

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail
de la Direction des Etudes et Synthèses Économiques*

G 9914

**Macro-commandes SAS
pour l'économétrie des panels
et des variables qualitatives**

Emmanuel DUGUET *

Septembre 1999

* EUREQua - Université de Paris I Panthéon-Sorbonne et INSEE - Département des Etudes Economiques d'Ensemble - Division « Marchés et Stratégies d'Entreprise » Timbre G231 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX
E-mail : D3E@insee.fr.

Département des Etudes Economiques d'Ensemble - Timbre G201 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF
CEDEX - France - Tél. : 33 (1) 41 17 60 68 - Fax : 33 (1) 41 17 60 45

*Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their author's views.*

Résumé

Ce document de travail présente un ensemble de programmes d'estimation facilement utilisables sous SAS. L'essentiel des programmes concerne l'estimation sur données de panel. Il comprend les méthodes traditionnelles (between, within, mcqg) et quelques tests de spécification. Ce premier ensemble est prolongé par la méthode des moments généralisés aussi bien sur un modèle statique que sur un modèle dynamique, avec test de Sargan. Un dernier programme vient compléter les macro-commandes déjà disponibles pour les variables qualitatives. Il permet d'estimer le modèle logit multinomial, un modèle de choix discret qui est adapté aux variables non ordonnées.

Mots-clés : données de panel, between, within, moments généralisés, logit multinomial, modèles de choix discret.

Abstract

This working paper presents a package of econometric programs that can be used easily with SAS. The main part of the programs deals with panel data. It includes the more widespread methods (between, within, feasible GLS) with some specification tests. This first package is then extended to the generalized method of moments both in the static and the dynamic cases and includes the Sargan test. The last program complements the already available package for qualitative and limited dependent variables. It provides estimates of the multinomial logit model, a discrete choice model for non ordered alternatives.

Keywords: panel data, between, within, generalized method of moments, multinomial logit, discrete choice models.

Classification JEL : C23, C25, C87

TABLE DES MATIERES

1. INTRODUCTION	5
<hr/>	
2. ESTIMATION SUR DONNEES DE PANEL DANS LE CAS STANDARD	7
<hr/>	
2.1 LE MODELE A EFFET INDIVIDUEL	7
2.2 ESTIMATION	7
2.2.1 REGRESSION SUR LES MOYENNES INDIVIDUELLES OU « BETWEEN »	8
2.2.2 REGRESSION SUR LES ECARTS AUX MOYENNES INDIVIDUELLES OU « WITHIN »	9
2.2.3 TESTS DE SPECIFICATION	9
2.2.4 MOINDRES CARRES QUASI-GENERALISES	10
2.3 PRESENTATION DES MACRO-COMMANDES	11
2.4 MACRO-COMMANDE %MCO ROB(TAB,Y,X); RUN;	11
2.5 MACRO-COMMANDE %BETWEEN(TAB,Y,X,I); RUN;	12
2.6 MACRO-COMMANDE %BW WITHIN(TAB,Y,X,I,T); RUN;	13
2.7 MACRO-COMMANDE %BWM CQG(TAB,Y,X,I,T); RUN;	13
2.8 EXEMPLE D'UTILISATION	15
3. ESTIMATION PAR LA METHODE DES MOMENTS GENERALISES	22
<hr/>	
3.1 PRESENTATION DU MODELE	22
3.2 MACRO-COMMANDE %MOMDYN	24
3.3 EXEMPLE D'UTILISATION	24
3.4 MACRO-COMMANDE %MOMSTA	27
4. LE MODELE LOGIT POLYTOMIQUE OU « MULTINOMIAL LOGIT »	29
<hr/>	
4.1 PRESENTATION DU MODELE	29
4.2 LA MACRO-COMMANDE %MULTINO	29
4.3 EXEMPLE D'UTILISATION	32
5. SOURCES DES MACRO-COMMANDES	37
<hr/>	
5.1 %MCO ROB	37
5.2 %BETWEEN	38
5.3 %BW WITHIN	38
5.4 %BWM CQG	41
5.5 %MOMSTA	43
5.6 %MOMDYN	46
5.7 %MULTINO	49

1. Introduction

La disponibilité de données de panel est de plus en plus fréquente. Dans le même temps, les méthodes de traitement des données se sont développées et les standards requis tant pour évaluer la pertinence des estimations effectuées que pour communiquer et publier les études empiriques se sont élevés. Or les logiciels disponibles ne fournissent pas de programme d'estimation adapté, d'où des pertes de temps assez importantes. Afin de régler ce problème, il est apparu nécessaire d'améliorer un certain nombre de programmes d'estimation écrits dans le cadre des études économétriques antérieures et de les mettre sous forme de macro-commandes SAS utilisables par tous. Le but de ce rapport est de présenter les méthodes d'estimation et les macro-commandes en question. D'autre part, ce travail vient compléter les programmes réalisés par le passé en y ajoutant le modèle logit polytomique (« multinomial logit ») et le calcul d'écarts-types robustes.

Les méthodes d'estimation sur données de panel concernent à la fois les méthodes traditionnelles et la méthode des moments généralisés dont l'utilisation est aujourd'hui standard. Dans un premier temps nous présentons donc les méthodes between, within et des moindres carrés quasi-généralisés. Toutefois, toutes ces méthodes traditionnelles sont biaisées dès lors qu'il existe une corrélation entre les variables explicatives et la perturbation. On peut alors recourir à la méthode des moments généralisés. Deux macro-commandes sont alors disponibles : une première traite le cas des modèles statiques et une seconde l'estimation des modèles dynamiques.

Enfin, à la bibliothèque déjà disponible pour l'estimation de modèles à variables qualitatives et de comptage, s'ajoute une nouvelle macro-commande qui permet de traiter les réponses résultant d'un choix exclusif entre différentes possibilités non ordonnées.

Chaque section présente dans une première partie un rappel des méthodes d'estimation puis, dans une deuxième partie, on donne les modes d'emploi des macro-commandes correspondantes et un exemple d'utilisation. La dernière section donne les programmes sources pour en faciliter l'étude.

Mise à disposition

Il est possible de se procurer l'ensemble de ces programmes par courrier électronique auprès de M^{me} Christine Deroin : christine.deroin@insee.fr. (tél. 01.41.17.60.69). Ces macro-commandes sont également disponibles sous le répertoire de la DESE « O:\echanges\d3e\dmse\Macros » qui contient l'archive « m9599.zip ». Cette archive contient à la fois le jeu de macro-commandes présenté dans ce document de travail et le jeu de macro-commandes pour les données de comptage et les variables qualitatives présenté dans un document de travail antérieur du CREST (Crépon et Duguet, 1995).

2. Estimation sur données de panel dans le cas standard

2.1 Le modèle à effet individuel

On dispose d'un panel de N individus observés à T dates consécutives. On souhaite estimer le modèle linéaire :

$$y_{it} = x_{it} \mathbf{b} + u_{it} \text{ avec } u_{it} = \alpha_i + \varepsilon_{it} \quad i = 1, \dots, N \quad t=1, \dots, T$$

où y_{it} désigne la valeur prise par la variable expliquée de l'individu i à la date t , x_{it} les variables explicatives et u_{it} la perturbation. Cette dernière comporte deux parties : un effet individuel α_i qui peut être fixe ou aléatoire, et un bruit blanc ε_{it} non corrélé avec l'effet individuel.

Habituellement les deux composantes de la variance de la perturbation sont supposées constantes. Toutefois, il est possible que la régression de départ soit hétéroscédastique. Dans ce dernier cas, il nous faut calculer des écarts-types robustes à l'hétéroscédasticité. La macro-commande effectue les deux calculs.

On écrit habituellement la relation précédente sous la forme plus compacte, regroupée par individu :

$$y_i = x_i \mathbf{b} + u_i \quad i = 1, \dots, N$$

$$\text{avec } \underset{(T,1)}{y_i'} = (y_{i1}, \dots, y_{iT}), \quad \underset{(T,p)}{x_i'} = (x_{i1}, \dots, x_{iT}) \text{ et } \underset{(T,1)}{u_i'} = (u_{i1}, \dots, u_{iT})$$

2.2 Estimation

Dans le modèle à erreurs composées, l'estimateur asymptotiquement efficace est donné par les moindres carrés quasi-généralisés. En effet la variance de la perturbation est égale à :

$$V(u_i) = (T\sigma_\alpha^2 + \sigma_\varepsilon^2)B_T + \sigma_\varepsilon^2 W_T = \Sigma \quad i = 1, \dots, N$$

où B_T représente l'opérateur « between », qui donne le vecteur de la moyenne répliquée T fois, et W_T l'opérateur « within » qui donne le vecteur des écarts des variables à leur moyenne.

Sur l'ensemble du modèle empilé, la variance peut également s'écrire de manière empilée puisque les observations sont indépendantes. De plus les variances sont identiques pour tous les individus. La variance de l'ensemble du modèle empilé, de format $NT \times NT$ est donc égale à :

$$V(\mathbf{u}) = I_N \otimes \Sigma$$

Il faut donc transformer le modèle initial par :

$$V(\mathbf{u})^{-1/2} = (I_N \otimes \Sigma)^{-1/2} = I_N \otimes \Sigma^{-1/2} = I_N \otimes \sigma_\varepsilon^{-1} (I_T - (1-\theta)B_T)$$

avec

$$\theta = \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_\alpha^2 + \sigma_\varepsilon^2}}$$

Comme la constante $\sigma_{\varepsilon}^{-1}$ multiplie les deux membres de l'équation, on peut l'omettre dans les calculs et effectuer une régression par les moindres carrés ordinaires de la quasi-différence $y_{it} - (1-\theta)\bar{y}_i$ sur les quasi-différences $x_{it} - (1-\theta)\bar{x}_i$. Il nous reste à estimer θ . Pour ce faire, on utilise les résidus des régressions « between » et « within ».

2.2.1 Régression sur les moyennes individuelles ou « between »

On peut considérer la régression avec ou sans duplication des moyennes. En effet, lorsque l'on applique la transformation « between », B_T , les moyennes individuelles sont répliquées T fois pour chaque individu. De ce fait on est amené à travailler sur $N \times T$ observations au lieu de N observations. Ceci possède deux inconvénients : les calculs sont plus longs et les degrés de liberté sont faux puisqu'ils se basent sur $N \times T$ observations au lieu de N observations.

Comme les duplications amènent à de faux degrés de liberté, et pour réduire le temps de calcul par le programme, nous utiliserons le modèle sans duplication des moyennes. En plus du cas usuel (i.e., homoscédastique), nous donnons les écarts-types robustes à l'hétéroscédasticité. Le modèle en moyennes, qui comporte N observations, s'écrit :

$$\bar{y}_i = \bar{x}_i b + u_{Bi}$$

La variance de la perturbation est alors égale à¹ :

$$V(u_{Bi}) = V(\alpha_i + \varepsilon_i) = \sigma_{\alpha}^2 + \frac{\sigma_{\varepsilon}^2}{T} \text{ et } \text{Cov}(u_{Bi}, u_{Bj}) = 0 \forall i \neq j$$

Les moindres carrés ordinaires sur le modèle en moyenne sont donc optimaux et l'on peut estimer la variance de la perturbation par celle du résidu :

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

Mais on peut également postuler que les variances des perturbations varient d'un individu à l'autre. Dans ce cas la matrice de covariance des moindres carrés ordinaires n'est plus donnée par la formule habituelle mais par :

$$V(\bar{u}) = (\bar{x}' \bar{x})^{-1} \bar{x}' \Sigma_B \bar{x} (\bar{x}' \bar{x})^{-1}$$

White (1980) propose d'estimer la matrice $\bar{x}' \Sigma_B \bar{x}$ par la méthode suivante. Puisque la matrice de covariance est diagonale du fait de l'indépendance des observations, on l'estime par :

$$\bar{x}' \text{Diag}(\hat{u}_{Bi}^2) \bar{x} = \sum_{i=1}^N \bar{x}_i' \bar{x}_i \hat{u}_{Bi}^2$$

d'où :

¹ Sur le modèle avec répliquions : $\text{TV}(\bar{u}_i) = T\sigma_{\alpha}^2 + \sigma_{\varepsilon}^2$.

$$\hat{V}(\bar{u}) = \left(\sum_{i=1}^N \bar{x}_i' \bar{x}_i \right)^{-1} \sum_{i=1}^N \bar{x}_i' \bar{x}_i \hat{u}_{Bi}^2 \left(\sum_{i=1}^N \bar{x}_i' \bar{x}_i \right)^{-1}$$

2.2.2 Régression sur les écarts aux moyennes individuelles ou « within »

Ici, on effectue les régressions sur les variables prises en en écarts à leurs moyennes. Le modèle transformé est donc :

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)b + u_{Wi}$$

Les moindres carrés ordinaires appliqués au modèle transformés restent convergents en présence d'autocorrélation des perturbations. La variance des perturbations est alors égale à² :

$$V(u_{Wi}) = V(\varepsilon_{it} - \bar{\varepsilon}_i) = \sigma_\varepsilon^2 \left(1 - \frac{1}{T} \right)$$

et

$$\text{Cov}(u_{Wit}, u_{Wis}) = \text{Cov}(\varepsilon_{it} - \bar{\varepsilon}_i, \varepsilon_{is} - \bar{\varepsilon}_i) = -\frac{\sigma_\varepsilon^2}{T}$$

La variance de la perturbation peut être estimée à partir de la variance du résidu :

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

quand il n'y a pas de terme constant dans le modèle de départ,

$$\hat{\sigma}_W^2 = \frac{\hat{u}_W' \hat{u}_W}{NT - N - p - 1} \quad \text{sinon.}$$

2.2.3 Tests de spécification

Deux tests de spécification sont utilisés pour le modèle à erreurs composées : le test de Fisher ($H_0: \sigma_\alpha = 0$) et le test d'Hausman, que l'on peut écrire comme un test d'égalité des estimateurs « between » et « within ».

