

L 3.11

M4

6601

ALBAN THOMAS

MANUEL ET EXERCICES CORRIGÉS

Économétrie des variables qualitatives



DUNOD

025529164

33

Économétrie des variables qualitatives

D2

2000 - 61135

Collection Éco Sup

Manuels et exercices corrigés

- R. Bourbonnais
Économétrie
- Y. Dodge
Analyse de régression appliquée
- B. Goldfarb, C. Pardoux
Introduction à la méthode statistique (Gestion. Économie)
- O. Jokung-Nguéna
Microéconomie de l'incertain (Risque et décisions)
- J.-P. Lecoutre
Statistique et probabilités
- P. Petauton
Théorie de l'assurance dommages
Théorie et pratique de l'assurance vie
- J.-L. Viviani
Gestion de portefeuille
- T. de Montbrial, E. Fauchart
Introduction à l'économie (Microéconomie. Macroéconomie)
- A. Thomas
Économétrie des variables qualitatives

Manuels

- B. Bernier, Y. Simon
Initiation à la macroéconomie
- B. Bernier, H. L. Védie
Initiation à la microéconomie
- J. Fourastié
Mathématiques appliquées à l'économie
- B. Grais
Statistique descriptive
Méthodes statistiques
- B. Guillochon
Économie internationale
- P. Kauffmann
Statistiques. Information, estimation, tests
- J. de Lagarde
Initiation à l'analyse des données
- A. Planche
Mathématiques pour économistes. Algèbre
Mathématiques pour économistes. Analyse

- F. Poulon
Économie générale

- A. Redslob
Introduction à la théorie macroéconomique
- B. et D. Saby
Les grandes théories économiques
- A. de Servigny, I. Zelenko
Économie financière
- J.-L. Sol
Mathématiques. Accès à l'université
- L. Stoléru
L'économie
(Comprendre l'avenir)
- D. Temam
La nouvelle comptabilité nationale
- A. Varoudakis
La politique macroéconomique

Exercices corrigés avec rappels de cours

- J. Fourastié
Mathématiques appliquées à l'économie
- B. Grais
Statistique descriptive
- J.-P. Lecoutre, P. Pilibossian
Algèbre
Analyse II
- J.-P. Lecoutre, S. Maille-Legait, P. Tassi
Statistique
- R. Sandretto
Probabilités

Travaux dirigés avec rappels de cours

- S. Brana, M.-C. Bergouignan
Macroéconomie
- S. Brana, M. Cazals, P. Kauffmann
Économie monétaire et financière
- B. Guillochon, A. Kawacki
Économie internationale
- J.-P. Lecoutre, P. Pilibossian
Analyse I
- P. Médan
Microéconomie

MANUEL ET EXERCICES CORRIGÉS

ALBAN THOMAS

Enseignant à l'université des sciences sociales de Toulouse
Directeur de recherche à l'INRA

Économétrie des variables qualitatives

DUNOD

DL- 20.04.2000

17265

Ce pictogramme mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du **photocopillage**.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les

établissements d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation du Centre français d'exploitation du droit de copie (CFC, 20 rue des Grands-Augustins, 75006 Paris).



LE PHOTOCOPIAGE
TUE LE LIVRE

© Dunod, Paris, 2000

ISBN 2 10 004665 9

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite selon le Code de la propriété intellectuelle (Art. L. 122-4) et constitue une contrefaçon réprimée par le Code pénal. * Seules sont autorisées (Art. L. 122-5) les copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations justifiées par le caractère critique, pédagogique ou d'information de l'œuvre à laquelle elles sont incorporées, sous réserve, toutefois, du respect des dispositions des articles L. 122-10 à L. 122-12 du même Code, relatives à la reproduction par reprographie.

BnF
DEF

Table des matières

Avant-propos	IX
1. Introduction aux variables qualitatives	1
I. Classification des variables qualitatives	2
II. Variables indicatrices et régression	4
III. Application	7
IV. Tableaux de contingence	8
V. Application : revenu et nombre d'enfants	10
VI. Modèle log-linéaire	13
VII. Application numérique	15
VIII. Le χ^2 minimum	19
IX. Modèle à probabilité linéaire	22
À retenir	23
<i>Exercices</i>	23
2. Méthodes d'estimation	27
I. Le maximum de vraisemblance	27
A. Définition de la vraisemblance	28
B. Dérivation et propriétés de l'estimateur	29
C. Vraisemblance concentrée	30
D. Inférence statistique	31
E. Exemple : régression hétéroscédastique	34
II. Les modèles de comptage	34
A. Modèle de Poisson	35
B. Extensions	36
C. Application : choix d'irrigation	37
III. Estimateurs des moments	39
A. L'estimateur par variable instrumentale	39
B. L'estimateur GMM	40
C. Exemple : les moindres carrés généralisés	43
D. Inférence statistique	44

E. Application : valeurs foncières	45
IV. Algorithmes d'optimisation	47
A. Position du problème	48
B. Méthodes de gradient	49
<i>À retenir</i>	50
3. Modèles à variable qualitative binaire	51
I. Variable binaire et variable latente	51
II. Modèles binaires Logit et Probit	54
A. Le modèle Probit	55
B. Le modèle Logit	57
C. Comparaison entre le Probit et le Logit	59
III. Inférence dans les modèles binaires	60
A. Interprétation des paramètres estimés	60
B. Calcul des effets marginaux	60
C. Test d'ajustement global	61
D. Critères de choix de modèles	62
E. Le problème des variables omises	64
F. Test de l'hétéroscédasticité	65
IV. Application : choix du mode de climatisation	66
V. Extensions	70
A. Les modèles binaires bivariés	70
B. Application	72
C. Méthodes d'estimation non paramétriques	73
D. Modèles avec données de panel	76
<i>À retenir</i>	78
<i>Exercices</i>	78
4. Les modèles multinomiaux	83
I. Introduction	84
II. Les modèles ordonnés	85
A. Exemple : enquête de consommation	85
B. Formulation courante d'un modèle ordonné	86
III. Application	87
IV. Les modèles séquentiels	90
V. Modèles non ordonnés : le Logit multinomial	91
A. Généralisation du Logit binaire au cas multinomial	91
B. Logit multinomial : spécification et estimation	92
C. Le Logit conditionnel de McFadden	94
D. Un modèle Logit plus général	96
VI. Application : mode de transport	97

