre, la matrice $X' \Omega X$ est de dimension (p, p) et contient donc un nombre maximum de $p(p+1)/2$ paramètres, nombre qui reste fini lorsque N ou T tendent vers l'infini. Plus précisément :

$$\begin{aligned} X' \Omega X &= \sum_{i=1}^N \begin{matrix} X_i' & \Omega_i & X_i \\ (p,T) & (T,T) & (T,p) \end{matrix} \\ &= \sum_{i=1}^N \begin{matrix} X_i' & E(u_i u_i') & X_i \\ (p,T) & (T,T) & (T,p) \end{matrix}. \end{aligned}$$

En utilisant ces propriétés, Halbert White (1980) a démontré que l'on peut estimer cette matrice par :

$$\widehat{X' \Omega X} = \sum_{i=1}^N X_i' \hat{u}_i \hat{u}_i' X_i,$$

où \hat{u}_i est le vecteur des résidus des moindres carrés ordinaires pour l'individu i :

$$\hat{u}_i = y_i - X_i \hat{b},$$

d'où l'estimateur de la covariance robuste de l'estimateur des moindres carrés ordinaires :

$$\widehat{V}_R(\hat{b} | X) = (X' X)^{-1} \widehat{X' \Omega X} (X' X)^{-1}.$$

2.1.6 Estimation par les moindres carrés généralisés

L'estimateur des moindres carrés généralisés est donné par :

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

il est sans biais et à variance minimale. Pratiquement, cet estimateur correspond à l'application des moindres carrés ordinaires sur des données transformées (Aitken). Le modèle transformé est défini par :

$$y^* = X^* b + u^*,$$

avec

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

La perturbation du modèle transformé vérifie en effet les hypothèses du théorème de Gauss-Markov puisque :

$$\begin{aligned} E(u^* | X) &= \Omega^{-1/2} E(u | X) \\ &= 0, \end{aligned}$$

$$\begin{aligned} V(u^* | X) &= \Omega^{-1/2} V(u | X) \Omega^{-1/2} \\ &= \Omega^{-1/2} \Omega \Omega^{-1/2} \\ &= \Omega^{-1/2} \Omega^{1/2} \Omega^{1/2} \Omega^{-1/2} \\ &= I_{NT}. \end{aligned}$$

L'estimateur optimal est donc égal à :

$$\begin{aligned} b^* &= (X^{*'} X^*)^{-1} X^{*'} y^* \\ &= \left((\Omega^{-1/2} X)' \Omega^{-1/2} X \right)^{-1} (\Omega^{-1/2} X)' \Omega^{-1/2} y \\ &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y, \end{aligned}$$

en utilisant la symétrie de $\Omega^{-1/2}$. On vérifie que sa matrice de covariance est égale à :

$$V(b^* | X) = (X' \Omega^{-1} X)^{-1}.$$

Il nous reste à trouver une forme explicite de cette transformation matricielle équivalente aux moindres carrés généralisés. Celle forme résultera des hypothèses qui sont faites sur les perturbations. Pour cela nous allons d'abord devoir étudier deux transformations élémentaires des données : en moyennes individuelles ("*between*") et en écart aux moyennes individuelles ("*within*"). Nous reviendront ensuite sur l'estimation par les moindres carrés généralisés et la valeur de $\Omega^{-1/2}$.

2.2 Programmation

Nous allons voir ici la programmation de la covariance robuste de White (1980). Sur données de panel, il faut faire attention à gérer séparément les observations de chaque individu, ce qui oblige à détailler un peu plus la programmation. Le programme suivant réalise l'estimation.

2.3 Le modèle à effet fixe

Le modèle à erreur composée a été critiqué par Mundlak en 1973 sur une base assez solide. Dès lors que des agents économiques optimisent, il ne peuvent ignorer la valeur de l'effet individuel, et la prennent en compte dans leur décision. Ceci peut impliquer, dans un grand nombre de cas, que l'effet individuel est corrélé avec les variables explicatives du modèle. L'estimateur des moindres carrés généralisés ne serait donc plus convergent. La modélisation avec un effet fixe remet en cause l'hypothèse d'indépendance entre l'effet individuel et les variables explicatives du modèle. On fait donc l'hypothèse que :

$$\exists i \text{ ou } t \mid \text{Cov}(X_{it}, \alpha_i | X) \neq 0.$$

2.3.1 Exemple : la fonction de production (3)

On considère toujours la fonction de production (1.1). Mais cette fois ci, nous souhaiterions que l'effet individuel puisse être corrélé avec (c_{it}, ℓ_{it}) . Nous pourrions faire cette hypothèse directement, mais encore faut-il pouvoir lui donner un sens économique. Ici, nous allons voir qu'il suffit que l'entreprise observe la valeur de α_i et maximise son espérance de profit pour que l'effet individuel soit *toujours* corrélé avec les variables explicatives. La fonction de production s'écrit :

$$Y_{it} = A_{it} C_{it}^{\gamma_C} L_{it}^{\gamma_L},$$

on pose :

$$\ln A_{it} = a + \alpha_i + \varepsilon_{it},$$

où a est la valeur moyenne du logarithme de la productivité, α_i la composante individuelle constante dans le temps de la productivité et $\exp(\varepsilon_{it})$ sa composante aléatoire. Pour simplifier les calculs, on pose :

$$E(\exp(a + \varepsilon_{it})) = d,$$

une constante positive quelconque, et on introduit la notation :

$$A_i = \exp(\alpha_i).$$

Le profit de l'entreprise est égal à :

$$\Pi_{it} = p_{it} Y_{it} - r_{it} C_{it} - w_{it} L_{it},$$

où p_{it} est le prix du bien produit, r_{it} le coût d'usage du capital et w_{it} le salaire horaire.² L'espérance de profit est égale à :

$$E(\Pi_{it}) = p_{it} A_i d C_{it}^{\gamma_C} L_{it}^{\gamma_L} - r_{it} C_{it} - w_{it} L_{it},$$

les conditions du premier ordre donnent les valeurs des variables explicatives :³

$$C_{it}^* = \left(\frac{p_{it} A_i d \gamma_C}{r_{it}} \right)^{\frac{1}{1-(\gamma_C+\gamma_L)}} \left(\frac{\gamma_L r_{it}}{\gamma_C w_{it}} \right)^{\frac{\gamma_L}{1-(\gamma_C+\gamma_L)}},$$

$$L_{it}^* = \left(\frac{p_{it} A_i d \gamma_L}{w_{it}} \right)^{\frac{1}{1-(\gamma_C+\gamma_L)}} \left(\frac{\gamma_C w_{it}}{\gamma_L r_{it}} \right)^{\frac{\gamma_C}{1-(\gamma_C+\gamma_L)}},$$

2. Notons que le fait de prendre le coût d'usage du capital évite d'avoir à effectuer une optimisation dynamique, sous l'hypothèse d'un marché parfait des capitaux.

3. La condition du second ordre est vérifiée sous l'hypothèse que les rendements d'échelle sont décroissants ($\alpha + \beta < 1$).

en prenant ces relations en logarithmes et en utilisant $\ln A_i = \alpha_i$ on obtient :

$$c_{it}^* = \frac{1}{1 - (\gamma_C + \gamma_L)} \times \alpha_i + \frac{1}{1 - (\gamma_C + \gamma_L)} \ln \left[\frac{p_{it} d \gamma_C}{r_{it}} \left(\frac{\gamma_L r_{it}}{\gamma_C w_{it}} \right)^{\gamma_L} \right],$$

$$\ell_{it}^* = \frac{1}{1 - (\gamma_C + \gamma_L)} \times \alpha_i + \frac{1}{1 - (\gamma_C + \gamma_L)} \ln \left[\frac{p_{it} d \gamma_L}{w_{it}} \left(\frac{\gamma_C w_{it}}{\gamma_L r_{it}} \right)^{\gamma_C} \right],$$

et l'on voit que, du fait de l'optimisation, l'effet individuel α_i est toujours positivement corrélé aux variables explicatives du modèle, puisque :

$$\text{Cov}(c_{it}^*, \alpha_i) = \frac{\sigma_\alpha^2}{1 - (\gamma_C + \gamma_L)} > 0.$$

En effet, les entreprises avec une grande productivité globale des facteurs ont intérêt à demander *plus* de facteurs de production que les autres, puisque leur productivité marginale est supérieure. Ainsi le modèle à effet fixe n'est pas une simple extension statistique du cas de base mais résulte bien souvent d'un programme d'optimisation qui peut varier avec la relation estimée. Il s'agit d'un argument assez général en économie, d'où la pertinence de la critique. Nous pouvons écrire notre fonction de production sous la forme :

$$y_{it} = a + \varepsilon_{it} + \alpha_i + \gamma_C c_{it} + \gamma_L \ell_{it}$$

$$= X_{it} b + u_{it},$$

avec :

$$X_{it} = (1, c_{it}, \ell_{it}), \quad b = \begin{pmatrix} a \\ \gamma_C \\ \gamma_L \end{pmatrix} \text{ et } u_{it} = \varepsilon_{it} + \alpha_i,$$

α_i corrélé avec (c_{it}, ℓ_{it}) ,

ε_{it} non corrélé avec $(\alpha_i, c_{it}, \ell_{it})$.

2.3.2 Présentation

Comme le montre l'exemple précédent, le modèle à erreurs composées repose sur l'hypothèse forte que l'effet individuel n'est pas corrélé aux variables explicatives. Or, quand cette hypothèse n'est plus vérifiée, les estimations des moindres carrés ordinaires et des moindres carrés quasi-généralisés ne sont plus convergentes et sont biaisées. Nous allons donc voir comment estimer un modèle quand les effets individuels sont corrélés avec les variables explicatives. Cette situation met en évidence une supériorité des données de panel

par rapport aux données individuelles ou temporelles puisqu'on peut interpréter un effet individuel corrélé comme une variable manquante. La seule condition de validité des estimations est alors que ces variables manquantes soient constantes dans le temps. Le modèle est le suivant :

$$\begin{aligned} y_{it} &= X_{it}b + u_{it}, \\ u_{it} &= \varepsilon_{it} + \alpha_i \\ i &= 1, \dots, N \\ t &= 1, \dots, T \end{aligned}$$

où α_i est un effet individuel tel que :

$$\exists i \text{ ou } t \mid \text{Cov}(X_{it}, \alpha_i \mid X) \neq 0.$$

2.3.3 Estimation par les moindres carrés ordinaires

Un moyen direct de parvenir à une estimation convergente et sans biais de la relation précédente est simple : éliminer l'effet individuel. Il existe de nombreux moyens d'y parvenir qui reposent tous sur l'intuition suivante : puisque l'effet est constant dans le temps, il ne peut pas influencer les variations de la variable explicative. En conséquence toute transformation linéaire du modèle original qui fait apparaître une différence entre deux dates différentes élimine l'effet individuel et mène, pour cette raison, à un estimateur sans biais et convergent. Un choix souvent utilisé, historiquement et en pratique, consiste à prendre l'écart aux moyennes individuelles. Le comportement moyen de chaque individu peut être résumé par :

$$y_{i\bullet} = \frac{1}{T} \sum_{t=1}^T y_{it} \quad \text{et} \quad X_{i\bullet} = \frac{1}{T} \sum_{t=1}^T X_{it} \quad i = 1, \dots, N.$$

L'écriture du modèle se fait alors en écart aux moyennes individuelles, ce qui donne :

$$y_{it} - y_{i\bullet} = (X_{it} - X_{i\bullet})b + u_{it} - u_{i\bullet}$$

avec

$$u_{it} - u_{i\bullet} = \varepsilon_{it} - \varepsilon_{i\bullet} + \alpha_i - \frac{1}{T} \sum_{t=1}^T \alpha_i$$

or α_i est constant dans le temps donc il est égal à sa moyenne individuelle :

$$\frac{1}{T} \sum_{t=1}^T \alpha_i = \frac{T\alpha_i}{T} = \alpha_i \Leftrightarrow \alpha_i - \frac{1}{T} \sum_{t=1}^T \alpha_i = 0.$$

Ainsi, on peut appliquer les moindres carrés ordinaires au modèle en écarts. Cet estimateur, que nous retrouverons plus loin est appelé *within*.

Chapitre 3

Estimation inter-individuelle

Cet estimateur, également appelé *between*, consiste à se focaliser sur les différences permanentes entre les individus, en éliminant les différences de nature conjoncturelle. Pour obtenir ce résultat, on calcule les valeurs moyennes associées à chaque individu puis on effectue une régression par les moindres carrés ordinaires sur les moyennes individuelles.

Il existe deux manières d'écrire le modèle : avec réplification des moyennes pour chaque individu ou sans réplification. Le cas le plus pratique sur le plan théorique est celui avec réplification pour des raisons que nous verrons dans ce chapitre. Dans la pratique, on estime toujours le modèle sans réplification afin de réduire le temps de calcul.

3.1 Ecriture du modèle

On part du modèle de base :

$$y_{it} = X_{it}b + u_{it},$$

$$u_{it} = \varepsilon_{it} + \alpha_i$$

$$i = 1, \dots, N$$

$$t = 1, \dots, T$$

Les moyennes individuelles sont définies par :

$$y_{i\bullet} = \frac{1}{T} \sum_{t=1}^T y_{it}, \quad i = 1, \dots, N$$

Pour obtenir les moyennes individuelles, il faut d'abord prendre la somme sur chaque individu (par rapport à t) puis diviser par le nombre d'observations cor-

respondant T . On remarque alors que :

$$e'_{(1,T)(T,1)} e_T = T,$$

et que :

$$e'_{(1,T)(T,1)} y_i = \sum_{t=1}^T y_{it}.$$

Pour des raisons de commodité, qui apparaîtront ultérieurement, on réplique T fois cette moyenne, ce qui revient à la multiplier par le vecteur colonne e_T :¹

$$e_T \times \frac{1}{T} e'_{(1,T)(T,1)} y_i = \frac{1}{T} \begin{pmatrix} \sum_{t=1}^T y_{it} \\ \vdots \\ \sum_{t=1}^T y_{it} \end{pmatrix} = \begin{pmatrix} y_{i\bullet} \\ \vdots \\ y_{i\bullet} \end{pmatrix}_{(T,1)},$$

ce qui se réécrit sous la forme d'une projection orthogonale :

$$\begin{pmatrix} y_{i\bullet} \\ \vdots \\ y_{i\bullet} \end{pmatrix} = e_T \underbrace{(e'_T e_T)^{-1}}_{1/T} e'_T y_i$$

Prendre la moyenne revient donc à effectuer une projection orthogonale sur le sous-espace vectoriel engendré par le terme constant $\text{Im}(e_T)$. On note B_T la matrice de projection orthogonale correspondante :

$$B_T = e_T (e'_T e_T)^{-1} e'_T,$$

on vérifie qu'elle est idempotente et symétrique :

$$\begin{aligned} B_T^2 &= e_T (e'_T e_T)^{-1} \underbrace{e'_T e_T (e'_T e_T)^{-1}}_1 e'_T \\ &= e_T (e'_T e_T)^{-1} e'_T \\ &= B_T, \end{aligned}$$

$$\begin{aligned} B'_T &= [e_T (e'_T e_T)^{-1} e'_T]' \\ &= e_T (e'_T e_T)^{-1} e'_T \\ &= B_T. \end{aligned}$$

1. On peut considérer que cette réplcation est celle qui serait faite sous un logiciel où l'on souhaite fusionner les NT observations y avec les N moyennes individuelles correspondantes.

Pour chaque individu, on peut donc écrire indifféremment :

$$y_{i\bullet} = X_{i\bullet} b + u_{i\bullet}, \quad i = 1, \dots, N$$

ou

$$B_T y_i = B_T X_i b + B_T u_i, \quad i = 1, \dots, N$$

En empilant les observations de tous les individus, on obtient donc :

$$\begin{pmatrix} B_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_T \end{pmatrix}_{(NT, NT)} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} B_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_T \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} b + \begin{pmatrix} B_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_T \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}.$$

Ceci nous amène à définir la matrice dite *between* ou *inter* suivante :

$$B_{(NT, NT)} = \begin{pmatrix} B_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_T \end{pmatrix} = I_N \otimes B_T,$$

ce qui amène à l'expression suivante du modèle empilé :

$$By = BXb + Bu. \tag{3.1}$$

La perturbation du modèle vérifie les propriétés suivantes :

$$\begin{aligned} E(Bu) &= BE(u) = 0 \\ V(u) &= BV(u)B = B\Omega B. \end{aligned}$$

La matrice *between* est une matrice de projection orthogonale. En effet :²

$$\begin{aligned} B^2 &= (I_N \otimes B_T) (I_N \otimes B_T) \\ &= (I_N^2 \otimes B_T^2) \\ &= I_N \otimes B_T \\ &= B, \end{aligned}$$

$$\begin{aligned} B' &= (I_N \otimes B_T)' \\ &= I_N' \otimes B_T' \\ &= I_N \otimes B_T \\ &= B. \end{aligned}$$

La projection orthogonale se fait sur le sous espace vectoriel engendré par les N vecteurs orthogonaux deux à deux (g_1, \dots, g_N) définis par :

$$g_1 = \begin{pmatrix} e_T \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(NT,1)}, \dots, g_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ e_T \end{pmatrix}_{(NT,1)},$$

Ces projections donnent toute une moyenne dupliquée T fois pour chaque individu et les observations relatives à deux individus ne sont jamais utilisées ensemble du fait de l'orthogonalité des vecteurs de cette base.

3.2 Estimateur Between

L'estimation du modèle 3.1 par les moindres carrés ordinaires définit l'estimateur *between* :

$$\begin{aligned} \hat{b}_B &= [(BX)'(BX)]^{-1} (BX)'By \\ &= (X'B'BX)^{-1} X'B'By, \end{aligned}$$

2. RAPPELS.

- $(A \otimes B)(C \otimes D) = (AC \otimes BD)$,
- $(A + B) \otimes C = A \otimes C + B \otimes C$,
- $A \otimes (B + C) = A \otimes B + A \otimes C$,
- $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$,
- $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$.

avec

$$B'B = B^2 = B,$$

ce qui donne :

$$\hat{b}_B = (X'BX)^{-1} X'By.$$

Cet estimateur est sans biais puisque :

$$\hat{b}_B = b + (X'BX)^{-1} X'Bu,$$

ce qui implique :

$$E(\hat{b}_B - b | X) = (X'BX)^{-1} X'BE(u|X) = 0.$$

Pour calculer sa variance, il est plus pratique de réécrire la variance de la perturbation. Nous avons montré que :

$$\Omega = I_N \otimes \Sigma$$

avec

$$\Sigma = \sigma_\varepsilon^2 I_T + \sigma_\alpha^2 e_T e_T',$$

ici on intercale un terme en T pour faire apparaître la matrice *between* individuelle :

$$\begin{aligned} \Sigma &= \sigma_\varepsilon^2 I_T + T\sigma_\alpha^2 \left(\frac{e_T e_T'}{T} \right) \\ &= \sigma_\varepsilon^2 I_T + T\sigma_\alpha^2 e_T (e_T' e_T)^{-1} e_T' \\ &= \sigma_\varepsilon^2 I_T + T\sigma_\alpha^2 B_T, \end{aligned}$$

ce qui permet d'écrire directement l'expression de Ω en fonction des matrices identité et *between* :

$$\begin{aligned} \Omega &= I_N \otimes (\sigma_\varepsilon^2 I_T + T\sigma_\alpha^2 B_T) \\ &= \sigma_\varepsilon^2 (I_N \otimes I_T) + T\sigma_\alpha^2 (I_N \otimes B_T) \\ &= \sigma_\varepsilon^2 I_{NT} + T\sigma_\alpha^2 B. \end{aligned}$$

La variance de l'estimateur *between* est donnée par :

$$\begin{aligned} V(\hat{b}_B | X) &= E \left[(\hat{b}_B - b) (\hat{b}_B - b)' | X \right] \\ &= (X'BX)^{-1} X'E(Bu (Bu)' | X) X (X'BX)^{-1} \\ &= (X'BX)^{-1} X'V(Bu | X) X (X'BX)^{-1} \end{aligned}$$

or on a :

$$\begin{aligned} V(Bu|X) &= B(\sigma_\varepsilon^2 I_{NT} + T\sigma_\alpha^2 B)B \\ &= (\sigma_\varepsilon^2 + T\sigma_\alpha^2)B, \end{aligned}$$

ce qui donne :

$$\begin{aligned} V(\hat{b}_B|X) &= (\sigma_\varepsilon^2 + T\sigma_\alpha^2)(X'BX)^{-1}X'BX(X'BX)^{-1} \\ &= (\sigma_\varepsilon^2 + T\sigma_\alpha^2)(X'BX)^{-1}. \end{aligned}$$

3.3 Questions complémentaires

3.3.1 Expression séparée du terme constant

Pour effectuer les tests de Hausman et de Mundlak, que nous verrons par la suite, nous aurons besoin de comparer seulement les coefficients autres que le terme constant. Pour cela, commençons par séparer le terme constant e_{NT} des autres variables explicatives \underline{X} de la manière suivante :

$$X = (e_{NT}, \underline{X}),$$

le modèle *between* peut s'écrire :

$$\begin{aligned} By &= B(e_{NT}, \underline{X})b + Bu \\ &= (Be_{NT}, B\underline{X})b + Bu \\ &= (e_{NT}, B\underline{X})b + Bu, \end{aligned}$$

car $Be_{NT} = e_{NT}$. En posant :

$$b = \begin{pmatrix} a \\ \underline{b} \end{pmatrix},$$

où a est le terme constant et \underline{b} le coefficient des variables \underline{X} , on obtient :

$$By = e_{NT}a + B\underline{X}b + Bu.$$

L'estimateur des MCO de ce modèle *between*, noté $\hat{b}_B = (\hat{a}_B, \hat{\underline{b}}_B)'$ est défini par la solution du système suivant :

$$\begin{pmatrix} e'_{NT} \\ (B\underline{X})' \end{pmatrix} (e_{NT}, B\underline{X}) \begin{pmatrix} \hat{a}_B \\ \hat{\underline{b}}_B \end{pmatrix} = \begin{pmatrix} e'_{NT} \\ (B\underline{X})' \end{pmatrix} y$$

$$\begin{pmatrix} e'_{NT}e_{NT} & e'_{NT}B\underline{X} \\ \underline{X}'Be_{NT} & \underline{X}'B\underline{X} \end{pmatrix} \begin{pmatrix} \hat{a}_B \\ \hat{\underline{b}}_B \end{pmatrix} = \begin{pmatrix} e'_{NT}y \\ \underline{X}'By \end{pmatrix}$$

$$\begin{cases} e'_{NT} e_{NT} \hat{a}_B + e'_{NT} \underline{X} \hat{b}_B = e'_{NT} y \\ \underline{X}' e_{NT} \hat{a}_B + \underline{X}' B \underline{X} \hat{b}_B = \underline{X}' B y \end{cases}$$

en utilisant la première équation de ce système, on trouve :

$$\hat{a}_B = (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \hat{b}_B),$$

et en reportant ce résultat dans la seconde équation, on trouve :

$$\underline{X}' e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \hat{b}_B) + \underline{X}' B \underline{X} \hat{b}_B = \underline{X}' B y,$$

ici on remarque que la matrice :

$$B_{NT} \triangleq e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT},$$

est symétrique et idempotente, et qu'elle prend la moyenne sur l'ensemble de l'échantillon au lieu de la prendre individu par individu contrairement à la matrice *between* B . On obtient finalement :

$$\hat{b}_B = (\underline{X}' (B - B_{NT}) \underline{X})^{-1} \underline{X}' (B - B_{NT}) y.$$

Ce résultat est intuitif : en centrant le modèle globalement par B_{NT} avant de prendre la transformation *between* par B , on élimine le terme constant du modèle, de sorte que les MCO donnent les coefficients des autres variables \underline{X} . La matrice $Q_I \triangleq B - B_{NT}$ est appelée matrice *between centrée*, nous la retrouverons dans le chapitre sur le modèle général à erreur composée.

3.3.2 Les variables qui ne varient que dans une dimension

Avant de lancer une régression sur les transformations *between*, il faut veiller à retirer du modèle les variables qui ne varient qu'avec la date. Soit une variable $X_{it} = X_t, \forall i = 1, \dots, N$ à la date t . Lorsque l'on effectue la transformation *between*, on obtient la variable :

$$X_{i\bullet} = \frac{1}{T} \sum_{t=1}^T X_{it} = \frac{1}{T} \sum_{t=1}^T X_t = X_{\bullet\bullet}, \forall i,$$

et cette moyenne globale est colinéaire au terme constant du modèle. On ne pourra identifier les coefficients de ce type de variable qu'avec les estimateurs que nous verrons plus loin.

Par contre, aucun problème n'apparaît pour les variables qui sont constantes dans le temps. Dans ce cas, on a $X_{it} = X_i, \forall t = 1, \dots, T$, ce qui implique la transformation *between* suivante :

$$X_{i\bullet} = \frac{1}{T} \sum_{t=1}^T X_{it} = \frac{1}{T} \sum_{t=1}^T X_i = X_i, \forall i,$$

qui ne pose pas de problème particulier.

3.4 Programmation

Comme pour les MCO, les programmes d'estimation standard fournissent les bons degrés de liberté et peuvent donc être utilisés. La seule différence avec les MCO vient de ce que l'on effectue la régression sur les N moyennes individuelles au lieu de l'effectuer sur les NT observations. Dans la partie IML nous introduirons le calcul des MCO sous forme de fonction, car les estimateurs *between*, *within* et des moindres carrés généralisés peuvent être calculés de cette manière. L'utilisation d'une fonction évitera donc de répéter des blocs de programme quand nous calculerons plusieurs estimateurs dans la même procédure IML.

3.4.1 Avec la procédure REG

Le seul point à voir est la manière de calculer les moyennes individuelles. Sur nos données, l'identifiant d'entreprise s'appelle `firm`. Il faut donc d'abord trier la base de données par rapport à cet identifiant puis faire une moyenne pour chacun des identifiants de la base de données. Ensuite on utilise la procédure REG pour réaliser l'estimation.

Programme 3.1.

```
proc sort data=tab; by firm;
proc means noprint data=tab; by firm;
var LogS LogL LogC LogK;
output out=between mean=;
run;

proc reg data=between;
model LogS=LogL LogC LogK;
run;
```

La procédure *sort* sert à trier les données du tableau `tab` par rapport à la variable `firm` (indiquée par l'instruction *by*). La procédure *means* qui suit, avec instruction *by* permet de faire une moyenne pour chacune des modalités de la variable `firm`, c'est à dire pour chaque individu de la base de données. Cette procédure est lancée avec l'option *noprint* afin d'éviter l'impression des statistiques pour chacune des 671 entreprises de la base des données. L'instruction *var* indique la liste des variables dont on fait les moyennes, ceci inclut

la variable expliquée et les variables explicatives. L'instruction *output* crée un tableau de nom *between* dans lequel on range les statistiques indiquées à la suite. L'instruction *mean=* stocke les moyennes individuelles dans le tableau *between* (ainsi que la variable *firm*), sous le même nom que dans l'instruction *var* qui la précède. C'est pour garder les mêmes noms de variables que dans la base de départ (*tab*) que l'on indique aucun nom à la suite de l'instruction *mean=*. On lance ensuite une procédure *reg* sur les moyennes individuelles stockées dans le tableau *between*. On obtient la sortie suivante :

Sortie 3.1.

```

                                The REG Procedure
                                Model: MODEL1
                                Dependent Variable: LogS

                                Number of Observations Read      671
                                Number of Observations Used        671

                                Analyse de variance

Source                            DF          Somme des          Carré          Valeur
                                DF          carrés            moyen          F
Model                             3          2467.45181        822.48394      6632.44
Error                             667        82.71418          0.12401
Corrected Total                    670        2550.16600

                                Analyse de variance

Source                            Pr > F
Model                             <.0001
Error
Corrected Total

                                Root MSE          0.35215          R-Square          0.9676
                                Dependent Mean    5.67592          Adj R-Sq          0.9674
                                Coeff Var         6.20427

                                Résultats estimés des paramètres

Variable  DF          Résultat estimé          Erreur
                                des paramètres          std  Valeur du test t

```

Intercept	1	3.24211	0.05865	55.28
LogL	1	0.58734	0.02321	25.30
LogC	1	0.37220	0.01727	21.56
LogK	1	0.03012	0.00960	3.14

Résultats estimés des paramètres

Variable	DF	Pr > t
Intercept	1	<.0001
LogL	1	<.0001
LogC	1	<.0001
LogK	1	0.0018

3.4.2 Avec la procédure IML

Le calcul des moyennes individuelles se fait de la même manière qu'avec la procédure *reg*. On lit simplement les moyennes individuelles dans le tableau *between* et on lance une estimation par les MCO. Ici, toutefois, nous allons utiliser une fonction pour calculer les MCO afin de pouvoir calculer plusieurs estimateurs à la suite dans les applications ultérieures. On peut utiliser un copier coller à partir du programme précédent pour gagner du temps. Ici, nous reproduisons tout le programme pour la clarté de l'exposé.

Programme 3.2.

1. `proc iml;`
Début de la procédure IML.
2. `ny={LogS}; nx={LogL LogC LogK}; nx=t(nx);`
Noms des variables expliquée et explicatives.
3. `use between;`
Ouvre le tableau *between*
4. `read all var (ny) into by;`
Lecture des observations de la variable expliquée (en *between*), qui sont mises dans le vecteur IML *by*.
5. `read all var (nx) into bx;`
Lecture des observations des variables explicatives (en *between*), qui sont mises dans la matrice IML *bx*.
6. `n=nrow(by);`
Calcul du nombre d'individus (*N*), mis dans la variable IML *n*.

7. `bx=J(n,1,1)||bx;`
Ajout du terme constant e_N parmi les variables explicatives.
8. `nx="Constante"//nx;`
Ajout du nom du terme constant dans le vecteur des noms.
9. `p=ncol(bx);`
Calcul du nombre de variables, mis dans la variable IML `p`.
10. `start mco(y,x) global(ixx);`
Début de la fonction `mco` dont les arguments sont y (vecteur de la variable expliquée) et x (matrice des variables explicatives). Les variables y et x seront remplacées par les quantités indiquées lors de l'appel de la fonction. L'instruction `global` indique que la quantité `ixx` sera transmise à l'extérieur de la fonction une fois qu'elle aura été exécutée. Ceci permet de calculer des statistiques en plus de l'estimateur des MCO et de les transmettre au reste du programme.
11. `ixx=inv(t(x)*x);`
Calcul de l'inverse de la matrice des produits croisés des variables mises en appel de la fonction soit `bx`. On obtient donc $(X'BX)^{-1}$. Elle sera transmise aux lignes de programmes situées après l'appel de la fonction. S'il existe déjà une matrice `ixx`, elle sera écrasée par la nouvelle valeur.
12. `b=ixx*t(x)*y;`
Calcul de l'estimation des MCO sur les moyennes individuelles.
13. `return (b);`
Renvoi de l'estimation des MCO vers l'appel de la fonction `mco`. Le résultat est stocké dans le vecteur `bet`.
14. `finish mco;`
Fin de la fonction `mco`.
15. `bet=mco(by,bx);`
Appel de la fonction MCO avec `by` à la place de y et `bx` à la place de x . Le renvoi de la fonction (i.e. l'estimation des MCO) est rangée dans le vecteur `bet`.
16. `ub=by-bx*bet;`
Calcul du résidu de la régression *between* \hat{u}_B , mais sans duplication.
17. `vub=ssq(ub)/(n-p);`
Calcul de la variance du résidu *between*, sans duplication. Il s'agit d'un estimateur sans biais de $\sigma_\varepsilon^2/T + \sigma_\alpha^2$.
18. `vbet=vub#ixx;`
Estimation de la variance de l'estimateur *between*.

19. `sbet=sqrt(vecdiag(vbet));`
Estimation des écarts types des estimateurs.
20. `tbet=abs(bet)/sbet;`
Estimation des t de Student correspondants.
21. `print "Estimation Between",`
Début de l'impression des résultats.
22. `"Variable expliquée :" ny, bet [rowname=nx] sbet`
`tbet;`
Fin de l'estimation des résultats.
23. `quit; run;`
Fin de la procédure IML.

On obtient la sortie suivante :

Sortie 3.2.

```

Estimation Between
                NY

Variable expliquée : LOGS
BET                SBET                TBET

Constante 3.2421137  0.058648  55.280868
LOGL      0.5873402  0.0232142  25.300922
LOGC      0.3721953  0.0172656  21.557104
LOGK      0.0301222  0.0095967  3.1387936

```

Chapitre 4

Estimation intra-individuelle

Cette estimateur, également appelé *within*, complète l'analyse effectuée à partir des moyennes individuelles. Au lieu de comparer les comportements moyens des individus, on étudie leurs écarts aux moyennes individuelles, c'est-à-dire leurs variations de comportements. Ceci revient, de fait, à éliminer l'effet individuel présent dans la perturbation avant d'effectuer les régressions.

4.1 Ecriture du modèle

On effectue la transformation suivante au niveau individuel :

$$y_{it} - y_{i\bullet} = (X_{it} - X_{i\bullet})b + u_{it} - u_{i\bullet}, \quad i = 1, \dots, N \quad t = 1, \dots, T$$

Une analyse détaillée de la perturbation montre que :

$$\begin{aligned} u_{it} - u_{i\bullet} &= \varepsilon_{it} - \varepsilon_{i\bullet} + \alpha_i - \frac{1}{T} \sum_{t=1}^T \alpha_i \\ &= \varepsilon_{it} - \varepsilon_{i\bullet} \end{aligned}$$

Ainsi, les effets individuels sont éliminés par cette transformation. La série temporelle de l'individu i est donc donnée par :

$$\begin{aligned} \begin{pmatrix} y_{i1} - y_{i\bullet} \\ \vdots \\ y_{iT} - y_{i\bullet} \end{pmatrix} &= \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} - \begin{pmatrix} y_{i\bullet} \\ \vdots \\ y_{i\bullet} \end{pmatrix} \\ &= \mathbf{I}_T y_i - \mathbf{B}_T y_i \\ &= (\mathbf{I}_T - \mathbf{B}_T) y_i, \end{aligned}$$

on effectue donc la transformation dite *within* ou *intra* de matrice W_T :

$$W_T = I_T - B_T.$$

Il s'agit de la matrice de projection orthogonale sur le sous-espace vectoriel orthogonal à $\text{Im}(e_T)$ que l'on note $\text{Im}^\perp(e_T)$. On vérifie que cette matrice de projection est idempotente, symétrique et orthogonale à B_T :

$$\begin{aligned} W_T^2 &= (I_T - B_T)(I_T - B_T) \\ &= I_T^2 - I_T B_T - B_T I_T + B_T^2 \\ &= I_T - B_T - B_T + B_T \\ &= I_T - B_T \\ &= W_T, \end{aligned}$$

$$\begin{aligned} W_T' &= (I_T - B_T)' \\ &= I_T' - B_T' \\ &= I_T - B_T \\ &= W_T, \end{aligned}$$

$$\begin{aligned} W_T B_T &= (I_T - B_T) B_T \\ &= B_T - B_T^2 \\ &= 0. \end{aligned}$$

Pour chaque individu, on peut donc écrire indifféremment :

$$y_{it} - y_{i\bullet} = (X_{it} - X_{i\bullet})b + u_{it} - u_{i\bullet}, \quad i = 1, \dots, N \quad t = 1, \dots, T$$

ou

$$W_T y_i = W_T X_i b + W_T u_i, \quad i = 1, \dots, N.$$

Ici, on remarque que :

$$\begin{aligned} u_{it} - u_{i\bullet} &= \varepsilon_{it} + \alpha_i - (\varepsilon_{i\bullet} + \alpha_i) = \varepsilon_{it} - \varepsilon_{i\bullet} \\ \Leftrightarrow W_T u_i &= W_T \varepsilon_i, \end{aligned}$$

de sorte que seul le bruit blanc du modèle intervient dans la variance de l'estimateur *within*. En empilant les observations de tous les individus, on obtient

donc :

$$\begin{pmatrix} W_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_T \end{pmatrix}_{(NT,NT)} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} W_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_T \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} b + \begin{pmatrix} W_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_T \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}.$$

Ceci nous amène à définir la matrice *within* suivante :

$$\begin{aligned} W_{(NT,NT)} &= \begin{pmatrix} W_T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_T \end{pmatrix} \\ &= I_N \otimes W_T \\ &= I_N \otimes (I_T - B_T) \\ &= I_N \otimes I_T - I_N \otimes B_T \\ &= I_{NT} - B. \end{aligned}$$

Le modèle transformé peut donc s'écrire :

$$Wy = WXb + Wu. \quad (4.1)$$

La perturbation du modèle vérifie les propriétés suivantes :

$$\begin{aligned} E(Wu|X) &= WE(u|X) = 0 \\ V(Wu|X) &= WV(u|X)W = W\Omega W. \end{aligned}$$

La matrice *within* est une matrice de projection orthogonale. En effet :¹

$$\begin{aligned} W^2 &= (I_N \otimes W_T) (I_N \otimes W_T) \\ &= (I_N^2 \otimes W_T^2) \\ &= I_N \otimes W_T \\ &= W, \end{aligned}$$

$$\begin{aligned} W' &= (I_N \otimes W_T)' \\ &= I_N' \otimes W_T' \\ &= I_N \otimes W_T \\ &= W. \end{aligned}$$

La projection orthogonale se fait sur le sous espace vectoriel orthogonal à celui sur lequel la transformation *between* projette les données, que l'on note $\text{Im}^\perp(g_1, \dots, g_N)$. Pour vérifier l'orthogonalité des transformations *within* et *between* entre elles, il suffit de vérifier que leur produit matriciel est nul :

$$WB = (I_{NT} - B)B = B - B^2 = 0.$$

On obtient donc une décomposition des vecteurs de \mathbb{R}^{NT} puisque :

$$W + B = I_{NT},$$

et que I_{NT} est la matrice de projection orthogonale sur les vecteurs de la base canonique de \mathbb{R}^{NT} .

4.2 Traitement du terme constant

On remarque alors que le modèle *within* ne peut pas contenir de terme constant. En effet :

$$We_{NT} = (I_N \otimes W_T) (e_N \otimes e_T) = (e_N \otimes W_T e_T),$$

1. On peut aussi utiliser :

$$W^2 = (I_{NT} - B)(I_{NT} - B) = I_{NT}^2 - I_{NT}B - BI_{NT} + B^2 = I_{NT} - B = W$$

et :

$$W' = (I_{NT} - B)' = I_{NT}' - B' = I_{NT} - B = W.$$

or :

$$W_T e_T = \left(I_T - e_T (e_T' e_T)^{-1} e_T' \right) e_T = e_T - e_T = 0,$$

donc

$$W e_{NT} = e_N \otimes 0 = 0.$$

Ce résultat implique que la matrice WX n'est pas de plein rang colonne, donc que $(WX)'WX$ n'est pas inversible. Il faut donc retirer le terme constant du modèle quand on réalise l'estimation *within*. Ce point est important car il implique le changement suivant des degrés de liberté : alors que la régression *between*, qui comporte toujours avec un terme constant, comporte p variables explicatives, la régression *within* ne comporte jamais de terme constant et comporte donc $p - 1$ variables explicatives. Pour clarifier les expressions, nous posons donc la notation suivante :

$$\underset{(NT,p)}{X} = (e_{NT}, \underline{X}),$$

où $\underline{X}_{(NT,p-1)}$ désigne les variables explicatives autres que le terme constant. La matrice $W\underline{X}$ est de plein rang colonne et $(W\underline{X})'(W\underline{X})$ est donc inversible.

4.3 Estimateur Within

L'estimation du modèle (4.1) par les moindres carrés ordinaires définit l'estimateur *within* :

$$\begin{aligned} \hat{\underline{b}}_W &= \left[(W\underline{X})'(W\underline{X}) \right]^{-1} (W\underline{X})' W y \\ &= (\underline{X}' W' W \underline{X})^{-1} \underline{X}' W' W y, \end{aligned}$$

avec

$$W'W = W^2 = W,$$

ce qui donne :

$$\hat{\underline{b}}_W = (\underline{X}' W \underline{X})^{-1} \underline{X}' W y.$$

Cet estimateur est sans biais puisque :

$$\hat{\underline{b}}_W = b + (\underline{X}' W \underline{X})^{-1} \underline{X}' W u,$$

ce qui implique :

$$E(\hat{\underline{b}}_W - b | X) = (\underline{X}' W \underline{X})^{-1} \underline{X}' W E(u | X) = 0.$$

Pour calculer sa variance, on utilise la variance des perturbations :

$$\Omega = \sigma_\varepsilon^2 \mathbf{I}_{NT} + T\sigma_\alpha^2 B$$

La variance de l'estimateur *within* est donnée par :

$$\begin{aligned} V(\hat{\underline{b}}_W | X) &= E \left[(\hat{\underline{b}}_W - b) (\hat{\underline{b}}_W - b)' | X \right] \\ &= (\underline{X}' W \underline{X})^{-1} \underline{X}' E(Wu(Wu)' | X) \underline{X} (\underline{X}' W \underline{X})^{-1} \\ &= (\underline{X}' W \underline{X})^{-1} \underline{X}' V(Wu | X) \underline{X} (\underline{X}' W \underline{X})^{-1} \end{aligned}$$

or on a :

$$\begin{aligned} V(Wu | X) &= W (\sigma_\varepsilon^2 \mathbf{I}_{NT} + T\sigma_\alpha^2 B) W \\ &= \sigma_\varepsilon^2 W \end{aligned}$$

ce qui donne :

$$\begin{aligned} V(\hat{\underline{b}}_W | X) &= \sigma_\varepsilon^2 (\underline{X}' W \underline{X})^{-1} \underline{X}' W \underline{X} (\underline{X}' W \underline{X})^{-1} \\ &= \sigma_\varepsilon^2 (\underline{X}' W \underline{X})^{-1}. \end{aligned}$$

4.4 Questions complémentaires

4.4.1 Optimalité asymptotique

Le modèle dans la dimension intra-individuelle s'écrit :

$$Wy = W\underline{X}b + W\varepsilon,$$

où \underline{X} est la matrice des variables explicatives sans terme constant, de dimensions $(NT, p-1)$. En appliquant les moindres carrés ordinaires à ces transformations, on obtient l'estimateur :

$$\hat{\underline{b}}_W = (\underline{X}' W \underline{X})^{-1} \underline{X}' W y,$$

qui est sans biais et convergent. L'estimateur *within* est donc un bon choix pour régler le problème de l'effet fixe. De plus la matrice de covariance pour un individu est égale à :

$$V(W_T \varepsilon_i | X) = \sigma_\varepsilon^2 (I_T - B_T) = \sigma_\varepsilon^2 \begin{pmatrix} 1 - 1/T & \cdots & -1/T \\ \vdots & \ddots & \vdots \\ -1/T & \cdots & 1 - 1/T \end{pmatrix}.$$

La somme des colonnes de cette matrice est nulle, de sorte qu'elle n'est pas inversible. Il existe toutefois un cas où la matrice de covariance devient scalaire : quand T est grand. En effet quand $T \rightarrow +\infty$, on a :

$$V(W_T \varepsilon_i | X) \rightarrow \sigma_\varepsilon^2 I_T,$$

ce qui est une condition suffisante d'optimalité. Dans le cas de notre application $T = 9$ et la matrice est de la forme suivante :

$$\sigma_\varepsilon^2 \begin{pmatrix} 0.89 & \cdots & -0.11 \\ \vdots & \ddots & \vdots \\ -0.11 & \cdots & 0.89 \end{pmatrix}.$$

4.4.2 Les variables qui ne varient que dans une dimension

Avant de lancer une régression sur les transformations *within*, il faut veiller à retirer du modèle les variables qui ne varient qu'avec les individus. Soit une variable $X_{it} = X_i, \forall t = 1, \dots, T$, pour l'individu i . Lorsque l'on effectue la transformation *within*, on obtient la variable :

$$X_{it} - X_{i\bullet} = X_i - \frac{1}{T} \sum_{t=1}^T X_i = 0, \forall i,$$

qui est égale au vecteur nul. On ne pourra identifier les coefficients de ce type de variable qu'en estimant d'abord les effets individuels à partir des estimations *within*, ou avec les estimateurs *between* et des moindres carrés généralisés, méthodes que nous verrons plus loin.

Par contre, aucun problème n'apparaît pour les variables qui ne varient que dans le temps. Dans ce cas, on a $X_{it} = X_t, \forall i = 1, \dots, N$, ce qui implique la transformation *within* suivante :

$$X_{it} - X_{i\bullet} = X_t - \frac{1}{T} \sum_{t=1}^T X_t = X_t - X_{\bullet\bullet}, \forall i,$$

qui ne pose pas de problème particulier.

4.5 Programmation

L'estimateur *within* est convergent même en présence d'effets individuels corrélés avec les variables explicatives individuelles, c'est donc un estimateur particulièrement important. On peut l'obtenir par les programmes standard de

MCO, toutefois les écarts types sont faux ainsi que les t de Student. Ceci vient du fait que l'ordinateur prend pour degrés de liberté l'écart entre le nombre d'observations et le nombre de variables explicatives. Même en enlevant le terme constant du modèle par l'option appropriée, on obtient donc $NT - (p - 1)$ au lieu de $N(T - 1) - (p - 1)$. Si l'on examine le ratio de ces deux quantités, on a :

$$\frac{NT - (p - 1)}{N(T - 1) - (p - 1)} \xrightarrow{N \rightarrow +\infty} \frac{T}{T - 1} > 1.$$

En surestimant les degrés de liberté, on sous-estime l'écart type du résidu de la régression, ce qui implique que l'on surestimera les t de Student avec un programme de MCO standard. Ce problème se pose surtout sur les panels courts. Le tableau suivant donne l'importance de la surestimation des t de Student quand $N \rightarrow +\infty$ en fonction du nombre d'années disponibles, donnée par :

$$\left(\sqrt{T/(T - 1)} - 1 \right) \times 100$$

T	surestimation %
2	41,4%
3	22,5%
4	15,5%
5	11,8%
6	9,5%
7	8,0%
8	6,9%
9	6,1%
10	5,4%

Notons que l'on peut toujours effectuer une correction "à la main" sous un tableur. Dans notre application $T = 9$ de sorte que l'erreur n'est que de 6,1%, mais ce n'est pas représentatif du cas général où, le plus souvent, $T \leq 4$.

4.5.1 Avec la procédure REG

Le point essentiel de programmation consiste à construire les écarts aux moyennes individuelles. Il s'agit donc de retrancher à chaque variable la moyenne de l'individu à laquelle elle est associée. Ceci peut être fait simplement grâce à la procédure *standard* avec instruction *by* de manière à centrer les variables du modèle pour chacune des modalités de l'identifiant individuel `firm`. On retire le terme constant du modèle par l'option *noint* de l'instruction *model* de la procédure *reg*. Le programme est le suivant :

Programme 4.1.

```

proc sort data=tab; by firm;
proc standard mean=0 data=tab out=within; by firm;
var LogS LogL LogC LogK;
run;

proc reg data=within;
model LogS=LogL LogC LogK/noint;
run;

```

et l'on obtient la sortie suivante :

Sortie 4.1. (écarts types faux)

The REG Procedure
 Model: MODEL1
 Dependent Variable: LogS

Number of Observations Read	6039
Number of Observations Used	6039

NOTE: No intercept in model. R-Square is redefined.

Analyse de variance

Source	DF	Somme des carrés	Carré moyen	Valeur F
Model	3	546.84752	182.28251	3855.32
Error	6036	285.38660	0.04728	
Uncorrected Total	6039	832.23412		

Analyse de variance

Source	Pr > F
Model	<.0001
Error	
Uncorrected Total	

Root MSE	0.21744	R-Square	0.6571
Dependent Mean	2.89735E-17	Adj R-Sq	0.6569
Coeff Var	7.504828E17		

Résultats estimés des paramètres

Variable	DF	Résultat estimé des paramètres	Erreur std	Valeur du test t
LogL	1	0.71440	0.01122	63.68
LogC	1	0.20410	0.00979	20.85
LogK	1	0.21294	0.00847	25.15

Résultats estimés des paramètres

Variable	DF	Pr > t
LogL	1	<.0001
LogC	1	<.0001
LogK	1	<.0001

Ici le lecteur doit faire attention au fait que la définition du R^2 change quand on estime le modèle sans terme constant, de sorte que la valeur indiquée sur la sortie ne peut pas être comparée à celle de l'estimation *between* du chapitre précédent.

