



# Une macro-commande SAS pour le modèle Probit à effets aléatoires

Nicolas Moreau

## ► To cite this version:

Nicolas Moreau. Une macro-commande SAS pour le modèle Probit à effets aléatoires. 2018. hal-01691529

**HAL Id: hal-01691529**

**<https://hal.science/hal-01691529>**

Preprint submitted on 24 Jan 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GREQAM

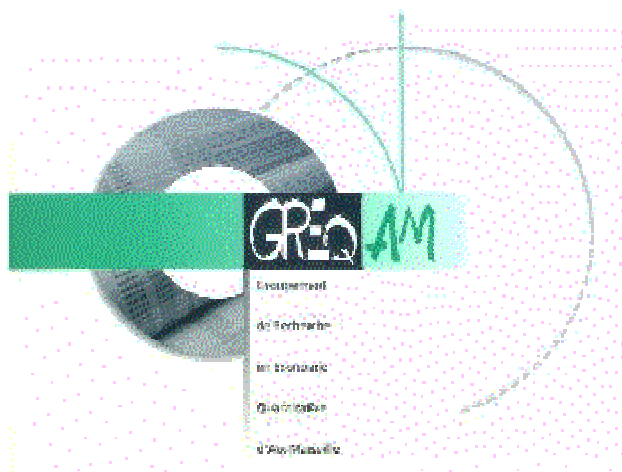
Groupement de Recherche en Economie  
Quantitative d'Aix-Marseille - UMR-CNRS 6579  
Ecole des Hautes Etudes en Sciences Sociales  
Universités d'Aix-Marseille II et III

Document de Travail

n°00B09

## UNE MACRO-COMMANDE SAS POUR LE MODÈLE PROBIT À EFFETS ALÉATOIRES

*N. MOREAU*



DT-GREQAM

# Une macro-commande SAS pour l'estimation du modèle probit à effets aléatoires individuels

Nicolas Moreau\*

---

\*Moreau: GREQAM, Université de la Méditerranée, 2 rue de la Charité, 13002 Marseille. e-mail: moreau@ehess.cnrs-mrs.fr. Je tiens à remercier S. Lollivier pour ses suggestions. Bien entendu, je reste seul responsable des erreurs et imprécisions pouvant subsister. La macro-commande “Probit” exposée dans ce document est disponible auprès de l’auteur.

## **Abstract**

In this paper we supply the reader with a SAS macro-command which computes estimation of the probit individual random effects model.

## **Résumé**

Dans ce document nous fournissons une macro-commande SAS permettant le calcul du modèle probit à effets aléatoires individuels.

**Keywords:** Panel Data, probit, random effects, SAS

**Mots-clés:** Données de panel, probit, effets aléatoires, SAS

**JEL code:** C23, C80

## 1. Introduction

Dans ce document nous fournissons un programme informatique, sous forme d'une macro-commande SAS, qui permet de calculer facilement le modèle probit à effets aléatoires individuels. Nous proposons des estimations à 5, 8, 9, 12, 16 et 20 points d'évaluation. Celles-ci peuvent être menées sur panel cylindré ou non cylindré.

Les effets marginaux, à la moyenne de l'échantillon, sont aussi calculés ainsi que les inverses des ratios de Mill généralisés aux données de panel (Verbeek et Nijman ; 1992, 1996).

Le modèle et sa procédure d'estimation sont brièvement présentés en section 2. Un exemple d'application de la macro-commande "probit" est donné en section 3. Le programme d'instructions est exposé en section 4. Quelques explications sont alors données sur les principales lignes de programmation. La section 5 conclut le texte.

## 2. Brefs rappels sur le modèle probit à effets aléatoires individuels

Nous présentons successivement le modèle, sa procédure d'estimation, le calcul des effets marginaux et le calcul des inverses des ratios de Mill.

### 2.1. Le modèle

$$\begin{aligned} y_{it}^* &= x_{it}\gamma + \xi_i + \eta_{it}, & i &= 1, \dots, n, & t &= 1, \dots, T_i \\ y_{it} &= 1 \text{ si } y_{it}^* > 0, \\ y_{it} &= 0 \text{ sinon} \end{aligned}$$

avec

$$\begin{pmatrix} \xi_i \\ \eta_{it} \end{pmatrix} \rightsquigarrow N \left( 0, \begin{pmatrix} \sigma_\xi^2 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

De l'hypothèse de normalité et de non corrélation de l'effet individuel et de l'aléa il découle que

$$\xi_i + \eta_{it} \rightsquigarrow N(0, \sigma_\xi^2 + 1)$$

avec

$$\text{Corr}(\xi_i + \eta_{it}, \xi_i + \eta_{is}) = \rho = \frac{\sigma_\xi^2}{1 + \sigma_\xi^2}.$$

Le coefficient de corrélation  $\rho$  mesure donc la part de la variance du terme d'erreur  $\xi_i + \eta_{it}$  due à la variance de l'effet individuel.

### 2.2. Son estimation

Soit  $\varepsilon_{it} = \xi_i + \eta_{it}$  et  $r_{it} = 2y_{it} - 1$ . Alors,

$$P(Y_{it} = y_{it}) = \int_{-\infty}^{r_{it}x_{it}\gamma} f(\varepsilon_{it})d\varepsilon_{it} = \Phi(r_{it}x_{it}\gamma).$$

La contribution de l'individu  $i$  à la vraisemblance s'écrit

$$L_i = P(Y_{i1} = y_{i1}, \dots, Y_{iT_i} = y_{iT_i}) = \int_{-\infty}^{r_{i1}x_{i1}\gamma} \int_{-\infty}^{r_{i2}x_{i2}\gamma} \dots \int_{-\infty}^{r_{iT_i}x_{iT_i}\gamma} f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) d\varepsilon_{iT_i} \dots d\varepsilon_{i2} d\varepsilon_{i1}.$$

Or

$$f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) = \int_{-\infty}^{+\infty} f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}/\xi_i) f(\xi_i) d\xi_i$$

et conditionnellement à  $\xi_i$ , les  $\varepsilon_{it}$  sont indépendants de sorte que

$$f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) = \int_{-\infty}^{+\infty} \left[ \prod_{t=1}^{T_i} f(\varepsilon_{it}/\xi_i) \right] f(\xi_i) d\xi_i.$$

Suivant Greene (1995), faisons le changement de variable suivant

$$h_i = \frac{\xi_i}{\sqrt{2}}.$$

On peut alors montrer que

$$L_i = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-h_i^2} \left[ \prod_{t=1}^{T_i} \Phi(r_{it}w_{it}) \right] dh_i$$

où

$$w_{it} = x_{it}\gamma + \theta h_i \text{ avec } \theta = \sqrt{\frac{2\rho}{1-\rho}}.$$

La contribution de l'individu  $i$  à la vraisemblance est de la forme

$$\int_{-\infty}^{+\infty} e^{-Z^2} f(z) dz.$$

Ce type d'expressions s'intègre généralement numériquement par la méthode de quadrature gaussienne

$$\sum_{j=1}^J w_j f(z_j)$$

où  $J$  est le nombre de points d'évaluation,  $w_j$  le poids accordé au  $j$ ème point d'évaluation et  $f(z_j)$  la valeur de  $f(Z)$  au  $j$ ème point de  $Z$ .