VII.	Modèles de choix probabilistes et hypothèse IIA	99
A.	Modèles probabilistes à choix discret	100
B.	L'hypothèse IIA	101
C.	Un test de la propriété IIA	103
D.	Application numérique	104
VIII.	Modèles multinomiaux alternatifs	107
A.	Le Logit multinomial emboîté	107
B.	Le Logit multinomial hiérarchisé	108
C.	Application : transport aérien ou terrestre	111
D.	Le Probit multinomial	113
IX.	Extensions	114
A.	Estimateur MSM de McFadden	114
B.	L'estimateur GHK	116
	<i>À retenir</i>	118
	<i>Exercices</i>	118
5.	Les modèles à variable dépendante limitée	121
I.	Introduction	121
II.	Modèles censurés et tronqués	123
III.	Le modèle Tobit simple	126
A.	Comparaison avec les modèles tronqués	127
B.	Estimation par maximum de vraisemblance	128
C.	Estimation en deux étapes	130
D.	Effets marginaux avec le modèle Tobit	131
IV.	Tests de spécification	132
A.	Hétéroscédasticité	132
B.	Normalité	133
C.	Égalité des paramètres	134
V.	Application : dépenses de restaurant	134
VI.	Extensions du modèle Tobit	137
A.	Modèles Tobit à censure multiple	138
B.	Exemple : les modèles à friction	139
C.	Modèles Tobit généralisés	140
D.	Modèles à seuils stochastiques	142
VII.	Modèles de sélection	144
A.	Mécanisme de sélection	144
B.	Troncature auxiliaire	145
C.	Application : dépenses en eau domestique	146
D.	Application : équation de salaire	148
E.	Modèles à régimes	149
F.	Application : dépenses en logement	151
	<i>À retenir</i>	155
	<i>Exercices</i>	155

Corrigés des exercices	159
Bibliographie	169
Index	177



Avant-propos

Les variables qualitatives constituent une partie importante de l'économétrie contemporaine. Il existe trois façons de considérer les variables qualitatives en économétrie et en statistique :

- les incorporer comme variables explicatives dans un modèle de régression ;
- étudier leur corrélation ou leur indépendance ;
- les traiter comme variables dépendantes (expliquées).

Ces variables à valeurs discrètes forment souvent la majorité des informations disponibles dans les enquêtes, en raison notamment de leur grande facilité de collecte. D'autre part, les modèles économiques expliquant des événements et non des réalisations de variables continues ont connu un développement important. Les théories du consommateur et du producteur, par exemple, fournissent à présent les bases d'une formalisation rigoureuse des mécanismes de choix des agents économiques : décision de travail, d'achat, choix d'une technologie, etc.

Mais paradoxalement, alors que les variables qualitatives sont de plus en plus utilisées, leur analyse est encore porteuse d'une réputation de complexité, probablement parce que les techniques d'inférence par moindres carrés ne sont pas adaptées. La méthode du *maximum de vraisemblance* devant être utilisée dans la plupart des cas, l'économétrie des variables qualitatives entraînerait donc une grande complexité dans les calculs. Cet ouvrage a pour objectif de montrer que les techniques de modélisation des variables qualitatives possèdent l'avantage d'une grande cohérence méthodologique. Une fois que les principes à la base de la construction de la vraisemblance ont été compris, il est très facile d'estimer des modèles à variable dépendante qualitative. Les modèles les plus courants sont maintenant gérés par la plupart des logiciels économétriques et, pour les modèles plus élaborés, le langage de programmation dédié des logiciels comme SAS, GAUSS, EVIEWS et LIMDEP peut être mis à profit. Ces logiciels modernes permettent en outre de maximiser toute vraisemblance avec des algorithmes de maximisation numériques optimisés, donc fiables et rapides.

Cet ouvrage se limite aux modèles les plus couramment utilisés en économétrie appliquée. Des extensions aux chapitres présentent cependant des résultats récents en économétrie des variables qualitatives. On ne traitera ni

des modèles de durée, ni des modèles multivariés. Le lecteur intéressé trouvera ces thèmes bien traités dans les ouvrages de Gouriéroux (1989), Madala (1983) et Lancaster (1990). L'optique de l'ouvrage est résolument appliquée : on illustre les principaux modèles et méthodes au moyen d'applications sur données réelles, avec les commandes pour les logiciels SAS, GAUSS, EVIEWS et LIMDEP. Dans le domaine des variables qualitatives, le logiciel le plus élaboré est certainement LIMDEP, permettant d'estimer quasiment tous les modèles décrits dans cet ouvrage. GAUSS possède des procédures standard pour les modèles de base (essentiellement ceux des chapitres 1, 3 et 4), mais son langage de programmation très intuitif le rend intéressant pour appréhender des modélisations plus avancées. EVIEWS pour sa part est assez comparable à GAUSS : son langage de programmation est moins immédiat, mais la gestion des fichiers de données est plus aisée. Enfin le logiciel SAS, s'il est très utile pour manipuler des bases de données importantes, est peut-être le moins adapté aux méthodes de ce livre : langage de programmation difficile à mettre en œuvre, langage matriciel vite limité, très peu de procédures standard pour les modèles à variable dépendante qualitative.

Prérequis

Cet ouvrage s'adressant à des étudiants de deuxième et troisième cycles universitaires et de grandes écoles, on demandera au lecteur des connaissances de base en statistique et économétrique : tests, principes de l'estimation paramétrique, algèbre linéaire de base, moindres carrés ordinaires. Le lecteur intéressé par l'acquisition de telles notions pourra se référer utilement aux ouvrages de Régis Bourbonnais, *Économétrie*, et Pascal Kauffmann, *Statistique*, dans la même collection.

Notations

On utilisera dans la mesure du possible des notations mathématiques cohérentes et invariantes entre les chapitres. L'entier N dénote la taille de l'échantillon, pour des individus indicés par i , $i = 1, 2, \dots, N$. Les modalités et choix sont repérés par les indices j, k, l , et le nombre de modalités par m ou m_i selon les cas. La notation $y_i = \{0, 1\}$ signifie que y_i peut prendre deux valeurs : 0 ou 1. Les variables expliquées, continues ou à valeurs discrètes sont notées y_i , les variables explicatives x_i , et un vecteur $(1 \times K)$ ou encore z_i . Les paramètres sont notés β (un vecteur $K \times 1$) ou γ (un vecteur de dimension $G \times 1$). Ainsi, la combinaison linéaire formée à partir d'un vecteur de variables explicatives pour l'individu i , x_i , et d'un vecteur de paramètres β sera : $x_i \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK}$.

Les variables aléatoires v_i , u_i ou v_i^* sont les résidus (termes d'erreur) et les variables latentes sont notées y_i^* . La vraisemblance est notée L , le gradient $g(\cdot)$, la matrice hessienne $H(\cdot)$. En ce qui concerne les dimensions des vecteurs et matrices, la notation $A (N \times K)$ indique que la matrice A a N lignes et K colonnes. De même, $B (N \times 1)$ signifie que le vecteur B a N lignes et 1 colonne : c'est un vecteur colonne.

La transposée d'une matrice ou vecteur x est notée x' , et on rappelle ici les règles de dérivation matricielle : $\partial(x\beta)/\partial\beta = x'$; $\partial(x'Ax)/\partial x = 2Ax$ si A est symétrique, $(A + A')x$ sinon. Rappelons de plus l'écriture de la dérivée seconde par rapport à une matrice ou un vecteur : $\partial^2 f(\beta)/\partial\beta\partial\beta'$. Par exemple, les dérivées de la fonction $f = (y_i - x_i\beta)^2$ par rapport au vecteur de paramètres β de dimension $K \times 1$ sont : $\partial f/\partial\beta = -2(y_i - x_i\beta) \times x'_i$, un vecteur $K \times 1$, et $\partial^2 f/\partial\beta\partial\beta' = 2x'_i x_i$, une matrice $K \times K$.

