MASTER ECONOMETRIE ET STATISTIQUE APPLIQUEE (ESA)

Université d'Orléans

Econométrie des Variables Qualitatives

Chapitre 2

Modèles Multinomiaux

Modèles Logit Multinomiaux Ordonnées et non Ordonnés

Christophe Hurlin

Polycopié de Cours

Master Econométrie et Statistique Appliquée (ESA)
Université d'Orléans
Faculté de Droit, d'Economie et de Gestion
Bureau A 224
Rue de Blois – BP 6739
45067 Orléans Cedex 2
www.univ-orleans.fr/deg/masters/ESA/

January 21, 2003

Contents

Introduction

Nous allons à présent envisager le cas des **modèle multinomiaux**, ou plus précisément *des modèles à variable expliquée qualitative multinomiale (ou polytomique).* Ce sont des modèles dans lesquels la variable expliquée peut prendre plus de deux modalités[1]. Nous allons voir qu'il existe trois catégories de modèles multinomiaux :

1. Modèles **multinomiaux ordonnés**
2. Modèles **multinomiaux séquentiels**
3. Modèles **multinomiaux non ordonnés**

Dans la pratique, les modèles non ordonnés sont les plus fréquents, c'est pourquoi nous leur accorderons une attention toute particulière. Dans cette catégorie, on trouve notamment le modèle **logit multinomial et le modèle logit conditionnel de McFadden** qui sont les modèles les plus utilisés et qui constituent une extension du logit binaire étudié dans le chapitre précédent. Nous verrons toutefois, que si ces modèles sont simples à utiliser, ils posent toutefois un problème de cohérence en raison d'un propriété peu réaliste d'indépendance des états non pertinents. C'est pourquoi, des modèles alternatifs ont été développés comme le modèle **logit hierarchisé** ou le modèle **probit multinomial**. Ces derniers requièrent toutefois des techniques d'estimation relativement complexes.

Commençons par décrire le cadre général de ces modèles multinomiaux.

Definition 0.1. ***On considère un modèle multinomial dans le lequel la variable dépendante qualitative observée pour le $i^{ème}$ individu $\forall i = 1, .., N$, notée y_i, peut prendre $m_i + 1$ modalités indicées $j = 0, 1, 2, .., m_i$, supposées mutuellement exclusives pour chaque individu i :***

$$\sum_{j=0}^{m_i+1} Prob\,(y_i = j) = 1 \quad \forall i = 1, .., N \tag{0.1}$$

La probabilité associée à chaque réponse est définie par :

$$Prob\,(y_i = j) = F_{ij}\,(x, \beta) \quad \forall i = 1.., N \;\; \forall j = 0, 1, .., m_i \tag{0.2}$$

Partant de cette définition, on peut faire quatre remarques.

1. On remarque tout d'abord que le nombre de modalités m_i prises par la variable dépendante y_i peut dépendre de l'individu : on peut avoir $m_z \neq m_k$.
2. La fonction de répartition $F_{ij}\,(x, \beta)$ correspond à la probabilité que l'individu i choisisse la modalité j en fonction des variables explicatives x et du vecteur de paramètres β. Cette fonction peut ainsi différer suivant les individus (indice i) mais aussi suivant les modalités (indice j).

[1] Voir l''introduction du premier chapitre pour la définition générale des modèles à variable qualitative.

3. Dans un modèle multinomial, la probabilité associée à la $m_i+1^{ème}$ modalité (généralement l'événement codé en 0) n'a pas besoin d'être spécifiée puisqu'elle peut être calculée à partir des m_i probabilité comme suit :

$$Prob\left(y_i = 0\right) = 1 - \sum_{j=1}^{m_i} F_{ij}\left(x, \beta\right) \quad \forall i = 1.., N$$

Si l'on définit $\sum_{i=1}^{N}\left(m_i + 1\right)$ variables binaires y_{ij} telles que :

$$y_{ij} = \begin{cases} 1 & \text{si } y_i = j \\ 0 & \text{sinon} \end{cases} \quad \forall i = 1.., N \;\; \forall j = 0, 1, .., m_i \tag{0.3}$$

alors on peut écrire la vraisemblance associé à l'échantillon $y = \left(y_{10}, ..., y_{1m_1}, ..., y_{N0}, ..., y_{Nm_N}\right)$ comme le produit des probabilités associées aux différentes modalités, et ceci pour tous les individus :

$$L\left(y, \beta\right) = \prod_{i=1}^{N}\prod_{j=0}^{m_i} F_{ij}\left(x, \beta\right)^{y_{ij}} \tag{0.4}$$

Les résultats généraux concernant les estimateurs du MV et les propriétés asymptotiques des estimateurs étudiées dans le chapitre 1 concernant les modèles binaires restent valables ici. Il n'y a donc pas de difficulté technique concernant l'estimation des paramètres de ces modèles. Mais les modèles polytomiques peuvent avoir des formes mathématiques très différentes suivant les hypothèses retenues. *C'est pourquoi la recherche d'une forme adaptée au problème économique posé constitue dans la plupart des cas, la plus grande difficulté bien avant les méthodes d'estimation et d'inférence.* Si ces modèles posent des problèmes, ce sont avant tout des **problèmes de choix de modélisation**. Pour mieux comprendre ces difficultés, nous allons à présent distinguer les modèles multinomiaux ordonnés, les modèles multinomiaux séquentiels et les modèles multinomiaux non ordonnés.

1. Modèles Multinomiaux Ordonnés

Avant de définir précisément ce que sont les modèles multinomiaux ordonnés, commençons par définir leur champ d'application. **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. On suppose que les modalités sont identiques pour tous les individus :

$$m_i = m \quad \forall i = 1.., N$$

Definition 1.1. ***Un modèle polytomique univarié ordonné peut s'écrire sous la forme :***

$$y_i = \begin{cases} 0 & si\ y_i^* < c_1 \\ 1 & si\ c_1 \leq y_i^* < c_2 \\ ... & ... \\ m & si\ y_i^* > c_m \end{cases} \quad \forall i = 1.., N \tag{1.1}$$

avec $c_{j+1} \geq c_j$ *et* ***où la variable latente*** y_i^* ***est défini par***

$$y_i^* = x_i\beta + \varepsilon_i \tag{1.2}$$

avec $x_i = \left(x_i^1..x_i^K\right), \forall i = 1,.., N$***,*** $\beta = (\beta_1...\beta_K)' \in \mathbb{R}^K$***,*** ε_i *i.i.d.* $\left(0, \sigma_\varepsilon^2\right)$ ***et où*** $\varepsilon_i/\sigma_\varepsilon$ ***suit une loi de fonction de répartition*** $F(.)$.

Naturellement, si la fonction $F(.)$ correspond à la loi logistique, $F(.) = \Lambda(.)$, le modèle est un modèle **logit multinomial ordonné**, tandis que si la fonction $F(.)$ correspond à la loi normale centrée réduite, $F(.) = \Phi(.)$, le modèle est un modèle **probit multinomial ordonné.** Du point de vue pratique, naturellement un tel découpage en classe sur y_i^* n'a de sens que si le nombre de classes est relativement faible.

Naturellement, à partir de la définition précédente on peut déduire la loi de la variable qualitative observée y_i qui nous servira par la suite à construire la fonction de vraisemblance. En effet, on a :

$$Prob\left(y_i = 0\right) = Prob\left(y_i^* < c_1\right) = F\left(\frac{c_1}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right)$$

$$Prob\left(y_i = 1\right) = Prob\left(c_1 \leq y_i^* < c_2\right) = F\left(\frac{c_2}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right) - F\left(\frac{c_1}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right)$$

$$...$$

$$Prob\left(y_i = m\right) = Prob\left(y_i^* > c_m\right) = 1 - F\left(\frac{c_m}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right)$$

De façon générale, on obtient le résultat suivant :

Proposition 1.2. ***Dans un modèle polytomique univarié ordonné satisfaisant la définition 1.1, la probabilité associée à l'événement*** $y_i = j, \forall j = 0, 1, .., m$ ***est définie par :***

$$Prob\left(y_i = j\right) = F\left(\frac{c_{j+1}}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right) - F\left(\frac{c_j}{\sigma_\varepsilon} - \frac{x_i\beta}{\sigma_\varepsilon}\right) \quad \forall i = 1.., N$$

avec par convention $c_0 = -\infty$ ***et*** $c_{m+1} = \infty$.

Il ne reste plus alors qu'à construire la vraisemblance associée à l'échantillon y comme suit :

$$\begin{aligned} L(y, \beta, c_1, .., c_m, \sigma_\varepsilon) &= \prod_{i=1}^{N} \prod_{j=0}^{m_i} Prob\,(y_i = j)^{y_{ij}} \\ &= \prod_{i=1}^{N} \prod_{j=0}^{m_i} \left[F\left(\frac{c_{j+1}}{\sigma_\varepsilon} - \frac{x_i \beta}{\sigma_\varepsilon} \right) - F\left(\frac{c_j}{\sigma_\varepsilon} - \frac{x_i \beta}{\sigma_\varepsilon} \right) \right]^{y_{ij}} \end{aligned} \tag{1.3}$$

où la variable dichotomique y_{ij} est définie par :

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

Généralement seuls les paramètres $\widetilde{\beta} = \beta/\sigma_\varepsilon$ et $\widetilde{c_j} = c_j/\sigma_\varepsilon$ sont identifiables. La vraisemblance s'écrit alors en fonction de ces paramètres comme suit :

$$L\left(y, \widetilde{\beta}, \widetilde{c}_1, .., \widetilde{c}_m\right) = \prod_{i=1}^{N} \prod_{j=0}^{m_i} \left[F\left(\widetilde{c}_{j+1} - x_i \widetilde{\beta} \right) - F\left(\widetilde{c}_j - x_i \widetilde{\beta} \right) \right]^{y_{ij}} \tag{1.4}$$

On vérifie en outre que le **vecteur x_i ne peut contenir de constante** pour les mêmes raison d'identification qui avaient été évoquées dans le cas du modèle dichotomique en ce qui concerne la normalisation du seuil γ. On ne peut identifier à la fois le paramètre associé à la constante et les seuils c_j.

Il n'y a alors aucune difficulté technique à maximiser la fonction de log-vraisemblance en $\widetilde{\beta}, \widetilde{c}_1$, ..et $\widetilde{c}_m$ pour obtenir les estimateurs du maximum de vraisemblance dont les propriétés sont identiques à celles étudiées dans le modèle dichotomique univarié, si ce n'est que l'on estime en outre les paramètres de seuil c_j :

$$\widehat{\beta} = \underset{\{\widetilde{\beta}\}}{\arg\max} \left[\log L\left(y, \widetilde{\beta}, \widetilde{c}_1, .., \widetilde{c}_m\right) \right] \tag{1.5}$$

$$\widehat{c}_j = \underset{\{\widetilde{c}_j\}}{\arg\max} \left[\log L\left(y, \widetilde{\beta}, \widetilde{c}_1, .., \widetilde{c}_m\right) \right] \tag{1.6}$$

avec

$$\log L\left(y, \widetilde{\beta}, \widetilde{c}_1, .., \widetilde{c}_m\right) = \sum_{i=1}^{N} \sum_{j=0}^{m} y_{ij} \log \left[F\left(\widetilde{c}_{j+1} - x_i \widetilde{\beta} \right) - F\left(\widetilde{c}_j - x_i \widetilde{\beta} \right) \right] \tag{1.7}$$

où $F(.)$ est une fonction de répartition donnée.