Le test de Fisher teste l'absence d'effet individuel ($H_0: \sigma_\alpha = 0$) et est basé sur le ratio de la variance « between » à la variance totale. On a, sous H_0 :

$$(NT - N - p) \frac{T \hat{\sigma}_W^2}{(T - 1) \sigma_\varepsilon^2} \text{ suit un } \chi_{NT - N - p}^2$$

Donc, sous H_0 :

² Sur le modèle avec réplifications : $TV(\bar{u}_i) = T\sigma_\alpha^2 + \sigma_\varepsilon^2$.

$(N - p) \frac{T \hat{\sigma}_B^2}{\hat{\sigma}_\varepsilon^2}$ suit un χ_{N-p}^2 (car $\sigma_\alpha = 0$ et absence de duplications)

Les deux χ^2 précédents étant indépendants :

$F = \frac{(N - p)(T - 1)}{N(T - 1) - p} \times \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$ suit une loi de Fisher $F(N - p, NT - N - p)$

Remarque importante : si le modèle comporte un terme constant on utilise

$F = \frac{(N - p - 1)(T - 1)}{N(T - 1) - p - 1} \times \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$ suit une loi de Fisher $F(N - p - 1, NT - N - p - 1)$ sous H_0 .

Le test de Hausman teste l'absence de corrélation entre l'effet individuel et les variables explicatives. La statistique s'obtient en comparant directement les estimateurs « between » et « within ». Il permet ainsi de savoir si l'estimateur optimal est bien celui des moindres carrés généralisés. Par définition, il se fait sans terme constant car il est absent de la régression within, on le retire donc du vecteur des paramètres de la régression « between » également, ainsi que la ligne et la colonne correspondante de la matrice de covariance du paramètre. On a :

$$H = (\hat{b}_B - \hat{b}_W)' (\hat{V}(\hat{b}_B) + \hat{V}(\hat{b}_W))^{-1} (\hat{b}_B - \hat{b}_W)$$

La statistique de Hausman suit asymptotiquement sous H_0 une loi χ_{p-1}^2 (car il n'y a pas de terme constant). Dans la macro-commande présentée ici, on utilise les matrices de covariance robustes à l'hétéroscédasticité et à l'autocorrélation pour effectuer ce test.

2.2.4 Moindres carrés quasi-généralisés

Les deux estimateurs précédents permettent d'estimer les variances du modèle à erreurs composées. On a :

$$\sigma_W^2 = \sigma_\varepsilon^2 \left(1 - \frac{1}{T}\right) \text{ et } \sigma_B^2 = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{T} \text{ (sans réplification des moyennes)}$$

d'où :

$$\sigma_\varepsilon^2 = \frac{T}{T-1} \sigma_W^2 \text{ et } \sigma_\alpha^2 = \sigma_B^2 - \frac{\sigma_\varepsilon^2}{T}$$

Ce qui donne les estimateurs convergents des composantes de la variance :

$$\hat{\sigma}_\varepsilon^2 = \frac{T}{T-1} \hat{\sigma}_W^2 \text{ et } \hat{\sigma}_\alpha^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_\varepsilon^2}{T}$$

On en déduit :

$$\hat{\theta} = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{T \hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2}}$$

On utilise cet estimateur de \mathbf{q} pour calculer les quasi-différences nécessaires aux

moindres carrés quasi-généralisés.

2.3 Présentation des macro-commandes

L'estimation est réalisée à partir de quatre macro-commandes emboîtées. La première macro-commande %bwmcgg contient la macro-commande %bwithin, qui contient la macro-commande %between, qui contient la macro-commande %mcorob. Elles peuvent être utilisées séparément. La dernière, %mcorob, peut être utilisée sur des données en coupe.

- %mcorob : réalise l'estimation d'un modèle linéaire en coupe avec écarts-types robustes à l'hétéroscédasticité;
- %between : effectue les moyennes individuelles et réalise les estimations des moindres carrés ordinaires avec les écarts-types traditionnels et robustes (utilisation de %mcorob)
- %bwithin : lance d'abord la macro-commande %between, puis estime le modèle en écart aux moyennes individuelles avec les écarts-types traditionnels et robustes (sans utiliser %mcorob, car les variables sont autocorrélées), calcule le tableau d'analyse de la variance et réalise les tests de spécification de Fisher et de Hausman;
- %bwmcgg : lance %bwithin puis estime le modèle par les moindres carrés quasi-généralisés.

2.4 Macro-commande %MCOROB(TAB,Y,X); RUN;

En entrée : les 3 informations suivantes sont obligatoires et doivent être fournies dans l'ordre indiqué ci dessus.

- 1) TAB : Nom du tableau SAS contenant les données
- 2) Y : Nom de la variable expliquée
- 3) X : Liste des noms des variables explicatives séparés par un espace

En sortie (options) :

COVA = nom de tableau SAS.

Sauvergarde l'estimation et la matrice de covariance robuste correspondante dans le fichier indiqué.

RES = nom de tableau SAS.

Sauvergarde le résidu dans le fichier indiqué

VIEUX = nom de tableau SAS.

Sauvergarde l'estimation et la matrice de covariance NON robuste correspondante dans le fichier indiqué (déconseillé).

Exemple :

```
%MCOROB(coucou,y,z,cova=c,res=u,vieux=vi); run;
```

...effectue une régression à partir des données du tableau work.coucou, en prenant y comme variable expliquée, z et un terme constant comme variables explicatives, et stocke les résultats de l'estimation robuste dans work.c, le résidu dans work.u et les

résultats de l'estimation non robuste dans work.vi.

N.B. : par défaut, les tableaux des options cov_a=, res= et vieu_x= sont stockés respectivement sous work.__c, work.__r et work.__v. On a mis deux caractères « souligné » (touche 8) au début du nom de ces tableaux afin d'éviter d'écraser des tableaux déjà existants. En interactif, on peut donc récupérer les résultats de la régression même si l'on a oublié de spécifier les options.



2.5 Macro-commande %BETWEEN(TAB,Y,X,I); RUN;

En entrée : les 4 informations suivantes sont obligatoires et doivent être fournies dans l'ordre indiqué ci dessus.

- 1) TAB : Nom du tableau SAS contenant les données
- 2) Y : Nom de la variable expliquée
- 3) X : Liste des noms des variables explicatives séparés par un espace
- 4) I : identifiant des individus (e.g., code SIREN pour une entreprise).

En sortie (options) : comme %MCO ROB, plus ...

TRANSF = nom de tableau SAS.

Contient les moyennes individuelles des variables, non répliquées, et sauvegardées sous leur nom original.

Exemple :

```
%BETWEEN(coucou,y,z,siren,cova=c,res=u,vieux=vi,transf=moy);run;
```

...effectue une régression à partir des données du tableau work.coucou, en prenant y comme variable expliquée, z et un terme constant comme variables explicatives. On effectue les moyennes individuelles par siren. On stocke les résultats de l'estimation robuste dans work.c, le résidu dans work.u et les résultats de l'estimation non robuste dans work.vi. Les moyennes individuelles sont stockées dans work.moy.

N.B. : par défaut, les tableaux des options cov_a=, res=, vieu_x= et transf= sont stockés respectivement sous work.__cbet, work.__rbet, work.__pbet et work.__between. Il n'y a qu'un seul caractère « souligné » (touche 8) pour le dernier nom de tableau (i.e., 8 caractères maximum). En interactif, on peut donc récupérer les résultats de la régression même si l'on a oublié de spécifier les options.



2.6 Macro-commande %BWITHIN(TAB,Y,X,I,T); RUN;

En entrée : les 5 informations suivantes sont obligatoires et doivent être fournies dans l'ordre indiqué ci dessus.

- 1) TAB : Nom du tableau SAS contenant les données.
- 2) Y : Nom de la variable expliquée.
- 3) X : Liste des noms des variables explicatives séparés par un espace.
- 4) I : identifiant des individus (e.g., code SIREN pour une une entreprise).
- 5) T : identifiant de date (e.g., l'année).

En sortie (options) : comme %BETWEEN

TRANSF = nom de tableau SAS.

Contient les écarts aux moyennes individuelles des variables repérées sous leur nom original.

Exemple :

```
%BWITHIN(coucou,y,z,siren,an,cova=c,res=u,vieux=vi, transf=moy); run;
```

...effectue une régression à partir des données du tableau work.coucou, en prenant y comme variable expliquée, et en prenant z, un terme constant ainsi que t07, t08, an89 et an90 comme variables explicatives.

On effectue les moyennes individuelles par siren et l'identifiant temporel est stocké dans la variable an. On stocke les résultats de l'estimation robuste dans work.c, le résidu dans work.u et les résultats de l'estimation non robuste dans work.vi. Les écarts aux moyennes individuelles sont stockées dans work.moy.

N.B. : par défaut, les tableaux des options cova=, res=, vieux= et transf= sont stockés respectivement sous work.__cwit, work.__rwit, work.__pwit et work.__within. Il n'y a qu'un seul caractère « souligné » (touche 8) pour le dernier nom de tableau. En interactif, on peut donc récupérer les résultats de la régression même si l'on a oublié de spécifier les options. Mieux, tous les résultats de la régression « between » sont également disponibles et sont stockés dans les tableaux dont les noms sont indiqués plus haut.



2.7 Macro-commande %BWMCQG(TAB,Y,X,I,T); RUN;

En entrée : les 5 informations suivantes sont obligatoires et doivent être fournies dans l'ordre indiqué ci dessus.

- 1) TAB : Nom du tableau SAS contenant les données.
- 2) Y : Nom de la variable expliquée.
- 3) X : Liste des noms des variables explicatives séparés par un espace.
- 4) I : identifiant des individus (e.g., code SIREN pour une une entreprise).

5) T : identifiant de date (e.g., l'année).

En sortie (options) : comme %BWITHIN, plus ...

TRANSF = nom de tableau SAS.

Contient les quasi-différences des variables sauvegardées sous leur nom original.

Exemple :

```
%BWITHIN(coucou,y,z,siren,an,cova=c,res=u,vieux=vi, transf=moy,s=t07
t08,dtemp=an89 an90); run;
```

...effectue une régression à partir des données du tableau work.coucou, en prenant y comme variable expliquée, et en prenant z, un terme constant ainsi que t07, t08, an89 et an90 comme variables explicatives.

On stocke les résultats de l'estimation robuste dans work.c, le résidu dans work.u et les résultats de l'estimation non robuste dans work.vi. quasi-différences sont stockées dans work.moy.

Par rapport à %bwithin, On remarque la présence de deux options supplémentaires :

- **S** = la liste des variables explicatives qui n'ont qu'une dimension individuelle. *Exemple* : indicatrices sectorielles ou toute information **constante dans le temps** pour chaque individu.
- **DTEMP** = la liste des variables explicatives qui n'ont qu'une dimension temporelle. *Exemple* : indicatrices temporelles ou toute information **constante sur les individus** à chaque date.

Utilisation des options S= et DTEMP=

Les variables indiquées dans l'option **S=** ne sont utilisées que pour les régressions « between » et par les moindres carrés quasi-généralisés (i.e., elles ne peuvent pas être utilisées dans la régression within car elles sont identiquement nulles dans cette dimension).

Les variables indiquées par l'option **DTEMP=** ne sont utilisées que pour les régressions « within » et par les moindres carrés quasi-généralisés (i.e., elles ne peuvent pas être utilisées dans la régression between car elles sont colinéaires au terme constant dans cette dimension).

N.B. : par défaut, les tableaux des options cova=, res=, vieux= et transf= sont stockés respectivement sous work.__cmcg, work.__rmcg, work.__pmcg et work._mcg. Il n'y a qu'un seul caractère « souligné » (touche 8) pour le dernier nom de tableau, et deux pour les autres. En interactif, on peut donc récupérer les résultats de la régression même si l'on a oublié de spécifier les options. Mieux, tous les résultats des régressions « between » et « within » sont également disponibles et stockés dans les tableaux dont les noms sont indiqués plus haut.



2.8 Exemple d'utilisation

Cette simulation montre clairement qu'en présence d'un effet aléatoire corrélé avec la variable explicative, les estimations between et des moindres carrés quasi-généralisés sont biaisées (i.e., supérieure à 1 pour une vraie valeur de 0.3). Par contre l'estimation within est proche de la vraie valeur (0.294 pour 0.3). De plus, le test de Fisher conclut à l'existence d'un effet aléatoire et le test de Hausman conclut à l'existence d'une corrélation entre la variable expliquée et l'effet individuel, ce qui est cohérent avec la simulation.

```
filename npanel 'c:\user\duguet\macros\dmse97\npanel.pcu';
libname simul 'c:\user\duguet\macros\dmse97\';
```

```
%include npanel;
run;
```

```
data tablo;
seed=26081966;
do siren=1 to 400;
effet=rannor(seed);
  do an=89 to 94;
  x=an-89+rannor(seed)/3+effet;
  y=0.3*x+rannor(seed)/3+effet;
  output;
  end; end;
```

```
data simul.tablo; set tablo;
d90=(an=90); d91=(an=91); d92=(an=92);
d93=(an=93); d94=(an=94);
run;
```

```
%bwmcqg(simul.tablo,y,x,siren,an,dtemp=d90 d91 d92 d93
d94); run;
```

CORRELATIONS DANS LA DIMENSION INTER-INDIVIDUELLE (BETWEEN)

1

Correlation Analysis

2 'VAR' Variables: Y X

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum
Maximum					
Y	400	0.81687	1.29552	326.74840	-2.48420

5.21447
 X 400 2.56033 0.99524 1024 0.09997
 5.85371

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 400

	Y	X
Y	1.00000 0.0	0.98838 0.0001
X	0.98838 0.0001	1.00000 0.0

ESTIMATION DANS LA DIMENSION INTER-INDIVIDUELLE (BETWEEN)

2

Model: MODEL1
 Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	1	654.19814	654.19814	16832.530
Error	398	15.46831	0.03887	
C Total	399	669.66646		
Root MSE	0.19714	R-square	0.9769	
Dep Mean	0.81687	Adj R-sq	0.9768	
C.V.	24.13384			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	-2.477228	0.02723624	-90.953	0.0001
X	1	1.286594	0.00991669	129.740	0.0001

ESTIMATION DANS LA DIMENSION INTER-INDIVIDUELLE (BETWEEN)