En ce qui concerne les lois statistiques, $f(\cdot)$ dénote une fonction de densité, $F(\cdot)$ une fonction de répartition, et dans le cas de la loi normale, $\phi(\cdot)$ et $\Phi(\cdot)$ respectivement. Enfin, on utilise dans le texte la notation décimale française avec virgule. En revanche, les exemples d'application étant illustrés de sorties de logiciels anglo-saxons, celles-ci utilisent la convention décimale avec un point.



1. Introduction aux variables qualitatives

Nous introduisons dans ce chapitre l'analyse des variables qualitatives, en discutant d'abord de leur provenance et intérêt. On présente l'étude de modèles de régression avec variables indicatrices en décrivant l'interprétation de leur coefficient. On étudie ensuite les modalités d'analyse de la corrélation entre les variables explicatives, dans un contexte purement descriptif, à l'aide de tableaux de contingence et du modèle log-linéaire. Les méthodes de régression dans le cas d'observations groupées seront ensuite développées autour de la méthode du χ^2 minimum. Ce chapitre s'achèvera par un aperçu des modèles linéaires avec variable dépendante qualitative, dont les inconvénients motiveront l'analyse par *maximum de vraisemblance* des chapitres 3, 4 et 5.

Objectifs du chapitre : proposer une classification des variables qualitatives ; présenter les principales procédures d'analyse et de test utilisées en statistique descriptive des variables qualitatives ; introduire les modèles à variable expliquée qualitative.

Concepts clés étudiés : enquêtes, tableaux de contingence, tests du χ^2 , modèle log-linéaire, méthode du χ^2 minimum, modèle à probabilité linéaire.

I. Classification des variables qualitatives

Les variables qualitatives sont synonymes de variables à valeurs discrètes (ou plus simplement variables discrètes), c'est-à-dire prenant leurs valeurs dans l'ensemble des entiers naturels. La nature d'une variable qualitative dépend de la façon dont elle a été codée sur un questionnaire ou enregistrée dans une base de données. Dans ce cas, elle apparaît souvent sous une forme aisée à saisir ou à interpréter ultérieurement lors de différents travaux statistiques. Qu'elles proviennent d'une enquête statistique ou d'un fichier constitué par un économètre ayant créé ses propres variables, les variables discrètes représentent souvent la vision qu'a leur créateur du problème à traiter¹.

On considère deux grandes classes de variables discrètes :

- les variables *binaires*, à deux modalités possibles seulement ;
- les variables *polytomiques*, à plus de deux valeurs possibles.

Dans le premier cas, on parle aussi de variables *dichotomiques*, et dans le second cas, de variables *multinomiales*. Les variables binaires sont souvent sous la forme 0 ou 1. Un exemple est la variable « Sexe de l'individu », égale à 0 ou 1 s'il s'agit d'un homme ou d'une femme, respectivement. On peut créer par exemple cette variable à partir d'une variable notée SECU et contenant le numéro de Sécurité sociale à 13 chiffres de l'individu :

$$SEXU = \begin{cases} 1 & \text{si SECU commence par 2,} \\ 0 & \text{si SECU commence par 1.} \end{cases}$$

La convention est de prendre deux valeurs : 0 ou 1, le 1 indiquant l'existence de la caractéristique, 0 son absence². Une variable binaire, admettant une valeur 0 ou 1, est caractérisée par le fait que son espérance est la proportion sur l'échantillon de cas où la variable vaut 1. En effet, l'espérance d'une variable binaire x est :

$$E(x) \approx \frac{1}{N} \sum_{i=1}^N x_i = \frac{\text{Nombre de cas où } x \text{ vaut 1}}{\text{Nombre total de cas}} = p(x)$$

où N est la taille de l'échantillon (nombre d'observations) et $p(x)$ la fréquence empirique de x (la proportion). Si la probabilité théorique que $x = 1$ peut être estimée par la fréquence empirique, alors l'espérance d'une variable binaire

1. Ceci a pour conséquence qu'une variable qualitative peut la plupart du temps être recodifiée de façon à mieux correspondre aux besoins des utilisateurs de la base de données.

2. D'où le nom de variable binaire, l'analogie étant possible avec l'électronique digitale où le 1 représente le passage de courant (« état haut »), le 0 son absence (« état bas »).

est comparable à la probabilité d'occurrence de la catégorie à laquelle x fait référence.

Dans le second cas, celui des variables polytomiques, la nature des variables dépend du besoin de représenter ou non des catégories. Si une variable décrivant un individu représente les mois passés à la recherche d'un emploi, on a bien affaire à une variable discrète, mais cette variable ne décrit nullement une caractéristique ou une catégorie. Si, au contraire, une variable contient la réponse à une question relative à une caractéristique de l'individu ou à un choix (catégorie socioprofessionnelle avant la perte d'emploi, secteurs dans lesquels la personne souhaiterait trouver un emploi), on parle de *variable catégorielle*. Prenons quelques exemples :

- $$REVENU = \begin{cases} 1 & \text{si l'individu gagne moins de 72 000 F par an ;} \\ 2 & \text{si l'individu gagne entre 72 000 F et 96 000 F par an ;} \\ 3 & \text{si l'individu gagne entre 96 000 F et 120 000 F par an ;} \\ 4 & \text{si l'individu gagne entre 120 000 F et 180 000 F par an ;} \\ 5 & \text{si l'individu gagne plus de 180 000 F par an.} \end{cases}$$
- $$EMPLOI = \begin{cases} 1 & \text{si exploitant agricole ;} \\ 2 & \text{si ouvrier ;} \\ 3 & \text{si employé ;} \\ 4 & \text{si cadre ;} \\ 5 & \text{si sans profession, étudiant.} \end{cases}$$
- $$\bar{EDUCATION} = \begin{cases} 1 & \text{si la personne n'a pas le baccalauréat ;} \\ 2 & \text{si la personne a le baccalauréat mais pas le DEUG ;} \\ 3 & \text{si la personne a le DEUG mais pas la licence ;} \\ 4 & \text{si la personne a la licence mais pas la maîtrise ;} \\ 5 & \text{si l'individu a la maîtrise mais pas de troisième cycle ;} \\ 6 & \text{si l'individu a un troisième cycle.} \end{cases}$$
- $$PRÉFÉRENCE = \begin{cases} 1 & \text{si la personne n'aime pas du tout le roquefort ;} \\ 2 & \text{si la personne n'aime pas trop le roquefort ;} \\ 3 & \text{si la personne est indifférente au roquefort ;} \\ 4 & \text{si la personne aime un peu le roquefort ;} \\ 5 & \text{si l'individu aime beaucoup le roquefort.} \end{cases}$$

REVENU est catégorielle ordonnée : les niveaux de revenu sont triés par ordre croissant.

EMPLOI est catégorielle non ordonnée.

EDUCATION est catégorielle séquentielle : un niveau est conditionné par l'obtention du précédent diplôme.

PRÉFÉRENCE est aussi catégorielle ordonnée quoique un peu différente de *REVENU* : on a trié non pas par rapport à une valeur numérique, mais par rapport à une échelle dans les goûts.

Voir Maddala (1983) pour une présentation intéressante de ces différentes catégories.