La macro-commande "probit" propose six qualités d'approximation, la première à cinq points, la deuxième à huit points, la troisième à neuf, la quatrième à douze, la cinquième à seize et la dernière à vingt points.

La log-vraisemblance s'écrit

$$l = \sum_{i=1}^n \log L_i$$

Pour calculer le gradient nous avons besoin des contributions individuelles  $G_i$ , définies par

$$G_i \equiv \frac{\partial \log L_i}{\partial \alpha} = \frac{\partial L_i / \partial \alpha}{L_i} \text{ avec } \alpha = (\gamma, \theta).$$

Or

$$\frac{\partial L_i}{\partial \alpha} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-h_i^2} \left[ \sum_{t=1}^{T_i} \left( r_{it} \frac{\phi(r_{it} w_{it})}{\Phi(r_{it} w_{it})} \left[ \prod_{t=1}^{T_i} \Phi(r_{it} w_{it}) \right] (x_{it}, h_i) \right) \right] dh_i,$$

expression de la forme

$$\int_{-\infty}^{+\infty} e^{-Z^2} f(z) dz.$$

Le gradient sera approximé numériquement par la procédure de quadrature gaussienne.

La matrice de covariance estimée des paramètres est obtenue à partir de la matrice d'information, soit

$$\text{Cov}(\alpha) = \left[ \sum_{i=1}^n G_i G_i' \right]^{-1}.$$

Le modèle étant non linéaire, la seule lecture des coefficients estimés ne suffit pas à déterminer l'effet marginal d'un régresseur sur la régressande. Il est cependant possible de calculer ces effets marginaux.

### 2.3. Le calcul des effets marginaux

Partons de l'écriture de l'espérance de la variable dépendante observée  $y$ ,

$$E(y_{it}/X_{it} = x_{it}) = P(y_{it}^* > 0) = P(x_{it}\gamma + \xi_i + \eta_{it} > 0) = P(\xi_i + \eta_{it} > -x_{it}\gamma).$$

Comme

$$\xi_i + \eta_{it} \rightsquigarrow N(0, \sigma_\xi^2 + 1),$$

l'espérance conditionnelle s'écrit

$$E(y_{it}/X_{it} = x_{it}) = \Phi\left(\frac{x_{it}\gamma}{\sigma^*}\right)$$

où  $\Phi(\cdot)$  est la loi normale centrée réduite et  $\sigma^* = \sqrt{\sigma_\xi^2 + 1}$ .

L'effet marginal  $f(\gamma)$  vaut donc

$$f(\gamma) \equiv \frac{\partial E(y_{it}/X_{it} = x_{it})}{\partial X_{it}} = \phi\left(\frac{x_{it}\gamma}{\sigma^*}\right) \frac{\gamma}{\sigma^*}$$

avec  $\phi(\cdot)$  est la fonction de densité de la loi normale centrée réduite.

L'écart-type asymptotique de l'effet marginal est calculé par la méthode "delta". Ainsi,

$$\text{Var}_{as}(f(\gamma)) = C \text{Var}(\gamma) C'$$

où

$$C \equiv \frac{\partial f(\gamma)}{\partial \gamma'},$$

soit,

$$\begin{aligned} C &= \frac{1}{\sigma^*} \times \phi\left(\frac{x_{it}\gamma}{\sigma^*}\right) \times I_k - \frac{\gamma}{\sigma^*} \times \left(\frac{x_{it}\gamma}{\sigma^*}\right) \times \phi\left(\frac{x_{it}\gamma}{\sigma^*}\right) \times \left(\frac{x_{it}}{\sigma^*}\right) \\ &= \frac{1}{\sigma^*} \times \phi\left(\frac{x_{it}\gamma}{\sigma^*}\right) \left( I_k - \left(\frac{x_{it}\gamma}{\sigma^*}\right) \gamma \left(\frac{x_{it}}{\sigma^*}\right) \right). \end{aligned} \quad (2.1)$$

Certaines applications empiriques nécessitent une estimation en deux étapes. Verbeek et Nijman (1992, 1996) généralisent la procédure de Heckman sur données de panel. La macro-commande “probit” fournit les grandeurs nécessaires à l’application de cette méthode.

## 2.4. Le calcul des inverses des ratios de Mill

La macro-commande SAS fournit, si demandées, les quantités

$$A_{1i} = \frac{1}{\sigma_\eta^2 + T_i \sigma_\xi^2} \sum_{s=1}^{T_i} E \left\{ \xi_i + \eta_{is} / \underline{y}_i \right\}$$

et

$$A_{2it} = \frac{1}{\sigma_\eta^2} \left[ E \left\{ \xi_i + \eta_{it} / \underline{y}_i \right\} - \frac{\sigma_\xi^2}{\sigma_\eta^2 + T_i \sigma_\xi^2} \sum_{s=1}^{T_i} E \left\{ \xi_i + \eta_{is} / \underline{y}_i \right\} \right]$$

où

$$E \left\{ \xi_i + \eta_{it} / \underline{y}_i \right\} = \int_{-\infty}^{+\infty} [\xi_i + E \{ \eta_{it} / y_i, \xi_i \}] f \left( \xi_i / \underline{y}_i \right) d\xi_i \quad (2.2)$$

avec

$$E \left\{ \eta_{it} / \underline{y}_i, \xi_i \right\} = r_{it} \sigma_\eta \frac{\Phi \left( \frac{x_{it} \gamma + \xi_i}{\sigma_\xi} \right)}{\Phi \left( r_{it} \frac{x_{it} \gamma + \xi_i}{\sigma_\xi} \right)}$$

et

$$f \left( \xi_i / \underline{y}_i \right) = \frac{\left[ \prod_{s=1}^{T_i} \Phi \left( r_{it} \frac{x_{it} \gamma + \xi_i}{\sigma_\xi} \right) \right] \frac{1}{\sigma_\xi} \phi \left( \frac{\xi_i}{\sigma_\xi} \right)}{\int_{-\infty}^{+\infty} \left[ \prod_{s=1}^{T_i} \Phi \left( r_{it} \frac{x_{it} \gamma + \xi_i}{\sigma_\xi} \right) \right] \frac{1}{\sigma_\xi} \phi \left( \frac{\xi_i}{\sigma_\xi} \right) d\xi_i}. \quad (2.3)$$

Le dénominateur de (2.3) ainsi que (2.2) sont calculés dans la macro-commande “probit” par la procédure de quadrature gaussienne.

## 3. Un exemple d’application : la participation des femmes au marché du travail

L’appel de la macro-commande “probit” est le suivant :

```
%probit(tab=,y=,x=,id=,points=,mill=,impr1=,impr2=);
```

Tous ces arguments doivent être renseignés.