3

Régresseur	Coefficient MCO	Ecart-Type Robuste	T de Student Robuste	Seuil de Significativité
INTERCEP	-2.47723	0.026409	93.803	0
X	1.28659	0.009677	132.957	0

CORRELATIONS DANS LA DIMENSION INTRA-INDIVIDUELLE (WITHIN)

4

Correlation Analysis

2 'VAR' Variables: Y X

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum
Maximum					
Y	2400	0	0.61292	0	-1.87325
1.82712					
X	2400	0	1.75034	0	-3.36720
3.58039					

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 2400

	Y	X
Y	1.00000 0.0	0.86363 0.0001
X	0.86363 0.0001	1.00000 0.0

ESTIMATION DANS LA DIMENSION INTRA-INDIVIDUELLE (WITHIN)

5

Model: MODEL1

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

Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	672.76620	112.12770	1174.861
Error	2394	228.48125	0.09544	
U Total	2400	901.24746		

Root MSE	0.30893	R-square	0.7465
Dep Mean	0.00000	Adj R-sq	0.7458
C.V.	1.5530869E19		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
X	1	0.294094	0.00586821	50.116	0.0001
D90	1	-0.027989	0.01781397	-1.571	0.1163
D91	1	-0.016246	0.01574507	-1.032	0.3023
D92	1	-0.012590	0.01574904	-0.799	0.4241
D93	1	0.028144	0.01782046	1.579	0.1144
D94	1	0.026333	0.02136082	1.233	0.2178

ESTIMATION DANS LA DIMENSION INTRA-INDIVIDUELLE (WITHIN)

6

Ecarts-types robuste a l'autocorrelation et a l'heteroscedasticite

Régresseur	Coefficient MCO	Ecart-Type Robuste	T de Student Robuste	Seuil de Significativité
X	0.29409	0.005983	49.1543	0.00000
D90	-0.02799	0.019907	1.4060	0.15972
D91	-0.01625	0.015644	1.0384	0.29906
D92	-0.01259	0.014835	0.8486	0.39609
D93	0.02814	0.016000	1.7590	0.07858

PROBA
Seuil de Significativite = 2.029E-13

Statistique de Hausman (H0 -> Modele a erreurs composees OK) :
Test de la difference entre within et between
(sans constante ni indicatrices, ecart types robustes)

K HAUSMAN
Khi-deux (1) = 7610.3128

PROBA
Seuil de Significativite = 0

Analyse de la Variance

9

THETA2
THETA CARRE = 0.5881768

THETA
THETA = 0.7669268

CORRELATIONS DES QUASI DIFFERENCES (MCQG)

10

Correlation Analysis

2 'VAR' Variables: Y X

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum
Maximum					
Y	2400	0.62648	1.16653	1504	-2.77454
5.48341					
X	2400	1.96358	1.90920	4713	-2.50295
7.06637					

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 2400

	Y	X
Y	1.00000 0.0	0.75187 0.0001
X	0.75187 0.0001	1.00000 0.0

ESTIMATION DU MODELE TRANSFORME (MCQG)

11

Il faut diviser le coefficient du terme constant par 'theta'

Model: MODEL1
Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	2789.98503	464.99751	2344.800
Error	2393	474.55596	0.19831	
C Total	2399	3264.54099		
Root MSE	0.44532	R-square	0.8546	
Dep Mean	0.62648	Adj R-sq	0.8543	
C.V.	71.08287			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	0.525779	0.02309818	22.763	0.0001
X	1	1.146803	0.01104317	103.847	0.0001
D90	1	-0.889717	0.03339787	-26.640	0.0001
D91	1	-1.724027	0.03846253	-44.824	0.0001
D92	1	-2.610055	0.04605526	-56.672	0.0001
D93	1	-3.414300	0.05455638	-62.583	0.0001
D94	1	-4.268739	0.06389223	-66.812	0.0001

ESTIMATION DU MODELE TRANSFORME (MCQG)

12

Ecart-types robuste a l'autocorrelation et a l'heteroscedasticite

Régresseur	Coefficient MCO	Ecart-Type Robuste	T de Student Robuste	Seuil de Significativité
INTERCEP	0.52578	0.022867	22.9928	0
X	1.14680	0.013238	86.6269	0
D90	-0.88972	0.033475	26.5789	0
D91	-1.72403	0.038928	44.2876	0
D92	-2.61006	0.049389	52.8471	0
D93	-3.41430	0.059630	57.2583	0
D94	-4.26874	0.071259	59.9045	0

3. Estimation par la méthode des moments généralisés

3.1 Présentation du modèle

On reprend le modèle à effet fixe individuel. Afin d'éliminer l'effet fixe, on prend la différence première de l'équation de départ. Lorsque le modèle est en logarithmes, il s'agit d'une approximation du modèle en taux de croissance. Dans la présentation qui suit, on considère que les variables sont déjà prises en variation et que l'ensemble des variables explicatives contient éventuellement un jeu complet d'indicatrices temporelles. Soit, en abrégé, le modèle linéaire :

$$y = x b + u$$

Notons que la matrice x peut très bien contenir la variable endogène retardée y_{-1} . L'essentiel est que la matrice x contienne des variables explicatives endogènes. Par hypothèse, on admet que les régresseurs sont corrélés avec la perturbation u^3 . On a

$\text{Plim } x'u/n \neq 0$. Dans ce cas, aucun des estimateurs de la section précédente n'est convergent. Nous devons donc utiliser une méthode à variables instrumentales. Ceci revient à utiliser un ensemble de variables z satisfaisant les trois conditions suivantes : $\text{Plim } z'u/n = 0$, $\text{Plim } z'z/n$ inversible et $\text{Plim } z'x/n$ de même rang que le nombre de variables explicatives du modèle.

Dans un premier temps, on effectue donc une estimation par les variables instrumentales (i.e., les doubles moindres carrés), ce qui nous donne un estimateur convergent :

$$\hat{\beta} = [x'Px]^{-1} x'Py$$

de covariance asymptotique :

$$\text{Vas}(\hat{\beta}) = [x'Px]^{-1} x'z(z'z)^{-1} \Phi(z'z)^{-1} z'x[x'Px]^{-1}$$

où P est la matrice de projection orthogonale sur $\text{Im}(z)$, $P = z(z'z)^{-1}z'$ et $\Phi = \text{Vas}(z'u)$.

L'estimateur des moments généralisés, noté β^* , fait un meilleur usage de l'information disponible, en tenant compte de la covariance des perturbations. On minimise la quantité suivante :

$$\beta^* = \arg \min_{\beta} u'z\Phi^{-1}z'u$$

Ce qui donne :

$$\beta^* = (x'z\Phi^{-1}z'x)^{-1} x'z\Phi^{-1}z'y$$

$$\text{Vas}(\beta^*) = (x'z\Phi^{-1}z'x)^{-1}.$$

Il nous faut donc un estimateur de Φ pour réaliser l'estimation. Il est facilement obtenu en effectuant d'abord une estimation par les doubles moindres carrés, parfois appelée estimateur des moments généralisés de première étape.

Soit \hat{u}_i le résidu (T,1) des doubles moindres carrés correspondant à l'individu i et z_i la matrice (T,p) des instruments pour ce même individu. Alors, comme les observations sont indépendamment distribuées dans la dimension individuelle, on obtient l'estimateur

³ Qui est elle-même autocorrélée dans un modèle en différences premières.

convergent suivant Φ (White, 1980, 1982) :

$$\hat{\Phi} = \frac{1}{n} \sum_{i=1}^n z_i' \hat{u}_i \hat{u}_i' z_i$$

A partir de cet estimateur, on recalcule l'estimateur des moments généralisés une seconde fois, ainsi que sa matrice de covariance asymptotique.

Il nous reste alors à tester la validité des instruments utilisés, c'est-à-dire du modèle lui-même. En utilisant l'estimateur de deuxième étape, on effectue le test de suridentification de Sargan (1958), dont la statistique est donnée par :

$$S = u^{*'} z \hat{\Phi}^{-1} z' u^*$$

Sous l'hypothèse nulle $E(z'u) = 0$ elle suit un χ^2 à p-k degrés de liberté où u^* désigne le résidu de deuxième étape.

Instruments utilisés pour l'estimation d'un modèle statique sous exogénéité faible

Le modèle en variations s'écrit :

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})b + d_t + u_{it} - u_{it-1} \text{ avec } d_t = \delta_t - \delta_{t-1}$$

$$i = 1, \dots, n \quad t = 1, \dots, T$$

Comme y et x sont déterminés simultanément, les instruments doivent être décalés d'au moins deux périodes. Les macro-commandes qui suivent considèrent que les instruments ne sont que faiblement exogènes, c'est-à-dire que les perturbations ne sont indépendantes que des valeurs strictement passées des instruments (i.e., sont corrélées avec les valeurs présentes et futures des instruments). On peut choisir librement la liste des variables instrumentales.

Ainsi, pour un décalage de deux ans, l'ensemble des variables instrumentales pour l'individu i est donné par :

$$z_i = \begin{bmatrix} z_{i1} & 0 & 0 & 0 \\ 0 & z_{i1} & z_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & z_{i1} \dots z_{iT-2} \end{bmatrix} \quad T \geq 3$$

La macro-commande réalise les opérations suivantes :

- Mise du modèle en différences premières.
- Les variables instrumentales sont conservées en niveaux. On peut facilement contourner cette option par défaut en mettant parmi les instruments des variables déjà en différences premières dans le tableau de départ (i.e., le programme n'effectue aucune transformation des variables instrumentales).
- Estimation par les doubles moindres carrés, avec matrice de covariance robuste à l'hétéroscédasticité.
- Estimation par les moments généralisés.
- Test de Sargan.
- Enfin, une macro-commande spécifique, présentée dans la section suivante, ajoute automatiquement la variable endogène décalée aux variables explicatives. Comme

l'écart est de deux ans, aucune modification des instruments n'est nécessaire pour le modèle de base, si l'on instrumente par les niveaux. On peut toutefois augmenter le décalage des instruments en utilisant les options.

3.2 Macro-commande %MOMDYN

La syntaxe de base est :

```
%MOMDYN(TAB,Y,X,I,T,Z=Z1 Z2,D=2) ; RUN ;
```

avec

TAB : nom du tableau SAS contenant les données.

Y : nom de la variable expliquée.

X : liste des variables explicatives (la valeur décalée de la variable endogène est ajoutée automatiquement, il ne faut donc surtout pas la mettre sous peine d'obtenir une multicolinéarité parfaite).

I : identifiant d'individu (e.g., le numéro SIREN pour une entreprise).

T : identifiant temporel (e.g., l'année), cette variable doit être NUMÉRIQUE.

Z=liste des instruments.

Dans l'exemple ci dessus, les instruments sont les variables Z1 et Z2 situées dans la base WORK.TAB.

D=nombre entier supérieur ou égal à 2.

Décalage employé pour les instruments, il est égal à 2 par défaut. Ce nombre est le minimum possible car le modèle est pris en variations. Il faut faire attention à ne pas dépasser la longueur du panel.

COVA1= nom de tableau SAS.

Contient l'estimation par les doubles moindres carrés avec sa matrice de covariance robuste. Par défaut, le tableau est sauvegardé dans WORK.__V1.

COVA2= nom de tableau SAS.

Contient l'estimation par les moments généralisés avec sa matrice de covariance robuste. Par défaut, le tableau est sauvegardé dans WORK.__V2.



3.3 Exemple d'utilisation

L'estimation du modèle en différences premières donne des résultats similaires au within, mais avec une précision beaucoup plus faible (coefficient non significatif). Le modèle dynamique est ensuite estimé avec une valeur initiale aléatoire et un coefficient de la variable retardée de 1/2. On trouve une bonne estimation pour la variable retardée, avec estimation par les moments généralisés du modèle en variation avec un décalage de 3 ans pour les instruments. Toutefois, le problème de précision se pose toujours sur la variable explicative. Son origine réside certainement dans les faibles variances obtenues en

différences premières.

```
filename npanel 'c:\user\duguet\macros\dmse97\momdyn.pcu';
libname simul 'c:\user\duguet\macros\dmse97\';
```

```
%include npanel;
run;
```

```
data tablo;
seed=26081966;
do siren=1 to 400;
effet=rannor(seed);
y0=rannor(seed);
  do an=89 to 94;
    base=an-89+rannor(seed)/3+effet;
    x=base+rannor(seed)/2;
    z=base+rannor(seed)/3;
    u=rannor(seed);
    output;
  end; end;
```

```
data simul.tablo2; set tablo; by siren;
retain y;
if first.siren then y=y0; else y=0.5*y+0.3*x+effet+u;
d90=(an=90); d91=(an=91); d92=(an=92);
d93=(an=93); d94=(an=94);
run;
```

```
%momdyn(simul.tablo2,y,x,siren,an); run;
```

```

                                                    NY
VARIABLE EXPLIQUEE = Y

                                                    NX
T
VARIABLES EXPLICATIVES = X  NOMBRE DE PERIODES =
6

                                                    N
NOMBRE D'INDIVIDUS = 400

                                                    NT
NOMBRE D'OBSERVATIONS = 2400

                                                    K
NOMBRE DE VARIABLES EXPLICATIVES = 1

                                                    P
NOMBRE D'INSTRUMENTS = 6

DECALAGE EMPLOYE = 3

SOUS HYPOTHESES D'HETEROSCEDASTICITE ET
D'AUTOCORRELATION
(PREMIERE ETAPE)
```

RESULTAT DE L'ESTIMATION :

NY

VARIABLE EXPLIQUEE : Y

	NX	NZ
variables explicatives =	LAG_Y Instruments =	X
	X	T_1
	T_1	T_2
	T_2	T_3
	T_3	

ESTIMATIONS :

PROBA	RES	COEFFICIENT	ECART_TYPE	STUDENT	
10	LAG_Y	0.2450825	0.0390932	6.2691791	3.63E-
0.2711632	X	0.5577065	0.5068273	1.1003877	
0.8821206	T_1	-0.069825	0.4708978	0.1482815	
0.6951488	T_2	-0.199381	0.5087844	0.3918774	
0.7102329	T_3	-0.195277	0.5255833	0.3715433	

ESTIMATION PAR LES MOMENTS GENERALISES
(SECONDE ETAPE)