Il existe deux utilisations de ces variables qualitatives, selon qu'elles sont *explicatives* (dans un modèle de régression par exemple) ou *dépendantes* (expliquées). Cette distinction conduit à une interprétation de ces variables comme correspondant à un *choix*, que l'on cherchera à expliquer, ou comme un facteur spécifique aux individus, entreprises, etc. Dans le premier cas les variables seront endogènes (engendrées par le modèle économique ou statistique), dans le second cas elles seront prédéterminées (au sens économétrique du terme).

II. Variables indicatrices et régression

Les variables discrètes sont souvent utilisées dans les modèles de régression pour capter l'impact d'une caractéristique des entités observées sur l'espérance de la variable dépendante. On parle de *variables indicatrices* ou variables muettes (*dummy variables*).

Les mêmes précautions doivent être prises lorsqu'on incorpore des variables indicatrices dans une équation de régression, ainsi qu'avec tout modèle comportant des variables continues. Tout d'abord, à moins de travailler avec des variables centrées par rapport à leur moyenne sur l'échantillon, il importe d'incorporer un terme constant dans le modèle. L'omission de ce dernier revient en effet à spécifier un modèle « passant par l'origine », c'est-à-dire une relation affine, ce qui n'est pas toujours pertinent. De plus, l'inférence basée sur le coefficient de détermination linéaire R^2 est faussée.

Il est également important, dans une régression avec terme constant, de veiller à ce que la somme des variables indicatrices soit différente de 1. Si non, l'estimateur des *moindres carrés ordinaires* ne pourra être calculé, la matrice des variables explicatives contenant deux colonnes non linéairement indépendantes. La solution usuelle est d'omettre le terme constant, lequel sera reconstitué par les variables indicatrices, ou bien encore de négliger l'une de celles-ci. Concernant les variables polytomiques à plus de deux modalités, une erreur courante consiste à insérer une telle variable qualitative dans une équation de régression. Or, cela n'a que rarement un sens. Considérons par exemple la régression du salaire sur une constante et une variable polytomique représentant la catégorie socioprofessionnelle (CSP). Le coefficient associé à CSP indiquera le gain marginal en termes salariaux de passer d'une catégorie à celle immédiatement supérieure, et ce gain sera constant. Mais puisque CSP est discrète, rien n'empêche de modifier la définition de cette variable dans la base de données, en considérant par exemple un agencement mettant en avant non

plus la progression attendue des salaires, mais la nature du travail demandé (direction, manuel, etc). La régression du salaire sur CSP aboutirait alors à un coefficient estimé dont il serait difficile de fournir une interprétation aisée. La bonne solution, lorsque l'on envisage d'introduire une variable polytomique dans une régression, est de créer une variable indicatrice pour chaque modalité. Une possibilité, pour limiter le nombre d'indicatrices, est de conserver les plus significatives. Pour ce faire, il est utile de disposer d'une mesure de corrélation entre la variable dépendante continue et une variable discrète. Le résultat suivant donne la mesure de corrélation entre une variable continue et une variable polytomique.

Définition

Corrélation entre une variable continue et une variable polytomique : pour un échantillon de taille N , la corrélation entre une variable continue z et une variable polytomique y admettant m modalités est calculée par :

$$\rho^2 = \frac{1/N \sum_{j=1}^m n_j (\bar{z}_j - \bar{z})^2}{s_z^2}$$

où n_j est l'effectif pour la modalité j , \bar{z}_j la moyenne empirique de z pour la catégorie j , \bar{z} la moyenne empirique de z sur l'échantillon, et s_z^2 la variance de z .

On voit que ρ^2 est nul si $\bar{z}_1 = \bar{z}_2 = \dots = \bar{z}_m$, c'est-à-dire si l'on a absence de dépendance en moyenne. Dans le cas opposé, $\rho^2 = 1$ si tous les individus de la catégorie j ont la même valeur de z , quel que soit j . Dans le cas de deux catégories seulement, le coefficient de corrélation est :

$$\rho^2 = (s_z^2)^{-1} \frac{n_1 n_2}{N^2} (\bar{z}_1 - \bar{z}_2)^2$$

Pour tester l'absence de corrélation ($\rho^2 = 0$), on fait l'hypothèse que la loi de z est normale dans chaque modalité. Sous l'hypothèse nulle que les z dans chaque modalité ont des moyennes et des variances identiques, on calcule la statistique de test suivante :

$$\frac{\rho^2 / (m - 1)}{(1 - \rho^2) / (N - m)} \sim F(m - 1, N - m)$$

qui suit une distribution de Fisher. En comparant la valeur calculée aux valeurs tabulées de la loi de Fisher à $\nu_1 = m - 1$ et $\nu_2 = N - m$ degrés de liberté, on rejette l'hypothèse de non-corrélation si la statistique de test est supérieure à la valeur théorique pour un niveau de confiance donné (95 % par exemple).

Considérons un échantillon constitué de y , une variable continue, et d'une variable indicatrice notée x . Pour l'échantillon complet, l'estimateur de l'espérance de y_i est $1/N \sum_{i=1}^N y_i$, la moyenne arithmétique de la variable dépendante, qui coïncide avec l'estimateur par moindres carrés ordinaires (MCO) obtenu par régression de y_i sur une constante. Considérons le partage de cet échantillon en deux sous-échantillons de taille N_1 et N_2 respectivement, d'après la règle suivante : les N_1 observations correspondent à $x_i = 0$, les N_2 observations à $x_i = 1$. Il est clair que la moyenne de y_i sur ces deux sous-échantillons correspondra à l'espérance conditionnelle $E(y_i|x_i = 0)$ et $E(y_i|x_i = 1)$ respectivement. On a par exemple :

$$E(y_i|x_i = 1) = \frac{E(y_i|x_i)}{\text{Prob}(x_i)} = \frac{E(y_i|x_i)}{E(x_i)}$$

qui peut être estimé par :

$$E(y_i|x_i = 1) \approx \frac{1/N \sum_{i=1}^N y_i x_i}{1/N \sum_{i=1}^N x_i}$$

Considérons maintenant le modèle de régression suivant :

$$y_i = \beta_0 + x_i \beta_1 + \varepsilon_i$$

où β_0 et β_1 sont deux paramètres, ε_i un terme d'erreur de moyenne 0. Pour le premier sous-échantillon, on a $x_i = 0$ et donc seule la valeur de β_0 devrait intervenir dans l'espérance conditionnelle. Pour l'autre sous-échantillon, cette dernière est calculée à partir des deux paramètres. Les estimateurs par MCO de cette équation correspondent aux quantités recherchées :

$$E(\hat{\beta}_0) = E(y_i|x_i = 0)$$

et :

$$E(\hat{\beta}_0 + \hat{\beta}_1) = E(y_i|x_i = 1)$$

Il est important de garder à l'esprit cette notion de conditionnement : une variable indicatrice fait référence à une certaine catégorie, un sous-échantillon des données d'origine. L'espérance de la variable dépendante n'est pas l'espérance du produit des deux variables quand la variable indicatrice vaut 1. Il faut diviser ce produit par la proportion de l'échantillon telle que l'indicatrice vaut 1, c'est-à-dire par son espérance.

On peut bien sûr croiser des variables indicatrices entre elles et les utiliser dans une régression, de la même façon que pour des variables continues. Les considérations relatives à l'espérance conditionnelle de y (cf. plus haut) restent valables (McCullagh et Nelder, 1983).