**tab= :** le nom de la table SAS contenant les données

**y= :** le nom de la régressande (la variable expliquée)

**x= :** la liste des régresseurs (séparés par un blanc) y-compris la constante, non calculée par le programme

**id= :** le nom de l’identifiant individuel (variable numérique)



**points= :** le nombre de points demandé pour l'intégration numérique (5, 8, 9, 12, 16 ou 20)

**mill=:1** si le calcul de l'inverse des ratios de Mill est demandé, 0 sinon

**impr1=: 1** si les résultats intermédiaires du modèle probit sans effet aléatoire sont demandés, 0 sinon

**impr2=: 2** si les résultats intermédiaires du modèle probit avec effets aléatoires sont demandés, 0 sinon.

Les effets marginaux, calculés à la moyenne de l'échantillon, sont systématiquement affichés ainsi que la moyenne du régresseur concerné. Est aussi imprimé un tableau croisant les valeurs prédites et observées de la régressande. Par convention,  $\hat{y}_{it}$  (yhat) =1 dès que  $x_{it}\hat{\gamma} > 0$ .

La macro-commande "probit" crée aussi un certain nombre de tables SAS. Les paramètres estimés du modèle sans hétérogénéité individuelle inobservée sont stockés dans la table **betahat**, la matrice de covariance estimée des paramètres dans **covar**. Les paramètres du modèle avec hétérogénéité sont dans la table **betahath** et leur matrice de covariance estimée dans **covarh**. Si le calcul des inverses des ratios de Mill est demandé, deux tables supplémentaires sont créées, **A1** et **A2**. Elles incluent l'identifiant individuel.

Dans notre exemple, la table SAS contenant les données s'appelle **particip**, la régressande **y**, l'identifiant individuel **identn**. La liste des régresseurs inclut outre la constante (**cst**), le nombre d'enfants de moins de trois ans (**enf02**), le nombre d'enfants entre 3 et 5 ans (**enf35**), ceux entre 6 et 11 ans (**enf611**), ceux de douze ans et plus (**enf1216**), le taux de chômage (**tch**) ainsi qu'un certain nombre de variables discrètes représentant le niveau d'études (**sdip**, **certif**, **cap** et **sup**), l'âge (**age2**, **age3**, **age4** et **age5**) l'année de naissance (**anais2**, **anais3** et **anais4**), le salaire mensuel du conjoint (**salc2**, **salc3**, **salc4** et **salc5**) et le lieu de résidence Ile de France/hors Ile de France (**idfrce**). Ces régresseurs sont similaires à ceux utilisés par Lollivier (1995).

Les données utilisées proviennent des vagues 1, 2 et 3 (soit les années 1994, 1995 et 1996) de l'enquête Europanel. Les estimations ont été réalisées sur le sous-échantillon des femmes vivant en couple à chaque date d'interrogation. La situation de référence correspond à une femme de moins de 30 ans, née après 1963, de niveau BAC, dont le conjoint gagne moins de 6000 francs par mois.

L'estimation est lancée par

```
%probit(tab = particip,
        y = y,
        x = cst enf02 enf35 enf611 enf1216 idfrce sdip certif cap sup age2
            age3 age4 age5 salc2 salc3 salc4 salc5 anais1 anais2 anais3 tch,
        id = identn,
        points = 5,
        mill = 0,
        impr1 = 0,
        impr2 = 0);
```

pour une intégration numérique à 5 points. Le calcul des inverses des ratios de Mill et l’affichage des résultats intermédiaires ne sont pas demandés.

Sont exposées à la suite les estimations du modèle avec hétérogénéité individuelle pour une intégration numérique à 20 points.

# ESTIMATION DU MODELE PROBIT EMPILE SUR DONNEES DE PANEL

NT

NOMBRE D'OBSERVATIONS: 4794

N

NOMBRE D'INDIVIDUS: 1598

TMIN TMAX

NOMBRE DE PERIODES MIN et MAX: 3 3

ITER

NOMBRE D'ITERATIONS: 7

LOGLHOOD

VALEUR DE LA LOG-VRAISEMBLANCE: -2707.861

PARAMETRES ESTIMES

PARAM	BHAT	ETYPE	TSTAT	PCRIT
CST	0.9283313	0.127583	7.2762933	5.389E-13
ENF02	-0.60424	0.0496896	-12.16029	0
ENF35	-0.35791	0.0449265	-7.966563	3.109E-15
ENF611	-0.385818	0.0331647	-11.63339	0
ENF1216	-0.269091	0.0273363	-9.843728	0
IDFRCE	0.3988759	0.0563052	7.0841723	2.101E-12
SDIP	-0.477772	0.0903094	-5.290386	1.3921E-7
CERTIF	-0.283268	0.0721046	-3.928569	0.0000891
CAP	-0.057382	0.0641812	-0.89406	0.3714263
SUP	0.2326548	0.0733788	3.1705996	0.0015505
AGE2	0.3469513	0.0862488	4.0226805	0.0000603
AGE3	0.4751074	0.1084316	4.3816313	0.0000126
AGE4	0.4337664	0.1146178	3.7844585	0.0001598
AGE5	0.1807081	0.1155512	1.5638795	0.1180466
SALC2	0.1074308	0.076471	1.4048574	0.1602607
SALC3	0.1721487	0.0777929	2.2129092	0.0270468
SALC4	0.1144421	0.0779165	1.4687787	0.1420924
SALC5	-0.270755	0.083646	-3.236909	0.0012334
ANAI51	-0.448214	0.120527	-3.718786	0.0002072
ANAI52	0.1225106	0.0988252	1.2396693	0.2152823
ANAI53	0.1088845	0.0819135	1.329263	0.1839536
TCH	-0.01606	0.0046303	-3.468444	0.0005376

## QUALITE DE L'AJUSTEMENT: REGRESSANDE OBSERVEE VS REGRESSANDE AJUSTEE

YOBS_AJU	yhat=0	yhat=1	total
yobs=0	621	1041	1662
yobs=1	323	2809	3132
total	944	3850	4794

# ESTIMATION DES EFFETS MARGINAUX A LA MOYENNE DE L'ECHANTILLON

PARAM	EFFET	ETYPE	TSTAT	PCRIT	MOYENNE
CST	0.3368967	0.0459117	7.3379353	3.457E-13	1
ENF02	-0.219282	0.0180501	-12.14851	0	0.217355
ENF35	-0.129887	0.0163169	-7.960285	3.331E-15	0.2507301
ENF611	-0.140016	0.0120443	-11.62502	0	0.4870672
ENF1216	-0.097655	0.0099161	-9.848112	0	0.6447643
IDFRCE	0.1447543	0.0204034	7.0946222	1.953E-12	0.1745932
SDIP	-0.173386	0.032797	-5.286646	1.4204E-7	0.1395494
CERTIF	-0.102799	0.0261629	-3.929206	0.0000889	0.189612
CAP	-0.020824	0.0232913	-0.894076	0.3714178	0.2991239
SUP	0.0844318	0.0266069	3.1733071	0.0015362	0.22403
AGE2	0.1259106	0.0312994	4.022785	0.0000602	0.2033792
AGE3	0.1724192	0.0393373	4.3830994	0.0000125	0.2169378
AGE4	0.1574163	0.0415861	3.785309	0.0001593	0.1587401
AGE5	0.06558	0.0419354	1.563836	0.1180568	0.2102628
SALC2	0.0389873	0.0277535	1.4047669	0.1602877	0.226116
SALC3	0.0624737	0.028235	2.2126322	0.0270659	0.2154777
SALC4	0.0415317	0.0282777	1.4687072	0.1421118	0.2569879
SALC5	-0.098258	0.0303367	-3.238925	0.0012248	0.2094285
ANAI1	-0.162659	0.0437391	-3.71886	0.0002071	0.071965
ANAI2	0.0444598	0.035863	1.2397135	0.2152659	0.2033792
ANAI3	0.0395148	0.0297292	1.3291576	0.1839884	0.4205257
TCH	-0.005828	0.0016807	-3.46776	0.0005389	15.438256