RESULTAT DE L'ESTIMATION :

NY

VARIABLE EXPLIQUEE : Y

	NX	NZ
variables explicatives =	LAG_Y Instruments =	X
	X	T_1
	T_1	T_2
	T_2	T_3
	T_3	

ESTIMATIONS :

PROBA	RES	COEFFICIENT	ECART_TYPE	STUDENT	
12	LAG_Y	0.2550816	0.0372128	6.8546692	7.148E-
0.222926	X	0.6084843	0.499255	1.2187846	
0.8063939	T_1	-0.113695	0.4639088	0.2450808	
0.6231993	T_2	-0.246416	0.501537	0.4913212	
0.6228954	T_3	-0.255407	0.5193834	0.491751	

SARGAN

STATISTIQUE DE SARGAN = 2.582183

PROBS
PROBA = 0.6299826

3.4 Macro-commande %MOMSTA

La syntaxe est exactement la même que pour %MOMDYN. La seule différence est que la variable endogène décalée n'est pas ajoutée aux variables explicatives.

```

                                                    NY
VARIABLE EXPLIQUEE = Y

                                                    NX
T
VARIABLES EXPLICATIVES = X  NOMBRE DE PERIODES =
6

                                                    N
NOMBRE D'INDIVIDUS = 400

                                                    NT
NOMBRE D'OBSERVATIONS = 2400

                                                    K
NOMBRE DE VARIABLES EXPLICATIVES = 1

                                                    P
NOMBRE D'INSTRUMENTS = 10

DECALAGE EMPLOYE = 2

SOUS HYPOTHESES D'HETEROSCEDASTICITE ET
D'AUTOCORRELATION
(PREMIERE ETAPE)

RESULTAT DE L'ESTIMATION :

                                                    NY
VARIABLE EXPLIQUEE : Y

variables explicatives = X  Instruments = X
T_1
T_2
T_3
T_4
                                                    NZ
T_1
T_2
T_3
T_4

ESTIMATIONS :

RES COEFFICIENT  ECART_TYPE  STUDENT  PROBA

```

X	0.333798	0.3067977	1.0880069	0.276592
T_1	-0.027651	0.3021259	0.09152	0.9270794
T_2	-0.037769	0.3242857	0.1164695	0.9072804
T_3	0.0013892	0.3037381	0.0045736	0.9963508
T_4	-0.041511	0.3060258	0.1356448	0.8921021

ESTIMATION PAR LES MOMENTS GENERALISES
(SECONDE ETAPE)

RESULTAT DE L'ESTIMATION :

NY
VARIABLE EXPLIQUEE : Y

	NX		NZ
variables explicatives = X		Instruments = X	
	T_1		T_1
	T_2		T_2
	T_3		T_3
	T_4		T_4

ESTIMATIONS :

RES	COEFFICIENT	ECART_TYPE	STUDENT	PROBA
X	0.3993668	0.2601358	1.535224	0.1247288
T_1	-0.095735	0.2560579	0.3738821	0.7084921
T_2	-0.107761	0.2748204	0.3921149	0.6949733
T_3	-0.057675	0.2584261	0.2231791	0.8233961
T_4	-0.108226	0.2598192	0.4165417	0.6770137

SARGAN
STATISTIQUE DE SARGAN = 8.7464936

PROBS
PROBA = 0.4609966

4. Le modèle logit polytomique ou « multinomial logit »

4.1 Présentation du modèle

Il s'agit d'un modèle de choix discret entre un nombre fixe de possibilités mutuellement exclusives. On suppose qu'un individu prend une décision sur la base d'une comparaison des utilités associées à chaque action. L'utilité de l'individu i quand il prend une décision j est supposée s'écrire $U_{ji} = x_{ji} b_j + v_{ji}$ où x_j désigne l'ensemble des variables explicatives, b_j le vecteur des coefficients qui leurs sont associés et v_{ji} un terme aléatoire représentant l'hétérogénéité des individus. Si les perturbation v_{ji} sont indépendantes et identiquement distribuées selon une loi de Gompertz, on obtient le modèle logit polytomique, aussi appelé « multinomial logit ». La probabilité de réalisation de l'action j par rapport à toutes les autres actions, regroupées dans l'ensemble d'indices J est donnée par :

$$P_{j/J} = \frac{\exp(x_j b_j)}{\sum_{m \in J} \exp(x_m b_m)}$$

Afin d'assurer l'identifiabilité des paramètres du modèle, on impose que les variables explicatives sont les mêmes pour toutes les décisions et que le paramètre de la première modalité est nulle. Cette première modalité devient alors la référence par rapport à laquelle il faut interpréter les estimations. Pour choisir une modalité particulière comme référence, il suffit donc de lui donner la valeur la plus faible (e.g., -100). Cette valeur n'apparaîtra pas dans le programme car il renumérote les modalités à partir de 1, c'est donc toujours la modalité 1 qui sera la référence dans le listing.

L'estimation est réalisée par le maximum de vraisemblance. Deux algorithmes d'optimisation sont disponibles : Berndt-Hall-Hausman et Newton-Raphson. Pour plus de détails sur le modèle et la méthode d'estimation, voir Gouriéroux 1989, page 51 et suivantes ainsi que le chapitre 3 du même ouvrage.

4.2 La macro-commande %MULTINO

La syntaxe d'utilisation est la suivante :

%MULTINO(TAB,ENDO,EXO); RUN ;

avec :

TAB

Nom du tableau SAS contenant les données

ENDO

Nom de la variable endogène

EXO

Noms des variables exogènes séparés par des espaces. Il faut mettre explicitement le terme constant, le programme n'en met pas automatiquement

IMPR=0

Annule l'impression des itérations (option par défaut).

IMPR=1

Active l'impression du détail des itérations.

ARRET=1E-8

Critère d'arrêt fixé à 10^{-8} (option par défaut). Le maximum des valeurs absolues des élasticités de la log-vraisemblance par rapport aux coefficients ne doit pas dépasser cette valeur pour que l'algorithme converge. On peut parfois mettre des valeurs plus élevées (i.e., le critère par défaut est strict et la convergence est souvent atteinte dans la pratique deux ou trois itérations avant que le programme ne s'arrête).

MAXITER=100

Arrêt de sécurité pour éviter une boucle infinie en cas de non convergence. Il s'agit du nombre maximum d'itérations permises. Dans la pratique le programme converge en beaucoup moins d'itérations. On peut fixer librement cette valeur.

MAXMOD=50

Nombre maximum de sous-itérations pour trouver une direction de croissance. Permet d'autoriser un pas minimal de 2^{50} . On peut fixer cette valeur librement.

SORB=nom de tableau SAS

Nom du tableau dans lequel sont sauvegardés les estimations des paramètres et de leur matrice de covariance.

SORI=nom de tableau SAS

Nom du tableau dans lequel sont sauvegardés la log-vraisemblance individu par individu (i.e., l'estimation du logarithme de la probabilité associée à l'événement considéré), la réalisation de la variable endogène recodée à partir de 1 et l'identifiant d'individu (car le fichier est retrié au début de l'exécution de la macro-commande). L'utilisation de cette option rend OBLIGATOIRE l'utilisation de l'option IDENT=.

SCORE=nom de tableau SAS

Nom du tableau dans lequel sont sauvegardés les dérivées de la log-vraisemblance individu par individu. Les colonnes sont repérées par le nom de la variable correspondante. Attention ! L'utilisation de cette option rend OBLIGATOIRE l'utilisation de l'option IDENT=.

METHODE=NRA

Option par défaut, on utilise un algorithme de Newton-Raphson

METHODE=B3H

Demande l'utilisation de l'algorithme de Berndt-Hall-Hall-Hausman. Cet algorithme utilise des pas plus courts que la méthode de Newton-Raphson. Il peut permettre de régler des problèmes de convergence lorsque l'algorithme de Newton-Raphson effectue des pas initiaux trop élevés.

N.B. : même quand l'option précédente est utilisée, c'est toujours le hessien analytique au point de convergence qui est utilisé pour calculer les écarts-types asymptotiques.

TRONC=15

Indique la valeur de troncature pour le terme en « x b ». Elle est fixée par défaut à exp(15). On peut fixer la valeur que l'on souhaite, en faisant attention de ne pas causer un dépassement de la capacité de traitement de l'ordinateur.

INF=1E-8

Valeur minimale autorisée pour une probabilité, car on doit diviser par cette quantité. Par symétrie, on calcule la valeur maximale autorisée comme $1-INF$. Cette valeur peut être fixée librement.

IDENT=nom de variable

Nom de l'identifiant d'individu dans la base de départ. Cette option **EST OBLIGATOIRE** lorsque l'on utilise les options `SORI=` ou `SCORE=`. La raison pour laquelle cette option est obligatoire est que le tableau est retérié en début de programme.

INB=nom de tableau SAS

Tableau contenant la valeur initiale du paramètre. Cette option est extrêmement utile pour travailler sur de grosses bases de données, car elle permet de travailler en plusieurs étapes sans perdre le travail déjà effectué. Sur gros-système, elle permet notamment de faire passer des programmes de la classe (4,59) en (0,59). Ce tableau doit avoir été généré par la macro-commande `multino`. Elle rend **OBLIGATOIRE** l'utilisation de l'option `MAXITER=`. Il faut fixer ce nombre après un premier essai qui permet de déterminer combien d'itérations le programme peut effectuer dans le temps imparti (e.g. `MAXITER=2` ou `3`).

OUTB=nom de tableau SAS

Tableau où l'on stocke la valeur du paramètre lors de l'utilisation de `INB=`. On utilise donc d'abord `OUTB=` avec `MAXITER=`, puis on initialise le paramètre avec l'instruction `INB=` dans le programme qui suit. On peut utiliser les deux options en même temps et avec le même nom de tableau. Dans ce dernier cas, la valeur du paramètre est mise à jour. Il est **IMPORTANT** de sauvegarder le tableau sur disque lorsque l'on travaille sur gros système.

4.3 Exemple d'utilisation

L'exemple montre que les termes constants ne sont pas significatifs et que le coefficient de la variable explicative, toujours significatif au seuil de 5%, est croissant avec la modalité, ce qui correspond bien au modèle simulé.

```
filename                                                    npanel
'c:\user\duguet\macros\dmse97\multino.pcu';
libname simul 'c:\user\duguet\macros\dmse97\';

%include npanel;
/* place la macro-commande au début du programme */
run;

data tablo; /* simulation */
  /* la loi de Gompertz n'est pas disponible */
  /* on en prend une autre */
seed=26081966;
do siren=1 to 600;
x=rannor(seed);
v1=x+rannor(seed)/4;
v2=2*x+rannor(seed)/4;
v3=3*x+rannor(seed)/4
output;
end;
```

```

data simul.tablo3; set tablo;
y=3;
if v1>v2 and v1>v3 then y=1;
if v2>v1 and v2>v3 then y=2;
cste=1;
run;

```

```
%multino(simul.tablo3,y,cste x,impr=1); run;
```

Percent	Y	Frequency	Cumulative Percent	Cumulative Frequency
46.5	1	279	46.5	279
52.2	2	34	5.7	313
100.0	3	287	47.8	600

NY
VARIABLE EXPLIQUEE : Y

NOMS
VARIABLES EXPLICATIVES : X
CSTE

ESTIMATION DU MODELE LOGIT MULTINOMIAL

ALGORITHMME : NRA

M{THODE : MAXIMUM DE VRAISEMBLANCE

NOMS2
B
VALEURS INITIALES : MODALITE_2:X -
0.004157
MODALITE_2:CSTE
0.2451433
MODALITE_3:X
1.2010957
MODALITE_3:CSTE
1.4992492

(NON CONVERGENTES)

LV
LOG-VRAISEMBLANCE AU POINT INITIAL = -
515.2277

RES	ITERATION	PAS LOG_VRAISEMBLANCE
	1	1 -227.6859

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	0.9755141	-0.319254	74.514127
MODALITE_2:CSTE	-1.777488	-0.701602	-89.87117
MODALITE_3:X	2.9388454	-1.291665	100.07123
MODALITE_3:CSTE	-0.264146	-0.107231	-92.43026

RES	ITERATION	PAS LOG_VRAISEMBLANCE
	2	1 -177.7197

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	2.0196942	-0.051956	4.5717605
MODALITE_2:CSTE	-1.33185	0.0011465	0.1529915
MODALITE_3:X	4.9558456	-0.977422	35.050963
MODALITE_3:CSTE	-0.226847	0.0167871	13.151614

RES	ITERATION	PAS LOG_VRAISEMBLANCE
	3	1 -149.7453

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	3.3323185	-0.041357	1.8584642
MODALITE_2:CSTE	-0.943934	-0.00974	-1.545081
MODALITE_3:X	7.6754403	-0.778731	15.192782
MODALITE_3:CSTE	-0.245222	0.0086906	5.306938

RES	ITERATION	PAS LOG_VRAISEMBLANCE
	4	1 -136.8709

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	4.8690984	-0.028839	0.8106663
MODALITE_2:CSTE	-0.612334	-0.007796	-1.742479
MODALITE_3:X	10.85488	-0.496296	6.2578779
MODALITE_3:CSTE	-0.280806	0.0050424	2.457775

RES	ITERATION	PAS LOG_VRAISEMBLANCE
	5	1 -133.2027

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.2025958	-0.015244	0.3273721
MODALITE_2:CSTE	-0.395978	-0.004114	-1.383823
MODALITE_3:X	13.557746	-0.230539	2.2650129
MODALITE_3:CSTE	-0.307064	0.0027807	1.2062641

RES	ITERATION	PAS LOG_VRAISEMBLANCE
-----	-----------	-----------------------

6 1 -132.8226

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.8167701	-0.005166	0.1006625
MODALITE_2:CSSTE	-0.316481	-0.001507	-0.632388
MODALITE_3:X	14.759582	-0.064691	0.5821556
MODALITE_3:CSSTE	-0.314152	0.0010527	0.4450766

RES	ITERATION	PAS	LOG_VRAISEMBLANCE
	7	1	-132.8174

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.9031241	-0.000784	0.0150798
MODALITE_2:CSSTE	-0.306906	-0.000226	-0.097868
MODALITE_3:X	14.919817	-0.007022	0.0625108
MODALITE_3:CSSTE	-0.314309	0.0001495	0.0631549

RES	ITERATION	PAS	LOG_VRAISEMBLANCE
	8	1	-132.8174

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.9045271	-0.000018	0.0003413
MODALITE_2:CSSTE	-0.306764	-3.918E-6	-0.001696
MODALITE_3:X	14.922231	-0.000099	0.0008812
MODALITE_3:CSSTE	-0.314296	2.4796E-6	0.0010478

RES	ITERATION	PAS	LOG_VRAISEMBLANCE
	9	1	-132.8174

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.9045275	-6.328E-9	1.2172E-7
MODALITE_2:CSSTE	-0.306764	-9.64E-10	-4.172E-7
MODALITE_3:X	14.922231	-2.057E-8	1.8309E-7
MODALITE_3:CSSTE	-0.314296	5.773E-10	2.4397E-7

RES	ITERATION	PAS	LOG_VRAISEMBLANCE
	10	1	-132.8174

RES	COEFFICIENT	ELASTICITE	GRADIENT
MODALITE_2:X	6.9045275	-6.37E-16	1.226E-14
MODALITE_2:CSSTE	-0.306764	-4.44E-17	-1.92E-14
MODALITE_3:X	14.922231	-1E-15	8.923E-15
MODALITE_3:CSSTE	-0.314296	6.279E-17	2.653E-14

PAS DE SAUVEGARDE DEMANDEE...