4.5.2 Avec la procédure IML

Nous allons maintenant ajouter l'estimateur *within* au programme écrit pour l'estimateur *between*. Ceci nous permettra d'effectuer des comparaisons entre ces deux estimateurs plus loin. Pour cela, il faut faire attention au fait qu'il y a maintenant deux listes de variables explicatives : *nx* sans le terme constant (pour l'estimateur *within*) et *nxc* avec le terme constant (pour l'estimateur *between*). Les noms ont été ajustés de manière assez directe. On notera que la fonction *mco* est appelée deux fois : lors du premier appel, la matrice *ixx* contient $(X'BX)^{-1}$ alors que lors du second appel elle contient $(\underline{X}'W\underline{X})^{-1}$. Le programme complet devient donc :

Programme 4.2.

```
proc means noprint data=tab; by firm;
```



```

var LogS LogL LogC LogK;
output out=between mean=;
proc standard mean=0 data=tab out=within; by firm;
var LogS LogL LogC LogK;
run;

proc iml;
ny={LogS}; nx={LogL LogC LogK}; nx=t(nx);
use between;
  read all var (ny) into by;
  read all var (nx) into bx;
use within;
  read all var (ny) into wy;
  read all var (nx) into wx;
nt=nrow(wy); n=nrow(by); t=nt/n;
bx=J(n,1,1)||bx;
nxc="Constante"//nx;
p=ncol(bx);
print "Estimation sur panel cylindré",
"Nombre d'observations =" nt,
"Nombre d'individus =" n,
"Nombre de dates =" t;
start mco(y,x) global(ixx);
ixx=inv(t(x)*x);
b=ixx*t(x)*y;
return (b);
finish mco;
bet=mco(by,bx);
ub=by-bx*bet;
vub=ssq(ub)/(n-p);
vbet=vub#ixx;
sbet=sqrt(vecdiag(vbet));
tbet=abs(bet)/sbet;
print "Estimation Between",
"Variable expliquée :" ny, bet [rowname=nxc] sbet
tbet;
wit=mco(wy,wx);
uw=wy-wx*wit;
vuw=ssq(uw)/(n*(t-1)-(p-1));/* attention */

```

```

vwit=vuw#ixx;
swit=sqrt(vecdiag(vwit));
twit=abs(wit)/swit;
print "Estimation Within",
"Variable expliquée :" ny, wit [rowname=nx] swit twit;
quit; run;

```

qui donne le résultat :

Sortie 4.2. (*écarts types corrects*)

```

Estimation sur panel cylindré
NT
Nombre d'observations =      6039
N
Nombre d'individus =      671
T
Nombre de dates =      9

Estimation Between
NY
Variable expliquée : LOGS
BET          SBET          TBET
Constante 3.2421137  0.058648  55.280868
LOGL      0.5873402  0.0232142  25.300922
LOGC      0.3721953  0.0172656  21.557104
LOGK      0.0301222  0.0095967  3.1387936

Estimation Within
NY
Variable expliquée : LOGS
WIT          SWIT          TWIT
LOGL 0.7143977  0.0119004  60.031621
LOGC 0.204099  0.010383  19.657056

```

LOGK 0.2129392 0.0089813 23.709215

Chapitre 5

Estimation optimale

5.1 Les moindres carrés généralisés

5.1.1 Estimation

La décomposition que nous venons de voir entre les variations inter-individuelles (*between*) et intra-individuelles (*within*) permet de réécrire beaucoup plus simplement le problème d'estimation par les moindres carrés généralisés. Pour cela il faut trouver la racine carrée de l'inverse de la matrice de covariance des perturbations. Or nous venons de voir que cette matrice peut se mettre sous la forme d'une somme pondérée de matrices orthogonales, ce qui simplifie considérablement son inversion. Plus précisément :

$$\begin{aligned}\Omega &= \sigma_\varepsilon^2 I_{NT} + T\sigma_\alpha^2 B & (5.1) \\ &= \sigma_\varepsilon^2 (W + B) + T\sigma_\alpha^2 B \\ &= \sigma_\varepsilon^2 W + (\sigma_\varepsilon^2 + T\sigma_\alpha^2) B,\end{aligned}$$

en conséquence l'inverse est obtenue en prenant simplement l'inverse des coefficients de pondération des matrices de la décomposition :

$$\Omega^{-1} = \frac{1}{\sigma_\varepsilon^2} W + \frac{1}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} B.$$

Vérifions-le :

$$\begin{aligned}
\Omega\Omega^{-1} &= (\sigma_\varepsilon^2 W + (\sigma_\varepsilon^2 + T\sigma_\alpha^2) B) \left(\frac{1}{\sigma_\varepsilon^2} W + \frac{1}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} B \right) \\
&= \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2} W^2 + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} WB + \frac{\sigma_\varepsilon^2 + T\sigma_\alpha^2}{\sigma_\varepsilon^2} BW + \frac{\sigma_\varepsilon^2 + T\sigma_\alpha^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} B^2 \\
&= W + B \\
&= I_{NT}.
\end{aligned}$$

Pour simplifier l'interprétation, on pose :

$$\Omega^{-1} = \frac{1}{\sigma_\varepsilon^2} (W + \theta^2 B) \quad \text{avec} \quad \theta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} \in]0, 1[.$$

Le paramètre θ^2 représente l'importance des variations intra individuelles dans les variations totales. Intuitivement, plus ce paramètre sera élevé, plus la régression effectuée en *within* aura une erreur importante et moins il sera optimal d'y recourir. C'est ce qui explique que la pondération de la transformation *between* soit augmentée quand θ est élevé. Inversement quand θ est faible, cela signifie que l'erreur de la régression effectuée en *within* est également faible c'est-à-dire que les estimations sont plus précises dans cette dimension. Il est alors optimal de réduire le poids de l'opérateur *between* dans la transformation pour se concentrer sur la transformation *within*. Enfin, il convient de remarquer que, pour calculer b^* , on peut enlever le terme $1/\sigma_\varepsilon^2$ pour la raison suivante :¹

$$\begin{aligned}
b^* &= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y \\
&= \left(X' \frac{1}{\sigma_\varepsilon^2} (W + \theta^2 B) X \right)^{-1} X' \frac{1}{\sigma_\varepsilon^2} (W + \theta^2 B) y \\
&= \sigma_\varepsilon^2 \frac{1}{\sigma_\varepsilon^2} (X'(W + \theta^2 B)X)^{-1} X'(W + \theta^2 B) y \\
&= (X'(W + \theta^2 B)X)^{-1} X'(W + \theta^2 B) y.
\end{aligned}$$

Il suffit donc de trouver une racine carrée de la matrice $W + \theta^2 B$ pour obtenir la transformation à faire sur les données avant de lancer l'estimation par les moindres carrés ordinaires. Comme les composantes de cette matrice sont orthogonales, il suffit de prendre la racine carrée des coefficients. En effet, on

1. Ce terme intervient toutefois dans la variance de b^* .

a :

$$\begin{aligned}(W + \theta B)(W + \theta B) &= W^2 + \theta WB + \theta BW + \theta^2 B^2 \\ &= W + \theta^2 B,\end{aligned}$$

ce qui est équivalent à dire que :

$$(W + \theta^2 B)^{1/2} = W + \theta B.$$

Ceci implique que l'estimateur des moindres carrés généralisés peut se ré-écrire :

$$\begin{aligned}b^* &= [(W + \theta B) X]' (W + \theta B) X)^{-1} ((W + \theta B) X)' (W + \theta B) y \\ &= (X^{*'} X^*)^{-1} X^{*'} y^*,\end{aligned}$$

avec

$$X^* = (W + \theta B) X \quad \text{et} \quad y^* = (W + \theta B) y.$$

Remarquons que pour la transformation X^* , le fait qu'il y ait une constante dans la matrice X ne pose pas de problème d'inversibilité contrairement à l'estimateur *within*. En effet, X^* ne comporte pas de colonne nulle puisque :

$$X^* = (W + \theta B) X = (W + \theta B) (e_{NT}, \underline{X}),$$

et que l'on a :

$$W e_{NT} = 0, \quad B e_{NT} = e_{NT},$$

ce qui implique :

$$X^* = (\theta e_{NT}, (W + \theta B) \underline{X}),$$

et la matrice $X^{*'} X^*$ est inversible sous les conditions usuelles (i.e., absence de multicollinéarité). On peut donc garder la notation en X pour la matrice des variables explicatives.

Pour le calcul de la variance de cette estimateur il faut juste faire attention au terme en σ_ε^2 car :

$$\begin{aligned}V(b^* | X) &= (X' \Omega^{-1} X)^{-1} \\ &= \left(X' \frac{1}{\sigma_\varepsilon^2} (W + \theta^2 B) X \right)^{-1} \\ &= \sigma_\varepsilon^2 (X' (W + \theta^2 B) X)^{-1} \\ &= \sigma_\varepsilon^2 (X^{*'} X^*)^{-1}\end{aligned}$$

5.1.2 Expression séparée du terme constant

Pour effectuer le test de Hausman que nous verrons plus loin, il faut comparer l'estimateurs *within* avec celui des moindres carrés généralisés. Comme l'estimateur *within* ne comporte pas de terme constant, il nous faut l'expression de l'estimateur des moindres carrés généralisés séparé du terme constant. L'estimateur des moindres carrés généralisés peut se réécrire comme la solution du système d'équations suivant :

$$X' \Omega^{-1} X b^* = X' \Omega^{-1} y,$$

afin de séparer les composantes, on pose :

$$X = (e_{NT}, \underline{X}), \quad b^* = \begin{pmatrix} a^* \\ \underline{b}^* \end{pmatrix},$$

sans oublier que :

$$\sigma_\varepsilon^2 \Omega^{-1} = W + \theta^2 B, \quad W e_{NT} = 0 \text{ et } B e_{NT} = e_{NT}.$$

On obtient le système suivant :

$$\begin{pmatrix} \theta^2 e'_{NT} e_{NT} & \theta^2 e'_{NT} \underline{X} \\ \theta^2 \underline{X}' e_{NT} & \underline{X}' W \underline{X} + \theta^2 \underline{X}' B \underline{X} \end{pmatrix} \begin{pmatrix} a^* \\ \underline{b}^* \end{pmatrix} = \begin{pmatrix} \theta^2 e'_{NT} y \\ \underline{X}' W y + \theta^2 \underline{X}' B y \end{pmatrix},$$

équivalent à :

$$\begin{cases} e'_{NT} e_{NT} a^* + e'_{NT} \underline{X} \underline{b}^* & = e'_{NT} y \\ \theta^2 \underline{X}' e_{NT} a^* + (\underline{X}' W \underline{X} + \theta^2 \underline{X}' B \underline{X}) \underline{b}^* & = \underline{X}' W y + \theta^2 \underline{X}' B y \end{cases}$$

de la première équation, on déduit :

$$a^* = (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \underline{b}^*),$$

que l'on reporte dans la seconde équation :

$$\theta^2 \underline{X}' B_{NT} (y - \underline{X} \underline{b}^*) + (\underline{X}' W \underline{X} + \theta^2 \underline{X}' B \underline{X}) \underline{b}^* = \underline{X}' W y + \theta^2 \underline{X}' B y,$$

avec $B_{NT} = e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT}$, l'opérateur de moyenne globale. On a finalement :

$$\begin{aligned} \underline{b}^* &= (\underline{X}' (W + \theta^2 (B - B_{NT})) \underline{X})^{-1} \underline{X}' (W + \theta^2 (B - B_{NT})) y \\ &= (\underline{X}' (W + \theta^2 Q_I) \underline{X})^{-1} \underline{X}' (W + \theta^2 Q_I) y \end{aligned}$$

Par rapport à l'estimateur global b^* , il suffit de centrer la matrice *between*. On remarquera que la matrice *within* est, en effet, déjà centrée. La formule est équivalente au cas où l'on aurait centré la transformation du modèle avant d'effectuer les MCO, ce qui élimine le terme constant du modèle ; on obtient alors directement les coefficients des variables \underline{X} .

5.1.3 Cas extrêmes : MCO et Within

Pour une observation de l'individu i à la date t , nous avons transformé la variable expliquée (donc les autres) selon la règle :

$$y_{it}^* = \underbrace{y_{it} - y_{i\bullet}}_{\text{WITHIN}} + \theta \underbrace{y_{i\bullet}}_{\text{BETWEEN}} = y_{it} - (1 - \theta) y_{i\bullet} \quad \text{avec } \theta \in]0, 1[\quad (5.2)$$

Le principe de l'estimation optimale est donc le suivant : quand une perturbation possède une forte variance dans une dimension, on réduit le poids de cette dimension dans l'estimation. Deux cas extrêmes sont particulièrement intéressants :

1. $\sigma_\alpha^2 = 0$.

Ce cas implique que $\alpha_i = 0 \forall i$ car les effets individuels sont d'espérance nulle et qu'une variable aléatoire qui a une variance nulle est en fait une constante égale à son espérance. Dans ce cas, nous avons vu que l'estimateur optimal est celui des moindres carrés ordinaires puisqu'il n'y a plus d'effet individuel et que nous sommes, pour cette raison, dans un modèle à erreur simple. L'estimateur des moindres carrés généralisés revient alors à faire des moindres carrés ordinaires sur la transformation (5.2) avec $\theta = 1$, ce qui donne :

$$y_{it}^* = y_{it} - (1 - 1) y_{i\bullet} = y_{it}$$

On retrouve l'estimateur des moindres carrés ordinaires quand il est optimal.

2. $\sigma_\varepsilon^2 = 0$.

Ce cas implique qu'il n'y a plus d'erreur dans la dimension intra-individuelle (*within*) de sorte que l'estimation dans cette dimension donne une précision maximale. Les moindres carrés généralisés consistent à appliquer la transformation (5.2) avec $\theta = 0$, ce qui donne :

$$y_{it}^* = y_{it} - (1 - 0) y_{i\bullet} = y_{it} - y_{i\bullet}$$

On retrouve l'estimateur *within* quand il est optimal.

Dans le cas général la pondération est comprise entre ces deux bornes et dépend de la précision relative de l'information individuelle par rapport à l'information totale. Plus elle est précise, plus on tendra vers les moindres carrés ordinaires, moins elle est précise plus on tendra vers l'estimation *within*. Le lecteur remarquera que l'on ne trouve jamais l'estimateur *between* comme cas particulier des MCG.

5.1.4 Ecriture comme une moyenne pondérée

Estimateur global

On peut également écrire l'estimateur des moindres carrés généralisés comme une moyenne pondérée des estimateurs *between* et *within*, qui sont indépendants. Ici, il faut faire attention au terme constant à cause de la partie *within* de la transformation optimale. On a :

$$\begin{aligned} b^* &= (X'(W + \theta^2 B)X)^{-1} X'(W + \theta^2 B)y \\ &= (X'(W + \theta^2 B)X)^{-1} X'Wy + (X'(W + \theta^2 B)X)^{-1} \theta^2 X'By, \end{aligned}$$

A ce stade on remarque que :

$$\begin{aligned} X'Wy &= (e_{NT}, \underline{X})' Wy = \begin{pmatrix} e'_{NT} \\ \underline{X}' \end{pmatrix} Wy \quad (5.3) \\ &= \begin{pmatrix} e'_{NT} Wy \\ \underline{X}' Wy \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \underline{X}' Wy \end{pmatrix}, \end{aligned}$$

et que :

$$\begin{aligned} X'WX &= \begin{pmatrix} e'_{NT} \\ \underline{X}' \end{pmatrix} W(e_{NT}, \underline{X}) = \begin{pmatrix} e'_{NT} We_{NT} & e'_{NT} W\underline{X} \\ \underline{X}' We_{NT} & \underline{X}' W\underline{X} \end{pmatrix} \quad (5.4) \\ &= \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & \underline{X}' W\underline{X} \end{pmatrix}, \end{aligned}$$

ce qui implique :

$$\begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & \underline{X}' W\underline{X} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & (\underline{X}' W\underline{X})^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & \mathbf{I}_{p-1} \end{pmatrix},$$

pour faire apparaître les estimateurs *between* et *within*, on intercale des termes en $\underline{X}' W\underline{X}$ et $X'By$, de la manière suivante :

$$\begin{aligned} b^* &= (X'(W + \theta^2 B)X)^{-1} \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & \underline{X}' W\underline{X} \end{pmatrix} \\ &\quad \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & (\underline{X}' W\underline{X})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \underline{X}' Wy \end{pmatrix} \\ &\quad + (X'(W + \theta^2 B)X)^{-1} \theta^2 X'BX (X'BX)^{-1} X'By \\ &= (X'(W + \theta^2 B)X)^{-1} \begin{pmatrix} \mathbf{0}_{1,1} & \mathbf{0}_{1,p-1} \\ \mathbf{0}_{p-1,1} & \underline{X}' W\underline{X} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \hat{b}_W \end{pmatrix} \\ &\quad + (X'(W + \theta^2 B)X)^{-1} \theta^2 X'BX \hat{b}_B, \end{aligned}$$

or nous avons :

$$X'(W + \theta^2 B)X = \begin{pmatrix} 0_{1,1} & 0_{1,p-1} \\ 0_{p-1,1} & \underline{X}'W\underline{X} \end{pmatrix} + \theta^2 X'BX,$$

de sorte que l'on a également :

$$(X'(W + \theta^2 B)X)^{-1} \left(\begin{pmatrix} 0_{1,1} & 0_{1,p-1} \\ 0_{p-1,1} & \underline{X}'W\underline{X} \end{pmatrix} + \theta^2 X'BX \right) = I_p,$$

on peut donc réécrire l'estimateur des moindres carrés généralisés sous la forme :

$$b^* = (I_p - \pi_B) \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} + \pi_B \hat{b}_B,$$

avec :

$$\pi_B = (X'(W + \theta^2 B)X)^{-1} \theta^2 X'BX.$$

La matrice π_B représente la pondération matricielle de l'estimateur *between* dans l'estimateur optimal et $I_p - \pi_B$ celle de l'estimateur *within*.

On retrouve les trois cas limites que nous avons vu précédemment en faisant varier la valeur de θ ou tendre T vers l'infini.

1. Si $\sigma_\alpha^2 = 0$, on a $\theta = 1$, ce qui implique :

$$\pi_B = (X'(W + B)X)^{-1} X'BX = (X'X)^{-1} X'BX,$$

de sorte que :

$$\pi_B \hat{b}_B = (X'X)^{-1} X'By,$$

et que

$$\begin{aligned} (I_p - \pi_B) \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} &= (I_p - (X'X)^{-1} X'BX) \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} \\ &= \left((X'X)^{-1} X'X - (X'X)^{-1} X'BX \right) \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} \\ &= (X'X)^{-1} (X'X - X'BX) \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} \\ &= (X'X)^{-1} X'WX \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix} \text{ d'après (5.4)} \\ &= (X'X)^{-1} \begin{pmatrix} 0_{1,1} & 0_{1,p-1} \\ 0_{p-1,1} & \underline{X}'W\underline{X} \end{pmatrix} \begin{pmatrix} 0_{1,1} \\ \hat{b}_W \end{pmatrix} \\ &= (X'X)^{-1} \begin{pmatrix} 0_{1,p-1} \\ \underline{X}'W\underline{y} \end{pmatrix} \\ &= (X'X)^{-1} X'W\underline{y} \text{ d'après (5.3)} \end{aligned}$$

Globalement, on a donc :

$$\begin{aligned} b^* &= (X'X)^{-1} (X'Wy + X'By) \\ &= (X'X)^{-1} X'y \\ &= \hat{b}, \end{aligned}$$

on retrouve l'estimateur des MCO quand il est optimal.

2. D'autre part lorsque $\sigma_\varepsilon^2 = 0$, on a $\theta = 0$, ce qui implique :

$$\pi_B = 0,$$

de sorte que :

$$b^* = \begin{pmatrix} 0 \\ \hat{b}_W \end{pmatrix},$$

et l'on retrouve l'estimateur *within* quand il est optimal. On remarque également que le terme constant est égal à 0, ce qui est normal puisqu'il n'y a pas de terme constant dans la régression en *within* et qu'exclure une variable d'un modèle revient à mettre son coefficient à 0.

3. Enfin, lorsque $T \rightarrow +\infty$, on a $\theta = 0$ et l'on retrouve l'estimateur *within*.

5.1.5 Estimateur avec terme constant séparé

Dans ce cas l'écriture est plus immédiate. On a :

$$\begin{aligned} \underline{b}^* &= (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \underline{X}'(W + \theta^2(B - B_{NT}))y \\ &= (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \underline{X}'Wy \\ &\quad + (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \theta^2 \underline{X}'(B - B_{NT})y \\ &= (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} (\underline{X}'W\underline{X})(\underline{X}'W\underline{X})^{-1} \underline{X}'Wy \\ &\quad + (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \theta^2 \underline{X}'(B - B_{NT})\underline{X} \\ &\quad (\underline{X}'(B - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})y \\ &= (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} (\underline{X}'W\underline{X})\hat{b}_W \\ &\quad + (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \theta^2 \underline{X}'(B - B_{NT})\underline{X}\hat{b}_B, \end{aligned}$$

on peut donc poser :

$$\underline{\pi}_B = \theta^2 (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X},$$

et l'on peut réécrire l'estimateur optimal comme :

$$\underline{b}^* = (\mathbf{I}_{p-1} - \underline{\pi}_B) \hat{b}_W + \underline{\pi}_B \hat{b}_B. \quad (5.5)$$

Les cas particuliers s'obtiennent alors plus rapidement que pour l'estimateur global :

1. Si $\sigma_\alpha^2 = 0$, on a $\theta = 1$, ce qui implique :

$$\begin{aligned} \underline{\pi}_B &= (\underline{X}'(W + B - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X} \\ &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X}, \end{aligned}$$

ainsi que :

$$\begin{aligned} \mathbf{I}_{p-1} - \underline{\pi}_B &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X} \\ &\quad - (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X} \\ &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(\mathbf{I}_{NT} - B)\underline{X} \\ &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'W\underline{X} \end{aligned}$$

et donc :

$$\begin{aligned} \underline{b}^* &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X} (\underline{X}'(B - B_{NT})\underline{X})^{-1} \\ &\quad \underline{X}'(B - B_{NT})y \\ &\quad + (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'W\underline{X} (\underline{X}'W\underline{X})^{-1} \underline{X}'W y \\ &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(W + B - B_{NT})y \\ &= (\underline{X}'(\mathbf{I}_{NT} - B_{NT})\underline{X})^{-1} \underline{X}'(\mathbf{I}_{NT} - B_{NT})y \\ &= \hat{b}, \end{aligned}$$

on retrouve les moindres carrés ordinaires.

2. Quand $\sigma_\alpha^2 = 0$, on a $\theta = 0$, ce qui implique $\underline{\pi}_B = 0$ et donc :

$$\underline{b}^* = \hat{b}_W,$$

on retrouve l'estimateur *within*.

5.2 Les moindres carrés quasi-généralisés

L'estimation des variances est importante car elle seule permet de calculer les statistiques de test de toutes les méthodes d'estimation que nous venons de voir. De plus, elle permet de procéder à l'estimation optimale du modèle à effets non corrélés. Dans ce chapitre, nous allons donc étudier les propriétés des sommes des carrés des résidus des modèles *between* et *within* avant de détailler l'estimation optimale du modèle à erreur composée.

5.2.1 Variance inter-individuelle (between)

La régression *between* peut s'écrire :

$$By = (BX)b + Bu$$

Le résidu de ce modèle est donné simplement par :²

$$\begin{aligned}\hat{u}_B &= By - BX\hat{b}_B \\ &= \left(By - BX(X'BX)^{-1}X'By \right) \\ &= \left(I_{NT} - BX(X'BX)^{-1}X'B \right) By \\ &= M_B By,\end{aligned}$$

où M_B est la matrice de projection orthogonale sur le sous-espace vectoriel $\text{Im}^\perp(BX)$. On vérifie que cette matrice est symétrique et idempotente :

$$\begin{aligned}M'_B &= I'_{NT} - \left(BX(X'BX)^{-1}X'B \right)' \\ &= I_{NT} - BX(X'BX)^{-1}X'B \\ &= M_B,\end{aligned}$$

$$\begin{aligned}M_B^2 &= \left(I_{NT} - BX(X'BX)^{-1}X'B \right) \left(I_{NT} - BX(X'BX)^{-1}X'B \right) \\ &= I_{NT}^2 - I_{NT}BX(X'BX)^{-1}X'B \\ &\quad - BX(X'BX)^{-1}X'BI_{NT} + BX(X'BX)^{-1} \underbrace{X'BBX}_{X'BX} \\ &\quad (X'BX)^{-1}X'B \\ &= I_{NT} - BX(X'BX)^{-1}X'B - BX(X'BX)^{-1}X'B \\ &\quad + BX(X'BX)^{-1}X'B \\ &= I_{NT} - BX(X'BX)^{-1}X'B \\ &= M_B.\end{aligned}$$

Cette matrice de projection vérifie également la propriété suivante :

$$\begin{aligned}M_B BX &= BX - BX(X'BX)^{-1}X'BBX \\ &= BX - BX \\ &= 0.\end{aligned}$$

2. On utilise le fait que $By = BB_y$.

Le résidu de la régression peut donc s'écrire :

$$\begin{aligned}\hat{u}_B &= M_B B y \\ &= M_B B (Xb + u) \\ &= M_B B X + M_B B u \\ &= M_B B u.\end{aligned}$$

Ceci nous permet d'exprimer la somme des carrés des résidus de la manière suivante :

$$\begin{aligned}S_B &= \begin{matrix} \hat{u}'_B & \hat{u}_B \\ (1, NT) & (NT, 1) \end{matrix} \\ &= u' B' M'_B M_B B u \\ &= u' B M_B B u\end{aligned}$$

Le calcul de l'espérance de cette somme des carrés fournit la solution à notre problème d'estimation :

$$E(S_B | X) = E(\text{tr}(S_B) | X) \quad \text{car } S_B \text{ est un nombre réel,}$$

donc :

$$\begin{aligned}E(S_B | X) &= E(\text{tr}(u' B M_B B u) | X) \\ &= E(\text{tr}(u u' B M_B B) | X) \\ &= E(\text{tr}(B u u' B M_B) | X) \\ &= \text{tr}(E(B u u' B | X) M_B) \\ &= \text{tr}(V(B u | X) M_B) \\ &= \text{tr}(M_B V(B u | X)) \\ &= (\sigma_\varepsilon^2 + T \sigma_\alpha^2) \text{tr}(M_B B),\end{aligned}$$

or

$$\begin{aligned}\text{tr}(M_B B) &= \text{tr}\left(\left(I - B X (X' B X)^{-1} X' B\right) B\right) \\ &= \text{tr} B - \text{tr}\left(X' B X (X' B X)^{-1}\right) \\ &= \text{tr} B - \text{tr} I_p \\ &= \text{tr} B - p\end{aligned}$$

Il nous reste donc à trouver la trace de la matrice *between* :

$$\begin{aligned}
 \text{tr}(B) &= \text{tr}(I_N \otimes B_T) \\
 &= \text{tr}(I_N) \text{tr}(B_T) \\
 &= N \text{tr}\left(e_T (e_T' e_T)^{-1} e_T'\right) \\
 &= N \text{tr}\left(\underbrace{e_T' e_T (e_T' e_T)^{-1}}_1\right) \\
 &= N.
 \end{aligned}$$

Nous pouvons donc écrire que :

$$E(S_B|X) = (\sigma_\varepsilon^2 + T\sigma_\alpha^2)(N - p),$$

ce qui est équivalent à dire que :

$$E\left(\frac{S_B}{N - p} \middle| X\right) = \sigma_\varepsilon^2 + T\sigma_\alpha^2.$$

On estimera cette variance par l'estimateur sans biais :

$$\widehat{\sigma}_B^2 = \frac{\hat{u}_B' \hat{u}_B}{N - p}$$

Nous utiliserons cet estimateur pour estimer θ^2 un peu plus loin.

5.2.2 Variance intra-individuelle (within)

La régression *within* peut s'écrire :

$$Wy = (W\underline{X})\underline{b} + Wu$$

Le résidu de ce modèle est donné simplement par :

$$\hat{u}_W = M_W Wu,$$

avec :

$$M_W = I_{NT} - W\underline{X}(\underline{X}'W\underline{X})^{-1}\underline{X}'W.$$

La démonstration de cette propriété est analogue à celle que nous avons faite pour la somme des carrés des résidus *between*. Il suffit de remplacer B par W . La matrice M_W représente une projection orthogonale sur le sous-espace

vectorel $\text{Im}^\perp(W\underline{X})$ et est donc symétrique et idempotente. Ceci nous permet d'écrire la somme des carrés des résidus *within* :

$$\begin{aligned} S_W &= \hat{u}'_W \hat{u}_W \\ &\quad (1,NT)(NT,1) \\ &= u' W' M'_W M_W W u \\ &= u' W M_W W u \end{aligned}$$

Le calcul de l'espérance de cette somme des carrés fournit la solution à notre problème d'estimation :

$$E(S_W|X) = E(\text{tr}(S_W)|X) \quad \text{car } S_W \text{ est un nombre réel,}$$

donc :

$$\begin{aligned} E(S_W|X) &= E(\text{tr}(u' W M_W W u)|X) \\ &= E(\text{tr}(u u' W M_W W)|X) \\ &= E(\text{tr}(M_W W u u' W')|X) \\ &= \text{tr}(M_W E(W u u' W'|X)) \\ &= \text{tr}(M_W V(W u|X)), \end{aligned}$$

or :

$$\begin{aligned} V(W u|X) &= W (\sigma_\varepsilon^2 I_{NT} + (\sigma_\varepsilon^2 + T\sigma_\alpha^2) B) W \\ &= \sigma_\varepsilon^2 W, \end{aligned}$$

ce qui implique :

$$E(S_W|X) = \sigma_\varepsilon^2 \text{tr}(M_W W).$$

Il nous reste donc à trouver la trace de la dernière matrice :

$$\begin{aligned} \text{tr}(M_W W) &= \text{tr}\left(\left(I_{NT} - W\underline{X}(\underline{X}'W\underline{X})^{-1}\underline{X}'W\right)W\right) \\ &= \text{tr}(W) - \text{tr}\left(W\underline{X}(\underline{X}'W\underline{X})^{-1}\underline{X}'W\right) \\ &= \text{tr}(W) - \text{tr}\left(\underline{X}'W\underline{X}(\underline{X}'W\underline{X})^{-1}\right) \\ &= \text{tr}(W) - \text{tr}(I_{p-1}) \\ &= \text{tr}(W) - (p-1). \end{aligned}$$

et :

$$\begin{aligned}
\text{tr}(W) &= \text{tr}(\mathbf{I}_{NT} - B) \\
&= \text{tr}(\mathbf{I}_{NT}) - \text{tr}(B) \\
&= NT - N \\
&= N(T - 1)
\end{aligned}$$

Nous pouvons donc écrire que :

$$\begin{aligned}
E(S_W|X) &= \sigma_\varepsilon^2 (N(T - 1) - (p - 1)) \\
&= \sigma_\varepsilon^2 (NT - N - p + 1)
\end{aligned}$$

ce qui est équivalent à dire que :

$$E\left(\frac{S_W}{NT - N - p + 1} \middle| X\right) = \sigma_\varepsilon^2.$$

On estimera cette variance par l'estimateur sans biais :

$$\widehat{\sigma}_W^2 = \frac{\hat{u}'_W \hat{u}_W}{NT - N - p + 1}.$$

5.2.3 Estimation des composantes de la variance

On estime les composantes de la variance par la méthode des moments. Nous avons montré que :

$$\begin{cases} E(\widehat{\sigma}_B^2 | X) = \sigma_\varepsilon^2 + T\sigma_\alpha^2 \\ E(\widehat{\sigma}_W^2 | X) = \sigma_\varepsilon^2 \end{cases}$$

En résolvant ce système par rapport à σ_ε^2 et σ_α^2 , on obtient le résultat suivant :

$$\begin{cases} \sigma_\alpha^2 = E\left(\frac{1}{T}(\widehat{\sigma}_B^2 - \widehat{\sigma}_W^2) \middle| X\right) \\ \sigma_\varepsilon^2 = E(\widehat{\sigma}_W^2 | X) \end{cases}$$

Comme le système ci-dessus est linéaire, les estimateurs suivants sont sans biais :

$$\widehat{\sigma}_\alpha^2 = \max\left(0, \frac{1}{T}(\widehat{\sigma}_B^2 - \widehat{\sigma}_W^2)\right) \quad \text{et} \quad \widehat{\sigma}_\varepsilon^2 = \widehat{\sigma}_W^2.$$

Ici, on voit qu'il n'est pas impossible a priori que $\widehat{\sigma}_B^2 - \widehat{\sigma}_W^2$ soit négatif ou nul sur des échantillons très particuliers (i.e., il ne peut pas être négatif en

moyenne). Dans ce cas, la solution est simple. On peut montrer que l'on se trouve dans le cas $\sigma_\alpha^2 = 0$ et qu'il faut donc appliquer les moindres carrés ordinaires puisqu'il n'y a pas d'effet individuel significatif (voir le test de Fisher, plus loin). Dans le cas inverse, le plus fréquent, on utilise les deux estimateurs précédents pour calculer la pondération qui correspond aux moindres carrés généralisés.

5.2.4 Estimation du paramètre

Dans la pratique la somme des carrés *between* est toujours calculée sans duplication. Ceci implique qu'elle est T fois plus petite que la somme des carrés données dans la section précédente. On note \tilde{S}_B la somme des carrés non dupliquée, définie par :

$$S_B = T \times \tilde{S}_B$$

On en déduit que :

$$\begin{aligned} E\left(\frac{T \times \tilde{S}_B}{N-p} \middle| X\right) &= \sigma_\varepsilon^2 + T\sigma_\alpha^2 \\ \Leftrightarrow E\left(\frac{\tilde{S}_B}{N-p} \middle| X\right) &= \frac{\sigma_\varepsilon^2}{T} + \sigma_\alpha^2, \end{aligned}$$

ce qui permet de rédéfinir la variance de la régression *between non dupliquée* comme :

$$\tilde{\sigma}_B^2 = \frac{\tilde{S}_B}{N-p} = \frac{1}{T} \widehat{\sigma}_B^2.$$

Avec cette nouvelle définition, on a :

$$\begin{aligned} E\left(\tilde{\sigma}_B^2 \middle| X\right) &= \frac{\sigma_\varepsilon^2}{T} + \sigma_\alpha^2 \\ \Leftrightarrow \sigma_\alpha^2 &= E\left(\tilde{\sigma}_B^2 - \frac{\sigma_\varepsilon^2}{T} \middle| X\right), \end{aligned}$$

de sorte que l'on peut définir l'estimateur de la variance de l'effet individuel de la manière suivante :

$$\widehat{\sigma}_\alpha^2 = \max\left(0, \tilde{\sigma}_B^2 - \frac{\widehat{\sigma}_W^2}{T}\right).$$

5.2.5 Application des moindres carrés quasi-généralisés

On pose :

$$\widehat{\theta^2} = \frac{\widehat{\sigma_\varepsilon^2}}{\widehat{\sigma_\varepsilon^2} + T\widehat{\sigma_\alpha^2}} = \frac{\widehat{\sigma_W^2}}{\widehat{\sigma_B^2}} = \frac{\widehat{\sigma_W^2}}{T \times \widehat{\sigma_B^2}},$$

cet estimateur n'est pas sans biais car l'espérance du ratio n'est pas égale au ratio des espérances. Par contre, il est convergent. De même, on utilise l'estimateur convergent suivant pour calculer les transformations menant à l'estimateur optimal :

$$\hat{\theta} = \sqrt{\widehat{\theta^2}} = \frac{\widehat{\sigma_W}}{\widehat{\sigma_B}}.$$

Appliquer les moindres carrés quasi-généralisés revient à remplacer θ par son estimateur convergent $\hat{\theta}$ dans le but d'effectuer les transformations suivantes :

$$\tilde{y}_{it} = y_{it} - (1 - \hat{\theta}) y_{i\bullet}, \quad \tilde{X}_{it} = X_{it} - (1 - \hat{\theta}) X_{i\bullet},$$

puis on applique les moindres carrés ordinaires aux données ainsi transformées. Aucune correction supplémentaire n'est nécessaire. Notons que nous employons la notation \tilde{y}_{it} au lieu de y_{it}^* car nous utilisons $\hat{\theta}$ au lieu de θ . Ce remplacement justifie la terminologie de moindres carrés *quasi* généralisés.³ L'estimateur est défini par :

$$\tilde{b} = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{y} \text{ et } \widehat{V}(\tilde{b}) = \widehat{\sigma_\varepsilon^2} (\tilde{X}'\tilde{X})^{-1}.$$

1. Remarquons que la transformation doit également être appliquée au terme constant. On a :

$$\tilde{e}_{it} = 1 - (1 - \hat{\theta}) 1 = \hat{\theta},$$

ce qui implique que l'on régresse \tilde{y} sur $\tilde{X} = (\hat{\theta} e_{NT}, \tilde{X})$ et non sur (e_{NT}, \tilde{X}) .

5.3 Programmation sous IML

Comme les MCQG sont une moyenne pondérée des estimations *between* et *within*, on reprend le programme précédent et l'on ajoute juste les calculs de $\widehat{\sigma_\alpha^2}$, $\widehat{\sigma_\varepsilon^2}$ et $\hat{\theta}$. On estime ensuite le modèle par les MCQG. On utilise donc la formulation :

$$\tilde{y}_{it} = y_{it} - y_{i\bullet} + \hat{\theta} y_{i\bullet},$$

3. La terminologie employée vient du fait que l'on peut interpréter cet estimateur comme un cas particulier du pseudo maximum de vraisemblance quasi-généralisé, sous hypothèse de pseudo loi normale.

Le programme est le suivant (le début est identique au programme précédent).

Programme 5.1.

```

proc iml;
ny={LogS}; nx={LogL LogC LogK}; nx=t(nx);
use between;
  read all var (ny) into by;
  read all var (nx) into bx;
use within;
  read all var (ny) into wy;
  read all var (nx) into wx;
nt=nrow(wy); n=nrow(by); t=nt/n;
bx=J(n,1,1)||bx;
nxc="Constante"/nx;
p=ncol(bx);
print "Estimation sur panel cylindré",
"Nombre d'observations =" nt,
"Nombre d'individus =" n,
"Nombre de dates =" t;
start mco(y,x) global(ixx);
ixx=inv(t(x)*x);
b=ixx*t(x)*y;
return (b);
finish mco;
bet=mco(by,bx);
ub=by-bx*bet;
vub=ssq(ub)/(n-p);
vbet=vub#ixx;
sbet=sqrt(vecdiag(vbet));
tbet=abs(bet)/sbet;
print "Estimation Between",
"Variable expliquée :" ny, bet [rowname=nxc] sbet
tbet;
wit=mco(wy,wx);
uw=wy-wx*wit;
vuw=ssq(uw)/(n*(t-1)-(p-1));/* attention */
vwit=vuw#ixx;
swit=sqrt(vecdiag(vwit));
twit=abs(wit)/swit;

```

```

print "Estimation Within",
"Variable expliquée :" ny, wit [rowname=nx] swit twit;
/* nouveau */
se2=vuw;
sa2=max(0,vub-se2/t);
theta2=se2/(se2+t#sa2);
theta=sqrt(theta2);
et=J(T,1,1);
oy=wy+theta#(by@et);
ox=(J(nt,1,0)||wx)+theta#(bx@et);
mcg=mco(oy,ox);
vmcg=se2#ixx;
smcg=sqrt(vecdiag(vmcg));
tmcg=abs(mcg)/smcg;
print "Variance idiosyncratique =" se2,
"Variance de l'effet individuel =" sa2,
"Theta au carré =" theta2, "Theta =" theta,,
"Estimation des MCQG",
"Variable expliquée :" ny, mcg [rowname=nxc] smcg
tmcg;
quit; run;

```

Les nouvelles lignes du programme sont :

1. `se2=vuw;`
On renomme la variable `vuw` en `se2`. Il s'agit d'un estimateur de $\widehat{\sigma_\varepsilon^2}$.
2. `sa2=max(0,vub-se2/t);`
On calcule $\widehat{\sigma_\alpha^2}$ que l'on met dans la variable `sa2`. Dans le cas, rarissime ou cette estimation est négative, on met sa valeur à 0, ce qui revient à estimer le modèle par les MCO. On utilise la formule sans duplication.
3. `theta2=se2/(se2+t#sa2);`
Calcul de $\widehat{\theta^2}$.
4. `theta=sqrt(theta2);`
Calcul de $\widehat{\theta}$.
5. `et=J(T,1,1);`
Calcul de e_T .
6. `oy=wy+theta#(by@et);`
Calcul de la transformation optimale de la variable expliquée. Comme chaque observation de la matrice `between by` n'est pas dupliquée T fois

pour chaque individu, on calcule $(by \otimes e_T)$. Le produit de Kronecker est noté @ sous SAS IML.

7. `ox=(J(nt,1,0)||wx)+theta#(bx@et);`
Calcul des transformations optimales des variables explicatives (y compris du terme constant). On commence par ajouter une colonne nulle à la matrice `wx` pour tenir compte de l'absence de terme constant en *within*, on procède ensuite comme pour la variable expliquée.
8. `mcg=mco(oy,ox);`
Calcul de l'estimateur des MCQG comme l'application des MCO aux données transformées.
9. `vmcg=se2#ixx;`
Calcul de la variance de l'estimateur des MCQG, on remarque que l'on utilise $\widehat{\sigma_\varepsilon^2}$.
10. `smcg=sqrt(vecdiag(vmcg));`
Estimation des écarts types des estimateurs.
11. `tmcg=abs(mcg)/smcg;`
Estimation des *t* de Student correspondants.
12. `print "Variance idiosyncratique =" se2,
"Variance de l'effet individuel =" sa2,
"Theta au carré =" theta2, "Theta =" theta,
"Estimation des MCQG", "Variable expliquée :" ny,
mcg [rowname=nxc] smcg tmcg;`
Impression des résultats.

On obtient les résultats supplémentaires suivants par rapport à la sortie du chapitre précédent :

Sortie 5.1.

```

SE2
Variance idiosyncratique = 0.0531941
SA2
Variance de l'effet individuel = 0.1180988
THETA2
Theta au carré = 0.0476614
THETA

```

Theta = 0.218315

Estimation des MCQG
NY

Variable expliquée : LOGS

MCG		SMCG	TMCG
Constante	3.39639	0.0305546	111.15815
LOGL	0.6634851	0.0100399	66.085077
LOGC	0.2607767	0.0084156	30.987135
LOGK	0.1142305	0.006385	17.890375

Chapitre 6

Les tests de spécification

Il existe deux types de test principaux que l'on pratique sur données de panel. Le premier test, de Fisher, vise à vérifier s'il existe un effet aléatoire individuel significatif. Le second type de test, de Hausman et de Mundlak, vise à vérifier s'il existe un effet individuel corrélé avec les variables explicatives et sert donc au choix de l'estimateur. Trois cas peuvent se présenter.

1. S'il n'y a pas d'effet individuel significatif, on utilisera les moindres carrés ordinaires.
2. Si l'effet est significatif et n'est pas corrélé, on utilisera les moindres carrés quasi-généralisés.
3. Si l'effet est significatif et corrélé, on utilisera l'estimateur *within*.

6.1 Le test de Fisher

On part du modèle à erreur composée et l'on souhaite savoir s'il existe un effet individuel significatif. Le problème de test (unilatéral) est donc le suivant :

$$\begin{cases} H_0 : \sigma_\alpha^2 = 0 \\ H_a : \sigma_\alpha^2 > 0 \end{cases}$$

Notons qu'on peut également le réécrire sous la forme suivante :

$$\begin{cases} H_0 : \rho_u = 0 \\ H_a : \rho_u > 0 \end{cases}$$

avec :

$$\rho_u = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} \geq 0.$$

L'hypothèse alternative est ici équivalente à dire que la variance de la perturbation *between* est supérieure à la variance de la perturbation *within*. On peut donc proposer une statistique de test qui se base sur les résidus des régressions *between* et *within*. Les deux estimateurs étant des fonctions affines de la perturbation u , la normalité de u impliquerait d'une part la normalité de ces estimateurs et d'autre part, le fait que les sommes des carrés des résidus correspondantes suivent des lois du khi-deux. On peut également considérer que les propriétés qui suivent sont valables quand $N \rightarrow +\infty$ dans le cas où les perturbations ne sont pas normales. D'après ce que nous avons vu dans les sections précédentes :

$$(N-p) \frac{\widehat{\sigma}_B^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} \rightsquigarrow \chi^2(N-p),$$

$$(N(T-1) - (p-1)) \frac{\widehat{\sigma}_W^2}{\sigma_\varepsilon^2} \rightsquigarrow \chi^2(N(T-1) - (p-1)),$$

or l'orthogonalité des opérateurs *between* et *within* implique que ces deux quantités sont indépendantes.¹ En conséquence, le rapport des ces quantités divisées par leurs degrés de liberté respectifs suit une loi de Fisher :

$$F = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2} \times \frac{\widehat{\sigma}_B^2}{\widehat{\sigma}_W^2} \rightsquigarrow F(N-p, N(T-1) - (p-1)),$$

sous l'hypothèse nulle $H_0 : \sigma_\alpha^2 = 0$, on a donc la statistique suivante :

$$F_0 = \frac{\widehat{\sigma}_B^2}{\widehat{\sigma}_W^2} \stackrel{H_0}{\rightsquigarrow} F(N-p, N(T-1) - (p-1)),$$

cette statistique tend vers 1 s'il n'y a pas d'effet individuel, et augmente si $\sigma_\alpha^2 > 0$. Pour un test au seuil β , on rejette l'hypothèse nulle si :

$$F_0 > F_{N-p, N(T-1)-(p-1), 1-\beta}$$

6.2 Le test de Hausman

Le test de Hausman repose sur une comparaison directe d'estimateurs. Pour pouvoir le réaliser, il faut d'une part un estimateur convergent et asymptotiquement efficace (i.e., à variance minimale quand $NT \rightarrow +\infty$) sous l'hypothèse nulle et d'autre part, un estimateur convergent aussi bien sous l'hypothèse nulle que sous l'hypothèse alternative. L'hypothèse que l'on veut tester

1. Cette propriété est démontrée dans la section sur le test de Mundlak.

est celle d'absence de corrélation entre l'effet individuel et les variables explicatives. Pour mieux poser cette hypothèse nulle, il faut examiner les conséquences de cette corrélation sur l'estimateur des moindres carrés généralisés. Le modèle à effet corrélé peut s'écrire :²

$$y = Xb + u, \text{ avec } u = \varepsilon + \alpha$$

et l'estimateur des moindres carrés généralisés est égal à :

$$b^* = b + (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}u,$$

en développant l'expression de la perturbation, on a :

$$\begin{aligned} b^* - b &= (X'(W + \theta^2 B)X)^{-1} X'(W + \theta^2 B)\varepsilon \\ &\quad + (X'(W + \theta^2 B)X)^{-1} X'(W + \theta^2 B)\alpha \end{aligned}$$

or ε est un bruit blanc indépendant aussi bien des moyennes individuelles des variables explicatives que des écarts des variables explicatives à leurs moyennes individuelles, ce qui implique que :

$$\text{Plim} \frac{1}{NT} X' B \varepsilon = \text{Plim} \frac{1}{NT} X' W \varepsilon = 0.$$

La seconde composante de la perturbation est constante dans le temps donc elle est bien indépendante des écarts des variables explicatives à leurs moyennes individuelles. Numériquement, on a $W\alpha = 0$. Par contre les effets individuels ne sont pas indépendants des moyennes individuelles puisque l'on a $B\alpha = \alpha$. On a donc :

$$\begin{aligned} \text{Plim} \frac{1}{NT} X' B \alpha &= \text{Plim} \frac{1}{NT} X' \alpha \neq 0 \\ \text{et } \text{Plim} \frac{1}{NT} X' W \alpha &= \text{Plim} \frac{1}{NT} X' 0 = 0. \end{aligned}$$

L'estimateur des moindres carrés généralisés n'est plus convergent en présence d'effets individuels corrélés puisque :³

$$\begin{aligned} \text{Plim} b^* - b &= \theta^2 \text{Plim} \left(\frac{X'\Omega^{-1}X}{NT} \right)^{-1} \times \text{Plim} \frac{1}{NT} X' \alpha \\ &= \theta^2 \text{Vas} \left(\sqrt{NT} b^* \right)^{-1} \text{Plim} \frac{1}{NT} X' \alpha \\ &\neq 0. \end{aligned}$$

2. Dans ce qui suit, on peut considérer aussi bien une corrélation entre un effet aléatoire et les variables explicatives qu'entre un effet fixe et les variables explicatives.