## ESTIMATION DU MODELE PROBIT A EFFETS ALEATOIRES

NT

NOMBRE D'OBSERVATIONS : 4794

N

NOMBRE D'INDIVIDUS : 1598

TMIN TMAX

NOMBRE DE PERIODES MIN et MAX : 3 3

ITER

NOMBRE D'ITERATIONS : 59

LOGLHOOD

VALEUR DE LA LOG-VRAISEMBLANCE : -1863.167

# PARAMETRES ESTIMES

PARAM	BHAT	ETYPE	TSTAT	PCRIT
CST	1.7269328	0.2226114	7.7576126	1.554E-14
ENF02	-1.230818	0.0923324	-13.33029	0
ENF35	-0.765539	0.0887396	-8.626803	0
ENF611	-0.776163	0.0757423	-10.24742	0
ENF1216	-0.593102	0.0694635	-8.538333	0
IDFRCE	0.5064364	0.1212735	4.1759852	0.0000313
SDIP	-0.954309	0.1863642	-5.120669	3.4193E-7
CERTIF	-0.405557	0.1644029	-2.466849	0.0137365
CAP	-0.041066	0.1419527	-0.289294	0.7723948
SUP	0.6484872	0.158667	4.0870957	0.0000459
AGE2	0.9275813	0.1832088	5.0629721	4.6129E-7
AGE3	1.1884843	0.2223433	5.3452665	1.0353E-7
AGE4	1.5619459	0.2718439	5.7457451	1.0966E-8
AGE5	1.0922435	0.2758666	3.9593184	0.0000785
SALC2	0.2573015	0.1455095	1.7682803	0.0772076
SALC3	0.192185	0.1526279	1.2591741	0.208154
SALC4	0.1101403	0.1620379	0.6797198	0.4967818
SALC5	-0.365574	0.1809186	-2.020656	0.0434843
ANAIS1	-1.082086	0.3185628	-3.396775	0.0006988
ANAIS2	0.0730722	0.243964	0.2995204	0.7645825
ANAIS3	0.1963842	0.1859031	1.0563793	0.2909569
TCH	-0.016143	0.0071433	-2.259923	0.0239621
RHO	0.9142242	0.004925	185.63106	0

## QUALITE DE L'AJUSTEMENT : REGRESSANDE OBSERVEE VS REGRESSANDE AJUSTEE

YOBS_AJU	yhat=0	yhat=1	total
yobs=0	442	1220	1662
yobs=1	156	2976	3132
total	598	4196	4794

# ESTIMATION DES EFFETS MARGINAUX A LA MOYENNE DE L'ECHANTILLON

PARAM	EFFET	ETYPE	TSTAT	PCRIT	MOYENNE
CST	0.1884099	0.0241373	7.8057635	1.066E-14	1
ENF02	-0.134283	0.0099845	-13.4492	0	0.217355
ENF35	-0.083521	0.0096435	-8.660865	0	0.2507301
ENF611	-0.08468	0.0081957	-10.33229	0	0.4870672
ENF1216	-0.064708	0.0075263	-8.597538	0	0.6447643
IDFRCE	0.0552527	0.0132053	4.1841209	0.0000302	0.1745932
SDIP	-0.104116	0.0203305	-5.121161	3.4105E-7	0.1395494
CERTIF	-0.044247	0.0179428	-2.465981	0.0137697	0.189612
CAP	-0.00448	0.0154877	-0.289283	0.7724027	0.2991239
SUP	0.0707505	0.0172502	4.1014213	0.0000432	0.22403
AGE2	0.1011999	0.0198822	5.0899895	4.0109E-7	0.2033792
AGE3	0.1296647	0.0241397	5.3714346	8.9806E-8	0.2169378
AGE4	0.1704097	0.029409	5.7944744	8.2617E-9	0.1587401
AGE5	0.1191647	0.0299576	3.9777802	0.0000727	0.2102628
SALC2	0.0280718	0.0158734	1.7684795	0.0771743	0.226116
SALC3	0.0209676	0.01665	1.259316	0.2081028	0.2154777
SALC4	0.0120164	0.0176777	0.6797505	0.4967623	0.2569879
SALC5	-0.039884	0.0197264	-2.021886	0.043357	0.2094285
ANAIS1	-0.118057	0.0346906	-3.403125	0.0006829	0.071965
ANAIS2	0.0079722	0.0266178	0.2995078	0.7645922	0.2033792
ANAIS3	0.0214257	1.05615	0.0202866	0.2910617	0.4205257
TCH	-0.001761	0.0007791	-2.260656	0.0239165	15.438256

L'estimation du Probit à effets aléatoires individuels avec intégration numérique à 20 points est :

## ESTIMATION DU MODELE PROBIT A EFFETS ALEATOIRES

NT

NOMBRE D'OBSERVATIONS : 4794

N

NOMBRE D'INDIVIDUS : 1598

TMIN TMAX

NOMBRE DE PERIODES MIN et MAX : 3 3

ITER

NOMBRE D'ITERATIONS : 123

LOGLHOOD

VALEUR DE LA LOG-VRAISEMBLANCE : -1829.131

# PARAMETRES ESTIMES

PARAM	BHAT	ETYPE	TSTAT	PCRIT
CST	2.0993828	0.360516	5.8232714	6.9815E-9
ENF02	-1.496141	0.1259888	-11.87519	0
ENF35	-1.007641	0.1308022	-7.70355	2.331E-14
ENF611	-1.000142	0.1238354	-8.076379	1.332E-15
ENF1216	-0.662443	0.100408	-6.597513	5.694E-11
IDFRCE	1.1851015	0.2329321	5.0877545	4.0577E-7
SDIP	-1.880518	0.343094	-5.481059	4.9159E-8
CERTIF	-0.916046	0.3089689	-2.964849	0.0030738
CAP	-0.241565	0.2782696	-0.868096	0.385474
SUP	1.0484474	0.3052438	3.4347864	0.0006084
AGE2	0.9733844	0.2416716	4.0277148	0.000059
AGE3	1.1799585	0.3015078	3.9135249	0.0000948
AGE4	1.3599567	0.3721898	3.6539332	0.0002667
AGE5	1.2574598	0.4189261	3.0016271	0.0027277
SALC2	0.3616633	0.2119514	1.7063502	0.08814
SALC3	0.3410291	0.2300944	1.4821267	0.1385066
SALC4	-0.014468	0.2436109	-0.059391	0.9526481
SALC5	-0.790349	0.2716818	-2.909097	0.0036756
ANAIS1	-1.969113	0.5038864	-3.907851	0.0000971
ANAIS2	0.321348	0.3907525	0.8223824	0.4109838
ANAIS3	0.6947888	0.2935775	2.366628	0.0180714
TCH	-0.014919	0.0095947	-1.554972	0.1201536
RHO	0.9508697	0.0045243	210.16835	0