NIT
10

L'ALGORITHME A CONVERGE EN

ITERATIONS

ELASTICITES DE L'OBJECTIF PAR RAPPORT AUX

COEFFICIENTS :

NOMS2	ELAS
MODALITE_2:X	-6.37E-16
MODALITE_2:CSTE	-4.44E-17
MODALITE_3:X	-1E-15
MODALITE_3:CSTE	6.279E-17

LE GRADIENT EST EGAL A :

NOMS2	DILV
MODALITE_2:X	1.226E-14
MODALITE_2:CSTE	-1.92E-14
MODALITE_3:X	8.923E-15
MODALITE_3:CSTE	2.653E-14

ET LES VALEURS PROPRES DU HESSIEN ANALYTIQUE

A :

VPH
-0.289196
-1.487991
-11.98654
-37.38754

NY

VARIABLE EXPLIQUEE : Y

NMOD

NOMBRE DE MODALITES : 3

N

NOMBRE D'OBSERVATIONS : 600

KK

NOMBRE DE VARIABLES : 2

K

NOMBRE DE PARAMETRES : 4

RLV

MAXIMUM DE LOG VRAISEMBLANCE : -132.8174

RESULTATS DE L'ESTIMATION DU MODELE LOGIT POLYTOMIQUE
PAR LE MAXIMUM DE VRAISEMBLANCE
(ECARTS-TYPE BASES SUR LES DERIVEES SECONDES)

RES	COEFFICIENT	ECART_TYPE	T_DE_STUDENT	WALD	PROB
MODALITE_2:X	6.9045275	1.2333186	5.5983323	31.341325	2.1642E-8
MODALITE_2:CSTE	-0.306764	5.2312965	-0.05864	0.0034387	0.9532387
MODALITE_3:X	14.922231	0.9476264	15.746956	247.96663	0
MODALITE_3:CSTE	-0.314296	4.6223079	-0.067996	0.0046234	0.9457892

5. Sources des macro-commandes

Le fichier npanel.pcu contient les macro-commandes %mcorob, %between, %bwithin et %bwmcqq. On peut les utiliser séparément. Il suffit d'une seule instruction %include pour l'ensemble des macro-commandes. Cette instruction se contente en effet de mettre le fichier indiqué au début du programme avant son exécution. On pourrait y mettre un programme SAS standard sans problème.

Les autres programmes sont séparés et sont repérés sous le même nom que la macro-commande correspondante.

5.1 %mcorob

```
%macro mcorob(tab,y,x,cova=__c,res=__r,vieux=__v);
  /* calcul des ecartis-type robustes */
  /* a l'heteroscedasticite */
  proc reg data=&tab outest=__b;
    model &y=&x;
    output out=__res r=res;
  data __b; set __b(keep=intercep &x);
  proc transpose data=__b out=__b(rename=(coll=coeff));
  data __x; merge &tab(keep=&x) __res;
  res2=res**2;
  drop res;
  /* calcul de I=X'(u**2)X */
  proc corr nocorr noprint data=__x sscp outp=__i;
    var &x;
    weight res2;
  data __i; set __i; if _type_='SSCP'; drop _type_ _name_;
  /* calcul de J=X'X */
  proc corr nocorr noprint data=__x sscp outp=__j;
    var &x;
    run;
  data __j; set __j; if _type_='SSCP'; drop _type_ _name_;

  /* calcul de la covariance robuste par iml : inv(J)*I*inv(J) */
  /* et des statistiques associee */

  proc iml;
  use __b; read all var{__NAME_} into nvar;
  use __b; read all var{COEFF} into b;
  use __i; read all into mi;
  use __j; read all into mj;
  use __res; read all var{res} into u;
  imj=solve(mj,i(nrow(mj))); /* inversion plus efficace de J */
  covar=imj*mi*imj;
  std=sqrt(vecdiag(covar)); /* ecartis-type robustes hetero */
  stu=abs(b/std); /* Valeur absolue du Student asymptotique */
  pro=2*(1-probnorm(stu)); /* Seuil de significativite valeur absolue */
  test=std||stu||pro;
  noms={'std','stu','pro'};
  create __test from test [colname=noms]; append from test;
  create &cova from covar[colname=nvar]; append from covar;
  /* statistiques non robustes a l'heteroscedasticite */
  su2=ssq(u)/(nrow(u)-nrow(imj)); /* avec degres de liberte */
  cova=su2#imj;
  create &vieux from cova[colname=nvar]; append from cova;
  quit;
  data __test; merge __b __test;
  label
  _name_='Régresseur'
  coeff='Coefficient MCO'
  std='Ecart-Type Robuste'
```

```

    stu='T de Student Robuste'
    pro='Seuil de Significativité'
;
proc print label; var coeff std stu pro; id _name_;
run;
data &cova; merge __b &cova;
run;
data &res; set __res; label res='Residu MCO';
run;
data &vieux; merge __b &vieux;
run;
%mend mcorob;

```

5.2 %between

```

%macro
between(tab,y,x,i,transf=_between,cova=__cbet,res=__rbet,vieux=__pbet,s=);
proc sort data=&tab; by &i;
proc means noprint data=&tab; by &i;
var &y &x &s;
output out=&transf(drop=_type_) mean=;
proc corr pearson data=&transf; var &y &x &s;
title 'CORRELATIONS DANS LA DIMENSION INTER-INDIVIDUELLE (BETWEEN)';
run;
title 'ESTIMATION DANS LA DIMENSION INTER-INDIVIDUELLE (BETWEEN)';
run;
%mcorob(&transf,&y,&x,cova=&cova,res=&res,vieux=&vieux);
run;
%mend between;

```

5.3 %bwithin

```

%macro
bwithin(tab,y,x,i,t,dtemp=,cova=__cwit,res=__rwit,vieux=__pwit,transf=_with
in,s=);

    %between(&tab,&y,&x,&i,s=&s); run; /* differences inter-individuelles */

proc iml; /* differences intra-individuelles */
start lecture;
    noms={&y &x &s};
    use _between; read all var{&y &x &s} into by; n=nrow(by);
    use &tab; read all var {&y &x &s} into y; nt=nrow(y); t=int(nt/n);
finish lecture;
start within;
    et=j(t,1,1); wy=y-(by@et);
    free by;
    create &transf from wy[colname=noms]; append from wy;
finish within;

run lecture;
run within;

quit;

    /* correlations intra-individuelles */
proc corr pearson data=&transf; var &y &x &s;
title 'CORRELATIONS DANS LA DIMENSION INTRA-INDIVIDUELLE (WITHIN)';
run;

    /* regression standard par les MCO */
    /* on ajoute les indicatrices temporelles */
data &transf; merge &transf &tab(keep=&i &t &dtemp); run;
title 'ESTIMATION DANS LA DIMENSION INTRA-INDIVIDUELLE (WITHIN)';
proc reg data=&transf outest=__b;

```

```

        model &y=&x &dtemp/noint;
        output out=&res r=res;
        data __b; set __b(keep=&x &dtemp);
        proc transpose data=__b out=__b(rename=(coll=coeff));
run;
/* Regression avec ecart-types robustes : */
/* on corrige la covariance pour l'heteroscedasticite */
/* ET l'autocorrelation d'ordre quelconque */
proc iml;
use &transf; read all var{&x &dtemp} into x;
use &res; read all var{res} into u;
use __rbet; read all var{res} into n;
use __b; read all var{__NAME_} into nvar;
use __b; read all var{COEFF} into b;
n=nrow(n);
nt=nrow(x); p=ncol(x);
t=int(nt/n);
xox=0; xx=0;
do i=1 to n;
    rg=1+(i-1)#t:i#t;
    xi=x[rg,]; ui=u[rg];
    xox=xox+(t(xi)*ui*t(ui)*xi)/nt;
    xx=xx+(t(xi)*xi)/nt;
end;
free x; /* economie de memoire */
su2=ssq(u)/(nt-p); /* l'ecart type mco (voir plus loin) */
free u; /* economie de memoire */
ixx=solve(xx,i(p));
covar=xmult(xmult(ixx,xox),ixx)/nt;
std=sqrt(vecdiag(covar)); /* ecart-types robustes hetero et
autocorrelation */
stu=abs(b/std); /* Valeur absolue du Student asymptotique */
pro=2#(1-probnorm(stu)); /* Seuil de significativite */
test=std||stu||pro;
noms={'std','stu','pro'};
create __test from test [colname=noms]; append from test;
create &cova from covar[colname=nvar]; append from covar; /* ecrase
l'ancienne valeur */
/* statistiques non robustes a l'heteroscedasticite */
cova=su2#ixx; /* covariance mco, pas bon */
create &vieux from cova[colname=nvar]; append from cova;
quit;

data __test; merge __b __test;
label
    _name_='Régresseur'
    coeff='Coefficient MCO'
    std='Ecart-Type Robuste'
    stu='T de Student Robuste'
    pro='Seuil de Significativité'
;
proc print label; var coeff std stu pro; id _name_;
title2 'Ecart-types robuste a l'autocorrelation et a
l'heteroscedasticite';
run;
data &cova; merge __b &cova;
run;
data &res; set &res; label res='Residu';
run;
data &vieux; merge __b &vieux;
run;

/* Decomposition des variances Between/Within */

proc means noprint data=_between; var &y &x;
output out=_varb(drop=_type_) uss=;

```

```

proc transpose out=_varb;
proc means noprint data=&transf; var &y &x;
output out=_varw(drop=_type_) uss=;
proc transpose out=_varw;
run;

proc iml;
use _varb; read all var{COL1} into vb;
      read all var{_NAME_} into noms;
use _varw; read all var{COL1} into vw;
n=vb[1];
  rg=2:nrow(vb);
  vb=vb[rg];
  noms=noms[rg];
nt=vw[1];
  t=int(nt/n);
  vb=t*vb; /* on duplique la variance between par T car */
           /* elle est calculee sur N obs seulement */
  vw=vw[rg];
  vt=vb+vw;
  pvb=(vb/vt)*100;
pvb=int((pvb+0.005)*100)/100; /* arrondi 2 decimales */
pvw=100-pvb;
res=pvb||pvw||(vt/nt);
etiqh={'% Between * T', '% Within', 'Variance Totale'};
print "Decompositions de la Variance Totale",
      res [colname=etiqh rowname=noms];
quit;

/* sommes des carres des residus */
/* et tests de specification = Fisher et Hausman */
title 'Analyse de la Variance'; run;

proc iml;
start lecture;
  use __rbet; read all var{res} into rbet;
  use &res; read all var{res} into rwit;
  use __cbet; read all into cbet;
  k=nrow(cbet);
  n=nrow(rbet);
  nt=nrow(rwit);
  t=int(nt/n);
  print "Nombre d'individus =" n, "Nombre de Perodes =" t,
        "Nombre d'Observations =" nt;
finish lecture;
/* important : on corrige les degres de liberte pour la presence */
/* obligatoire d'un terme constant dans le modele */
start scr;
mcb=ssq(rbet)/n; /* moyenne des carres between, on ne divise que par N */
dlb=n-k; /* degres de liberte de la somme des carres */
mcw=ssq(rwit)/nt; /* moyenne des carres within, on divise par N*T */
dlw=n*(t-1)-k+1; /* degres de liberte de la somme des carres */
mct=mcb+mcw; /* moyenne des carres totaux */
print
"Moyenne des Carres Between / N =" mcb ", Degres de Liberte =" dlb,
"Moyenne des Carres Within / (N*T) =" mcw ", Degres de Liberte =" dlw;
s2b=mcb*(n/dlb); /* variance regression between sur N obs */
sb=sqrt(s2b); /* ecart-type */
s2w=mcw*(nt/dlw); /* variance regression within sur N*T obs */
sw=sqrt(s2w); /* ecart-type */
s2e=s2w*T/(T-1); /* variance perturbation bruit blanc */
se=sqrt(s2e); /* ecart-type */
s2a=s2b-s2e/T; /* variance effet aleatoire individuel */
if s2a<0 then s2a=0; /* d'apres le maximum vraisemblance */
sa=sqrt(s2a); /* ecart-type */

