

# Université d'Orléans - Master Econométrie et Statistique Appliquée

## Econométrie des Variables Qualitatives

Examen Mai 2007. C. Hurlin

### Exercice 1 : Scoring et Défaillances d'Entreprises (12 points)

Une société souhaite construire un modèle de scoring sur les défaillances d'entreprises. On considère pour cela une base de données 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, 81 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

On considère un modèle de scoring fondé sur la spécification suivante :

$$YD_i = \begin{cases} 1 & YD_i^* > 0 \\ 0 & \text{sinon} \end{cases} \quad i = 1, \dots, N \quad (1)$$

avec  $N = 181$  et où la variable latente  $YD_i^*$  est définie par

$$YD_i^* = \alpha_0 + \alpha_1 TDTA_i + \alpha_2 GEMPL_i + \alpha_3 OPITA_i + \alpha_4 INVSLS_i + \alpha_5 LSLS_i + \varepsilon_i \quad (2)$$

où  $\varepsilon_i$  est *i.i.d.*  $(0, \sigma_\varepsilon^2)$ . On note

$$YD_i^* = x_i \alpha + \varepsilon_i \quad (3)$$

avec  $\alpha = (\alpha_0 \ \alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4 \ \alpha_5)'$  et  $x_i = (1 \ TDTA_i \ GEMPL_i \ OPITA_i \ INVSLS_i \ LSLS_i)$ .

**Question 1** (0.5 point) : Peut on donner une interprétation à la variable latente  $YD_i^*$  ? Si oui, laquelle ?

**Question 2** (1.5 points) : On considère un modèle de scoring de type logit. Sous cette contrainte, écrivez la probabilité qu'une entreprise  $i$  connaisse une défaillance en fonction des variables explicatives  $X_i$  et des paramètres  $\alpha$  et  $\sigma_\varepsilon$ .

**Question 3** (1.5 point) : A partir des résultats d'estimation du modèle logit (Figure 1), peut on estimer les paramètres  $\alpha$  et  $\sigma_\varepsilon$  ? *Vous justifiez très précisément votre réponse.*

**Question 4** (1 point) : A partir des résultats d'estimation du modèle logit (Figure 1), **évaluez la probabilité de défaillance et la cote** d'une firme ayant pour caractéristiques :

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

- Question 5** (1 point) : Commentez économiquement les résultats d'estimation du modèle logit (Figure 1).
- Question 6** (1.5 point) : Calculez l'effet marginal sur la probabilité de défaillance de l'entreprise décrite à la question 4 d'une hausse marginale du ratio dette sur actif total. Commentez vos résultats.
- Question 7** (1 point) : A partir des éléments de la Figure 2 construisez deux points de la ROC curve in sample pour ce modèle logit.
- Question 8** (1.5 point) : Commentez la sortie SAS de ce modèle logit (Figure 3) en insistant plus particulièrement (i) les problèmes de convergence, (ii) l'analyse des cotes et (iii) les propriétés prédictives du modèle.
- Question 9** (2.5 points) : On considère les 87<sup>ème</sup>, 88<sup>ème</sup> et 89<sup>ème</sup> firmes de l'échantillon pour lesquelles on a obtenu dans le modèle Logit les valeurs suivantes :

i	$YD_i$	$Index_i$	$\Pr [YD_i = 1]$
87	0	-3.06	0.044
88	1	-0.99	0.269
89	1	-0.66	0.339

On admet que les coefficients estimés du modèle Logit sont corrects. 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  $h = 1.72$ . Comparez cette probabilité à celle estimée par le modèle Logit.

### Exercice 2 : Rating et Modèle Probit Ordonné (5 points)

On considère un modèle de rating portant sur des émetteurs de titres obligataires représenté. Soit  $y_i$  une variable multinomiale associée à un émetteur  $i = 1, \dots, N$  pouvant prendre trois modalités : "AAA", "AA" et "A". On cherche à modéliser la probabilité d'attribution de la note en fonction de deux variables de bilan, notées  $x_{1,i}$  et  $x_{2,i}$  et d'une constante.