Nous allons à présent étudier plusieurs exemples de modèles qualitatifs multinomiaux ordonnées afin de mieux appréhender le type de problèmes économiques auxquels cette modélisation s'adapte. Ces exemples sont repris de Amemiya (1981).

1.1. Exemples de Modèles Multinomiaux Ordonnés

Considérons plusieurs exemples de modèles multinomiaux ordonnés.

1.1.1. Dosage d'insecticide : Gurland, Lee et Dahm (1960)

Le premier exemple relève de la bio-économétrie, domaine privilégié des premières applications des modèles qualitatifs.. Il s'agit de l'étude de *Gurland, Lee et Dahm (1960)* parue dans *Biometrics* qui constitue une extension en multinomial de l'exemple de l'insecticide étudié dans le modèle dichotomique. On considère un dosage d'insecticide x_i vaporisé sur le $i^{ème}$ individu et l'on note y_i^* la tolérance de cet individu au produit. Naturellement, on suppose que la tolérance est inobservable et continue. On suppose pour simplifier que la tolérance y_i^* est distribué selon une loi $N\left(\mu,\sigma^2\right)$ de paramètre inconnus. On suppose en outre que la variable observée y_i qui traduit l'état de l'insecte peut prendre à présent trois valeurs, $m=3$ et $\forall i=1..,N$:

$$y_i=\begin{cases}0 & \text{si l'individu } i \text{ est vivant}\\ 1 & \text{si l'individu } i \text{ est moribond}\\ 2 & \text{si l'individu } i \text{ est mort}\end{cases}$$

On suppose que l'individu meurt si et seulement si le dosage de l'insecticide dépasse la tolérance $x_i > y_i^*$, et qu'il reste vivant si au contraire sa tolérance dépasse d'un montant γ le dosage d'insecticide, c'est à dire si et seulement si $y_i^* > x_i+\gamma$, où γ est un paramètre inconnu. Entre les deux, l'insecte est mal en point. Un tel problème correspond bien à la structure d'un modèle polytomique ordonné, **puisque les valeurs prises par la variable multinomiale ($y_i=0,1,2$) correspondent à des intervalles dans lesquels va se trouver une seule variable latente inobservable continue, à savoir la tolérance y_i^*.** En effet, un tel problème peut se modéliser sous la forme :

$$y_i=\begin{cases}0 & \text{si } y_i^*>x_i+\gamma\\ 1 & \text{si } x_i<y_i^*\leq x_i+\gamma\\ 2 & \text{si } y_i^*\leq x_i\end{cases} \tag{1.8}$$

Les probabilités associées au trois modalités sont donc égales à :

$$Prob\left(y_i=0\right)=Prob\left(y_i^*>x_i+\gamma\right)=1-\Phi\left(\frac{x_i+\gamma-\mu}{\sigma}\right)=\Phi\left(\frac{\mu-\gamma-x_i}{\sigma}\right)$$

$$Prob\left(y_i=1\right)=Prob\left(x_i<y_i^*\leq x_i+\gamma\right)=\Phi\left(\frac{x_i-\mu+\gamma}{\sigma}\right)-\Phi\left(\frac{x_i-\mu}{\sigma}\right)$$

$$Prob\left(y_i=2\right)=Prob\left(y_i^*<x_i\right)=\Phi\left(\frac{x_i-\mu}{\sigma}\right)$$

où $\Phi\left(x\right)$ désigne la fonction de répartition de la loi normale centrée réduite. Naturellement, puisque la somme de ces trois probabilité est égale à l'unité, seules deux probabilités seront effectivement estimées. Considérons seulement $Prob\left(y_i=0\right)$ et $Prob\left(y_i=2\right)$, c'est à dire les probabilités de rester vivant et de décès. L'estimation de la probabilité $Prob\left(y_i=2\right)$, sous la forme $Prob\left(y_i=2\right)=\Phi\left(\beta_0+\beta_1 x\right)$, fournit un estimateur $\widehat{\beta}_0$ du coefficient associé à la constante et un estimateur $\widehat{\beta}_1$ du coefficient associé à x_i. On peut alors identifier les paramètres structurels σ et μ en résolvant le système trivial $\beta_0=1/\sigma$ et $\beta_1=-\mu/\sigma$ et proposer deux estimateurs $\widehat{\sigma}=1/\widehat{\beta}_0$ et $\widehat{\mu}=\widehat{\beta}_1\widehat{\beta}_0$. De la même façon, l'estimation de la probabilité $Prob\left(y_i=0\right)$, sous la forme $Prob\left(y_i=0\right)=\Phi\left(\alpha_0+\alpha_1 x\right)$, fournit un estimateur $\widehat{\alpha}_0$ du coefficient associé à la constante et un estimateur $\widehat{\alpha}_1$ du coefficient associé à x_i. On peut alors identifier le paramètre structurel manquant à savoir le seuil γ. En effet, on sait que $\alpha_0=\left(\mu-\gamma\right)/\sigma$, connaissant une valeur de μ et de σ on peut en déduire un estimateur pour γ, tel que $\widehat{\gamma}=\widehat{\mu}-\widehat{\sigma}\widehat{\alpha}_0$. Naturellement, l'estimateur du paramètre $\alpha_1=-1/\sigma$ est obtenu sous la contrainte $\alpha_1=-\beta_0$. Donc dans ce modèle, l'estimation des probabilités $Prob\,(y_i = 0)$ et $Prob\,(y_i = 2)$ permet d'identifier les trois paramètres structurels μ, σ et γ. Notre bio-économètre connaît alors la distribution de la tolérance y_i^* des insectes et le seuil γ d'insecticide.

1.1.2. Acquisition d'un bien immobilier : David et Legg (1975)

L'étude de David et Legg (1975) parue dans le *Journal of Business and Economic Statistics,* est une tentative de modélisation du prix des biens immobiliers en fonction d'un certain nombre de caractéristiques comme la taille du bien immobilier, l'âge de l'acquéreur, le revenu de l'acquéreur, le nombre d'années d'études de l'acquéreur etc.. Les données de David et Legg sont présentées de la façon suivante. On observe si l'acquisition d'un bien immobilier a eu lieu, les biens étant regroupés en trois catégories suivant leur prix. Etant donné la nature des données utilisées, le prix de chaque bien est inobservable et seul son appartenance à l'une des trois catégories est observée :

$$y_i = \begin{cases} 0 & \text{si le prix du bien } i \text{ acquis est inférieur à \$28,999} \\ 1 & \text{si le prix du bien } i \text{ acquis est compris entre \$29,000 et \$54,999} \\ 2 & \text{si le prix du bien } i \text{ acquis est supérieur à \$55,000} \end{cases} \tag{1.9}$$

David et Legg (1975) proposent de modéliser la variable polytomique $y_i = 0, 1, 2$ selon l'appartenance d'une variable inobservable y_i^* à trois classes distinctes :

$$y_i = \begin{cases} 0 & \text{si } y_i^* < c_1 \\ 1 & \text{si } c_1 \leq y_i^* < c_2 \quad \forall i = 1.., N \\ 2 & \text{si } y_i^* > c_2 \end{cases} \tag{1.10}$$

où la variable latente y_i^* est distribuée selon une loi normale $N\left(x_i\beta, \sigma^2\right)$, où le vecteur x_i comporte l'ensemble des caractéristiques du bien citées précédemment.. On suppose que le vecteur x_i ne comporte pas de constante. Le problème consiste donc à estimer les paramètres structurels c_1, c_2, σ et les K paramètres du vecteur β. On a donc $K+3$ paramètres structurels à estimer. Les probabilités associées aux trois modalités sont définies de la façon suivante :

$$Prob\,(y_i = 0) = Prob\,(y_i^* < c_1) = \Phi\left(\frac{c_1}{\sigma} - x_i\frac{\beta}{\sigma}\right)$$

$$Prob\,(y_i = 1) = Prob\,(c_1 \leq y_i^* < c_2) = \Phi\left(\frac{c_2}{\sigma} - x_i\frac{\beta}{\sigma}\right) - \Phi\left(\frac{c_1}{\sigma} - x_i\frac{\beta}{\sigma}\right)$$

$$Prob\,(y_i = 2) = Prob\,(y_i^* > c_2) = 1 - \Phi\left(\frac{c2}{\sigma} - x_i\frac{\beta}{\sigma}\right) = \Phi\left(-\frac{c2}{\sigma} + x_i\frac{\beta}{\sigma}\right)$$

Dès lors, la log-vraisemblance de l'échantillon est définie par la fonction :

$$\begin{aligned} \log L\,(y, \beta, c_1, c_2, \sigma) \quad = \quad & \sum_{i=1}^{N} \left\{ y_{i0} \log\left[\Phi\left(\frac{c_1}{\sigma} - x_i\frac{\beta}{\sigma}\right)\right] + y_{i1} \log\left[\Phi\left(\frac{c_2}{\sigma} - x_i\frac{\beta}{\sigma}\right) - \Phi\left(\frac{c_1}{\sigma} - x_i\frac{\beta}{\sigma}\right)\right] \right. \\ & \left. + y_{i2} \log\left[\Phi\left(-\frac{c2}{\sigma} + x_i\frac{\beta}{\sigma}\right)\right] \right\} \end{aligned}$$

Ce qui peut s'écrire sous une forme log-linéaire dans les paramètres :

$$\begin{aligned} \log L\left(y, \widetilde{\beta}, \widetilde{c}_1, \widetilde{c}_2\right) \quad = \quad & \sum_{i=1}^{N} \left\{ y_{i0} \log\left[\Phi\left(\widetilde{c}_1 - x_i\widetilde{\beta}\right)\right] + y_{i1} \log\left[\Phi\left(\widetilde{c}_2 - x_i\widetilde{\beta}\right) - \Phi\left(\widetilde{c}_1 - x_i\widetilde{\beta}\right)\right] \right. \\ & \left. + y_{i2} \log\left[\Phi\left(-\widetilde{c}_2 + x_i\widetilde{\beta}\right)\right] \right\} \end{aligned} \tag{1.11}$$

avec $\widetilde{x} = x/\sigma$. La procédure du maximum de vraisemblance fournit alors une estimation pour les $K+2$ paramètres $\widetilde{\beta}, \widetilde{c}_1$ et $\widetilde{c}_2$. Dès lors, on ne peut pas dans ce cas identifier la variance σ du fait de la normalisation imposée par le choix de la distribution normale. Par conséquent l'estimation de ce modèle ne permet pas d'identifier les paramètres de seuil c_1 et c_2 et les paramètres β, mais seulement les transformées $\widetilde{\beta}, \widetilde{c}_1$ et $\widetilde{c}_2$. *Ceci n'a pas d'importance dès lors que l'on s'intéresse uniquement aux effets marginaux des variables x_i sur la probabilité d'acheter des biens immobiliers appartenant aux trois catégories.* On peut par exemple calculer l'impact de la taille du logement sur la probabilité d'acheter un logement dont le prix est inférieur à \$28,999 et comparer celle-ci avec l'impact de la taille du logement sur la probabilité d'acheter un logement dont le prix est supérieur à \$55,000. En effet, on peut calculer $3*N*K$ dérivées suivantes :

$$\frac{\partial Prob\left(y_i = j\right)}{\partial x_i^{[k]}} \quad \forall i = 1..,N, \;\; \forall j = 0,1,2 \;\; \forall k = 1..,K$$

Dans cette étude, tout comme dans les modèles multinomiaux en général, il y a finalement plusieurs façons d'interpréter la variable latente y_i^*.