3. La notation Vas désigne la variance asymptotique, c'est-à-dire de la distribution vers laquelle tend la variable aléatoire étudiée lorsque $NT \rightarrow +\infty$.

Le test de Hausman repose donc sur l'hypothèse nulle suivante :

$$\begin{cases} H_0 : \theta^2 \text{Vas}(\sqrt{NT}b^*)^{-1} \text{Plim} \frac{1}{NT} X' \alpha = 0 \\ H_a : \theta^2 \text{Vas}(\sqrt{NT}b^*)^{-1} \text{Plim} \frac{1}{NT} X' \alpha \neq 0 \end{cases}$$

L'hypothèse nulle se simplifie dans le cas usuel où la matrice $\text{Vas}(\sqrt{NT}b^*)^{-1}$ est inversible puisque l'on a :

$$\begin{aligned} \theta^2 > 0, \quad \text{Vas}(\sqrt{NT}b^*)^{-1} \text{ de plein rang} \\ \text{et } \theta^2 \text{Vas}(\sqrt{NT}b^*)^{-1} \text{Plim} \frac{1}{NT} X' \alpha = 0 \\ \Rightarrow \text{Plim} \frac{1}{NT} X' \alpha = 0. \end{aligned}$$

On réécrit donc souvent le problème du test de Hausman comme :

$$\begin{cases} H_0 : \text{Plim} \frac{1}{NT} X' \alpha = 0 \\ H_a : \text{Plim} \frac{1}{NT} X' \alpha \neq 0 \end{cases}$$

Si l'hypothèse nulle est vérifiée, l'estimateur des moindres carrés quasi généralisés est convergent et asymptotiquement efficace. Dans le cas contraire, il n'est plus convergent. D'autre part, l'estimateur *within* est toujours convergent, aussi bien sous l'hypothèse nulle que sous l'hypothèse alternative.

La statistique de test est donc simple à calculer car elle repose sur une comparaison directe des estimateurs *within* et des moindres carrés généralisés. Comme chacun de ces estimateurs est asymptotiquement normal, on construit une statistique du khi-deux.

Toutefois, il faut faire attention au point suivant : il n'y a pas d'estimateur du terme constant dans la régression *within*. On utilise donc uniquement les $p - 1$ paramètres qui correspondent aux autres variables explicatives, c'est-à-dire les estimateurs \hat{b}_W et \underline{b}^* . Si $\hat{\delta}$ est une variable aléatoire normale centrée de $p - 1$ composantes indépendantes, la quantité $\hat{\delta}' V(\hat{\delta})^{-1} \hat{\delta}$ converge en loi vers une distribution du khi-deux à $p - 1$ degrés de liberté. Ici, on pose :

$$\hat{\delta}_{W*} = \hat{b}_W - \underline{b}^*,$$

la variance de cette quantité s'écrit :

$$\begin{aligned} V(\hat{\delta}_{W*} | X) &= V(\hat{b}_W - \underline{b}^* | X) = V(\hat{b}_W | X) + V(\underline{b}^* | X) \\ &\quad - 2\text{Cov}(\hat{b}_W, \underline{b}^* | X). \end{aligned}$$

Le seul terme à calculer est la covariance entre les deux estimateurs. Pour cela on utilise leur réécriture en fonction de la perturbation :

$$\begin{aligned}\hat{\underline{b}}_W &= b + (\underline{X}'W\underline{X})^{-1} \underline{X}'Wu \quad \text{et} \\ \underline{b}^* &= b + (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \underline{X}'(W + \theta^2(B - B_{NT}))u\end{aligned}$$

ce qui implique :

$$\begin{aligned}\text{Cov}(\hat{\underline{b}}_W, \underline{b}^* | X) &= E\left((\hat{\underline{b}}_W - b)(\underline{b}^* - b)' | X\right) \\ &= E\left((\underline{X}'W\underline{X})^{-1} \underline{X}'Wuu'(W + \theta^2(B - B_{NT}))\underline{X}\right. \\ &\quad \left. (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} | X\right) \\ &= (\underline{X}'W\underline{X})^{-1} \underline{X}'WE(uu' | X)(W + \theta^2(B - B_{NT}))\underline{X} \\ &\quad (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1}\end{aligned}$$

avec

$$\begin{aligned}WE(uu') &= W(\sigma_\varepsilon^2 W + (\sigma_\varepsilon^2 + T\sigma_\alpha^2)B) \\ &= \sigma_\varepsilon^2 W,\end{aligned}$$

ce qui implique :

$$\begin{aligned}\text{Cov}(\hat{\underline{b}}_W, \underline{b}^* | X) &= \sigma_\varepsilon^2 (\underline{X}'W\underline{X})^{-1} \underline{X}'W(W + \theta^2(B - B_{NT}))\underline{X} \\ &\quad (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1},\end{aligned}$$

or :

$$W(B - B_{NT}) = \mathbf{0},$$

car $WB = 0$ et $We_{NT} = 0$. Il reste donc :

$$\begin{aligned}\text{Cov}(\hat{\underline{b}}_W, \underline{b}^* | X) &= \sigma_\varepsilon^2 (\underline{X}'W\underline{X})^{-1} \underline{X}'W\underline{X}(\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \\ &= \sigma_\varepsilon^2 (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \\ &= V(\underline{b}^* | X),\end{aligned}$$

on en déduit que :

$$\begin{aligned}V(\hat{\delta}_{W^*} | X) &= V(\hat{\underline{b}}_W | X) + V(\underline{b}^* | X) - 2V(\underline{b}^* | X) \\ &= V(\hat{\underline{b}}_W | X) - V(\underline{b}^* | X).\end{aligned}$$

La statistique du test de Hausman est donc donnée par :

$$\begin{aligned} H &= \hat{\delta}'_{W*} V(\hat{\delta}_{W*} | X)^{-1} \hat{\delta}_{W*} \\ &= (\hat{\underline{b}}_W - \underline{b}^*)' [V(\hat{\underline{b}}_W | X) - V(\underline{b}^* | X)]^{-1} (\hat{\underline{b}}_W - \underline{b}^*) \xrightarrow{H_0} \chi^2(p-1), \end{aligned}$$

mais cette forme n'est pas très pratique car \underline{b}^* dépend des paramètres théoriques de variance. Ce point ne pose toutefois pas de problème car \underline{b}^* et $\tilde{\underline{b}}$ ont la même distribution asymptotique. La statistique que l'on utilisera dans la pratique est donc :

$$\tilde{H} = (\hat{\underline{b}}_W - \tilde{\underline{b}})' [V(\hat{\underline{b}}_W | X) - V(\tilde{\underline{b}} | X)]^{-1} (\hat{\underline{b}}_W - \tilde{\underline{b}}) \xrightarrow{H_0} \chi^2(p-1).$$

6.3 Le test de Mundlak

6.3.1 Test direct

On utilise ici le fait que l'estimateur *within* est convergent sous l'hypothèse nulle et sous l'hypothèse alternative, alors que l'estimateur *between* n'est convergent que sous l'hypothèse nulle. On peut donc baser un test sur l'écart de ces deux estimateurs. On compare directement les estimateurs *between* et *within*, sans terme constant, ce qui mène à la comparaison suivante :

$$\hat{\delta}_{WB} = \hat{\underline{b}}_W - \hat{\underline{b}}_B.$$

On a :

$$\begin{aligned} \hat{\underline{b}}_W &= \underline{b} + (\underline{X}' W \underline{X})^{-1} \underline{X}' W u \\ \text{et } \hat{\underline{b}}_B &= (\underline{X}' (B - B_{NT}) \underline{X})^{-1} \underline{X}' (B - B_{NT}) u, \end{aligned}$$

ce qui implique :

$$\begin{aligned} \text{Cov}(\hat{\underline{b}}_W, \hat{\underline{b}}_B | X) &= E\left((\hat{\underline{b}}_W - \underline{b})(\hat{\underline{b}}_B - \underline{b})' | X\right) \\ &= E\left((\underline{X}' W \underline{X})^{-1} \underline{X}' W u u' (B - B_{NT}) \underline{X}\right. \\ &\quad \left. (\underline{X}' (B - B_{NT}) \underline{X})^{-1} | X\right) \\ &= (\underline{X}' W \underline{X})^{-1} \underline{X}' W E(u u' | X) (B - B_{NT}) \underline{X} \\ &\quad (\underline{X}' (B - B_{NT}) \underline{X})^{-1} \\ &= (\underline{X}' W \underline{X})^{-1} \underline{X}' W \Omega (B - B_{NT}) \underline{X} (\underline{X}' (B - B_{NT}) \underline{X})^{-1} \end{aligned}$$

or

$$W\Omega(B - B_{NT}) = W(\sigma_\varepsilon^2 I_{NT} + T\sigma_\alpha^2 B)(B - B_{NT}) = 0$$

En conséquence :

$$\text{Cov}(\hat{\underline{b}}_W, \hat{\underline{b}}_B | X) = 0.$$

D'où la statistique du test :

$$\begin{aligned} M &= \hat{\delta}'_{WB} V(\hat{\delta}_{WB} | X)^{-1} \hat{\delta}_{WB} \\ &= (\hat{\underline{b}}_W - \hat{\underline{b}}_B)' [V(\hat{\underline{b}}_W | X) + V(\hat{\underline{b}}_B | X)]^{-1} (\hat{\underline{b}}_W - \hat{\underline{b}}_B) \stackrel{H_0}{\rightsquigarrow} \chi^2(p-1). \end{aligned} \quad (6.1)$$

Notons qu'ici on *additionne* les variances au lieu de prendre leur différence. Ce résultat provient de l'indépendance des deux estimateurs.⁴

6.3.2 Test par régression auxiliaire

Le test de Mundlak est particulièrement pratique à mettre en oeuvre car on peut le réécrire facilement sous forme d'un test de nullité des coefficients d'une régression auxiliaire.⁵ Il suffit d'estimer la relation suivante :

$$y = e_{NT}a + \underline{X}c + (B\underline{X})d + v, \quad (6.2)$$

où v est une perturbation qui dépend de u . Ici, on remarque qu'il faut impérativement séparer le terme constant e_{NT} des autres variables explicatives du modèle \underline{X} car $Be_{NT} = e_{NT}$, de sorte que l'on retrouve le même vecteur en niveau et en moyennes individuelles. De ce fait, si l'on ne séparait pas le terme constant des autres variables, on aurait deux fois la même variable explicative e_{NT} dans le membre de droite de l'équation, ce qui donnerait un modèle parfaitement colinéaire. Ce modèle peut se réécrire :

$$y = Z\gamma + v \quad \text{avec} \quad Z = (e_{NT}, \underline{X}, B\underline{X}) \quad \text{et} \quad \gamma = \begin{pmatrix} a \\ c \\ d \end{pmatrix}$$

L'estimateur des moindres carrés ordinaires de ce modèle est donné par la solution du système d'équations suivant :

$$(Z'Z) \begin{pmatrix} \hat{a} \\ \hat{c} \\ \hat{d} \end{pmatrix} = Z'y,$$

4. En effet, les estimateurs *within* et des moindres carrés généralisés ne sont pas indépendants, ce qui explique que l'on ne prenne pas la somme de leurs variances.

5. On parle de régression auxiliaire car son objectif est de calculer une statistique de test, et non de faire une régression à proprement parler.

or, en utilisant $Be_{NT} = e_{NT}$, les matrices des produits croisés se simplifient en :

$$Z'Z = \begin{pmatrix} e'_{NT}e_{NT} & e'_{NT}\underline{X} & e'_{NT}\underline{X} \\ \underline{X}'e_{NT} & \underline{X}'\underline{X} & \underline{X}'B\underline{X} \\ \underline{X}'e_{NT} & \underline{X}'B\underline{X} & \underline{X}'B\underline{X} \end{pmatrix} \quad \text{et} \quad Z'y = \begin{pmatrix} e'_{NT}y \\ \underline{X}'y \\ \underline{X}'By \end{pmatrix},$$

le système se ramène donc à :

$$\begin{cases} e'_{NT}e_{NT} \hat{a} + e'_{NT}\underline{X} \hat{c} + e'_{NT}\underline{X} \hat{d} = e'_{NT}y \\ \underline{X}'e_{NT} \hat{a} + \underline{X}'\underline{X} \hat{c} + \underline{X}'B\underline{X} \hat{d} = \underline{X}'y \\ \underline{X}'e_{NT} \hat{a} + \underline{X}'B\underline{X} \hat{c} + \underline{X}'B\underline{X} \hat{d} = \underline{X}'By \end{cases}$$

La première équation de ce système donne :

$$\hat{a} = (e'_{NT}e_{NT})^{-1} e'_{NT} (y - \underline{X} \hat{c} - \underline{X} \hat{d}),$$

en reportant ce résultat dans les deuxième et troisième équations, on obtient le sous système qui nous intéresse en (\hat{c}, \hat{d}) :

$$\begin{cases} \underline{X}'(I_{NT} - B_{NT})\underline{X} \hat{c} + \underline{X}'(B - B_{NT})\underline{X} \hat{d} = \underline{X}'(I_{NT} - B_{NT})y \\ \underline{X}'(B - B_{NT})\underline{X} \hat{c} + \underline{X}'(B - B_{NT})\underline{X} \hat{d} = \underline{X}'(B - B_{NT})y \end{cases}$$

la dernière équation de ce sous-système donne directement :

$$\begin{aligned} \underline{X}'(B - B_{NT})\underline{X}(\hat{c} + \hat{d}) &= \underline{X}'(B - B_{NT})y \\ \Leftrightarrow \hat{c} + \hat{d} &= (\underline{X}'(B - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})y \\ \Leftrightarrow \hat{c} + \hat{d} &= \hat{b}_B, \end{aligned}$$

et en faisant la différence entre la première et la seconde équation du sous-système, on obtient :

$$\begin{aligned} \underline{X}'(I_{NT} - B)\underline{X} \hat{c} &= \underline{X}'(I_{NT} - B)y \\ \Leftrightarrow \underline{X}'W\underline{X} \hat{c} &= \underline{X}'Wy \\ \Leftrightarrow \hat{c} &= (\underline{X}'W\underline{X})^{-1} \underline{X}'Wy = \hat{b}_W. \end{aligned}$$

Dans l'ensemble, on obtient :

$$\hat{c} = \hat{b}_W \quad \text{et} \quad \hat{d} = \hat{b}_B - \hat{b}_W.$$

On peut donc effectuer un test de Wald sur \hat{d} qui est numériquement identique au test (6.1). De plus quand $N \rightarrow +\infty$, ce test est équivalent à un test de Fisher de nullité jointe du coefficient d . Ceci revient à tester la significativité des

moyennes individuelles dans la régression auxiliaire (6.2). Notons bien que le terme constant n'intervient pas dans ce test.

Cette méthode possède un avantage par rapport à l'application directe du test de Mundlak : en cas de rejet de l'hypothèse nulle elle fournit automatiquement les estimations *within*, qui sont justement valables dans ce cas. Notons que les estimations *within* sont obtenues comme les coefficients des variables explicatives *en niveaux*. Ce résultat provient du fait que les transformations orthogonales *between* et *within* permettent de décomposer tous les vecteurs de \mathbb{R}^{NT} de manière unique. Enfin, si l'on résoud en \hat{a} , on obtient l'estimateur suivant du terme constant :

$$\hat{a} = (e'_{NT} e_{NT})^{-1} e'_{NT} (y - \underline{X} \hat{\underline{b}}_B) = \hat{a}_B,$$

qui n'est autre que l'estimateur *between* du terme constant. Nous verrons qu'il existe un meilleur estimateur du terme constant dans la section sur le calcul des prévisions.

6.4 Equivalence entre les tests de Hausman et Mundlak

Les tests de Hausman et de Mundlak sont numériquement identiques. Pour s'en convaincre, il suffit de remarquer que l'estimateur des moindres carrés généralisés s'écrit (relation 5.5) :

$$\underline{b}^* = \underline{\pi}_B \hat{\underline{b}}_B + (I_{p-1} - \underline{\pi}_B) \hat{\underline{b}}_W$$

On en déduit la propriété suivante :

$$\begin{aligned} \hat{\underline{b}}_W - \underline{b}^* &= \hat{\underline{b}}_W - [\underline{\pi}_B \hat{\underline{b}}_B + (I_{p-1} - \underline{\pi}_B) \hat{\underline{b}}_W] \\ &= \underline{\pi}_B (\hat{\underline{b}}_W - \hat{\underline{b}}_B) \end{aligned}$$

ce qui implique :

$$V(\hat{\underline{b}}_W - \underline{b}^* | X) = \underline{\pi}_B V(\hat{\underline{b}}_W - \hat{\underline{b}}_B | X) \underline{\pi}'_B$$

et

$$V(\hat{\underline{b}}_W - \underline{b}^*)^{-1} = (\underline{\pi}'_B)^{-1} V(\hat{\underline{b}}_W - \hat{\underline{b}}_B | X)^{-1} \underline{\pi}_B^{-1},$$

on en déduit la statistique du test de Hausman :

$$\begin{aligned} H &= (\hat{\underline{b}}_W - \underline{b}^*)' V(\hat{\underline{b}}_W - \underline{b}^* | X)^{-1} (\hat{\underline{b}}_W - \underline{b}^*) \\ &= (\hat{\underline{b}}_W - \hat{\underline{b}}_B)' \underline{\pi}'_B (\underline{\pi}'_B)^{-1} V(\hat{\underline{b}}_W - \hat{\underline{b}}_B | X)^{-1} \underline{\pi}_B^{-1} \underline{\pi}_B (\hat{\underline{b}}_W - \hat{\underline{b}}_B) \\ &= (\hat{\underline{b}}_W - \hat{\underline{b}}_B)' [V(\hat{\underline{b}}_W | X) + V(\hat{\underline{b}}_B | X)]^{-1} (\hat{\underline{b}}_W - \hat{\underline{b}}_B) \\ &= M. \end{aligned}$$

6.5 Un autre test

On peut voir le test qui suit comme une curiosité du modèle à erreur composée. A priori, une comparaison de l'estimateur des MCG et *between* ne devrait pas être un bon choix pour tester la présence d'effets individuels car ils sont tous les deux non convergents sous l'hypothèse alternative. Toutefois, comme l'estimateur des moindres carrés généralisés est une moyenne pondérée des estimateurs *between* et *within*, l'écart entre les deux estimateurs *between* et *within* va apparaître. La différence que l'on étudie est :

$$\begin{aligned}\hat{\underline{b}}_B - \underline{b}^* &= \hat{\underline{b}}_B - [\underline{\pi}_B \hat{\underline{b}}_B + (\mathbf{I}_{p-1} - \underline{\pi}_B) \hat{\underline{b}}_W] \\ &= (\mathbf{I}_{p-1} - \underline{\pi}_B) (\hat{\underline{b}}_B - \hat{\underline{b}}_W),\end{aligned}$$

ce qui change tout car étudier l'écart entre les estimateurs *between* et des MCG devient équivalent à étudier l'écart entre les estimateurs *between* et *within*. La statistique la plus simple que l'on puisse prendre est :

$$D_B = (\hat{\underline{b}}_B - \underline{b}^*)' \mathbf{V}(\hat{\underline{b}}_B - \underline{b}^*)^{-1} (\hat{\underline{b}}_B - \underline{b}^*),$$

or :

$$\begin{aligned}\mathbf{V}(\hat{\underline{b}}_B - \underline{b}^* | X) &= \mathbf{V}[(\mathbf{I}_{p-1} - \underline{\pi}_B) (\hat{\underline{b}}_B - \hat{\underline{b}}_W) | X] \\ &= (\mathbf{I}_{p-1} - \underline{\pi}_B) \mathbf{V}(\hat{\underline{b}}_B - \hat{\underline{b}}_W | X) (\mathbf{I}_{p-1} - \underline{\pi}_B)' \\ &= (\mathbf{I}_{p-1} - \underline{\pi}_B) [\mathbf{V}(\hat{\underline{b}}_B | X) + \mathbf{V}(\hat{\underline{b}}_W | X)] (\mathbf{I}_{p-1} - \underline{\pi}_B)',\end{aligned}$$

de sorte que l'on obtient :

$$\begin{aligned}D_B &= (\hat{\underline{b}}_B - \hat{\underline{b}}_W)' (\mathbf{I}_{p-1} - \underline{\pi}_B)' (\mathbf{I}_{p-1} - \underline{\pi}_B)^{-1} [\mathbf{V}(\hat{\underline{b}}_B | X) + \mathbf{V}(\hat{\underline{b}}_W | X)]^{-1} \\ &\quad (\mathbf{I}_{p-1} - \underline{\pi}_B) (\hat{\underline{b}}_B - \hat{\underline{b}}_W) \\ &= (\hat{\underline{b}}_B - \hat{\underline{b}}_W)' [\mathbf{V}(\hat{\underline{b}}_B | X) + \mathbf{V}(\hat{\underline{b}}_W | X)]^{-1} (\hat{\underline{b}}_B - \hat{\underline{b}}_W) \\ &= M,\end{aligned}$$

elle est donc numériquement égale à celle des tests de Hausman et de Mundlak. Pour effectuer le test, on peut toutefois faire directement *la différence* des matrices de covariance des deux estimateurs $\hat{\underline{b}}_B$ et \underline{b}^* car on a :

$$\begin{aligned}\hat{\underline{b}}_B &= b + (\underline{X}' (B - B_{NT}) \underline{X})^{-1} \underline{X}' (B - B_{NT}) u \\ \underline{b}^* &= b + (\underline{X}' (W + \theta^2 (B - B_{NT})) \underline{X})^{-1} \underline{X}' (W + \theta^2 (B - B_{NT})) u,\end{aligned}$$

ce qui implique :

$$\text{Cov}(\hat{\underline{b}}_B, \underline{b}^* | X) = (\underline{X}'(B - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT}) E(uu' | X) \\ (W + \theta^2(B - B_{NT}))\underline{X}(\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1},$$

or

$$E(uu' | X) = \sigma_\varepsilon^2 \left(W + \frac{1}{\theta^2} B \right),$$

à ce stade, on vérifie que :⁶

$$B_{NT}B = B_{NT}, B_{NT}W = 0,$$

de sorte que :

$$(B - B_{NT})E(uu' | X) = \sigma_\varepsilon^2 \left((B - B_{NT})W + \frac{1}{\theta^2} (B - B_{NT})B \right) \\ = \frac{\sigma_\varepsilon^2}{\theta^2} (B - B_{NT}),$$

et que :

$$(B - B_{NT})E(uu' | X)(W + \theta^2(B - B_{NT})) = \frac{\sigma_\varepsilon^2}{\theta^2} (B - B_{NT}) \\ (W + \theta^2(B - B_{NT})) \\ = \sigma_\varepsilon^2 (B - B_{NT}).$$

On en déduit :

$$\text{Cov}(\hat{\underline{b}}_B, \underline{b}^* | X) = \sigma_\varepsilon^2 (\underline{X}'(B - B_{NT})\underline{X})^{-1} \underline{X}'(B - B_{NT})\underline{X} \\ (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \\ = \sigma_\varepsilon^2 (\underline{X}'(W + \theta^2(B - B_{NT}))\underline{X})^{-1} \\ = V(\underline{b}^* | X)$$

$$V(\hat{\underline{b}}_B - \underline{b}^* | X) = V(\hat{\underline{b}}_B | X) + V(\underline{b}^* | X) - 2\text{Cov}(\hat{\underline{b}}_B, \underline{b}^* | X) \\ = V(\hat{\underline{b}}_B | X) - V(\underline{b}^* | X),$$

d'où la statistique :

$$D_B = (\hat{\underline{b}}_B - \underline{b}^*)' [V(\hat{\underline{b}}_B | X) - V(\underline{b}^* | X)]^{-1} (\hat{\underline{b}}_B - \underline{b}^*).$$

6. On vérifie que $B_{NT} = B_N \otimes B_T$, avec $B_N = e_N(e'_N e_N)^{-1} e'_N$ et $B_T = e_T(e'_T e_T)^{-1} e'_T$.

6.6 Programmation sous IML

Pour comprendre l'exécution du programme qui suit, il suffit de posséder les trois informations suivantes :

- La fonction *probf*(*x*, *dln*, *dld*) donne la fonction de répartition de la loi de Fisher (*dln*, *dld*) au point *x* (*dln* : degrés de liberté du numérateur, *dld* : degrés de liberté du dénominateur) ;
- L'expression *a* : *b* donne la liste des nombres entiers de *a* à *b*. Par exemple *1* : *4* donne le vecteur (1,2,3,4). Ici la liste *rg=2* : *p* donne la position des variables autres que le terme constant dans le vecteur des paramètres estimés. Comme les tests ne portent que sur les vecteurs sans le terme constant il faut utiliser *rg* avec tous les estimateurs sauf le *within*. L'expression *mcg[rg]* signifie que l'on ne garde que les éléments 2 à *p* du vecteur *mcg*. De même *vmcg[rg,rg]* signifie que l'on ne garde que les lignes et les colonnes 2 à *p* de la matrice *vmcg*, c'est-à-dire le bloc correspondant aux paramètres autres que le terme constant.
- La fonction *probchi*(*x*, *dl*) donne la fonction de répartition de la loi du Khi-Deux à *dl* degrés de liberté au point *x*.

La programmation des tests est immédiate. Il suffit d'entrer la formule théorique. Voici le module qu'il faut ajouter à la suite du programme qui réalise l'estimation par les MCQG (avant l'instruction *quit*) :

```

1. /* test de Fisher */
2. fisher=vub/vuw;
3. pfisher=1-probf(fisher,n-p,n#(t-1)-(p-1));
4. print "Test de Fisher (existence de l'effet
   individuel)",
5. "Statistique =" fisher,
6. "Probabilité critique =" pfisher;
7. /* Test de Hausman */
8. rg=2:p;
9. h=t(wit-mcg[rg])*inv(vwit-vmcg[rg,rg])*
   (wit-mcg[rg]);
10. ph=1-probchi(h##2,p-1);
11. print "Test de Hausman (effet individuel corrélé)",
12. "Statistique (Within - MCQG) =" h,
13. "Probabilité critique =" ph;

```

```

14. /* Test de Mundlak */
15. m=t(wit-bet[rg])*inv(vwit+vbet[rg,rg])*
    (wit-bet[rg]);
16. pm=1-probchi(m##2,p-1);
17. print "Test de Mundlak (effet individuel
    corrélé)",
18. "Statistique (Within - Between) =" m,
19. "Probabilité critique =" pm;
20. /* Curiosité */
21. c=t(bet[rg]-mcg[rg])*inv(vbet[rg,rg]-vmcg[rg,
    rg])*(bet[rg]-mcg[rg]);
22. pc=1-probchi(c##2,p-1);
23. print "Autre test (effet individuel corrélé)",
24. "Statistique (Between - MCQG) =" c,
25. "Probabilité critique =" pc;
26. quit;run;

```

On obtient la sortie suivante, où l'on peut vérifier trois propriétés :

- Il existe bien un effet individuel significatif au seuil de 5% ;
- L'effet individuel est corrélé avec les variables explicatives au seuil de 5% ;
- Les statistiques de Hausman, Mundlak et *Between*-MCQG sont numériquement identiques.

Sortie 6.1.

Test de Fisher (existence de l'effet individuel)

FISHER

Statistique = 2.3312578

PFISHER

Probabilité critique = 0

Test de Hausman (effet individuel corrélé)

H

Statistique (Within - MCQG) = 274.20757
PH

Probabilité critique = 0

Test de Mundlak (effet individuel corrélé)
M

Statistique (Within - Between) = 274.20757
PM

Probabilité critique = 0

Autre test (effet individuel corrélé)
C

Statistique (Between - MCQG) = 274.20757
PC

Probabilité critique = 0

Chapitre 7

Effet individuel et prévision

7.1 Estimateur brut de l'effet individuel

Pour estimer l'effet individuel du modèle, il suffit *a priori* d'estimer la relation suivante par les moindres carrés ordinaires :

$$y = \underline{X} \underline{b} + Ga + \varepsilon,$$

avec :

$$G = (g_1, \dots, g_N) = \begin{pmatrix} e_T & 0_{T,1} & \cdots & 0_{T,1} \\ 0_{T,1} & e_T & \cdots & 0_{T,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{T,1} & 0_{T,1} & \cdots & e_T \end{pmatrix} \text{ et } a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}.$$

La matrice G est appelée matrice des indicatrices individuelles. Elle est définie par :

$$\forall t, g_{it} = \begin{cases} 1 & \text{si l'observation appartient à l'individu } i \\ 0 & \text{sinon} \end{cases},$$

on peut réécrire cette matrice sous la forme :

$$\underset{(NT,N)}{G} = I_N \otimes e_T.$$

Le vecteur a représente ici l'ensemble des effets individuels *non centrés*. En notant \underline{a} la moyenne des effets individuels a_i , on obtient la relation :

$$a_i - \underline{a} = \alpha_i$$

On remarque que la somme des colonnes de la matrice G est égale au terme constant du modèle, de sorte qu'il ne faut pas l'ajouter au modèle. On devra

estimer le terme constant du modèle en prenant la moyenne des estimations des effets individuels non centrés \hat{a}_i . L'estimateur des MCO de ce modèle est défini par :

$$\begin{pmatrix} \underline{X}'\underline{X} & \underline{X}'G \\ G'\underline{X} & G'G \end{pmatrix} \begin{pmatrix} \hat{\underline{b}} \\ \hat{\underline{a}} \end{pmatrix} = \begin{pmatrix} \underline{X}'y \\ G'y \end{pmatrix},$$

ce qui donne les deux équations matricielles :

$$\begin{aligned} \underline{X}'\underline{X} \hat{\underline{b}} + \underline{X}'G \hat{\underline{a}} &= \underline{X}'y \\ G'\underline{X} \hat{\underline{b}} + G'G \hat{\underline{a}} &= G'y \end{aligned}$$

En résolvant la seconde équation, on obtient :

$$\hat{\underline{a}} = (G'G)^{-1} G' (y - \underline{X} \hat{\underline{b}}),$$

et en reportant dans la première équation :

$$\begin{aligned} \underline{X}'\underline{X} \hat{\underline{b}} + \underbrace{\underline{X}'G(G'G)^{-1}G'}_{\hat{\underline{a}}} (y - \underline{X} \hat{\underline{b}}) &= \underline{X}'y \\ \Leftrightarrow \hat{\underline{b}} &= (\underline{X}'(\mathbf{I}_{NT} - P_G)\underline{X})^{-1} \underline{X}'(\mathbf{I}_{NT} - P_G)y, \end{aligned}$$

avec :

$$P_G = G(G'G)^{-1}G',$$

la matrice de projection orthogonale sur $\text{Im}(G)$. Or on a :

$$\begin{aligned} G'G &= (\mathbf{I}'_N \otimes e'_T)(\mathbf{I}_N \otimes e_T) \\ &= \mathbf{I}_N \otimes e'_T e_T, \end{aligned}$$

de sorte que :

$$(G'G)^{-1} = \mathbf{I}_N \otimes (e'_T e_T)^{-1},$$

d'où :

$$\begin{aligned} P_G &= (\mathbf{I}_N \otimes e_T) \left(\mathbf{I}_N \otimes (e'_T e_T)^{-1} \right) (\mathbf{I}_N \otimes e_T)' \\ &= \mathbf{I}_N \otimes e_T (e'_T e_T)^{-1} e'_T \\ &= \mathbf{I}_N \otimes B_T \\ &= B. \end{aligned}$$

en consequence on retrouve l'estimateur *within* :

$$\begin{aligned} \hat{\underline{b}} &= (\underline{X}'(\mathbf{I}_{NT} - B)\underline{X})^{-1} \underline{X}'(\mathbf{I}_{NT} - B)y \\ &= (\underline{X}'W\underline{X})^{-1} \underline{X}'W y \\ &= \hat{\underline{b}}_W, \end{aligned}$$

on en déduit que l'on peut estimer les effets individuels (sans duplication) par :

$$\hat{a} = (G'G)^{-1} G' (y - \underline{X} \hat{\underline{b}}_W) = [y_{i\bullet} - \underline{X}_{i\bullet} \hat{\underline{b}}_W]_{i=1, \dots, N}.$$

Dans la littérature on désigne donc également l'estimateur *within* comme l'estimateur LSDV ("Least Square Dummy Variables" estimator), ce qui signifie estimateur des moindres carrés avec variables indicatrices. La variance des effets individuels s'obtient à partir de :

$$\begin{aligned} \hat{a} &= (G'G)^{-1} G' (\underline{X} \underline{b} + Ga + \varepsilon - \underline{X} \hat{\underline{b}}_W) \\ &= a + (G'G)^{-1} G' (\varepsilon + \underline{X} (\underline{b} - \hat{\underline{b}}_W)), \end{aligned}$$

avec :

$$\underline{b} - \hat{\underline{b}}_W = -(\underline{X}'W\underline{X})^{-1} \underline{X}'W\varepsilon,$$

d'où :

$$\hat{a} - a = (G'G)^{-1} G' \left(\mathbf{I}_{NT} - \underline{X} (\underline{X}'W\underline{X})^{-1} \underline{X}'W \right) \varepsilon$$

ce qui implique :

$$V(\hat{a}|X) = E \left[(\hat{a} - a) (\hat{a} - a)' | X \right].$$

Ici, on remarque que :

$$WG = (\mathbf{I}_N \otimes W_T) (\mathbf{I}_N \otimes e_T) = (\mathbf{I}_N \otimes W_T e_T) = 0.$$

En conséquence :

$$\begin{aligned} V(\hat{a}|X) &= \sigma_\varepsilon^2 (G'G)^{-1} G' \left(\mathbf{I}_{NT} - \underline{X} (\underline{X}'W\underline{X})^{-1} \underline{X}'W \right) \\ &\quad \left(\mathbf{I}_{NT} - W\underline{X} (\underline{X}'W\underline{X})^{-1} \underline{X}' \right) G (G'G)^{-1} \\ &= (G'G)^{-1} G' \left(\mathbf{I}_{NT} - \underline{X} (\underline{X}'W\underline{X})^{-1} \underline{X}' \right) G (G'G)^{-1}. \end{aligned}$$

Il est possible ensuite d'étudier la distribution des \hat{a}_i et des $\hat{\varepsilon}_{it}$ séparément car le résidu du modèle s'écrit :

$$\hat{\varepsilon} = y - \underline{X} \hat{\underline{b}}_W - G \hat{a}.$$

Notons également que cette estimation est valable que l'effet individuel soit corrélé ou non aux variables explicatives, puisque $\hat{\underline{b}}_W$ et \hat{a} sont sans biais et convergents dans les deux cas.

7.2 Les variables constantes dans le temps

Certaines variables individuelles, comme le secteur d'activité, le genre, le plus diplôme obtenu, la profession des parents, ne varient pas dans le temps. Ceci pose un problème d'identification évident car ces variables sont colinéaires à l'effet individuel. Afin de mieux comprendre la structure du terme d'erreur, il est toutefois souhaitable de retirer l'effet de ces variables de l'effet individuel. Supposons que notre modèle soit composé d'un effet individuel en deux parties :

$$a_i = F_i d + \mu_i,$$

où F_i sont les r variables observables constantes dans le temps et μ_i l'effet individuel à proprement parler, puisque F_i est observable. Si l'on applique la transformation *within*, la partie en $F_i d$ est éliminée en même temps que μ_i et l'on obtient seulement un estimateur du paramètre b . Le problème d'estimation consiste ici à trouver un estimateur sans biais de d ainsi que de μ_i . La méthode que nous avons vu dans ce chapitre permet d'estimer a_i . Il est donc naturel d'utiliser la décomposition :

$$\hat{a}_i = F_i \hat{d} + \hat{\mu}_i, \quad i = 1, \dots, N$$

car elle permet de rester cohérent avec les estimateurs traditionnels. Ici, il n'est pas nécessaire de dupliquer les observations. On pose :

$$\underset{(N,r)}{F} = \begin{pmatrix} F_1 \\ \vdots \\ F_N \end{pmatrix}, \quad \underset{(N,1)}{\hat{a}} = \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_N \end{pmatrix}, \quad \underset{(N,1)}{\hat{\mu}} = \begin{pmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_N \end{pmatrix}$$

En minimisant la variance de l'estimateur \hat{d} , on obtient l'estimateur des moindres carrés ordinaires :

$$\hat{d} = (F'F)^{-1} F' \hat{a}$$

et

$$\hat{\mu} = \hat{a} - F \hat{d}.$$

Il nous reste à trouver la variance de \hat{d} . Elle est donnée par :

$$V(\hat{d}) = (F'F)^{-1} F' V(\hat{a}|X) F (F'F)^{-1}.$$

7.3 Estimateur centré de l'effet individuel

Avec l'estimation brute toutefois, rien ne garantit que la moyenne des \hat{a} soit nulle. Or on peut souhaiter normaliser cette moyenne à 0 pour être cohérent avec l'hypothèse $E(\alpha_i|X) = 0$. Notons $\hat{\alpha}_i$ le nouvel estimateur de l'effet individuel. Il suffit de remarquer qu'on peut l'obtenir en estimant le modèle *avec* un terme constant *et* la contrainte suivante :

$$\frac{1}{N} \sum_{i=1}^N \hat{\alpha}_i = 0$$

En effet, sans cette contrainte, le coefficient du terme constant n'est pas identifiable dans le modèle puisque le terme constant est égal à la somme des colonnes de G . Mais cette dernière propriété permet également de recalculer les estimateurs correspondants facilement. Le modèle que nous avons estimé est équivalent au modèle suivant, au sens où l'on projette y sur le même sous espace vectoriel $\text{Im}(\underline{X}, G)$:

$$y = \underline{X} \underline{b} + \sum_{i=1}^N g_i a_i + u, \text{ avec } \sum_{i=1}^N g_i = e_{NT},$$

donc, on peut écrire :

$$y = e_{NT} \underline{a} + \underline{X} \underline{b} + \sum_{i=1}^N g_i ((a_i - \underline{a}) + u), \text{ avec } \underline{a} = \frac{1}{N} \sum_{i=1}^N a_i,$$

Il suffit donc de poser :

$$\alpha_i = a_i - \underline{a},$$

et de prendre la moyenne des effets individuels bruts comme terme constant. Par analogie, on utilisera l'estimateur suivant pour le terme constant :

$$\hat{\underline{a}} = \frac{1}{N} \sum_{i=1}^N \hat{a}_i \triangleq \hat{a}_W,$$

On note cet estimateur \hat{a}_W car il ne dépend que de l'estimateur *within*. On peut donc associer un terme constant au modèle *within* alors même que le terme constant disparaît avec cette transformation. A priori, il s'agit d'un meilleur choix que \hat{a}_B , car il est robuste à l'existence d'une corrélation entre l'effet individuel et les variables explicatives, contrairement à \hat{a}_B . Pour l'effet individuel centré, on utilisera :

$$\hat{\alpha}_i = \hat{a}_i - \hat{\underline{a}}.$$

On utilisera cet estimateur en prévision, et pour tester la spécification du modèle.

7.4 Prédiction et résidu du modèle

Les estimations précédentes permettent de faire la prédiction suivante de la variable expliquée du modèle :

$$\hat{y}_{it} = \hat{a}_W + \underline{X}_{it} \hat{b}_W + \hat{\alpha}_i,$$

on peut alors examiner empiriquement la matrice de corrélation empirique entre les $\hat{\alpha}_i$ et les variables incluses dans \underline{X} . Ceci permet de voir si la corrélation se limite à certaines variables explicatives. Enfin, le résidu du modèle est donné par :

$$\hat{\varepsilon}_{it} = y_{it} - \hat{y}_{it} = y_{it} - \hat{a}_W - \underline{X}_{it} \hat{b}_W - \hat{\alpha}_i,$$

et l'on peut étudier sa distribution empirique à des fins de modélisation.

7.5 Programmation sous IML

Pour effectuer cette application, on a besoin des transformations *between* et *within* seulement. Après avoir estimé l'effet individuel, nous vérifierons directement la corrélation entre l'effet individuel et les variables explicatives. Le programme est le suivant. On étudie toujours la fonction de production mais, en anticipant sur le chapitre suivant, on ajoute une tendance t (variable *year* de la base) à l'équation :

$$\ln y_{it} = a + \alpha_L \ln L_{it} + \alpha_C \ln C_{it} + \alpha_K \ln K_{it} + \beta \times t + \alpha_i + \varepsilon_{it}$$

Programme 7.1.

```
%let y=LogS;
%let x=LogL LogC LogK year;
proc means noprint data=tab; by firm;
var &y &x;
output out=between mean=;

proc standard mean=0 data=tab out=within; by firm;
var &y &x;

proc iml;
ny=&y; nx=&x; nx=t(nx);
```

```

use between;
  read all var (ny) into by;
  read all var (nx) into bx;
use within;
  read all var (ny) into wy;
  read all var (nx) into wx;
use tab;
  read all var (ny) into y;
  read all var (nx) into x;
nt=nrow(wy); n=nrow(by); t=nt/n;
p=ncol(bx);
start mco(y,x) global(ixx);
ixx=inv(t(x)*x);
b=ixx*t(x)*y;
return (b);
finish mco;
wit=mco(wy,wx);
et=J(T,1,1);
abrut=(by-bx*wit)@et;
acent=abrut-sum(abrut)/nt;
y=by@et+wy; x=bx@et+wx;
epsi=y-x*wit-abrut;
py=y-epsi;
res=acent || epsi || py || y || x;
nres="acent"//"epsi"//"prev"//ny//nx;
create res from res[colname=nres]; append from res;
quit;
run;

```

Seules les dernière lignes du programme appellent un commentaire :

1. `wit=mco(wy,wx);`
Calcul de l'estimateur *within*, qui est convergent que l'effet individuel soit corrélé ou non;
2. `et=J(T,1,1);`
Calcul de e_T afin de pouvoir dupliquer les données qui ne sont disponibles dans la dimension individuelle (*between* et effet individuel estimé);
3. `abrut=(by-bx*wit)@et;`
Calcul de l'effet individuel non centré à partir de la combinaison des

transformation *between* des données et de l'estimateur *within*. Le produit de Kronecker @ sert à dupliquer chaque effet individuel T fois ;

4. `acent=abrut-sum(abrut)/nt;`
Calcul de l'effet individuel centré ;
5. `y=by@et+wy; x=bx@et+wx;`
Calcul des variables en niveaux, en n'oubliant pas de dupliquer les transformations *between* ;
6. `epsi=y-x*wit-abrut;`
Calcul du résidu idiosyncratique. On enlève l'effet individuel non centré (dupliqué) parce qu'il inclut le terme constant, ce qui garantit que le résidu est centré.
7. `py=y-epsi;`
Calcul de la prévision en utilisant la relation : $\hat{y}_{it} = y_{it} - \hat{\varepsilon}_{it}$.

On crée ensuite un tableau SAS contenant les estimations des composantes du terme d'erreur et l'ensemble des variables du modèle en niveaux. Les graphiques 7.1 et 7.2 donnent respectivement les densités de $\hat{\alpha}_i$ et $\hat{\varepsilon}_{it}$, ainsi que la comparaison des prévisions \hat{y}_{it} avec les vraies valeurs y_{it} , calculées par le programme ci-dessous.¹

Programme 7.2.

```
proc kde data=res;
univar acent/out=acent;
run;
symbol1 c=black i=join v=none;
proc gplot data=acent;
plot density*value=var/href=0;
run;
proc kde data=res;
univar epsi/out=epsi;
run;
symbol1 c=black i=join v=none;
proc gplot data=epsi;
plot density*value=var/href=0;
run;
```

1. L'exposé des procédures graphique et d'estimation non paramétrique dépasse le cadre de cet ouvrage. Sur les graphiques, voir Duguet (2004) ; sur l'estimation des densités, voir Silverman (1986). Le descriptif de la procédure KDE est disponible sur le site internet de SAS Institute.

```

symbol1 c=gray i=none v=plus;
symbol2 c=black i=join v=none;
proc gplot data=res;
plot (prev LogS)*LogS/overlay;
run;

```

On constate que la densité de l'effet individuel estimé est multimodale et fortement asymétrique. Ceci est sans doute relié au fait que les tests concluent à la présence d'une corrélation avec les variables explicatives du modèle.

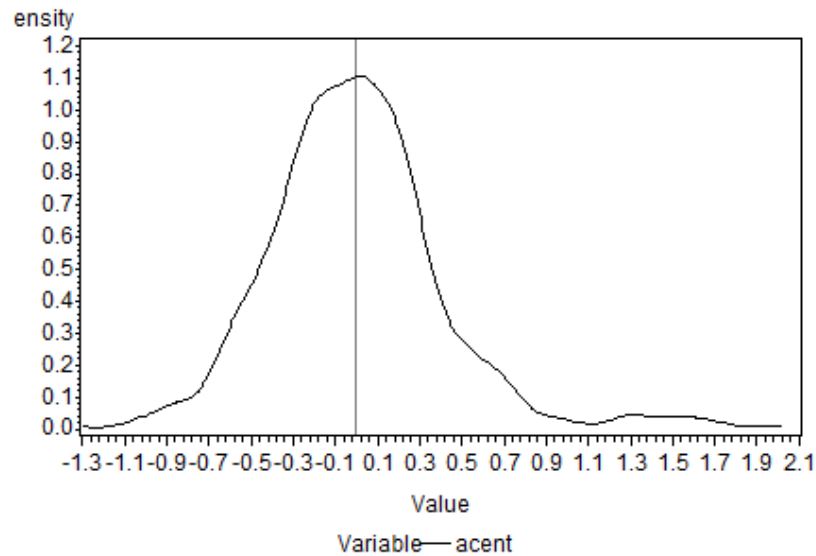


FIGURE 7.1 – Estimation de l'effet individuel, $\hat{\alpha}_{it}$

On peut creuser cette idée en calculant directement les coefficients de corrélation linéaires entre $\hat{\alpha}_i$ et les variables explicatives du modèle. Le lecteur remarquera qu'il est inutile de calculer les corrélations avec $\hat{\epsilon}_{it}$ car elles sont nulles par construction. C'est d'ailleurs ce que montre le programme ci-dessous.

Programme 7.3.

```

proc corr data=res;
var LogL LogC LogK;
with acent epsi;

```

run;

On obtient les corrélations suivantes :

Sortie 7.1.

Procédure CORR

2 Avec les variables : acent epsi
3 Variables : LOGL LOGC LOGK

Statistiques simples

Variable	N	Moyenne	Ecart-type	Somme
acent	6039	0	0.42478	0
epsi	6039	0	0.17981	0
LOGL	6039	1.13568	1.81638	6858
LOGC	6039	4.49649	2.28798	27154
LOGK	6039	3.09411	2.32898	18685

Statistiques simples

Variable	Minimum	Maximum
acent	-1.33709	2.01536
epsi	-4.41396	1.47857
LOGL	-5.80914	6.77537
LOGC	-3.72970	11.21666
LOGK	-4.71276	10.10108

Coefficients de corrélation de Pearson, N = 6039
Proba > |r| sous H0: Rho=0

	LOGL	LOGC	LOGK
acent	0.40571 <.0001	0.51235 <.0001	0.34836 <.0001
epsi	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000

On constate qu'il existe effectivement une forte corrélation entre $\hat{\alpha}_i$ et toutes les variables explicatives, de sorte que l'estimation *within* est à préférer dans cette application.