III. Application

Pour un échantillon de 3 177 ménages, on observe le revenu annuel *REV*, et une indicatrice *SEX* égale à 1 si le chef de famille est un homme, 0 s'il s'agit d'une femme. Les moyennes sur l'échantillon complet sont calculées avec les commandes GAUSS :

```
meanc(rev) ; meanc(sex) ; meanc((rev.*sex)) ; meanc((rev.*(1-sex)) ;
meanc((rev.*sex)/meanc(sex)) ; meanc((rev.*(1-sex))/meanc(1-sex)) ;
```

	Moyenne
<i>REV</i>	19954,576
<i>SEX</i>	0,803 588 29
<i>REV × SEX</i>	17 595,703
<i>REV × (1 - SEX)</i>	2 358,873 2
<i>REV × SEX/SEX</i>	21 896,415
<i>REV × (1 - SEX)/1 - SEX</i>	12 009,840

Comparons maintenant avec la régression de *REV* sur *SEX* et une constante. Les résultats avec le logiciel GAUSS sont :

Valid cases [1]	3177	Dependent variable [7]	Y
Missing cases [2]	0	Deletion method [8]	None
Total SS [3]	739534945367,892	Degrees of freedom [9]	3175
R-squared [4]	0,066	Rbar-squared [10]	0,066
Residual SS [5]	690522094969,865	Std error of est [11]	14747,450
F(1,3175) [6]	225,360	Probability of F [12]	0,000
Variable	Estimate [13]	Standard Error [14]	t-value [15]
CONSTANT	12009,839744	590,370475	20,342887
X1	9886,575454	658,578927	15,011983
			Prob > t [16]
			Standardized Estimate [17]

Les indications fournies par le logiciel sont :

[1] : nombre d'observations utilisées ; [2] : valeurs manquantes ; [3] : somme des carrés totale (*Sum of Squares*) ; [4] : coefficient R^2 ; [5] : somme des carrés des résidus (*Residual Sum of Squares*) ; [6] : statistique de significativité globale de Fisher ; [7] : variable dépendante (expliquée) ; [8] : méthode de sélection des observations ; [9] : degrés de liberté ; [10] : coefficient de détermination ajusté R^2 ; [11] : écart type de la variable dépendante estimée ; [12] : valeur critique

associée au test de Fisher (voir plus loin dans ce chapitre sur la définition de la p-value).

Le tableau de résultats fournit les coefficients estimés [13], leur écart type [14], la statistique de test de Student [15], la valeur critique associée à cette dernière [16], et le coefficient standardisé [17] défini comme le paramètre estimé multiplié par le rapport des écarts types de la variable explicative et de la variable dépendante.

On vérifie bien que les paramètres estimés permettent de retrouver les espérances conditionnelles de *REV* dans les deux sous-échantillons. Par conséquent, il est équivalent de séparer un échantillon en deux et d'estimer le terme constant de l'équation dans chacun des deux sous-échantillons, ou bien d'introduire une variable indicatrice dans le modèle sur l'échantillon complet et d'en déduire les espérances conditionnelles de la variable dépendante.

IV. Tableaux de contingence

Dans cette section, on présente une analyse statistique des liaisons entre deux variables qualitatives, sans introduire de variables continues dans le modèle.

Supposons que l'on dispose d'un échantillon de taille N et de deux variables d'enquête polytomiques. Notons x et y ces deux variables, x admettant r modalités et y , s modalités. L'effectif de la cellule (i,j) , donnant le nombre d'individus tels que x est caractérisé par la modalité i et y par la modalité j , est noté n_{ij} . On a donc $N = \sum_{i=1}^r \sum_{j=1}^s n_{ij}$. On représente généralement les effectifs sous la forme d'un *tableau de contingence*, c'est-à-dire un tableau croisé à deux entrées :

y	y_1	y_2	...	y_j	...	y_s	
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2\cdot}$
...
x_i	n_{ij}	$n_{i\cdot}$
...
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rs}	$n_{r\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot j}$...	$n_{\cdot s}$	N

Les $n_{i\cdot}$ [1] sont appelées *marges en ligne* et les $n_{\cdot j}$ [2] *marges en colonne*, avec $n_{i\cdot} = \sum_j n_{ij} \forall i$ et $n_{\cdot j} = \sum_i n_{ij} \forall j$. Ces marges indiquent les effectifs totaux de chaque ligne et colonne.

Des indications intéressantes peuvent être déduites de ce type de tableaux si l'on ramène les effectifs en termes relatifs par rapport aux marges en ligne

et en colonne. Plutôt que de travailler avec les effectifs de départ, il est en effet plus instructif de travailler en termes relatifs. En remplaçant dans le tableau ci-dessus les n_{ij} par des n_{ij}/n_i , on obtient le *tableau des profils-lignes*. De façon similaire, les n_{ij}/n_j fournissent le *tableau des profils-colonnes*. Les profils-lignes indiquent la répartition des individus de l'échantillon caractérisés par la modalité de la ligne, selon les différentes modalités en colonne. Les profils-colonnes expriment de façon similaire la répartition des individus de l'échantillon caractérisés par la modalité de la colonne, selon les différentes modalités en ligne (Saporta, 1990).

La moyenne pondérée des profils-lignes (respectivement des profils-colonnes) est égale au profil marginal des colonnes (respectivement profil marginal des lignes) :

$$\sum_{i=1}^r \frac{n_{ij}}{n_i} \left(\frac{n_i}{N} \right) = \frac{n_j}{N} \quad \text{et} \quad \sum_{j=1}^s \frac{n_{ij}}{n_j} \left(\frac{n_j}{N} \right) = \frac{n_i}{N}$$

Les profils présentés ci-dessus sont en fait des distributions (empiriques) des fréquences conditionnelles. Ainsi, n_{ij}/n_i indique la probabilité d'avoir choisi (ou d'être caractérisé par) la modalité j , une fois que la modalité i s'applique. Si x et y sont indépendants, il s'ensuit que les profils-lignes doivent être tous identiques, et par conséquent les profils-colonnes également (en utilisant le résultat ci-dessus), c'est-à-dire :

$$\frac{n_{1j}}{n_1} = \frac{n_{2j}}{n_2} = \cdots = \frac{n_{rj}}{n_r} \quad \forall j \quad \Leftrightarrow \quad \frac{n_{ij}}{n_i} = \frac{n_j}{N}$$

Résultat

Indépendance des modalités : l'indépendance des modalités se traduit par $n_{ij} = n_i n_j / N$, où $n_i n_j / N$ est l'*effectif théorique* sous l'hypothèse d'indépendance.

Ce résultat est à la base du test d'indépendance des modalités de x et y . Si ces variables sont indépendantes, les effectifs de chaque cellule doivent être très proches du produit des marges en lignes et en colonnes. On construit la statistique suivante à partir du tableau de contingence :

$$k^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_i n_j / N)^2}{n_i n_j / N}$$

On montre que cette statistique a comme borne supérieure $N \times \inf(s-1, r-1)$, et que cette borne est atteinte dans le cas d'une liaison parfaite³. Plusieurs

3. Ceci correspond au cas extrême où il n'existe qu'une cellule non vide par ligne.

mesures d'indépendance calculées à partir de la statistique k^2 ont été proposées. Citons les coefficients de contingence de Pearson et de Cramer, dont les expressions sont les suivantes :

$$\text{Coefficient de Pearson : } (k^2/(N + k^2))^{1/2}$$

$$\text{Coefficient de Cramer : } (k^2/(N \times \min(s-1, r-1)))^{1/2}$$

Une valeur proche de 0 indique l'indépendance, une valeur proche de 1 une liaison. Notons également que le signe de la corrélation entre les modalités i et j est le même que celui de la différence $n_{ij} - n_i n_j / N$.