## QUALITE DE L'AJUSTEMENT : REGRESSANDE OBSERVEE VS REGRESSANDE AJUSTEE

YOBS_AJU	yhat=0	yhat=1	total
yobs=0	552	1110	1662
yobs=1	276	2856	3132
total	828	3966	4794

## ESTIMATION DES EFFETS MARGINAUX A LA MOYENNE DE L'ECHANTILLON

PARAM	EFFET	ETYPE	TSTAT	PCRIT	MOYENNE
CST	0.176342	0.030044	5.8694595	5.3209E-9	1
ENF02	-0.125671	0.0105588	-11.90205	0	0.217355
ENF35	-0.084639	0.010991	-7.700739	2.376E-14	0.2507301
ENF611	-0.084009	0.010374	-8.098016	1.11E-15	0.4870672
ENF1216	-0.055643	0.0084043	-6.620781	4.886E-11	0.6447643
IDFRCE	0.0995451	0.0195011	5.1045796	3.7182E-7	0.1745932
SDIP	-0.157958	0.0286234	-5.518489	3.9917E-8	0.1395494
CERTIF	-0.076945	0.0258977	-2.971118	0.0030121	0.189612
CAP	-0.020291	0.0233674	-0.868333	0.3853443	0.2991239
SUP	0.0880665	0.0255665	3.4446006	0.0005869	0.22403
AGE2	0.0817614	0.0202507	4.0374637	0.0000566	0.2033792
AGE3	0.0991131	0.0252852	3.9198045	0.0000924	0.2169378
AGE4	0.1142324	0.0311761	3.6640961	0.0002564	0.1587401
AGE5	0.1056229	0.035096	3.0095434	0.002658	0.2102628
SALC2	0.0303787	0.0177958	1.707072	0.0880058	0.226116
SALC3	0.0286454	0.0193226	1.4824853	0.1384113	0.2154777
SALC4	-0.001215	0.0204626	-0.059391	0.9526482	0.2569879
SALC5	-0.066387	0.0228044	-2.911153	0.0036516	0.2094285
ANAIS1	-0.1654	0.0422515	-3.914648	0.0000944	0.071965
ANAIS2	0.0269923	0.0328121	0.822633	0.4108412	0.2033792
ANAIS3	0.0583602	0.0246264	2.3698266	0.0179163	0.4205257
TCH	-0.001253	0.0008064	-1.553987	0.1203885	15.438256

Les paramètres et les écart-types estimés des modèles sans et avec hétérogénéité ne peuvent être directement comparés. Il faut préalablement normaliser les coefficients et les écarts-types du modèle avec hétérogénéité en les divisant par  $\sqrt{\hat{\sigma}_\xi^2 + 1}$ . La comparaison peut être néanmoins directement réalisée à partir de l'estimation des effets marginaux.

Des différences apparaissent entre les estimations issues des intégrations numériques à 5 et 20 points.

La variable ANAIS3 devient significative avec l'approximation à 20 points au contraire du taux de chômage dont l'impact devient nul. De plus, le nombre de  $y_{it} = 0$  prédits comme tel augmente tandis que celui des  $y_{it} = 1$  diminue. Le choix du nombre de points d'évaluation peut s'avérer important : **le nombre minimal de points nécessaires dépend beaucoup de la fonction à intégrer**. Parfois il suffit de prendre 5 points, parfois il en faut 20.

Présentons maintenant le programme de la macro-commande.

## 4. L'écriture de la macro "probit"

Afin de rendre plus compréhensible l'écriture du programme, explicitons certaines lignes d'instructions concernant respectivement le calcul de l'estimateur, des effets marginaux et des inverses des ratios de Mill.

#### 4.1. L'estimation du modèle

Toutes les opérations de calcul du modèle avec hétérogénéité individuelle sont réalisées pour différentes valeurs de  $\xi_i$ , appelées **abscisse**. Le nombre de points dépend de la valeur de la macro-variable **points** choisi. Le vecteur **ri** correspond aux  $T_i$  éléments  $r_{it}$ . L'instruction

$$\text{ri}\#(\text{x}[\text{ti\_1}+1:\text{ti\_1}+\text{ti\_i}]\*\text{bhat}+\text{teta}\#\text{abscisse}[\text{j},]\#\text{j}(\text{ti},1,1)/\text{sqrt}(2))$$

calcule le vecteur **z** constitué des  $T_i$  éléments

$$r_{it}(x_{it}\gamma + \theta h_i).$$

Le vecteur **cdf** correspond aux  $T_i$  éléments  $\Phi(r_{it}w_{it})$  et **pcdf** au produit pour l'individu  $i$  de ces  $T_i$  éléments, à savoir

$$\prod_{t=1}^{T_i} \Phi(r_{it}w_{it}).$$

Le vecteur **pdf** représente les  $T_i$  éléments de  $\phi(r_{it}w_{it})$  tandis que **cgrad** définit les  $T_i$  termes

$$r_{it} \frac{\phi(r_{it}w_{it})}{\Phi(r_{it}w_{it})} \left[ \prod_{t=1}^{T_i} \Phi(r_{it}w_{it}) \right] (x_{it}, h_i)$$

de la contribution individuelle au gradient et que **sgrad** en fait la somme,

$$\sum_{t=1}^{T_i} \left( r_{it} \frac{\phi(r_{it}w_{it})}{\Phi(r_{it}w_{it})} \left[ \prod_{t=1}^{T_i} \Phi(r_{it}w_{it}) \right] (x_{it}, h_i) \right).$$

Les instructions

$$\begin{aligned} \text{Li} &= \text{Li} + \text{poids}[\text{j},] \# \text{pcdf}; \\ \text{gradi} &= \text{gradi} + \text{poids}[\text{j},] \# \text{sgrad}; \end{aligned}$$

réalisent respectivement les approximations numériques de  $L_i$  et de  $\frac{\partial L_i}{\partial \alpha}$  par l'application de la formule de quadrature gaussienne (Butler et Moffitt, 1982)

$$\sum_{j=1}^J w_j f(z_j)$$

où  $\text{poids}[j,]$  représente  $w_j$ . Le gradient **grad** est alors obtenu par l'instruction

$$\text{grad} = \text{grad} + (\text{gradi}/\text{Li})';$$

et la log-vraisemblance par

$$\text{loglhood} = \text{loglhood} + \log(\text{Li}/\text{sqrt}(4 * \text{atan}(1)));$$

où  $\text{sqrt}(4 * \text{atan}(1)) = \frac{1}{\sqrt{\pi}}$ .