```

```

    fnum=(T-1)#s2b;      /* avec correction pour l'absence de duplication
*/
    fden=s2w;           /* degres de liberte corriges pour la constante
*/
    fisher=fnum/fden;  proba=1-probf(fisher,dlb,dlw);
print
"Variance Regression Between (sur N obs.) =" s2b ", Ecart-type =" sb,
"Variance Regression Within (sur N*T obs.) =" s2w ", Ecart-type =" sw,
"Variance du bruit blanc =" S2e ", Ecart-type =" se,
"Variance Effet Individuel  =" s2a ", Ecart-type =" sa;
print
"Test d'existence d'un effet aléatoire individuel (Dormont, pp. 70-71) :",
"F(" dlb ", " dlw ") =" fisher,
"Seuil de Significativite =" proba;
    create __s2e from s2e; append from s2e;
    create __s2a from s2a; append from s2a;
    create __t from t; append from t;
    finish scr;

start hausman; /* calcul de la statistique d'Hausman */
                /* (avec covariances robustes) */
    use __cbet; read all into between;
    use &cova; read all into within;
    k=nrow(between);
    rg=2:k;
    between=between[rg,];
    rg=3:(k+1);
    bb=between[,1]; vb=between[,rg];
    k=nrow(bb);
    rg=1:k;
    bw=within[rg,1];
    rg2=2:(k+1);
    vw=within[rg,rg2];
    q=bw-bb; vq=vw+vb;
    hausman=t(q)*inv(vq)*q; proba=1-probchi(hausman,k);
    print "Statistique de Hausman (H0 -> Modele a erreurs composees OK) :",
          "Test de la difference entre within et between",
          "(sans constante ni indicatrices, ecart types robustes)",
          "Khi-deux ( " k ") =" hausman,
          "Seuil de Significativite =" proba,
          ;
    finish hausman;

run lecture;
run scr;
run hausman;

quit;
%mend bwithin;

```

5.4 %bwmcqg

```

%macro
bwmcqg(tab,y,x,i,t,cova=__cmcg,res=__rmcg,vieux=__pmcg,transf=__mcqg,dtemp=,
s=);

    %bwithin(&tab,&y,&x,&i,&t,dtemp=&dtemp,s=&s); run;

proc iml; /* Quasi-differences MCQG */
use __s2e; read all into s2e; /* Variance bruit blanc */
use __s2a; read all into s2a; /* Variance effet aleatoire individuel */
use __t; read all into t;
theta2=s2e/(T#s2a+s2e);
theta=sqrt(theta2);
noms={&y &x &s};

```

```

use _between; read all var{&y &x &s} into by;
use &tab; read all var{&y &x &s} into y;
    et=j(t,1,1); my=y-(1-theta)#(by@et);
    free by;
    create &transf from my[colname=noms]; append from my;
print "THETA CARRE = " theta2,
      "THETA =" theta;
quit;

/* correlations des quasi-differences */
proc corr pearson data=&transf; var &y &x &s;
title 'CORRELATIONS DES QUASI DIFFERENCES (MCQG)';
run;

/* regression standard par les MCO */
title 'ESTIMATION DU MODELE TRANSFORME (MCQG)';
title2 'Il faut diviser le coefficient du terme constant par ''theta'';
data &transf; merge &transf &tab(keep=&dtemp); run;
proc reg data=&transf outest=__b;
    model &y=&x &dtemp;
    output out=&res r=res;
data &transf; set &transf; intercep=1;
%let x=INTERCEP &x &dtemp; run;
data __b; set __b(keep=&x);
proc transpose data=__b out=__b(rename=(coll=coeff));
run;
/* Regression avec ecart-types robustes : */
/* on corrige la covariance pour l'heteroscedasticite */
/* ET l'autocorrelation d'ordre quelconque */
proc iml;
use &transf; read all var{&x} into x;
use &res; read all var{res} into u;
use __t; read all into t;
use __b; read all var{__NAME__} into nvar;
use __b; read all var{COEFF} into b;
nt=nrow(x); p=ncol(x);
n=int(nt/t);
xox=0; xx=0;
do i=1 to n;
    rg=1+(i-1)#t:i#t;
    xi=x[rg,]; ui=u[rg];
    xox=xox+(t(xi)*ui*t(ui)*xi)/nt;
    xx=xx+(t(xi)*xi)/nt;
end;
free x; /* economie de memoire */
su2=ssq(u)/(nt-p); /* l'ecart type mco (voir plus loin) */
free u; /* economie de memoire */
ixx=solve(xx,i(p));
covar=xmult(xmult(ixx,xox),ixx)/nt;
std=sqrt(vecdiag(covar)); /* ecart-type robustes hetero et
autocorrelation */
stu=abs(b/std); /* Student asymptotiques */
pro=2#(1-probnorm(stu)); /* Seuil de significativite */
test=std||stu||pro;
noms={'std','stu','pro'};
create __test from test [colname=noms]; append from test;
create &cova from covar[colname=nvar]; append from covar; /* ecrase
l'ancienne valeur */
/* statistiques non robustes a l'heteroscedasticite */
cova=su2#ixx; /* covariance mco, pas bon */
create &vieux from cova[colname=nvar]; append from cova;
quit;
data __test; merge __b __test;
label
    __name_='Régresseur'
    coeff='Coefficient MCO'
    std='Ecart-Type Robuste'
    stu='T de Student Robuste'

```

```

    pro='Seuil de Significativité'
;
proc print label; var coeff std stu pro; id _name_;
title2 'Ecart-types robuste a l'autocorrelation et a
l'heteroscedasticite';
run;
data &cova; merge __b &cova;
run;
data &res; set &res; label res='Residu';
run;
data &vieux; merge __b &vieux;
run;
%mend bwmcqg;

```

5.5 %momsta

```

%MACRO MOMSTA(TAB,Y,X,I,T,COVA1=__V1,COVA2=__V2,D=2,Z=&X);
/* ESTIMATION PAR LA METHODE DES MOMENTS GENERALISES
*/
*/
/* APPLIQUEE AUX DONNEES DE PANEL SUR UN MODELE EN DIFFERENCES PREMIERES
*/
OPTIONS LINESIZE=100;
/* ON TRIE LE FICHIER */
PROC SORT DATA=&TAB; BY &I &T; RUN;
/* POUR DETERMINER LE NOMBRE DE PERIODES */
PROC FREQ NOPRINT; TABLES &T/OUT=__T(KEEP=&T); RUN;

PROC IML;
START LECTURE; /* LECTURE DES DONNEES */
NOMS={&Y &X}; NOMS=T(NOMS); NX={&X}; NX=T(NX); NY={&Y};
nz={&z}; nz=t(nz);
USE &TAB; READ ALL VAR{&Y} INTO Y; /* VARIABLE EXPLIQUEE */
                READ ALL VAR{&X} INTO X; /* VARIABLES EXPLICATIVES */
                read all var{&z} into zz; /* instruments */
USE __T; READ ALL INTO T; T=NROW(T);
K=NCOL(X); /* NOMBRE DE VARIABLES EXPLICATIVES */
NT=NROW(Y); /* NT = NOMBRE D'OBSERVATIONS */
N=INT(NT/T); /* NOMBRE D'INDIVIDUS */
kz=ncol(zz); /* nombre variables dans les instruments */
P=(KZ)#(T-&D+1)#(T-&D)/2;
PRINT
"VARIABLE EXPLIQUEE =" NY, "VARIABLES EXPLICATIVES =" NX
"NOMBRE DE PERIODES =" T, "NOMBRE D'INDIVIDUS =" N,
"NOMBRE D'OBSERVATIONS =" NT, "NOMBRE DE VARIABLES EXPLICATIVES =" K,
"NOMBRE D'INSTRUMENTS =" P,
"DECALAGE EMPLOYE =" &D;
FINISH LECTURE;

START MATRICES; /* CALCUL DES DIFFERENCES PREMIERES ET DES
INSTRUMENTS*/
/* D = MATRICE DE TRANSFORMATION DU VECTEUR DES OBSERVATIONS D'UN
INDIVIDU */
/* EN DIFFERENCES PREMIERES AVEC ELIMINATION DE LA PREMIERE
OBSERVATION */
D=J(T-&D,&D-1,0) || (-I(T-&D) || J(T-&D,1,0))+(J(T-&D,1,0) || I(T-&D));
DY=J(N#(T-&D),1,0); DX=J(N#(T-&D),K,0);
DO I=1 TO N;
    RG=(I-1)#T+1:I#T; RGD=(I-1)#(T-&D)+1:I#(T-&D);
    DY[RGD]=D*Y[RG];
    DX[RGD,]=D*X[RG,];
END;
/* L2 = MATRICE DE DECALAGE DE DEUX ANS AVEC ELIMINATION DES DEUX
PREMIERES */
/* OBSERVATIONS, PREPARE LES INSTRUMENTS DONT ON AURA BESOIN */
L2=(I(T-&D) || J(T-&D,&D,0));

```

```

LYX=J(N#(T-&D),KZ,0);
DO I=1 TO N;
  RG=(I-1)#T+1:I#T; RGL=(I-1)#(T-&D)+1:I#(T-&D);
  LYX[RGL,]=L2*(ZZ[RG,]);
END;
Z=J(N#(T-&D),P,0);
/* MATRICE DIAGONALE COMPRENANT LES INSTRUMENTS SEPARES EN LIGNE */
DO I=1 TO N;
  RGI=(I-1)#(T-&D)+1:I#(T-&D);
  Z[RGI,]=INSTRU(LYX[RGI,]);
END;
Z=Z; /* MATRICE DES INSTRUMENTS */
FREE LYX X;
/* CALCUL DE LA DIFFERENCE DE L'ENDOGENE DECALEE */
D=(-I(T-&D) || J(T-&D,1,0)+(J(T-&D,1,0) || I(T-&D)) || J(T-&D,&D-1,0);
DY1=J(N#(T-&D),1,0);
DO I=1 TO N;
  RG=(I-1)#T+1:I#T; RGD=(I-1)#(T-&D)+1:I#(T-&D);
  DY1[RGD]=D*Y[RG];
END;
FREE Y;
FINISH MATRICES;

START VECL(M);
M1=NROW(M); M2=NCOL(M);
V=J(1,M1*M2,0);
I0=0;
DO I1=1 TO M1;
  DO I2=1 TO M2;
    I0=I0+1;
    V[I0]=M[I1,I2];
  END; END;
RETURN(V);
FINISH VECL;

START INSTRU(M);
M1=NROW(M); M2=NCOL(M);
W=J(M1,M2#M1#(M1+1)/2,0);
DO I1=1 TO M1;
  RG1=1:I1;
  LIGNE=VECL(M[RG1,]);
  SI=I1#(I1+1)/2;
  RG2=(SI-I1)#M2+1:SI#M2;
  W[I1,RG2]=LIGNE;
END;
RETURN(W);
FINISH INSTRU;

START DMC; /* DOUBLES MOINDRES CARRES */
ZX=0; /* PRODUITS CROISES INSTRUMENTS * DIF VAR EXPLICATIVES */
ZY=0; /* PRODUITS CROISES INSTRUMENTS * DIF VAR EXPLIQUEE */
ZZ=0; /* PRODUITS CROISES DES INSTRUMENTS */
T=T-&D; /* NOMBRE D'ANNEES EFFECTIVEMENT DISPONIBLES POUR LA REGRESSION */
NDUM=COMPRESS(J(T,1,'T_')+CHAR(T(1:T))); /* INDICATRICES TEMPORELLES */
NX=NX//NDUM; /* NOUVEAUX NOMS DE VARIABLES EXPLICATIVES */
NZ=NZ//NDUM;
DX=DX || (J(N,1,1)@I(T)); /* on ajoute les indicatrices */
K=NCOL(DX);
Z=Z || (J(N,1,1)@I(T));
P=NCOL(Z);
DO I=1 TO N;
  RGI=(1+(I-1)#T):(I#T);
  YI=DY[RGI];
  XI=DX[RGI,];
  ZI=Z[RGI,];
  ZX=ZX+(T(ZI)*XI)/N;

```

```

      ZY=ZY+(T(ZI)*YI)/N;
      ZZ=ZZ+(T(ZI)*ZI)/N;
END;
      IZZ=INV(ZZ);
      PX=Z*IZZ*ZX; /* PROJECTIONS EXPLICATIVES SUR INSTRUMENTS */
      XPX=(T(PX)*PX)/N;
      XPY=(T(PX)*DY)/N;
      IXPX=INV(XPX);
      B=IXPX*XPY; /* ESTIMATION DOUBLES MOINDRES CARRES */
      U=DY-DX*B; /* RESIDU DOUBLES MOINDRES CARRES */
      ZOZ=0; /* HETERO ET AUTO CO */
      DO I=1 TO N;
          RGI=(1+(I-1)#T):(I#T);
          UI=U[RGI];
          ZI=Z[RGI,];
          ZOZ=ZOZ+(T(ZI)*UI*T(UI)*ZI)/N;
      END;
      RES=J(NROW(B),4,0); /* MATRICE DES RESULTATS */
      COVROB=(IXPX*T(ZX)*IZZ*ZOZ*IZZ*ZX*IXPX)/N;
      STDROB=SQRT(VECDIAG(COVROB)); /* ECARTS TYPES CORRESPONDANTS */
      STUROB=ABS(B/STDROB); /* STUDENT CORRESPONDANT */
      PROROB=2#(1-PROBNORM(STUROB)); /* SEUIL DE SIGNIF */
      RES[,1]=B;
      RES[,2]=STDROB;
      RES[,3]=STUROB;
      RES[,4]=PROROB;
      PRINT ,,, "SOUS HYPOTHESES D'HETEROSCEDASTICITE ET D'AUTOCORRELATION",
      "(PREMIERE ETAPE)";
      RUN IMPR;
      FINISH DMC;

START OPTIMAL; /* ESTIMATEUR DES MOMENTS GENERALISES */
      PRINT ,,, "ESTIMATION PAR LES MOMENTS GENERALISES",
      "(SECONDE ETAPE)";
      IZOZ=INV(ZOZ);
      GMM=INV(T(ZX)*IZOZ*ZX)*(T(ZX)*IZOZ*ZY);
      U=DY-DX*GMM; /* REDEFINITION RESIDU = GMM */
      ZOZ=0; /* HETERO ET AUTO CO */
      DO I=1 TO N;
          RGI=(1+(I-1)#T):(I#T);
          UI=U[RGI];
          ZI=Z[RGI,];
          ZOZ=ZOZ+(T(ZI)*UI*T(UI)*ZI)/N;
      END;
      COVGMM=INV(T(ZX)*IZOZ*ZX)/N;
      STDGMM=SQRT(VECDIAG(COVGMM)); /* ECARTS TYPES CORRESPONDANTS */
      STUGMM=ABS(GMM/STDGMM); /* STUDENT CORRESPONDANT */
      PROGMM=2#(1-PROBNORM(STUGMM)); /* SEUIL DE SIGNIF */
      RES[,1]=GMM;
      RES[,2]=STDGMM;
      RES[,3]=STUGMM;
      RES[,4]=PROGMM;
      RUN IMPR;
      FINISH OPTIMAL;

START IMPR;
      PRES={'COEFFICIENT','ECART_TYPE','STUDENT','PROBA'};
      PRINT "RESULTAT DE L'ESTIMATION :",
          "VARIABLE EXPLIQUEE : " NY,
          "variables explicatives = " nx "Instruments = " nz,
          "ESTIMATIONS :",
          RES [COLNAME=PRES ROWNAME=NX]
      ;
FINISH IMPR;

START TESTS;