Remarque Dans un modèle multinomial, il n'y a généralement aucune nécessité de donner un nom et d'expliquer réellement ce qu'est la variable latente y_i^*. Peu importe ce qu'elle représente, il suffit juste de supposer que c'est une variable continue qui affecte la variable polytomique observée y_i. Le fait de nommer y_i^* permet simplement de faciliter la justification économique du choix des variables explicatives x_i.

Dans le cas présent de l'étude David et Legg (1975), on peut dire que la variable latente y_i^* correspond au prix du bien immobilier. Dans ce cas, si cette variable est par ailleurs observable, il aurait été préférable au lieu d'estimer les probabilités, d'estimer directement la relation entre le prix du bien et les caractéristiques x_i observées. Mais rien ne nous contraint dans le fait d'assimiler y_i^* au prix des biens : *cette variable inobservable y_i^* peut représenter n'importe quelle grandeur économique susceptible d'affecter l'achat de biens des trois catégories de prix et qui elle même dépend des variables x_i (taille du logement, revenu et âge de l'acquéreur etc..).* Ce peut être par exemple, la disponibilité à payer le bien immobilier.

1.2. Application

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*". Le fichier de données disponible sur le sitye web (??) est intitulé Gunther.xls. 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[2]. Toutes les banques sont ainsi évaluées par des instances de contrôle qui sont les suivantes : 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). En effet, la loi recquiert que les agences de contrôle mentionnées, évaluent régulièrement les performance des institutions au regard des objectifs du CRA.

[2] Pour un historique et une présentation plus global du CRA, consulter le site http://www.ffiec.gov/cra/history.htm.

La performance des institutions en ce qui concerne le fait de favoriser les besoins en crédit de la communauté est évalué dans le contexte des informations disponibles sur cette institution (capacités, contraintes diverses, stratégie..), des informations sur la communauté (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 : $y_i = 1$ pour performance remarquable, $y_i = 2$ pour performance satisfaisante,; $y_i = 3$ pour performance à améliorer, $y_i = 4$ pour performance déplorable.

L'étude de Gunther porte sur 350 observations de ces rating et propose de modéliser ces ratings en fonction de plusieurs variables explicatives intitulées respectivement *loa*, *prl*, *equ*, *roa*, *sec*, *ass*, *metro* et *growth*. La variable *loa* désigne le ratio prêts sur actif total de la banque, la variable *prl* désigne le ratio actifs douteux sur actif total, la variable *equ* désigne le ratio capitaux propres sur actif, la variable *roa* désigne le ratio dividende sur actifs, la variable *sec* désigne le ratio investissements de valeurs sur actifs, la variable *ass* le logarithme de l'actif de la banque, la variable *metro* prend une valeur 1 si la banque à son siège dans une MSA ($\cap$) et 0 sinon, et enfin la variable *growth* désigne le taux de croissance du Pib de l'état dans lequel la banque opère. Dans le tableau ci dessous sont reproduit les valeurs ces différentes variables pour les 10 premiers individus de l'échantillon.

CRA	EQU	GROWTH	LOA	METRO	PRL	ROA	SEC
CRA	EQU	GROWTH	LOA	METRO	PRL	ROA	SEC
1.000000	0.068826	0.055908	0.398216	0.000000	0.012070	0.003437	0.395789
1.000000	0.109697	0.037609	0.622300	0.000000	0.005999	0.008949	0.329522
1.000000	0.062175	0.059176	0.520229	1.000000	0.002058	0.013376	0.316544
1.000000	0.071209	0.061640	0.559965	0.000000	0.009662	0.013883	0.323823
1.000000	0.111563	0.037609	0.653102	0.000000	0.003479	0.011397	0.269915
1.000000	0.107558	0.077307	0.654570	1.000000	0.000995	0.013553	0.166809
1.000000	0.108573	0.031546	0.581331	0.000000	0.009516	0.014816	0.351508
1.000000	0.091564	0.060335	0.544357	0.000000	0.000905	0.014636	0.171662
1.000000	0.075470	0.101828	0.633069	1.000000	0.001361	0.010469	0.247411
1.000000	0.121436	0.018210	0.250721	0.000000	0.039951	-0.015693	0.495006

Figure 1.1:

Nous allons à présent modéliser le rating sous la forme d'un probit ordonné. En effet, dans ce cas précis, les valeurs prises par la variable multinomiale peuvent correspondre à des intervalles dans lesquels va se trouver une variable latente inobservable. On a ici un ordre naturel sur les modalités allant de la satisfaction la plus complète au regard des objectifs du CRA à la performance déplorable. Pour modéliser ce probit ordonné sous Eviews, on choisit *Estimate Equation* dans le menu *Quick*, et l'on retient la méthode *ORDERED* - *Ordered Choice* avec une *Error Distribution* de type Normal. On indique ensuite la variable polytomique ainsi que les variables explicatives, la constante ne pouvant être introduite pour une raison de colinéarité. Les coefficients des variables *roa* et sec sont alors non significativement différents de zéro. On retire donc ces variables et les résultats obtenus pour le probit ordonnés sont alors les suivants :

On observe cette fois que toutes les variables sont significatives de même que les trois seuils

Figure 1.2: Estimation d'un Probit Ordonné

Dependent Variable: CRA
Method: ML - Ordered Probit
Date: 10/15/02 Time: 11:07
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	-0.250326	0.053268	-4.699375	0.0000
EQU	5.249539	1.720978	3.050324	0.0023
GROWTH	-6.107240	2.790589	-2.188513	0.0286
LOA	-1.724511	0.338800	-5.090059	0.0000
METRO	0.768531	0.133197	5.769866	0.0000
PRL	10.74869	2.697211	3.985113	0.0001
Limit Points				
LIMIT_2:C(7)	-3.645509	0.689564	-5.286686	0.0000
LIMIT_3:C(8)	-2.725115	0.678875	-4.014162	0.0001
LIMIT_4:C(9)	-1.614912	0.677212	-2.384648	0.0171

Akaike info criterion	2.397249	Schwarz criterion	2.496453
Log likelihood	-410.5185	Hannan-Quinn criter.	2.436735
Restr. log likelihood	-473.1244	Avg. log likelihood	-1.172910
LR statistic (6 df)	125.2117	LR index (Pseudo-R2)	0.132324
Probability(LR stat)	0.000000		

c_1, c_2 et c_3 tels que :

$$y_i = \begin{cases} 1 & \text{si } y_i^* < c_1 \\ 2 & \text{si } c_1 \leq y_i^* < c_2 \\ 3 & \text{si } c_2 \leq y_i^* < c_3 \\ 4 & \text{si } y_i^* > c_3 \end{cases} \quad \forall i = 1.., N \quad y_i^* = x_i\beta + \varepsilon_i \tag{1.12}$$

On obtient ainsi des réalisations $\widehat{c}_1 = -3.645$, $\widehat{c}_2 = -2.725$ et $\widehat{c}_3 = -1.61$. On peut alors calculer pour chaque banque la probabilité d'obtenir un rating très satisfaisant ($y_i = 1$) de la façon suivante :

$$Prob\,(y_i = 1) = Prob\,(y_i^* < c_1) = \Phi\left(\widetilde{c}_1 - x_i\widetilde{\beta}\right)$$

où $\widetilde{c}_1 = c_1/\sigma_\varepsilon$, $\widetilde{\beta} = \beta/\sigma_\varepsilon$. On peut donc estimer cette probabilité de la façon suivante :

$$\widehat{Prob}\,(y_i = 1) = \Phi\left(\widehat{c}_1 - x_i\widehat{\beta}\right) \tag{1.13}$$

où $\widehat{c}_1$ est un estimateur convergent de $\widetilde{c}_1$ et où $\widehat{\beta}$ est un estimateur convergent de $\widetilde{\beta}$. Ainsi dans le cas de notre modèle pour l'individu 1, on montre que la réalisation de l'estimation de la variable latente est :

$$\begin{aligned} x_1\widehat{\beta} &= 0.068 * 5.24 - 6.138 * 0.0559 - 11.273 * 0.250 \\ &\quad -1.7245 * 0.3982 + 0.7685 * 0 + 10.748 * 0.0120 \\ &= -3.359 \end{aligned}$$

Dès lors, on a :

$$\begin{aligned}
\widehat{Prob}\left(y_1 = 1\right) &= \Phi\left(\widehat{c}_1 - x_1\widehat{\beta}\right) \\
&\simeq \Phi\left(-3.645 + 3.359\right) \\
&\simeq \Phi\left(-0.286\right) \\
&\simeq 0.387
\end{aligned}$$

De la même façon, on peut calculer pour cette banque la probabilité que $y_i = 2$:

$$\begin{aligned}
\widehat{Prob}\left(y_1 = 2\right) &= \Phi\left(\widehat{c}_2 - x_1\widehat{\beta}\right) - \Phi\left(\widehat{c}_1 - x_1\widehat{\beta}\right) \\
&\simeq \Phi\left(-2.725 + 3.359\right) - \Phi\left(-0.286\right) \\
&\simeq \Phi\left(0.634\right) - 0.387 \\
&\simeq 0.736 - 0.387 \\
&\simeq 0.349
\end{aligned}$$

Ainsi pour tous les individus on peut calculer les probabilités associées aux quatre modalités. On obtient les résultats suivants pour les dix premiers individus :

Figure 1.3: Probabilités Estimées du Probit Ordonné

CRA_1F	CRA_2F	CRA_3F	CRA_4F	I_CRAF
0.387312	0.349668	0.222464	0.040556	-3.359178
0.465902	0.332188	0.176024	0.025886	-3.559933
0.322511	0.354625	0.264646	0.058218	-3.184819
0.453040	0.335802	0.183259	0.027899	-3.527524
0.413969	0.345018	0.206117	0.034897	-3.428162
0.267365	0.350391	0.302944	0.079300	-3.024708
0.401455	0.347375	0.213715	0.037455	-3.395926
0.401473	0.347372	0.213704	0.037451	-3.395974
0.460104	0.333851	0.179267	0.026777	-3.545338
0.093806	0.251774	0.416473	0.237947	-2.327836

2. Modèles Multinomiaux Séquentiels

Avant de définir précisément ce que sont les modèles multinomiaux séquentiels, commençons par définir leur champ d'application. **Les modèles séquentiels sont utilisés pour rendre compte de choix effectués ou d'événements selon une séquence bien précise, le plus souvent dans le temps, et dont les réalisations successives conditionnent naturellement l'ensemble des modalités futures**. Ces modèles possèdent la particularité de construire la séquence des événements comme le produit des probabilités élémentaires associées à la réalisation d'un seul évenement à chaque étape.

