

# Université d'Orléans - Maîtrise d'Econométrie

## Econométrie des Variables Qualitatives

Examen Terminal Janvier 2003. C. Hurlin

Exercice 1 (20 points) : Modèle Logit Multinomiel

On considère un **modèle logit multinomial non ordonné** dans lequel la variable dépendante observée pour le  $i^{\text{ème}}$  individu, notée  $y_i$ , peut prendre 3 modalités codées respectivement 0, 1 et 2. La probabilité que l'individu  $i$  choisisse la modalité  $j$ ,  $\forall j = 0, 1, 2$ , est définie par :

$$Prob(y_i = j) = \frac{\exp[v(x_{i,j})]}{\exp[v(x_{i,0})] + \exp[v(x_{i,1})] + \exp[v(x_{i,2})]} \quad (1)$$

où  $x_{i,j}$  désigne un vecteur  $(1, K)$  de variables explicatives conditionnant le choix de la  $j^{\text{ème}}$  modalité pour l'individu  $i$ . Nous allons successivement envisager le cas où la fonction  $v(\cdot)$  est définie par  $v(x_{i,j}) = x_i\beta_j$  (modèle logit indépendant) et où la fonction  $v(\cdot)$  est définie par  $v(x_{i,j}) = x_{i,j}\beta$  (modèle logit conditionnel).

**Partie I : Hypothèse IIA (5 points)** On s'intéresse tout d'abord à l'hypothèse d'Indépendance des Alternatives Non Pertinentes (IIA en anglais pour *Independance of Irrelevant Alternative*) dans le cadre des modèles logit multinomiaux.

**Question 1 (1.5 point)** : Définissez de façon générale l'hypothèse d'Indépendance des Alternatives Non Pertinentes et montrez que, dans le modèle logit, la condition nécessaire à cette hypothèse se ramène à la condition  $\forall (z, k), v(x_{i,z}) - v(x_{i,k})$  indépendant de  $j, \forall j \neq z, j \neq k$

**Question 2 (2 points)** : Montrez que cette condition est satisfaite dans le cas du modèle logit multinomial indépendant et dans le cas du modèle logit multinomial conditionnel. Quelles implications ont ces résultats notamment dans le cas du modèle conditionnel ?

**Question 3 (1.5 point)** : Trouvez un exemple de fonction  $v(x_{i,j})$  pour laquelle le modèle logit multinomial universel ne satisfait pas l'hypothèse IIA.

**Partie II : Modèle Logit Indépendant (9 points)** On considère à présent un modèle logit multinomial indépendant dans lequel la probabilité que l'individu  $i$  choisisse la modalité  $j, \forall j = 0, 1, 2$ , est définie par :

$$Prob(y_i = j) = \frac{\exp(x_i\beta_j)}{\exp(x_i\beta_0) + \exp(x_i\beta_1) + \exp(x_i\beta_2)} \quad (2)$$

**Question 1 (1.5 point)** : En divisant le numérateur et le dénominateur des probabilités  $p_{i,j} = Prob(y_i = j), j = 0, 1, 2$ , par la quantité  $\exp(x_i\beta_0)$ , montrez que dans le modèle logit indépendant, les paramètres  $\beta_0, \beta_1$  et  $\beta_2$  ne sont pas identifiables simultanément. Quelle solution adopte-t-on généralement ?

**Question 2 (1.5 point)** : Montrez que la log-vraisemblance associée à un échantillon de  $N$  observations  $y = (y_1, \dots, y_N)$  dans ce modèle s'écrit, en l'absence de toute normalisation des paramètres, sous la forme suivante :

$$\log L(y, \beta_0, \beta_1, \beta_2) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_i \beta_j - \sum_{i=1}^N \log \left[ \sum_{k=0}^2 \exp(x_i \beta_k) \right] \quad (3)$$

**Question 3 (3 points)** : En posant  $\beta_0 = 0$ , dérivez à partir de l'expression précédente le vecteur du gradient ainsi que la matrice hessienne de la log-vraisemblance. Montrez que cette fonction est globalement concave. Qu'en déduisez concernant la phase d'estimation par  $MV$  des paramètres du modèle ?

**Question 4 (3 points)** : On suppose qu'il n'y a qu'une seule variable explicative ( $K = 1$ ), notée  $x_i$ . Montrer que l'effet marginal associé à une variation de cette variable sur la probabilité que l'individu  $i$  choisisse la  $j^{\text{ème}}$  modalité,  $\forall j = 0, 1, 2$ , est défini par

$$\frac{\partial p_{i,j}}{\partial x_i} = p_{i,j} (\beta_j - p_{i,1} \beta_1 - p_{i,2} \beta_2) \quad (4)$$

où  $p_{i,j} = \text{Prob}(y_i = j)$ . Qu'en déduisez vous concernant la variable exogène  $x_i$ , si l'on suppose que les paramètres vérifient

$$\beta_1 > \left( \frac{p_{i,2}}{1 - p_{i,1}} \right) \beta_2 \quad (5)$$

**Partie III : Modèle Logit Conditionnel (6 points)** On considère à présent un modèle logit multinomial conditionnel dans lequel la probabilité que l'individu  $i$  choisisse la modalité  $j$ ,  $\forall j = 0, 1, 2$ , est définie par :

$$\text{Prob}(y_i = j) = \frac{\exp(x_{i,j} \beta)}{\exp(x_{i,0} \beta) + \exp(x_{i,1} \beta) + \exp(x_{i,2} \beta)} \quad (6)$$