La densité de l'erreur idiosyncratique (7.2) présente un profil plus standard que celle de l'effet individuel, mais présente une asymétrie qui suggère que la distribution de la perturbation n'est pas normale. Ceci n'est toutefois pas un problème sur les grands échantillons. Il nous reste à vérifier la qualité des prévisions.

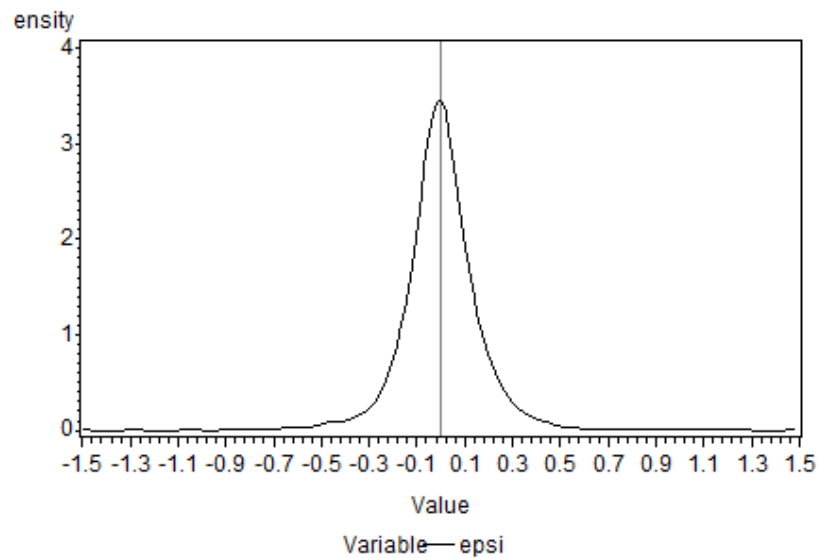
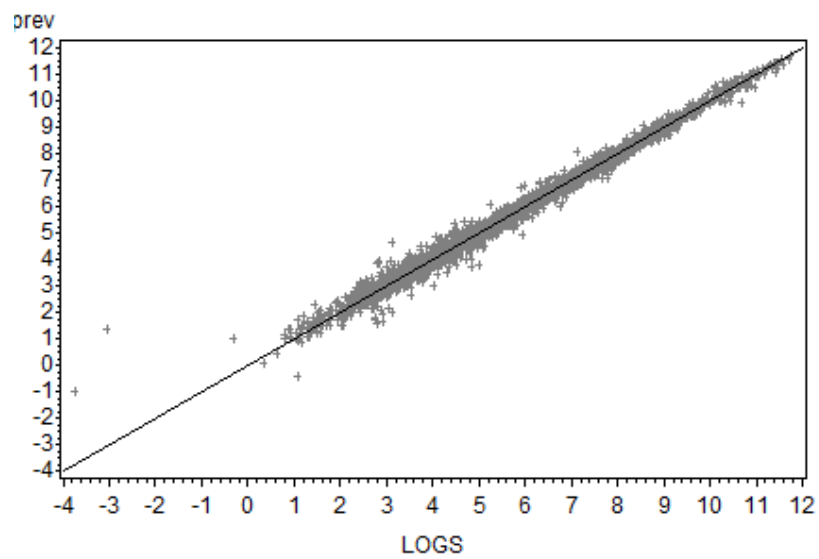


FIGURE 7.2 – Résidu idiosyncratique, $\hat{\epsilon}_{it}$

Le graphique 7.3 représente la prévision en fonction de la vraie valeur. Un point situé sur la première bissectrice représente une prévision parfaite. On constate que les prévisions se concentrent autour des vraies valeurs, à l'exception de quelques points sur la gauche du graphique, qui correspondent soit aux valeurs extrêmes des plus petites entreprises (LogS mesure la taille par les ventes) soit aux observations d'une seule entreprise dans le temps.

FIGURE 7.3 – Prédiction et vraie valeur, \hat{y}_{it} et y_{it}

Chapitre 8

Modèles estimés en différences

8.1 Estimation

Considérons le modèle statique défini par :

$$y_{it} = X_{it}b + \alpha_i + \varepsilon_{it}, \quad (8.1)$$

où la perturbation $u_{it} = \alpha_i + \varepsilon_{it}$ est potentiellement corrélée avec les variables explicatives X_{it} . Pour rendre la discussion plus intéressante, nous allons faire l'hypothèse que la perturbation ε_{it} est potentiellement autocorrélée :

$$\varepsilon_{it} = r\varepsilon_{it-1} + \eta_{it},$$

avec

$$|r| \leq 1, \eta_{it} \rightsquigarrow \text{BB}(0, \sigma_\eta^2),$$

où η_{it} est un bruit blanc indépendant de α_i et des variables explicatives X_{it} . On retrouve deux cas extrêmes intéressants :

1. $r = 0$. L'erreur idiosyncratique du modèle est un bruit blanc, et l'on se retrouve dans le cas standard du modèle à effet individuel (corrélé ou non) ;
2. $r = 1$. L'erreur idiosyncratique est une marche aléatoire et il faut estimer le modèle en différences.

Appliquée à la relation 8.1, la méthode des moindres carrés ordinaires ne permet pas d'obtenir un estimateur convergent de b en raison de la présence de α_i dans l'équation. Une méthode simple pour éliminer cet effet individuel consiste à écrire le modèle en différences :

$$\begin{aligned} y_{it} - y_{it-1} &= X_{it}b + \alpha_i + \varepsilon_{it} - (X_{it-1}b + \alpha_i + \varepsilon_{it-1}) \\ &= (\underline{X}_{it} - \underline{X}_{it-1})\underline{b} + \varepsilon_{it} - \varepsilon_{it-1}, \end{aligned}$$

où \underline{X}_{it} désigne les variables explicatives sans le terme constant et \underline{b} le paramètre sans la constante, car elle est éliminée par la différence. Il n'y a plus maintenant de corrélation entre les variables explicatives et le terme d'erreur. De ce fait, ce modèle peut être estimé par les moindres carrés ordinaires et ne nécessite pas d'appliquer des méthodes à variables instrumentales sous certaines hypothèses. L'hypothèse qui garantit la convergence de l'estimateur des MCO est celle d'*exogénéité forte* des perturbations, selon laquelle les valeurs de $\Delta\varepsilon_{it}$ à toutes les dates sont sans corrélation avec les valeurs de ΔX_{it} à toutes les dates. On résume cette propriété par :

$$E(\Delta X_{it} \Delta \varepsilon_{is}) = 0, \forall (t, s).$$

On peut réécrire le modèle en différences sous sa forme matricielle, pour un individu :

$$\begin{aligned} \begin{pmatrix} y_{i2} - y_{i1} \\ \vdots \\ y_{iT} - y_{iT-1} \end{pmatrix}_{(T-1,1)} &= \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}_{(T-1,T)} \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT-1} \\ y_{iT} \end{pmatrix}_{(T,1)} \\ &= Dy_i \end{aligned}$$

l'équation pour un individu s'écrit donc :

$$Dy_i = D\underline{X}_i \underline{b} + D\varepsilon_i,$$

et pour l'ensemble des individus :

$$\underbrace{(\mathbf{I}_N \otimes D)}_{(N(T-1), NT)} \underbrace{y}_{(NT, 1)} = (\mathbf{I}_N \otimes D) \underline{X} \underline{b} + (\mathbf{I}_N \otimes D) \varepsilon,$$

on peut donc poser la notation :

$$\Delta = \mathbf{I}_N \otimes D,$$

ce qui donne le modèle :

$$\Delta y = \Delta \underline{X} \underline{b} + \Delta \varepsilon, \quad (8.2)$$

et l'estimateur en différences s'écrit :

$$\hat{\underline{b}}_D = \left((\Delta \underline{X})' \Delta \underline{X} \right)^{-1} (\Delta \underline{X})' \Delta y,$$

Ici il faut toutefois faire attention à l'écriture de la matrice de covariance car elle n'est scalaire que sous l'hypothèse que ε_{it} suit une marche aléatoire ($r = 0$). Plus précisément :

$$\begin{aligned}\varepsilon_{it} - \varepsilon_{it-1} &= r\varepsilon_{it-1} + \eta_{it} - \varepsilon_{it-1} \\ &= (r-1)\varepsilon_{it-1} + \eta_{it}\end{aligned}\quad (8.3)$$

Si ε_{it} est une marche aléatoire, on a :

$$\begin{aligned}V(\varepsilon_{it} - \varepsilon_{it-1}|X) &= V(\eta_{it}|X) = \sigma_\eta^2, \\ \text{Cov}(\varepsilon_{it} - \varepsilon_{it-1}, \varepsilon_{it-1} - \varepsilon_{it-2}) &= \text{Cov}(\eta_{it}, \eta_{it-1}) = 0,\end{aligned}$$

de sorte que la matrice de covariance est scalaire :

$$V(\Delta\varepsilon) = \sigma_\eta^2 \times I_{NT},$$

ce qui implique que l'on ne peut utiliser les écarts types des moindres carrés ordinaires que lorsque l'erreur idiosyncratique est une marche aléatoire. Dans le cas où ε_{it} est un bruit blanc, on a :

$$\begin{aligned}V(\Delta\varepsilon|X) &= \Delta V(\varepsilon|X) \Delta' \\ &= \sigma_\varepsilon^2 \Delta \Delta' \\ &= \sigma_\varepsilon^2 (I_N \otimes DD')\end{aligned}$$

avec :

$$\begin{aligned}DD' &= \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ & & & \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{pmatrix}\end{aligned}$$

ceci correspond aux propriétés suivantes :

$$\begin{aligned}
V(\Delta\varepsilon_{it}|X) &= V(\varepsilon_{it} - \varepsilon_{it-1}|X) \\
&= V(\varepsilon_{it}|X) + V(\varepsilon_{it-1}|X) \\
&= 2\sigma_\varepsilon^2, \\
\text{Cov}(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}|X) &= \text{Cov}(\varepsilon_{it} - \varepsilon_{it-1}, \varepsilon_{it-1} - \varepsilon_{it-2}|X) \\
&= -V(\varepsilon_{it-1}|X) \\
&= -\sigma_\varepsilon^2. \\
\text{Corr}(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}|X) &= -\frac{1}{2}.
\end{aligned}$$

En conséquence :

$$\begin{aligned}
V(\hat{\underline{b}}_D|X) &= \sigma_\varepsilon^2 \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1} (\Delta\underline{X})' \Delta \Delta' (\Delta\underline{X}) \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1} \\
&= \sigma_\varepsilon^2 \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1} \underline{X}' \left(I_N \otimes (D'D)^2 \right) \underline{X} \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1}.
\end{aligned}$$

qui ne se simplifie pas. Dans le cas général (8.3), toutes les différences sont corrélées entre elles. Heureusement, il existe une méthode qui permet d'estimer $V(\hat{\underline{b}}_D|X)$ de manière convergente dans tous les cas : la covariance robuste de White. Elle est définie par :

$$\hat{V}_R(\hat{\underline{b}}_D|X) = \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1} \left[\sum_{i=1}^N (D\underline{X}_i)' D\hat{\varepsilon}_i (D\hat{\varepsilon}_i)' D\underline{X}_i \right] \left((\Delta\underline{X})' \Delta\underline{X} \right)^{-1},$$

avec $D\hat{\varepsilon}_i = Dy_i - D\underline{X}_i \hat{\underline{b}}_D$. Cet estimateur tient compte de l'autocorrélation du terme d'erreur. Une fois l'estimation réalisée, on peut estimer la somme de l'effet individuel et du terme constant par la relation :

$$\hat{a}_{Di} = y_{i\bullet} - \underline{X}_{i\bullet} \hat{\underline{b}}_D, \quad i = 1, \dots, N.$$

Le terme constant du modèle s'obtient par :

$$\hat{\underline{a}}_D = \frac{1}{N} \sum_{i=1}^N \hat{a}_{Di},$$

l'effet individuel par :

$$\hat{\alpha}_i = \hat{a}_{Di} - \hat{\underline{a}}_D,$$

et le résidu du modèle s'obtient par :

$$\hat{\varepsilon}_{it} = y_{it} - \underline{X}_{it} \hat{\underline{b}}_D - \hat{a}_{Di} = y_{it} - \underline{X}_{it} \hat{\underline{b}}_D - \hat{\underline{a}}_D - \hat{\alpha}_i, \quad (8.4)$$

ce qui permet d'étudier l'autocorrélation de cette série. Si le coefficient de corrélation empirique entre $\hat{\varepsilon}_{it}$ et $\hat{\varepsilon}_{it-1}$ est proche de zéro, on obtient les écarts types des moindres carrés ordinaires (cas de la marche aléatoire), sinon les écarts types seront différents (cas du modèle à effet individuel). On peut avoir une idée plus précise du processus des perturbations en calculant le coefficient de corrélation linéaire *moyen* entre $\Delta\hat{\varepsilon}_{it}$ et $\hat{\varepsilon}_{it-1}$. Notons que l'on perd une année de données à cause du décalage temporel. Toutes les sommes se font donc de $t = 2$ à $t = T$, sur $T - 1$ valeurs :

$$\hat{r} = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{\sum_{t=2}^T (\Delta\hat{\varepsilon}_{it} - \Delta\hat{\varepsilon}_{i\bullet}) (\hat{\varepsilon}_{it-1} - \hat{\varepsilon}_{i\bullet}^{(-1)})}{\sqrt{\sum_{t=2}^T (\Delta\hat{\varepsilon}_{it} - \Delta\hat{\varepsilon}_{i\bullet})^2} \sqrt{\sum_{t=2}^T (\hat{\varepsilon}_{it-1} - \hat{\varepsilon}_{i\bullet}^{(-1)})^2}} \right\},$$

avec :

$$\Delta\hat{\varepsilon}_{i\bullet} = \frac{1}{T-1} \sum_{t=2}^T \Delta\hat{\varepsilon}_{it}, \quad \hat{\varepsilon}_{i\bullet}^{(-1)} = \frac{1}{T-1} \sum_{t=2}^T \hat{\varepsilon}_{it-1}, \quad i = 1, \dots, N.$$

Enfin, notons que si le résidu $\Delta\hat{\varepsilon}$ est orthogonal à ΔX , il est possible que $\hat{\varepsilon}$ ne soit pas orthogonal à X . Dans notre application la corrélation est très faible, mais elle existe entre les niveaux, alors qu'elle est inexistante, par construction, entre les différences.

8.2 Application

Nous reprenons l'estimation de la fonction de production. Le programme suivant réalise l'estimation du modèle. Pour estimer l'effet individuel, on peut utiliser la même méthode que pour l'estimation *within*. Le terme constant s'obtient en faisant la moyenne des effets individuels, qu'il convient alors de centrer pour faire la prévision. On se contente ici de l'estimation de départ, car il s'agit de la seule partie nouvelle.

1. `data tab; set rep.bal_hmuspan; proc sort;`
`by firm year;`
 Lecture et tri de la base de données. Ce dernier point est important pour prendre les différences;
2. `data tab; set tab; by firm year;`
 On lit les données dans le bon ordre chronologique. L'instruction `by` permet de générer la variable booléenne `first.firm` que nous utiliserons plus loin.

3. `dLogS=dif(LogS); dLogL=dif(LogL);`
`dLogC=dif(LogC); dLogK=dif(LogK);`
 La fonction *dif* fait la différence entre l'observation courante de la variable en argument, et l'observation qui la précède dans la base de données. Ceci pose un problème quand on utilise la première observation d'une entreprise, car la différence se fait alors avec la dernière observation de l'entreprise précédente, ce qu'il faut corriger en enlevant la première observation de chaque entreprise;
4. `if first.firm then delete;`
 La variable booléenne *first.firm* vaut "vrai" quand l'entreprise (indiquée par *firm*) apparaît pour la première fois dans la base de données. Dans ce cas, on supprime la ligne d'observation correspondante grâce à l'instruction *delete*. Ainsi, on ne mélange pas les données des entreprises. On remarque que chaque entreprise a maintenant $T - 1$ observations et la base de données $N(T - 1)$ observations;
5. `proc freq data=tab; tables year/out=tab2;`
 La procédure *freq* donne le nombre d'observations de chaque modalité de la variable d'année *year*. Il s'agit du nombre d'entreprise et, comme l'échantillon est cylindré, ce nombre est le même pour toutes les modalités de *year*. On stocke le résultat dans un tableau *tab2*;
6. `data tab2; set tab2; if _N_=1;`
 On garde la première ligne du tableau *tab2*. La variable *count*, générée automatiquement par la procédure *freq*, contient le nombre d'entreprises du panel;
7. `proc iml;`
 On lance la procédure *iml*;
8. `ny={dLogS}; nx={dLogL dLogC dLogK}; nx=t(nx);`
 On range le nom de la variable expliquée dans un vecteur de caractères *ny*, ceux des variables explicatives dans un vecteur *nx*. On transpose ensuite *nx* afin d'avoir les noms dans une colonne pour l'affichage des résultats effectué à la fin du programme;
9. `use tab;`
 On ouvre le tableau *tab*, qui contient les données;
10. `read all var (ny) into dy;`
 On lit les valeurs de *Dy* que l'on range dans un vecteur *dy*;
11. `read all var (nx) into dx;`
 On lit les valeurs de *DX* que l'on range dans une matrice *dx*. On n'ajoute pas de terme constant au modèle;

12. use tab2;
On ouvre le tableau tab2, qui contient le nombre d'entreprises;
13. read all var {count} into n;
On lit le nombre d'entreprises que l'on range dans la variable n;
14. nt1=nrow(dy); t1=nt1/n;
Le nombre de lignes de dy est égal à $N(T-1)$, on range ce nombre dans la variable nt1. On calcule ensuite le nombre de différences par entreprise en divisant ce nombre par N . On range ce nombre dans la variable t1;
15. idxdx=inv(t(dx)*dx);
On calcule la matrice $((DX)'DX)^{-1}$ que l'on range dans idxdx;
16. b=idxdx*t(dx)*dy;
Calcul de l'estimateur des moindres carrés ordinaires, que l'on range dans b;
17. de=dy-dx*b;
Calcul de $D\hat{\epsilon}$ que l'on range dans de;
18. somme=0;
On initialise une somme à 0. Après la boucle qui suit, elle contiendra $S = \sum_{i=1}^N (DX_i)' D\hat{\epsilon}_i (D\hat{\epsilon}_i)' DX_i$
19. do i=1 to n;
Début de la boucle, réalisée individu par individu. Cette boucle portera sur des blocs de dimension $T-1$;
20. rgi=(1+t1#(i-1)): (t1#i);
rgi contient l'emplacement des lignes de données de la i -ème entreprise. Elles se situent de la position $a = 1 + (T-1)(i-1)$ à la position $b = (T-1)i$ (bornes incluses). On obtient donc les positions : 1 à $T-1$, T à $2(T-1)$ etc. On remarque que le nombre d'éléments de a à b est égal à $b-a+1 = (T-1)i - (1 + (T-1)(i-1)) + 1 = T-1$. La notation $a : b$ fournit la suite des nombres entiers de a à b ;
21. somme=somme+t(dx[rgi,])*de[rgi]*t(de[rgi])*dx[rgi,];
Calcul de la somme S . La notation $dy[rgi]$ signifie que l'on ne retient que les lignes de dy indiquée par rgi; de même la notation $x[rgi,]$ signifie que l'on retient les lignes indiquées par rgi et toutes les colonnes car il n'y a pas d'indication en deuxième position (i.e., après la virgule) pour dx;
22. end;
Fin de la boucle;

23. `vb=idxdx*somme*idxdx;`
Calcul de la matrice de covariance robuste de l'estimateur en différences ;
24. `sb=sqrt(vecdiag(vb));`
Calcul des estimations des écarts-types des estimateurs ;
25. `tb=abs(b)/sb;`
Calcul du *t* de Student ;
26. `print "Estimation du modèle en différences :",`
`b[rowname=nx] sb tb;`
Impression des résultats ;
27. `quit; run;` Sortie de la procédure IML. Il faut ajouter une instruction *run* pour exécuter ce programme.

Sortie 8.1.

```
Estimation du modèle en différences :
      b                sb                tb

DLOGL 0.5995625 0.0374567 16.006799
DLOGC 0.1189873 0.0178776 6.6556654
DLOGK 0.1558388 0.0320384 4.8641196
```

Chapitre 9

Modèle général à erreur composée

Les méthodes précédentes s'appliquent en présence d'un effet individuel. Dans la pratique, il est également possible qu'il existe un effet temporel dans le modèle, c'est-à-dire une variable aléatoire qui affecte la totalité des individus à une même date. On peut penser à un effet de type conjoncturel. Les estimateurs que nous avons vu doivent être alors modifiés. Il faut également veiller à avoir suffisamment d'observations dans la dimension temporelle pour que le modèle général à erreur composée soit utile. Le modèle devient :

$$\begin{aligned}y_{it} &= X_{it}b + u_{it}, \\u_{it} &= \alpha_i + \beta_t + \varepsilon_{it}, \\i &= 1, \dots, N \quad t = 1, \dots, T\end{aligned}$$

avec :

$$\begin{aligned}E(\alpha_i|X) &= E(\beta_t|X) = E(\varepsilon_{it}|X) = 0, \\V(\alpha_i|X) &= \sigma_\alpha^2, V(\beta_t|X) = \sigma_\beta^2, V(\varepsilon_{it}|X) = \sigma_\varepsilon^2, \\E(\alpha_i\alpha_j|X) &= 0 \quad \forall i \neq j, E(\alpha_i\beta_t|X) = 0 \quad \forall (i, t), E(\beta_s\beta_t|X) = 0 \quad \forall s \neq t, \\E(\alpha_i\varepsilon_{jt}|X) &= 0 \quad \forall (i, j, t), E(\beta_s\varepsilon_{it}|X) = 0 \quad \forall (s, i, t), \\E(\varepsilon_{is}\varepsilon_{jt}|X) &= 0 \quad \forall i \neq j \text{ ou } s \neq t.\end{aligned}$$

On empile alors les données comme dans les cas précédents, ce qui donne :

$$y = Xb + u,$$

avec une structure particulière pour le terme d'erreur :

$$u = \begin{pmatrix} \varepsilon_{11} + \alpha_1 + \beta_1 \\ \vdots \\ \varepsilon_{1T} + \alpha_1 + \beta_T \\ \vdots \\ \varepsilon_{i1} + \alpha_i + \beta_1 \\ \vdots \\ \varepsilon_{iT} + \alpha_i + \beta_T \\ \vdots \\ \varepsilon_{N1} + \alpha_N + \beta_1 \\ \vdots \\ \varepsilon_{NT} + \alpha_N + \beta_T \end{pmatrix}$$

cette structure a pour conséquence de créer une corrélation entre les perturbations de deux individus différents, via les effets temporels β_t . Pour un individu, la matrice de covariance de la perturbation est donnée par :

$$\begin{aligned} \underbrace{V(u_i|X)}_{(T,T)} &= \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 + \sigma_\beta^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\alpha^2 + \sigma_\beta^2 \end{pmatrix} \\ &= T\sigma_\alpha^2 B_T + (\sigma_\varepsilon^2 + \sigma_\beta^2) I_T \end{aligned}$$

et la matrice de covariance entre deux individus est égale à :

$$\underbrace{\text{Cov}(u_i, u_j|X)}_{(T,T)} = \begin{pmatrix} \sigma_\beta^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_\beta^2 \end{pmatrix} = \sigma_\beta^2 I_T, \forall i \neq j,$$

ce qui donne globalement :

$$\underbrace{V(u|X)}_{(NT,NT)} = \begin{pmatrix} T\sigma_\alpha^2 B_T + (\sigma_\varepsilon^2 + \sigma_\beta^2) I_T & \cdots & \sigma_\beta^2 I_T \\ \vdots & \ddots & \vdots \\ \sigma_\beta^2 I_T & \cdots & T\sigma_\alpha^2 B_T + (\sigma_\varepsilon^2 + \sigma_\beta^2) I_T \end{pmatrix}.$$

On peut réécrire cette matrice de covariance sous la forme plus pratique :

$$\Omega = V(u|X) = \sigma_\varepsilon^2 I_{NT} + T\sigma_\alpha^2 (I_N \otimes B_T) + N\sigma_\beta^2 (B_N \otimes I_T)$$

avec :

$$B_N = e_N \underbrace{(e'_N e_N)^{-1}}_{1/N} e'_N.$$

9.1 Les transformations

On distingue trois estimateurs dont les perturbations sont orthogonales : le *between* individuel, le *between* temporel et le *double within*. Nous verrons plus loin qu'il est également possible d'obtenir des estimateurs *within* individuel et temporel. Afin d'assurer l'orthogonalité des matrices de projection que nous utiliserons plus loin, il est nécessaire de centrer tous les estimateurs. On définit l'opérateur qui effectue la moyenne sur toutes les observations de la manière suivante :

$$B_{NT} = e_{NT} (e'_{NT} e_{NT})^{-1} e'_{NT},$$

cette matrice effectue également une duplication car la moyenne est représentée par un vecteur constant de dimension $(NT, 1)$. On vérifie que cette matrice est symétrique ($B_{NT} = B'_{NT}$) et idempotente ($B^2_{NT} = B_{NT}$). La régression dans la dimension totale est centrée et s'effectue par la transformation $I_{NT} - B_{NT}$. Ceci correspond à une régression de $y_{it} - y_{..}$ sur $X_{it} - X_{..}$ où les points indiquent les indices sur lesquels on a pris les moyennes. Ici :

$$y_{..} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it} \text{ et } X_{..} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it}.$$

La matrice de projection B_{NT} peut se réécrire d'une manière plus pratique pour certains calculs. On remarque tout d'abord que :

$$e_{NT} = e_N \otimes e_T,$$

ce qui implique :

$$\begin{aligned} B_{NT} &= (e_N \otimes e_T) ((e_N \otimes e_T)' (e_N \otimes e_T))^{-1} (e_N \otimes e_T)' \\ &= (e_N \otimes e_T) ((e'_N \otimes e'_T) (e_N \otimes e_T))^{-1} (e'_N \otimes e'_T) \\ &= (e_N \otimes e_T) (e'_N e_N \otimes e'_T e_T)^{-1} (e'_N \otimes e'_T) \\ &= (e_N \otimes e_T) \left((e'_N e_N)^{-1} \otimes (e'_T e_T)^{-1} \right) (e'_N \otimes e'_T) \\ &= \left(e_N (e'_N e_N)^{-1} \otimes e_T (e'_T e_T)^{-1} \right) (e'_N \otimes e'_T) \\ &= \left(e_N (e'_N e_N)^{-1} e'_N \otimes e_T (e'_T e_T)^{-1} e'_T \right) \\ &= B_N \otimes B_T, \end{aligned}$$

d'autre part, on a également :

$$e_{NT} = e_T \otimes e_N,$$

de sorte que :

$$B_{NT} = B_T \otimes B_N.$$

Pour le *between* individuel on procède comme précédemment après avoir centré les variables par rapport à la moyenne globale. La transformation est donnée par $(I_N \otimes B_T) - B_{NT}$ et correspond à une régression de $y_{i\bullet} - y_{\bullet\bullet}$ sur $X_{i\bullet} - X_{\bullet\bullet}$. L'opérateur *between* individuel sera noté :

$$Q_I = B - B_{NT} = (I_N \otimes B_T) - B_{NT}.$$

De la même manière on définit l'estimateur *between* temporel comme l'application des moindres carrés ordinaires aux moyennes temporelles centrées. La matrice de projection est donnée par $(B_N \otimes I_T) - B_{NT}$ et correspond à une régression de $y_{\bullet t} - y_{\bullet\bullet}$ sur $X_{\bullet t} - X_{\bullet\bullet}$. L'opérateur *between* temporel sera noté :

$$Q_T = (B_N \otimes I_T) - B_{NT}.$$

Enfin l'estimateur *double within* est obtenu en enlevant les moyennes individuelles et temporelles à la série de départ. Une petite correction, par la moyenne globale, intervient car il nous faut préserver l'orthogonalité entre les différents projecteurs. Le *double within* est obtenu par la transformation :

$$Q_W = I_{NT} - B_{NT} - Q_I - Q_T,$$

elle consiste à travailler sur les transformations :

$$y_{it} - y_{\bullet\bullet} - (y_{i\bullet} - y_{\bullet\bullet}) - (y_{\bullet t} - y_{\bullet\bullet}) = y_{it} - y_{i\bullet} - y_{\bullet t} + y_{\bullet\bullet}$$

Vérifions que les matrices précédentes sont symétriques et idempotentes :

$$\begin{aligned} Q_I' &= (I_N \otimes B_T)' - B_{NT}' \\ &= (I_N' \otimes B_T') - B_{NT} \\ &= (I_N \otimes B_T) - B_{NT} \\ &= Q_I, \end{aligned}$$

$$\begin{aligned} Q_I^2 &= (I_N \otimes B_T)^2 - (I_N \otimes B_T) B_{NT} - B_{NT} (I_N \otimes B_T) + B_{NT}^2 \\ &= (I_N^2 \otimes B_T^2) - (I_N \otimes B_T) (B_N \otimes B_T) - (B_N \otimes B_T) (I_N \otimes B_T) + B_{NT} \\ &= I_N \otimes B_T - B_N \otimes B_T^2 - B_N \otimes B_T^2 + B_{NT} \\ &= I_N \otimes B_T - B_N \otimes B_T - B_N \otimes B_T + B_N \otimes B_T \\ &= I_N \otimes B_T - B_{NT} - B_{NT} + B_{NT} \\ &= Q_I. \end{aligned}$$

Pour le *between* temporel, la preuve est similaire :

$$\begin{aligned}
 Q'_T &= (B_N \otimes I_T)' - B'_{NT} \\
 &= (B'_N \otimes I'_T) - B_{NT} \\
 &= (B_N \otimes I_T) - B_{NT} \\
 &= Q_T,
 \end{aligned}$$

$$\begin{aligned}
 Q_T^2 &= (B_N \otimes I_T)^2 - (B_N \otimes I_T) B_{NT} - B_{NT} (B_N \otimes I_T) + B_{NT}^2 \\
 &= B_N^2 \otimes I_T^2 - (B_N \otimes I_T) (B_N \otimes B_T) - (B_N \otimes B_T) (B_N \otimes I_T) + B_{NT} \\
 &= B_N \otimes I_T - B_N^2 \otimes B_T - B_N^2 \otimes B_T + B_{NT} \\
 &= B_N \otimes I_T - B_N \otimes B_T - B_N \otimes B_T + B_N \otimes B_T \\
 &= B_N \otimes I_T - B_{NT} - B_{NT} + B_{NT} \\
 &= Q_T.
 \end{aligned}$$

Vérifions maintenant l'orthogonalité des opérateurs entre eux :

$$\begin{aligned}
 Q_I B_{NT} &= (I_N \otimes B_T - B_{NT}) B_{NT} \\
 &= (I_N \otimes B_T - B_N \otimes B_T) (B_N \otimes B_T) \\
 &= B_N \otimes B_T^2 - B_N^2 \otimes B_T^2 \\
 &= B_N \otimes B_T - B_N \otimes B_T \\
 &= 0,
 \end{aligned}$$

et, en utilisant les propriétés de symétrie :

$$B_{NT} Q_I = B'_{NT} Q'_I = (Q_I B_{NT})' = \begin{matrix} \mathbf{0}' \\ (NT, NT) \end{matrix} = \begin{matrix} \mathbf{0} \\ (NT, NT) \end{matrix}.$$

De même :

$$\begin{aligned}
 Q_T B_{NT} &= (B_N \otimes I_T - B_{NT}) B_{NT} \\
 &= (B_N \otimes I_T - B_N \otimes B_T) (B_N \otimes B_T) \\
 &= B_N^2 \otimes B_T - B_N^2 \otimes B_T^2 \\
 &= B_N \otimes B_T - B_N \otimes B_T \\
 &= 0,
 \end{aligned}$$

ainsi que $B_{NT}Q_T = 0$,

$$\begin{aligned}
 Q_T Q_I &= (B_N \otimes I_T - B_{NT})(I_N \otimes B_T - B_{NT}) \\
 &= (B_N \otimes I_T - B_N \otimes B_T)(I_N \otimes B_T - B_N \otimes B_T) \\
 &= B_N \otimes B_T - B_N^2 \otimes B_T - B_N \otimes B_T^2 + B_N^2 \otimes B_T^2 \\
 &= B_N \otimes B_T - B_N \otimes B_T - B_N \otimes B_T + B_N \otimes B_T \\
 &= 0,
 \end{aligned}$$

et $Q_I Q_T = 0$. Les propriétés du *double within* peuvent maintenant être étudiées simplement :

$$\begin{aligned}
 Q'_W &= I'_{NT} - B'_{NT} - Q'_I - Q'_T \\
 &= I_{NT} - B_{NT} - Q_I - Q_T \\
 &= Q_W,
 \end{aligned}$$

et

$$\begin{aligned}
 Q_W^2 &= (I_{NT} - B_{NT} - Q_I - Q_T)(I_{NT} - B_{NT} - Q_I - Q_T) \\
 &= I_{NT} - B_{NT} - Q_I - Q_T - B_{NT} + B_{NT}^2 + B_{NT}Q_I + B_{NT}Q_T \\
 &\quad - Q_I + Q_I B_{NT} + Q_I^2 + Q_I Q_T - Q_T + Q_T B_{NT} + Q_T Q_I + Q_T^2 \\
 &= I_{NT} - B_{NT} - Q_I - Q_T - B_{NT} + B_{NT} - Q_I + Q_I - Q_T + Q_T \\
 &= I_{NT} - B_{NT} - Q_I - Q_T \\
 &= Q_W
 \end{aligned}$$

cette matrice de projection est également orthogonale aux autres matrices :

$$\begin{aligned}
 Q_W B_{NT} &= (I_{NT} - B_{NT} - Q_I - Q_T) B_{NT} \\
 &= B_{NT} - B_{NT} - \underbrace{Q_I B_{NT}}_0 - \underbrace{Q_T B_{NT}}_0 \\
 &= 0,
 \end{aligned}$$

$$\begin{aligned}
 Q_W Q_I &= (I_{NT} - B_{NT} - Q_I - Q_T) Q_I \\
 &= Q_I - \underbrace{B_{NT} Q_I}_0 - \underbrace{Q_I^2}_{Q_I} + \underbrace{Q_T Q_I}_0 \\
 &= 0,
 \end{aligned}$$

$$\begin{aligned}
Q_W Q_T &= (I_{NT} - B_{NT} - Q_I - Q_T) Q_T \\
&= Q_T - \underbrace{B_{NT} Q_T}_0 - \underbrace{Q_I Q_T}_0 + \underbrace{Q_T^2}_{Q_T} \\
&= 0.
\end{aligned}$$

L'orthogonalité de ces matrices s'avérera très pratique lorsque l'on devra inverser Ω .

9.2 Les estimateurs

Tous les estimateurs s'obtiennent en appliquant les moindres carrés ordinaires aux données transformées.

9.2.1 Le between individuel

Il s'agit de l'estimateur que nous avons vu dans les chapitres précédents, à ceci près que l'on n'estime plus le terme constant en raison du centrage des données.¹ On remarquera que chaque moyenne individuelle centrée est dupliquée T fois :

$$\begin{aligned}
\hat{b}_I &= \left((Q_I \underline{X})' Q_I \underline{X} \right)^{-1} (Q_I \underline{X})' Q_I y \\
&= (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I y,
\end{aligned}$$

cet estimateur est sans biais car $E(u|X) = 0$. Pour calculer sa variance, on remplace y par sa valeur $Xb + u$, ce qui donne :

$$\hat{b}_I = b + (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I u,$$

d'où la variance :

$$\begin{aligned}
V(\hat{b}_I | X) &= E \left[(\hat{b}_I - b) (\hat{b}_I - b)' \middle| X \right] \\
&= (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I \Omega Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1}.
\end{aligned}$$

La matrice de covariance Ω peut elle même se réécrire en fonction des matrices de projection orthogonales :

$$\Omega = \underbrace{\sigma_\varepsilon^2 (B_{NT} + Q_W + Q_I + Q_T)}_{I_{NT}} + T \underbrace{\sigma_\alpha^2 (Q_I + B_{NT})}_{I_N \otimes B_T} + N \underbrace{\sigma_\beta^2 (Q_T + B_{NT})}_{B_N \otimes I_T},$$

1. La seule différence entre \hat{b}_B et l'estimateur \hat{b}_I est donc l'absence d'estimateur du terme constant dans le second. Il est toutefois possible d'estimer un terme constant par une méthode analogue à celle du chapitre précédent.

en réarrangeant on obtient :

$$\Omega = \begin{pmatrix} (\sigma_\varepsilon^2 + T\sigma_\alpha^2 + N\sigma_\beta^2)B_{NT} & +\sigma_\varepsilon^2Q_W \\ +(\sigma_\varepsilon^2 + T\sigma_\alpha^2)Q_I & +(\sigma_\varepsilon^2 + N\sigma_\beta^2)Q_T, \end{pmatrix} \quad (9.1)$$

on en déduit que :

$$Q_I\Omega Q_I = (\sigma_\varepsilon^2 + T\sigma_\alpha^2)Q_I,$$

de sorte que :

$$V(\hat{b}_I|X) = (\sigma_\varepsilon^2 + T\sigma_\alpha^2)(\underline{X}'Q_I\underline{X})^{-1}.$$

9.2.2 Le between temporel

Pour cette estimation, on effectue d'abord les moyennes des variables sur tous les individus afin d'obtenir des séries temporelles, puis on les centre. Enfin, on applique les moindres carrés ordinaires aux données ainsi transformées. Cette régression ne comporte pas non plus de terme constant. On remarquera que chaque moyenne temporelle centrée est répliquée N fois :

$$\begin{aligned} \hat{b}_T &= \left((Q_T\underline{X})' Q_T\underline{X} \right)^{-1} (Q_T\underline{X})' Q_T y \\ &= (\underline{X}' Q_T \underline{X})^{-1} \underline{X}' Q_T y, \end{aligned}$$

cet estimateur est sans biais car $E(u|X) = 0$. La variance est donnée par :

$$\begin{aligned} V(\hat{b}_T|X) &= E \left[(\hat{b}_T - b)(\hat{b}_T - b)' \middle| X \right] \\ &= (\underline{X}' Q_T \underline{X})^{-1} \underline{X}' Q_T \Omega Q_T \underline{X} (\underline{X}' Q_T \underline{X})^{-1}. \end{aligned}$$

En utilisant (9.1) on obtient :

$$Q_T \Omega Q_T = (\sigma_\varepsilon^2 + N\sigma_\beta^2) Q_T,$$

de sorte que :

$$V(\hat{b}_T|X) = (\sigma_\varepsilon^2 + N\sigma_\beta^2)(\underline{X}' Q_T \underline{X})^{-1}.$$

9.2.3 Le double within

Pour cette estimation, on retranche aux variables leurs moyennes individuelles et temporelles puis on ajoute leur moyenne globale. On applique ensuite les moindres carrés ordinaires aux données transformées. Cette régression ne comporte pas de terme constant. L'estimateur est défini par :

$$\begin{aligned} \hat{b}_{DW} &= \left((Q_W X)' Q_W X \right)^{-1} (Q_W X)' Q_W y \\ &= (X' Q_W X)^{-1} X' Q_W y, \end{aligned}$$

cet estimateur est sans biais car $E(u) = 0$. La variance est donnée par :

$$V(\hat{b}_{DW}|X) = (\underline{X}'Q_W\underline{X})^{-1} \underline{X}'Q_W\Omega Q_W\underline{X}(\underline{X}'Q_W\underline{X})^{-1}.$$

En utilisant (9.1) on obtient :

$$Q_W\Omega Q_W = \sigma_\varepsilon^2 Q_W,$$

de sorte que :

$$V(\hat{b}_{DW}|X) = \sigma_\varepsilon^2 (\underline{X}'Q_W\underline{X})^{-1}.$$

9.2.4 Les moindres carrés généralisés

Cet estimateur est défini par :

$$b_{IT}^* = (\underline{X}'\Omega^{-1}\underline{X})^{-1} \underline{X}'\Omega^{-1}y,$$

de matrice de covariance :

$$V(b_{IT}^*|X) = (\underline{X}'\Omega^{-1}\underline{X})^{-1}.$$

Pour obtenir cet estimateur il suffit de réaliser des moindres carrés ordinaires sur les données pré-multipliées par $\Omega^{-1/2}$. En utilisant le fait que l'expression (9.1) de Ω est une somme pondérée de matrices de projection orthogonales, orthogonales entre elles et sommante à l'identité, il suffit d'élever à la puissance $-\frac{1}{2}$ les coefficients des ces matrices. On multiplie cette matrice par σ_ε pour obtenir des expressions plus faciles à interpréter :

$$\begin{aligned} \sigma_\varepsilon \times \Omega^{-1/2} = & \underbrace{\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2 + N\sigma_\beta^2}}}_{\theta_3} B_{NT} + Q_W \\ & + \underbrace{\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}}}_{\theta_1} Q_I + \underbrace{\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + N\sigma_\beta^2}}}_{\theta_2} Q_T, \end{aligned}$$

ceci est équivalent à travailler sur les transformations suivantes des données :

$$\begin{aligned} y_{it}^* &= \theta_3 y_{i\bullet} + y_{it} - (y_{i\bullet} - y_{\bullet\bullet}) - (y_{\bullet t} - y_{\bullet\bullet}) + \theta_1 (y_{i\bullet} - y_{\bullet\bullet}) + \theta_2 (y_{\bullet t} - y_{\bullet\bullet}) \\ &= y_{it} - (1 - \theta_1) y_{i\bullet} - (1 - \theta_2) y_{\bullet t} + (1 - \theta_1 - \theta_2 + \theta_3) y_{\bullet\bullet} \end{aligned}$$

On peut étudier quelques cas extrêmes qui aboutissent à des estimateurs usuels :

1. Pas d'effet individuel ($\sigma_\alpha^2 = 0$).

(a) Ceci implique $\theta_1 = 1$ et $\theta_2 = \theta_3$. On a alors : $y_{it}^* = y_{it} - (1 - \theta_2)y_{\bullet t}$. Dans ce cas, les moyennes individuelles n'interviennent plus dans le calcul de l'estimateur optimal. Ceci provient du fait que l'effet aléatoire individuel n'influence plus la variance de l'estimateur.

(b) Si on ajoute de plus que $N \rightarrow \infty$, alors $\theta_2 = 0$ et l'on obtient l'estimateur *within* temporel basé sur $y_{it}^* = y_{it} - y_{\bullet t}$. Ceci provient du fait que, sous cette hypothèse supplémentaire, la matrice de covariance de la perturbation du modèle en *within* temporel est scalaire. Cet estimateur est alors optimal.

2. Pas d'effet temporel ($\sigma_\beta^2 = 0$)

(a) Ceci implique $\theta_2 = 1$ et $\theta_1 = \theta_3$. La transformation devient $y_{it}^* = y_{it} - (1 - \theta_1)y_{i\bullet}$ et l'on retrouve l'estimateur optimal du modèle à effet aléatoire individuel que nous avons déjà vu.

(b) Si l'on a de plus $T \rightarrow \infty$, alors $\theta_1 = 0$ et l'on obtient l'estimateur *within* individuel avec $y_{it}^* = y_{it} - y_{i\bullet}$, dont nous avons vu qu'il était optimal dans ce cas.

3. Ni effet temporel ni effet individuel ($\sigma_\alpha^2 = 0$ et $\sigma_\beta^2 = 0$). Dans ce cas $\theta_1 = \theta_2 = \theta_3 = 1$ ce qui donne la transformation $y_{it}^* = y_{it}$ et l'on retrouve l'application des moindres carrés ordinaires, ce qui était attendu car le modèle se réduit alors à un modèle à erreur simple.

Plus généralement, l'estimateur des moindres carrés généralisés s'obtient en appliquant les moindres carrés ordinaires aux transformations suivantes :

$$y^* = (Q_W + \theta_1 Q_I + \theta_2 Q_T + \theta_3 B_{NT}) y$$

$$\text{et } \underline{X}^* = (Q_W + \theta_1 Q_I + \theta_2 Q_T + \theta_3 B_{NT}) \underline{X},$$

ce qui donne :

$$b_{IT}^* = (\underline{X}^{*'} \underline{X}^*)^{-1} \underline{X}^{*'} y^*,$$

pour la variance, il suffit de ne pas oublier le terme en σ_ε^2 :

$$\begin{aligned} V(b_{IT}^* | X) &= (\underline{X}' \Omega^{-1} \underline{X})^{-1} \\ &= \left(\underline{X}' \frac{1}{\sigma_\varepsilon^2} (Q_W + \theta_1^2 Q_I + \theta_2^2 Q_T + \theta_3^2 B_{NT}) \underline{X} \right)^{-1} \\ &= \sigma_\varepsilon^2 (\underline{X}^{*'} \underline{X}^*)^{-1}. \end{aligned}$$

9.3 Estimation des composantes de la variance

Pour pouvoir estimer les matrices de covariance des estimateurs précédentes ainsi que pour pouvoir réaliser l'estimation optimale par les moindres carrés quasi généralisés, il faut disposer d'estimateurs de σ_a^2 , σ_β^2 et σ_ε^2 . Pour cela, nous allons nous baser sur les sommes des carrés des résidus des trois estimateurs de base du modèle général à erreur composée : le *between* individuel, le *between* temporel et le *double within*.