Definition 2.1. ***Soit*** T ***le nombre d'étapes et*** $y_i = 1, .., N$ ***une variable polytomique dont les modalités sont*** $1, 2, .., T$. ***On écrit alors la probabilité de s'arrêter à l'étape*** t ***comme une fonction*** $F_t(x_i\beta)$, $t = 1, 2, .., T$ ***:***

$$Prob\left(y_i = j\right) = \prod_{s=1}^{j-1} \left[1 - F_s\left(x_i\beta\right)\right] \times F_j\left(x_i\beta\right) \tag{2.1}$$

L'exemple typique est celui de la réussite aux examens, qui est bien entendu conditionnée par la réussite aux examens antérieurs dans le cursus. Considérons l'exemple de la réussite au master. On cherche à modéliser la probabilité qu'un étudiant obtienne son Master en fonction de caractéristiques individuelles, comme le revenu moyen des parents, la moyenne des notes au baccalauréat, la série du baccalauréat etc.. On note $y_i = 1$ si le $i^{ème}$ étudiant a obtenu le baccalauréat mais pas la licence, $y_i = 2$ si l'étudiant a obtenu la licence mais pas le master et $y_i = 3$ si l'étudiant a obtenu le master. Naturellement, la probabilité associée à l'obtention du master s'écrit :

$$\begin{aligned} Prob\left(y_i = 3\right) &= \prod_{s=1}^{2} \left[1 - F_s\left(x_i\beta\right)\right] \times F_3\left(x_i\beta\right) \\ &= \left[1 - F_1\left(x_i\beta\right)\right] \times \left[1 - F_2\left(x_i\beta\right)\right] \times F_3\left(x_i\beta\right) \end{aligned}$$

De la même façon, la probabilité associée à l'obtention de la licence s'écrit :

$$Prob\left(y_i = 2\right) = \prod_{s=1}^{1} \left[1 - F_s\left(x_i\beta\right)\right] \times F_2\left(x_i\beta\right) = \left[1 - F_1\left(x_i\beta\right)\right] \times F_2\left(x_i\beta\right)$$

La probabilité que les individus n'obtiennent pas leur baccalauréat, c'est à dire que $y_i = 1$, est calculée en utilisant tout l'échantillon constitué des deux sous groupes : les étudiants ayant obtenu le baccalauréat et ceux qui ont échoué ($y_i = 0$ non modélisée). On utilise ensuite le sous échantillon des étudiants ayant obtenu le baccalauréat pour déterminer les caractéristiques de la probabilité d'obtenir la licence $y_i = 2$. Et enfin, on utilise le sous échantillon des étudiants ayant obtenu la licence pour déterminer les caractéristiques de la probabilité d'obtenir le master $y_i = 3$.

2.1. Application

****

***** Application Eviews ****

****

cf. Ordered Models

- Chandek, Meghan Stroshine (1999). Race, expectations and evaluations of police performance: An empirical assessment. Policing 22(4):675

- http://faculty.smu.edu/tfomby/eco6352/data/

3. Modèles Multinomiaux Non Ordonnés

Nous allons à présent envisager la classe des modèles multinomiaux les plus fréquents en économie : les **modèles multinomiaux non ordonnés**. Il existe deux grandes classes de modèles multinomiaux non ordonnés suivant que ces modèles satisfont ou ne satisfont pas une hypothèse particulière qui est l'hypothèse **d'Indépendence des Alternatives Non Pertinentes** ($IANP$ ou IIA en anglais pour *Independance of Irrelevant Alternative*). Cette hypothèse traduit le fait que le rapport de deux probabilités associés à deux évenements particuliers est indépendant des autres événements. Ainsi, la première grande classe de modèle est constitué par les modèles logit multinomiaux non ordonnés qui comprend notamment :

1. Les modèles logit multinomiaux indépendants ou **modèles logit multinomiaux**.
2. Les modèles logit multinomiaux conditionnels ou **modèles logit conditionnels**.
3. Les modèles logit multinomiaux universels ou **modèles logit universels**.

Tous ces modèles satisfont l'hypothèse IIA, or nous verrons qu'une telle hypothèse pose des problèmes de cohérence dans certaines modélisations économiques. C'est pourquoi des modèles alternatifs ont été développés de sorte à ne pas satisfaire cette hypothèse contestable : cette seconde classe de modèles comprend notamment le modèle logit hierarchisé, le modèle probit multinomial.

Toutefois, dans la pratique les modèles multinomiaux les plus fréquemment utilisés restent les modèles logit satisfaisant l'hypothèse IIA. C'est pourquoi dans cette section, nous **nous limiterons à l'étude de cette classe de modèles multinomiaux non ordonnés**. Nous étudierons les modèles alternatifs dans la prochaine section. Mais avant de présenter la classe des logit multinomiaux non ordonnés, nous allons introduire ces modèles en décrivant leur utilisation essentielle, **à savoir celle de rendre compte de choix probabilistes.** Les modèles multinomiaux non ordonnés sont en effet avant tout des modèles permettant de décrire des choix individuels en présence d'utilité stochastique.

3.1. Des modèles de choix probabilistes

Supposons qu'un individu ait a effectuer un choix rationnel entre $m+1$ modalités procurant $m+1$ niveaux de satisfaction différents pour l'individu. On postule que les choix rationnels peuvent être représentés par une fonction d'utilité. Rappelons qu'une fonction d'utilité $U(\omega)$ est définie à une transformée croissante près : si la fonction $h(.)$ est une fonction croissante et continue, $h(U(\omega))$ est une autre fonction d'utilité associée au même préordre que celui de $U(\omega)$. On considère le cas où **le niveau d'utilité est stochastique** et décrit par une fonction $U(.)$ dépendant d'un terme aléatoire. Ce choix peut se justifier par ne mauvaise perception de la qualité des différentes modalités ou en raison d'une grande difficulté à évaluer de façon certaine les niveaux d'utilité. On pose que pour chaque modalité $j = 0, 1, .., m$, l'utilité de l'individu s'exprime sous la forme suivante :

$$U_j = U(x_j, \varepsilon_j) = v(x_j) + \varepsilon_j \quad \forall j = 0, 1, .., m \tag{3.1}$$

où $v(.)$ est une fonction continue déterministe et où ε_j est une variable aléatoire *i.i.d.* dont la loi est décrite par la fonction de densité $f(.)$ et la fonction de répartition $F(.)$. **On suppose que les perturbations $\varepsilon_j, \forall j = 0, 1, .., m$ sont indépendantes**. Ainsi l'utilité aléatoire associée à la $j^{ème}$modalité dépend des caractéristiques propres à cette modalité

On définit une variable polytomique y qui prend $m+1$ modalités suivant les choix de l'individu :

$$y = j \text{ si l'individu choisit la } j^{ème} \text{ modalité } \forall j = 0, 1, .., m$$

Dès lors, *la probabilité que notre individu choisisse la modalité j correspond à la probabilité que cette modalité lui confère un niveau d'utilité supérieure à toutes les autres modalités qui s'offrent à lui.* En effet, la probabilité que l'individu choisisse la modalité j est définie par :

$$\begin{aligned} Prob(y = j) &= Prob\{U(x_j, \varepsilon_j) > U(x_0, \varepsilon_0), U(x_j, \varepsilon_j) > U(x_1, \varepsilon_1), \\ &\quad ....., U(x_j, \varepsilon_j) > U(x_k, \varepsilon_k), ..., U(x_j, \varepsilon_j) > U(x_m, \varepsilon_m)\} \end{aligned} \tag{3.2}$$

Prenons par exemple la probabilité que l'agent choisisse la modalité 0 :

$$\begin{aligned} Prob(y = 0) &= Prob\{U(x_0, \varepsilon_0) > U(x_1, \varepsilon_1), U(x_0, \varepsilon_0) > U(x_2, \varepsilon_2), \\ &\quad ....., U(x_0, \varepsilon_0) > U(x_k, \varepsilon_k), ..., U(x_0, \varepsilon_0) > U(x_m, \varepsilon_m)\} \end{aligned}$$

ce qui peut se réécrire sous la forme suivante :

$$\begin{aligned} Prob(y = 0) &= Prob[U(x_0, \varepsilon_0) > U(x_k, \varepsilon_k), \forall k = 1, .., m] \\ &= Prob[v(x_0) + \varepsilon_0 > v(x_k) + \varepsilon_k, \forall k = 1, .., m] \\ &= Prob[\varepsilon_k < v(x_0) - v(x_k) + \varepsilon_0, \forall k = 1, .., m] \end{aligned}$$

Pour calculer cette probabilité à partir des fonctions de densité $f(.)$, rappelons la définition de la densité jointe.

Definition 3.1. ***La densité jointe de deux v.a.r. continues X et Y, notée $f_{X,Y}(x, y) \geq 0$, satisfait les propriétés suivantes :***

$$\forall (a, b, c, d) \in \mathbb{R}^4 \quad P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y)\, dy\, dx \tag{3.3}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dx = 1 \tag{3.4}$$

La fonction de distribution cumulative jointe, notée $F_{X,Y}(x, y)$ est alors définie par, $\forall (a, b) \in \mathbb{R}^2$:

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx \tag{3.5}$$

On sait en outre que si les variables X et Y sont dites indépendantes si et seulement si $f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$. L'indépendance implique par conséquent que $F_{X,Y}(x, y) = F_X(x)\, F_Y(y)$. Ainsi, si l'on suppose que les perturbations ε_j sont indépendantes et distribuées selon une loi de distribution $f(.)$, et si l'on note $v_j = v(x_j)$, la probabilité $Prob(y=0)$ peut s'écrire sous la forme :

$$\begin{aligned} Prob(y=0) &= \int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{m} Prob\left[\varepsilon_k < v_0 - v_k + \varepsilon_0\right] \right\} f(\varepsilon_0)\, d\varepsilon_0 \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{v_0-v_1+\varepsilon_0} f(\varepsilon_1)\, d\varepsilon_1 \times \int_{-\infty}^{v_0-v_2+\varepsilon_0} f(\varepsilon_2)\, d\varepsilon_2 \times .. \right. \\ &\quad \left. .. \times \int_{-\infty}^{v_0-v_m+\varepsilon_0} f(\varepsilon_m)\, d\varepsilon_m \right\} f(\varepsilon_0)\, d\varepsilon_0 \end{aligned}$$

En d'autres termes, si l'on note $F(.)$ la fonction de répartition associée à la loi des perturbations ε_j, on a:

$$Prob(y=0) = \int_{-\infty}^{\infty} \left(\prod_{k=1}^{m} F\left[v(x_0) - v(x_k) + \varepsilon_0\right] \right) f(\varepsilon_0)\, d\varepsilon_0$$

De façon générale, quelque soit la modalité $j = 0, 1, .., m$ on montre ainsi que :

$$Prob(y=j) = \int_{-\infty}^{\infty} \left(\prod_{\substack{k=0 \\ k\neq j}}^{m} F\left[v(x_j) - v(x_k) + \varepsilon_j\right] \right) f(\varepsilon_j)\, d\varepsilon_j \tag{3.6}$$

Supposons à présent que la loi des perturbations soit une **loi de Gompertz**[3] :

$$F(z) = \exp\left[-\exp(-z)\right] \tag{3.7}$$

$$f(z) = \exp\left[-z - \exp(-z)\right] \tag{3.8}$$

Alors, il est possible de donner une expression analytique à la probabilité que l'agent choisisse la modalité 0.

$$\begin{aligned} Prob(y=0) &= \int_{-\infty}^{\infty} \left(\prod_{k=1}^{m} \exp\left[-\exp(-v_0 + v_k - \varepsilon_0)\right] \right) \exp\left[-\varepsilon_0 - \exp(-\varepsilon_0)\right] d\varepsilon_0 \\ &= \int_{-\infty}^{\infty} \exp\left(-\sum_{k=1}^{m} \exp(-v_0 + v_k - \varepsilon_0) \right) \exp\left[-\varepsilon_0 - \exp(-\varepsilon_0)\right] d\varepsilon_0 \end{aligned}$$

****
*** FINIR DEMONSTRATION ***
****

Finalement, on montre que :

$$Prob(y=0) = \frac{\exp\left[v(x_0)\right]}{\sum\limits_{k=0}^{m} \exp\left[v(x_k)\right]} = \frac{1}{1 + \sum\limits_{k=1}^{m} \exp\left[v(x_k) - v(x_0)\right]}$$

avec par convention $v(x_0) = 0$. De façon générale, la probabilité que l'individu choisisse la modalité j est définie par :

$$Prob(y=j) = \frac{\exp\left[v(x_j)\right]}{\sum\limits_{k=0}^{m} \exp\left[v(x_k)\right]} \tag{3.9}$$

[3] Dite aussi loi des valeurs extrêmes.