## 4.2. Le calcul des effets marginaux

L'instruction

$$\text{moyenne} = (\text{j}(1, \text{nt}, 1) * \text{x}) / \text{nt};$$

calcule la moyenne des régresseurs. L'effet marginal  $f(\gamma)$  est calculé par

$$\text{pdfz} = \exp(-z \# z / 2) / \text{sqrt}(8 * \text{atan}(1));$$

et (2.1) par

$$C = \text{pdfz} * \text{i}(\text{k}) / \text{sqrt}(1 + \text{sig}^2) - (\text{bhat}[1:\text{k},] / (1 + \text{sig}^2)) * z * \text{pdfz} * \text{moyenne};$$

où **sig2** représente  $\sigma_\xi^2$  et **bhat[1:k,]**  $\gamma$ .

## 4.3. Le calcul des inverses des ratios de Mill

La variable **adenomf** correspond à l'approximation numérique du dénominateur de (2.3), **espeteps** à  $E \{ \xi_i + \eta_{it} / \underline{y}_i \}$  et **espereta** à  $E \{ \eta_{it} / y_i, \xi_i \}$ . Par ailleurs **cdf1** est égal à  $\Phi \left( \frac{x_{it}\gamma + \xi_i}{\sigma_\xi} \right)$  où **sig** correspondent à  $\sigma_\xi$  tandis que **cdf2** représente  $\Phi \left( r_{it} \frac{x_{it}\gamma + \xi_i}{\sigma_\xi} \right)$ .

L'instruction **cdf2[#,]** réalise le produit sur les  $T_i$  observations de l'individu  $i$  de  $\Phi \left( r_{it} \frac{x_{it}\gamma + \xi_i}{\sigma_\xi} \right)$  afin de donner  $\prod_{s=1}^{T_i} \Phi \left( r_{it} \frac{x_{it}\gamma + \xi_i}{\sigma_\xi} \right)$  (**pcdf2**). Les instructions

$$\begin{aligned} \text{espeteps} &= \text{espeteps} + \text{poids}[\text{j},] \# ((\text{abscisse}[\text{j},] \# \text{j}(\text{ti}, 1, 1) + \text{espereta}) \# \text{pcdf2}); \\ \text{adenomf} &= \text{adenomf} + \text{poids}[\text{j},] \# \text{pcdf2}; \end{aligned}$$

réalisent les sommations

$$\sum_{j=1}^J w_j f(z_j)$$

nécessaires à l'approximation numérique de la fonction  $f(z)$  concernée.

## 4.4. L'écriture du programme

```
%macro probit(tab=,y=,x=,id=,points=,mill=,impr1=1,impr2=);
```

```
data deb;
```

```
deb=time();      /* pour calcul du temps d'exécution */
```

```
cst=1;
```

```
/* comptabilisation du nombre de périodes par individu */
```

```
proc summary data=&tab nway;
```

```
class &id;
```

```
output out=nobs(keep=_freq-);
```

```
data nobs;
```

```
set nobs;
```

```
obs=1;
```

```
data n(keep=n);
```

```
set nobs;
```

```

by obs;                                /* nombre total d'individus */
n+1;
if last.obs;
run;
proc iml;
start probit;

/*****
*****                                lecture des données                                *****/
*****/

parm={&x rho};                          /* nom des paramètres à estimer */
use &tab;
read all var{&y} into y;
read all var{&x} into x;
close &tab;

use n;
read all var{n} into n;
close n;

use nobs;
read all var{_freq_} into t;
close nobs;

tmax=max(t);                            /* nombre max de périodes */
tmin=min(t);                            /* nombre min de périodes */
free t;
nt=nrow(y);

/*****
*****                                estimation du Probit sans facteur d'hétérogénéité                                *****/
*****/

crit=1;
lo=0;
k=ncol(x);
maxiter=40;
bhat=inv(x'*x)*x'*y;                    /* paramètres initiaux */

do iter=1 to maxiter while(crit>1.0e-6);
  loglhood=0;
  z=x*bhat;
  pdf=exp(-z#z/2)/sqrt(8*atan(1));      /* calcul de la densité f(z) */
  cdf=probnorm(z);                      /* calcul de la fonction de répartition F(z) */
  mcdf=probnorm(-z);

```

```

grad=(y#(pdf/cdf)-(1-y)#(pdf/(1-cdf)))#x;          /* calcule du gradient */
grad=grad[+,,]; /*fait la somme sur toutes les lignes pour chaque colonne et trans-
pose*/
phi=pdf##2/(cdf#(1-cdf));
inf=(x#phi)'*x;          /* matrice d'information */
free phi pdf;
loglhood=sum(y#log(cdf)+(1-y)#log(mcdf));
free cdf mcdf;
invinf=inv(inf);
free inf;
delta=invinf*grad;          /* calcul du pas */
/* critères de convergence */
crit1=abs(loglhood-lo);
crit2=ssq(grad);
crit3=ssq(delta);
critall=crit1//crit2//crit3;
crit=max(critall);
free grad crit1 crit2 crit3 critall;
if &impr1=1 then print iter loglhood;
bhat=bhat+delta;          /* mise à jour des paramètres estimés */
lo=loglhood;          /* mise à jour de la vraisemblance */
end;

/* inférence statistique */
etype=sqrt(vecdiag(invinf));
tstat=bhat/etype;
pcrit=2#(1-probt(abs(tstat),n-k));

/* impression des résultats */
param=t(parm[,1:k]);
print "ESTIMATION DU MODELE PROBIT EMPILE SUR DONNEES DE PANEL",,
"NOMBRE D'OBSERVATIONS:" NT,
"NOMBRE D'INDIVIDUS:" N,
"NOMBRE DE PERIODES MIN et MAX:" tmin tmax,
"NOMBRE D'ITERATIONS:" iter,
"VALEUR DE LA LOG-VRAISEMBLANCE:" loglhood,,
"PARAMETRES ESTIMES",
param bhat etype tstat pcrit;

/* y réalisés et prédits */
yhat=x*bhat>j(nt,1,0);
start predic;
y1=sum(y);
y0=nt-y1;
yhat1=sum(x*bhat[1:k,]>j(nt,1,0));
yhat0=nt-yhat1;