**Question 1 (1.5 point)** : Quel est le principal avantage du modèle logit conditionnel ?

**Question 2 (1.5 point)** : Montrez que la log-vraisemblance associée à un échantillon de  $N$  observations  $y = (y_1, \dots, y_N)$  dans ce modèle s'écrit :

$$\log L(y, \beta) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_{i,j} \beta - \sum_{i=1}^N \log \left[ \sum_{k=0}^2 \exp(x_{i,k} \beta) \right] \quad (7)$$

**Question 3 (3 points)** : On suppose que l'on dispose pour chacun des  $N$  individus de l'échantillon d'une évaluation, notée  $\tilde{x}_{i,3}$ , du vecteur des variables explicatives pour une quatrième modalité codée 3, non encore disponible dans l'espace de choix des agents. (i) Décrivez précisément la démarche permettant d'estimer la probabilité que chaque individu  $i$  choisisse cette modalité 3 lorsqu'elle sera effectivement mise en place. On notera  $\hat{p}_{i,3}$  l'estimateur obtenu. (ii) A partir d'un intervalle de confiance sur l'estimateur du maximum de vraisemblance de  $\beta$ , proposez un encadrement de la vraie probabilité  $p_{i,3}$ .

## Exercice 2 (10 points) : Modèle Probit Ordonné

On considère une application tirée d'une étude de J. Gunther de la *Federal Reserve Bank de Dallas*, intitulée "*Between a Rock and a Hard Place : The CRA-Safety and Soundness Pinch*". Cette étude porte sur le *Community Reinvestment Act (CRA)*, loi promulgué aux Etats Unis en 1977 et visant à encourager les institutions de dépôts (banques et autres institutions financières) à répondre aux besoins en crédit des communautés dans lesquelles elles opèrent. Dans le cadre de cette loi, les banques commerciales sont régulièrement évaluées par différentes instances de contrôle<sup>1</sup> au regard des objectifs du CRA. La performance d'une institution de dépôt est contrôlée à partir des informations disponibles sur cette institution (capacités, contraintes diverses, stratégie..), des informations sur la communauté dans laquelle elle opère (démographie, données économiques, prêts, investissements..) et des informations sur ses concurrents et sur l'état du marché. Une notation (ou rating) est alors attribué selon quatre modalité. Si l'on considère une institution  $i$ , la note  $y_i$  est codée de la façon suivante :

$$y_i = \begin{cases} 1 & \text{performances remarquables} \\ 2 & \text{performances satisfaisantes} \\ 3 & \text{performances à améliorer} \\ 4 & \text{performances déplorables} \end{cases}$$

On dispose d'un échantillon de 350 observations et l'on se propose de modéliser ces rating en fonction de plusieurs variables explicatives :

- *ASS* : logarithme de l'actif de l'institution de dépôt.
- *EQU* : ratio capitaux propres sur actif.
- *GROWTH* : taux de croissance du P.I.B. de l'état dans lequel la banque opère.
- *LOA* : ratio prêts sur actif total de la banque

Afin de modéliser la notation des institutions en fonction de ces caractéristiques, on utilise un **modèle probit multinomial ordonné**. On suppose que l'attribution du rating dépend de la valeur prise par une variable latente continue inobservable  $y_i^*$  telle que :

$$y_i^* = x_i\beta + \varepsilon_i \quad (8)$$

où  $x_i$  désigne le vecteur (1,4) des quatre caractéristiques de la banque  $i$  et  $\varepsilon_i$  est *i.i.d*  $N(0, \sigma_\varepsilon^2)$ . On suppose que le rating est déterminé par le modèle :

$$y_i = \begin{cases} 1 & \text{si } y_i^* < c_1 \\ 2 & \text{si } c_1 \leq y_i^* < c_2 \\ 3 & \text{si } c_2 \leq y_i^* < c_3 \\ 4 & \text{si } y_i^* > c_3 \end{cases} \quad (9)$$

où  $(c_1, c_2, c_3)^3 \in \mathbb{R}$ .

---

<sup>1</sup>Office of the Comptroller of the Currency (OCC), Board of Governors of the Federal Reserve System (FRB), Office of Thrift Supervision (OTS), and Federal Deposit Insurance Corporation (FDIC).