Or, si l'on se restreint à une classe des fonctions $v(.)$ affines, avec $v(x_j) = \beta_j x_j$, cette formulation de la probabilité associée à la $j^{ème}$ modalité est précisément la modélisation de la probabilité que l'on retiendra dans les modèles logit multinomiaux non ordonnés.

Considérons à présent le cas où l'on dispose d'un ensemble de N individus indicés $i = 1, .., N$ ayant les mêmes préférences que l'individu de référence de l'exemple précédent. De la même façon sous les hypothèse d'indépendance des perturbations ε_j et sous une hypothèse particulière sur la distribution des ces perturbations, on montre que la probabilité que l'individu i choisisse la modalité j, $\forall j = 0, .., m$, est définie par :

$$Prob\,(y_i = j) = \frac{\exp[v(x_{i,j})]}{\sum_{k=0}^{m} \exp[v(x_{i,k})]} \tag{3.10}$$

où $x_{i,j}$ désigne la valeur du vecteur de variable explicative pour l'individu i conditionnant le choix de la $j^{ème}$ modalité.

Suivant la forme de la fonction $v(x_{i,j})$ plusieurs modèles peuvent être envisagés.

1. **Le modèle logit multinomial indépendant (ou logit multinomial) est obtenu lorsque la fonction $v(.)$ est une fonction linéaire dont les paramètres β_j diffèrent selon les modalités et pour laquelle les variables explicatives varient uniquement en fonction des individus :**

$$v(x_{i,j}) = x_i \beta_j \tag{3.11}$$

2. **Le modèle logit multinomial conditionnel (ou logit conditionnel) est obtenu lorsque la fonction $v(.)$ est linéaire, les paramètres β sont indépendants des modalités et que les variables explicatives diffèrent selon les modalités et les individus :**

$$v(x_{i,j}) = x_{i,j} \beta \tag{3.12}$$

3. **Le modèle logit multinomial universel (ou logit universel) est obtenu pour toute fonction $v(.)$ continue dépendant de paramètres β_j conditionnels aux modalités et de l'ensemble des variables explicatives du modèle :**

$$v(x_{i,j}) = v\left(\beta_j, x_{ij}\right) \tag{3.13}$$

3.2. Logit multinomial indépendant

Comme nous l'avons dit précédemment, le modèle **logit multinomial indépendant**[4] (ou logit multinomial) est obtenu lorsque la fonction $v(.)$ est linéaire, les paramètres β_j diffèrent selon les modalités et que les variables explicatives varient uniquement en fonction des individus, c'est à dire lorsque $v(x_{i,j}) = x_i \beta_j$. Dès lors, on peut définir la forme générale de la probabilité que l'individu i choisisse la modalité j de la façon suivante :

[4] Il convient de faire attention ici sur la terminologie employée. Amemiya (1981) qualifie le modèle logit conditionnel de modèle indépendant puisque construit sur l'indépendance des perturbations. Donc pour Amemiya, le modèle indépendant ici étudié n'est qu'un cas particulier sans nom du logit universel..

Definition 3.2. ***Dans un modèle logit multinomial, la probabilité que l'individu*** i ***choisisse la modalité*** j***,*** $\forall j = 0, .., m$***, est définie par :***

$$Prob\left(y_i = j\right) = \frac{\exp\left(x_i \beta_j\right)}{\sum\limits_{k=0}^{m} \exp\left(x_i \beta_k\right)} = \frac{\exp\left(x_i \beta_j\right)}{1 + \sum\limits_{k=1}^{m} \exp\left(x_i \beta_k\right)} \tag{3.14}$$

où le vecteur β_0 ***est normalisé à zéro :*** $\beta_0 = 0$.

Sous l'hypothèse de normalisation $\beta_0 = 0$, la probabilité associée à la modalité de référence 0 est définie par :

$$Prob\left(y_i = 0\right) = \frac{1}{\sum\limits_{k=0}^{m} \exp\left(x_i \beta_k\right)} = \frac{1}{1 + \sum\limits_{k=1}^{m} \exp\left(x_i \beta_k\right)} \tag{3.15}$$

Exemple : Considérons à présent un exemple de modèle logit multinomial. On cherche à modéliser la probabilité de défaillance d'abonnés en fonction de leur revenu, noté r_i, et d'une constante. Supposons que la variable observable y_i puisse prendre trois modalités : défaillance totale (modalité 0), défaillance partielle (modalité 1) et remboursement intégral de la créance (modalité 2). L'hypothèse selon laquelle les paramètres β_j diffèrent selon les modalités revient en fait à poser que :

$$Prob\left(y_i = 1\right) = \frac{\exp\left(\beta_1^{[1]} + \beta_1^{[2]} r_i\right)}{1 + \exp\left(\beta_1^{[1]} + \beta_1^{[2]} r_i\right) + \exp\left(\beta_2^{[1]} + \beta_2^{[2]} r_i\right)}$$

$$Prob\left(y_i = 2\right) = \frac{\exp\left(\beta_2^{[1]} + \beta_2^{[2]} r_i\right)}{1 + \exp\left(\beta_1^{[1]} + \beta_1^{[2]} r_i\right) + \exp\left(\beta_2^{[1]} + \beta_2^{[2]} r_i\right)}$$

où $\beta_1^{[2]}$ désigne le coefficient associé au revenu ($2^{ème}$ variable explicative) dans la modélisation de la probabilité de la défaillance partielle ($1^{ère}$ modalité), tandis que $\beta_2^{[2]}$ désigne le coefficient associé au revenu ($2^{ème}$ variable explicative) dans la modélisation de la probabilité de l'absence de défaillance ($2^{ème}$ modalité). On suppose ici que le revenu n'affecte pas de façon identique la probabilité de défaillance partielle et la probabilité d'absence de défaillance. Par contre le revenu du couple ne diffère suivant que ce dernier rembourse totalement ou de façon partielle son prêt. Essayons à présent de recenser les implications de la spécification d'un modèle logit multinomial indépendant.

3.2.1. Spécification du Logit Multinomial

On considère un modèle multinomial pour lequel la probabilité que l'individu i choisisse la modalité $j = 0, .., m$ s'écrit sous la forme :

$$Prob\left(y_i = j\right) = \frac{\exp\left(x_i \beta_j\right)}{\sum\limits_{k=0}^{m} \exp\left(x_i \beta_k\right)}$$

où les vecteurs de paramètres $\beta_j \in \mathbb{R}^K$ peuvent différer selon les modalités j. La première remarque sur cette spécification concerne la normalisation $\beta_0 = 0$. En effet, supposons que cette normalisation ne soit pas imposée a priori. On obtient alors une expression équivalente à la probabilité (3.14) en divisant les deux membres de la probabilité par $\exp(\beta_0 x_i)$:

$$Prob\,(y_i = j) = \frac{\exp\left(x_i\beta_j\right)/\exp\left(x_i\beta_0\right)}{\left[\sum\limits_{k=0}^{m}\exp\left(x_i\beta_k\right)\right]/\exp\left(x_i\beta_0\right)} = \frac{\exp\left[x_i\left(\beta_j-\beta_0\right)\right]}{\sum\limits_{k=0}^{m}\exp\left[x_i\left(\beta_k-\beta_0\right)\right]}$$

En posant $\beta_j^* = \beta_j - \beta_0$, $\forall j$, et donc $\beta_0^* = 0$, on obtient alors une expression de la probabilité $Prob\,(y_i = j)$ similaire à celle de la définition, sans imposer la normalisation $\beta_0 = 0$.

$$Prob\,(y_i = j) = \frac{\exp\left(x_i\beta_j^*\right)}{\sum\limits_{k=0}^{m}\exp\left(x_i\beta_k^*\right)} = \frac{\exp\left(x_i\beta_j^*\right)}{1+\sum\limits_{k=1}^{m}\exp\left(x_i\beta_k^*\right)} \tag{3.16}$$

Ainsi écrire la probabilité sous la forme $\exp\left(\beta_j^* x_i\right)/\left[1+\sum_{k=1}^{m}\exp\left(x_i\beta_k^*\right)\right]$ revient en fait à normaliser les paramètres du modèle qui correspondent en fait aux différences entre les paramètres originaux β et le vecteur de paramètres de la modalité de référence, ici en l'occurrence β_0. **Ainsi, les paramètres s'interprètent comme des écarts au référentiel (c'est à dire aux paramètres de la modalité 0)**. On peut exprimer cette propriété de la façon suivante.

Proposition 3.3. ***Dans un modèle logit multinomial à** $m+1$ **modalités, la probabilité associée à la** $j^{ème}$ **modalité dépend des écarts** $\beta_k - \beta_j$ avec $k \neq j$ et $k = 0, 1, ..m$ **:***

$$Prob\,(y_i = j) = \frac{\exp\left(x_i\beta_j\right)}{\sum\limits_{k=0}^{m}\exp\left(x_i\beta_k\right)} = \frac{1}{1+\sum\limits_{\substack{k=0\\k\neq j}}^{m}\exp\left[x_i\left(\beta_k-\beta_j\right)\right]} \tag{3.17}$$

***Les** m **probabilité indépendantes dépendent alors de** $(m+1)\,m/2$ **différences de paramètres** $\left(\beta_k - \beta_j\right)$**. Les paramètres ne sont pas identifiables à moins d'imposer une contrainte de normalisation : par exemple** $\beta_0 = 0$.*

Exemple : considérons un modèle avec 3 modalités ($m = 2$), $j = 0, 1, 2$. On a alors :

$$Prob\,(y_i = 0) = \frac{1}{1+\exp\left[x_i\left(\beta_1-\beta_0\right)\right]+\exp\left[x_i\left(\beta_2-\beta_0\right)\right]}$$

$$Prob\,(y_i = 1) = \frac{1}{1+\exp\left[x_i\left(\beta_0-\beta_1\right)\right]+\exp\left[x_i\left(\beta_2-\beta_1\right)\right]}$$

$$Prob\,(y_i = 2) = \frac{1}{1+\exp\left[x_i\left(\beta_0-\beta_2\right)\right]+\exp\left[x_i\left(\beta_1-\beta_2\right)\right]}$$

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 β_1 et β_2. Ces paramètres s'interprètent comme des écarts au vecteur β_0.