```

```

y1hat0=sum(y>yhat);
y0hat1=sum(y<yhat);
y1hat1=y1-y1hat0;
y0hat0=y0-y0hat1;
yobs_aju=(y0hat0||y0hat1||y0)/(y1hat0||y1hat1||y1)/(yhat0||yhat1||nt);
nom1={"yobs=0","yobs=1","total"};
nom2={"yhat=0" "yhat=1" "total"};
print "QUALITE DE L'AJUSTEMENT: REGRESSANDE OBSERVEE VS REGRES-
SANDE AJUSTEE"
yobs_aju [rowname=nom1 colname=nom2];
free yhat y0 y1 yhat1 yhat0 y1hat0 y1hat1 y0hat0 y0hat1 nom1 nom2 yobs_aju;
finish predic;
run predic;

/* calcul et impression des effets marginaux la moyenne de l'échantillon */
moyenne=(j(1,nt,1)*x)/nt;
z=moyenne*bhat;
pdfz=exp(-z#z/2)/sqrt(8*atan(1));
effet=pdfz*bhat;
C=-z*pdfz*bhat*moyenne+pdfz*i(k);
var=C*invinf*C';
etype=sqrt(vecdiag(var));
tstat=effet/etype;
pcrit=2#(1-probt(abs(tstat),n-k));
moyenne=moyenne';
print "ESTIMATION DES EFFETS MARGINAUX A LA MOYENNE DE L'ECHANTILLON",
param effet etype tstat pcrit moyenne;
free C z pdfz effet etype tstat pcrit moyenne var;

/* création de la table incluant les paramètres estimés et de la table*/
/* comprenant la matrice de covariance estimée des paramètres */
tbhat=bhat';
parm1=parm[,1:k];
create betahat from tbhat [colname=parm1];
append from tbhat;
create covar from invinf [colname=parm1];
append from invinf;
free invinf tbhat crit lo loglikelihood delta param parm1;

/*****
/***** estimation du Probit avec facteur d'hétérogénéité *****/
/*****

/* initialisation des abscisses et poids nécessaires l'approximation numérique */
/* de la vraisemblance et du gradient par l'intégration hermitienne */
if &points=5 then do;

```

```

abscisse={-2.02018,-0.95857,0,0.95857,2.02018};
poids={1.99532e-2,3.93619e-1,9.45308e-1,3.93619e-1,1.99532e-2};
end;
if &points=8 then do;
    abscisse={-2.93063,-1.98165,-1.15719,-0.38118,0.38118,1.15719,1.98165,2.93063};
    poids={1.99604e-4,1.70779e-2,2.07802e-1,6.61147e-1,6.61147e-1,2.07802e-1,1.70779e-2,
    1.99604e-4};
end;
end;
if &points=9 then do;
    abscisse={-3.19099,-2.26658,-1.46855,-0.72355,0,0.72355,1.46855,2.26658,3.19099};
    poids={3.96069e-5,4.94362e-3,8.84745e-2,4.32651e-1,7.20235e-1,4.32651e-1,8.84745e-2,
    4.94362e-3,3.96069e-5};
end;
if &points=12 then do;
    abscisse={-3.88972,-3.02063,-2.27950,-1.59768,-0.94778,-0.31424,0.31424,0.94778,1.59768,
    2.27950,3.02063,3.88972};
    poids={2.65855e-7,8.57368e-5,3.90539e-3,5.16079e-2,2.60492e-1,5.70135e-1,5.70135e-1,
    2.60492e-1,5.16079e-2,3.90539e-3,8.57368e-5,2.65855e-7};
end;
if &points=16 then do;
    abscisse={-4.68873,-3.86944,-3.17699,-2.54620,-1.95178,-1.38025,-0.82295,-0.27348,0.27348,
    0.82295,1.38025,1.95178,2.54620,3.17699,3.86944,4.68873};
    poids={2.65480e-10,2.32098e-7,2.71186e-5,9.32284e-4,1.28803e-2,8.38100e-2,2.71186e-5,
    2.80647e-1,5.07929e-1,5.07929e-1,2.80647e-1,8.38100e-2,1.28803e-2,9.32284e-4,
    2.32098e-7,2.65480e-10};
end;
if &points=20 then do;
    abscisse={-5.38748,-4.60368,-3.94476,-3.34785,-2.78880,-2.25497,-1.73853,-1.23407,
    -0.73747,-0.24534,0.24534,0.73747,1.23407,1.73853,2.25497,2.78880,3.34785,3.94476,
    4.60368,5.38748};
    poids={2.22939e-13,4.39934e-10,1.08606e-7,7.80255e-6,2.28338e-4,3.24377e-3,2.48105e-
    2,
    1.09017e-1,2.86675e-1,4.62243e-1,4.62243e-1,2.86675e-1,1.09017e-1,2.48105e-2,3.24377e-
    3,
    2.28338e-4,7.80255e-6,1.08606e-7,4.39934e-10,2.22939e-13};
end;
crit=1;
lo=0;
teta=0.29;
alpha=bhat//teta;
maxiter=200;
use nob;
read all var{_freq_} into t;
close nob;
do iter=1 to maxiter while(crit>1.0e-6);

```

```

loglikelihood=0;
inf=0;
grad=0;
ti_1=0;
do i=1 to n;
  Li=0;
  gradi=0;
  ti=t[i,];
  ri=2#y[ti_1+1:ti_1+ti,]-j(ti,1,1);
  seuil=j(ti,1,-5);
  /* intégration numérique */
  do j=1 to &points;
    z=ri#(x[ti_1+1:ti_1+ti,]*bhat+teta#abscisse[j,]#j(ti,1,1)/sqrt(2));
    cdf=probnorm(z);
    pcdf=cdf[#,];
    Li=Li+poids[j,]#pcdf;
    pdf=exp(-z#z/2)/sqrt(8*atan(1));
    cond=z>=seuil; /* valeurs extrêmes */
    ratio=cond#(pdf/cdf)+(j(ti,1,1)-cond)#(-z);
    sgrad=(ri#ratio)*(x[ti_1+1:ti_1+ti,]|(abscisse[j,]#j(ti,1,1)));
    free cdf pdf cond ratio;
    gradi=gradi+poids[j,]#pcdf#sgrad;
    free pcdf sgrad;
  end;
  ti_1=ti_1+ti;
  grad=grad+(gradi/Li);
  loglikelihood=loglikelihood+log(Li/sqrt(4*atan(1)));
  inf=inf+(gradi/Li)*(gradi/Li);
  free gradi Li;
end;
cov=inv(inf);
free inf;
delta=cov*grad; /* calcul du pas */
/* critères de convergence */
crit1=abs(loglikelihood-lo);
crit2=ssq(grad);
crit3=ssq(delta);
critall=crit1//crit2//crit3;
crit=max(critall);
free grad crit1 crit2 crit3;
if &impr2=1 then print iter loglikelihood;
alpha=alpha+delta; /* mise à jour des paramètres estimés */
bhat=alpha[1:k,];
teta=alpha[k+1,];
lo=loglikelihood; /* mise à jour de la vraisemblance */