```

```

UZ=(T(U)*Z)/N; /* RESIDU DE SECONDE ETAPE */
SARGAN=N#(UZ*IZOZ*T(UZ));
PROBS=1-PROBCHI(SARGAN,P-K);
PRINT "STATISTIQUE DE SARGAN =" SARGAN, "PROBA =" PROBS;
FINISH TESTS;

/* APPEL DES MODULES */
RUN LECTURE;
RUN MATRICES;
RUN DMC;
RUN OPTIMAL;
RUN TESTS;
/* SAUVEGARDE DES DMC AVEC ECARTS TYPES ROBUSTES */
/* AINSI QUE L'ESTIMATEUR DES MOMENTS GENERALISES CORRESPONDANT */

B=T(B); /* ECRITURE EN LIGNE POUR LA SAUVEGARDE, TRANSPOSITION PLUS LOIN */
CREATE __B1 FROM B[COLNAME=NX]; APPEND FROM B;
CREATE &COVA1 FROM COVROB[COLNAME=NX]; APPEND FROM COVROB;
GMM=T(GMM); /* MEME REMARQUE */
CREATE __B2 FROM GMM[COLNAME=NX]; APPEND FROM GMM;
CREATE &COVA2 FROM COVGMM[COLNAME=NX]; APPEND FROM COVGMM;
QUIT;

PROC TRANSPOSE DATA=__B1 OUT=__B1(RENAME=(COL1=COEFF)); RUN;
PROC TRANSPOSE DATA=__B2 OUT=__B2(RENAME=(COL1=COEFF)); RUN;
DATA &COVA1; MERGE __B1 &COVA1; RUN;
DATA &COVA2; MERGE __B2 &COVA2; RUN;
%MEND MOMSTA;

```

5.6 %momdyn

```

%MACRO MOMDYN(TAB,Y,X,I,T,COVA1=__V1,COVA2=__V2,D=2,Z=&X);
/* ESTIMATION PAR LA METHODE DES MOMENTS GENERALISES
*/
/* APPLIQUEE AUX DONNEES DE PANEL SUR UN MODELE EN DIFFERENCES PREMIERES
*/
OPTIONS LINESIZE=100;
/* ON TRIE LE FICHER */
PROC SORT DATA=&TAB; BY &I &T; RUN;
/* POUR DETERMINER LE NOMBRE DE PERIODES */
PROC FREQ NOPRINT; TABLES &T/OUT=__T(KEEP=&T); RUN;

PROC IML;
START LECTURE; /* LECTURE DES DONNEES */
NOMS={&Y &X}; NOMS=T(NOMS); NX={&X}; NX=T(NX); NY={&Y};
nz={&z}; nz=t(nz);
USE &TAB; READ ALL VAR{&Y} INTO Y; /* VARIABLE EXPLIQUEE */
READ ALL VAR{&X} INTO X; /* VARIABLES EXPLICATIVES */
read all var{&z} into zz; /* instruments */
USE __T; READ ALL INTO T; T=NROW(T);
K=NCOL(X); /* NOMBRE DE VARIABLES EXPLICATIVES */
NT=NROW(Y); /* NT = NOMBRE D'OBSERVATIONS */
N=INT(NT/T); /* NOMBRE D'INDIVIDUS */
kz=ncol(zz); /* nombre variables dans les instruments */
P=(KZ)#(T-&D+1)#(T-&D)/2;
PRINT
"VARIABLE EXPLIQUEE =" NY, "VARIABLES EXPLICATIVES =" NX
"NOMBRE DE PERIODES =" T, "NOMBRE D'INDIVIDUS =" N,
"NOMBRE D'OBSERVATIONS =" NT, "NOMBRE DE VARIABLES EXPLICATIVES =" K,
"NOMBRE D'INSTRUMENTS =" P,
"DECALAGE EMPLOYE =" &D;
FINISH LECTURE;

START MATRICES; /* CALCUL DES DIFFERENCES PREMIERES ET DES
INSTRUMENTS*/

```

```

/* D = MATRICE DE TRANSFORMATION DU VECTEUR DES OBSERVATIONS D'UN
INDIVIDU */
/* EN DIFFERENCES PREMIERES AVEC ELIMINATION DE LA PREMIERE
OBSERVATION */
D=J(T-&D,&D-1,0) || (-I(T-&D) || J(T-&D,1,0))+(J(T-&D,1,0) || I(T-&D));
DY=J(N#(T-&D),1,0); DX=J(N#(T-&D),K,0);
DO I=1 TO N;
  RG=(I-1)#T+1:I#T; RGD=(I-1)#(T-&D)+1:I#(T-&D);
  DY[RGD]=D*Y[RG];
  DX[RGD,]=D*X[RG,];
END;
/* L2 = MATRICE DE DECALAGE DE DEUX ANS AVEC ELIMINATION DES DEUX
PREMIERES */
/* OBSERVATIONS, PREPARE LES INSTRUMENTS DONT ON AURA BESOIN */
L2=(I(T-&D) || J(T-&D,&D,0));
LYX=J(N#(T-&D),KZ,0);
DO I=1 TO N;
  RG=(I-1)#T+1:I#T; RGL=(I-1)#(T-&D)+1:I#(T-&D);
  LYX[RGL,]=L2*(ZZ[RG,]);
END;
Z=J(N#(T-&D),P,0);
/* MATRICE DIAGONALE COMPRENANT LES INSTRUMENTS SEPARES EN LIGNE */
DO I=1 TO N;
  RGI=(I-1)#(T-&D)+1:I#(T-&D);
  Z[RGI,]=INSTRU(LYX[RGI,]);
END;
Z=Z; /* MATRICE DES INSTRUMENTS */
FREE LYX X;
/* CALCUL DE LA DIFFERENCE DE L'ENDOGENE DECALEE */
D=(-I(T-&D) || J(T-&D,1,0))+(J(T-&D,1,0) || I(T-&D)) || J(T-&D,&D-1,0);
DY1=J(N#(T-&D),1,0);
DO I=1 TO N;
  RG=(I-1)#T+1:I#T; RGD=(I-1)#(T-&D)+1:I#(T-&D);
  DY1[RGD]=D*Y[RG];
  END;
DX=DY1 || DX;
NX=COMPRESS('LAG_'+NY)//NX;
K=K+1;
FREE Y;
FINISH MATRICES;

START VECL(M);
M1=NROW(M); M2=NCOL(M);
V=J(1,M1*M2,0);
I0=0;
DO I1=1 TO M1;
  DO I2=1 TO M2;
    I0=I0+1;
    V[I0]=M[I1,I2];
  END; END;
RETURN(V);
FINISH VECL;

START INSTRU(M);
M1=NROW(M); M2=NCOL(M);
W=J(M1,M2#M1#(M1+1)/2,0);
DO I1=1 TO M1;
  RG1=1:I1;
  LIGNE=VECL(M[RG1,]);
  SI=I1#(I1+1)/2;
  RG2=(SI-I1)#M2+1:SI#M2;
  W[I1, RG2]=LIGNE;
END;
RETURN(W);
FINISH INSTRU;

```

```

START DMC; /* DOUBLES MOINDRES CARRES */
ZX=0; /* PRODUITS CROISES INSTRUMENTS * DIF VAR EXPLICATIVES */
ZY=0; /* PRODUITS CROISES INSTRUMENTS * DIF VAR EXPLIQUEE */
ZZ=0; /* PRODUITS CROISES DES INSTRUMENTS */
T=T-&D; /* NOMBRE D'ANNEES EFFECTIVEMENT DISPONIBLES POUR LA REGRESSION */
NDUM=COMPRESS(J(T,1,'T_')+CHAR(T(1:T))); /*INDICATRICES TEMPORELLES */
NX=NX//NDUM; /* NOUVEAUX NOMS DE VARIABLES EXPLICATIVES */
DX=DX|(J(N,1,1)@I(T));
K=NCOL(DX);
NZ=nz//ndum;
Z=Z|(J(N,1,1)@I(T));
P=NCOL(Z);
DO I=1 TO N;
  RGI=(1+(I-1)#T):(I#T);
  YI=DY[RGI];
  XI=DX[RGI,];
  ZI=Z[RGI,];
  ZX=ZX+(T(ZI)*XI)/N;
  ZY=ZY+(T(ZI)*YI)/N;
  ZZ=ZZ+(T(ZI)*ZI)/N;
END;
IZZ=INV(ZZ);
PX=Z*IZZ*ZX; /* PROJECTIONS EXPLICATIVES SUR INSTRUMENTS */
XPX=(T(PX)*PX)/N;
XPY=(T(PX)*DY)/N;
IXPX=INV(XPX);
B=IXPX*XPY; /* ESTIMATION DOUBLES MOINDRES CARRES */
U=DY-DX*B; /* RESIDU DOUBLES MOINDRES CARRES */
ZOZ=0; /* HETERO ET AUTOOCO */
DO I=1 TO N;
  RGI=(1+(I-1)#T):(I#T);
  UI=U[RGI];
  ZI=Z[RGI,];
  ZOZ=ZOZ+(T(ZI)*UI*T(UI)*ZI)/N;
END;
RES=J(NROW(B),4,0); /* MATRICE DES RESULTATS */
COVROB=(IXPX*T(ZX)*IZZ*ZOZ*IZZ*ZX*IXPX)/N;
STDROB=SQRT(VECDIAG(COVROB)); /* ECARTS TYPES CORRESPONDANTS */
STUROB=ABS(B/STDROB); /* STUDENT CORRESPONDANT */
PROROB=2#(1-PROBNORM(STUROB)); /* SEUIL DE SIGNIF */
RES[,1]=B;
RES[,2]=STDROB;
RES[,3]=STUROB;
RES[,4]=PROROB;
PRINT ,,, "SOUS HYPOTHESES D'HETEROSCEDASTICITE ET D'AUTOCORRELATION",
" (PREMIERE ETAPE) ";
RUN IMPR;
FINISH DMC;

START OPTIMAL; /* ESTIMATEUR DES MOMENTS GENERALISES */
PRINT ,,, "ESTIMATION PAR LES MOMENTS GENERALISES",
" (SECONDE ETAPE) ";
IZOZ=INV(ZOZ);
GMM=INV(T(ZX)*IZOZ*ZX)*(T(ZX)*IZOZ*ZY);
U=DY-DX*GMM; /* REDEFINITION RESIDU = GMM */
ZOZ=0; /* HETERO ET AUTOOCO */
DO I=1 TO N;
  RGI=(1+(I-1)#T):(I#T);
  UI=U[RGI];
  ZI=Z[RGI,];
  ZOZ=ZOZ+(T(ZI)*UI*T(UI)*ZI)/N;
END;
COVGMM=INV(T(ZX)*IZOZ*ZX)/N;
STDGMM=SQRT(VECDIAG(COVGMM)); /* ECARTS TYPES CORRESPONDANTS */
STUGMM=ABS(GMM/STDGMM); /* STUDENT CORRESPONDANT */
PROGMM=2#(1-PROBNORM(STUGMM)); /* SEUIL DE SIGNIF */

```

```

RES[,1]=GMM;
RES[,2]=STDGMM;
RES[,3]=STUGMM;
RES[,4]=PROGMM;
RUN IMPR;
    FINISH OPTIMAL;

START IMPR;
PRES={'COEFFICIENT','ECART_TYPE','STUDENT','PROBA'};
PRINT "RESULTAT DE L'ESTIMATION :",
      "VARIABLE EXPLIQUEE :" NY,
      "variables explicatives =" nx "Instruments =" nz,
      "ESTIMATIONS :",
      RES [COLNAME=PRES ROWNAME=NX]
      ;
FINISH IMPR;

START TESTS;
UZ=(T(U)*Z)/N; /* RESIDU DE SECONDE ETAPE */
SARGAN=N#(UZ*IZOZ*T(UZ));
PROBS=1-PROBCHI(SARGAN,P-K);
PRINT "STATISTIQUE DE SARGAN =" SARGAN, "PROBA =" PROBS;
FINISH TESTS;

/* APPEL DES MODULES */
RUN LECTURE;
RUN MATRICES;
RUN DMC;
RUN OPTIMAL;
RUN TESTS;
/* SAUVEGARDE DES DMC AVEC ECARTS TYPES ROBUSTES */
/* AINSI QUE L'ESTIMATEUR DES MOMENTS GENERALISES CORRESPONDANT */
B=T(B); /* ECRITURE EN LIGNE POUR LA SAUVEGARDE, TRANSPOSITION PLUS LOIN */
CREATE __B1 FROM B[COLNAME=NX]; APPEND FROM B;
CREATE &COVA1 FROM COVROB[COLNAME=NX]; APPEND FROM COVROB;
GMM=T(GMM); /* MEME REMARQUE */
CREATE __B2 FROM GMM[COLNAME=NX]; APPEND FROM GMM;
CREATE &COVA2 FROM COVGMM[COLNAME=NX]; APPEND FROM COVGMM;
QUIT;
PROC TRANSPOSE DATA=__B1 OUT=__B1(RENAME=(COL1=COEFF)); RUN;
PROC TRANSPOSE DATA=__B2 OUT=__B2(RENAME=(COL1=COEFF)); RUN;
DATA &COVA1; MERGE __B1 &COVA1; RUN;
DATA &COVA2; MERGE __B2 &COVA2; RUN;
%MEND MOMDYN;