Corollary 3.4. ***Dans un modèle logit mutinomial à** $m+1$ **modalités :***

- ***Les paramètres associés à la modalité de référence, généralement** 0**, sont normalisés à zéro : seuls les paramètres associées à** m **modalités peuvent être estimés.***

- ***Les paramètres du modèle s'interprètent comme des écarts au référentiel (c'est à dire aux paramètres β_0 de la modalité 0)***
- ***La log vraisemblance s'écrit uniquement en fonction des vecteurs $\beta_1, ..., \beta_m$***

La seconde propriété centrale dans l'analyse des modèles logit multinomiaux concerne les ratio de probabilités p_j/p_k où j et k sont deux modalités distinctes. En effet, on montre la propriété suivante :

Proposition 3.5. ***Dans le cas d'un modèle logit multinomial, le rapport des probabilités associés à deux modalités j et k distinctes, $\forall j = 0, .., m$ et $\forall k = 0, .., m$, s'écrit sous la forme :***

$$\frac{p_j}{p_k} = \frac{Prob\left(y_i = j\right)}{Prob\left(y_i = k\right)} = \frac{\exp\left(x_i\beta_j\right)}{\exp\left(x_i\beta_k\right)} = \exp\left[x_i\left(\beta_j - \beta_k\right)\right] \tag{3.18}$$

Ce rapport de probabilités est indépendant des alternatives autres que j et k.

Ainsi, on illustre une hypothèse très particulière de ces modèles logit multinomiaux indépendants : à savoir l'hypothèse **d'Indépendence des Alternatives Non Pertinentes** ($IANP$ ou IIA en anglais pour *Independance of Irrelevant Alternative*). Cette hypothèse traduit le fait que le rapport de deux probabilités associés à deux évenements particuliers est indépendant des autres événements. La question que se pose est alors de savoir si une telle hypothèse est satisfaite en pratique.

Pour montrer à quel point cette hypothèse peut s'avérer inadéquate, reprenons l'exemple du choix de transport proposé par Mac Fadden et connu sous le nom de "*bus bleu, bus rouge*". On considère qu'un individu pour se rendre à son travail à le choix entre deux modes de transports.

$$y_i = \begin{cases} 0 & \text{si l'individu prend le métro} \\ 1 & \text{si l'individu prend un bus bleu} \end{cases}$$

On note $p_b = p_0$ la probabilité que l'individu prenne le bus bleu et $p_m = p_1$ la probabilité que l'individu prenne le métro dans cette configuration de choix. Supposons maintenant que l'on offre à l'individu la possibilité soit de prendre un bus bleu, soit de prendre un bus rouge.

$$y_i = \begin{cases} 0 & \text{si l'individu prend le métro} \\ 1 & \text{si l'individu prend un bus bleu} \\ 2 & \text{si l'individu prend un bus rouge} \end{cases}$$

La probabilité que l'agent prenne le bus s'écrit donc sous la forme :

$$p_b = Prob\left(y_i = 1\right) + Prob\left(y_i = 2\right) \tag{3.19}$$

La probabilité que l'agent prenne le métro demeure $p_m = p_0$ et reste à un niveau inchangé par rapport au cas précédent étant donné les nouvelles alternatives proposées. Si l'on admet que la couleur du bus a vraisemblablement peu de chance d'affecter le choix du mode de transport, on doit avoir que les probabilités de sélection p_1 et p_2 doivent être égales :

$$Prob\left(y_i = 1\right) = Prob\left(y_i = 2\right) \tag{3.20}$$

Maintenant, si l'hypothèse IIA est satisfaite le rapport entre la probabilité de prendre le métro et la probabilité de prendre le bus devrait être la même dans les deux modèles : en effet, ce ratio doit être indépendant des alternatives. Or, ici ce ratio vaut p_0/p_1 dans la première modélisation et vaut :

$$\frac{p_m}{p_b} = \frac{p_0}{p_1 + p_2} = \frac{1}{2}\frac{p_0}{p_1}$$

Ce ratio diffère de celui que l'on avait obtenu en l'absence de l'alternative "*bus rouge*" : l'hypothèse IIA n'est donc pas satisfaite.

En conclusion, l'hypothèse IIA n'est que rarement satisfaite, ce qui pose le problème de la cohérence d'une modélisation de type logit multinomial pour rendre compte de choix probabilistes. Nous reviendrons dans la prochaine section sur cette hypothèse IIA et les modèles alternatifs qui en rendement compte. Toutefois, le modèle logit multinomial indépendant est très souvent utilisé compte tenu de la simplicité de sa mise en oeuvre pratique. C'est ce que nous allons voir à présent en ce qui concerne l'estimation des paramètres de ce type de modèles..

3.2.2. Estimation des paramètres du logit multinomial

Tout comme dans le cas du modèle logit dichotomique, l'estimation des paramètres des modèles logit multinomiaux peut se faire de différentes façons :

1. Méthodes du maximum de vraisemblance
2. Méthodes de moments : GMM, moments simulés etc..
3. Méthodes non paramétriques et semi-paramétriques

Nous n'étudierons ici que la méthode du maximum de vraisemblance à information complète.

Comme nous l'avons vu précédemment la vraisemblance associée à un modèle logit multinomial indépendant à $m+1$ modalités s'écrit en fonction de m vecteur de paramètres β_j, $j = 1, .., m$ du fait de la normalisation $\beta_0 = 0$. Ainsi l'estimation des paramètres du modèle logit multinomial s'effectue alors en maximisant la log-vraisemblance par rapport aux vecteurs de paramètres $(\beta_1, \beta_2, ..., \beta_m)$:

$$\log L(y, \beta_1, \beta_2, ..., \beta_m) = \sum_{i=1}^{N} \sum_{j=0}^{m} y_{i,j} \log [Prob(y_i = j)]$$

avec $y_{i,j} = 1$ si $y_i = j$ et 0 sinon, et où les probabilités $Prob(y_i = j)$ sont définies par :

$$Prob(y_i = j) = \frac{\exp(x_i\beta_j)}{\sum_{k=0}^{m} \exp(x_i\beta_k)} = \frac{\exp(x_i\beta_j)}{1 + \sum_{k=1}^{m} \exp(x_i\beta_k)}$$

$$Prob(y_i = 0) = \frac{1}{1 + \sum_{k=1}^{m} \exp(x_i\beta_k)}$$

avec β_0 par convention.

Proposition 3.6. ***La log vraisemblance associée à un modèle logit multinomial à*** *$m+1$ modalités $j=0,1,..,m$* ***s'écrit :***

$$\log L(y,\beta_1,\beta_2,...,\beta_m)=\sum_{i=1}^{N}\sum_{j=1}^{m}y_{i,j}x_i\beta_j-\sum_{i=1}^{N}\log\left[1+\sum_{k=1}^{m}\exp(x_i\beta_k)\right] \tag{3.21}$$

avec $\beta_0=0$ ***par convention.***

En effet, on sait que la log-vraisemblance est définie par la relation suivante :

$$\begin{aligned}\log L(y,\beta_1,\beta_2,...,\beta_m) &= \sum_{i=1}^{N}\sum_{j=0}^{m}y_{i,j}\log(p_{i,j})\\ &= \sum_{i=1}^{N}\sum_{j=0}^{m}y_{i,j}\log\left[\frac{\exp(x_i\beta_j)}{1+\sum_{k=1}^{m}\exp(x_i\beta_k)}\right]\end{aligned}$$

Si l'on pose $\widetilde{H}_i=\log\left[1+\sum_{k=1}^{m}\exp(x_i\beta_k)\right]$, cette expression devient :

$$\log L(y,\beta_1,\beta_2,...,\beta_m)=\sum_{i=1}^{N}\sum_{j=0}^{m}y_{i,j}x_i\beta_j-\sum_{i=1}^{N}\widetilde{H}_i\left(\sum_{j=0}^{m}y_{i,j}\right)$$

Sachant que par convention$\beta_0=0$ on a donc :

$$\sum_{i=1}^{N}\sum_{j=0}^{m}y_{i,j}x_i\beta_j=\sum_{i=1}^{N}\sum_{j=1}^{m}y_{i,j}x_i\beta_j$$

Etant donnée la définition de la variable $y_{i,j}$ qui prend la valeur 1 si $y_i=j$ et 0 sinon, on a immédiatement que :

$$\sum_{j=0}^{m}y_{i,j}=y_{i,0}+y_{i,1}+..+y_{i,m-1}+y_{i,m}=1$$

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

$$\sum_{i=1}^{N}\widetilde{H}_i\left(\sum_{j=0}^{m}y_{i,j}\right)=\sum_{i=1}^{N}\widetilde{H}_i=\sum_{i=1}^{N}\log\left[1+\sum_{k=1}^{m}\exp(x_i\beta_k)\right]$$

Puisque en effet $\sum_{i=1}^{N}y_{i,0}=1$. On en déduit alors finalement que :

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

On retrouve alors l'expression (3.21) de la log-vraisemblance. Notons au passage que **la fonction de log-vraisemblance d'un modèle logit multinomial indépendant est globalement concave**[5] et que par conséquent on peut utiliser différents algorithmes d'optimisation numérique propres à ce type de problème (Newton Raphson par exemple) et que les résultats ne sont pas sensibles au choix des conditions initiales de ces algorithmes.

[5] La démonstration de ce résultat est laissée au lecteur à titre d'exercice.

Definition 3.7. ***Le gradient associé à la log-vraisemblance d'un modèle logit multinomial est défini*** $\forall z = 1, .., m$ **:**

$$\frac{\partial \log L(y, \beta_1, \beta_2, ..., \beta_m)}{\partial \beta_z} = \sum_{i=1}^{N} (y_{i,z} - p_{i,z}) x_i' \tag{3.22}$$

avec $p_{i,z} = Prob(y_i = z)$.

En effet, on a :

$$\begin{aligned}
\frac{\partial \log L(y, \beta_1, \beta_2, ..., \beta_m)}{\partial \beta_z} &= \frac{\partial}{\partial \beta_z}\left(\sum_{i=1}^{N}\sum_{j=1}^{m} y_{i,j} x_i \beta_j\right) - \frac{\partial}{\partial \beta_z}\left\{\sum_{i=1}^{N} \log\left[1 + \sum_{k=1}^{m} \exp(x_i\beta_k)\right]\right\} \\
&= \sum_{i=1}^{N} y_{i,z} x_i' - \sum_{i=1}^{N}\left[\frac{\exp(x_i\beta_z)}{1 + \sum_{k=1}^{m}\exp(x_i\beta_k)}\right] x_i'
\end{aligned}$$

Connaissant la définition de $p_{i,z} = Prob(y_i = z)$, on a donc :

$$\frac{\partial \log L(y, \beta_1, \beta_2, ..., \beta_m)}{\partial \beta_z} = \sum_{i=1}^{N} y_{i,z} x_i' - \sum_{i=1}^{N} p_{i,z} x_i' = \sum_{i=1}^{N} (y_{i,z} - p_{i,z}) x_i'$$

On retrouve naturellement la même expression que dans le cas du modèle logit bivarié. De la même façon, la matrice hessienne est définie par :

$$\frac{\partial^2 \log L(y, \beta_1, \beta_2, ..., \beta_m)}{\partial \beta_j \partial \beta_k'} = -\sum_{i=1}^{N} p_{i,j} (\mathbb{I}_{j,k} - p_{i,k}) x_i' x_i \tag{3.23}$$

où la fonction indicatrice $\mathbb{I}_{j,k}$ est telle que $\mathbb{I}_{j,k} = 1$ si $k = j$ et 0 sinon.