**Question 1** (1 point) : Justifiez l'utilisation du modèle probit ordonné dans ce cas.

**Question 2** (1.5 points) : Donnez, en fonction des paramètres  $\beta$ ,  $c_1$ ,  $c_2$ ,  $c_3$  et  $\sigma_\varepsilon^2$ , la formulation des probabilités théoriques que le rating de l'institution de dépôt  $i$ ,  $\forall i = 1, \dots, N$ , prenne les valeurs 1, 2, 3 et 4 :

$$Prob(y_i = 1) \quad Prob(y_i = 2) \quad Prob(y_i = 3) \quad Prob(y_i = 4)$$

**Question 3** (1.5 points) : Commentez les résultats d'estimation par maximum de vraisemblance reportés sur la figure (1) :

Figure 1: Probit Ordonné

Dependent Variable: Y				
Method: ML - Ordered Probit				
Date: 01/06/03 Time: 12:07				
Sample: 1 350				
Included observations: 350				
Number of ordered indicator values: 4				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
	Coefficient	Std. Error	z-Statistic	Prob.
ASS	-0.163769	0.050154	-3.265342	0.0011
EQU	4.036511	1.627355	2.480412	0.0131
GROWTH	-8.170760	2.742002	-2.979852	0.0029
LOA	-1.100523	0.324646	-3.389923	0.0007
Limit Points				
LIMIT_2:C(5)	-3.076116	0.656269	-4.687282	0.0000
LIMIT_3:C(6)	-2.263611	0.648269	-3.491672	0.0005
LIMIT_4:C(7)	-1.276862	0.648353	-1.969394	0.0489
Akaike info criterion	2.566294	Schwarz criterion	2.643453	
Log likelihood	-442.1015	Hannan-Quinn criter.	2.597006	
Restr. log likelihood	-473.1244	Avg. log likelihood	-1.263147	
LR statistic (4 df)	62.04580	LR index (Pseudo-R2)	0.065570	
Probability(LR stat)	1.08E-12			

**Question 4** (3 points) : Calculez les réalisations des estimateurs des probabilités que la 70<sup>ème</sup> institution de dépôt reçoive une notation "performances remarquables", "performances satisfaisantes", "performances à améliorer" ou "performances déplorables", sachant que pour cette institution a les caractéristiques suivantes :

$$ASS = 9.949 \quad EQU = 0.066 \quad GROWTH = 0.046 \quad LOA = 0.660$$

Qu'en concluez vous pour le rating de cette banque ?

**Question 5** (3 points) : En utilisant les résultats du premier chapitre de cours, déterminez la variation de la probabilité que cette banque obtienne la rating "performances remarquables" engendrée par une augmentation d'une unité de son ratio capitaux propres sur actif.

# Université d'Orléans - Maîtrise Econométrie

## Econométrie des Variables Qualitatives

Correction Examen Terminal Janvier 2003. C. Hurlin

Exercice 1 (20 points) : Modèle Logit Multinomial

### Partie I : Hypothèse IIA (5 points)

#### Question 1 :

Cette hypothèse traduit le fait que le rapport de deux probabilités associés à deux événements particuliers est indépendant des autres événements (*0.5 point*). Ainsi, l'hypothèse est satisfaite si  $\forall (z, k)$  le rapport de probabilité

$$\frac{p_{i,z}}{p_{i,k}} = \frac{Prob(y_i = j)}{Prob(y_i = k)}$$

est indépendant des alternatives  $j$  autres que  $z$  et  $k$ . Dans le modèle logit multinomial, on a :

$$\frac{p_{i,z}}{p_{i,k}} = \frac{\exp[v(x_{i,z})]}{\exp[v(x_{i,k})]} = \exp[v(x_{i,z}) - v(x_{i,k})]$$

Ainsi, la condition garantissant l'hypothèse IIA dans un modèle logit multinomial se ramène à

$$\forall (z, k), v(x_{i,z}) - v(x_{i,k}) \text{ indépendant de } j, \forall j \neq z, j \neq k \quad (1 \text{ point})$$

#### Question 2

Dans le cas du modèle logit multinomial indépendant, on a :

$$v(x_{i,z}) - v(x_{i,k}) = x_i \beta_z - x_i \beta_k = x_i (\beta_z - \beta_k)$$

Cette différence est indépendante de la valeur des paramètres  $\beta_j$  associés aux modalités  $j$  autres que  $z$  et  $k$ . L'hypothèse IIA est vérifiée dans le modèle logit multinomial indépendant (*0.5 point*).

Dans le cas du modèle logit multinomial conditionnel, on a :

$$v(x_{i,z}) - v(x_{i,k}) = x_{i,z} \beta - x_{i,k} \beta = (x_{i,z} - x_{i,k}) \beta$$

Cette différence est indépendante de la valeur des variables explicatives  $x_{i,j}$  associées aux modalités  $j$  autres que  $z$  et  $k$ . L'hypothèse IIA est vérifiée dans le modèle logit multinomial conditionnel (*0.5 point*).

On a vu dans le cours que l'hypothèse IIA n'est que rarement (exemple bus bleu, bus rouge) satisfaite, ce qui pose le problème de la cohérence d'une modélisation de type logit multinomial pour rendre compte de choix probabilistes (*0.5 point*). Toutefois, ces modèles logit multinomial sont très souvent utilisés compte tenu de la simplicité de leur mise en oeuvre pratique. Dans le cadre du modèle conditionnel, c'est cette hypothèse IIA qui permet notamment de faire des prévisions sur les probabilités que les agents retiennent des modalités virtuelles, c'est à dire non encore disponible dans le cadre des choix courants. L'exemple typique est le modèle hypothétique de choix de transport (*0.5 point*).

