Université d'Orléans - Master Econométrie et Statistique Appliquée
Econométrie des Variables Qualitatives

Examen Mai 2008. C. Hurlin

Exercice 1 : Modèle Logit Ordonné (9 points)

Soit y_i une variable multinomiale associée à l'individu i pouvant prendre 3 modalités, notées respectivement 0,1 et 2. On suppose que cette variable admet une représentation telle que :

$$y_i = \begin{cases} 0 & \text{si } y_i^* < \gamma_1 \\ 1 & \text{si } \gamma_1 \leq y_i^* < \gamma_2 \\ 2 & \text{si } y_i^* \geq \gamma_2 \end{cases} \qquad \forall i = 1, .., N \tag{1}$$

où la variable latente y_i^* vérifie :

$$y_i^* = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i \tag{2}$$

où x_i et z_i désignent deux variables explicatives. On admet que les termes d'erreur ε_i sont *i.i.d.* et sont distribués selon une loi logistique de variance égale à σ_ε^2.

On note $\Lambda(.)$ la fonction de répartition de la loi logistique standard de variance égale à $\pi^2/3$, telle que :

$$\Lambda(\omega) = \frac{1}{1 + \exp(-\omega)} \tag{3}$$

Question 1 *(2 points)* : Donnez l'expression des probabilités $\Pr(y_i = 0)$, $\Pr(y_i = 1)$ et $\Pr(y_i = 2)$ en fonction de $\Lambda(.)$ et des paramètres $\beta_0,\ \beta_1,\ \beta_2,\ \gamma_1,\ \gamma_2$ et σ_ε.

Question 2 *(1 point)* : On considère une représentation de type modèle logit multinomial ordonné pouvant s'écrire sous la forme suivante :

$$\Pr(y_i = j) = \Lambda(c_{j+1} - w_i\alpha) - \Lambda(c_j - w_i\alpha) \quad \forall j \in \{0, 1, 2\} \tag{4}$$

où $\alpha = (\alpha_0\ \alpha_1\ \alpha_2)'$ désigne le vecteur des paramètres, $w_i = (1\ x_i\ z_i)$ désigne le vecteur des variables explicatives et où les seuils c_j vérifient $c_0 = -\infty$ et $c_3 = +\infty$. Exprimez les paramètres du logit multinomial ordonnée $(\alpha_0,\ \alpha_1, \alpha_2, c_1, c_2)$ en fonction des paramètres structurels du modèle précédent $(\beta_0,\ \beta_1,\ \beta_2,\ \gamma_1,\ \gamma_2,\ \sigma_\varepsilon)$. *Remarque : Vous discuterez précisèment le cas de la constante α_0.*

Question 3 *(1 point)* : On considère les estimateurs du maximum de vraisemblance des paramètres du modèle logit multinomial ordonné (4) obtenus à partir d'un N-échantillon $\{y_i, x_i, z_i\}_{i=1}^N$. On admet que les réalisations de ces estimateurs sont :

$$\widehat{\alpha}_0 = 0 \quad \widehat{\alpha}_1 = 4 \quad \widehat{\alpha}_2 = -2 \quad \widehat{c}_1 = 0.5 \quad \widehat{c}_2 = 1 \tag{5}$$

Pour un individu i tel que $x_i = 1$ et $z_i = 1.75$, déterminez les probabilités $\Pr(y_i = 0)$, $\Pr(y_i = 1)$ et $\Pr(y_i = 2)$.

Question 4 *(2 points)* : Ecrivez la vraisemblance associée à un individu i en fonction de ses caractéristiques x_i et z_i, ainsi que des paramètres du modèle logit multinomial ordonné (4). En déduire la valeur de cette vraisemblance pour un individu pour lequel on observe $y_i = 1$, $x_i = 1$, $z_i = 1.75$ lorsque l'on suppose que les paramètres sont égaux à leur valeur estimée (5).

Question 5 *(1 point)* : On suppose que les paramètres sont égaux à leur valeur estimée (5). Sans effectuer de calcul, que pouvez conclure quant à l'évolution de la probabilité $\Pr(y_i = 2)$ lorsque la variable x augmente d'une unité ? *Vous justifierez précisèment votre réponse.*

Question 6 *(2 points)* : Quels sont les effets marginaux associés à la hausse d'une unité de la variable x. On effectuera les calculs pour l'individu de référence tel que $x_i = 1$ et $z_i = 1.75$.