Enfin, on peut naturellement étudier **les effets marginaux dans un modèle logit multinomial indépendant** de la façon suivante.

Definition 3.8. ***Les effets marginaux d'une variation de la variable exogène*** $x_i^{[k]}$, $\forall k = 1, .., K$ ***sur la probabilité que l'individu*** i ***choisisse la*** $j^{\text{ème}}$ ***modalité,*** $\forall j = 0, 1, .., m$, ***sont définis par :***

$$\frac{\partial p_{i,j}}{\partial x_i^{[k]}} = p_{i,j}\left[\beta_j^{[k]} - \sum_{z=0}^{m} p_{i,z}\beta_z^{[k]}\right] \tag{3.24}$$

où $\beta_j^{[k]}$ ***est la*** $k^{\text{ème}}$***composante de*** β_j ***associé à la variable explicative*** $x_i^{[k]}$ ***et où*** $p_{i,j} = Prob(y_i = j)$ ***désigne la probabilité que l'individu*** i ***choisisse la*** $j^{\text{ème}}$ ***modalité :***

$$p_{i,j} = \frac{\exp(x_i\beta_j)}{\sum\limits_{z=0}^{m}\exp(x_i\beta_z)} \tag{3.25}$$

Pour démontrer ce résultat, on pose $H(x_i) = \sum_{z=0}^{m}\exp(x_i\beta_z)$ En effet, on sait que

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

En simplifiant cette expression, on fait apparaître les probabilités $p_{i,j} = \exp\left(x_i\beta_j\right)/H$, et on trouve :

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

On retrouve ainsi l'expression (3.24) des effets marginaux. Il convient de remarquer que, pour chaque individu, pour une variable explicative quelconque $x_i^{[k]}$, on doit calculer $m+1$ effets marginaux associés aux probabilités $p_{i,j}$ pour $j = 0, 1, .., m$.

Dans le cas d'un modèle dichotomique à deux modalités ($m = 1$), on retrouve évidemment la formule proposée dans le premier chapitre. En effet, nous avions vu que pour un modèle logit univarié :

$$\frac{\partial p_i}{\partial x_i^{[k]}} = \frac{e^{x_i\beta}}{\left(1+e^{x_i\beta}\right)^2}\beta_k \tag{3.26}$$

Selon la formule (3.24), dans un modèle à 2 modalités on doit avoir

$$\frac{\partial p_{i,1}}{\partial x_i^{[k]}} = p_{i,1}\left[\beta_1^{[k]} - p_{i,0}\beta_0^{[k]} - p_{i,1}\beta_1^{[k]}\right]$$

avec $p_{i,1} = Prob\left(y_i = 1\right)$ et $p_{i,0} = Prob\left(y_i = 0\right) = 1 - p_{i,1}$ et $p_{i,1} = e^{x_i\beta}/\left(1+e^{x_i\beta}\right)$. Par normalisation, on pose $\beta_0^{[k]} = 0$, $\forall k$, dès lors, il vient $\partial p_{i,1}/\partial x_i^{[k]} = p_{i,1}\beta_1^{[k]} - p_{i,1}^2\beta_1^{[k]}$ ou encore :

$$\frac{\partial p_{i,1}}{\partial x_i^{[k]}} = \beta_1^{[k]}\left[\frac{e^{x_i\beta}}{1+e^{x_i\beta}} - \left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right)^2\right] = \beta_1^{[k]}\left[\frac{e^{x_i\beta}}{\left(1+e^{x_i\beta}\right)^2}\right]$$

On retrouve naturellement la formule proposée dans le cadre du premier chapitre pour le modèle logit dichotomique.

3.2.3. Exemples de modèles logit multinomial

Le exemple de modèle logit multinomial est tiré de Perloff et Watcher (1979). Aux Etats-Unis en 1977-1978 est mis en place un programme de subventions destiné à lutter contre le chômage endémique lié à la crise économique du début des années 70. Ce programme vise à proposer des réductions de taxes aux entreprises qui embauchent de nouveaux salariés et plus particulièrement des salariés non qualifiés. Perloff et Watcher (1979) proposent d'évaluer l'utilité de ce programme en régressant la variation en pourcentage de l'emploi dans la $i^{ème}$ firme, notée y_i^*, sur différentes variables explicatives incluant notamment une variable dummy indiquant si l'entreprise participe au programme. Etant donnés que les premiers résultats de ces régressions n'étaient pas satisfaisant, les auteurs ont alors regroupées les entreprises en cinq classes : celles pour lesquels y_i^* était compris dans les intervalles $S_0 = \left]-\infty, -1\right]$, $S_1 = \left]-1, 2\right]$, $S_2 = \left]2, 30\right]$, $S_3 = \left]30, 45\right]$ et $S_4 = \left[45, +\infty\right[$. Ils ont alors estimé un modèle non ordonné de type logit multinomial en essayant de modéliser la probabilité

$$p_{i,j} = Prob\left(y_i^* \in S_j\right) \tag{3.27}$$

en utilisant les mêmes variables explicatives que dans leurs premières estimations. Dans leur modèle les variables $x_{i,j}$ sont indépendantes des modalités ($x_{i,j} = x_i$) et les coefficients β_j varient avec les modalités. **On a vu que généralement lorsque la variable latente y_i^* est continue, distribuée selon une loi normale et observable, il est préférable de régresser directement y_i^* sur un ensemble de variables explicatives plutôt que de vouloir modéliser les probabilités que y_i^* tombe dans certains intervalles.** Pourtant, si y_i^* est distribuée selon une loi inconnue et non normale qui dépend de x_i d'une façon plus compliquée qu'un simple problème de localisation dans un intervalle, la procédure de Perloff et Watcher peut donner de meilleurs résultats que les MCO de y_i^* sur x_i.

****
*** Chercher d'autres applications sur EconLit **
****

3.3. Application

****
*** Application Eviews ou Limdep à programmer ***
****

3.4. Logit Conditionnel

Comme nous l'avons vu, le modèle logit multinomial indépendant permet d'envisager une modélisation où les paramètres diffèrent selon les modalités, mais où les variables explicatives sont les mêmes quelles que soient les modalités. Ces dernières ne varient qu'avec les individus. Si l'on reprend la notation de l'utilité stochastique utilisée précédemment, on a un modèle où la probabilité que l'individu i choisisse la modalité j s'écrit sous la forme :

$$Prob\,(y_i = j) = \frac{\exp\left[v\left(x_{i,j}\right)\right]}{\sum\limits_{k=0}^{m} \exp\left[v\left(x_{i,k}\right)\right]} \tag{3.28}$$

où la fonction $v\,(.)$ est linéaire, les paramètres β_j diffèrent selon les modalités et que les variables explicatives varient uniquement en fonction des individus :

$$v\left(x_{i,j}\right) = x_i \beta_j \tag{3.29}$$

Une alternative à ce modèle logit multinomial consiste **à supposer qu'au contraire les paramètres β sont indépendants des modalités et que ce sont les variables explicatives qui diffèrent selon les modalités et les individus :**

$$v\left(x_{i,j}\right) = x_{i,j} \beta \tag{3.30}$$

On obtient alors, un modèle logit multinomial conditionnel (ou logit conditionnel) introduit par McFadden (1973). La définition d'un modèle logit conditionnel est ainsi la suivante :

Definition 3.9. ***Dans un modèle logit conditionnel, la probabilité que l'individu i choisisse la modalité j, $\forall j = 0, .., m$, est définie par :***

$$Prob\,(y_i = j) = \frac{\exp\left(x_{i,j}\beta\right)}{\sum\limits_{k=0}^{m} \exp\left(x_{i,k}\beta\right)} = \frac{\exp\left(x_{i,j}^{*}\beta\right)}{1 + \sum\limits_{k=1}^{m} \exp\left(x_{i,k}^{*}\beta\right)} \tag{3.31}$$

où par convention $x_{i,j}^{*}=x_{i,j}-x_{i,0}$.

Nous allons à présent étudier quels sont les avantages et limites de cette spécification avant de proposer différentes applications.

3.4.1. Spécification du logit conditionnel

Commençons par étudier le rapport de probabilités dans un modèle logit conditionnel. On constate immédiatement que ce modèle vérifie l'hypothèse IIA tout comme le modèle logit multinomial.

Proposition 3.10. ***Dans le cas d'un modèle logit conditionnel, le rapport des probabilités associés à deux modalités*** j ***et*** k ***distinctes,*** $\forall j=0,..,m$ ***et*** $\forall k=0,..,m$, ***s'écrit sous la forme :***

$$\frac{p_j}{p_k}=\frac{Prob\left(y_i=j\right)}{Prob\left(y_i=k\right)}=\frac{\exp\left(x_{i,j}\beta\right)}{\exp\left(x_{i,k}\beta\right)}=\exp\left[\left(x_{i,j}-x_{i,k}\right)\beta\right] \tag{3.32}$$

Ce rapport de probabilités est indépendant des alternatives autres que j ***et*** k.

Ainsi, le modèle logit conditionnel de McFadden vérifie l'hypothèse **d'Indépendence des Alternatives Non Pertinentes** ($IANP$ ou IIA en anglais pour *Independance of Irrelevant Alternative*). Donc l'avantage de ce modèle ne se situe pas dans ses propriétés vis à vis de l'hypothèse d'IIA.

L'avantage de ce modèle se situe d'avantage dans la possibilité qui est offerte de prédire la probabilité d'une nouvelle modalité (virtuelle) en fonction de variables explicatives simulées.

Proposition 3.11. ***Le modèle logit conditionnel permet d'estimer la probabilité associée à une modalité virtuelle de la façon suivante :***

$$\widehat{P}_{m+1}=\frac{\exp\left(\widehat{x}_{i,m+1}^{*}\widehat{\beta}\right)}{1+\sum\limits_{k=1}^{m}\exp\left(x_{i,k}^{*}\widehat{\beta}\right)+\exp\left(\widehat{x}_{i,m+1}^{*}\widehat{\beta}\right)} \tag{3.33}$$

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

C'est l'exemple typique du modèle hypothétique de choix de transport cité dans Amemiya (1981) et Alban (2000). Prenons l'exemple d'une collectivité territoriale envisageant la mise en place d'un nouveau mode de transport public, le tramway, en plus des modes de transports collectifs existant (le bus pour simplifier). Pour évaluer la probabilité que les administrés choisissent le tramway, on conduit tout d'abord une enquête sur les choix des modes de transport existant : le but est de calculer la probabilité que l'individu choisisse le bus (modalité 1), la voiture (modalité 2), ou le vélo (modalité 3). On a ici $m+1=4$ modalités, la modalité de référence (codée 0) étant les autres modes de transports (marche à pieds, roller, auto-stop etc..). Les variables explicatives sont exprimées en différences par rapport à leur valeur prises dans la modalité 0. Il s'agit par exemple du temps de transport moyen du domicile au lieu de travail pour le mode j, noté $t_{i,j}=x_{i,j}^{[1]}$ et le coût au kilomètre de ce mode, noté $c_{i,j}=x_{i,j}^{[2]}$, pour l'individu i. Le modèle donne alors la probabilité qu'un individu caractérisé par des temps relatifs $(t_{i,1}, t_{i,2}, t_{i3})$ et des coûts $(c_{i,1}, c_{i,2}, c_{i3})$ choisisse le mode de transport $j = 1, 2, 3$. La probabilité que l'individu i choisisse le bus est par exemple égale à :

$$Prob\,(y_i = 1) = \frac{\exp\left(\beta_0 + \beta_1\, t_{i,1} + \beta_2\, c_{i,1}\right)}{1 + \sum\limits_{k=1}^{3} \exp\left(\beta_0 + \beta_1\, t_{i,j} + \beta_2\, c_{i,j}\right)} = \frac{\exp\left(x_{i,1}^{*}\beta\right)}{1 + \sum\limits_{k=1}^{3} \exp\left(x_{i,k}^{*}\beta\right)}$$

avec $\beta = (\beta_0\, \beta_1\, \beta_2)'$ et $x_{i,j}^{*} = (1\ t_{i,j}\ c_{i,j})$.