### Question 3

Il existe bien évidemment une infinité de fonctions  $v(x_{i,j})$  pour laquelle le modèle logit multinomial universel ne satisfait pas l'hypothèse *IIA* : on peut par exemple considérer la fonction :

$$v(x_{i,j}) = \left( x_{i,j} - \frac{1}{m+1} \sum_{s=0}^m x_{i,s} \right)^\beta \quad (1 \text{ point})$$

où  $m+1$  désigne le nombre de modalités, ici  $m=2$ . Dans ce cas, on a en effet :

$$\begin{aligned} \frac{p_{i,z}}{p_{i,k}} &= \exp[v(x_{i,z}) - v(x_{i,k})] \\ &= \exp \left[ \left( x_{i,z} - \frac{1}{m+1} \sum_{s=0}^m x_{i,s} \right)^\beta - \left( x_{i,k} - \frac{1}{m+1} \sum_{s=0}^m x_{i,s} \right)^\beta \right] \end{aligned}$$

Cette expression dépend des valeurs des variables explicatives pour les autres modalités  $j$  différentes de  $z$  et  $k$ . L'hypothèse *IIA* n'est pas satisfaite : ainsi un modèle logit universel ne satisfait pas nécessairement l'hypothèse *IIA*. Les modèles logit conditionnels et logit indépendant, qui sont des cas particuliers du modèle logit universel, vérifient cette hypothèse, mais ce n'est pas un résultat général. (0.5 point)

## Partie II : Modèle Logit Indépendant (9 points)

### Question 1

Les probabilités s'écrivent :

$$\begin{aligned} Prob(y_i = j) &= \frac{\exp(x_i \beta_j)}{\sum_{k=0}^2 \exp(x_i \beta_k)} \quad (0.5 \text{ point}) \\ &= \frac{\exp(x_i \beta_j) / \exp(x_i \beta_0)}{\sum_{k=0}^2 \exp(x_i \beta_k) / \exp(x_i \beta_0)} \end{aligned}$$

On obtient alors :

$$\begin{aligned} Prob(y_i = 0) &= \frac{1}{1 + \exp[x_i(\beta_1 - \beta_0)] + \exp[x_i(\beta_2 - \beta_0)]} \\ Prob(y_i = 1) &= \frac{1}{1 + \exp[x_i(\beta_0 - \beta_1)] + \exp[x_i(\beta_2 - \beta_1)]} \\ Prob(y_i = 2) &= \frac{1}{1 + \exp[x_i(\beta_0 - \beta_2)] + \exp[x_i(\beta_1 - \beta_2)]} \end{aligned}$$

Par construction  $\sum_{j=0}^2 p_j = 1$ , on dispose ainsi de deux probabilités indépendantes pour déterminer trois différences de paramètres  $(\beta_1 - \beta_0)$ ,  $(\beta_2 - \beta_0)$  et  $(\beta_2 - \beta_1)$ . Naturellement, ces différences de paramètres ne sont identifiables que si l'on impose une contrainte de normalisation du type  $\beta_0 = 0$ . Dès lors, on a deux probabilités indépendantes qui nous permettent d'identifier deux paramètres  $\beta_1$  et  $\beta_2$ . Ces paramètres s'interprètent comme des écarts au vecteur  $\beta_0$ . (1 point)

### Question 2

La log-vraisemblance associée à un échantillon de  $N$  observations  $y = (y_1, \dots, y_N)$  dans ce modèle logit indépendant à trois modalités s'écrit :

$$\log L(y, \beta_0, \beta_1, \beta_2) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log [\text{Prob}(y_i = j)] \quad (0.5 \text{ point})$$

$$\sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log \left[ \frac{\exp(x_i \beta_j)}{\sum_{k=0}^2 \exp(x_i \beta_k)} \right]$$

avec

$$y_{ij} = \begin{cases} 1 & \text{si } y_i = j \\ 0 & \text{sinon} \end{cases} \quad \forall i = 1.., N \quad \forall j = 0, 1, 2$$

On obtient alors :

$$\log L(y, \beta_0, \beta_1, \beta_2) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_i \beta_j - \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log \left[ \sum_{k=0}^2 \exp(x_i \beta_k) \right]$$

$$= \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_i \beta_j - \sum_{i=1}^N h_i \log \left[ \sum_{k=0}^2 \exp(x_i \beta_k) \right]$$

avec  $\forall i = 1.., N$  :

$$h_i = \sum_{j=0}^2 y_{i,j} = y_{i,0} + y_{i,1} + y_{i,2} = 1 \quad (0.5 \text{ point})$$

En effet, on sait que la variable  $y_i$  ne peut prendre qu'une seule et même valeur parmi les 3 modalités, dès lors  $\sum_{j=0}^2 y_{i,j} = 1$ . Ainsi, on obtient finalement :

$$\log L(y, \beta_0, \beta_1, \beta_2) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_i \beta_j - \sum_{i=1}^N \log \left[ \sum_{k=0}^2 \exp(x_i \beta_k) \right] \quad (0.5 \text{ point})$$