Afin de tester la significativité statistique de l'indépendance entre x et y , on utilise le résultat suivant :

Résultat

Distribution de k^2 : la statistique de test k^2 suit une loi du χ^2 (chi-2) sous l'hypothèse d'indépendance entre x et y , avec $(r-1)(s-1)$ degrés de liberté.

On teste donc l'indépendance des variables polytomiques x et y en comparant k^2 à la valeur théorique de la loi χ^2 , pour un risque d'erreur α donné (5 %). D'autres mesures de dépendance ont été proposées, à partir de la statistique k^2 : le chi-2 de vraisemblance (appelé aussi G^2) et le τ_b de Goodman et Kruskal (voir Goodman et Kruskal, 1979, Marcotorchino *et al.*, 1985). Ces indices sont :

$$G^2 = 2 \sum_{i=1}^r \sum_{j=1}^s n_{ij} \log \left(\frac{n_{ij}}{n_i n_j / N} \right)$$

$$\tau_b = \frac{\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{N n_i} - \sum_{j=1}^s \left(\frac{n_{ij}}{N} \right)^2}{1 - \sum_{j=1}^s \left(\frac{n_{ij}}{N} \right)^2}$$

V. Application : revenu et nombre d'enfants

À partir de l'échantillon de la section III, on étudie la relation entre le revenu et le nombre d'enfants du foyer. Pour cela, on construit un tableau de contingence à deux entrées. La première variable (*REV*), en ligne, est une variable polytomique donnant la classe de revenu. La seconde (*ENF*) indique le nombre

d'enfants du ménage. Les variables sont définies par :

$$REV = \begin{cases} 1 & \text{si revenu} \leq 5\,000 \\ 2 & " " \in [5\,000, 7\,500] \\ 3 & " " \in [7\,500, 10\,000] \\ 4 & " " \in [10\,000, 15\,000] \\ 5 & " " > 15\,000 \end{cases}, \quad ENF = \begin{cases} 0 & \text{si enfants} = 0 \\ 1 & " " \quad 1 \\ 2 & " " \quad 2 \\ 3 & " " \quad 3 \\ 4 & " " \geq 4 \end{cases}$$

Le tableau de contingence est construit avec la procédure Freq du logiciel SAS. Les commandes sont :

```
proc freq data=enquete.enfrev;
tables rev*enf/ chisq;
title 'Tableau de contingence REV vs. ENF';
run;
```

Les résultats figurent ci-dessous. La variable *REV* figure en ligne, *ENF* en colonne. Le logiciel rapporte dans le tableau de contingence quatre expressions : l'effectif \square (*Frequency*), la proportion \square (*Percent*), la proportion en ligne \square (*Row Pct*) et en colonne \square (*Col Pct*). Ces deux dernières expressions sont donc les profils-lignes et profils-colonnes vus plus haut. Les marges en lignes et en colonnes (n_i et n_j) ainsi que les pourcentages correspondants (n_i/N et n_j/N) figurent respectivement à droite de chaque ligne et en bas de chaque colonne. Les deux dernières quantités \square et \square fournissent enfin la somme des marges-lignes (ou colonnes), qui est égale à l'effectif total (3517), et la somme des proportions correspondantes, qui vaut 100 %.

Les statistiques de test figurent dans le tableau du bas (\square). \square indique le type de statistique de test, \square le nombre de degrés de liberté du test (*degrees of freedom*), \square la valeur calculée de la statistique, et \square la p-value associée (voir plus loin). Le logiciel SAS donne cinq types de statistiques : le χ^2 du texte (*Chi-square*), les χ^2 du rapport des vraisemblances (*Likelihood Ratio Chi-Square*) et de Mantel-Hänszel (*Mantel-Haenszel*), les coefficients ϕ (*Phi Coefficient*), de contingence (*Contingency Coefficient*) et de Cramer (*Cramer's V*). Enfin, \square dénote la taille de l'échantillon.

Concernant le test du χ^2 , on voit que l'hypothèse d'indépendance entre les variables *REV* et *ENF* est clairement rejetée : la statistique de test vaut 56,95 pour 16 degrés de liberté. Ces derniers correspondent bien à la valeur $(r - 1) \times (s - 1)$, les deux variables ayant chacune 5 modalités possibles.

Tableau de contingence REV vs. ENF

15:02 Friday, August 27, 1999

TABLE OF REV BY ENF

REV ENF

Frequency

Percent

Row Pct

Col Pct

		1	2	3	4	Total
1	1 348	171	95	69	128	811
	7 9.89	4.86	2.70	1.96	3.64	23.06
	3 42.91	21.09	11.71	8.51	15.78	
	4 21.03	28.64	21.06	21.84	25.70	
2	345	131	88	48	96	708
	9.81	3.72	2.50	1.36	2.73	20.13
	48.73	18.50	12.43	6.78	13.56	
	20.85	21.94	19.51	15.19	19.28	
3	302	88	94	52	90	626
	8.59	2.50	2.67	1.48	2.56	17.80
	48.24	14.06	15.02	8.31	14.38	
	18.25	14.74	20.84	16.46	18.07	
4	443	125	84	84	118	854
	12.60	3.55	2.39	2.39	3.36	24.28
	51.87	14.64	9.84	9.84	13.82	
	26.77	20.94	18.63	26.58	23.69	
5	217	82	90	63	66	518
	6.17	2.33	2.56	1.79	1.88	14.73
	41.89	15.83	17.37	12.16	12.74	
	13.11	13.74	19.96	19.94	13.25	
Total	1655	597	451	316	498	3517 8
	47.06	16.97	12.82	8.98	14.16	100.00 6

STATISTICS FOR TABLE OF REV BY ENF 7

Statistic 8	DF 9	Value 10	Prob 11
Chi-Square	16	56.956	0.001
Likelihood Ratio Chi-Square	16	56.069	0.001
Mantel-Haenszel Chi-Square	1	0.004	0.952
Phi Coefficient		0.127	
Contingency Coefficient		0.126	
Cramer's V		0.064	

12 Sample Size = 3517

VI. Modèle log-linéaire

Il est possible d'utiliser l'information contenue dans un tableau de contingence pour estimer les probabilités conjointes associées aux différentes cellules. Considérons un tableau de trois variables dichotomiques y_1 , y_2 et y_3 , avec les probabilités théoriques associées notées P_{jkl} :

	$y_2 = 0$	$y_2 = 1$
$y_1 = 0$	$(y_3 = 0)P_{000}$ $(y_3 = 1)P_{001}$	$(y_3 = 0)P_{010}$ $(y_3 = 1)P_{011}$
$y_1 = 1$	$(y_3 = 0)P_{100}$ $(y_3 = 1)P_{101}$	$(y_3 = 0)P_{110}$ $(y_3 = 1)P_{111}$