9.3.1 Estimation de σ_ε^2

Pour réaliser cette estimation, nous allons nous baser sur les résidus de la régression en *double within*. Le résidu de cette régression s'écrit :

$$\begin{aligned}\hat{u}_{DW} &= Q_W y - Q_W X \hat{b}_{DW} \\ &= Q_W y - Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W y \\ &= Q_W y - Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W Q_W y \\ &= \left[I_{NT} - Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W \right] Q_W y,\end{aligned}$$

et l'on vérifie que la matrice suivante :

$$M_{DW} = I_{NT} - Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W,$$

est symétrique ($M'_{DW} = M_{DW}$) et idempotente ($M^2_{DW} = M_{DW}$). En développant l'expression de y , on obtient :

$$\hat{u}_{DW} = M_{DW} Q_W \underline{X} b + M_{DW} Q_W u,$$

or on a :

$$M_{DW} Q_W \underline{X} = Q_W \underline{X} - Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W \underline{X} = 0,$$

ce qui implique :

$$\hat{u}_{DW} = M_{DW} Q_W u,$$

d'où la somme des carrés des résidus :

$$\begin{aligned}\hat{S}_{DW} &= \hat{u}'_{DW} \hat{u}_{DW} \\ &= u' Q'_W M'_{DW} M_{DW} Q_W u \\ &= u' Q_W M_{DW} Q_W u,\end{aligned}$$

et son espérance mathématique :

$$\begin{aligned}
 E(\hat{S}_{DW} | X) &= E(\text{tr}(\hat{S}_{DW}) | X) \\
 &= E[\text{tr}(u' Q_W M_{DW} Q_W u) | X] \\
 &= E[\text{tr}(u u' Q_W M_{DW} Q_W) | X] \\
 &= \text{tr}[E(u u' Q_W M_{DW} Q_W) | X] \\
 &= \text{tr}(\Omega Q_W M_{DW} Q_W) \\
 &= \text{tr}(Q_W \Omega Q_W M_{DW}),
 \end{aligned}$$

or

$$Q_W \Omega Q_W = \sigma_\varepsilon^2 Q_W,$$

ce qui implique :

$$E(\hat{S}_{DW} | X) = \sigma_\varepsilon^2 \text{tr}(Q_W M_{DW}).$$

La trace de l'expression précédente est égale aux degrés de libertés de la somme des carrés des résidus. Pour la calculer, il va nous falloir déterminer les traces de toutes les matrices de projections parce que Q_W est définie comme une différence :

$$Q_W = I_{NT} - B_{NT} - Q_I - Q_T.$$

9.3.2 Trace des matrices de projection

On a :

$$\text{tr}(I_{NT}) = NT,$$

car une matrice identité est composée de 1 sur sa diagonale. Sa trace est donc égale à sa dimension. D'autre part :

$$\begin{aligned}
 \text{tr}(B_{NT}) &= \text{tr}_{NT}(e'_{NT} e_{NT})^{-1} e'_{NT} \\
 &= \text{tr}_{NT} e'_{NT} e_{NT} (e'_{NT} e_{NT})^{-1} \\
 &= \text{tr}(1) \\
 &= 1,
 \end{aligned}$$

ceci est vrai de l'ensemble des matrices de type *between* car la trace ne dépend pas de la dimension du vecteur unité e utilisé. En utilisant ces résultats, on obtient :

$$\begin{aligned}
 \text{tr}(Q_I) &= \text{tr}(I_N \otimes B_T) - \text{tr}(B_{NT}) \\
 &= \text{tr}(I_N) \times \underbrace{\text{tr}(B_T)}_1 - 1 \\
 &= N - 1,
 \end{aligned}$$

ainsi que :

$$\begin{aligned}\text{tr}(Q_T) &= \text{tr}(B_N \otimes I_T) - \text{tr}(B_{NT}) \\ &= \underbrace{\text{tr}(B_N)}_1 \times \text{tr}(I_T) - 1 \\ &= T - 1.\end{aligned}$$

On en déduit la trace recherchée :

$$\begin{aligned}\text{tr}(Q_W) &= \text{tr}(I_{NT}) - \text{tr}(B_{NT}) - \text{tr}(Q_I) - \text{tr}(Q_T) \\ &= NT - 1 - (N - 1) - (T - 1) \\ &= NT - N - T + 1 \\ &= (N - 1)(T - 1).\end{aligned}$$

Les degrés de liberté de la somme des carrés de la régression en *double within* sont donc égaux à :

$$\begin{aligned}\text{tr}(Q_W M_{DW}) &= \text{tr}\left(Q_W - Q_W^2 \underline{X} (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W\right) \\ &= \text{tr}(Q_W) - \underbrace{\text{tr}\left(\underline{X}' Q_W \underline{X} (\underline{X}' Q_W \underline{X})^{-1}\right)}_{I_{p-1}} \\ &= (N - 1)(T - 1) - (p - 1),\end{aligned}$$

où le terme en $p - 1$ provient du fait qu'il n'y a pas de terme constant dans la régression. Globalement, on a donc :

$$E(\hat{S}_{DW} | X) = \sigma_\varepsilon^2 [(N - 1)(T - 1) - (p - 1)],$$

soit :

$$\sigma_\varepsilon^2 = E\left[\frac{\hat{S}_{DW}}{(N - 1)(T - 1) - (p - 1)} \middle| X\right],$$

ce qui permet de sélectionner l'estimateur sans biais suivant de σ_ε^2 :

$$\widehat{\sigma_\varepsilon^2} = \frac{\hat{S}_{DW}}{(N - 1)(T - 1) - (p - 1)}.$$

9.3.3 Estimation de σ_α^2

Pour réaliser cette estimation, nous allons nous baser sur les résidus de la régression en *between* individuel. Le résidu de cette régression s'écrit :

$$\begin{aligned}\hat{u}_I &= Q_I y - Q_I \underline{X} \hat{b}_I \\ &= Q_I y - Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I y \\ &= Q_I y - Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I Q_I y \\ &= \left[I_{NT} - Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I \right] Q_I y,\end{aligned}$$

et l'on vérifie que la matrice suivante :

$$M_I = I_{NT} - Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I,$$

est symétrique ($M_I' = M_I$) et idempotente ($M_I^2 = M_I$). En développant l'expression de y , on obtient :

$$\hat{u}_I = M_I Q_I \underline{X} b + M_I Q_I u,$$

or on a :

$$M_I Q_I \underline{X} = Q_I \underline{X} - Q_I \underline{X} (\underline{X}' Q_I \underline{X})^{-1} \underline{X}' Q_I \underline{X} = 0,$$

ce qui implique :

$$\hat{u}_I = M_I Q_I u,$$

d'où la somme des carrés des résidus :

$$\begin{aligned}\hat{S}_I &= \hat{u}_I' \hat{u}_I \\ &= u' Q_I' M_I' M_I Q_I u \\ &= u' Q_I M_I Q_I u,\end{aligned}$$

et son espérance mathématique :

$$\begin{aligned}E(\hat{S}_I | X) &= E[\text{tr}(\hat{S}_I) | X] \\ &= E[\text{tr}(u' Q_I M_I Q_I u) | X] \\ &= E[\text{tr}(u u' Q_I M_I Q_I) | X] \\ &= \text{tr}[E(u u' | X) Q_I M_I Q_I] \\ &= \text{tr}(\Omega Q_I M_I Q_I) \\ &= \text{tr}(Q_I \Omega Q_I M_I),\end{aligned}$$

or

$$Q_I \Omega Q_I = (\sigma_\varepsilon^2 + T \sigma_\alpha^2) Q_I,$$

ce qui implique :

$$E(\hat{S}_I | X) = (\sigma_\varepsilon^2 + T\sigma_\alpha^2) \text{tr}(Q_I M_I)$$

et

$$\begin{aligned} \text{tr}(Q_I M_I) &= \text{tr}\left(Q_I - Q_I^2 X (X' Q_I X)^{-1} X' Q_I\right) \\ &= \text{tr}Q_I - \text{tr}I_{p-1} \\ &= (N-1) - (p-1) \\ &= N-p, \end{aligned}$$

on en déduit que :

$$\sigma_\alpha^2 = \frac{1}{T} \left[E\left(\frac{\hat{S}_I}{N-p} \middle| X\right) - \sigma_\varepsilon^2 \right],$$

ce qui permet de définir l'estimateur sans biais suivant :

$$\widehat{\sigma}_\alpha^2 = \max\left(0, \frac{1}{T} \left(\frac{\hat{S}_I}{N-p} - \widehat{\sigma}_\varepsilon^2\right)\right).$$

9.3.4 Estimation de σ_β^2

Pour réaliser cette estimation, nous allons nous baser sur les résidus de la régression en *between* temporel. Le résidu de cette régression s'écrit :

$$\hat{u}_T = Q_T y - Q_T \underline{X} \hat{b}_T = M_T Q_T y$$

avec :

$$M_T = I_{NT} - Q_T \underline{X} (\underline{X}' Q_T \underline{X})^{-1} \underline{X}' Q_T,$$

symétrique ($M_T' = M_T$) et idempotente ($M_T^2 = M_T$). En développant l'expression de y , on obtient :

$$\hat{u}_T = M_T Q_T \underline{X} b + M_T Q_T u,$$

or on a :

$$M_T Q_T \underline{X} = Q_T \underline{X} - Q_T \underline{X} (\underline{X}' Q_T \underline{X})^{-1} \underline{X}' Q_T \underline{X} = 0,$$

ce qui implique :

$$\hat{u}_T = M_T Q_T u,$$

d'où la somme des carrés des résidus :

$$\hat{S}_T = \hat{u}_T' \hat{u}_T = u' Q_T M_T Q_T u,$$

et son espérance mathématique :

$$E(\hat{S}_I | X) = \text{tr}(Q_T \Omega Q_T M_T),$$

or

$$Q_T \Omega Q_T = (\sigma_\varepsilon^2 + N\sigma_\beta^2) Q_T,$$

ce qui implique :

$$E(\hat{S}_T | X) = (\sigma_\varepsilon^2 + N\sigma_\beta^2) \text{tr}(Q_T M_T)$$

et

$$\begin{aligned} \text{tr}(Q_I M_I) &= \text{tr}\left(Q_T - Q_T^2 \underline{X} (\underline{X}' Q_T \underline{X})^{-1} \underline{X}' Q_T\right) \\ &= \text{tr} Q_T - \text{tr} I_{p-1} \\ &= (T-1) - (p-1) \\ &= T-p, \end{aligned}$$

on en déduit que :

$$\sigma_\beta^2 = \frac{1}{N} \left[E\left(\frac{\hat{S}_T}{T-p} \middle| X\right) - \sigma_\varepsilon^2 \right],$$

ce qui permet de définir l'estimateur sans biais suivant :

$$\widehat{\sigma}_\beta^2 = \max\left(0, \frac{1}{N} \left(\frac{\hat{S}_T}{T-p} - \widehat{\sigma}_\varepsilon^2\right)\right).$$

On utilise ensuite les estimateurs $\widehat{\sigma}_\varepsilon^2$, $\widehat{\sigma}_\alpha^2$ et $\widehat{\sigma}_\beta^2$ pour effectuer des moindres carrés quasi généralisés, ce qui donne les estimateurs convergents :

$$\hat{\theta}_1 = \sqrt{\frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_\varepsilon^2 + T \widehat{\sigma}_\alpha^2}}, \hat{\theta}_2 = \sqrt{\frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_\varepsilon^2 + N \widehat{\sigma}_\beta^2}} \text{ et } \hat{\theta}_3 = \sqrt{\frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_\varepsilon^2 + T \widehat{\sigma}_\alpha^2 + N \widehat{\sigma}_\beta^2}}.$$

Plus précisément, on réalise les transformations :

$$\begin{aligned} \tilde{y}_{it} &= y_{it} - (1-\hat{\theta}_1) y_{i\bullet} - (1-\hat{\theta}_2) y_{\bullet t} + (1-\hat{\theta}_1 - \hat{\theta}_2 + \hat{\theta}_3) y_{\bullet\bullet}, \\ \tilde{\underline{X}}_{it} &= X_{it} - (1-\hat{\theta}_1) \underline{X}_{i\bullet} - (1-\hat{\theta}_2) \underline{X}_{\bullet t} + (1-\hat{\theta}_1 - \hat{\theta}_2 + \hat{\theta}_3) \underline{X}_{\bullet\bullet}, \end{aligned}$$

ce qui donne l'estimateur :

$$\tilde{b}_{IT} = (\tilde{\underline{X}}' \tilde{\underline{X}})^{-1} \tilde{\underline{X}}' \tilde{y},$$

et l'estimateur de sa variance :

$$V(\tilde{b}_{IT} | X) = \widehat{\sigma}_\varepsilon^2 (\tilde{\underline{X}}' \tilde{\underline{X}})^{-1}.$$

9.4 Tests de spécification

9.4.1 Test de Hausman

L'hypothèse nulle comporte maintenant deux possibilités car il y a deux effets potentiellement corrélés aux variables explicatives dans le modèle :

$$\begin{cases} H_0 : \text{Plim} \frac{1}{NT} \underline{X}' \alpha = 0 \text{ et } \text{Plim} \frac{1}{NT} \underline{X}' \beta = 0 \\ H_a : \text{Plim} \frac{1}{NT} \underline{X}' \alpha \neq 0 \text{ ou } \text{Plim} \frac{1}{NT} \underline{X}' \beta \neq 0 \end{cases}$$

L'estimateur du *double within* est convergent sous les deux hypothèses, alors que l'estimateur des MCQG est optimal sous l'hypothèse nulle et non convergent sous l'hypothèse alternative. Examinons la covariance de ces deux estimateurs :

$$\begin{aligned} \text{Cov}(\hat{b}_{\text{DW}}, b_{\text{IT}}^* | X) &= E\left((\hat{b}_{\text{DW}} - b)(b_{\text{IT}}^* - b)' \middle| X\right) \\ &= E\left((\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W u u' \Omega^{-1} \underline{X} (\underline{X}' \Omega^{-1} \underline{X})^{-1} \middle| X\right) \\ &= (\underline{X}' Q_W \underline{X})^{-1} \underline{X}' Q_W \Omega \Omega^{-1} \underline{X} (\underline{X}' \Omega^{-1} \underline{X})^{-1} \\ &= (\underline{X}' \Omega^{-1} \underline{X})^{-1}, \end{aligned}$$

ce qui implique :

$$\begin{aligned} V(\hat{b}_{\text{DW}} - b_{\text{IT}}^* | X) &= V(\hat{b}_{\text{DW}} | X) + V(b_{\text{IT}}^* | X) - 2\text{Cov}(\hat{b}_{\text{DW}}, b_{\text{IT}}^* | X) \\ &= V(\hat{b}_{\text{DW}} | X) - V(b_{\text{IT}}^* | X), \end{aligned}$$

on peut donc utiliser la statistique :

$$H = (\hat{b}_{\text{DW}} - b_{\text{IT}}^*)' [V(\hat{b}_{\text{DW}} | X) - V(b_{\text{IT}}^* | X)]^{-1} (\hat{b}_{\text{DW}} - b_{\text{IT}}^*) \xrightarrow{L} \chi_{p-1}^2,$$

ou la statistique asymptotiquement équivalente :

$$\tilde{H} = (\hat{b}_{\text{DW}} - \tilde{b}_{\text{IT}})' [V(\hat{b}_{\text{DW}} | X) - V(\tilde{b}_{\text{IT}} | X)]^{-1} (\hat{b}_{\text{DW}} - \tilde{b}_{\text{IT}}) \xrightarrow{L} \chi_{p-1}^2,$$

car b_{IT}^* et \tilde{b}_{IT} ont la même distribution asymptotique.

9.4.2 Tests à la Mundlak

L'idée des tests qui suivent est d'exploiter l'absence de corrélation entre les estimateurs \hat{b}_{DW} , \hat{b}_I et \hat{b}_T afin de simplifier le calcul des matrices de covariance.

Corrélation avec l'effet individuel

L'estimateur *double within* est toujours convergent en présence de corrélation des variables explicatives avec les effets individuels ou temporels. D'autre part l'estimateur *between* individuel n'est pas convergent en présence de corrélation entre les effets individuels et les variables explicatives tout en restant convergent en présence d'effets temporels. En effet, le modèle s'écrit :

$$y_{it} = X_{it}b + \alpha_i + \beta_t + \varepsilon_{it},$$

ce qui implique :

$$\begin{aligned} y_{i\bullet} &= X_{i\bullet}b + \alpha_i + \frac{1}{T} \sum_{t=1}^T \beta_t + \varepsilon_{i\bullet} \\ &= X_{i\bullet}b + \alpha_i + \beta_{\bullet} + \varepsilon_{i\bullet}, \end{aligned}$$

ainsi que

$$\begin{aligned} y_{\bullet\bullet} &= X_{\bullet\bullet}b + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \alpha_i + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \beta_t + \varepsilon_{\bullet\bullet} \\ &= X_{\bullet\bullet}b + \frac{1}{N} \sum_{i=1}^N \frac{T\alpha_i}{T} + \frac{1}{T} \sum_{t=1}^T \frac{N\beta_t}{N} + \varepsilon_{\bullet\bullet} \\ &= X_{\bullet\bullet}b + \frac{1}{N} \sum_{i=1}^N \alpha_i + \frac{1}{T} \sum_{t=1}^T \beta_t + \varepsilon_{\bullet\bullet} \\ &= X_{\bullet\bullet}b + \alpha_{\bullet} + \beta_{\bullet} + \varepsilon_{\bullet\bullet} \end{aligned}$$

en faisant la différence des deux dernières équations, on obtient la régression correspondant au *between* individuel :

$$y_{i\bullet} - y_{\bullet\bullet} = (X_{i\bullet} - X_{\bullet\bullet})b + \alpha_i - \alpha_{\bullet} + \varepsilon_{i\bullet} - \varepsilon_{\bullet\bullet},$$

et l'on constate qu'il n'y a plus de termes en β . L'effet temporel étant éliminé, il ne peut pas influencer la convergence de l'estimateur de b que l'on obtient dans cette régression. On peut donc tester séparément l'hypothèse d'absence de corrélation avec les effets individuels :

$$\begin{cases} H_0 : \text{Plim} \frac{1}{NT} X' \alpha = 0 \\ H_a : \text{Plim} \frac{1}{NT} X' \alpha \neq 0 \end{cases}$$

et la statistique est donnée par :

$$M_I = (\hat{b}_{DW} - \hat{b}_I)' [V(\hat{b}_{DW}|X) + V(\hat{b}_I|X)]^{-1} (\hat{b}_{DW} - \hat{b}_I) \xrightarrow{L} \chi_{p-1}^2,$$

car :

$$\text{Cov}(\hat{b}_{DW}, \hat{b}_I|X) = (X'Q_W X)^{-1} X' \underbrace{Q_W \Omega Q_I X}_0 (X'Q_W X)^{-1} = 0$$

Corrélation avec l'effet temporel

Il s'agit du cas symétrique du précédent. Cette fois-ci on effectue le centrage par rapport aux moyennes temporelles. Le modèle en moyennes temporelles s'écrit :

$$y_{\bullet t} = X_{\bullet t} b + \alpha_{\bullet} + \beta_t + \varepsilon_{\bullet t},$$

une fois centré par rapport à la moyenne globale, on obtient la régression correspondant au *between* temporel :

$$y_{\bullet t} - y_{\bullet\bullet} = (X_{\bullet t} - X_{\bullet\bullet}) b + \beta_t - \beta_{\bullet} + \varepsilon_{\bullet t} - \varepsilon_{\bullet\bullet},$$

qui ne contient aucune fonction des effets individuels α_i . On peut donc effectuer séparément le test d'absence de corrélation avec les effets temporels :

$$\begin{cases} H_0 : \text{Plim}_{\frac{1}{NT}} X' \beta = 0 \\ H_a : \text{Plim}_{\frac{1}{NT}} X' \beta \neq 0 \end{cases}$$

où la statistique est donnée par :

$$M_T = (\hat{b}_{DW} - \hat{b}_T)' [V(\hat{b}_{DW}|X) + V(\hat{b}_T|X)]^{-1} (\hat{b}_{DW} - \hat{b}_T) \xrightarrow{L} \chi_{p-1}^2,$$

car :

$$\text{Cov}(\hat{b}_{DW}, \hat{b}_T|X) = (X'Q_W X)^{-1} X' \underbrace{Q_W \Omega Q_T X}_0 (X'Q_T X)^{-1} = 0$$

9.5 Programmation sous IML

Pour effectuer cette estimation, on prend le panel le plus long possible : 1960 – 1990, soit $T = 31$. Ceci fait évidemment chuter le nombre d'entreprises, ici à $N = 41$. Inutile de signaler que cela réduit considérablement la représentativité de l'échantillon, mais c'est suffisant pour donner un exemple de programmation. Au total, nous avons donc $31 \times 41 = 1271$ observations. Le programme qui suit est une simple extension de la programmation du chapitre précédent.

Programme 9.1.

```
data tab; set rep.bal_hmuspan_2;
proc sort data=tab; by firm;

proc means noprint data=tab; by firm;
var LogS LogL LogC LogK;
```

```

output out=Qi mean=;

proc standard mean=0 data=Qi out=Qi;
var LogS LogL LogC LogK;
proc sort data=tab; by year;

proc means noprint data=tab; by year;
var LogS LogL LogC LogK;
output out=Qt mean=;

proc standard mean=0 data=Qt out=Qt;
var LogS LogL LogC LogK;

proc sort data=tab; by firm year; /* important */

proc iml;
ny=LogS; nx=LogL LogC LogK; nx=t(nx);
use Qi;
read all var (ny) into biy;
read all var (nx) into bix;
close Qi;
use Qt;
read all var (ny) into bty;
read all var (nx) into btx;
close Qt;
use tab;
read all var (ny) into y;
read all var (nx) into x;
close tab;
n=nrow(biy); t=nrow(bty); nt=n#t;
ent=J(nt,1,1);
nxc="Constante"//nx; p=ncol(x)+1;
print "Estimation sur panel cylindré",
"Nombre d'observations =" nt,
"Nombre d'individus =" n,
"Nombre de dates =" t;
start mco(y,x) global(ixx);
ixx=inv(t(x)*x);
b=ixx*t(x)*y;

```

```

return (b);
finish mco;
/* between individuel */
bi=mco(biy,bix);
ubi=biy-bix*bi;
si=ssq(ubi);
ibi=ixx;
/* between temporel */
bt=mco(bty,btx);
ubt=bty-btx*bt;
st=ssq(ubt);
ibt=ixx;
/* double within */
my=y[+]/nt; mx=x[+,]/nt;
et=J(T,1,1); en=J(N,1,1);
dwy=y-ent@my-biy@et-en@bty;
dwx=x-ent@mx-bix@et-en@btx;
bdw=mco(dwy,dwx);
udw=dwy-dwx*bdw;
sdw=ssq(udw);
idw=ixx;
/* composantes de la variance */
se2=sdw/((N-1)*(T-1)-(p-1));
sa2=max(0,si/(N-p)-se2/T);
sb2=max(0,st/(T-p)-se2/N);
/* estimateur MCQG */
theta1=sqrt(se2/(se2+T*sa2));
theta2=sqrt(se2/(se2+N*sb2));
theta3=sqrt(se2/(se2+T*sa2+N*sb2));
oy=dwy+theta1*(biy@et)+theta2*(en@bty)+
theta3*(ent@my);
ox=theta3*ent+theta1*(bix@et)+theta2*(en@btx)
+theta3*(ent@mx);
bo=mco(oy,ox);
io=ixx;
/* variance des estimateurs et résultats */
vbi=(si/(N-p))*ibi;
vbt=(st/(T-p))*ibt;
vdw=se2*idw;

```

```

vbo=se2#io;
/* résultats */
sbi=sqrt(vecdiag(vbi));
tbi=abs(bi)/sbi;
print "Estimation Between Individuel",
"N=" N,
"Variable expliquée :" ny, bi [rowname=nx] sbi tbi;
sbt=sqrt(vecdiag(vbt));
tbt=abs(bt)/sbt;
print "Estimation Between Temporel",
"T=" T,
"Variable expliquée :" ny, bt [rowname=nx] sbt tbt;
sdw=sqrt(vecdiag(vdw));
tdw=abs(bdw)/sdw;
print "Estimation Double Within",
"NT=" nt,
"Variable expliquée :" ny, bdw [rowname=nx] sdw tdw;
sbo=sqrt(vecdiag(vbo));
tbo=abs(bo)/sbo;
print "Variance idiosyncratique =" se2,
"Variance de l'effet individuel =" sa2,
"Variance de l'effet temporel =" sb2,,
"Estimation des MCQG",
"Variable expliquée :" ny, bo [rowname=nxc] sbo tbo;
/* test de Hausman */
rg=2:p;
test=vdw-vbo[rg,rg];
h=t(bdw-bo[rg])*ginv(vdw-vbo[rg,rg])*(bdw-bo[rg]);
ph=1-probchi(h,p-1);
print "Test de Hausman", "statistique =" h,
"probabilité critique =" ph;
/* Tests de Mundlak */
m1=t(bdw-bi)*ginv(vdw+vbi)*(bdw-bi);
pml=1-probchi(m1,p-1);
print "Mundlak : Corrélation avec l'effet individuel",
"statistique =" m1, "probabilité critique =" pml;
m2=t(bdw-bt)*ginv(vdw+vbt)*(bdw-bt);
pm2=1-probchi(m2,p-1);
print "Mundlak : Corrélation avec l'effet temporel",

```



```
"statistique =" m2, "probabilité critique =" pm2;
quit;
run;
```

Sortie 9.1.

```
Estimation sur panel cylindré
NT

Nombre d'observations =      1271
N

Nombre d'individus =        41
T

Nombre de dates =          31

Estimation Between Individuel
N

N=          41
NY

Variable expliquée : LOGS
BI          SBI          TBI

LOGL 0.6030371 0.0811327 7.4327285
LOGC 0.455305 0.0496934 9.1622752
LOGK -0.049375 0.0638922 0.7727838

Estimation Between Temporel
T

T=          31
NY

Variable expliquée : LOGS
BT          SBT          TBT

LOGL 0.5180195 0.090003 5.7555801
LOGC 0.0637951 0.0871724 0.7318267
LOGK 0.7644023 0.079504 9.6146379
```

Estimation Double Within
NT

NT= 1271
NY

Variable expliquée : LOGS
BDW SDW TDW
LOGL 0.6888775 0.019992 34.457708
LOGC 0.1187953 0.0172674 6.8797484
LOGK 0.1289523 0.0178655 7.2179368

SE2

Variance idiosyncratique = 0.0178619
SA2

Variance de l'effet individuel = 0.0548294
SB2

Variance de l'effet temporel = 0.0031511

Estimation des MCQG
NY

Variable expliquée : LOGS
BO SBO TBO
Constante 1.9785197 0.0570619 34.673229
LOGL 0.3504993 0.0142837 24.538492
LOGC 0.3493614 0.0140109 24.934903
LOGK 0.3523259 0.0143598 24.535552

Test de Hausman
H

statistique = 622.21892
PH

probabilité critique = 0

Mundlak : Corrélation avec l'effet individuel
M1

statistique = 50.324864

PM1

probabilité critique = 6.813E-11

Mundlak : Corrélation avec l'effet temporel
M2

statistique = 657.06405

PM2

probabilité critique = 0

Les résultats produits par les estimateurs *between* individuels et temporels semblent assez irréalistes pour une fonction de production. On peut toutefois les comprendre de la manière suivante : en présence d'un effet individuel corrélé, le *between* individuel n'est pas convergent, et en présence d'un effet temporel corrélé, le *between* temporel n'est pas convergent non plus. Or les tests que nous effectuons à la fin du programme montrent clairement que les effets individuels et temporels sont tous les deux corrélés. On doit donc rejeter les estimations *between* individuelle et temporelle. Le seul estimateur convergent dans ce contexte est le *double within*. On constate effectivement que c'est l'estimateur qui donne les estimations les plus réalistes (pour un échantillon de 41 entreprises).

Chapitre 10

Estimation sur panels non cylindrés

Les individus d'un panel ne sont pas forcément présents à toutes les dates. Ceci peut provenir d'une réponse non parvenue, d'une erreur de saisie de l'identifiant de l'individu ou encore du mode de collecte des données. Ainsi, on peut décider d'interroger la moitié de l'échantillon à la date suivante afin de renouveler la moitié de l'échantillon à chaque période. Dans ce dernier cas, on perdrait une très grande partie des informations si l'on imposait la présence à toutes les dates pour réaliser l'estimation. Dans ce chapitre, on suppose qu'un individu i est présent à T_i dates différentes, qui ne se suivent pas forcément.

Le point important est que les absences du fichier sont *purement aléatoires*, c'est-à-dire qu'elles sont indépendantes des variables du modèle que l'on estime. Le modèle que l'on estime est identique au modèle précédent, seuls les nombres d'observations par individu sont différents.

Les dates auxquelles les individus sont présents sont données par des fonctions $h_i(s)$, qui indiquent la s -ième date à laquelle l'individu i est présent. Les fonctions h_i sont implicitement fournies par la base de données. Si, par exemple, un individu i est présent aux dates $t = 3, 5, 9$, on aura :

$$h_i(1) = 3, h_i(2) = 5 \text{ et } h_i(3) = 9.$$

Dans tous les cas $h_i(1)$ donne la première date d'apparition de l'individu i et $h_i(T_i)$ sa dernière date d'apparition dans le fichier. Pour l'individu i on

observe le vecteur :

$$y_i = \begin{pmatrix} y_{i,h_i(1)} \\ \vdots \\ y_{i,h_i(t)} \\ \vdots \\ y_{i,h_i(T_i)} \end{pmatrix}.$$

En utilisant la même convention pour X_i et pour u_i , on peut écrire le modèle individu par individu :

$$y_i = X_i b + u_i,$$

la seule différence par rapport au cas usuel est que le nombre d'observations varie avec chaque individu. La variance de la perturbation s'écrit donc maintenant :

$$\begin{aligned} V(u_i|X) &= \sigma_\varepsilon^2 I_{T_i} + \sigma_\alpha^2 e_{T_i} e'_{T_i} \\ &= \sigma_\varepsilon^2 I_{T_i} + T_i \sigma_\alpha^2 B_{T_i}, \end{aligned}$$

un terme en T_i multiplie maintenant la variance individuelle σ_α^2 .

10.1 Estimation inter-individuelle (between)

On effectue maintenant les moyennes sur les T_i observations disponibles. Pour cela, on définit l'opérateur *between* pour l'individu i à partir d'un vecteur unité de dimension T_i . On a :

$$B_{T_i} = e_{T_i} \left(e'_{T_i} e_{T_i} \right)^{-1} e'_{T_i},$$

ce que l'on peut écrire :

$$B_{T_i} = \frac{1}{T_i} e_{T_i} e'_{T_i} = \frac{1}{T_i} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$

On vérifie que cette matrice est symétrique et idempotente :

$$B'_{T_i} = B_{T_i} \text{ et } B^2_{T_i} = B_{T_i}.$$

En appliquant cette transformation à une variable, par exemple y_i , on obtient :

$$\begin{aligned} \underbrace{B_{T_i} y_i}_{(T_i,1)} &= \frac{1}{T_i} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_{i,h_i(1)} \\ \vdots \\ y_{i,h_i(T_i)} \end{pmatrix} \\ &= \frac{1}{T_i} \begin{pmatrix} \sum_{s=1}^{T_i} y_{i,h_i(s)} \\ \vdots \\ \sum_{s=1}^{T_i} y_{i,h_i(s)} \end{pmatrix} \\ &= \begin{pmatrix} y_{i\bullet} \\ \vdots \\ y_{i\bullet} \end{pmatrix}, \end{aligned}$$

où $y_{i\bullet}$ désigne la moyenne arithmétique sur les observations disponibles. Le modèle par individu s'écrit donc comme dans le cas usuel :

$$B_{T_i} y_i = B_{T_i} X_i b + B_{T_i} u_i, \quad (10.1)$$

le seul changement significatif concerne la perturbation du modèle. Sous les hypothèses du modèle à erreurs composées, la perturbation est toujours d'espérance nulle mais sa variance s'écrit :

$$\begin{aligned} V(B_{T_i} u_i | X) &= B_{T_i} V(u_i | X) B_{T_i} \\ &= B_{T_i} (\sigma_\varepsilon^2 I_{T_i} + T_i \sigma_\alpha^2 B_{T_i}) B_{T_i} \\ &= (\sigma_\varepsilon^2 + T_i \sigma_\alpha^2) B_{T_i}. \end{aligned}$$

On peut alors empiler ces modèles individuels (10.1) sur l'ensemble de l'échantillon :

$$\begin{aligned} B_{T_1} y_1 &= B_{T_1} X_1 b + B_{T_1} u_1 \\ &\vdots \\ B_{T_i} y_i &= B_{T_i} X_i b + B_{T_i} u_i \\ &\vdots \\ B_{T_N} y_N &= B_{T_N} X_N b + B_{T_N} u_N \end{aligned}$$

ce qui se réécrit sous la forme :

$$\underbrace{\begin{pmatrix} B_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{T_N} \end{pmatrix}}_{(\sum_{i=1}^N T_i, \sum_{i=1}^N T_i)} \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, 1)} =$$

$$\begin{pmatrix} B_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{T_N} \end{pmatrix} \underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, p)} b + \begin{pmatrix} B_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{T_N} \end{pmatrix} \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, 1)}$$

soit en abrégé :

$$By = BXb + Bu, \quad (10.2)$$

avec

$$B = \text{diag}(B_{T_i}).$$

On vérifie que :

$$B^2 = \text{diag}(B_{T_i})^2 = \text{diag}(B_{T_i}^2) = \text{diag}(B_{T_i}) = B,$$

et que :

$$B' = \text{diag}(B_{T_i})' = \text{diag}(B_{T_i}') = \text{diag}(B_{T_i}) = B.$$

La perturbation est d'espérance nulle et de variance conditionnelle :

$$\begin{aligned} V(Bu|X) &= \text{diag}(V(B_{T_i} u_i | X)) \\ &= \text{diag}((\sigma_\varepsilon^2 + T_i \sigma_\alpha^2) B_{T_i}) \\ &= \sigma_\varepsilon^2 B + \sigma_\alpha^2 \text{diag}(T_i B_{T_i}). \end{aligned}$$

L'estimation inter-individuelle consiste à appliquer les moindres carrés ordinaires à la relation (10.2). On obtient :

$$\begin{aligned} \hat{b}_B &= (X'BX)^{-1} X'By \\ &= b + (X'BX)^{-1} X'Bu. \end{aligned}$$

L'estimateur *between* est donc sans biais. Sa variance est donnée par :¹

$$\begin{aligned}
V(\hat{b}_B | X) &= E \left[(\hat{b}_B - b)(\hat{b}_B - b)' \middle| X \right] \\
&= (X'BX)^{-1} X' E(Bu u' B' | X) X (X'BX)^{-1} \\
&= (X'BX)^{-1} X' V(Bu | X) X (X'BX)^{-1} \\
&= (X'BX)^{-1} X' (\sigma_\varepsilon^2 B + \sigma_\alpha^2 \text{diag}(T_i B_{T_i})) X (X'BX)^{-1} \\
&= \sigma_\varepsilon^2 (X'BX)^{-1} + \sigma_\alpha^2 (X'BX)^{-1} X' \text{diag}(T_i B_{T_i}) X (X'BX)^{-1} \\
&= \sigma_\varepsilon^2 (X'BX)^{-1} \\
&\quad + \sigma_\alpha^2 (X'BX)^{-1} (BX)' \text{diag}(T_i I_{T_i}) BX (X'BX)^{-1}
\end{aligned}$$

10.2 Estimation intra-individuelle (within)

On calcule les écarts à partir des moyennes individuelles précédentes. Ceci revient donc à définir la matrice *between* comme :

$$W_{T_i} = I_{T_i} - B_{T_i},$$

on vérifie que cette matrice définit une projection orthogonale :

$$W_{T_i}^2 = W_{T_i} \text{ et } W_{T_i}' = W_{T_i}.$$

En appliquant cette transformation à une variable, par exemple y_i , on obtient :

$$\underbrace{W_{T_i} y_i}_{(T_i, 1)} = \begin{pmatrix} y_{i, h_i(1)} \\ \vdots \\ y_{i, h_i(T_i)} \end{pmatrix} - \begin{pmatrix} y_{i \bullet} \\ \vdots \\ y_{i \bullet} \end{pmatrix} = \begin{pmatrix} y_{i, h_i(1)} - y_{i \bullet} \\ \vdots \\ y_{i, h_i(T_i)} - y_{i \bullet} \end{pmatrix}$$

Le modèle par individu s'écrit donc comme dans le cas usuel :

$$W_{T_i} y_i = W_{T_i} \underline{X}_i b + W_{T_i} u_i, \quad (10.3)$$

1. On utilise :

$$\begin{aligned}
\text{diag}(T_i B_{T_i}) &= \text{diag}(T_i B_{T_i}^2) \\
&= \text{diag}(B_{T_i}) \text{diag}(T_i I_{T_i}) \text{diag}(B_{T_i}) \\
&= B \text{diag}(T_i I_{T_i}) B.
\end{aligned}$$

le seul changement significatif concerne la perturbation du modèle. Sous les hypothèses du modèle à erreurs composées, la perturbation est toujours d'espérance nulle mais sa variance s'écrit :

$$\begin{aligned} V(W_{T_i} u_i | X) &= W_{T_i} V(u_i | X) W_{T_i} \\ &= W_{T_i} (\sigma_\varepsilon^2 I_{T_i} + T_i \sigma_\alpha^2 B_{T_i}) W_{T_i} \\ &= \sigma_\varepsilon^2 W_{T_i}. \end{aligned}$$

On peut alors empiler ces modèles individuels (10.1) sur l'ensemble de l'échantillon :

$$\begin{aligned} W_{T_1} y_1 &= W_{T_1} X_1 b + W_{T_1} u_1 \\ &\vdots \\ W_{T_i} y_i &= W_{T_i} X_i b + W_{T_i} u_i \\ &\vdots \\ W_{T_N} y_N &= W_{T_N} X_N b + W_{T_N} u_N \end{aligned}$$

ce qui se réécrit sous la forme :

$$\underbrace{\begin{pmatrix} W_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_{T_N} \end{pmatrix}}_{(\sum_{i=1}^N T_i, \sum_{i=1}^N T_i)} \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, 1)} = \begin{pmatrix} W_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_{T_N} \end{pmatrix} \underbrace{\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, p)} b + \begin{pmatrix} W_{T_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_{T_N} \end{pmatrix} \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}}_{(\sum_{i=1}^N T_i, 1)}$$

soit en abrégé :

$$Wy = W \underline{X} b + Wu, \quad (10.4)$$

avec

$$W = \text{diag}(W_{T_i}).$$

On vérifie que :

$$W^2 = \text{diag}(W_{T_i})^2 = \text{diag}(W_{T_i}^2) = \text{diag}(W_{T_i}) = W,$$

et que :

$$W' = \text{diag}(W_{T_i})' = \text{diag}(W_{T_i}') = \text{diag}(W_{T_i}) = W.$$

La perturbation est d'espérance nulle et de variance conditionnelle :

$$\begin{aligned} V(Wu|X) &= \text{diag}(V(W_{T_i}u_i|X)) \\ &= \text{diag}((\sigma_\varepsilon^2 + T_i\sigma_\alpha^2)W_{T_i}) \\ &= \sigma_\varepsilon^2 \text{diag}(W_{T_i}) \\ &= \sigma_\varepsilon^2 W. \end{aligned}$$

L'estimation inter-individuelle consiste à appliquer les moindres carrés ordinaires à la relation (10.4). On obtient :

$$\begin{aligned} \hat{b}_W &= (\underline{X}'W\underline{X})^{-1} \underline{X}'Wy \\ &= b + (\underline{X}'W\underline{X})^{-1} \underline{X}'Wu. \end{aligned}$$

L'estimateur *between* est donc sans biais. Sa variance est donnée par :

$$\begin{aligned} V(\hat{b}_W|X) &= E \left[(\hat{b}_W - b)(\hat{b}_W - b)' | X \right] \\ &= (\underline{X}'W\underline{X})^{-1} \underline{X}'E(Wu u'W'|X) \underline{X} (\underline{X}'W\underline{X})^{-1} \\ &= (\underline{X}'W\underline{X})^{-1} \underline{X}'V(Wu) \underline{X} (\underline{X}'W\underline{X})^{-1} \\ &= (\underline{X}'W\underline{X})^{-1} \underline{X}'\sigma_\varepsilon^2 W \underline{X} (\underline{X}'W\underline{X})^{-1} \\ &= \sigma_\varepsilon^2 (\underline{X}'W\underline{X})^{-1}. \end{aligned}$$

10.3 Estimation des composantes de la variance

10.3.1 Variance inter-individuelle (between)

La somme des carrés des résidus dans la dimension inter-individuelle est donnée par :

$$S_B = \hat{u}'_B \hat{u}_B = u' B M_B B u,$$

son espérance mathématique est donc égale à :

$$\begin{aligned}
 E(S_B|X) &= E(\text{tr}(u'BM_BBu)|X) \\
 &= E(\text{tr}(Bu u'BM_B)|X) \\
 &= E(\text{tr}(M_BBu u'B)|X) \\
 &= \text{tr}(M_BE(Bu u'B|X)) \\
 &= \text{tr}(M_BV(Bu|X)) \\
 &= \text{tr}(M_B(\sigma_\varepsilon^2B + \sigma_\alpha^2\text{diag}(T_iB_{T_i}))) \\
 &= \sigma_\varepsilon^2\text{tr}(M_BB) + \sigma_\alpha^2\text{tr}(M_B\text{diag}(T_iB_{T_i})).
 \end{aligned}$$

or

$$\begin{aligned}
 \text{tr}(M_BB) &= \text{tr}\left(\left(I - BX(X'BX)^{-1}X'B\right)B\right) \\
 &= \text{tr}B - \text{tr}\left(X'BX(X'BX)^{-1}\right) \\
 &= \text{tr}B - \text{tr}I_p
 \end{aligned}$$

et l'on a :

$$\begin{aligned}
 \text{tr}B &= \text{tr}(\text{diag}B_{T_i}) \\
 &= \sum_{i=1}^N \text{tr}B_{T_i} \\
 &= \sum_{i=1}^N \text{tr}\left(e_{T_i}\left(e'_{T_i}e_{T_i}\right)^{-1}e'_{T_i}\right) \\
 &= \sum_{i=1}^N \text{tr}\left(\underbrace{e'_{T_i}e_{T_i}\left(e'_{T_i}e_{T_i}\right)^{-1}}_1\right) \\
 &= N
 \end{aligned}$$

donc

$$E(S_B|X) = \sigma_\varepsilon^2(N - p) + \sigma_\alpha^2\text{tr}(M_B\text{diag}(T_iB_{T_i})).$$

Le second terme de cette expression ne peut être que partiellement simpli-

fié. On a :

$$\begin{aligned}
\text{tr}(M_B \text{diag}(T_i B_{T_i})) &= \text{tr}(\text{diag}(T_i B_{T_i})) \\
&\quad - \text{tr}\left(BX(X'BX)^{-1} X' B \text{diag}(T_i B_{T_i})\right) \\
&= \sum_{i=1}^N \text{tr}(T_i B_{T_i}) \\
&\quad - \text{tr}\left(X' B \text{diag}(T_i B_{T_i}) BX(X'BX)^{-1}\right) \\
&= \sum_{i=1}^N T_i \text{tr}(B_{T_i}) \\
&\quad - \text{tr}\left(X' \text{diag}(B_{T_i}) \text{diag}(T_i B_{T_i}) \text{diag}(B_{T_i}) X(X'BX)^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^N T_i - \text{tr}\left(X' \text{diag}(T_i B_{T_i}) X(X'BX)^{-1}\right), \\
&= \sum_{i=1}^N T_i - \underbrace{\text{tr}\left((BX)' \text{diag}(T_i I_{T_i}) BX(X'BX)^{-1}\right)}_{d_{BX}}
\end{aligned}$$

et la quantité d_{BX} doit être calculée sur la base de données. On remarque que les degrés de liberté vont maintenant dépendre des valeurs des variables explicatives.²

On a finalement :

$$E(S_B) = \sigma_\varepsilon^2(N - p) + \sigma_\alpha^2 \left(\sum_{i=1}^N T_i - d_{BX} \right)$$

10.3.2 Variance intra-individuelle (within)

La somme des carrés des résidus dans la dimension intra-individuelle est donnée par :

$$S_W = \hat{u}'_W \hat{u}_W = u' W M_W W u,$$

2. Si l'échantillon était cylindré, on aurait $d_{BX} = T \times p$.

son espérance mathématique est donc égale à :

$$\begin{aligned}
 E(S_W | X) &= E(\text{tr}(u' W M_W W u) | X) \\
 &= E(\text{tr}(W u u' W M_W) | X) \\
 &= E(\text{tr}(M_W W u u' W) | X) \\
 &= \text{tr}(M_W E(W u u' W | X)) \\
 &= \text{tr}(M_W V(W u | X)) \\
 &= \text{tr}(M_W \sigma_\varepsilon^2 W) \\
 &= \sigma_\varepsilon^2 \text{tr}(M_W W).
 \end{aligned}$$

or

$$\begin{aligned}
 \text{tr}(M_W W) &= \text{tr}\left(\left(I - W \underline{X} (\underline{X}' W \underline{X})^{-1} \underline{X}' W\right) W\right) \\
 &= \text{tr}W - \text{tr}\left(\underline{X}' W \underline{X} (\underline{X}' W \underline{X})^{-1}\right) \\
 &= \text{tr}W - \text{tr}I_{p-1}
 \end{aligned}$$

et l'on a :

$$\begin{aligned}
 \text{tr}W &= \text{tr}(\text{diag}W_{T_i}) \\
 &= \sum_{i=1}^N \text{tr}W_{T_i} \\
 &= \sum_{i=1}^N \text{tr}(I_{T_i} - B_{T_i}) \\
 &= \sum_{i=1}^N (T_i - 1) \\
 &= \sum_{i=1}^N T_i - N
 \end{aligned}$$

donc

$$E(S_W | X) = \sigma_\varepsilon^2 \left(\sum_{i=1}^N T_i - N - p + 1 \right).$$

On peut donc estimer la composante σ_ε^2 par l'estimateur sans biais :

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{u}'_W \hat{u}_W}{\sum_{i=1}^N T_i - N - p + 1},$$

et l'autre composante s'obtient par l'estimateur sans biais :

$$\hat{\sigma}_\alpha^2 = \max\left(0, \frac{\hat{u}'_B \hat{u}_B - \hat{\sigma}_\varepsilon^2 (N-p)}{\sum_{i=1}^N T_i - d_{BX}}\right),$$

où \hat{u}_B est obtenu sur données dupliquées. Le terme d_{BX} peut être calculé facilement à partir d'un langage matriciel. Toutefois, quand ce n'est pas possible, on peut utiliser un autre estimateur qui est convergent. Pour voir cela, regardons la variance de la perturbation *between* pour un individu, *sans réplication* :

$$V(u_{i\bullet} | X) = \frac{\sigma_\varepsilon^2}{T_i} + \sigma_\alpha^2 \Leftrightarrow \sigma_\alpha^2 = V(u_{i\bullet} | X) - \frac{\sigma_\varepsilon^2}{T_i},$$

ce qui implique :³

$$\sigma_\alpha^2 = \frac{1}{N} \sum_{i=1}^N \left(E(u_{i\bullet}^2 | X) - \frac{\sigma_\varepsilon^2}{T_i} \right)$$

on en déduit l'estimateur :

$$\tilde{\sigma}_\alpha^2 = \max\left(0, \frac{1}{N} \sum_{i=1}^N \left(\hat{u}_{Bi}^2 - \frac{\hat{\sigma}_\varepsilon^2}{T_i} \right)\right),$$

il est biaisé mais convergent, et peut être calculé facilement.

10.4 Estimation par les moindres carrés généralisés

Le modèle pour un individu i s'écrit

$$y_i = X_i b + u_i, \quad V(u_i | X) \triangleq \Omega_i$$

(T_{i,1}) (T_{i,1}) (T_{i,1})

avec :

$$\begin{aligned} \Omega_i &= \sigma_\varepsilon^2 I_{T_i} + T_i \sigma_\alpha^2 B_{T_i} \\ &= \sigma_\varepsilon^2 W_{T_i} + (\sigma_\varepsilon^2 + T_i \sigma_\alpha^2) B_{T_i}, \end{aligned}$$

dont l'inverse est donnée par :⁴

$$\Omega_i^{-1} = \frac{1}{\sigma_\varepsilon^2} \left(W_{T_i} + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2} B_{T_i} \right),$$

3. La perturbation $u_{i\bullet}$ est d'espérance nulle.

4. On utilise le fait que B_{T_i} et W_{T_i} sont des matrices de projection orthogonales, qui sont orthogonales entre elles.

la seule différence par rapport au cas usuel est que le nombre d'observations varie avec chaque individu. Comme pour un panel cylindré, on introduit une pondération optimale pour la transformation *between*, mais elle varie avec chaque individu :

$$\theta_i^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2}, \quad i = 1, \dots, N$$

L'ensemble du modèle s'obtient en empilant les observations les unes au dessus des autres et s'écrit donc :

$$\underset{(\sum T_i, 1)}{y} = \underset{(\sum T_i, p)}{X} \underset{(p, 1)}{b} + \underset{(\sum T_i, 1)}{u}, \quad V(u|X) = \underset{(\sum T_i, \sum T_i)}{\Omega}$$

avec :

$$\Omega = \text{diag}(\Omega_i)$$

L'estimateur optimal est donc donné par :

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

et admet pour matrice de covariance :

$$V(b^* | X) = (X' \Omega^{-1} X)^{-1}.$$

L'inverse de la matrice de covariance des perturbations s'obtient facilement :

$$\begin{aligned} \Omega^{-1} &= \text{diag}(\Omega_i)^{-1} \\ &= \text{diag}(\Omega_i^{-1}) \\ &= \frac{1}{\sigma_\varepsilon^2} (\text{diag}(W_{T_i}) + \text{diag}(\theta_i^2 B_{T_i})) \\ &= \frac{1}{\sigma_\varepsilon^2} (W + \text{diag}(\theta_i^2 I_{T_i}) B). \end{aligned}$$

Pour obtenir la transformation à appliquer aux données, on prend la racine carrée de la matrice précédente, ce qui donne :

$$\Omega^{-\frac{1}{2}} = \frac{1}{\sigma_\varepsilon} (W + \text{diag}(\theta_i I_{T_i}) B).$$

Dans la pratique, on utilise les transformations suivantes des données du modèle :

$$y^* = (W + \text{diag}(\theta_i I_{T_i}) B) y,$$

et

$$X^* = (W + \text{diag}(\theta_i I_{T_i}) B) X,$$

ce qui permet d'écrire l'estimateur des moindres carrés généralisés comme un estimateur des moindres carrés ordinaires sur données transformées :⁵

$$b^* = (X^{*'} X^*)^{-1} X^{*'} y^*,$$

et la covariance se réécrit :

$$V(b^* | X) = \sigma_\varepsilon^2 (X^{*'} X^*)^{-1}.$$

10.5 Estimation par les moindres carrés quasi-généralisés

Dans la pratique, on ne connaît pas les valeurs de θ_i ($i = 1, \dots, N$) et de σ_ε^2 , on utilise donc les estimateurs suivants :

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{u}'_W \hat{u}_W}{\sum_{i=1}^N T_i - N - p + 1},$$

et

$$\hat{\sigma}_\alpha^2 = \max\left(0, \frac{\hat{u}'_B \hat{u}_B - \hat{\sigma}_\varepsilon^2 (N - p)}{\sum_{i=1}^N T_i - d_B}\right),$$

où les \hat{u}_B sont obtenus dans la régression *sans* duplication. Ces deux estimateurs sont sans biais. Par contre, comme dans le cas cylindré, les estimateurs de $\hat{\theta}_i$ sont biaisés mais convergents :

$$\hat{\theta}_i = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_\alpha^2}}, \quad i = 1, \dots, N$$

ceci permet d'effectuer l'estimation par les moindres carrés quasi-généralisés avec les données transformées :

$$\hat{y}^* = (W + \text{diag}(\hat{\theta}_i I_{T_i}) B) y,$$

et

$$\hat{X}^* = (W + \text{diag}(\hat{\theta}_i I_{T_i}) B) X,$$

ce qui donne les estimateurs :

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

et

$$\hat{V}(b^* | X) = \hat{\sigma}_\varepsilon^2 (\hat{X}^{*'} \hat{X}^*)^{-1}.$$

5. Le terme en σ_ε se simplifie dans la formule de b^* . Par contre, il ne se simplifie pas dans la formule de la matrice de covariance, comme indiqué ci-dessous.