Exercice 2 : Scoring et Défaillances d'Entreprises (9 points)

Une société souhaite construire un modèle de scoring sur les défaillances d'entreprises. On considère pour cela une base de donnés portant sur 181 firmes du secteur manufacturier cotées au NYSE ou à l'AMEX de 1981 à 1989 (*Using SAS in Financial Research, Boehmer*, Broussard and Kallunki, , SAS Press, 2002). Parmi ces firmes, 86 ont été mises en faillites à sur la période et l'on dispose pour ces firmes de données portant sur leur bilan une année avant la mise en faillite. Pour chaque firme, on considère les variables suivantes :

YD	Variable dummie valant 1 en cas de faillite et 0 sinon
TDTA	Ratio dette sur actif total
GEMPL	Taux de croissance annuel des effectifs
OPITA	Excédent Brut d'Exploitation sur Actif Total
INVSLS	Ratio Stock sur Ventes
LSLS	Logarithme des Ventes

Question 1 *(1 point)* : On se propose tout d'abord d'estimer un modèle probit. Commentez économiquement les résultats d'estimation de ce modèle (Figure 1).

Question 2 *(1 point)* : A partir des résultats d'estimation du modèle probit (Figure 1), **évaluez la probabilité de défaillance et la cote** d'une firme ayant pour caractéristiques :

$$\begin{aligned} TDTA &= 0.462 \qquad GEMPL = 0.00728 \qquad OPITA = 0.09584 \\ INVSLS &= 0.0822 \qquad LSLS = 5.044. \end{aligned}$$

Question 3 *(1 point)* : Calculez l'effet marginal sur la probabilité de défaillance de l'entreprise décrite à la question précédente d'une hausse marginale du ratio dette sur actif total. Commentez vos résultats.

Question 4 *(3 points)* : Dans le tableau suivant, sont reportées les probabilités estimées de défaillance issues d'un modèle probit et d'un modèle logit obtenues pour un échantillon test de 5 individus.

i	YD_i	Probit $\Pr(y_i = 1)$	Logit $\Pr(y_i = 1)$
1	1	0.8	0.7
2	1	0.6	0.4
3	1	0.5	0.3
4	0	0.3	0.2
5	0	0.1	0.1

En utilisant la ROC curve comme critère de comparaison, déterminez le meilleur modèle prédictif.

Question 5 *(3 points)* : Pour un vecteur de paramètres donné, on suppose que les index associés aux $87^{ème,}$ $88^{ème}$ et $89^{ème}$ firmes de l'échantillon sont les suivants :

i	YD_i	$Index_i$
87	0	-3.06
88	1	-0.99
89	1	-0.66

En ne considérant uniquement que ces trois firmes, calculez la probabilité de défaillance de la firme $i = 88$ par une approche semi-paramétrique de type Klein et Spady (1993) en utilisant un kernel de type gaussien et une valeur du bandwidth parameter $h = 1.72$.

Exercice 3 : Modèle Tobit (3 points)

On considère le modèle suivant :

$$y_i^* = 1 - 2\ x_i + \varepsilon_i \tag{6}$$

où les résidus ε_i sont $N.i.d.$ avec $E(\varepsilon_i) = 0$ et $E\left(\varepsilon_i^2\right) = 4$. On suppose que les variables x_i sont indépendantes et distribuées selon une loi normale, avec $E(x_i) = 0$ et $E\left(x_i^2\right) = 9$. On suppose que la variable y_i^* n'est observable que lorsqu'elle est positive et l'on note :

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* \geq 0 \\ 0 & \text{sinon} \end{cases} \tag{7}$$

Enfin, on considère un N-échantillon d'observations $\{y_i, x_i\}$, $i = 1, .., N$.

Question 1 *(1 point)* : Vers quelle valeur converge l'estimateur MCO du coefficient de la variable x_i obtenu à partir de l'échantillon complet $\{y_i, x_i\}$?

Question 2 *(1 point)* : Vers quelle valeur converge l'estimateur MCO du coefficient de la variable x_i obtenu à partir de l'échantillon construit à partir des seules valeurs positives de y_i.

Question 3 *(1 point)* : Proposez dans ce cas précis un estimateur non biaisé.

Figure 1: Modèle Probit

Dependent Variable: YD
Method: ML - Binary Probit
Date: 04/28/08 Time: 10:54
Sample: 1 181
Included observations: 181
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.997374	0.605556	-1.647040	0.0995
TDTA	2.927908	0.702700	4.166653	0.0000
GEMPL	-3.747136	1.188848	-3.151904	0.0016
OPITA	-2.861211	1.322973	-2.162714	0.0306
INVSLS	2.248068	1.403348	1.601932	0.1092
LSLS	-0.125057	0.073824	-1.693984	0.0903

Mean dependent var	0.475138	S.D. dependent var	0.500767
S.E. of regression	0.418522	Akaike info criterion	1.063933
Sum squared resid	30.65311	Schwarz criterion	1.169960
Log likelihood	-90.28593	Hannan-Quinn criter.	1.106919
Restr. log likelihood	-125.2358	Avg. log likelihood	-0.498817
LR statistic (5 df)	69.89973	McFadden R-squared	0.279072
Probability(LR stat)	1.07E-13		

Obs with Dep=0	95	Total obs	181
Obs with Dep=1	86		