Comme le mode de transport métro (modalité 4) n'existe pas encore, les variables de temps de trajet $t_{i,4}$ et de coût $c_{i,4}$ ne sont pas disponibles. Mais elles peuvent être simulées à partir d'une évaluation du temps de trajet du métro et du coût au kilomètre de ce mode de transport dans d'autres villes. Soient $\widehat{t}_{i,4}$ et $\widehat{c}_{i,4}$ les évaluations correspondantes. Si l'on dispose en outre d'un estimateur convergent $\widehat{\beta}$ du vecteur de paramètres β, on peut alors calculer la probabilité qu'un individu i prenne le métro lorsque celui-ci sera effectivement mis en place :

$$\widehat{Prob}\,(y_i = 4) = \frac{\exp\left(\widehat{\beta}_0 + \widehat{\beta}_1\, t_{i,4} + \widehat{\beta}_2\, c_{i,4}\right)}{1 + \sum\limits_{k=1}^{3} \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1\, t_{i,j} + \widehat{\beta}_2\, c_{i,j}\right) + \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1\, t_{i,4} + \widehat{\beta}_2\, c_{i,4}\right)}$$

Cela donne la probabilité que l'individu i choisisse le métro plutôt que les autres modes de transport.

3.4.2. Estimations des paramètres du logit conditionnel

Tout comme dans le cas du modèle logit multinomial, plusieurs méthodes peuvent être utilisées pour estimer les paramètres d'un modèle logit conditionnel : méthodes du maximum de vraisemblance, méthodes de moments, méthodes non paramétriques et semi-paramétriques. Nous n'étudierons ici que la méthode du maximum de vraisemblance à information complète.

La vraisemblance associée à un modèle logit conditionnel à $m+1$ modalités s'écrit en fonction d'un vecteur $\beta \in \mathbb{R}^K$ de K paramètres.

$$\log L\,(y, \beta) = \sum_{i=1}^{N} \sum_{j=0}^{m} y_{i,j} \log\left[Prob\,(y_i = j)\right]$$

avec $y_{i,j} = 1$ si $y_i = j$ et 0 sinon, et où les probabilités $Prob\,(y_i = j)$ sont définies par :

$$Prob\,(y_i = j) = \frac{\exp\left(x_{i,j}\beta\right)}{\sum\limits_{k=0}^{m} \exp\left(x_{i,k}\beta\right)} = \frac{\exp\left(x_{i,j}^{*}\beta\right)}{1 + \sum\limits_{k=1}^{m} \exp\left(x_{i,k}^{*}\beta\right)}$$

avec par convention $x_{i,j}^{*} = x_{i,j} - x_{i,0}$.

Proposition 3.12. ***La log vraisemblance associée à un modèle logit conditionnel s'écrit alors :***

$$\log L\,(y, \beta) = \sum_{i=1}^{N} \sum_{j=0}^{m} y_{i,j} x_{i,j}\beta - \sum_{i=1}^{N} \log\left[\sum_{k=0}^{m} \exp\left(x_{i,j}\beta\right)\right] \tag{3.34}$$

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

avec $x^*_{i,j} = x_{i,j} - x_{i,0}$ ***par convention.***

En effet, on sait que la log-vraisemblance est définie par la relation suivante :

$$\begin{aligned}\log L(y,\beta) &= \sum_{i=1}^{N}\sum_{j=0}^{m} y_{i,j}\log(p_{i,j}) \\ &= \sum_{i=1}^{N}\sum_{j=0}^{m} y_{i,j}\log\left[\frac{\exp\left(x^*_{i,j}\beta\right)}{1+\sum_{k=1}^{m}\exp\left(x^*_{i,j}\beta\right)}\right]\end{aligned}$$

On en déduit la relation suivante :

$$\begin{aligned}\log L(y,\beta) &= \sum_{i=1}^{N}\sum_{j=0}^{m} y_{i,j}x^*_{i,j}\beta - \sum_{i=1}^{N}\left(\sum_{j=0}^{m} y_{i,j}\right)\log\left[1+\sum_{k=1}^{m}\exp\left(x^*_{i,j}\beta\right)\right] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{m} y_{i,j}x^*_{i,j}\beta - \sum_{i=1}^{N}\log\left[1+\sum_{k=1}^{m}\exp\left(x^*_{i,j}\beta\right)\right]\end{aligned}$$

puisque par construction $\sum_{j=0}^{m} y_{i,j} = 1$ et que $x^*_{i,0} = 0$. On retrouve alors l'expression (3.34) de la log-vraisemblance. De la même façon que pour un modèle logit multinomial, **la fonction de log-vraisemblance d'un modèle logit conditionnel est globalement concave.**

3.4.3. Exemples de modèles logit conditionnel

Un premier exemple de modèle logit conditionnel est donné dans l'étude de 1976 de McFadden (prix Nobel 2000). Dans cette étude, préalablement réalisée en 1968, McFadden utilise un logit conditionnel pour analyser la sélection des projets autoroutiers faite par la division californienne des autoroutes (*California Division of Highways*) pour les districts de San Fransisco et de Los Angles durant les années 1958-1966. L'échantillon porte sur $N = 65$ projets. Le $i^{ème} projet$ peut être choisi parmi m_i routes possibles et la probabilité de sélection est donnée par :

$$Prob(y_i = j) = \frac{\exp(x_{i,j}\beta)}{\sum_{k=0}^{m_i}\exp(x_{i,k}\beta)} \quad \forall j = 0,1,..,m_i \tag{3.35}$$

où $x_{i,j}$ est un ensemble de caractéristiques attribuées à la route j dans le projet i (durée de trajet, nombre de kilomètre, difficulté de construction etc..). Naturellement pour chaque projet, le nombre de routes possibles diffère : d'où la présence d'un terme m_i pour le nombre de modalités. Naturellement, l'ensemble des caractéristiques $x_{i,j}$ diffère avec la nature du projet mais aussi selon les routes envisagées pour ce projet.

McFadden considère deux ensembles de variables explicatives $x_{i,j}$: un premier ensemble ne comportant que des variables de coûts et de bénéfices, le second ensemble regroupe les variables du premier ainsi que des variables qui expriment les sentiments de la population sur le projet et le degré selon lequel la population sera affectée par le projet. A chaque ensemble de variables explicatives, McFadden associe un modèle logit conditionnel différent et le choix du modèle est fait selon le critère de la log-vraisemblance et le critère du nombre de prédictions fausses. Ce que montre ainsi McFadden c'est que pour chaque projet le classement entre les différentes routes (selon la probabilité de sélection $p_{i,j}$) diffère selon le modèle utilisé. Suivi que l'on ne considère que des variables de coûts - bénéfices ou des variables liées aux souhaits de la population, les probabilités affectés pour un même projet aux différentes routes varient.

3.4.4. Applications

****
*** Application Eviews ou Limdep à programmer ***
****

3.5. Logit Universel

La dernière catégorie des modèles logit multinomiaux est celle du logit universel, qui comme son nom l'indique englobe le logit multinomial indépendant et le logit multinomial conditionnel de McFadden. Si l'on reprend la notation de l'utilité stochastique utilisée précédemment, on a un modèle où la probabilité que l'individu i choisisse la modalité j s'écrit sous la forme :

$$Prob\left(y_i = j\right) = \frac{\exp\left[v\left(x_{i,j}\right)\right]}{\sum\limits_{k=0}^{m} \exp\left[v\left(x_{i,k}\right)\right]} \tag{3.36}$$

où la fonction $v\left(.\right)$ est linéaire, les paramètres β_j diffèrent selon les modalités et que les variables explicatives varient uniquement en fonction des individus :

Definition 3.13. ***Le modèle logit multinomial universel (ou logit universel) est obtenu pour toute fonction*** $v\left(.\right)$ ***continue dépendant de paramètres*** β_j ***conditionnels aux modalités et de l'ensemble des variables explicatives du modèle :***

$$v\left(x_{i,j}\right) = v\left(\beta_j,\, x_{ij}\right) \tag{3.37}$$

La probabilité que l'individu i ***choisisse la modalité*** j***,*** $\forall j = 0, .., m$, ***est alors définie par :***

$$Prob\left(y_i = j\right) = \frac{\exp\left[v\left(\beta_j,\, x_{ij}\right)\right]}{\sum\limits_{k=0}^{m} \exp\left[v\left(\beta_j,\, x_{ij}\right)\right]} \tag{3.38}$$

On peut montrer que si les fonctions $v\left(\beta_j,\, x_{ij}\right)$ dépendent de l'ensemble des caractéristiques, le modèle logit universel ne satisfait pas l'hypothèse d'indépendance des alternatives non pertinentes (IIA). Lorsque $v\left(.\right)$ est linéaire dans les paramètres β_j, on retrouve alors le modèle logit indépendant.

4. L'hypothèse d'indépendance des alternatives non pertinentes

Tests + modèles alternatifs.

4.1. Test de l'hypothèse IIA

4.2. Modèle Alternatifs

4.2.1. Probit multinomial

4.2.2. Logit Hierarchisé

cf amemiya

Bibliographie

Amemiya T. (1981), "Qualitative Response Models : A Survey", *Journal of Economic Litterature,* 19(4), 481-536

Alban T. (2000), "Econométrie des Variables Qualitatives", *Dunod.*

David J.M. et Legg W.E. (1975), "An application of Multivariate Probit Analysis to the Demand for Housing", *Journal of Business and Economic Statistics,* August 1975, 295-300

Gourieroux C. (1989), "Econométrie des Variables Qualitatives", *Economica.*

Gurland J., Lee I.et Dahm P. (1960), "Polytchotomous Quantal Response in Biological Assay", *Biometrics*, Sept. 1960, 382-388.

Greene W.H. (1997), "Econometric Analysis", *Londres, Prentice Hall.*

Judge G.G., Miller D.J. et Mittelhammer R.C. (2000), "Econometric Foundations", *Cambridge University Press.*

Maddala. G.S. (1983), "Limited-dependent and Qualitative Variables in Econometrics", *Econometric Society Monographs,* 3, Cambrige University Press.

McFadden D. (1976), "The Revealed Prefrences of a Government Bureaucracy: Empirical Evidence", *Bell Journal of Economic Mangament*, Spring 1976, 55-72

Spector L.C. et MazzeoMcFadden M. (1980), "Probit Analysis and Economic Education", *Journal of Economic Education*, 11(2), 37-44

Tobin J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24-36.