Les probabilités P_{jkl} peuvent être estimées à partir des fréquences empiriques du tableau de contingence :

$$\hat{P}_{jkl} = \frac{n_{jkl}}{N} \quad j, k, l = 0, 1$$

Le modèle log-linéaire est un moyen très utile de modélisation de l'impact des différentes variables dichotomiques sur les probabilités associées aux cellules (McCullagh et Nelder, 1983). On spécifie pour cela une décomposition en effets marginaux et croisés entre les différentes variables, et on modélise la relation entre le logarithme de la probabilité conjointe P_{jkl} et les effets marginaux et croisés associés aux différentes modalités comme suit :

$$\begin{aligned}\log(P_{jkl}) = & \mu_0 + \mu_1(y_1) + \mu_2(y_2) + \mu_3(y_3) + \mu_{12}(y_1y_2) + \mu_{13}(y_1y_3) \\ & + \mu_{23}(y_2y_3) + \mu_{123}(y_1y_2y_3)\end{aligned}$$

Les termes $\mu_j, j = 1, 2, 3$ représentent les *effets marginaux*, c'est-à-dire les impacts des variables dichotomiques sur la probabilité conjointe. Les *effets croisés* entre les variables sont captés par les termes μ_{jk} et μ_{jkl} . μ_0 est appelé *l'effet principal*.

Les contraintes sur les P_{jkl} pour $j, k, l = 0, 1$ sont :

$$P_{jkl} \geq 0, \quad \sum_{j=0,1} \sum_{k=0,1} \sum_{l=0,1} P_{jkl} = 1$$

Afin d'assurer une représentation unique du système, on impose la contrainte suivante au système :

Résultat

Représentation unique d'un modèle log-linéaire : pour que le modèle log-linéaire admette une représentation unique, il faut que les effets marginaux et croisés soient, en moyenne, nuls :

$$\left\{ \begin{array}{l} \mu_j(y_j = 1) + \mu_j(y_j = 0) = 0 \forall j \\ \mu_{jk}(y_j y_k | y_j = 1) + \mu_{jk}(y_j y_k | y_j = 0) = 0 \forall k \\ \mu_{jk}(y_j y_k | y_k = 1) + \mu_{jk}(y_j y_k | y_k = 0) = 0 \forall j \\ \mu_{jkl}(y_j y_k y_l | y_j = 1) + \mu_{jkl}(y_j y_k y_l | y_j = 0) = 0 \forall k, l \\ \mu_{jkl}(y_j y_k y_l | y_k = 1) + \mu_{jkl}(y_j y_k y_l | y_k = 0) = 0 \forall j, l \\ \mu_{jkl}(y_j y_k y_l | y_l = 1) + \mu_{jkl}(y_j y_k y_l | y_l = 0) = 0 \forall j, k \end{array} \right.$$

Une façon simple de réécrire le système est alors de poser la forme suivante pour les effets :

$$\boxed{\begin{aligned} \mu_j(y_j) &= \mu_j \times y_j - \mu_j \times (1 - y_j) = \mu_j(2y_j - 1) \\ \mu_{jk}(y_j y_k) &= \mu_{jk}(2y_j - 1)(2y_k - 1) \\ \mu_{jkl}(y_j y_k y_l) &= \mu_{jkl}(2y_j - 1)(2y_k - 1)(2y_l - 1) \end{aligned}}$$

L'effet marginal sera donc égal à μ_j si $y_j = 1$ et à $-\mu_j$ si $y_j = 0$. Les effets croisés d'ordre 2 seront positifs si $y_j = y_k$ et négatifs sinon. Enfin, l'effet croisé d'ordre 3 sera positif si $y_j = y_k = y_l = 1$ ou $y_j = y_k = 0$ et $y_l = 1$, et négatif sinon⁴.

Par exemple, dans le cas de trois variables dichotomiques, le système devient :

$$\left\{ \begin{array}{l} \log(P_{000}) = \mu_0 - \mu_1 - \mu_2 - \mu_3 + \mu_{12} + \mu_{13} + \mu_{23} - \mu_{123} \\ \log(P_{001}) = \mu_0 - \mu_1 - \mu_2 + \mu_3 + \mu_{12} - \mu_{13} - \mu_{23} + \mu_{123} \\ \log(P_{010}) = \mu_0 - \mu_1 + \mu_2 - \mu_3 - \mu_{12} + \mu_{13} - \mu_{23} + \mu_{123} \\ \log(P_{100}) = \mu_0 + \mu_1 - \mu_2 - \mu_3 - \mu_{12} - \mu_{13} + \mu_{23} + \mu_{123} \\ \log(P_{101}) = \mu_0 + \mu_1 - \mu_2 + \mu_3 - \mu_{12} + \mu_{13} - \mu_{23} - \mu_{123} \\ \log(P_{110}) = \mu_0 + \mu_1 + \mu_2 - \mu_3 + \mu_{12} - \mu_{13} - \mu_{23} - \mu_{123} \\ \log(P_{011}) = \mu_0 - \mu_1 + \mu_2 + \mu_3 - \mu_{12} - \mu_{13} + \mu_{23} - \mu_{123} \\ \log(P_{111}) = \mu_0 + \mu_1 + \mu_2 + \mu_3 + \mu_{12} + \mu_{13} + \mu_{23} + \mu_{123} \end{array} \right.$$

Le modèle ci-dessus est appelé *modèle saturé*, en ce sens qu'aucune autre contrainte n'est imposée sur les paramètres. Le nombre de paramètres du modèle est égal au nombre de probabilités conjointes. Dans le cas d'un modèle avec trois variables binaires, on a 27 effets (y compris μ_0) pour 8 probabilités conjointes. Le nombre de contraintes imposées est donc égal à $27 - 8 = 19$.

4. Notons que l'on peut remplacer le terme $(2y_j - 1)$ par $sgn(y_j)$ dans les expressions où $sgn(X) = +1$ si X est positif, et $sgn(X) = -1$ si X est négatif.

Le modèle log-linéaire permet de tester l'indépendance entre les variables qualitatives, en étudiant la significativité des paramètres μ (Gouriéroux, 1989 ; Goodman, 1986). Toujours avec notre exemple à trois variables dichotomiques, on a les résultats suivants sur la relation entre indépendance des variables qualitatives et paramètres :

Définition

Indépendance dans un modèle log-linéaire : les variables y_1 , y_2 et y_3 sont indépendantes si :

$$\mu_{12} = \mu_{13} = \mu_{23} = \mu_{123} = 0$$

Les variables y_2 et y_3 sont indépendantes, conditionnellement à y_1 , si :

$$\mu_{23} = \mu_{123} = 0$$

La démarche d'estimation est la suivante. Dans une première étape, on estime le modèle saturé sur les N observations. Dans ce cas, une solution analytique existe et correspond à celle de l'estimateur du maximum de vraisemblance que l'on verra au chapitre 2. On élimine ensuite successivement les termes d'interaction et l'on teste les restrictions paramétriques correspondantes. Il est important de conserver les termes d'ordre « inférieur » associés à des variables si les termes d'ordre « supérieur » sont conservés. Par exemple, si on conserve l'interaction y_1 et y_2 , mais pas celle entre y_1 et y_3 , le paramètre μ_{12} sera estimé et μ_{13} sera contraint à 0. Mais les paramètres μ_1 , μ_2 et μ_3 seront tous conservés. Pour les modèles non saturés, une solution analytique, comme celle que nous avons plus haut, n'existe pas et il faut recourir à un algorithme itératif de résolution.