### Question 3

Sous l'hypothèse  $\beta_0 = 0$ , on a :

$$\log L(y, \beta_1, \beta_2) = \sum_{i=1}^N \sum_{j=1}^2 y_{i,j} x_i \beta_j - \sum_{i=1}^N \log \left[ 1 + \sum_{k=1}^2 \exp(x_i \beta_k) \right]$$

Dès lors, le vecteur du gradient s'écrit :

$$G(y, \beta_1, \beta_2) = \begin{pmatrix} \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_1} \\ \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_2} \end{pmatrix}$$

avec pour  $z = 1, 2$  :

$$\frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_z} = \sum_{i=1}^N y_{i,z} x'_i - \frac{\partial}{\partial \beta_z} \left( \sum_{i=1}^N \log \left[ 1 + \sum_{k=1}^2 \exp(x_i \beta_k) \right] \right) \quad (1 \text{ point})$$

$$= \sum_{i=1}^N y_{i,z} x'_i - \sum_{i=1}^N \left( \frac{\exp(x_i \beta_z)}{1 + \sum_{k=1}^2 \exp(x_i \beta_k)} \right) x'_i$$

$$= \sum_{i=1}^N (y_{i,z} - p_{i,z}) x'_i \quad (1)$$

avec  $p_{i,z} = \text{Prob}(y_i = z)$ . De la même façon, on peut dériver la matrice hessienne :

$$H(y, \beta_1, \beta_2) = \begin{pmatrix} \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_1^2} & \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_1 \partial \beta_2} \\ \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_1 \partial \beta_2} & \frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_2^2} \end{pmatrix}$$

Les éléments de cette matrice pour  $z = 1, 2$  et  $k = 1, 2$  sont définis par :

$$\frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_z \partial \beta_k} = \frac{\partial}{\partial \beta_k} \left( \sum_{i=1}^N (y_{i,z} - p_{i,z}) x'_i \right) = - \sum_{i=1}^N \left( \frac{\partial p_{i,z}}{\partial \beta_k} \right) x'_i$$

avec :

$$\frac{\partial p_{i,z}}{\partial \beta_k} = \frac{\partial}{\partial \beta_k} \left[ \frac{\exp(x_i \beta_z)}{\sum_{h=0}^2 \exp(x_i \beta_h)} \right]$$

Si  $k \neq z$ , alors cette expression devient :

$$\frac{\partial p_{i,z}}{\partial \beta_k} = -x_i \left( \frac{\exp(x_i \beta_z) \exp(x_i \beta_k)}{\left[ \sum_{h=0}^2 \exp(x_i \beta_h) \right]^2} \right) = -x_i p_{iz} p_{ik} \quad \forall k \neq z$$

Si  $k = z$ , alors cette expression devient :

$$\begin{aligned} \frac{\partial p_{i,z}}{\partial \beta_k} &= -x_i \left( \frac{\exp(x_i \beta_z) \left[ \sum_{h=0}^2 \exp(x_i \beta_h) \right] - \exp(x_i \beta_z) \exp(x_i \beta_k)}{\left[ \sum_{h=0}^2 \exp(x_i \beta_h) \right]^2} \right) \\ &= -x_i (p_{iz} - p_{i,z} p_{ik}) \quad \forall k = z \end{aligned}$$

On a donc finalement :

$$\frac{\partial \log L(y, \beta_1, \beta_2)}{\partial \beta_z \partial \beta_k} = \begin{cases} - \sum_{i=1}^N p_{iz} p_{ik} x_i x'_i & \forall k \neq z \\ - \sum_{i=1}^N p_{iz} (1 - p_{ik}) x_i x'_i & \forall k = z \end{cases}$$

On retrouve ainsi l'expression du cours :

$$\frac{\partial^2 \log L(y, \beta_1, \beta_2, \dots, \beta_m)}{\partial \beta_j \partial \beta'_k} = - \sum_{i=1}^N p_{i,j} (\mathbb{I}_{j,k} - p_{i,k}) x'_i x_i \quad (1 \text{ point})$$

On peut alors montrer que la fonction de log-vraisemblance est globalement concave en  $(\beta_1, \beta_2)$ . Ainsi, la fonction de log-vraisemblance d'un modèle logit multinomial indépendant est globalement concave. Par conséquent, on peut utiliser différents algorithmes d'optimisation numérique propres à ce type de problème (Newton Raphson par exemple) et les résultats ne sont pas sensibles au choix des conditions initiales de ces algorithmes.