- Question 1** (0.5 point) : Quel(s) modèle(s) proposeriez vous pour répondre à la question posée. Donnez la forme du modèle proposé.
- Question 2** (1 point) : On suppose que l'on a estimé un modèle probit ordonné. On note  $\alpha_0$  la constante de ce modèle,  $\alpha_1$  le coefficient de la variable  $x_1$  et  $\alpha_2$  le coefficient de la variable  $x_2$ . On admet que les résultats d'estimation sont :

$$\hat{\alpha}_0 = -0.3 \quad \hat{\alpha}_1 = 0.8 \quad \hat{\alpha}_2 = -0.5 \tag{4}$$

Les seuils estimés sont alors les suivants :  $\hat{c}_1 = -0.43$ ,  $\hat{c}_2 = 0.43$ . Pour un émetteur caractérisé par  $x_{1,i} = x_{2,i} = 1$  quelle est la probabilité d'obtenir une note "AAA", "AA" et "A" ?

- Question 3** (1 point) : Sans effectuer de calcul, pouvez conclure quant à l'évolution de la probabilité d'obtenir une note "AAA" lorsque la variable  $x_1$  augmente de une unité ? Vous justifierez précisément votre réponse.
- Question 4** (1.5 point) : Quel sont les effets marginaux associés à la hausse de une unité de la variable  $x_1$ . On effectuera les calculs pour l'individu de référence avec  $x_{1,i} = x_{2,i} = 1$ .

**Question 5** (1 point) : Ecrivez la log vraisemblance de ce modèle associée à un échantillon  $\{y_1, \dots, y_N\}$ .

**Exercice 3 : Modèle Tobit (3 points)**

On considère le modèle suivant :

$$y_i^* = 0.8 - 0.9x_i + \varepsilon_i \quad (5)$$

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

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (6)$$

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

**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) : Ecrivez la log-vraisemblance de ce modèle.

Figure 1: Modèle de Scoring : Régression Logistique

Dependent Variable: YD  
 Method: ML - Binary Logit  
 Date: 05/10/07 Time: 15:07  
 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	-1.508675	1.043633	-1.445599	0.1483
TDTA	4.813332	1.217061	3.954881	0.0001
GEMPL	-6.122710	1.981094	-3.090571	0.0020
OPITA	-5.142973	2.357698	-2.181354	0.0292
INVSLS	3.619992	2.360702	1.533439	0.1252
LSLS	-0.215150	0.125149	-1.719149	0.0856

  

Mean dependent var	0.475138	S.D. dependent var	0.500767
S.E. of regression	0.418309	Akaike info criterion	1.066921
Sum squared resid	30.62194	Schwarz criterion	1.172948
Log likelihood	-90.55633	Hannan-Quinn criter.	1.109907
Restr. log likelihood	-125.2358	Avg. log likelihood	-0.500311
LR statistic (5 df)	69.35893	McFadden R-squared	0.276913
Probability(LR stat)	1.39E-13		

  

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

Figure 2 : Prévisions in Sample

Dependent Variable: YD	Dependent Variable: YD		
Method: ML - Binary Logit	Method: ML - Binary Logit		
Date: 05/10/07 Time: 15:53	Date: 05/10/07 Time: 15:53		
Sample: 1 181	Sample: 1 181		
Included observations: 181	Included observations: 181		
Prediction Evaluation (success cutoff C = 0.5)	Prediction Evaluation (success cutoff C = 0.3)		

  

	Estimated Equation				Estimated Equation		
	Dep=0	Dep=1	Total		Dep=0	Dep=1	Total
P(Dep=1)<=C	81	30	111	P(Dep=1)<=C	52	12	64
P(Dep=1)>C	14	56	70	P(Dep=1)>C	43	74	117
Total	95	86	181	Total	95	86	181
Correct	81	56	137	Correct	52	74	126
% Correct	85.26	65.12	75.69	% Correct	54.74	86.05	69.61
% Incorrect	14.74	34.88	24.31	% Incorrect	45.26	13.95	30.39
Total Gain*	-14.74	65.12	23.20	Total Gain*	54.74	-13.95	22.10
Percent Gain*	NA	65.12	48.84	Percent Gain*	54.74	NA	42.11