10.6 Tests de spécification

Le test de Hausman peut toujours être appliqué tel quel car le calcul de la covariance entre les estimateurs *within* et des MCQG n'est pas affecté. De même, on peut toujours effectuer le test de Mundlak car l'estimateur *between* reste sans corrélation avec l'estimateur *within*. Pour bien voir ce dernier cas, il suffit de montrer que $WB = 0$ et l'on a :

$$\begin{aligned} WB &= \text{diag}(W_{T_i}) \text{diag}(B_{T_i}) \\ &= \text{diag}(W_{T_i} B_{T_i}) \\ &= \text{diag}(0_{(T_i, T_i)}) \\ &= 0. \end{aligned}$$

on peut donc faire les tests comme précédemment. La seule question qui demeure est celle de l'égalité éventuelle entre les statistiques de Hausman et de Mundlak. Pour y parvenir, il faudrait d'abord décomposer l'estimation des moindres carrés généralisés comme une moyenne pondérée des estimateurs *between* et *within*, mais on a :

$$\begin{aligned} X' \Omega^{-1} y &= \frac{1}{\sigma_\varepsilon^2} X' (W + \text{diag}(\theta_i^2 I_{T_i}) B) y \\ &= \frac{1}{\sigma_\varepsilon^2} (X' W y + X' \text{diag}(\theta_i^2 I_{T_i}) B y), \end{aligned}$$

et il n'est pas possible de sortir le terme en $\text{diag}(\theta_i^2 I_{T_i})$, ce qui empêche de faire apparaître l'estimateur *between* séparément. Il est donc possible que les deux statistiques soient différentes dans le cas non cylindré.

10.7 Programmation

Nous considérons maintenant un échantillon, toujours sur la période 1981-1989, où l'on garde toutes les entreprises qui ont au moins 2 ans de présence, au lieu d'imposer neuf ans de présence. La taille de l'échantillon augmente de $NT = 6039$ observations à $\sum_{i=1}^N T_i = 9964$ observations. Ceci représente un gain de 3995 observations soit 65% de la taille initiale de l'échantillon. La programmation sur panel non cylindré doit prendre un compte un certain nombre d'éléments :

- On ne peut estimer la variance de l'estimateur *between* qu'après avoir estimé les composantes de la variance ;
- Il faut prendre en compte la distribution des T_i lorsque l'on estime les termes de variance ;

- Il faut estimer le terme d_{BX} pour obtenir un estimateur sans biais de σ_α^2 ;
- Il est plus pratique de travailler sur des données dupliquées pour les estimateurs *between* et des MCQG, parce que le nombre de duplications varie avec chaque individu, ce qui empêche de recourir au produit de Kronecker de manière habituelle.

Commençons par examiner la distribution des nombres d'observations. Pour cela, on entre les commandes :

Programme 10.1.

```
data tab; set rep.unb_hmuspan;
proc sort; by firm;
proc freq noprint data=tab;
tables firm/out=tab2;
run;
proc freq data=tab2;
tables count;
run;
goptions htext=1.7;
pattern1 v=L2 c=black;
proc gchart data=tab2;
hbar count/discrete;
run;
```

Le résultat est donné par le graphique 10.1 qui représente le nombre d'entreprises présentes pour chaque valeur de T .

La colonne **FREQ** donne le nombre d'entreprises pour chaque modalité, **CUM FREQ** donne le nombre cumulé d'entreprises, **PCT** le pourcentage de chaque modalité et **CUM PCT** le pourcentage cumulé correspondant. La taille du fichier cylindré est donc de $N = 671$ entreprises présentes $T = 9$ ans et correspond au dernier bâton du graphique. Dans le fichier non cylindré on dispose de $N = 1463$ entreprises pour $2 \leq T_i \leq 9$ (dernière ligne de **CUM FREQ**). Le nombre d'observations ne correspond toutefois pas au nombre d'entreprises car il faut pondérer le nombre d'entreprises par le nombre d'années de présence. Pour l'échantillon cylindré on a donc $NT = 671 \times 9 = 6039$ observations, et pour l'échantillon non cylindré :

$$\sum_{i=1}^N T_i = 2 \times 91 + 3 \times 128 + \dots + 9 \times 671 = 9964 \text{ entreprises.}$$

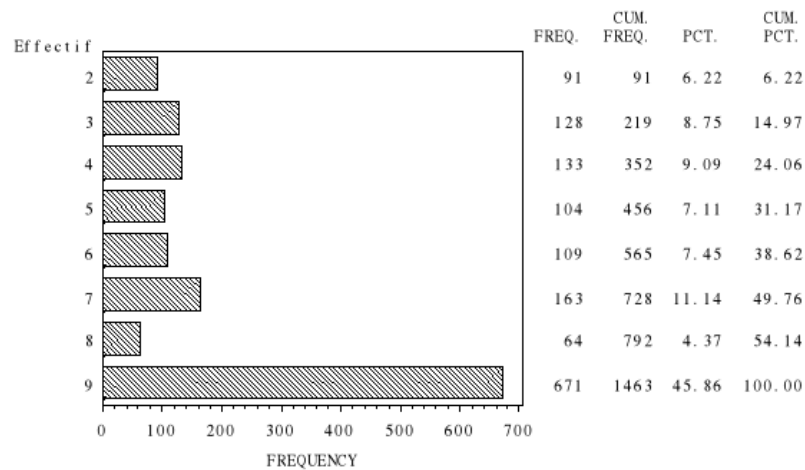


FIGURE 10.1 – Nombre d'entreprises selon la durée de présence

Le programme suivant réalise les estimations *between*, *within*, des composantes de la variance et des composantes du terme d'erreur. Il effectue également le test de Mundlak. La différence principale avec les programmes précédents porte sur le traitement de la duplication, variable pour chaque individu, des observations. Nous ne commenterons que les parties nouvelles. Dans un premier temps, on calcule les transformations *between* et *within* des variables, puis les nombres de présences que l'on stocke dans un nouveau tableau de nom *Ti*. La variable de ce tableau qui contient le nombre de présence pour chaque entreprises s'appelle *COUNT* par défaut :

Programme 10.2.

```
proc means noprint data=tab; by firm;
var LogS LogL LogC LogK;
output out=between mean=;
proc standard mean=0 data=tab out=within; by firm;
var LogS LogL LogC LogK;
proc freq noprint data=tab;
tables firm/out=Ti;
run;
```

Le tableau T_i se présente de la manière suivante et donne le nombre de réplifications qu'il faudra effectuer, entreprise par entreprise :⁶

Sortie 10.1.

firm	COUNT	PERCENT
1	3	0.030108
2	9	0.090325
3	3	0.030108
4	7	0.070253
5	9	0.090325
6	4	0.040145
7	7	0.070253
8	2	0.020072
9	6	0.060217
10	4	0.040145
11	9	0.090325
12	5	0.050181
13	9	0.090325
14	2	0.020072
15	4	0.040145
16	7	0.070253
17	9	0.090325
18	5	0.050181
19	5	0.050181
20	9	0.090325
21	9	0.090325
22	9	0.090325
23	4	0.040145
24	9	0.090325
25	4	0.040145
26	9	0.090325
27	8	0.080289
28	4	0.040145
29	3	0.030108
30	4	0.040145

Voici le programme d'estimation. On commente les parties nouvelles.

Programme 10.3.

6. Le tableau complet comporte 1463 lignes.

```
1. proc means noprint data=tab; by firm;
   var LogS LogL LogC LogK year;
   output out=between mean=;
```

La procédure *means* calcule les moyennes individuelles sur l'ensemble des variables de la régression. Les statistiques sont stockées dans le tableau *between* sous les mêmes noms que dans le tableau originel *tab*

```
2. proc standard mean=0 data=tab out=within; by firm;
   var LogS LogL LogC LogK year;
```

La procédure *standard* centre les variables par rapport à leurs moyennes individuelles (i.e., le centrage est fait individu par individu). Les variables sont stockées dans le tableau *within* sous les mêmes noms que dans le tableau originel *tab*.

```
3. proc freq noprint data=tab;
   tables firm/out=Ti;
```

La procédure *freq* compte le nombre d'occurrence de chaque entreprises dans la base de données, il s'agit donc d'un tableau contenant les T_i . Le tableau en sortie s'appelle *Ti* et contient une variable nommée *COUNT* qui contient le nombre d'observations de chaque entreprise.

```
4. proc iml;
   ny=LogS;
   nx=LogL LogC LogK;
   nt=year;
   nxw=nx||nt;
   p=ncol(nxw)+1;
```

On stocke les noms des variables dans des vecteurs de caractères *ny*, *nx* et *nt*. La tendance *nt* est stockée à part car elle ne varie que dans le temps, pas entre individus. On ne pourra donc pas l'utiliser dans la régression *between*, seulement dans les régressions *within* et *MCQG*. La matrice *nwt* comprend les variables explicatives et la tendance, elle sera utilisée pour la régression *within*. Enfin, la variable *p* contient le nombre de variables avec la tendance et le terme constant (qui justifie le +1).

```
5. use BETWEEN; read all var (ny) into By;
   read all var (nx) into Bx;
   read all var (nxw) into Bx2;
   close BETWEEN;
   N=nrow(By);
   Bx=J(N,1,1)||Bx; Bx2=J(N,1,1)||Bx2; nxc="Constante"||nx;
   nxc2=nxc||nt;
```

Lecture du tableau *between*. N contient le nombre d'individus, la matrice Bx est celle qui sera utilisée pour la régression *between* et la matrice $Bx2$ sera utilisée pour la régression par les *MCQG*. Ces deux matrices contiennent un terme constant.

6. `use WITHIN;`
`read all var (ny) into Wy;`
`read all var (nxw) into Wx;`
`close WITHIN;`
`Sti=nrow(Wy);`
 Lecture du tableau *within*. Le nombre Sti contient le nombre total d'observations $\sum_{i=1}^N T_i$.
7. `use Ti;`
`read all var COUNT into T;`
`close Ti;`
 Lecture des nombres d'observations de chaque individu, rangés dans le vecteur T .
8. `print "Estimation sur panel non cylindré",`
`"Nombre d'individus =" n,`
`"Nombre d'observations =" sti,;`
 Impression des nombres d'individus et d'observations. La virgule représente un saut de ligne.
9. `do i=1 to n;`
`Ti=T[i];`
`byd=byd//J(Ti,1,by[i]);`
`bxd=bxd//(J(Ti,1,1)@bx[i,]);`
`tibxd=tibxd//(J(Ti,1,Ti)@bx[i,]);`
`bx2=bx2//(J(Ti,1,1)@bx2[i,]);`
`end;`
 Cette boucle calcule les matrices *between* avec duplication des observations. La notation $@$ désigne le produit de Kronecker (\otimes). Ainsi l'instruction $J(Ti,1,by[i])$ calcule le vecteur :

$$\begin{pmatrix} y_{i\bullet} \\ \vdots \\ y_{i\bullet} \end{pmatrix}_{(Ti,1)}$$

et $J(T_i, 1, 1) @ bx[i,]$ donne :

$$\begin{pmatrix} X_{i\bullet} \\ \vdots \\ X_{i\bullet} \end{pmatrix}_{(T_i, p)}$$

et $J(T_i, 1, T_i) @ bx[i,]$ donne :

$$\begin{pmatrix} T_i X_{i\bullet} \\ \vdots \\ T_i X_{i\bullet} \end{pmatrix}_{(T_i, p)}$$

```
10. start MCO(y, x) global(ixx);
    ixx=inv(t(x)*x);
    b=ixx*t(x)*y;
    return (b);
    finish MCO;
```

Fonction calculant l'estimateur des moindres carrés ordinaires à partir d'une variables expliquée stockée dans un vecteur y et de variables explicatives stockées dans une matrice x . L'estimateur est renvoyé vers l'appel de la fonction, et la matrice $(x'x)^{-1}$ est rendue disponible après son exécution (via l'option *global*).

```
11. bbet=MCO(Byd, Bxd);
```

Calcul de l'estimation *between* avec duplication, car les données ne sont pas cylindrées.

```
12. ubet=Byd-Bxd*bbet;
```

Calcul du résidu *between*.

```
13. sub=ssq(ubet);
```

Calcul de la somme des carrés des résidus *between*.

```
14. ixbx=ixx;
```

Récupération de la matrice $(X'BX)^{-1}$.

```
15. db=trace(t(bxd)*tibxd*ixbx);
```

print db;

Calcul de d_{BX} .

```
16. bwit=MCO(Wy, Wx);
```

uwit=Wy-Wx*bwit;

suw=ssq(uwit);

ixwx=ixx;

Estimation *within*, calcul du résidu, de la somme des carrés des résidus et récupération de $(X'WX)^{-1}$.

- ```

17. se2=suw/(sti-n-(p-1));
 sa2=max(0,(sub-se2#(n-p))/(sti-db));
 rho=sa2/(sa2+se2);
 print "Variance de l'effet individuel =" sa2,
 "Variance de l'erreur idiosyncratique =" se2,
 "Autocorrélation =" rho;
 Estimation et impression des composantes de la variance et du coeffi-
 cient d'autocorrélation.
18. vbet=se2#ixbx+sa2#(ixbx*t(bxd)*tibxd*ixbx);
 sbet=sqrt(vecdiag(vbet));
 tbet=abs(bbet)/sbet;
 print "Régression Between","Variable expliquée : "
 ny, bbet[rowname=nxc] sbet tbet;
 Impression des résultats de la régression between.
19. vwit=se2#ixwx;
 swit=sqrt(vecdiag(vwit));
 twit=abs(bwit)/swit;
 print "Régression Within","Variable expliquée : " ny,
 bwit[rowname=nxw] swit twit;
 Impression des résultats de la régression within.
20. rgb=2:(p-1);
 Liste des indices des variables différentes de la constante et de la ten-
 dence dans l'estimateur between. En effet, le test ne peut pas porter sur
 le terme constant car la constante disparaît lors de la transformation wi-
 thin et la tendance est colinéaire au terme constant après la transforma-
 tion between.
21. rgw=1:(p-2);
 Liste des indices des variables différentes de la tendance dans l'estima-
 teur within. Il n'est pas nécessaire d'éliminer la constante car c'est déjà
 fait.
22. m=t(bbet[rgb]-bwit[rgw])*inv(vbet[rgb,rgb]+
 vwit[rgw,rgw])*(bbet[rgb]-bwit[rgw]);
 pm=1-probchi(m,p-2);
 print "Test de Mundlak", "Statistique =" m,
 "Probabilité critique =" pm;
 Calcul de la statistique de Mundlak, de sa probabilité critique et impres-
 sion.
23. theta=sqrt(se2/(se2+t#sa2));
 do i=1 to n;

```

```
thetad=thetad//J(T[i],1,theta[i]);
end;
```

Calcul du vecteur des pondération. La quantité  $J(T[i],1,theta[i])$  donne :

$$\begin{pmatrix} \hat{\theta}_i \\ \vdots \\ \hat{\theta}_i \end{pmatrix}_{(T_i,1)}$$

- ```
24. y=byd+wy;
x=bx2+(j(sti,1,0)||wx);
oy=y-(1-thetad)#byd;
ox=x-(1-thetad)#bx2;
bopt=mco(oy,ox);
vopt=SE2#ixx;
sopt=sqrt(vecdiag(vopt));
topt=abs(bopt)/sopt;
print "Régression MCQG","Variable expliquée : " ny,
bopt[rowname=nxc2] sopt topt;
```
- On reconstitue les séries complètes en additionnant les transformations *between* et *within*. Dans le cas de la transformation *within* on ajoute une colonne nulle car il n'y a pas de constante et qu'il faut que les formats des matrices *between* et *within* coïncident. On remarque également que l'on doit utiliser la matrice *between* dupliquée complète $B \times 2$ pour obtenir la liste complète des variables : terme constant, variables explicatives classiques et tendance (constante dans la dimension individuelle). On calcule ensuite l'estimateur des MCQG puis on imprime les résultats.
- ```
25. rgo=2:p;
h=t(bwit-bopt[rgo])*inv(vwit-vopt[rgo,rgo])*(bwit-
bopt[rgo]);
ph=1-probchi(h,p-1);
print "Test de Hausman", "Statistique =" h,
"Probabilité critique =" ph;
```
- Réalisation du test de Hausman. On enlève le terme constant de l'estimation des MCQG.
- ```
26. rg=2:p;
a=by-bx2[,rg]*bwit;
ma=sum(a)/N;
print "Terme constant déduit du Within =" ma;
alpha=a-ma;
```

```
create alpha from alpha[colname="Alpha"];
append from alpha;
```

Estimation des effets individuels sans duplication, $\hat{\alpha}_i$ et sauvegarde dans un tableau SAS de nom alpha.

```
27. ad=byd-bxd2[,rg]*bwit;
mad=sum(ad)/sti;
alphad=ad-mad;
epsi=y-x[,rg]*bwit-ad;
py=y-epsi;
res=y||x||py||epsi||alphad;
nres=ny||nxc2||"prev"||"epsi"||"alpha";
create RES from res[colname=nres];
append from res;
quit; run;
```

Calcul de la prévision, du résidu $\hat{\epsilon}_{it}$ et stockage des résultats dans un tableau SAS de nom res.

L'exécution du programme donne le résultats suivant :

Sortie 10.2.

```
Estimation sur panel non cylindré

                                N
Nombre d'individus =           1463
                                Sti
Nombre d'observations =       9964

                                db
                                30.729393

                                sa2
Variance de l'effet individuel = 0.1703809
                                se2
Variance de l'erreur idiosyncratique = 0.0548732
                                rho
Autocorrélation = 0.7563941

Régression Between
                                ny
Variable expliquée : LOGS
bbet                                sbet                                tbet
```

Constante	3.419719	0.0498262	68.632905
LOGL	0.6642606	0.0193446	34.338232
LOGC	0.3111107	0.0149353	20.830494
LOGK	0.0308521	0.0084247	3.6620973

Régression Within

			ny
	Variable expliquée : LOGS		
bwit		swit	twit
LOGL	0.7795193	0.0090413	86.217824
LOGC	0.1396125	0.0076129	18.338934
LOGK	0.0244897	0.0086721	2.8239839
YEAR	0.0563187	0.0012536	44.926052

Test de Mundlak

		m
	Statistique = 158.54862	
		pm
Probabilité critique =		0

Régression MCQG

			ny
	Variable expliquée : LOGS		
bopt		sopt	topt
Constante	-0.666046	0.0909715	7.3214782
LOGL	0.7954689	0.0076985	103.32793
LOGC	0.1809648	0.0063508	28.495031
LOGK	0.0302756	0.005748	5.2671157
YEAR	0.0531161	0.0011162	47.588536

Test de Hausman

		h
	Statistique = 163.67666	
		ph
Probabilité critique =		0

	malpha
Terme constant déduit du Within =	-0.759096

Globalement, on rejette l'hypothèse d'absence de corrélation entre les variables explicatives et l'effet individuel, ce qui incite à retenir l'estimation *within*, où l'on constate que l'effet de la recherche développement sur la production serait fort, et non faible comme le suggère la régression *between*. On

peut examiner directement cette question en calculant les coefficients de corrélation linéaire entre les variables explicatives et l'effet individuel estimé. On constate que les corrélations sont significatives, mais que la corrélation entre l'effet individuel et les dépenses de recherche et développement est particulièrement importante. On vérifie également que les corrélations avec $\hat{\varepsilon}_{it}$ sont nulles par construction.

Sortie 10.3.

```

                                Procédure CORR
2 Avec les variables :      alpha      epsi
4      Variables :      LOGL      LOGC      LOGK      YEAR

                                Statistiques simples

Variable          N          Moyenne      Ecart-type      Somme
alpha             9964             0          0.46915          0
epsi             9964             0          0.21633          0
LOGL             9964          0.71161          1.86728          7090
LOGC             9964          3.95377          2.33145          39395
LOGK             9964          2.76279          2.19311          27528
YEAR             9964          84.86130          2.52520          845558

                                Statistiques simples

Variable          Minimum      Maximum
alpha             -2.96902      2.28388
epsi             -7.09284      2.27645
LOGL             -6.21461      6.77537
LOGC             -4.34281      11.21666
LOGK             -4.71276      10.10108
YEAR             81.00000      89.00000

                                Coefficients de corrélation de Pearson, N = 9964
                                Proba > |r| sous H0: Rho=0

                                LOGL          LOGC          LOGK          YEAR
alpha      0.36670      0.44226      0.29263      0.00400
           <.0001      <.0001      <.0001      0.6898

epsi       0.00000      0.00000      0.00000      0.00000
           1.0000      1.0000      1.0000      1.0000

```

Le graphique 10.2 représente la distribution empirique des $\hat{\alpha}_i$. On constate qu'elle possède un profil moins heurté que dans l'estimation cylindrée. Ce point peut s'expliquer par le fait que le cylindrage des données tend parfois à éliminer les plus petites entreprises de l'échantillon, ce qui peut accentuer l'asymétrie des distributions empiriques sur l'échantillon. Le graphique 10.3 représente la distribution empirique du résidu de la régression, qui présente une forme symétrique.

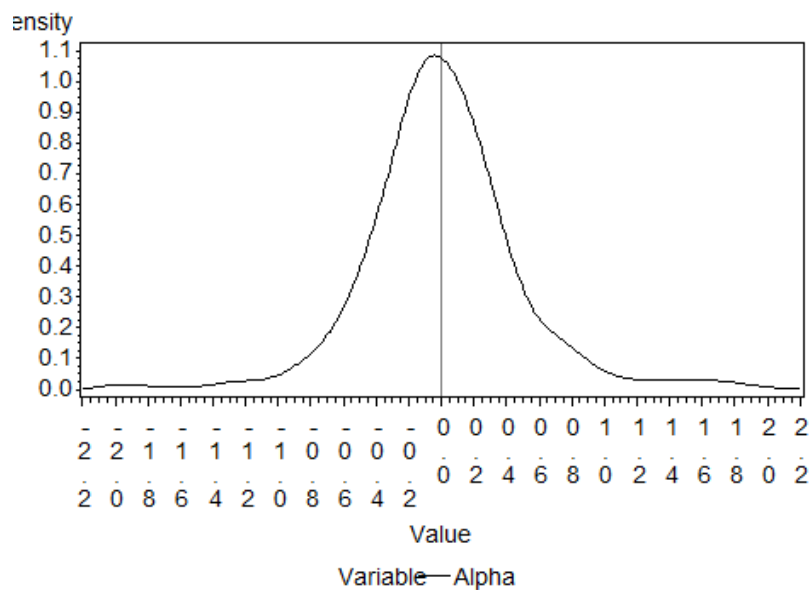
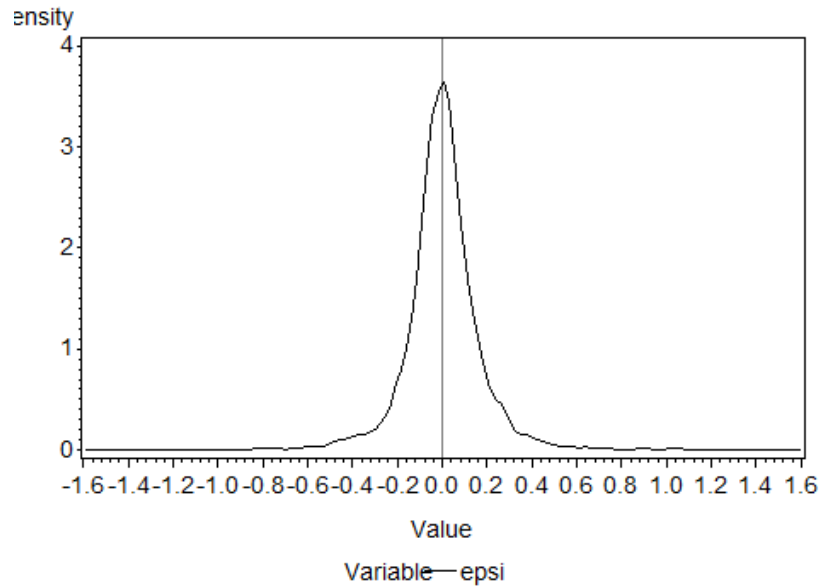


FIGURE 10.2 – Estimation des effets individuels, $\hat{\alpha}_i$

Il s'agit d'une simple extension du cas cylindrée. La seule différence importante tient au fait que la pondération varie avec chaque individu, avec le nombre d'années de présence. Plus précisément, on ajoute les lignes suivantes au programme précédent :

1. `theta=sqrt(se2/(se2+t#sa2));`
Calcul des pondérations $\hat{\theta}_i = (\hat{\sigma}_\varepsilon^2 / (\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_\alpha^2))^{\frac{1}{2}}$. Il ne faut pas oublier que `t` est un vecteur contenant les années de présence.
2. `do i=1 to n;`
3. `thetad=thetad//J(T[i],1,theta[i]);`

FIGURE 10.3 – Résidus idiosyncratiques, $\hat{\epsilon}_{it}$

4. end ;
Cette boucle duplique chaque $\hat{\theta}_i$ autant de fois qu'il y a d'années de présence (T_i fois) et empile ces vecteurs individuels dans un vecteur global de pondération nommé thetad
5. oy=y-(1-thetad)#byd; ox=x-(1-thetad)#bxd;
Transformations optimales de la variable expliquée et des variables explicatives, y compris le terme constant
6. bopt=mco(oy,ox);
On obtient l'estimateur optimal comme des moindres carrés ordinaires appliqués aux transformations optimales.
7. vopt=SE2#ixx;
8. sopt=sqrt(vecdiag(vopt));
9. topt=abs(bopt)/sopt;
10. print "Régression MCQG", "Variable expliquée :" ny,
11. bopt[rowname=nxc] sopt topt;

```

12. /* Test de Hausman */
    Le test de Hausman est appliqué directement aux estimateurs des MCQG
    et within
13. h=t(bwit-bopt[rg])*inv(vwit-vopt[rg,rg])*(bwit-bopt[rg]);
14. ph=1-probchi(h,p-1);
15. print "Test de Hausman", "Statistique =" h,
16. "Probabilité critique =" ph;
17. quit;

```

On obtient la sortie suivante :

Sortie 10.4.

```

                                                    STI
Nombre d'observations =          9964
                                                    N
Nombre d'individus =          1463

                                                    DB
30.729393

                                                    SA2
Variance de l'effet individuel =  0.168462
                                                    SE2
Variance de l'erreur idiosyncratique = 0.0678996

Régression Between
                                                    NY
Variable expliquée : LOGS
BBET          SBET          TBET
Constante    3.419719  0.0497979  68.672008
LOGC         0.3111107  0.0149262  20.843217
LOGL         0.6642606  0.0193341  34.356896
LOGK         0.0308521  0.008419  3.6645706

```


Régression Within

NY

Variable expliquée : LOGS

	BWIT	SWIT	TWIT
LOGC	0.1695643	0.0084359	20.100299
LOGL	0.7558187	0.0100402	75.279174
LOGK	0.2065444	0.0085286	24.217715

Test de Mundlak

M

Statistique = 273.4369

PM

Probabilité critique =

0

Régression MCQG

NY

Variable expliquée : LOGS

	BOPT	SOPT	TOPT
Constante	3.5358792	0.0236505	149.50534
LOGC	0.2165751	0.0068234	31.739873
LOGL	0.7267629	0.0082135	88.483533
LOGK	0.1091717	0.0057382	19.025346

Test de Hausman

H

Statistique = 251.53587

PH

Probabilité critique =

0

On constate que la statistique de Hausman est maintenant différente de celle de Mundlak, mais qu'elles sont du même ordre de grandeur. On rejette l'hypothèse nulle d'absence de corrélation des effets individuels au seuil de 5%. Il faut donc utiliser l'estimation *within*.

Chapitre 11

Le modèle apparemment linéaire

On parle de modèle apparemment linéaire quand la perturbation du modèle est corrélée avec au moins une variable explicative du modèle. Un premier type de modèle apparemment linéaire peut émerger dans un modèle statique : quand une variable est mesurée avec erreur ou quand une des variables explicatives est corrélée avec le terme d'erreur (modèle à effets fixes). Un second type de modèle apparemment linéaire émerge quasi-systématiquement dans les modèles dynamiques de panel. Même sous des hypothèses restrictives portant sur l'erreur du modèle, les variables explicatives sont corrélées à la perturbation du modèle. La conséquence principale d'une corrélation entre les variables explicatives et le terme d'erreur est simple : les estimateurs que nous avons vu jusqu'à présent ne sont plus convergents. Pour régler ce problème, il nous faudra recourir à la méthode des variables instrumentales puis à son extension, la méthode des moments généralisés.

11.1 Le modèle avec erreurs de mesure

Les données sur lesquelles on travaille ne sont pas forcément mesurées de manière parfaite. Certaines variables explicatives sont mesurées avec erreur. L'objectif de cette section est de montrer que, même en l'absence de corrélation entre l'erreur de mesure et le terme d'erreur du modèle, l'estimation par les moindres carrés ordinaires n'est plus convergente. Ce résultat est valable même dans un modèle statique. Pour fixer les idées, nous traiterons le cas de la fonction de production. Dans toute cette section, on raisonne sur des variables centrées.

Soit une fonction de production à rendements d'échelle constants, en loga-

rithmes et sans effet individuel, donnée par :¹

$$y_{it} = \alpha c_{it}^* + \varepsilon_{it}, \quad (11.1)$$

avec $y_{it} = \ln(Y_{it}/L_{it})$, le logarithme de la productivité apparente du travail, $c_{it}^* = \ln(C_{it}^*/L_{it})$, le logarithme de l'intensité capitalistique, α l'élasticité de la production au capital (que l'on cherche à estimer) et ε_{it} un bruit blanc d'espérance nulle et de variance constante σ_ε^2 . La variable c_{it}^* est calculée à partir de la vraie valeur du capital C_{it}^* qui est inobservable.

La raison pour laquelle on n'observe pas la véritable quantité de capital est que les équipements de production ne sont pas homogènes (on ne peut pas additionner les équipements productifs des entreprises) de sorte qu'il faut procéder en divisant la valeur des équipements productifs par un indice de prix du capital. Toute erreur sur l'indice de prix peut alors se répercuter sur la mesure de quantité de capital que l'on utilise. Notons bien que ce problème ne se pose pas pour le nombre d'heures travaillées, qui est bien une mesure de quantité.² Plus précisément, on suppose que le capital observé est relié à la vraie valeur du capital par la relation :

$$c_{it} = c_{it}^* + \zeta_{it},$$

où ζ_{it} est une erreur de mesure dont les propriétés sont les suivantes :

1. $E(\zeta_{it}) = 0$, il n'y a pas d'erreur systématique dans la mesure du capital, on utilise un estimateur sans biais ; on pose également $V(\zeta_{it}) = \sigma_\zeta^2$, de sorte que la précision de la mesure ne dépend pas de l'individu ;
2. Les ζ_{it} sont indépendamment et identiquement distribués, ce qui revient à dire que les erreurs de mesure ne sont pas corrélées entre elles ;
3. Les ζ_{it} sont indépendants des c_{it}^* , de sorte que l'erreur ne dépend pas de la valeur de la variable que l'on mesure ;
4. Les ζ_{it} sont indépendants de la perturbation du modèle que l'on estime ε_{it} ; ceci revient à dire que les variables que l'on se procure sont collectées indépendamment du modèle que l'on estime.

Ces hypothèses impliquent :

$$E(c_{it}) = c_{it}^*, \quad V(c_{it}) = \sigma_{c^*}^2 + \sigma_\zeta^2,$$

1. On a :

$$Y_{it} = A_{it} C_{it}^{*\alpha} L_{it}^{1-\alpha} \Leftrightarrow \frac{Y_{it}}{L_{it}} = A_{it} \left(\frac{C_{it}^*}{L_{it}} \right)^\alpha,$$

avec $A_{it} = \exp(\varepsilon_{it})$.

2. On peut également décomposer ce nombre d'heures en distinguant le travail qualifié du travail non qualifié.

où :

$$\sigma_{c^*}^2 = V(c_{it}^*) = \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^{*2}.$$

Malgré toutes les hypothèses d'indépendance sur ζ_{it} , l'estimateur des MCO n'est plus convergent. En effet, considérons le seul estimateur que peut utiliser l'économètre. Comme seul c_{it} est observable, l'économètre régresse y_{it} sur c_{it} , ce qui donne l'estimateur :

$$\hat{\alpha} = \left(\sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T c_{it} y_{it},$$

pour obtenir sa limite en probabilité, on remplace $y_{it} = \alpha c_{it}^* + \varepsilon_{it}$ par sa valeur, ce qui donne :

$$\begin{aligned} \hat{\alpha} &= \left(\sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \left(\alpha \sum_{i=1}^N \sum_{t=1}^T c_{it} c_{it}^* + \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} \right) \\ &= \left(\sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \left(\alpha \sum_{i=1}^N \sum_{t=1}^T c_{it} (c_{it} - \zeta_{it}) + \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} \right) \\ &= \left(\sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \left[\alpha \sum_{i=1}^N \sum_{t=1}^T c_{it}^2 - \alpha \sum_{i=1}^N \sum_{t=1}^T c_{it} \zeta_{it} + \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} \right] \\ &= \alpha + \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \left[-\alpha \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \zeta_{it} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} \right]. \end{aligned}$$

Examinons les limites en probabilité des différents termes de l'expression précédente :

$$\begin{aligned}
\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^2 &= \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (c_{it}^* + \zeta_{it})^2 \\
&= \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^{*2} + \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \zeta_{it}^2 \\
&\quad + \underbrace{2 \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^* \zeta_{it}}_{=0 \text{ (indépendance)}} \\
&= \sigma_{c^*}^2 + \sigma_{\zeta}^2.
\end{aligned}$$

Pour le numérateur, on obtient :

$$\begin{aligned}
\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} &= \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (c_{it}^* + \zeta_{it}) \varepsilon_{it} \\
&= \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^* \varepsilon_{it} + \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \zeta_{it} \varepsilon_{it} \\
&= 0.
\end{aligned}$$

car l'erreur de mesure est indépendante de la vraie valeur du capital et de l'erreur du modèle. Il nous reste le terme qui pose problème :

$$\begin{aligned}
\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \zeta_{it} &= \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (c_{it}^* + \zeta_{it}) \zeta_{it} \\
&= \underbrace{\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^* \zeta_{it}}_{=0 \text{ (indépendance)}} + \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \zeta_{it}^2 \\
&= \sigma_{\zeta}^2.
\end{aligned}$$

Globalement, on obtient donc :

$$\text{Plim}\hat{\alpha} = \alpha - \alpha \times \frac{\sigma_{\zeta}^2}{\sigma_{c^*}^2 + \sigma_{\zeta}^2},$$

le ratio ϕ :

$$\phi = \frac{\sigma_{\zeta}^2}{\sigma_{c^*}^2 + \sigma_{\zeta}^2} \in [0, 1],$$

représente l'importance du bruit dans la variance de la mesure du capital.³ Plus ce ratio est proche de l'unité, plus la série est mesurée avec erreur. Inversement, plus il est proche de zéro, plus la mesure du capital est bonne. On a donc :

$$\text{Plim}\hat{\alpha} = \alpha(1 - \phi).$$

Le résultat principal est que l'estimateur des moindres carrés ordinaires est "biaisé vers zéro". Cette expression signifie que la valeur absolue de $\hat{\alpha}$ est plus proche de 0 que la vraie valeur. Pour le voir, considérons les cas $\alpha < 0$ et $\alpha > 0$:

1. $\alpha < 0 \Rightarrow \text{Plim}\hat{\alpha} = \alpha(1 - \phi) > \alpha$, donc la valeur estimée se rapproche de 0 par rapport à la vraie valeur ;
2. $\alpha > 0 \Rightarrow \text{Plim}\hat{\alpha} = \alpha(1 - \phi) < \alpha$, d'où le même résultat.

Dans le cas de la fonction de production, ceci signifie que l'on tendra à sous-estimer la valeur de l'élasticité du capital. Sur données françaises, les études microéconométriques tendent à trouver $\alpha \simeq 0,2$ alors que la comptabilité nationale se rapproche de $\alpha \simeq 0,3$. Les erreurs de mesure sur la quantité de capital sont parfois évoquées pour expliquer ce résultat.⁴ Plus généralement, on note qu'avec une variable explicative :

$$\text{Plim} \frac{X'\varepsilon}{NT} < 0,$$

de sorte que le biais va toujours dans le même sens.⁵

3. En anglais : "noise-to-signal ratio".

4. Il existe d'autres sources de biais possibles.

5. Un autre exemple souvent cité est celui de la demande de travail des entreprises. Dans ce cas la demande de travail est une fonction décroissante du salaire réel. On obtient donc le résultat que le coefficient estimé peut apparaître comme moins négatif que le vrai.

11.2 Le modèle avec simultanéité

On considère toujours la fonction de production de la section précédente, mais cette fois-ci le capital est mesuré sans erreur :

$$y_{it} = \alpha c_{it} + \varepsilon_{it},$$

Toutefois, suite à une contrainte de débouchés, l'investissement est déterminé par un modèle de type accélérateur.⁶ On peut donc écrire :

$$c_{it} = \gamma y_{it} + \xi_{it}, \gamma > 0$$

Pour bien montrer d'où vient le problème, on fait les hypothèses les plus favorables possibles sur les perturbations :

$$\begin{aligned} E(\varepsilon_{it}) &= 0, E(\xi_{it}) = 0, V(\varepsilon_{it}) = \sigma_\varepsilon^2, V(\xi_{it}) = \sigma_\xi^2, \\ \varepsilon \text{ et } \xi &\text{ indépendants entre eux.} \end{aligned}$$

L'estimateur des moindres carrés ordinaires est toujours défini par :

$$\hat{\alpha} = \left(\sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T c_{it} y_{it},$$

mais cette fois-ci, en résolvant, on obtient :

$$y_{it} = \frac{\alpha \xi_{it} + \varepsilon_{it}}{1 - \alpha \gamma} \text{ et } c_{it} = \frac{\xi_{it} + \gamma \varepsilon_{it}}{1 - \alpha \gamma}, \quad (11.2)$$

donc le capital par tête est corrélé avec le terme d'erreur :

$$\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} = \frac{\gamma \sigma_\varepsilon^2}{1 - \alpha \gamma} > 0 \text{ si } \alpha \gamma < 1.$$

et le biais asymptotique est donné par :

$$\begin{aligned} \text{Plim} \hat{\alpha} - \alpha &= \left(\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it}^2 \right)^{-1} \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T c_{it} \varepsilon_{it} \\ &= \left(\frac{\sigma_\xi^2 + \gamma^2 \sigma_\varepsilon^2}{(1 - \alpha \gamma)^2} \right)^{-1} \frac{\gamma \sigma_\varepsilon^2}{1 - \alpha \gamma} \\ &= (1 - \alpha \gamma) \frac{\gamma \sigma_\varepsilon^2}{\sigma_\xi^2 + \gamma^2 \sigma_\varepsilon^2} > 0 \end{aligned}$$

6. Le capital est proportionnel à la production, donc l'investissement est proportionnel à la variation de la production.

on tendra maintenant à surestimer l'élasticité du capital, cas peu réaliste en pratique sur données individuelles. Toutefois, le message général reste valable : l'estimateur des moindres carrés ordinaires n'est plus convergent en cas de simultanéité.

11.3 Le modèle autorégressif

La définition même du modèle autorégressif implique un problème de corrélation entre le terme d'erreur et la variable expliquée retardée du modèle. Ce résultat s'obtient même dans le modèle à erreur composée, où il n'existe pourtant aucune corrélation entre la perturbation du modèle et les variables explicatives. Dans un modèle autorégressif, tous les estimateurs que nous avons vu (MCO, *Between*, *Within*, MCQG) ne sont plus convergents, ce qui implique de changer de méthode d'estimation.

Trois cas principaux impliquent la non convergence des estimateurs usuels d'un modèle auto-régressif :

1. Quand il existe un effet fixe corrélé avec les variables explicatives, la transformation *within* ne règle plus le problème d'estimation ;
2. Quand il existe un effet individuel *non* corrélé aux variables explicatives ;
3. Quand il n'y a pas d'effet individuel dans le modèle et que la perturbation du modèle est autocorrélée.

11.3.1 Modèle avec effet fixe

Considérons le modèle dynamique suivant :

$$y_{it} = \rho_1 y_{it-1} + \tilde{X}_{it}c + \mu_i + \varepsilon_{it}, \quad (11.3)$$

où μ_i est un effet fixe corrélé avec les variables explicatives \tilde{X}_{it} et ε_{it} un bruit blanc. Pour estimer ce modèle, on pense d'abord à éliminer l'effet individuel μ_i , sinon l'estimation ne peut pas être convergente. Toutefois, ici, cela ne règle pas le problème d'estimation mais en pose un nouveau. En écrivant le modèle dans la dimension intra-individuelle, on obtient :

$$y_{it} - y_{i\bullet} = \rho_1 (y_{it-1} - y_{i\bullet}) + (\tilde{X}_{it} - \tilde{X}_{i\bullet})c + \varepsilon_{it} - \varepsilon_{i\bullet},$$

l'effet individuel a bien été éliminé mais la variable explicative $y_{it} - y_{i\bullet}$ vérifie d'après (11.3) :

$$y_{it-1} - y_{i\bullet} = \rho_1 (y_{it-2} - y_{i\bullet}) + (\tilde{X}_{it-1} - \tilde{X}_{i\bullet})c + \varepsilon_{it-1} - \varepsilon_{i\bullet},$$

ce qui implique que $y_{it-1} - y_{i\bullet}$ est corrélé avec $\varepsilon_{it-1} - \varepsilon_{i\bullet}$, c'est-à-dire avec tous les ε_{it} ($t = 1, \dots, T$). Une application des moindres carrés ordinaires à la relation (11.4) ne fournira donc pas d'estimateur convergent des paramètres. Notons que même des transformations plus simples ne règlent pas le problème. Si, par exemple, on écrit le modèle en différences, on obtient :

$$y_{it} - y_{it-1} = \rho_1 (y_{it-1} - y_{it-2}) + (\tilde{X}_{it} - \tilde{X}_{it-1})c + \varepsilon_{it} - \varepsilon_{it-1},$$

or ceci implique que :

$$y_{it-1} - y_{it-2} = \rho_1 (y_{it-2} - y_{it-3}) + (\tilde{X}_{it-1} - \tilde{X}_{it-2})c + \varepsilon_{it-1} - \varepsilon_{it-2},$$

qui fait intervenir un terme en ε_{it-1} . Quelque soit la différence effectuée, il reste toujours une corrélation entre la variable explicative retardée et la perturbation du modèle.

11.3.2 Modèle à erreur composée

Considérons le modèle à erreur composée suivant

$$y_{it} = \rho_1 y_{it-1} + \tilde{X}_{it}c + \alpha_i + \varepsilon_{it}, \quad (11.4)$$

où $(\alpha_i, \varepsilon_{it})$ sont des aléas indépendants des variables explicatives \tilde{X}_{it} du modèle. En décalant cette relation d'une période, on obtient :

$$y_{it-1} = \rho_1 y_{it-2} + \tilde{X}_{it-1}c + \alpha_i + \varepsilon_{it-1},$$

ce qui montre la corrélation entre y_{it-1} et α_i . L'estimation ne peut donc être convergente. On pourrait alors penser régler le problème en éliminant l'effet individuel dans la relation (11.4) :

$$y_{it} - y_{it-1} = \rho_1 (y_{it-1} - y_{it-2}) + (\tilde{X}_{it} - \tilde{X}_{it-1})c + \varepsilon_{it} - \varepsilon_{it-1}, \quad (11.5)$$

mais si l'on décale cette relation d'une période, on obtient :

$$y_{it-1} - y_{it-2} = \rho_1 (y_{it-2} - y_{it-3}) + (\tilde{X}_{it-1} - \tilde{X}_{it-2})c + \varepsilon_{it-1} - \varepsilon_{it-2},$$

ce qui montre la corrélation entre $y_{it-1} - y_{it-2}$ et ε_{it-1} , terme qui intervient dans la perturbation de la relation en différences (11.5). On n'obtient donc pas d'estimateur convergent par cette méthode.

11.3.3 Modèle à perturbations autocorrélées

Ce dernier cas n'est pas spécifique aux données de panel. Le problème se pose dans les mêmes termes que pour une série temporelle. On considère un modèle sans effet individuel du tout :

$$y_{it} = \rho_1 y_{it-1} + \tilde{X}_{it}c + \varepsilon_{it}, \quad (11.6)$$

dont la perturbation est autocorrélée :

$$\varepsilon_{it-1} = r \varepsilon_{it-1} + \eta_{it}, \quad (11.7)$$

où η_{it} est un bruit blanc. En décalant la relation (11.6) d'une période, on obtient :

$$y_{it-1} = \rho_1 y_{it-2} + \tilde{X}_{it-1}c + \varepsilon_{it-1},$$

donc y_{it-1} est corrélé avec ε_{it-1} , qui se trouve lui-même corrélé à la perturbation ε_{it} via la relation (11.7). L'estimation de ce modèle par les estimateurs usuels n'est donc pas convergente.