#### Question 4

L'effet marginal associé à une variation de cette variable sur la probabilité que l'individu  $i$  choisisse la  $j^{\text{ème}}$  modalité,  $\forall j = 0, 1, 2$ , est défini par :

$$\frac{\partial p_{i,j}}{\partial x_i} = \frac{\partial}{\partial x_i} \left[ \frac{\exp(x_i \beta_j)}{\sum_{h=0}^2 \exp(x_i \beta_h)} \right] = \frac{\partial}{\partial x_i} \left[ \frac{\exp(x_i \beta_j)}{H(x_i)} \right] \quad (2)$$

avec  $H(x_i) = \sum_{h=0}^2 \exp(x_i \beta_h)$ . On a donc :

$$\begin{aligned} \frac{\partial p_{i,j}}{\partial x_i} &= \frac{1}{H(x_i)^2} \left[ \frac{\partial \exp(x_i \beta_j)}{\partial x_i} \times H(x_i) - \exp(x_i \beta_j) \frac{\partial H(x_i)}{\partial x_i} \right] \\ &= \frac{1}{H(x_i)^2} \left[ \beta_j \exp(x_i \beta_j) \times H(x_i) - \exp(x_i \beta_j) \sum_{z=0}^2 \beta_z \exp(x_i \beta_z) \right] \\ &= \beta_j \left[ \frac{\exp(x_i \beta_j)}{H(x_i)} \right] - \frac{\exp(x_i \beta_j)}{H(x_i)} \sum_{z=0}^2 \beta_z \frac{\exp(x_i \beta_z)}{H(x_i)} \end{aligned}$$

On obtient finalement :

$$\frac{\partial p_{i,j}}{\partial x_i} = \beta_j p_{i,j} - p_{i,j} \sum_{z=0}^2 \beta_z p_{i,z} = p_{i,j} (\beta_j - p_{i,1} \beta_1 - p_{i,2} \beta_2) \quad (2 \text{ points})$$

Ainsi si

$$\beta_1 > \frac{p_{i,2} \beta_2}{1 - p_{i,1}} \iff \beta_1 - p_{i,1} \beta_1 - p_{i,2} \beta_2 > 0$$

Puisque  $p_{i,1} \geq 0$ , on en déduit que :

$$\frac{\partial p_{i,1}}{\partial x_i} = p_{i,1} (\beta_1 - p_{i,1} \beta_1 - p_{i,2} \beta_2) \geq 0 \quad (0.5 \text{ point})$$

Ainsi une augmentation d'une unité de la variable explicative induit ici une augmentation de la probabilité que l'agent  $i$  choisisse la modalité 1 (*0.5 point*). C'est en effet la comparaison de la valeur du coefficient obtenue dans la modalité 1, c'est à dire  $\beta_1$ , à la moyenne pondérée par les probabilités des autres coefficients dans les différents régimes :

$$p_{i,0} \beta_0 + p_{i,1} \beta_1 + p_{i,2} \beta_2 = p_{i,1} \beta_1 + p_{i,2} \beta_2$$

qui permet de déterminer si l'effet marginal est positif ou négatif.

### Partie III : Modèle Logit Conditionnel (6 points)

#### Question 1

L'avantage de ce modèle se situe dans la possibilité qui est offerte de prédire la probabilité d'une nouvelle modalité (virtuelle) en fonction de variables explicatives simulées. Le modèle logit conditionnel permet en effet d'estimer la probabilité associée à une modalité virtuelle de la façon suivante :

$$\widehat{P}_{m+1} = \frac{\exp(\widehat{x}_{i,m+1}^* \widehat{\beta})}{1 + \sum_{k=1}^m \exp(x_{i,k}^* \widehat{\beta}) + \exp(\widehat{x}_{i,m+1}^* \widehat{\beta})} \quad (3)$$

où  $\widehat{\beta}$  désigne un estimateur convergent de  $\beta$  obtenu sur la base des modalités  $j = 0, \dots, m$  existant et où  $\widehat{x}_{i,m+1}^*$  est une estimation des caractéristiques exogènes associées à la  $m + 1^{\text{ème}}$  modalité virtuelle. (*1.5 point*)

#### Question 2

La log-vraisemblance associée à un échantillon de  $N$  observations  $y = (y_1, \dots, y_N)$  dans ce modèle logit conditionnel à trois modalités s'écrit :

$$\log L(y, \beta) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log [\text{Prob}(y_i = j)] \quad (0.5 \text{ point})$$

$$\sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log \left[ \frac{\exp(x_{i,j}\beta)}{\sum_{k=0}^2 \exp(x_{i,k}\beta)} \right]$$

avec

$$y_{ij} = \begin{cases} 1 & \text{si } y_i = j \\ 0 & \text{sinon} \end{cases} \quad \forall i = 1.., N \quad \forall j = 0, 1, 2$$

On obtient alors :

$$\log L(y, \beta) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_{i,j} \beta - \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} \log \left[ \sum_{k=0}^2 \exp(x_{i,k}\beta) \right]$$

$$= \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_{i,j} \beta - \sum_{i=1}^N h_i \log \left[ \sum_{k=0}^2 \exp(x_{i,k}\beta) \right]$$

avec  $\forall i = 1.., N$  :

$$h_i = \sum_{j=0}^2 y_{i,j} = y_{i,0} + y_{i,1} + y_{i,2} = 1 \quad (0.5 \text{ point})$$