```

5.7 %multino

```

%MACRO MULTINO(TAB,ENDO,EXOG,IMPR=0,ARRET=1E-8,MAXITER=100,
MAXMOD=50,SORB=A,SORI=A,SCORE=A,METHODE=NRA,TRONC=15,
INF=1E-8,IDENT=0,INB=A,OUTB=A,EPSILON=1E-8);
option linesize=110 pagesize=65;
PROC SORT DATA=&TAB; BY &ENDO;
DATA __TAB; SET &TAB(KEEP=&EXOG); IF _N_=1;
PROC TRANSPOSE DATA=__TAB OUT=__TAB(KEEP=_NAME_);
PROC FREQ DATA=&TAB;
    TABLES &ENDO/OUT=__STAT;
PROC IML;
    START LECTURE;
        /*****/;
        /* LECTURE DES VARIABLES */;
        /*****/;
        /* NY : NOM DE LA VARIABLE ENDO}NE */;
        /* NOMS: NOMS DES VARIABLES EXOG}NES */;
        /* Y : VECTEUR DE LA VARIABLE ENDO}NE */;
        /* X : MATRICE DES VARIABLES EXOG}NES */;
        /*****/;

```

```

INF={&INF};
NY={&ENDOG};
INB={&INB}; INB=COMPRESS(INB);
OUTB={&OUTB}; OUTB=COMPRESS(OUTB);
USE __TAB;
  READ ALL VAR { _NAME_ } INTO NOMS;
  PRINT "VARIABLE EXPLIQUEE :" NY;
  PRINT "VARIABLES EXPLICATIVES :" NOMS;
USE &TAB;
  READ ALL VAR(NY) INTO Y;
  READ ALL VAR(NOMS) INTO X;
USE __STAT;
  READ ALL VAR(NY) INTO MODALITE;
  READ ALL VAR { COUNT } INTO COMPTE;
  NMOD=NROW(COMPTE);
  N=NROW(X); KK=NCOL(X);
FINISH LECTURE;

START DEBUT;
  /*****/;
  /* INITIALISATION GENERALE DU PROGRAMME */;
  /*****/;
TITRES={ COEFFICIENT, ECART_TYPE, T_DE_STUDENT, WALD, PROB };
EPSILON=&EPSILON;
INFINI=1/EPSILON;
RUN LECTURE;
NM=NMOD-1; K=NM*KK;
LAR=NM/10;
  IF LAR<1 THEN LAR=7; ELSE LAR=6;
  LL=8-LAR;
NOMS2=J(K,1,' ');
NOMS3=J(K,1,' ');
DO I=2 TO NMOD; RGI=((I-2)#KK+1):(I-1)#KK;
  NOMS2[RGI]=J(KK,1,'MODALITE_'+CHAR(I,LL,0)+':')+NOMS;
  NOMS3[RGI]=COMPRESS(SUBSTR(NOMS,1,LAR)+J(KK,1,CHAR(I,LL,0)));
END;
RUN INITIAL;
FINISH DEBUT;

START MATB(B,NVAR,NMOD);
  C=J(NVAR,NMOD-1,0);
  DO M=2 TO NMOD;
  RG=((M-2)#NVAR+1):(M-1)#NVAR;
  C[,M-1]=B[RG];
  END;
RETURN (C);
FINISH MATB;

START INITIAL;
B0=J(K,1,0);
  /*****/;
  /* RENUMEROTATION DE Y */;
  /*****/;
  YY=J(N,1,0);
  DO I=1 TO N;

  DO MOD=1 TO NMOD;
  IF Y[I]=MODALITE[MOD] THEN YY[I]=MOD;
  END;

  END; Y=YY; YY=0;

IF INB='A' THEN DO;

  DO MOD=2 TO NMOD;

```

```

DUM=J(N,1,0);
DO I=1 TO N;
  IF Y[I]=MOD THEN DUM[I]=1;
END;

B=GINV(X)*DUM;
B=B/SQRT(SSQ(DUM-X*B)/N);
RG=((MOD-2)#KK+1):((MOD-1)#KK);
B0[RG]=B;
END;

END;
      ELSE DO;
        USE &INB; READ ALL INTO B0;
        END;
B=B0;
FINISH INITIAL;

START INIT;
  /****** /;
  /* INITIALISATION DE L'ALGORITHME DE MAXIMISATION */;
  /****** /;
  /* LV : VALEUR COURANTE DE LA LOG-VRAISEMBLANCE */;
  /* LVB: VARIABLE DE STOCKAGE DE LA LOG-VRAISEMBLANCE */;
  /* B : VALEUR COURANTE DE L'ESTIMATION */;
  /* B1 : VALEUR EXPLORATOIRE DE L'ESTIMATION */;
  /****** /;
LVB=0; LV=1; B=B0; B1=B0; RES=J(KK,5,0);
TITER1={ ITERATION,PAS,LOG_VRAISEMBLANCE };
TITER2={ COEFFICIENT,ELASTICITE,GRADIENT };
FINISH;

START NEWTON;
  /****** /;
  /* ITERATIONS SELON LA METHODE DE NEWTON-RAPHSON */;
  /****** /;
  /* LAMBDA : PARAMETRE D'AJUSTEMENT DE L'ITERATION */;
  /****** /;
LAMBDA=1;ELAS=J(K,1,1);
DO ITER=1 TO &MAXITER WHILE ( (LAMBDA>0)*(MAX(ABS(ELAS))>&ARRET) ); /*
iteration */
  LVB=LV; RUN CLV; RUN &METHODE; DELTA=GINV(D2LV)*D1LV;
  LAMBDA=1;
  RLV=LV-1; TESTS=1;
  DO J=1 TO &MAXMOD WHILE (TESTS>0); /* sous-iteration */
    B1=B-LAMBDA#DELTA; RUN RECLV;
    IF RLV<LV THEN TESTS=1; ELSE TESTS=0;
    IF TESTS>0 THEN LAMBDA=LAMBDA/2*(1-INT(J/&MAXMOD));
  END; /* sous-iteration */
  B=B-LAMBDA#DELTA;
  /* B6=B;RUN GRADNUM; B=B6; */
  ELAS=B#D1LV/RLV;
  IF &IMPR=1 THEN DO; /* impression */
    RES=ITER||LAMBDA||RLV;
    PRINT RES [COLNAME=TITER1],;
    RES=B||ELAS||D1LV;
    PRINT RES [COLNAME=TITER2 ROWNAME=NOMS2],,;
  END; /* impression */
END; /* iteration */
IF LAMBDA=0 THEN NIT=ITER-3; ELSE NIT=ITER-1;

IF NIT=&MAXITER THEN PRINT "L'ALGORITHME N'A PAS CONVERGE EN" NIT
"ITERATIONS";
IF OUTB='A' THEN PRINT "PAS DE SAUVEGARDE DEMANDEE...";
ELSE DO;
  CREATE &OUTB FROM B; APPEND FROM B;

```

```

PRINT "LA VALEUR DU PARAMETRE EST SAUVEGARDEE DANS LE TABLEAU SAS
&OUTB";
END;
IF NIT<&MAXITER THEN DO;
PRINT "L'ALGORITHME A CONVERGE EN" NIT "ITERATIONS";
IF LAMBDA=0 THEN PRINT "LE PAS DE LA DERNIERE ITERATION EST NUL",
"LES RESULTATS PEUVENT DONC SE REVELER FANTASISTES",;
PRINT "ELASTICITES DE L'OBJECTIF PAR RAPPORT AUX COEFFICIENTS :",
NOMS2 ELAS;
PRINT "LE GRADIENT EST EGAL A : ", NOMS2 D1LV;
RUN NRA;
VPH=EIGVAL(D2LV);
PRINT "ET LES VALEURS PROPRES DU HESSIEN ANALYTIQUE A : " , VPH,;
VAS=-D2LV;
PRINT "VARIABLE EXPLIQUEE : " NY;
PRINT "NOMBRE DE MODALITES : " NMOD;
PRINT "NOMBRE D'OBSERVATIONS : " N;
PRINT "NOMBRE DE VARIABLES : " KK;
PRINT "NOMBRE DE PARAMETRES : " K;
PRINT "MAXIMUM DE LOG VRAISEMBLANCE : " RLV,;
RES=J(K,5,0);
RES[,1]=B;
RES[,2]=SQRT(VECDIAG(VAS));
RES[,3]=RES[,1]/RES[,2];
RES[,4]=RES[,3]##2;
RES[,5]=1-PROBCHI(RES[,4],1);
PRINT "RESULTATS DE L'ESTIMATION DU MODELE LOGIT POLYTOMIQUE",
"PAR LE MAXIMUM DE VRAISEMBLANCE",
"(ECARTS-TYPE BASES SUR LES DERIVEES SECONDES)",;
PRINT RES [ROWNAME=NOMS2 COLNAME=TITRES],,,;
END;
FINISH;

START CLV;
/*****/;
/* CALCUL DE LA LOG-VRAISEMBLANCE */;
/*****/;
PRO=J(N,NMOD,0);
C=MATB(B,KK,NMOD);
PRO[,1]=J(N,1,1);
DO M=2 TO NMOD;
XC=X*C[,M-1];
IF XC>&TRONC THEN XC=&TRONC;
PRO[,M]=EXP(XC);
END;
PRO=PRO#(1/PRO[,+]);
PRK=J(N,1,0);
DO I=1 TO N;
PR=PRO[I,Y[I]];
IF PR<=INF THEN PR=INF;
IF PR>=1-INF THEN PR=1-INF;
PRK[I]=PR;
END;
LV=SUM(LOG(PRK));
FINISH;

START RECLV;
PRO=J(N,NMOD,0);
C=MATB(B1,KK,NMOD);
PRO[,1]=J(N,1,1);
DO M=2 TO NMOD;
XC=X*C[,M-1];
IF XC>&TRONC THEN XC=&TRONC;
PRO[,M]=EXP(XC);
END;

```

```

PRO=PRO#(1/PRO[,+]);
PRK=J(N,1,0);
DO I=1 TO N;
PR=PRO[I,Y[I]];
  IF PR<=INF THEN PR=INF;
  IF PR>=1-INF THEN PR=1-INF;
PRK[I]=PR;
END;
RLV=SUM(LOG(PRK));
FINISH;

START B3H;
D1LV=J(K,1,0); D1LVI=J(N,K,0);
D2LV=J(K,K,0); RG=2:NMOD;
DO I=1 TO N;
XI=T(X[I,]); PRI=T(PRO[I, RG]); YI=Y[I];
P=-PRI; IF YI>=2 THEN P[YI-1]=1-PRI[YI-1];
DLI=P@XI;
D1LV=D1LV+DLI;
D1LVI[I,]=T(DLI);
END;
D2LV=-T(D1LVI)*D1LVI;
FINISH B3H;

START NRA;
D1LV=J(K,1,0);
D2LV=J(K,K,0); RG=2:NMOD;
DO I=1 TO N;
XI=T(X[I,]); PRI=T(PRO[I, RG]); YI=Y[I];
P=-PRI; IF YI>=2 THEN P[YI-1]=1-PRI[YI-1];
DLI=P@XI;
D1LV=D1LV+P@XI;
  /* D2LV=D2LV-(PRI@XI)*(T(1-PRI)@T(XI)); */
D2LV=D2LV+(PRI*T(PRI)-DIAG(PRI))@(XI*T(XI));
END;
FINISH NRA;

START TITRE;
/*          */;
/* PRESENTATION */;
/*          */;
PRINT "ESTIMATION DU MODELE LOGIT MULTINOMIAL";
PRINT "ALGORITHME : &METHODE";
PRINT "M{THODE : MAXIMUM DE VRAISEMBLANCE";
PRINT "VALEURS INITIALES : " NOMS2 B;
IF INB='A' THEN PRINT "(NON CONVERGENTES)";
ELSE PRINT "(DU TABLEAU SAS &INB)";
RUN CLV;
PRINT "LOG-VRAISEMBLANCE AU POINT INITIAL =" LV;
FINISH;

START SORTIES;
IF %LENGTH(&SORB)>1 THEN DO;
RUN B3H; II=INV(II);
  MAT=J(1,K,N)//J(1,K,RLV)//
  T(RES[,1])//VMV//II//VPMV;
  LAB1={ NOBS, LV, B };
  LAB1=LAB1//NOMS3//NOMS3//NOMS3;
  LAB2=J(2,1,'GLOBAL')//J(1,1,'PARAMETRE')//
  J(K,1,'INV(A)')//J(K,1,'INV(B)')//J(K,1,'INV(A)*B*INV(A)');
  LAB3=J(NROW(LAB2),1,'%MULTINO2');
  LABR=LAB3||LAB1||LAB2;
  CREATE &SORB FROM MAT [COLNAME=NOMS3];
  APPEND FROM MAT;
  CREATE __A FROM LABR;
  APPEND FROM LABR;

```

```

      END;
IF %LENGTH(&SORI)>1 THEN DO;
  RUN CLV;
  MAT=(Y||LOG(PRK));
  LAB1={ Y,LVI };
  CREATE &SORI FROM MAT [COLNAME=LAB1];
  APPEND FROM MAT;
  END;
IF %LENGTH(&SCORE)>1 THEN DO;
  RUN B3H;
  CREATE &SCORE FROM D1LVI [COLNAME=NOMS3];
  APPEND FROM D1LVI;
  END;
FINISH SORTIES;
  /****** */;
  /* PROGRAMME PRINCIPAL */;
  /****** */;
RUN DEBUT;
  RUN INIT;
  RUN TITRE;
  RUN NEWTON;
  RUN SORTIES;
QUIT;
  %IF %LENGTH(&SORB)>1 %THEN %DO;
  DATA &SORB;
  MERGE ___A(RENAME=(COL1=MACRO COL2=VARIABLE COL3=TYPE)) &SORB;
  %END;
  %IF %LENGTH(&SORI)>1 %THEN %DO;
  DATA &SORI;
  SET &SORI;
  LABEL
    LVI='LOG-VRAISEMBLANCE INDIVIDUELLE'
    Y='VARIABLE ENDOGENE (RECODAGE A PARTIR DE 1)'
    ;
  %END;
%MEND MULTINO;
RUN;

```