Chapitre 12

Estimation par variables instrumentales

12.1 Le modèle

Soit le modèle linéaire suivant :

$$y_{(NT,1)} = X_{(NT,p)} b_{(p,1)} + u_{(NT,1)} \quad (12.1)$$

avec $E(u|X) = 0$ et $V(u|X)$ finie et inversible. On dit que ce modèle est apparemment linéaire si :

$$\text{Plim} \frac{X'u}{NT} \neq 0.$$

Cette hypothèse implique que l'estimateur des moindres carrés ordinaires \hat{b} n'est plus convergent. En effet :

$$\begin{aligned} \hat{b} &= (X'X)^{-1} X'y \\ &= b + (X'X)^{-1} X'u \\ &= b + \left(\frac{X'X}{NT} \right)^{-1} \frac{X'u}{NT}, \end{aligned}$$

en prenant les limites en probabilité de chaque membre de l'équation, on ob-

tient :

$$\begin{aligned} \text{Plim} \hat{b} &= b + \text{Plim} \left(\frac{X'X}{NT} \right)^{-1} \frac{X'u}{NT} \\ &= b + \underbrace{\left(\text{Plim} \frac{X'X}{NT} \right)^{-1}}_{\text{finie par hypothèse}} \underbrace{\text{Plim} \frac{X'u}{NT}}_{\neq 0} \\ &\neq b. \end{aligned}$$

Nous avons vu trois exemples de modèles autorégressifs apparemment linéaires. Pour simplifier leur exposé, nous posons les notations suivantes :

$$X_{it} = (y_{it-1}, \tilde{X}_{it}) \text{ et } b = \begin{pmatrix} \rho_1 \\ c \end{pmatrix},$$

les trois modèles qui suivent peuvent se mettre sous la forme :

$$y_{it} = X_{it}b + u_{it},$$

en adaptant la définition de u_{it} à chaque cas particulier.

1. Modèle à effets fixes (corrélacion de \tilde{X}_{it} avec l'effet fixe) :

$$u_{it} = \mu_i + \varepsilon_{it},$$

avec :

$$\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{it} \mu_i \neq 0 \Rightarrow \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} u_{it} \neq 0.$$

2. Modèle à erreur composée (corrélacion de y_{it-1} avec l'effet individuel) :

$$u_{it} = \alpha_i + \varepsilon_{it},$$

avec :

$$\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it-1} \alpha_i \neq 0 \Rightarrow \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} u_{it} \neq 0.$$

3. Modèle à erreur autocorrélée (corrélacion de y_{it-1} avec une erreur simple) :

$$u_{it} = \varepsilon_{it},$$

$$\varepsilon_{it} = r \varepsilon_{it-1} + \eta_{it},$$

avec :

$$\text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it-1} \varepsilon_{it} \neq 0 \Rightarrow \text{Plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} u_{it} \neq 0.$$

12.2 Les variables instrumentales

Pour bien comprendre la méthode à variables instrumentales, examinons le cas où les MCO fonctionnent. Soit le modèle :

$$y = Xb + u,$$

en prémultipliant ce modèle par X , on obtient :

$$\begin{aligned} X'y &= X'Xb + X'u \\ \Leftrightarrow (X'X)^{-1} X'y &= b + (X'X)^{-1} X'u \\ \Leftrightarrow \hat{b} &= b + \left(\frac{X'X}{NT} \right)^{-1} \frac{X'u}{NT} \end{aligned}$$

ce qui implique :

$$\text{Plim} \hat{b} = b \Leftrightarrow \text{Plim} \frac{X'u}{NT} = 0.$$

Cette relation s'appelle une *condition d'orthogonalité*, que l'on peut obtenir à partir des relations :

$$E(X'_{it} u_{it}) = 0 \quad \forall i, t.$$

La méthode des variables instrumentales consiste à définir cette condition d'orthogonalité par rapport à d'autres variables que X , que l'on appelle des *variables instrumentales*. Les variables instrumentales peuvent comprendre toutes les variables de X qui ne sont pas corrélées avec la perturbation. Par exemple, le terme constant du modèle est une variable instrumentale.¹ Nous allons maintenant définir plus précisément les propriétés que doivent vérifier ces variables. Dans la suite, on note Z l'ensemble des variables instrumentales :

$$Z_{(NT,m)} = \begin{bmatrix} Z_1 \\ (T,m) \\ \vdots \\ Z_N \\ (T,m) \end{bmatrix}$$

Soit un modèle linéaire (12.1), la matrice Z de dimension $NT \times m$ avec $m \geq$

1. Vérifions-le : $\text{Plim} e'_{NT} u / NT = E(u) = 0$, d'où le résultat. Il ne s'agit toutefois pas d'un instrument intéressant, puisqu'il ne repose pas sur une corrélation.

p constitue un instrument pour X si elle vérifie les trois conditions suivantes :

$$\text{Plim} \frac{Z'Z}{NT} = \Omega_{ZZ} \text{ inversible}$$

(m,m)

$$\text{Plim} \frac{Z'X}{NT} = \Omega_{ZX} \text{ de rang } p$$

(m,p)

$$\text{Plim} \frac{Z'u}{NT} = \mathbf{0}$$

(m,1)

Il s'agit donc d'un ensemble de variables linéairement indépendantes, asymptotiquement corrélées avec les variables explicatives et asymptotiquement orthogonales à la perturbation du modèle. Nous indiquerons comment les trouver dans la dernière section de ce chapitre.

12.3 L'estimateur à variables instrumentales

Cet estimateur permet de présenter le principe général d'estimation, mais constitue également la première étape de l'estimation optimale que nous présenterons plus loin. Nous allons considérer le cas où il y a plus d'instruments que de variables explicatives. Dans un premier temps, nous considérerons une perturbation homoscédastique, le modèle à erreur simple. Soit le modèle apparemment linéaire :

$$y = Xb + \varepsilon,$$

$$\text{Plim} \frac{X'\varepsilon}{NT} \neq 0, \text{Plim} \frac{Z'\varepsilon}{NT} = 0, E(\varepsilon|X, Z) = 0, V(\varepsilon|X, Z) = \sigma_\varepsilon^2 I_{NT}.$$

En pré-multipliant le modèle par Z , afin de faire apparaître $Z'\varepsilon$, on obtient un modèle auxiliaire pour l'estimation du paramètre b :

$$Z'y = Z'Xb + Z'\varepsilon.$$

La nouvelle perturbation du modèle vérifie donc :

$$E(Z'\varepsilon|Z, X) = Z'E(\varepsilon|Z, X) = 0,$$

$$V(Z'\varepsilon|Z, X) = Z'V(\varepsilon|Z, X)Z = \sigma_\varepsilon^2 Z'Z.$$

Comme $V(\varepsilon|Z, X)$ n'est pas scalaire, minimiser la somme des carrés de ce modèle ne serait pas, dans ce cas, optimal. Il faut d'abord pré-multiplier le modèle par $(Z'Z)^{-1/2}$ afin d'obtenir une matrice de covariance scalaire pour la

perturbation. Il s'agit d'une condition suffisante pour obtenir l'estimateur optimal. Ce qui donne finalement :

$$y^* = X^* b + \varepsilon^*,$$

$$\text{avec } y^* = (Z'Z)^{-1/2} Z' y, \quad X^* = (Z'Z)^{-1/2} Z' X, \quad \varepsilon^* = (Z'Z)^{-1/2} Z' \varepsilon.$$

La perturbation ε^* vérifie toutes les bonnes propriétés, on a :

$$E(\varepsilon^* | Z, X) = (Z'Z)^{-1/2} Z' E(\varepsilon | Z, X) = 0$$

$$V(\varepsilon^* | Z, X) = (Z'Z)^{-1/2} Z' \underbrace{V(\varepsilon | Z, X)}_{\sigma_\varepsilon^2 I_{NT}} Z (Z'Z)^{-1/2} = \sigma_\varepsilon^2 I_{NT}$$

De plus :

$$\begin{aligned} \text{Plim} \frac{1}{NT} X^{*'} \varepsilon^* &= \text{Plim} \frac{1}{NT} \left((Z'Z)^{-1/2} Z' X \right)' (Z'Z)^{-1/2} Z' \varepsilon \\ &= \text{Plim} \frac{1}{NT} X' Z (Z'Z)^{-1/2} (Z'Z)^{-1/2} Z' \varepsilon \\ &= \text{Plim} \frac{1}{NT} X' Z (Z'Z)^{-1} Z' \varepsilon \\ &= \text{Plim} \frac{X' Z}{NT} \left(\frac{Z' Z}{NT} \right)^{-1} \frac{Z' \varepsilon}{NT} \\ &= \Omega'_{ZX} \Omega_{ZZ}^{-1} \times \underbrace{\text{Plim} \frac{Z' \varepsilon}{NT}}_0 \\ &= 0. \end{aligned}$$

On définit la matrice de projection orthogonale sur $\text{Im}(Z)$ par :

$$P_Z = Z (Z'Z)^{-1} Z'.$$

On vérifie que cette matrice est idempotente (projection) $P_Z = P_Z^2$ et symétrique (orthogonale) $P_Z' = P_Z$. Elle permet d'obtenir les prévisions d'une variable par les vecteurs colonne de Z . On voit que dans la limite en probabilité

précédente² :

$$\begin{aligned}\text{Plim} \frac{1}{NT} X^{*'} u^* &= \text{Plim} \frac{1}{NT} X' P_Z \varepsilon \\ &= \text{Plim} \frac{1}{NT} X' P'_Z P_Z \varepsilon \\ &= \text{Plim} \frac{1}{NT} (P_Z X)' P_Z \varepsilon.\end{aligned}$$

On a donc remplacé X par sa prévision en fonction des variables de Z . Cette transformation a pour conséquence d'éliminer la corrélation entre les variables explicatives (transformées) et la perturbation, puisque Z est une matrice de variables exogènes. Dans la suite, on pose :

$$y_V = P_Z y \quad \text{et} \quad X_V = P_Z X,$$

qui sont les projections orthogonales de y et X sur $\text{Im}(Z)$. Nous avons montré que le modèle transformé vérifie les propriétés des moindres carrés ordinaires. L'estimateur qui minimise la somme des carrés des résidus de ce modèle est donc donné par la formule des moindres carrés ordinaires appliquée aux données transformées :

$$\begin{aligned}\hat{b}_V &= (X^{*'} X^*)^{-1} X^{*'} y \\ &= [X' Z (Z' Z)^{-1/2} (Z' Z)^{-1/2} Z' X]^{-1} X' Z (Z' Z)^{-1/2} (Z' Z)^{-1/2} Z' y \\ &= [X' Z (Z' Z)^{-1} Z' X]^{-1} X' Z (Z' Z)^{-1} Z' y \\ &= (X' P_Z X)^{-1} X' P_Z y.\end{aligned}$$

On l'appelle \hat{b}_V estimateur *par variables instrumentales*. Il se réécrit :

$$\hat{b}_V = [X'_V X_V]^{-1} X'_V y_V.$$

Cet estimateur est parfois appelé "estimateur des doubles moindres carrés" car il peut être obtenu en appliquant deux fois les moindres carrés.³ Dans un premier temps on régresse chaque colonne de X et y sur les colonnes de Z , pour obtenir les prévisions X_V . Dans un second temps, on régresse y_V sur les

2. D'après ce qui précède $P_Z = P'_Z P_Z$.

3. Habituellement on utilise le terme d'estimateur des doubles moindres carrés dans le cadre des systèmes à équations simultanées et celui d'estimateur par variables instrumentales lors de l'estimation d'un modèle à une seule équation.

prévisions de X par Z , égales à $X_V = P_Z X = (P_Z X^{(1)} | \dots | P_Z X^{(p)})$. En fait, il est *inutile* de remplacer y par sa prévision y_V , car on a :

$$\begin{aligned} X'_V y_V &= (P_Z X)' (P_Z y) = X' P'_Z P_Z y = X' P'_Z y = (P_Z X)' y = X'_V y \\ \Rightarrow \hat{b}_V &= [X'_V X_V]^{-1} X'_V y_V = [X'_V X_V]^{-1} X'_V y. \end{aligned}$$

Une deuxième façon d'interpréter cet estimateur consiste à dire que l'on impose une condition d'orthogonalité entre les variables instrumentales et le résidu. On a en effet :

$$(X_V)' \hat{u}_V = 0 \text{ avec } X_V = P_Z X \text{ et } \hat{u}_V = y - X \hat{b}_V,$$

ce qui implique :

$$(P_Z X)' \hat{u}_V = 0 \Leftrightarrow \hat{u}_V \perp \text{Im}(P_Z X),$$

où $\text{Im}(P_Z X)$ est l'espace vectoriel engendré par les prévisions de X par Z .

Afin d'étudier ses propriétés asymptotiques, on réécrit l'estimateur par variables instrumentales sous la forme :

$$\begin{aligned} \hat{b}_V &= (X' P_Z X)^{-1} X' P_Z (Xb + \varepsilon) \\ &= (X' P_Z X)^{-1} X' P_Z X b + (X' P_Z X)^{-1} X' P_Z \varepsilon \\ &= b + (X' P_Z X)^{-1} X' P_Z \varepsilon. \end{aligned}$$

On a :

$$\begin{aligned} \text{Plim} \hat{b}_V - b &= \text{Plim} \left(\frac{X' P_Z X}{NT} \right)^{-1} \frac{X' P_Z \varepsilon}{NT} \\ &= \left(\text{Plim} \frac{X' Z}{NT} \left(\text{Plim} \frac{Z' Z}{NT} \right)^{-1} \text{Plim} \frac{Z' X}{NT} \right)^{-1} \\ &\quad \text{Plim} \frac{X' Z}{NT} \left(\text{Plim} \frac{Z' Z}{NT} \right)^{-1} \text{Plim} \frac{Z' \varepsilon}{NT} \\ &= (\Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_{ZX})^{-1} \Omega'_{ZX} \Omega_{ZZ}^{-1} \text{Plim} \frac{Z' \varepsilon}{NT} \\ &= (\Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_{ZX})^{-1} \Omega'_{ZX} \Omega_{ZZ}^{-1} \times 0 \\ &= 0. \end{aligned}$$

L'estimateur par variables instrumentales est donc convergent. D'autre part, l'indépendance des perturbations ε implique celle des perturbations :

$$\varepsilon^* = (Z'Z)^{-1/2} Z'\varepsilon.$$

Donc cet estimateur est asymptotiquement normal. Il nous suffit de calculer sa variance asymptotique (i.e., celle de la distribution théorique vers laquelle l'estimateur converge). En utilisant $V(\varepsilon) = \sigma_\varepsilon^2 I_{NT}$ et $P_Z^2 = P_Z$, on a :

$$\begin{aligned} V\left[\sqrt{NT}(\hat{b}_V - b)\right] &= NT [X'P_ZX]^{-1} X'P_ZV(\varepsilon|X, Z) P_ZX [X'P_ZX]^{-1} \\ &= NT\sigma_\varepsilon^2 [X'P_ZX]^{-1} X'P_ZX [X'P_ZX]^{-1} \\ &= NT\sigma_\varepsilon^2 [X'P_ZX]^{-1} \\ &= \sigma_\varepsilon^2 \left[\frac{X'P_ZX}{NT} \right]^{-1}. \end{aligned}$$

En conséquence :

$$\begin{aligned} \text{Vas}\left[\sqrt{NT}(\hat{b}_V - b)\right] &= \text{Plim}\sigma_\varepsilon^2 \left[\frac{X'P_ZX}{NT} \right]^{-1} \\ &= \sigma_\varepsilon^2 (\Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_{ZX})^{-1} \end{aligned}$$

Un estimateur convergent de σ_ε^2 peut être obtenu à partir des résidus :

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{NT-p} (y - X\hat{b}_V)' (y - X\hat{b}_V) \quad (12.2)$$

Ici il convient de bien remarquer que *l'on ne calcule pas* cette variance à partir des résidus du modèle transformé car ceux-ci correspondent à :

$$\varepsilon^* = (Z'Z)^{-1/2} Z'\varepsilon \text{ et non à } \varepsilon.$$

L'estimateur de la variance asymptotique de $\sqrt{NT}(\hat{b}_V - b)$ est donc donné par :

$$\widehat{\text{Vas}}\left[\sqrt{NT}(\hat{b}_V - b)\right] = \widehat{\sigma}_\varepsilon^2 \left[\frac{X'P_ZX}{NT} \right]^{-1}.$$

Dans la pratique, on estimera la variance de l'estimateur à variables instrumentales par :

$$\widehat{\text{Vas}}(\hat{b}_V) = \widehat{\sigma}_\varepsilon^2 [X'P_ZX]^{-1}.$$

12.4 Cas particuliers

1. Avec autant d'instruments que de variables explicatives. Il s'agit du cas particulier $m = p$. Dans ce cas, la matrice $Z'X$ est inversible d'après nos hypothèses donc :⁴

$$\begin{aligned}\hat{b}_V &= [X'P_ZX]^{-1} X'P_Zy \\ &= [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1} Z'Z \underbrace{(X'Z)^{-1}(X'Z)}_{I_p} (Z'Z)^{-1}Z'y = (Z'X)^{-1}Z'y\end{aligned}$$

On peut donc se contenter d'étudier le cas $m \geq p$.

2. Quand X vérifie l'hypothèse $\text{Plim} X'\varepsilon/NT = 0$, il constitue un instrument admissible et l'on a $m = p$ par définition. En posant $Z = X$ on obtient :

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

on retrouve donc l'estimateur des moindres carrés ordinaires quand il est optimal.

12.5 Comment trouve t-on une variable instrumentale ?

Les méthodes précédentes ne peuvent être appliquée que si l'on trouve des instruments valides. Il faut qu'ils vérifient les trois conditions suivantes :

$$\text{Plim} \frac{Z'Z}{NT} \text{ inversible, } \text{Plim} \frac{Z'X}{NT} \text{ de rang } p \text{ et } \text{Plim} \frac{Z'u}{NT} = 0,$$

la première condition signifie que les instruments ne doivent pas être colinéaires, la seconde condition impose que les instruments soient corrélés avec X ; la troisième qu'ils ne doivent pas être corrélés avec u . Les deux dernières conditions impliquent que les instruments ne doivent influencer la variable expliquée y que *via* les variables explicatives du modèle X et non directement.⁵ Il existe deux manières, complémentaires, de trouver des instruments.

4. On rappelle que $(AB)^{-1} = B^{-1}A^{-1}$ et que $(A')^{-1} = (A^{-1})'$.

5. Ces conditions n'impliquent donc pas que les instruments ne soient pas corrélés avec la variable y , comme on l'entend parfois, mais que la corrélation *partielle* entre y et Z sachant X soit nulle.

1. La première méthode vise à vérifier une propriété de cohérence externe du modèle, c'est-à-dire à le confronter à l'ensemble de la littérature antérieure. Elle consiste à étudier la littérature sur les déterminants de X et à éliminer ceux qui sont déjà présents dans l'équation. Ici, il faut consulter à la fois la littérature théorique et la littérature empirique, pour savoir quels sont les bons instruments pour X , c'est-à-dire les plus corrélés. On parle de cohérence externe car on va chercher de l'information en dehors du modèle estimé et, plus précisément, sur les articles qui étudient la relation entre X et Z plutôt qu'entre y et X . Pour fixer les idées, prenons un exemple. Supposons que l'on veuille estimer une fonction de production reliant la productivité y_{it} à l'intensité capitalistique k_{it} , à l'emploi ℓ_{it} et à l'innovation I_{it} , cette dernière variables mesurant directement les changements technologiques apportés au processus de production :

$$y_{it} = \alpha c_{it} + \beta \ell_{it} + \gamma I_{it} + u$$

A priori, on pourrait penser que les changements technologiques causent la production. Toutefois, les données que l'on obtient sont généralement agrégées sur plusieurs années (3 ans, pour les enquêtes innovation) de sorte que la productivité à la date $t - 2$ (y_{it-2}) pourrait avoir eu le temps d'influencer la mise en oeuvre d'innovations aux dates $t - 2$, $t - 1$ ou t . Ceci impliquerait une corrélation entre I_{it} et $(u_{it}, u_{it-1}, u_{it-2})$ surtout si cette dernière perturbation est autocorrélée. Pour régler ce problème d'estimation ou, comme nous le verrons plus loin, pour tester que l'innovation n'est pas corrélée avec la perturbation u du modèle il nous faut une variable instrumentale qui doit être corrélée à l'innovation, qui soit différente du capital et de l'emploi, et qui ne soit pas corrélée avec u . En examinant la littérature sur les déterminants de l'innovation, on trouve que les investissements en recherche et développement et les achats de licence sont des déterminants de l'innovation. Or ces variables n'interviennent pas directement dans la fonction de production. Seul le résultat de la recherche, l'innovation, et non le simple fait de chercher, modifie la fonction de production. Les dépenses de recherche et plus généralement, tous les déterminants de l'innovation, peuvent être utilisés comme instruments. Pour parvenir à cette conclusion, toutefois, il faut consulter la littérature sur les déterminants de l'innovation, et non sur la fonction de production.

2. La seconde méthode complète la première. Elle consiste à vérifier une certaine cohérence interne au modèle. Prenons le cas du modèle autoré-

gressif :

$$y_{it} = \rho_1 y_{it-1} + \tilde{X}_{it} c + u_{it}$$

$$= \underbrace{(y_{it-1}, \tilde{X}_{it})}_{X_{it}} \underbrace{\begin{pmatrix} \rho_1 \\ c \end{pmatrix}}_b + u_{it},$$

avec

$$u_{it} = \alpha_i + \varepsilon_{it},$$

si l'on décale le modèle d'une période, on obtient :

$$y_{it-1} = \rho_1 y_{it-2} + X_{it-1} c + \alpha_i + \varepsilon_{it-1},$$

cette seconde relation met deux points en évidence. D'une part, la variable y_{it-1} est corrélée avec l'effet individuel α_i . D'autre part, les variables X_{it-1} constituent des instruments valables pour y_{it-1} car ils font partie de ses déterminants, sans avoir d'influence directe sur y_{it} , puisque seules les variables explicatives prises à la date t influencent directement y_{it} (et non celles prises en $t-1$). On pourrait donc prendre comme instrument la variable :

$$Z_{it} = (X_{it-1}, X_{it}).$$

En effet, les variables présentes dans X_{it} ne sont pas endogènes et constituent donc leur meilleur instrument, la variable y_{it-1} est endogène et peut être instrumentée par les toutes les variables explicatives exogènes retardées du modèle.

Arrivé à ce stade, il demeure toutefois un argument qui confère une supériorité à la première méthode sur la seconde : la cohérence interne repose sur l'hypothèse importante que le modèle est juste alors, qu'en toute rigueur, il ne s'agit *que* d'une hypothèse. Ce dernier point permet d'affirmer que les tests de spécification jouent un rôle crucial. Il faut les faire et, s'ils ne rejettent pas le modèle, penser que le *minimum* est fait. Pour avoir des estimation plus solides, il faut également effectuer l'examen de cohérence externe.

12.6 Les tests de spécification

L'expression "tests de spécification" regroupe l'ensemble des tests qui permettent d'évaluer le degré de vraisemblance des hypothèses du modèle. Les tests les plus importants sont ceux qui portent sur les hypothèses garantissant la convergence de l'estimateur. Nous allons voir trois de ces tests : le test de

Wu-Hausman, qui indique s'il vaut mieux faire des MCO ou des variables instrumentales; le test de Hausman-Taylor qui indique si l'on peut ajouter des variables instrumentales à un ensemble d'instruments déjà existant et le test de Sargan-Hansen qui permet de tester la validité des instruments.

12.6.1 Le test de Wu-Hausman

Savoir s'il vaut mieux faire des moindres carrés ordinaires ou des variables instrumentales est important, parce que si les variables explicatives sont exogènes, les moindres carrés ordinaires sont toujours plus efficaces que les variables instrumentales. Nous commencerons donc par démontrer cette propriété. Nous passerons ensuite au test lui-même.

L'efficacité des MCO

On prend un modèle linéaire simple :

$$y = Xb + u, \quad E(u|X, Z) = 0, \quad V(u|X, Z) = \sigma_u^2 I_{NT},$$

où Z est un ensemble de variables instrumentales non corrélées avec la perturbation u et corrélées avec X . L'estimateur des MCO est défini par :

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

sa variance est donnée par :

$$\begin{aligned} V(\hat{b}|X, Z) &= E\left[(\hat{b} - b)(\hat{b} - b)' \mid X, Z\right] \\ &= (X'X)^{-1} X' \underbrace{E(uu' \mid X, Z)}_{\sigma_u^2 I_{NT}} X (X'X)^{-1} \\ &= \sigma_u^2 (X'X)^{-1}. \end{aligned}$$

L'estimateur par variables instrumentales \tilde{b} utilise les prévisions de X par Z à la place de X , on a :

$$\tilde{b} = (X'P_Z X)^{-1} X'P_Z y = b + (X'P_Z X)^{-1} X'P_Z u,$$

d'où la variance :

$$\begin{aligned} V(\tilde{b}|X, Z) &= E\left[(\tilde{b} - b)(\tilde{b} - b)' \mid X, Z\right] \\ &= (X'P_Z X)^{-1} X'P_Z \underbrace{E(uu' \mid X, Z)}_{\sigma_u^2 I_{NT}} P_Z X (X'X)^{-1} \\ &= \sigma_u^2 (X'P_Z X)^{-1}. \end{aligned}$$

Le résultat sur l'efficacité de l'estimateur des MCO vient du fait que :⁶

$$\begin{aligned} X'X - X'P_ZX &= X'(I_{NT} - P_Z)X \\ &= X'(I_{NT} - P_Z)'(I_{NT} - P_Z)X \\ &= ((I_{NT} - P_Z)X)'(I_{NT} - P_Z)X \gg 0, \end{aligned}$$

où $A \gg 0$ désigne une matrice définie positive.⁷ Ceci est équivalent à :

$$\begin{aligned} X'X &\gg X'P_ZX \\ \Leftrightarrow (X'X)^{-1} &\ll (X'P_ZX)^{-1} \\ \Leftrightarrow \sigma_u^2 (X'X)^{-1} &\ll \sigma_u^2 (X'P_ZX)^{-1} \\ \Leftrightarrow V(\hat{b}|X, Z) &\ll V(\tilde{b}|X, Z). \end{aligned}$$

Ainsi, quand on peut faire des moindres carrés ordinaires, il ne faut pas utiliser l'estimateur par variables instrumentales.

La statistique de test

Le test vise à vérifier si certaines variables du modèle sont endogènes. Dans le cas où elles ne le sont pas, il faut faire des MCO, dans le cas contraire il faut faire des variables instrumentales. On considère le modèle suivant :

$$y = X_1 b_1 + X_2 b_2 + u,$$

où X_1 est la matrice des variables explicatives qui présentent potentiellement une corrélation avec le terme d'erreur u , et X_2 les variables explicatives exogènes du modèle. On pose :

$$X = (X_1, X_2) \text{ et } b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

6. On utilise le fait que $I_{NT} - P_Z$ est une matrice de projection orthogonale, elle est donc symétrique et idempotente.

7. Rappel : une matrice $A_{(p,p)}$ est définie positive si :

$$\forall v \neq 0, v \in \mathbb{R}^p, \underset{(1,p)(p,p)(p,1)}{v' A v} > 0.$$

Dans cette application, on a :

$$\underset{(1,p)}{v'} \underset{(p,NT)}{((I_{NT} - P_Z)X)'} \underset{(NT,p)}{(I_{NT} - P_Z)X} \underset{(p,1)}{v} = \|(I_{NT} - P_Z)Xv\|^2 > 0.$$

de sorte que le modèle s'écrit aussi :

$$y = Xb + u.$$

Le test est le suivant :

$$\begin{cases} H_0 : \text{Plim} X_1' u / NT = 0 \\ H_a : \text{Plim} X_1' u / NT \neq 0 \end{cases}$$

Pour réaliser ce test, on calcule une distance entre deux estimateurs :

1. L'estimateur des MCO de b_1 , noté \hat{b}_1 ;
2. L'estimateur à variables instrumentales de b_1 , noté \tilde{b}_1 ;
3. On ne considère pas les estimateurs de b_2 car les variables correspondantes n'ont pas besoin d'être instrumentées et que leur prise en compte génère généralement des problèmes d'inversibilité des matrices de covariances.

La statistique de test est simplement égale à :

$$\xi_{WH} = (\hat{b}_1 - \tilde{b}_1)' \hat{V} (\hat{b}_1 - \tilde{b}_1)^{-1} (\hat{b}_1 - \tilde{b}_1) \xrightarrow[H_0]{L} \chi_{p_1}^2,$$

où p_1 est le nombre de variables présentes dans X_1 . Pour calculer cette statistique de test, il nous faut un ensemble de variables instrumentales, en nombre $m_1 \geq p_1$, de matrice Z_1 , vérifiant :

$$\text{Plim} \frac{Z_1' X_1}{NT} \text{ de rang } p_1, \text{Plim} \frac{Z_1' u}{NT} = 0, \text{Plim} \frac{Z_1' Z_1}{NT} \text{ de rang } m_1.$$

Commençons par calculer l'estimateur des MCO de b_1 , il est défini par la solution du système suivant :

$$\begin{aligned} (X'X) \hat{b} &= X'y \\ \Leftrightarrow \begin{pmatrix} X_1' X_1 & X_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{pmatrix} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} &= \begin{pmatrix} X_1' y \\ X_2' y \end{pmatrix} \\ \Leftrightarrow \begin{cases} X_1' X_1 \hat{b}_1 + X_1' X_2 \hat{b}_2 &= X_1' y \\ X_2' X_1 \hat{b}_1 + X_2' X_2 \hat{b}_2 &= X_2' y \end{cases} \end{aligned}$$

En résolvant la seconde équation du système, on obtient :

$$\hat{b}_2 = (X_2' X_2)^{-1} X_2' (y - X_1 \hat{b}_1),$$

et en reportant ce résultat dans la première équation on obtient :

$$X_1' X_1 \hat{b}_1 + \underbrace{X_1' X_2 (X_2' X_2)^{-1} X_2'}_{P_{X_2}} (y - X_1 \hat{b}_1) = X_1' y,$$

où P_{X_2} est la matrice de projection orthogonale sur $\text{Im}(X_2)$. On obtient finalement :

$$\hat{b}_1 = (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (I_{NT} - P_{X_2}) y.$$

La matrice $I_{NT} - P_{X_2}$ réalise une projection orthogonale sur $\text{Im}^\perp(X_2)$. Ainsi, l'estimateur \hat{b}_1 est obtenu en enlevant à X_1 et à y leurs parties qui peuvent être expliquées par X_2 , puis en faisant une régression par les MCO avec les variables ainsi filtrées. On en déduit que \hat{b}_1 repose sur la corrélation partielle de y et X_1 . La variance de cet estimateur des MCO s'obtient en remplaçant y par sa valeur dans l'expression précédente :

$$\begin{aligned} \hat{b}_1 &= (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (I_{NT} - P_{X_2}) (X_1 b_1 + X_2 b_2 + u) \\ &= b_1 + (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (I_{NT} - P_{X_2}) (X_2 b_2 + u), \end{aligned}$$

le premier terme du membre de droite s'annule car :

$$\begin{aligned} (I_{NT} - P_{X_2}) X_2 &= X_2 - X_2 (X_2' X_2)^{-1} X_2' X_2 \\ &= X_2 - X_2 \\ &= 0, \end{aligned}$$

il reste donc :

$$\hat{b}_1 = b_1 + (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (I_{NT} - P_{X_2}) u,$$

ce qui implique :

$$V(\hat{b}_1 | X, Z) = \sigma_u^2 (X_1' (I_{NT} - P_{X_2}) X_1)^{-1}.$$

Pour l'estimateur par variables instrumentales, on introduit les variables qui instrumentent X_1 . On a :

$$X = (X_1, X_2) \text{ et } Z = (Z_1, X_2),$$

on remarque que l'on a gardé X_2 parmi les instruments car ces variables sont exogènes. Pour aller plus avant, on a besoin de la propriété suivante :

$$X_2 \subset \text{Im}(Z) \Rightarrow P_Z X_2 = X_2. \quad (12.3)$$

Il suffit de remarquer la projection de X_2 sur un espace incluant cette variable s'obtient par les moindres carrés ordinaires. La projection est de la forme générale : $P_Z X_2 = Z_1 \hat{\pi}_1 + X_2 \hat{\pi}_2$ et les coefficients de cette projection s'obtiennent en régressant X_2 sur Z par les moindres carrés ordinaires :

$$\begin{aligned} Z' Z \hat{\pi} &= Z' X_2 \\ \Leftrightarrow \begin{pmatrix} Z_1' Z_1 & Z_1' X_2 \\ X_2' Z_1 & X_2' X_2 \end{pmatrix} \begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} &= \begin{pmatrix} Z_1' X_2 \\ X_2' X_2 \end{pmatrix} \\ \Leftrightarrow \begin{cases} Z_1' Z_1 \hat{\pi}_1 + Z_1' X_2 \hat{\pi}_2 &= Z_1' X_2 \\ X_2' Z_1 \hat{\pi}_1 + X_2' X_2 \hat{\pi}_2 &= X_2' X_2 \end{cases} \end{aligned}$$

dont la solution est triviale :⁸

$$\hat{\pi}_1 = \mathbf{0}_{p_1}, \hat{\pi}_2 = \mathbf{I}_{p_2},$$

on en déduit que :

$$P_Z X_2 = Z_1 \mathbf{0}_{p_1} + X_2 \mathbf{I}_{p_2} = X_2.$$

Nous pouvons maintenant définir l'estimateur par variables instrumentales \tilde{b} :

$$\begin{aligned} X' P_Z X \tilde{b} &= X' P_Z y \\ \Leftrightarrow \begin{pmatrix} X_1' P_Z X_1 & X_1' P_Z X_2 \\ X_2' P_Z X_1 & X_2' P_Z X_2 \end{pmatrix} \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{pmatrix} &= \begin{pmatrix} X_1' P_Z y \\ X_2' P_Z y \end{pmatrix}, \end{aligned}$$

d'après la propriété (12.3) :

$$P_Z X_2 = X_2,$$

donc on obtient le système :

$$\begin{cases} X_1' P_Z X_1 \tilde{b}_1 + X_1' X_2 \tilde{b}_2 &= X_1' P_Z y \\ X_2' X_1 \tilde{b}_1 + X_2' X_2 \tilde{b}_2 &= X_2' y \end{cases}$$

dont la seconde équation implique :

$$\tilde{b}_2 = (X_2' X_2)^{-1} X_2' (y - X_1 \tilde{b}_1),$$

en reportant dans la première équation, on obtient :

$$X_1' P_Z X_1 \tilde{b}_1 + X_1' \underbrace{X_2 (X_2' X_2)^{-1} X_2'}_{P_{X_2}} (y - X_1 \tilde{b}_1) = X_1' P_Z y,$$

8. Le lecteur peut aussi la vérifier par le calcul.

ce qui donne finalement :

$$\tilde{b}_1 = (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) y,$$

l'interprétation généralise celle des moindres carrés ordinaires. Ici, on purge X_1 et y de la partie des instruments présents dans l'équation X_2 , de sorte que l'estimation se base en réalité sur les instruments *ajoutés* aux variables exogènes du modèle Z_1 . Cet estimateur peut également s'écrire :

$$\begin{aligned} \tilde{b}_1 &= (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) (X_1 b_1 + X_2 b_2 + u) \\ &= b_1 + (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) (X_2 b_2 + u), \end{aligned}$$

et

$$\begin{aligned} (P_Z - P_{X_2}) X_2 &= P_Z X_2 - P_{X_2} X_2 \\ &= X_2 - X_2 \\ &= 0, \end{aligned}$$

d'où l'expression :

$$\tilde{b}_1 = b_1 + (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) u.$$

La variance de cet estimateur est donnée par :

$$V(\tilde{b}_1 | X, Z) = \sigma_u^2 (X_1' (P_Z - P_{X_2}) X_1)^{-1}.$$

Pour obtenir la statistique de test, on a besoin de la contrepartie empirique de la quantité suivante :

$$\begin{aligned} \text{Cov}(\hat{b}_1, \tilde{b}_1 | X, Z) &= E \left[(\hat{b}_1 - b_1) (\tilde{b}_1 - b_1)' \right] \\ &= (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (I_{NT} - P_{X_2}) E(u u' | X, Z) \\ &\quad (P_Z - P_{X_2}) X_1 (X_1' (P_Z - P_{X_2}) X_1)^{-1}, \end{aligned}$$

ce qui s'écrit :

$$\begin{aligned} \text{Cov}(\hat{b}_1, \tilde{b}_1 | X, Z) &= \sigma_u^2 (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' \\ &\quad (I_{NT} - P_{X_2}) (P_Z - P_{X_2}) X_1 (X_1' (P_Z - P_{X_2}) X_1)^{-1} \end{aligned}$$

Pour simplifier, il nous faut la propriété suivante :

$$\text{Im}(X_2) \subset \text{Im}(Z) \Rightarrow P_Z P_{X_2} = P_{X_2}. \quad (12.4)$$

D'après la propriété (12.3) :

$$P_Z P_{X_2} = \underbrace{P_Z X_2}_{X_2} (X_2' X_2)^{-1} X_2 = P_{X_2}.$$

Ceci implique :

$$\begin{aligned} (I_{NT} - P_{X_2})(P_Z - P_{X_2}) &= P_Z - P_{X_2} - P_{X_2} P_Z + P_{X_2}^2 \\ &= P_Z - P_{X_2} - P_{X_2} + P_{X_2} \\ &= P_Z - P_{X_2}, \end{aligned}$$

d'où la covariance :

$$\begin{aligned} \text{Cov}(\hat{b}_1, \tilde{b}_1 | X, Z) &= \sigma_u^2 (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) X_1 \\ &\quad (X_1' (P_Z - P_{X_2}) X_1)^{-1} \\ &= \sigma_u^2 (X_1' (I_{NT} - P_{X_2}) X_1)^{-1} \\ &= V(\hat{b}_1 | X, Z). \end{aligned}$$

La variance dont on a besoin pour la statistique de test est donc la contrepartie empirique de :

$$\begin{aligned} V(\hat{b}_1 - \tilde{b}_1) &= V(\hat{b}_1) + V(\tilde{b}_1) - 2\text{Cov}(\hat{b}_1, \tilde{b}_1) \\ &= V(\hat{b}_1) + V(\tilde{b}_1) - 2V(\hat{b}_1) \\ &= V(\tilde{b}_1) - V(\hat{b}_1) \gg 0, \end{aligned}$$

cette matrice est définie positive car l'estimateur des MCO est toujours plus précis que l'estimateur à variables instrumentales. Pour les contreparties empiriques ceci n'est toujours vrai que si l'on utilise *le même estimateur* de σ_u^2 pour les deux variances empiriques :

$$\begin{aligned} \hat{V}(\hat{b}_1) &= \hat{\sigma}_u^2 (X_1' (I_{NT} - P_{X_2}) X_1)^{-1}, \\ \hat{V}(\tilde{b}_1) &= \hat{\sigma}_u^2 (X_1' (P_Z - P_{X_2}) X_1)^{-1}, \end{aligned}$$

pour $\hat{\sigma}_u^2$ on peut prendre l'estimateur des MCO, a priori préférable sous l'hypothèse nulle :

$$\hat{\sigma}_u^2 = \frac{1}{NT - p} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - X_{it} \hat{b})^2.$$

On obtient donc un test de Hausman, d'où son nom :⁹

9. L'article de Wu date de 1973 et celui de Hausman de 1976, d'où l'ordre des noms.

1. L'estimateur des MCO \hat{b}_1 est optimal sous H_0 ;
2. L'estimateur à variables instrumentales \tilde{b}_1 est convergent sous les deux hypothèses, et n'est pas optimal sous H_0 ;
3. C'est ce qui explique que la statistique se simplifie ; toutefois, on ne peut pas utiliser tous les coefficients de régression pour faire le test, car la différence globale des matrices de covariances ne serait plus inversible ; il faut se limiter aux coefficients des variables que l'on considère comme endogènes.

12.6.2 Le test de Hausman-Taylor

Un estimateur à variables instrumentales est plus efficace qu'un autre quand il possède un plus grand nombre d'instruments. Il est donc important de savoir tester la validité d'instruments supplémentaires. Nous allons commencer par démontrer cette propriété avant de développer le test lui-même.

Efficacité et nombre d'instruments

On prend un modèle linéaire simple :

$$y = Xb + u, \quad E(u|X, Z) = 0, \quad V(u|X, Z) = \sigma_u^2 I_{NT},$$

où Z est un ensemble de variables instrumentales non corrélées avec la perturbation u et corrélées avec X . Supposons que l'on dispose d'un premier estimateur par variables instrumentales \tilde{b}_1 , défini à partir d'un sous-ensemble d'instruments Z_1 et d'un second estimateur par variables instrumentales \tilde{b} à partir de tous les instruments disponibles $Z = (Z_1, Z_2)$. Le premier estimateur est donné par :

$$\tilde{b}_1 = (X'P_{Z_1}X)^{-1} X'P_{Z_1}y,$$

de variance :

$$V(\tilde{b}_1|X, Z) = \sigma_u^2 (X'P_{Z_1}X)^{-1}$$

Avec le second ensemble de variables instrumentales, on définit l'estimateur :

$$\tilde{b} = (X'P_ZX)^{-1} X'P_Zy,$$

de variance :

$$V(\tilde{b}|X, Z) = \sigma_u^2 (X'P_ZX)^{-1}.$$

Le résultat sur les estimateurs à variables instrumentales vient du fait que :

$$X'P_ZX - X'P_{Z_1}X = X'(P_Z - P_{Z_1})X,$$

or :

$$(P_Z - P_{Z_1})' = P_Z - P_{Z_1},$$

et, d'après la propriété (12.4) :

$$\begin{aligned} (P_Z - P_{Z_1})^2 &= P_Z^2 - P_Z P_{Z_1} - P_{Z_1} P_Z + P_{Z_1}^2 \\ &= P_Z - P_{Z_1} - P_{Z_1} + P_{Z_1} \\ &= P_Z - P_{Z_1}, \end{aligned}$$

on peut donc écrire :

$$X' P_Z X - X' P_{Z_1} X = [(P_Z - P_{Z_1}) X]' (P_Z - P_{Z_1}) X \gg 0,$$

on en déduit que :

$$\begin{aligned} X' P_Z X &\gg X' P_{Z_1} X \\ \Leftrightarrow (X' P_Z X)^{-1} &\ll (X' P_{Z_1} X)^{-1} \\ \Leftrightarrow \sigma_u^2 (X' P_Z X)^{-1} &\ll \sigma_u^2 (X' P_{Z_1} X)^{-1} \\ \Leftrightarrow V(\tilde{b} | X, Z) &\ll V(\tilde{b}_1 | X, Z). \end{aligned}$$

Ainsi, il faut toujours utiliser la totalité des instruments disponibles.

La statistique de test

Le test vise à vérifier la validité de nouveaux instruments dans le but d'améliorer l'efficacité de l'estimateur par variables instrumentales. On considère le modèle suivant :

$$y = X_1 b_1 + X_2 b_2 + u,$$

où X_1 est la matrice des p_1 variables explicatives *endogènes*, X_2 les p_2 variables explicatives exogènes du modèle. Ce modèle a été estimé par la méthode des variables instrumentales à l'aide de la matrice Z_1 , de rang colonne $m_1 \geq p_1$. On pose :

$$X = (X_1, X_2), \text{ et } b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

l'ensemble d'instruments supplémentaires que l'on veut tester est Z_2 . Le test est le suivant :

$$\begin{cases} H_0: & \text{Plim} Z_2' u / NT = 0 \\ H_a: & \text{Plim} Z_2' u / NT \neq 0 \end{cases}$$

Pour réaliser ce test, on calcule une distance entre deux estimateurs :

1. L'estimateur à variables instrumentales basé sur $Z = (Z_1, X_2)$, noté \tilde{b}_1 ;
2. L'estimateur à variables instrumentales basé sur $\bar{Z} = (Z_1, Z_2, X_2)$, noté \bar{b}_1 ;
3. On ne considère pas les estimateurs de b_2 car les variables correspondantes n'ont pas besoin d'être instrumentées et que leur prise en compte génère généralement des problèmes d'inversibilité des matrices de covariances.

La statistique de test est égale à :

$$\xi_{HT} = (\tilde{b}_1 - \bar{b}_1)' \widehat{V}(\tilde{b}_1 - \bar{b}_1)^{-1} (\tilde{b}_1 - \bar{b}_1) \xrightarrow[H_0]{L} \chi_{p_1}^2,$$

où p_1 est le nombre de variables présentes dans X_1 . Commençons par calculer le premier estimateur à variables instrumentales de b_1 , il est défini par la solution du système suivant :

$$\begin{aligned} (X'P_Z X) \tilde{b} &= X'P_Z y \\ \Leftrightarrow \begin{pmatrix} X'_1 P_Z X_1 & X'_1 P_Z X_2 \\ X'_2 P_Z X_1 & X'_2 P_Z X_2 \end{pmatrix} \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{pmatrix} &= \begin{pmatrix} X'_1 P_Z y \\ X'_2 P_Z y \end{pmatrix} \\ \Leftrightarrow \begin{cases} X'_1 P_Z X_1 \tilde{b}_1 + X'_1 X_2 \tilde{b}_2 &= X'_1 P_Z y \\ X'_2 X_1 \tilde{b}_1 + X'_2 X_2 \tilde{b}_2 &= X'_2 y \end{cases} \end{aligned}$$

car :

$$\text{Im}(X_2) \subset \text{Im}(Z) \Rightarrow P_Z X_2 = X_2$$

En résolvant la seconde équation du système, on obtient :

$$\tilde{b}_2 = (X'_2 X_2)^{-1} X'_2 (y - X_1 \tilde{b}_1),$$

et en reportant ce résultat dans la première équation on obtient :

$$X'_1 P_Z X_1 \tilde{b}_1 + \underbrace{X'_1 X_2 (X'_2 X_2)^{-1} X'_2}_{P_{X_2}} (y - X_1 \tilde{b}_1) = X'_1 P_Z y,$$

où P_{X_2} est la matrice de projection orthogonale sur $\text{Im}(X_2)$. On obtient finalement :

$$\tilde{b}_1 = (X'_1 (P_Z - P_{X_2}) X_1)^{-1} X'_1 (P_Z - P_{X_2}) y.$$

La variance de cet estimateur s'obtient en remplaçant y par sa valeur dans l'expression précédente :

$$\begin{aligned} \hat{b}_1 &= (X'_1 (P_Z - P_{X_2}) X_1)^{-1} X'_1 (P_Z - P_{X_2}) (X_1 b_1 + X_2 b_2 + u) \\ &= b_1 + (X'_1 (P_Z - P_{X_2}) X_1)^{-1} X'_1 (P_Z - P_{X_2}) (X_2 b_2 + u), \end{aligned}$$

le premier terme du membre de droite s'annule car :

$$\begin{aligned}(P_Z - P_{X_2}) X_2 &= X_2 - X_2 (X_2' X_2)^{-1} X_2' X_2 \\ &= X_2 - X_2 \\ &= 0,\end{aligned}$$

il reste donc :

$$\tilde{b}_1 = b_1 + (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) u,$$

ce qui implique :

$$V(\tilde{b}_1 | X, Z) = \sigma_u^2 (X_1' (P_Z - P_{X_2}) X_1)^{-1}.$$

Pour le second estimateur par variables instrumentales, on ajoute les instruments Z_2 . On a :

$$\bar{Z} = (Z_1, Z_2, X_2),$$

en raisonnant par analogie, on obtient :

$$\begin{aligned}\bar{b}_1 &= (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1} X_1' (P_{\bar{Z}} - P_{X_2}) y \\ &= b_1 + (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1} X_1' (P_{\bar{Z}} - P_{X_2}) u\end{aligned}$$

de variance :

$$V(\bar{b}_1 | X, Z) = \sigma_u^2 (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1}.$$

Pour obtenir la statistique de test, on a besoin de la contrepartie empirique de la quantité suivante :

$$\begin{aligned}\text{Cov}(\tilde{b}_1, \bar{b}_1 | X, Z) &= E \left[(\tilde{b}_1 - b_1) (\bar{b}_1 - b_1)' \right] \\ &= (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' (P_Z - P_{X_2}) E(u u' | X, Z) \\ &\quad (P_{\bar{Z}} - P_{X_2}) X_1 (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1},\end{aligned}$$

ce qui s'écrit :

$$\begin{aligned}\text{Cov}(\tilde{b}_1, \bar{b}_1 | X, Z) &= \sigma_u^2 (X_1' (P_Z - P_{X_2}) X_1)^{-1} X_1' \\ &\quad (P_Z - P_{X_2}) (P_{\bar{Z}} - P_{X_2}) X_1 (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1},\end{aligned}$$

avec :

$$\begin{aligned}(P_Z - P_{X_2}) (P_{\bar{Z}} - P_{X_2}) &= P_Z P_{\bar{Z}} - P_Z P_{X_2} - P_{X_2} P_{\bar{Z}} + P_{X_2}^2 \\ &= P_Z - P_{X_2} - P_{X_2} + P_{X_2} \\ &= P_Z - P_{X_2},\end{aligned}$$

car

$$\text{Im}(X_2) \subset \text{Im}(Z) \subset \text{Im}(\bar{Z}),$$

d'où la covariance :

$$\begin{aligned} \text{Cov}(\tilde{b}_1, \bar{b}_1 | X, Z) &= \sigma_u^2 (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1} \\ &= V(\bar{b}_1 | X, Z). \end{aligned}$$

La variance dont on a besoin pour la statistique de test est donc la contrepartie empirique de :

$$\begin{aligned} V(\tilde{b}_1 - \bar{b}_1) &= V(\tilde{b}_1) + V(\bar{b}_1) - 2\text{Cov}(\tilde{b}_1, \bar{b}_1) \\ &= V(\tilde{b}_1) + V(\bar{b}_1) - 2V(\bar{b}_1) \\ &= V(\tilde{b}_1) - V(\bar{b}_1) \gg 0, \end{aligned}$$

cette matrice est définie positive car l'estimateur avec le plus d'instruments est toujours plus précis. Pour les contreparties empiriques ceci n'est toujours vrai que si l'on utilise *le même estimateur* de σ_u^2 pour les deux variances empiriques :

$$\begin{aligned} \hat{V}(\tilde{b}_1) &= \hat{\sigma}_u^2 (X_1' (P_Z - P_{X_2}) X_1)^{-1}, \\ \hat{V}(\bar{b}_1) &= \hat{\sigma}_u^2 (X_1' (P_{\bar{Z}} - P_{X_2}) X_1)^{-1}, \end{aligned}$$

pour $\hat{\sigma}_u^2$ on peut prendre l'estimateur sous l'hypothèse nulle :

$$\hat{\sigma}_u^2 = \frac{1}{NT - p} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - X_{it} \bar{b})^2.$$

On obtient donc un test de Hausman :¹⁰

1. L'estimateur avec toutes les variables instrumentales \bar{b}_1 est optimal sous H_0 ;
2. L'estimateur avec une partie des variables instrumentales \tilde{b}_1 est convergent sous les deux hypothèses, et n'est pas optimal sous H_0 ;
3. C'est ce qui explique que la statistique se simplifie ; toutefois, on ne peut pas utiliser tous les coefficients de régression pour faire le test, car la différence globale des matrices de covariances ne serait plus inversible ; il faut se limiter aux coefficients des variables que l'on considère comme endogènes (X_1).