En effet, on sait que la variable  $y_i$  ne peut prendre qu'une seule et même valeur parmi les 3 modalités, dès lors  $\sum_{j=0}^2 y_{i,j} = 1$ . Ainsi, on obtient finalement :

$$\log L(y, \beta) = \sum_{i=1}^N \sum_{j=0}^2 y_{i,j} x_{i,j} \beta - \sum_{i=1}^N \log \left[ \sum_{k=0}^2 \exp(x_{i,k}\beta) \right] \quad (0.5 \text{ point})$$

### Question 3

La démarche se fait en trois étapes : (i) On estime le vecteur  $\beta$  sur la base des 3 modalités existant par une procédure de maximum de vraisemblance. Soit  $\hat{\beta}$  l'estimateur ainsi obtenu :

$$\hat{\beta} = \arg \max_{\{\beta\}} \log L(y, \beta) \quad (4)$$

On note  $\Omega$  la matrice de variance covariance asymptotique de cet estimateur. Sous des conditions standard on a :

$$\sqrt{N} (\hat{\beta} - \beta_0) \xrightarrow{N \rightarrow \infty} N \left[ 0, I(\beta_0)^{-1} \right]$$

$$\Omega = I(\beta_0) = -E \left[ \frac{\partial^2 \log L(y, \beta)}{\partial \beta \partial \beta'} \right]_{\beta = \beta_0}$$

où  $\beta_0$  désigne la vraie valeur du vecteur de paramètres (1 point). (ii) On estime alors la probabilité que l'agent  $i$  choisisse la modalité 3 comme suit :

$$\hat{p}_{i,3} = \frac{\exp(\tilde{x}_{i,3}\hat{\beta})}{\sum_{k=0}^2 \exp(x_{i,k}\hat{\beta}) + \exp(\tilde{x}_{i,3}\hat{\beta})} \quad (5)$$

On connaît ainsi pour chaque individu  $i$  une estimation de la probabilité que cet agent choisisse la modalité 3 lorsqu'elle sera effectivement disponible (1 point). (iii) On cherche alors la forme de l'intervalle de confiance sur  $\hat{p}_{i,3}$ . Pour simplifier on admet que les évaluations sur les variables explicatives sont déterministes. On sait que sous des conditions standard pour un risque de premier espèce de  $\alpha\%$  :

$$Prob(\underline{\beta} \leq \beta \leq \bar{\beta}) = 1 - \alpha$$

avec  $\underline{\beta} = \beta - C_\alpha \text{diag}(\Omega^{1/2})$  et  $\bar{\beta} = \beta + C_\alpha \text{diag}(\Omega^{1/2})$ . La probabilité estimée s'écrit alors :

$$\hat{p}_{i,3} = \frac{1}{1 + \sum_{k=0}^2 \exp[(x_{i,k} - \tilde{x}_{i,3})\hat{\beta}]} \quad (6)$$

Si l'on suppose que c'est une fonction croissante de  $\hat{\beta}$  on a alors un encadrement du type (1 point) :

$$\left\{ 1 + \sum_{k=0}^2 \exp[(x_{i,k} - \tilde{x}_{i,3})\bar{\beta}] \right\}^{-1} \leq p_{i,3} \leq \left\{ 1 + \sum_{k=0}^2 \exp[(x_{i,k} - \tilde{x}_{i,3})\underline{\beta}] \right\}^{-1}$$

## Exercice 2 (10 points) : Modèle Logit Multinomial

### Question 1 :

Les modèles ordonnés sont utilisés lorsque les valeurs prises par la variable multinomiale correspondent à des intervalles dans lesquels va se trouver une seule variable latente inobservable continue. Ainsi, un modèle polytomique univarié ordonné est un modèle dans lequel on a une variable, plusieurs modalités, et un ordre naturel sur ces modalités. Ici tel est le cas pour la variable de rating (1 point).

### Question 2 :

Si l'on note  $\Phi$  la fonction de répartition de la loi normale centrée réduite :

$$Prob(y_i = 1) = Prob(y_i^* < c_1) = \Phi\left(\frac{c_1}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) \quad (0.5 \text{ point})$$

$$Prob(y_i = 2) = Prob(c_1 \leq y_i^* < c_2) = \Phi\left(\frac{c_2}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) - \Phi\left(\frac{c_1}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) \quad (7)$$

$$Prob(y_i = 3) = Prob(c_2 \leq y_i^* < c_3) = \Phi\left(\frac{c_3}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) - \Phi\left(\frac{c_2}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) \quad (0.5 \text{ point})$$

$$Prob(y_i = 4) = Prob(y_i^* > c_3) = 1 - \Phi\left(\frac{c_3}{\sigma_\varepsilon} - x_i \frac{\beta}{\sigma_\varepsilon}\right) \quad (0.5 \text{ point})$$