```

```

end;

/* inférence statistique et impression des résultats */
rho=teta##2/(2+teta##2);
/* calcul de l'écart-type de rho par la méthode delta */
etyprho=cov[k+1,k+1]#sqrt(4#teta/((2+teta##2)##2));
etype=sqrt(vecdiag(cov[1:k,1:k]));
free alpha;
etype=etype//etyprho;
bhat=bhat//rho;
tstat=bhat/etype;
pcrit=2#(1-probt(abs(tstat),n-k-1));
param=parm';
print "ESTIMATION DU MODELE PROBIT A EFFETS ALEATOIRES" ,,
"NOMBRE D'OBSERVATIONS:" NT,
"NOMBRE D'INDIVIDUS:" N,
"NOMBRE DE PERIODES MIN et MAX:" tmin tmax,
"NOMBRE D'ITERATIONS:" iter,
"VALEUR DE LA LOG-VRAISEMBLANCE:" loglikelihood,,
"PARAMETRES ESTIMES",
param bhat etype tstat pcrit;

/* y réalisés et prédits */
yhat=x*bhat[1:k,]>j(nt,1,0);
run predic;

/* calcul et impression des effets marginaux la moyenne de l'échantillon */
sig2=rho/(1-rho); /* variance de l'effet fixe */
moyenne=(j(1,nt,1)*x)/(nt);
z=(moyenne*bhat[1:k,])/sqrt(1+sig2);
pdfz=exp(-z#z/2)/sqrt(8*atan(1));
effet=pdfz*bhat[1:k,]/sqrt(1+sig2);
C=pdfz*i(k)/sqrt(1+sig2)-(bhat[1:k,]/(1+sig2))*z*pdfz*moyenne;
var=C*cov[1:k,1:k]*C';
etype=sqrt(vecdiag(var));
tstat=effet/etype;
pcrit=2#(1-probt(abs(tstat),n-k-1));
moyenne=moyenne';
print "ESTIMATION DES EFFETS MARGINAUX A LA MOYENNE DE L'ECHANTILLON",
param effet etype tstat pcrit moyenne;
free C z pdfz effet etype tstat pcrit moyenne var;

/* création de la table incluant les paramètres estimés et de la table */
/* comprenant la matrice de covariance estimée des paramètres */
bhath=bhat';
create betahath from bhath [colname=parm];

```

```

append from bhath;
create covarh from cov [colname=parm];
append from cov;
free cov etype tstat pcrit crit lo loglhood delta etyprho bhath param;

if &mill=1 then do;
/*****
/*****          calcul des inverses des ratios de Mill          *****/
/*****
/*****

A1=j(nt,1,0);          /* initialisation des vecteurs incluant les ratios */
A2=j(nt,1,0);
sig=sqrt(sig2);        /* écart-type de l'effet fixe */
ti_1=0;
do i=1 to n;
  ti=t[i,];
  ri=2#y[ti_1+1:ti_1+ti,]-j(ti,1,1);
  adenomf=0;
  espeteps=0;
  do j=1 to &points;
    cdf1=probnorm((x[ti_1+1:ti_1+ti,]*bhat[1:k,]+abscisse[j,]#j(ti,1,1))/sig);
    cdf2=probnorm(ri#(x[ti_1+1:ti_1+ti,]*bhat[1:k,]+abscisse[j,]#j(ti,1,1))/sig);
    pcdf2=cdf2[#,];
    espereta=ri#cdf1/cdf2;
    free cdf1 cdf2;
    espeteps=espeteps+poids[j,]#((abscisse[j,]#j(ti,1,1)+espereta)#pcdf2);
    adenomf=adenomf+poids[j,]#pcdf2;
    free espereta pcdf2;
  end;
  espeteps=espeteps/adenomf;
  A1[ti_1+1:ti_1+ti,]=(j(1,ti,1)*espeteps/(1+ti#sig##2))#j(ti,1,1);
  A2[ti_1+1:ti_1+ti,]=espeteps-A1[ti_1+1:ti_1+ti,];
end;
free x y;

/* création des tables incluant les inverses des ratios de Mill */
/* A1 correspond celui de l'effet fixe et A2 celui de l'aléa. */
/* Est ajouté dans ces tables l'identifiant individuel*/
use &tab;
read all var{&id} into id;
close &tab;
A1=id||A1;
A2=id||A2;
nom1={id A1};
nom2={id A2};
create A1 from A1 [colname=nom1];

```



```
append from A1;  
create A2 from A2 [colname=nom2];  
append from A2;  
end;  
finish probit;  
run probit;  
quit;  
%mend probit;
```

## 5. Conclusion

Dans ce document nous proposons une macro-commande SAS pour le modèle Probit à effets aléatoires. Elle a été écrite à l'aide du langage matriciel de SAS, IML. Ce choix est discutable. L'écriture IML est relativement transparente. Elle suit de près la formalisation initiale du modèle. Par contre, la mauvaise gestion de la mémoire virtuelle de SAS conduit, pour des fichiers de grande taille, à un temps d'exécution plus long que ceux engendrés par d'autres choix d'écriture sous SAS.

## Références

- [1] Butler J.S. and R. Moffitt 1982, A computationally efficient quadrature procedure for the one-factor multinomial probit model, *Econometrica*, vol. 50, n° 3, pp 761-764
- [2] Davidson R. et J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York, Oxford University Press.
- [3] Greene W. (1995), “LIMDEP, Version 7 : User’s Manual”, NY : Econometric Software
- [4] Lollivier S. (1995), “Activité des femmes mariées et hétérogénéité : estimation sur données de panel”, *Annales d’économie et de statistique*, n° 39, pp 93-106
- [5] Verbeek M. et T. Nijman (1992), “Testing for selectivity bias in panel data models”, *International Economic Review*, vol. 33, n° 3, pp 681-703
- [6] Verbeek M. et T. Nijman (1996), “Incomplete panels and selection bias” dans Mátyás L. et P. Sevestre, ed., “*The Econometrics of Panel Data*”, Kluwer Academic Publishers

## *Working Papers / Documents de Travail*

Le GREQAM. diffuse une série de documents de travail qui prend la suite de celle diffusée depuis 1982 par le GREQE. Cette série comporte 4 sous séries - respectivement : "théories", "applications", "méthodologie", "Tirés à part"

Certains papiers sont disponibles sous la forme de fichiers PDF. Adobe Acrobat vous permet de les lire, de les transférer et de les imprimer.

**<http://durandal.cnrs-mrs.fr/GREQAM/dt/dt.htm>**

---

### **Adresses du GREQAM**

#### **GREQAM**

Centre de la Vieille Charité  
2 Rue de La Charité  
13002 MARSEILLE  
tél. 04.91.14.07.70  
fax. 04.91.90.02.27  
E-mail : [greqam@ehess.cnrs-mrs.fr](mailto:greqam@ehess.cnrs-mrs.fr)

#### **GREQAM/ LEQAM**

Château La Farge  
Route des Milles  
13290 LES MILLES  
tél. 04.42.93.59.80  
fax. 04.42.93.09.68  
E-mail : [leqam@romarin.univ-aix.fr](mailto:leqam@romarin.univ-aix.fr)

#### **GREQAM/ C.R.I.D.E.S.O.P.E.**

Faculté d'Économie Appliquée  
Bât. Austerlitz  
15-19 Allée Claude Forbin  
13627 Aix-en-Provence Cedex 1  
Tel : 04 42 96 12 31  
Fax : 04 42 96 80 00