10. L'article commun de Hausman et Taylor date de 1981.

12.7 Programmation 1 : test de Wu-Hausman

Nous allons, dans un premier temps, appliquer l'approche par variables instrumentales sans rechercher tous les instruments disponibles. Nous serons plus complets dans les chapitres suivants. Le modèle que nous souhaitons estimer vise à évaluer l'impact de la recherche et développement sur l'efficacité productive des entreprises. Pour cela, on considère une fonction de production de Cobb-Douglas :

$$Y_{it} = A_{it} C_{it}^{\gamma_C} L_{it}^{\gamma_L}, \quad (12.5)$$

où a est le terme constant, α_i est un effet individuel et ε_{it} un bruit blanc. La productivité globale des facteurs, peut s'écrire, en logarithmes :

$$y_{it} - \gamma_C c_{it} - \gamma_L \ell_{it} = a_{it}.$$

En fixant, comme pour beaucoup d'études, $\gamma_C = 1/3$ et $\gamma_L = 2/3$, nous obtenons une estimation de cette productivité globale des facteurs. Pour notre application, on pose donc :

$$a_{it} = y_{it} - \frac{1}{3} c_{it} - \frac{2}{3} \ell_{it}.$$

Ici, nous voulons tester le modèle dynamique :

$$a_{it} = (1 - \rho_1) a_{it-1} + \gamma_K k_{it} + a + \alpha_i + \varepsilon_{it},$$

où a est le terme constant du modèle, α_i est un effet individuel et ε_{it} un bruit blanc. Il est clair que a_{it-1} est corrélé avec α_i . Pour procéder à l'estimation par la méthode des variables instrumentales, il faut des variables qui soit corrélées avec a_{it-1} sans être corrélée avec α_i . On prend $\{\Delta c_{it}, \Delta \ell_{it}\}$. Comme ces variables sont prises en différences, elles ne devraient pas être corrélées avec α_i , mais elles restent corrélée avec la variation du logarithme de la productivité globale des facteurs. On lance donc le test avec :

$$X_1 = \{a_{it-1}\}, X_2 = \{1, k_{it}\}, Z_1 = \{\Delta c_{it}, \Delta \ell_{it}\}.$$

Le programme d'estimation est le suivant :

1. `proc iml;`
2. `ny={&y};`
3. `nx1={&x1}; p1=ncol(nx1);`
4. `nx={&x1 &x2};`

```

5. nz={&z1 &x2};
6. nz1={&z1};
7. use &table; read all var (ny) into y;
   read all var (nx) into x;
8.         read all var (nz) into z;
   close &table;
9. nt=nrow(y); ent=j(nt,1,1);
10. x=ent||x; z=ent||z; nx="Constante"||nx; nz="Constante"||nz;
11. start ols(y,x); b=ginv(x)*y; return (b);
   finish ols;
12. btot=ols(y,x); u=y-x*btot; ssu=ssq(u); p=ncol(x);
13. vu=ssu/(nt-p); su=sqrt(vu); print "Ecart-type du
   résidu (MCO) =" su;
14. vtot=vu#inv(t(x)*x); sbtot=sqrt(vecdiag(vtot));
   tbtot=abs(btot)/sbtot;
15. pbtot=1-probchi(tbtot##2,1);
16. print "Estimation par les MCO", "Variable
   expliquée =" ny,, btot[rowname=nx] sbtot tbtot pbtot;

17. izz=inv(t(z)*z); pzx=z*izz*t(z)*x; biv=ols(y,pxz);
18. uiv=y-x*biv; ssuiv=ssq(uiv);
19. vuiv=ssuiv/(nt-p); suiv=sqrt(vuiv); print
   "Ecart-type du résidu (VI) =" suiv;
20. viv=vuiv#inv(t(pzx)*pxz); siv=sqrt(vecdiag(viv));
   tiv=abs(biv)/siv;
21. piv=1-probchi(tiv##2,1);
22. print "Estimation par les variables instrumentales",
   "Variable expliquée =" ny, "Instruments =" nz,, biv[rowname=nx]
   siv tiv piv;
23. rg=2:(1+p1);
24. btot1=btot[rg]; vtot1=vtot[rg,rg];
25. biv1=biv[rg]; viv1=(vu/vuiv)#viv[rg,rg];
26. hw=t(biv1-btot1)*ginv(viv1-vtot1)*(biv1-btot1);
27. phw=1-probchi(hw,p1);

```

```
28. print "Statistique de Hausman-Wu =" hw, "p-value ="
    phw;
29. quit;
```

et son commentaire :

1. Début de la procédure IML;
2. `ny` contient le nom de la variable expliquée;
3. `nx1` contient les noms des variables explicatives dont on veut tester l'endogénéité, et `p1` le nombre de variables correspondant;
4. `nx` contient les noms des variables explicatives, qu'elles soient potentiellement corrélées avec l'effet individuel (`nx1`) ou non (`nx2`);
5. `nz` contient les noms des variables instrumentales, qu'elles fassent partie de l'équation (`nx2`) ou non (`nz1`);
6. `nz1` contient les noms des variables instrumentales externes à l'équation;
7. Lecture des valeurs de la variable expliquée (`ny`) et des variables explicatives (`nx`);
8. Lecture des valeurs des variables instrumentales (`nz`);
9. Le nombre d'observations est stocké dans la variable IML `nt`, `ent` contient le terme constant du modèle;
10. Un terme constant est ajouté aux variables explicatives et aux variables instrumentales;
11. Fonction qui calcule l'estimateur des moindres carrés ordinaires;
12. L'estimateur des MCO est stocké dans le vecteur IML `btot`, on calcule le résidu, stocké dans `u`, la somme des carrés des résidus, stockée dans `ssu` et le nombre de variables explicatives (avec le terme constant), qui est stocké dans `p`;
13. Calcul de la variance (`vu`) et de l'écart-type (`su`) du résidu; impression de l'écart-type du résidu;
14. Calcul de l'estimation de la variance de la perturbation déduite de l'estimateur des MCO (`vtot`), estimations des écarts-types de l'estimateur des MCO (`sbtot`) et des t de Student correspondants (`tbtot`);
15. Calcul des probabilités critiques associées aux estimations des paramètres par la méthode des MCO;
16. Impression des résultats de l'estimation par les MCO;

17. Calcul des matrices $(Z'Z)^{-1}$, P_Z et de l'estimateur par variables instrumentales (b_{iv});
18. Calcul du résidu de l'estimation par les variables instrumentales (u_{iv}) et de la somme des carrés des résidus correspondante (ss_{uiv});
19. Calcul de la variance (v_{uiv}) et de l'écart-type (s_{uiv}) du résidu; impression de l'écart-type du résidu;
20. Calcul de l'estimation de la variance de la perturbation déduite de l'estimateur par variables instrumentales (v_{iv}), estimations des écarts-types de l'estimateur par variables instrumentales (s_{iv}) et des t de Student correspondants (t_{iv});
21. Calcul des probabilités critiques associées aux estimations des paramètres par la méthode des variables instrumentales;
22. Impression des résultats de l'estimation par les variables instrumentales;
23. rg contient la position des variables explicatives potentiellement endogènes (i.e., instrumentées);
24. $btot1$ contient l'estimation par les MCO des coefficients des variables potentiellement endogènes et $vtot1$ l'estimation du bloc correspondant de la matrice de covariance;
25. $biv1$ contient l'estimation par les variables instrumentales des variables potentiellement endogènes et $viv1$ le bloc correspondant de la matrice de covariance, à laquelle on applique une correction pour que l'écart type de référence soit celui valable sous l'hypothèse nulle;
26. Calcul de la statistique de Wu-Hausman (hw);
27. Calcul de la probabilité critique associée à la statistique de Wu-Hausman;
28. Impression des résultats du test;
29. sortie d'IML.

Chapitre 13

Le modèle non sphérique

13.1 Motivation

Cette fois-ci les variables explicatives sont corrélées à une perturbation qui n'est pas homoscédastique. On a :

$$y = Xb + u,$$

$$\text{Plim} \frac{X'u}{n} \neq 0, \quad \text{Plim} \frac{Z'u}{n} = 0, \quad E(u|X, Z) = 0 \quad \text{et} \quad V(u|X, Z) = \Omega.$$

La perturbation u peut donc maintenant provenir d'un modèle à erreur composée du type, $u_{it} = \alpha_i + \varepsilon_{it}$, ce qui implique :

$$\Omega = \sigma_\varepsilon^2 \mathbf{1}_{NT} + T\sigma_\alpha^2 B.$$

La solution à ce problème d'estimation consiste à régler d'abord le problème d'instrumentation, puis le problème de variance. Prémultiplions ce modèle par Z , on obtient le modèle auxiliaire :

$$Z'y = Z'Xb + Z'u.$$

Il n'y a donc plus de corrélation asymptotique entre les variables explicatives et la perturbation, mais elle est encore hétéroscédastique. Pour trouver la bonne transformation à effectuer, regardons la matrice de covariance de la perturbation du modèle auxiliaire :

$$V(Z'u|Z, X) = Z'\Omega Z \triangleq W$$

Ce qui amène à poser l'hypothèse :

$$\text{Plim} \frac{W}{NT} = \Omega_W \text{ inversible}$$

Cette matrice est définie positive. On peut donc transformer le modèle par $(Z'\Omega Z)^{-1/2}$. Ce qui nous donne le modèle transformé :

$$y^* = X^* b + u^*,$$

avec :

$$\begin{aligned} y^* &= (Z'\Omega Z)^{-1/2} Z' y, & X^* &= (Z'\Omega Z)^{-1/2} Z' X, \\ u^* &= (Z'\Omega Z)^{-1/2} Z' u. \end{aligned}$$

La perturbation de ce modèle vérifie :

$$\begin{aligned} V(u^* | Z, X) &= (Z'\Omega Z)^{-1/2} V(Z' u | Z, X) (Z'\Omega Z)^{-1/2} \\ &= (Z'\Omega Z)^{-1/2} Z' \Omega Z (Z'\Omega Z)^{-1/2} \\ &= I_{NT}, \end{aligned}$$

donc l'application des moindres carrés ordinaires à ce modèle va bien minimiser la somme des carrés des résidus. On obtient ainsi l'estimateur optimal par variables instrumentales. Soit :

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

Ses propriétés se déduisent exactement de la même façon que pour l'estimateur par variables instrumentales. Il est convergent et asymptotiquement normal. On a en effet :

$$\begin{aligned} \tilde{b} - b &= [X' Z W^{-1} Z' X]^{-1} X' Z W^{-1} Z' u \\ \Rightarrow \text{Plim} \tilde{b} - b &= [\Omega'_{ZX} \Omega_W^{-1} \Omega_{ZX}]^{-1} \Omega'_{ZX} \Omega_W^{-1} \text{Plim} \frac{Z' u}{NT} = 0 \end{aligned}$$

La variance asymptotique est donc donnée par :

$$\text{Vas} \left[\sqrt{NT} (\tilde{b} - b) \right] = [\Omega'_{ZX} \Omega_W^{-1} \Omega_{ZX}]^{-1}$$

en utilisant $V(u) = \Omega$ et en effectuant deux simplifications. Dans la pratique, on estime la variance de cet estimateur par la quantité :

$$\widehat{\text{Vas}}(\tilde{b}) = [X' Z \widehat{W}^{-1} Z' X]^{-1},$$

l'expression de ce premier estimateur pourrait laisser supposer que l'on connaisse Ω , or ce n'est pas nécessaire. Nous allons voir qu'il est possible d'estimer directement $W = Z' \Omega Z$, de sorte que l'on peut effectivement calculer cet estimateur optimal sans faire d'hypothèse trop forte sur la matrice de covariance des perturbations.

13.2 L'estimateur des moments généralisés

L'estimateur optimal par variables instrumentales ne peut être calculé que lorsque l'on connaît l'expression de la matrice de covariance des perturbations Ω , ou quand on dispose d'un estimateur convergent de cette matrice. En fait, comme pour le modèle linéaire, il est possible que l'on ne connaisse pas Ω . Dans ce cas il faut d'une part, corriger les écarts-types de l'estimateur par variables instrumentales et d'autre part, on peut définir un meilleur estimateur, analogue à l'estimateur des moments généralisés (White, 1982 ; Hansen, 1982).

13.2.1 Covariance robuste de l'estimateur à variables instrumentales

Notre modèle est le même que dans la section précédente et l'estimateur par variables instrumentales n'est donc plus optimal. Mais il reste convergent car $\text{Plim } Z'u/NT = 0$. Sa variance asymptotique est toutefois différente. On a :

$$\begin{aligned} V(\hat{b}_V) &= [X' P_Z X]^{-1} X' P_Z V(u|X, Z) P_Z X [X' P_Z X]^{-1} \\ &= [X' P_Z X]^{-1} X' P_Z \Omega P_Z X [X' P_Z X]^{-1} \\ &= [X' P_Z X]^{-1} X' Z (Z' Z)^{-1} \underbrace{Z' \Omega Z}_W (Z' Z)^{-1} Z' X [X' P_Z X]^{-1} \end{aligned}$$

On en déduit que

$$\begin{aligned} \text{Plim} V \left[\sqrt{NT} (\hat{b}_V - b) \right] &= \text{Plim} \left[\frac{X' P_Z X}{NT} \right]^{-1} \frac{X' Z}{NT} \left(\frac{Z' Z}{NT} \right)^{-1} \frac{W}{NT} \\ &\quad \left(\frac{Z' Z}{NT} \right)^{-1} \frac{Z' X}{NT} \left[\frac{X' P_Z X}{NT} \right]^{-1} \\ &= [\Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_{ZX}]^{-1} \Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_W \Omega_{ZZ}^{-1} \Omega_{ZX} \\ &\quad [\Omega'_{ZX} \Omega_{ZZ}^{-1} \Omega_{ZX}]^{-1} \end{aligned}$$

Pour estimer cette quantité, il faut un estimateur convergent de Ω_W , car toutes les autres matrices des produits croisés peuvent être estimées directement à partir des moments empiriques qui leur correspondent.

Nous considérons à partir de maintenant le cas du modèle à effet individuel. Le modèle dont la structure de covariance est la plus générale autorise une autocorrélation au sein de la série temporelle de chaque individu :

$$\Omega = \text{diag}(\Omega_i) = \begin{pmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \Omega_N \end{pmatrix}.$$

On a donc :

$$\begin{aligned} Z' \Omega Z &= Z' \text{diag}(\Omega_i) Z \\ &= \sum_{i=1}^N Z_i' \Omega_i Z_i \\ &= \sum_{i=1}^N Z_i' \underset{(T,T)}{\mathbb{E}(u_i u_i')} Z_i \underset{(T,m)}{} \end{aligned}$$

On estime donc W par :

$$\widehat{W} = \widehat{Z' \Omega Z} = \sum_{i=1}^N Z_i' \hat{u}_{Vi} \hat{u}'_{Vi} Z_i,$$

avec :

$$\hat{u}_{Vi} = y_i - X_i \hat{b}_V.$$

Cette matrice permet d'estimer la matrice de covariance robuste de l'estimateur par variables instrumentales. Elle est robuste à l'hétéroscédasticité et à l'autocorrélation individuelles :

$$\widehat{\text{Vas}}(\tilde{b}) = [X' Z \widehat{W}^{-1} Z' X]^{-1}.$$

13.2.2 L'estimateur des moments généralisés

L'estimateur de W permet de faire beaucoup mieux que d'estimer la covariance robuste de l'estimateur par variables instrumentales. Il permet d'obtenir l'estimateur optimal du modèle non sphérique, sans connaître la forme de Ω . L'estimateur obtenu par cette méthode est appelé estimateur des moments généralisés. On réalise l'estimation en deux étapes, au minimum. Dans la pratique, on ajoute une troisième étape.

- *Première étape.* On estime le modèle par les variables instrumentales, \hat{b}_V , ce qui permet d'estimer :

$$\widehat{W} = \sum_{i=1}^N Z_i' \hat{u}_{Vi} \hat{u}_{Vi}' Z_i,$$

avec $\hat{u}_{Vi} = y_i - X_i \hat{b}_V$.

- *Seconde étape.* On utilise l'estimateur \widehat{W} pour obtenir l'estimateur optimal par variables instrumentales \tilde{b}^* . On obtient :

$$\tilde{b}^* = [X' Z \widehat{W}^{-1} Z' X]^{-1} X' Z \widehat{W}^{-1} Z' y,$$

et l'on utilise les résidus de cette estimation, $\tilde{u}_{Vi}^* = y_i - X_i \tilde{b}^*$, pour estimer sa matrice de covariance par :

$$\widehat{\text{Vas}}(\tilde{b}^*) = [X' Z \widetilde{W}^{*-1} Z' X]^{-1}, \quad (13.1)$$

avec :

$$\widetilde{W}^* = \sum_{i=1}^N Z_i' \tilde{u}_{Vi}^* \tilde{u}_{Vi}^{*'} Z_i,$$

- *Troisième étape.* L'estimateur précédent de \tilde{b}^* repose sur \widehat{W} et présente, de ce fait, l'inconvénient de ne pas être basé sur un estimateur optimal. On refait donc une étape en utilisant \widetilde{W}^* ce qui donne l'estimateur de troisième étape :

$$\bar{b}^* = [X' Z \widetilde{W}^{*-1} Z' X]^{-1} X' Z \widetilde{W}^{*-1} Z' y,$$

et l'on utilise les résidus de cette estimation, $\bar{u}_{Vi}^* = y_i - X_i \bar{b}^*$, pour estimer sa matrice de covariance par :

$$\widehat{\text{Vas}}(\bar{b}^*) = [X' Z \overline{W}^{*-1} Z' X]^{-1},$$

avec :

$$\overline{W}^* = \sum_{i=1}^N Z_i' \bar{u}_{Vi}^* \bar{u}_{Vi}^{*'} Z_i.$$

13.3 Les tests de spécification

13.3.1 Le test de Sargan-Hansen

Le test de Sargan (1958) et Hansen (1982) permet de tester la validité partielle des instruments par un autre moyen que le test de Wu-Hausman. Dans

cette section, on considère le modèle apparemment linéaire, avec p variables explicatives et $m > p$ instruments. Si le nombre d'instruments est égal au nombre de variables explicatives, on ne peut pas faire le test. Le modèle s'écrit :

$$y = Xb + u,$$

avec

$$\begin{aligned} \text{Plim} \frac{X'u}{NT} &\neq 0, \text{ Plim} \frac{Z'X}{NT} \text{ de rang } p, \\ \text{Plim} \frac{Z'u}{NT} &= 0, \text{ Plim} \frac{Z'Z}{NT} \text{ de rang } m \end{aligned}$$

13.3.2 Les conditions d'orthogonalité

L'estimateur par variables instrumentales \hat{b}_V repose sur la condition d'orthogonalité suivante :

$$\begin{matrix} X' & P_Z & \hat{u}_V & = & \mathbf{0} \\ \text{(p,NT)} & \text{(NT,NT)} & \text{(NT,1)} & & \text{(p,1)} \end{matrix}, \quad (13.2)$$

avec

$$\hat{u}_V = y - X\hat{b}_V.$$

Les conditions d'orthogonalité empiriques précédentes imposent p contraintes sur le résidu d'estimation \hat{u}_V . Or, en toute rigueur, les conditions d'orthogonalité théoriques imposent $m > p$ contraintes sur les moments de la perturbation u :

$$\text{Plim} \frac{Z'u}{NT} = E_{(m,1)}(Z'u) = 0,$$

il doit donc y avoir un moyen de tester partiellement la validité de l'estimation, en vérifiant que la perturbation n'est pas corrélée aux m instruments. En effet, sur un plan empirique, on n'utilise que les projections des p variables explicatives sur les variables instrumentales, ce qui donne un espace de dimension $p < m$. Plus précisément :

$$\text{Plim} \frac{Z'u}{NT} = 0 \Rightarrow \text{Plim} \frac{X'P_Z u}{NT} = 0, \quad (13.3)$$

mais la réciproque peut être fautive si $m > p$. Le test est le suivant :

$$\begin{cases} H_0 : \text{Plim} Z'u/NT = 0 \\ H_a : \text{Plim} Z'u/NT \neq 0 \end{cases}$$

13.3.3 La statistique de test

On se base sur les contreparties empiriques des moments théoriques. La statistique est donc égale à :

$$S = \hat{u}'_V Z \left(\sum_{i=1}^N Z'_i \hat{u}'_{Vi} \hat{u}_{Vi} Z \right)^{-1} Z' \hat{u}_V \xrightarrow[H_0]{L} \chi^2_{m-p}$$

le nombre de degrés de liberté est égal à $m - p$ parce que \hat{u}_V est défini à partir des p contraintes (13.2). Notons que cette statistique est valable aussi bien dans les modèles sphériques que non sphériques. On remarque aussi que le test sur les instruments est partiel, de sorte que si l'hypothèse nulle du test est acceptée, on dit souvent que les instruments "ne sont pas rejetés" plutôt qu'acceptés par les données. La forme de la statistique peut varier si l'on dispose de plus de précision sur la matrice de covariance du terme d'erreur.

Chapitre 14

Le choix des instruments

Dans cette section, nous discuterons des instruments suggérés par le modèle lui-même ; on peut également ajouter des instruments supplémentaires après avoir étudié la littérature sur les déterminants des variables explicatives. Nous distinguerons le cas des modèles estimés en niveaux de celui des modèles estimés en différences. Nous ne discutons dans ce chapitre que des instruments standard ; le lecteur est informé que de nombreux articles de recherche discutent de la pertinence d'instruments plus spécifiques.

14.1 Modèles estimés en niveaux

On considère un modèle à erreur composée exprimé en niveaux :

$$y_{it} = \rho_1 y_{it-1} + \tilde{X}_{it}c + u_{it},$$

bien que α_i ne soit pas corrélé aux variables explicatives \tilde{X}_{it} , le caractère dynamique du modèle suffit à le rendre apparemment linéaire. Pour estimer ce modèle il faut utiliser des variables instrumentales. Mais comme le modèle est à erreur composée, la matrice de covariance des perturbations $u_{it} = \alpha_i + \varepsilon_{it}$ n'est pas scalaire, et il faut recourir à la méthode des moments généralisés. La variable expliquée retardée y_{it-1} est la seule qui a besoin d'être instrumentée.

Pour obtenir les bons instruments, il suffit de développer le modèle lui-même :

$$\begin{aligned}
y_{it-1} &= \rho_1 y_{it-2} + \tilde{X}_{it-1} c + u_{it-1} \\
&= \rho_1 (\rho_1 y_{it-3} + \tilde{X}_{it-2} c + u_{it-2}) + \tilde{X}_{it-1} c + u_{it-1} \\
&= \rho_1^2 y_{it-3} + (\rho_1 \tilde{X}_{it-2} + \tilde{X}_{it-1}) c + \rho_1 u_{it-2} + u_{it-1} \\
&= \rho_1^{t-1} y_{i1} + (\rho_1^{t-2} \tilde{X}_{i2} + \dots + \rho_1 \tilde{X}_{it-2} + \tilde{X}_{it-1}) c \\
&\quad + \rho_1^{t-2} u_{i2} + \dots + \rho_1 u_{it-2} + u_{it-1} \\
&= \rho_1^t y_{i0} + (\rho_1^{t-1} \tilde{X}_{i1} + \rho_1^{t-2} \tilde{X}_{i2} + \dots + \rho_1 \tilde{X}_{it-2} + \tilde{X}_{it-1}) c \\
&\quad + \rho_1^{t-1} u_{i1} + \rho_1^{t-2} u_{i2} + \dots + \rho_1 u_{it-2} + u_{it-1}
\end{aligned}$$

cette relation montre que y_{it-1} peut être instrumenté par $(\tilde{X}_{i1}, \dots, \tilde{X}_{it-1})$. Le nombre d'instruments dépend donc de la datation de la variable, élément qui est spécifique à la méthode des moments généralisés. L'examen de cette relation montre également que l'on ne peut pas instrumenter la première date de la variable. On perd donc nécessairement une unité temporelle (e.g., une année de données) quand on réalise l'instrumentation. Pour la date $t = 2$, on peut instrumenter $y_{it-1} = y_{i1}$ par \tilde{X}_{i1} , pour la date $t = 3$, on peut instrumenter $y_{it-1} = y_{i2}$ par $(\tilde{X}_{i1}, \tilde{X}_{i2})$ et pour la date $t = T$, on peut instrumenter $y_{it-1} = y_{iT-1}$ par $(\tilde{X}_{i1}, \tilde{X}_{i2}, \dots, \tilde{X}_{iT-1})$. Globalement, le nombre d'instruments en $t = 2$ est égal à p , en $t = 3$ à $2p$ et en $t = T$ à $(T-1)p$. A ces instruments, nous devons ajouter \tilde{X}_{it} qui est exogène et s'instrumente donc elle-même. On instrumente donc $(y_{it-1}, \tilde{X}_{it})$ par $(\tilde{X}_{i1}, \dots, \tilde{X}_{it-1}, \tilde{X}_{it})$. Nous avons donc un nombre d'instruments total de :

$$m = (1 + \dots + T) p = \frac{T(T+1)}{2} \times p.$$

Pour que ce nombre soit suffisant, il faut que $m \geq p + 1$. Par exemple, si $T = 4$, on dispose de :

$$\frac{4 \times 5}{2} \times p = 10 \times p \text{ instruments.}$$

Il nous reste à écrire la matrice des instruments. Pour un individu, le modèle s'écrit :

$$\begin{aligned}
\begin{pmatrix} y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} &= \rho_1 \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-1} \end{pmatrix} + \begin{pmatrix} \tilde{X}_{i2} \\ \vdots \\ \tilde{X}_{iT} \end{pmatrix} b + \begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix} \\
&= \begin{pmatrix} y_{i1} & \tilde{X}_{i2} \\ \vdots & \vdots \\ y_{iT-1} & \tilde{X}_{iT} \end{pmatrix} \begin{pmatrix} \rho_1 \\ b \end{pmatrix} + \begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix},
\end{aligned}$$

la matrice des instruments peut donc s'écrire :

$$Z_i = \begin{pmatrix} \tilde{X}_{i1} & \tilde{X}_{i2} & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \tilde{X}_{i1} & \tilde{X}_{i2} & \tilde{X}_{i3} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \tilde{X}_{i1} & \tilde{X}_{i2} & \cdots & \tilde{X}_{iT} \end{pmatrix}$$

Cet instrument vérifie la condition d'orthogonalité :

$$E(Z_i' u_i) = 0 \text{ avec } u_i = \begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}.$$

Concrètement, cela donne :

$$E \left(\underbrace{\begin{pmatrix} \tilde{X}'_{i1} & 0 & \cdots & 0 \\ \tilde{X}'_{i2} & 0 & \cdots & 0 \\ 0' & \tilde{X}'_{i1} & \cdots & 0 \\ 0' & \tilde{X}'_{i2} & \cdots & 0 \\ 0' & \tilde{X}'_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \cdots & \tilde{X}'_{i1} \\ 0' & 0' & \cdots & \tilde{X}'_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \cdots & \tilde{X}'_{iT-1} \end{pmatrix}}_{(m, T-1)} \underbrace{\begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}}_{(T-1, 1)} \right) = 0,$$

ce qui est équivalent à :

$$\begin{aligned} E(\tilde{X}'_{i1} u_{i2}) &= 0, E(\tilde{X}'_{i2} u_{i2}) = 0, \\ E(\tilde{X}'_{i1} u_{i3}) &= 0, E(\tilde{X}'_{i2} u_{i3}) = 0, E(\tilde{X}'_{i3} u_{i3}) = 0, \\ &\vdots \\ E(\tilde{X}'_{i1} u_{iT}) &= 0, E(\tilde{X}'_{i2} u_{iT}) = 0, \dots, E(\tilde{X}'_{iT} u_{iT}) = 0. \end{aligned}$$

Pour estimer le modèle, on utilise simplement :

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_N \end{pmatrix}.$$

Il est clair que toute la difficulté de ce type d'estimation réside dans la capacité à programmer les matrices d'instruments sans faire d'erreur.

14.2 Modèles estimés en différences

Cette section s'inspire de l'approche d'Arellano et Bondt (1991). Dans un premier temps, nous estimerons un modèle statique puis, dans un second temps, un modèle autorégressif. Le problème qui est posé est de trouver tous les instruments disponibles pour réaliser l'estimation car c'est cette condition qui permet d'obtenir l'estimateur le plus efficace.

Considérons le modèle statique défini par :

$$y_{it} = X_{it}b + \alpha_i + \varepsilon_{it},$$

où la perturbation α_i est potentiellement corrélée avec les variables explicatives X_{it} . Le terme ε_{it} est un bruit blanc indépendant de α_i et des variables explicatives. Appliquée à ce modèle, la méthode des moindres carrés ordinaires ne permet pas d'obtenir un estimateur convergent de b en raison de la présence de α_i dans l'équation. Une manière simple d'éliminer cet effet consiste à écrire le modèle en différences :

$$\Delta y_{it} = \Delta X_{it}b + \Delta \varepsilon_{it},$$

où il n'y a plus maintenant de corrélation entre les variables explicatives et le terme d'erreur. De fait, ce modèle peut être estimé par les moindres carrés ordinaires et ne nécessite pas d'appliquer des méthodes à variables instrumentales sous certaines hypothèses. L'hypothèse qui garantit la convergence de l'estimateur des MCO et celle d'*exogénéité forte* des perturbations, selon laquelle les valeurs de ε_{it} à toutes les dates sont sans corrélation avec les valeurs de X_{it} à toutes les dates. On résume cette propriété par :

$$E(X_{it}\varepsilon_{is}) = 0, \forall (t, s).$$

Dans ce cas l'estimateur des MCO pourrait s'écrire :

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

Ici il faut toutefois faire attention à l'écriture de la matrice de covariance car elle n'est jamais scalaire. En effet, on a :

$$V(\Delta \varepsilon_{it}) = V(\varepsilon_{it} - \varepsilon_{it-1}) = V(\varepsilon_{it}) + V(\varepsilon_{it-1}) = 2\sigma_\varepsilon^2,$$

et

$$\begin{aligned} \text{Cov}(\Delta \varepsilon_{it}, \Delta \varepsilon_{it-1}) &= \text{Cov}(\varepsilon_{it} - \varepsilon_{it-1}, \varepsilon_{it-1} - \varepsilon_{it-2}) \\ &= -\text{Cov}(\varepsilon_{it-1}, \varepsilon_{it-1}) \\ &= -\sigma_\varepsilon^2, \end{aligned}$$

et les autres covariances sont nulles. La matrice de covariance pour un individu est donc égale à :

$$V(\Delta\varepsilon) = \sigma_\varepsilon^2 \begin{pmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{pmatrix} = \sigma_\varepsilon^2 DD',$$

où D est la matrice de différenciation. En conséquence :

$$V(\hat{b}) = \sigma_\varepsilon^2 (X'D'DX)^{-1} X'(D'D)^2 X(X'D'DX)^{-1},$$

qui ne se simplifie pas.

Chapitre 15

Le modèle VAR

Une extension immédiate du modèle autorégressif est le modèle vectoriel autorégressif. On considère ici la première équation d'un seul système, mais on peut inverser les notations des variables pour obtenir l'estimation de la seconde équation.

15.1 Cas général

On cherche à estimer une relation dynamique entre deux variables y et \tilde{x} . Pour cela on dispose de données de panel, que l'on supposera cylindrées.¹ Le modèle comporte un effet individuel α_i et un effet temporel ψ_t multiplicatif, tous deux potentiellement corrélés avec les variables explicatives. On considère l'estimation de l'équation suivante, tirée du système VAR :

$$\begin{aligned} y_{it} &= \sum_{\ell=1}^m \rho_{\ell} y_{it-\ell} + \sum_{\ell=1}^m c_{\ell} \tilde{x}_{it-\ell} + u_{it}, \\ i &= 1, \dots, N, t = (m+1), \dots, T \end{aligned} \quad (15.1)$$

avec

$$u_{it} = \psi_t \alpha_i + \varepsilon_{it},$$

où α_i est un effet individuel corrélé et ε_{it} un bruit blanc. On retrouve le modèle à effet individuel standard lorsque $\psi_t = 1, \forall t$. Comme il est possible d'estimer un modèle plus général, nous le faisons, afin d'éviter d'obtenir un estimateur non convergent en cas de violation de cette hypothèse. Nous estimerons ce modèle en suivant les méthodes de Chamberlain (1983), et de Holtz-Eakin, Newey et Rosen (1988). L'estimation de ce modèle se fait en deux étapes :

1. On peut estimer également certains modèles non cylindrés en étendant légèrement la méthode présentée ici (Duguet et Fremigacci, 2009).

1. On élimine l'effet individuel grâce à une quasi-différence, puis on estime la forme réduite (voir plus bas) par la méthode des moments généralisée ;
2. On remonte de la forme réduite du modèle à sa forme structurelle (15.1) par la méthode des moindres carrés asymptotiques (Gouriéroux, Monfort et Trognon, 1985).

15.1.1 Ecriture et estimation de la forme réduite

Pour éliminer l'effet individuel de la forme structurelle (15.1), nous lui appliquons la quasi-différence de Chamberlain (1983). Ceci consiste à multiplier l'équation prise en $t - 1$ par $r_t = \psi_t / \psi_{t-1}$ puis à la soustraire à l'équation t . On obtient :

$$y_{it} - \frac{\psi_t}{\psi_{t-1}} y_{it-1} = \sum_{\ell=1}^m \rho_\ell y_{it-\ell} + \sum_{\ell=1}^m c_\ell \tilde{x}_{it-\ell} + \psi_t \alpha_i + \varepsilon_{it} \\ - \frac{\psi_t}{\psi_{t-1}} \left\{ \sum_{\ell=1}^m \rho_\ell y_{it-\ell-1} + \sum_{\ell=1}^m c_\ell \tilde{x}_{it-\ell-1} + \psi_{t-1} \alpha_i + \varepsilon_{it-1} \right\}$$

ce qui peut se réécrire :

$$y_{it} = \sum_{\ell=1}^{m+1} \pi_{\ell,t}^y y_{it-\ell} + \sum_{\ell=1}^{m+1} \pi_{\ell,t}^x \tilde{x}_{it-\ell} + v_{it}, \quad (15.2) \\ i = 1, \dots, N, \quad t = (m+3), \dots, T$$

avec :

$$\begin{aligned} \pi_{1,t}^y &= r_t + \rho_1, \\ \pi_{\ell,t}^y &= \rho_\ell - r_t \rho_{\ell-1}, \\ \pi_{(m+1),t}^y &= -r_t \rho_m, \\ \pi_{1,t}^x &= c_1, \\ \pi_{\ell,t}^x &= c_\ell - r_t c_{\ell-1}, \\ \pi_{(m+1),t}^x &= -r_t c_m, \\ v_{it} &= \varepsilon_{it} - r_t \varepsilon_{it-1}. \end{aligned} \quad (15.3)$$

Nous allons d'abord estimer les paramètres de la forme réduite :

$$\pi_t = \left(\pi_{1,t}^y, \dots, \pi_{(m+1),t}^y, \pi_{1,t}^x, \dots, \pi_{(m+1),t}^x \right)'$$

avant de revenir à la forme structurelle

$$b = (\rho_1, \dots, \rho_\ell, c_1, \dots, c_\ell)' \text{ et } r = (r_{T-(m+1)}, \dots, r_T)'$$

Ce système doit être estimé par les moments généralisés en raison de variables explicatives endogènes. On applique donc la méthode de Chamberlain (voir Crépon et Mairesse, 1996). Pour procéder à l'estimation, on réécrit l'équation (15.2) en empilant tous les individus à la même date. Soit :

$$\begin{aligned} Y_t &= \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \quad \tilde{X}_t = \begin{pmatrix} \tilde{x}_{1t} \\ \vdots \\ \tilde{x}_{Nt} \end{pmatrix}, \\ X_t &= (Y_{t-1}, \dots, Y_{t-(m+1)}, \tilde{X}_{t-1}, \dots, \tilde{X}_{t-(m+1)}) \end{aligned}$$

où Y_t est le vecteur de la variable expliquée de la première équation pour tous les individus à la date t , et X_t la matrice correspondante des variables explicatives (endogènes et exogènes). Le vecteur des termes d'erreur est égal à :

$$V_t = \begin{pmatrix} v_{1t} \\ \vdots \\ v_{Nt} \end{pmatrix},$$

et le vecteur des coefficients de la forme réduite en t est donné par :

$$\pi_t = (\pi_{1,t}^y, \dots, \pi_{(m+1),t}^y, \pi_{1,t}^x, \dots, \pi_{(m+1),t}^x)'$$

L'équation (15.2) peut donc se réécrire :

$$Y_t = X_t \pi_t + V_t, \quad t = (m+3), \dots, T$$

ce qui suggère d'empiler toutes les dates disponibles :

$$\begin{aligned} Y &= \begin{pmatrix} Y_{m+3} \\ \vdots \\ Y_T \end{pmatrix}, \quad V = \begin{pmatrix} V_{m+3} \\ \vdots \\ V_T \end{pmatrix}, \\ \pi &= \begin{pmatrix} \pi_{m+3} \\ \vdots \\ \pi_T \end{pmatrix}, \end{aligned}$$

et

$$X = \text{diag}(X_{m+3}, \dots, X_T),$$

de sorte que :

$$Y = X\pi + V. \quad (15.4)$$

Sous hypothèse d'exogénéité faible, la matrice des variables instrumentales à la date t est donnée par :

$$\underset{(N, 2(t-2))}{Z_t} = (Y_{t-2}, \dots, Y_1, X_{t-2}, \dots, X_1),$$

et l'on remarque que le nombre d'instruments disponibles varie à chaque date, car plus une variable est éloignée de la date de départ plus il y a de retards disponibles pour l'instrumenter. La matrice globale des instruments est définie par :

$$Z = \text{diag}(Z_{m+3}, \dots, Z_T).$$

Afin d'obtenir l'estimateur, on prémultiplie l'équation (15.4) par Z' , ce qui donne :

$$Z'Y = Z'X\pi + Z'V,$$

comme $\text{Plim} Z'V/N = 0$ nous pouvons appliquer les moindres carrés généralisés à ces équations, ce qui revient à appliquer les moindres carrés asymptotiques. Pour estimer la covariance de $Z'V$ il nous faut un estimateur de première étape ; nous prenons simplement l'estimateur par variables instrumentales, pris année par année :

$$\tilde{\pi}_t = \left[X_t' Z_t (Z_t' Z_t)^{-1} Z_t' X_t \right]^{-1} X_t' Z_t (Z_t' Z_t)^{-1} Z_t' Y_t,$$

qui donne le vecteur des résidus suivant à la date t :

$$\tilde{V}_t = Y_t - X_t \tilde{\pi}_t,$$

de sorte que l'on peut estimer Ω par :

$$\tilde{\Omega}_{r,s} = \sum_{i=1}^N Z_{ir}' Z_{is} \tilde{v}_{ir} \tilde{v}_{is},$$

où \tilde{v}_{it} est le i -ème élément de \tilde{V}_t et Z_{it} la i -ème ligne de Z_t . Finalement, l'estimateur optimal du vecteur global π est obtenu par :

$$\hat{\pi} = [X' Z \tilde{\Omega}^{-1} Z' X]^{-1} X' Z \tilde{\Omega}^{-1} Z' Y.$$

15.1.2 Estimation de la forme structurelle

Ici, nous traitons un cas un peu plus général que celui discuté dans Holtz-Eakin, Newey et Rosen (1988), qui posent la contrainte $r_t = 1, \forall t$. Ceci implique que cette étape sera non linéaire, au lieu d'être linéaire. Pour amorcer l'algorithme d'optimisation, il nous faut un estimateur convergent simple à calculer.

Il en existe plusieurs, nous en proposons un. La première série de contraintes identifiantes est la suivante :

$$\pi_{1,t}^y = r_t + \rho_1, \quad t = m+3, \dots, T \quad (15.5)$$

ce qui implique :

$$\pi_{1,t}^y - \pi_{1,t-1}^y = r_t - r_{t-1}, \quad (15.6)$$

considérons un second ensemble de contraintes :

$$\pi_{\ell,t}^y = \rho_{\ell} - r_t \rho_{\ell-1},$$

ceci implique :

$$\pi_{\ell,t}^y - \pi_{\ell,t-1}^y = -(r_t - r_{t-1}) \rho_{\ell-1},$$

en additionnant de chaque côté et en utilisant (15.6), on obtient :

$$\begin{aligned} \sum_t (\pi_{\ell,t}^y - \pi_{\ell,t-1}^y) &= - \sum_t (r_t - r_{t-1}) \rho_{\ell-1} \\ &= - \rho_{\ell-1} \sum_t (\pi_{1,t}^y - \pi_{1,t-1}^y) \end{aligned}$$

de sorte que :

$$\rho_{\ell-1} = - \frac{\sum_t (\pi_{\ell,t}^y - \pi_{\ell,t-1}^y)}{\sum_t (\pi_{1,t}^y - \pi_{1,t-1}^y)}, \quad \ell = 2, \dots, m$$

ce qui suggère l'estimateur convergent suivant :

$$\hat{\rho}_{\ell-1} = - \frac{\sum_t (\hat{\pi}_{\ell,t}^y - \hat{\pi}_{\ell,t-1}^y)}{\sum_t (\hat{\pi}_{1,t}^y - \hat{\pi}_{1,t-1}^y)}, \quad \ell = 2, \dots, m$$

qui implique que l'on peut estimer ρ_1 par :

$$\hat{\rho}_1 = - \frac{\sum_t (\hat{\pi}_{2,t}^y - \hat{\pi}_{2,t-1}^y)}{\sum_t (\hat{\pi}_{1,t}^y - \hat{\pi}_{1,t-1}^y)},$$

de sorte que l'on peut estimer les coefficients r_t par :

$$\hat{r}_t = \hat{\pi}_{1,t}^y - \hat{\rho}_1,$$

en utilisant (15.5). Pour le coefficient du dernier retard, il suffit d'additionner les contraintes suivantes :

$$\begin{aligned}\pi_{(m+1),t}^y &= -r_t \rho_m \Rightarrow \sum_t \pi_{(m+1),t}^y = -\sum_t r_t \rho_m \\ \Leftrightarrow \rho_m &= -\frac{\sum_t \pi_{(m+1),t}^y}{\sum_t r_t},\end{aligned}$$

d'où l'estimateur :

$$\hat{\rho}_m = -\frac{\sum_t \hat{\pi}_{(m+1),t}^y}{\sum_t \hat{r}_t}.$$

Pour les coefficients de c , on obtient les estimateurs par la même méthode :

$$\begin{aligned}\hat{c}_1 &= \frac{1}{T-(m+2)} \sum_{t=m+3}^T \hat{\pi}_{1t}^x, \\ \hat{c}_{\ell-1} &= -\frac{\sum_t (\hat{\pi}_{\ell,t}^x - \hat{\pi}_{\ell,t-1}^x)}{\sum_t (\hat{\pi}_{1,t}^x - \hat{\pi}_{1,t-1}^x)}, \ell = 3, \dots, m, \\ \hat{c}_m &= -\frac{\sum_t \hat{\pi}_{(m+1),t}^x}{\sum_t \hat{r}_t}.\end{aligned}$$

Cet estimateur est convergent et nous l'utilisons comme valeur initiale d'un algorithme de Newton-Raphson. Afin de procéder à l'estimation optimale du modèle, nous devons introduire les notations suivantes, adaptées aux moindres carrés asymptotiques :

- π est le paramètre auxiliaire. Dans ce chapitre, il s'agit du paramètre de la forme réduite du modèle ; il est défini par :

$$\pi = \begin{pmatrix} \pi^y \\ \pi^x \end{pmatrix}$$

- θ est le paramètre d'intérêt. Dans ce chapitre, il s'agit du paramètre de la forme structurelle du modèle ; il est défini par :

$$\theta = \begin{pmatrix} \rho \\ c \\ r \end{pmatrix}$$

- Afin d'estimer θ par les moindres carrés asymptotiques, nous devons remplir deux conditions. D'une part, avoir un estimateur convergent et asymptotiquement normal de π (noté $\hat{\pi}$) et, d'autre part, qu'il existe une relation différentiable entre π et θ . Ces deux conditions sont remplies car

$\hat{\pi}$ est obtenu par les moments généralisés "classiques" et les contraintes (15.3) sont différentiables. Dans notre cas, on peut même les écrire sous la forme simplifiée :

$$\Psi = \pi - g(\theta) = 0,$$

de sorte que $\partial\Psi/\partial\pi'$ est égale à la matrice identité, ce qui simplifie toutes les expressions usuelles. Les moindres carrés asymptotiques consistent alors à résoudre le programme non linéaire suivant :

$$\min_{\theta} \Psi' S \Psi,$$

où S est une matrice de pondération. La matrice optimale est donnée par :

$$S = \text{Vas}(\hat{\pi})^{-1},$$

de sorte que nous devons résoudre :

$$\hat{\theta} = \arg \min_{\theta} (\hat{\pi} - g(\theta))' \text{Vas}(\hat{\pi})^{-1} (\hat{\pi} - g(\theta)),$$

l'estimateur de la forme structurelle $\hat{\theta}$ est convergent et asymptotiquement normal, et l'on estime sa matrice de covariance par la statistique suivante :

$$\widehat{\text{Vas}}(\hat{\theta}) = \left[\frac{\partial g}{\partial \theta'}(\hat{\pi}, g(\hat{\theta})) \text{Vas}(\hat{\pi})^{-1} \frac{\partial g}{\partial \theta}(\hat{\pi}, g(\hat{\theta})) \right]^{-1}.$$