### Question 3 :

Il s'agit de l'estimation d'un probit ordonné par MV. Tout d'abord on remarque que la procédure d'optimisation de la vraisemblance a convergé. On dispose des estimateurs des paramètres  $\beta/\sigma_\varepsilon$  pour les quatre variables explicatives, mais encore une fois on ne peut pas les interpréter directement (0.5 point). On constate qu'il n'y a pas de constante dans le vecteur des variables explicatives : le vecteur  $x_i$  ne peut contenir de constante pour les mêmes raisons d'identification qui avaient été évoquées dans le cas du modèle dichotomique en ce qui concerne la normalisation du seuil  $\gamma$ . On ne peut identifier à la fois le paramètre associé à la constante et les seuils  $c_j$ . Enfin, on dispose des estimateurs des paramètres  $c_1/\sigma_\varepsilon$ ,  $c_2/\sigma_\varepsilon$  et  $c_3/\sigma_\varepsilon$ . (0.5 point). Tous les coefficients sont statistiquement différents de zéro pour un risque de première espèce de 5% (z-statistiques) et le test joint de la nullité de tous les coefficients (LR statistique) permet de rejeter  $H_0$  (0.5 point)

### Question 4 :

Si l'on note  $\hat{\beta}$  et  $\hat{c}_j$  les estimateurs du MV des paramètres  $\beta/\sigma_\varepsilon$  et  $c_j/\sigma_\varepsilon$ , on a (1 point) :

$$\hat{p}_{i,1} = \Phi\left(\hat{c}_1 - x_i \hat{\beta}\right) \quad (8)$$

$$\hat{p}_{i,2} = \Phi(\hat{c}_2 - x_i\hat{\beta}) - \Phi(\hat{c}_1 - x_i\hat{\beta}) \quad (9)$$

$$\hat{p}_{i,3} = \Phi(\hat{c}_3 - x_i\hat{\beta}) - \Phi(\hat{c}_1 - x_i\hat{\beta}) \quad (10)$$

$$\hat{p}_{i,4} = 1 - \Phi(\hat{c}_3 - x_i\hat{\beta}) \quad (11)$$

où  $\hat{p}_{i,j} = Prob(y_i = j)$ ,  $j = 1, 2, 3, 4$ . Après évaluations numériques on a (2 points) :

$$\hat{p}_{i,1} \simeq 0.272 \quad (12)$$

$$\hat{p}_{i,2} \simeq 0.309 \quad (13)$$

$$\hat{p}_{i,3} \simeq 0.302 \quad (14)$$

$$\hat{p}_{i,4} \simeq 0.117 \quad (15)$$

On en déduit que la plus grande probabilité est que cette banque obtienne un rating "performances satisfaisantes" ou un rating "performances à améliorer".

### Question 5 :

On suppose ici pour simplifier les notations que  $\sigma_\varepsilon = 1$ . On sait que la probabilité que cette banque obtienne "performances remarquables" est donnée par  $p_{i,1} = \Phi(c_1 - x_i\beta)$ . On cherche donc à déterminer :

$$\frac{\partial p_{i,1}}{\partial equ_i} = \frac{\partial \Phi(c_1 - x_i\beta)}{\partial equ_i} \quad (16)$$

Si l'on note  $f(\cdot)$  la fonction de densité des résidus du modèle, on a vu dans le cadre des modèles dichotomiques que :

$$\frac{\partial F(x_i\beta)}{\partial x_i^{[j]}} = \frac{\partial F(x_i\beta)}{\partial(x_i\beta)} \frac{\partial(x_i\beta)}{\partial x_i^{[j]}} = \frac{\partial F(x_i\beta)}{\partial(x_i\beta)} \beta_j = f(x_i\beta) \cdot \beta_j$$

Donc ici, on a :

$$\frac{\partial \Phi(c_1 - x_i\beta)}{\partial equ_i} = \frac{\partial \Phi(c_1 - x_i\beta)}{\partial(x_i\beta)} \frac{\partial(x_i\beta)}{\partial equ_i} = \frac{\partial \Phi(c_1 - x_i\beta)}{\partial(x_i\beta)} \beta_{equ} = -\phi(c_1 - x_i\beta) \beta_{equ}$$

où  $\beta_{equ}$  désigne le coefficient associé à la variable  $EQU$  et  $\phi(\cdot)$  la fonction de densité de la loi normale centrée réduite. Ainsi dans notre application un estimateur de l'effet marginal de  $EQU$  sur la probabilité du rating "performances remarquables" est :

$$\frac{\partial p_{i,1}}{\partial equ_i} = -\phi(\hat{c}_1 - x_i\hat{\beta}) \hat{\beta}_{equ}$$

où  $\hat{\beta}$  est un estimateur de  $\beta/\sigma_\varepsilon$ ,  $\hat{c}_1$  est un estimateur de  $c_1/\sigma_\varepsilon$  et  $\hat{\beta}_{equ}$  un estimateur de  $\beta_{equ}/\sigma_\varepsilon$ . On a alors :

$$\frac{\partial p_{i,1}}{\partial equ_i} = -\phi(-3.076 - x_i\hat{\beta}) \times 4.036$$

avec  $x_i\hat{\beta} = -2.465$ . On a donc :

$$\frac{\partial p_{i,1}}{\partial equ_i} = -1.335$$

Une augmentation du ratio capitaux propres sur actif réduit donc la probabilité que la banque obtienne le rating "performances remarquables" au regard des objectifs du *CRA*.