VII. Application numérique

On considère un tableau de contingence à trois entrées, dont les effectifs sont : (32, 86, 11, 35, 61, 73, 41, 70). Le modèle saturé est d'abord estimé par la procédure log-linéaire LLEST du logiciel GAUSS. Il faut fournir au logiciel le vecteur des effectifs (ici, un vecteur 8×1) et une matrice de spécification (*design matrix*) indiquant les contraintes imposées aux effets. Cette matrice a autant de lignes que de cellules (*cells*) et, pour le modèle saturé, autant de colonnes que de paramètres (c'est-à-dire d'effets). Comme dans la spécification donnée plus haut, on place un 1 ou un -1 selon les cas. Enfin, il faut fournir une matrice d'indices au logiciel. Cette matrice contiendra autant de lignes que de cellules et, en colonnes, les valeurs des variables dichotomiques (0 ou 1).

La matrice des indices donne les combinaisons dans le même ordre que le vecteur des effectifs ; elle est donc utile pour vérifier que le modèle a bien été spécifié. De plus, la procédure permet de donner un poids différent aux diverses cellules (par défaut, ce poids vaut 1). Les commandes GAUSS sont les suivantes. On note MATDES la matrice de spécification, CELLCOUNT le vecteur des effectifs, INDEX la matrice des indices, et LBL le vecteur des noms donnés aux paramètres.

```
library loglin;
#include loglin.ext;
llset;
output file=loglin1.out reset;
cellcount={32,86,11,35,61,73,41,70};
lbl={"mu0","mu1","mu2","mu3","mu12","mu13","mu23","mu123"};
{ m,b,v,gsq,df,tol } = lllest(cellcount,0,matdes);
call llout(index,m,b,v,gsq,df,tol,cellcount,0,matdes,0,0,lbl);
```

Les résultats sont les suivants :

LOGLIN Version 3.1.2 8/26/99 4:42 pm

DESIGN MATRIX :

Parameter # :

Cell	1	2	3	4	5	6	7	8
1	1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	1	1	-1	-1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	-1	-1	-1	-1	1	1
5	1	1	1	-1	1	-1	-1	1
6	1	1	-1	1	-1	-1	-1	1
7	1	-1	1	1	-1	-1	1	1
8	1	1	1	1	1	1	1	-1

PARAMETER ESTIMATES :

Parameter	Coeff.	Std. Error	z-value	p> z
mu0	1	3.831785	0.127666	30.01 0.00
mu1	2	0.156134	0.078461	1.99 0.05
mu2	3	-0.046312	0.071144	-0.65 0.52
mu3	4	0.281558	0.067804	4.15 0.00
mu12	5	0.238078	0.111748	2.13 0.03
mu13	6	-0.462278	0.240378	-1.92 0.05
mu23	7	-0.085997	0.123146	-0.70 0.48
mu123	8	-0.335528	0.348773	-0.96 0.34

TABLE OF OBSERVED, EXPECTED COUNTS, RESIDUALS AND WEIGHTS : ⑩

Cell	Index			OBS N	%	EXP N	%	RESIDUAL	STD RES	CELL WEIGHT
	④	⑤	⑥							
1	0	0	0	32.00	0.08	32.00	0.08	0.000	0.000	1.00
2	0	0	1	86.00	0.21	86.00	0.21	0.000	0.000	1.00
3	0	1	0	11.00	0.03	11.00	0.03	-0.000	-0.000	1.00
4	1	0	0	35.00	0.09	35.00	0.09	-0.000	-0.000	1.00
5	1	1	0	61.00	0.15	61.00	0.15	-0.000	-0.000	1.00
6	1	0	1	73.00	0.18	73.00	0.18	0.000	0.000	1.00
7	0	1	1	41.00	0.10	41.00	0.10	0.000	0.000	1.00
8	1	1	1	70.00	0.17	70.00	0.17	0.000	0.000	1.00

Number of cases : ⑬ 409.00

GOODNESS OF FIT STATISTICS (tolerance of 0.0000 achieved) : ⑭

Likelihood Ratio Chi-square : 0.0000 df : 0 Prob : 1.000

Pearson's Chi-square : 0.0000 df : 0 Prob : 1.000

Le logiciel commence par rappeler l'expression de la matrice de spécification (*design matrix* ⑮), puis fournit les estimations des paramètres (par maximum de vraisemblance, voir le chapitre 2). Le tableau des paramètres estimés ⑯ présente, dans l'ordre, le nom du paramètre (*Parameter*, donné par le vecteur LBL, voir plus haut), son numéro, la valeur estimée (*Coeff.*), son écart type (*Std. Error*), la statistique de Student (*z-value*) et la p-value ($p > |z|$). Cette dernière information est associée à la statistique de Student ; elle indique la probabilité qu'une variable normale centrée réduite soit supérieure en valeur absolue au t de Student calculé. Ainsi, si on se fixe un seuil d'erreur de 5 %, on rejette l'hypothèse de nullité du paramètre si cette p-value est inférieure à 0,05. Si cette dernière est supérieure au niveau de risque fixé, on ne peut rejeter l'hypothèse que le paramètre est nul.

La procédure donne ensuite les effectifs observés et prédits par le modèle dans le tableau ⑰ (tableau des effectifs observés et prédits, des résidus et des poids). Les colonnes indiquent la cellule ④, l'indice ⑮ (les variables ont par défaut les noms A, B et C), l'effectif observé ⑯, la fréquence ⑯. Les colonnes ⑯ et ⑯ indiquent l'effectif et la fréquence prédits par le modèle. Les deux colonnes ⑯ et ⑯ présentent les résidus bruts et standardisés (différence entre l'effectif observé et prédit, divisée par son écart type). La valeur du poids de la cellule, dans ⑯, est égale à 1 par défaut. Dans le cas du modèle saturé, les fréquences estimées correspondent aux fréquences empiriques n_{jkl}/N ; les effectifs sont donc estimés sans erreur, et les résidus sont tous nuls. ⑯ indique le nombre de cas retenus.

La dernière table ⑯ est relative aux statistiques d'ajustement du modèle. Nous introduirons les procédures de test basées sur la vraisemblance au cha-

pitre 2. Mentionnons simplement que le logiciel fournit deux statistiques de test de la spécification choisie (traduite par l'écriture de la matrice MATDES). Ces statistiques suivent une loi du χ^2 sous l'hypothèse nulle dépendant des restrictions imposées. Comme dans ce premier exemple, le modèle est saturé, le nombre de contraintes imposées est égal à 0 (on a 0 degrés de liberté, df) et donc les deux statistiques du χ^2 également. Comme décrit plus haut, on peut imposer des restrictions sur les paramètres du modèle, et les statistiques de test permettront de déduire si les contraintes sont valides ou non. Il est important de noter que les restrictions seront évaluées en comparant le modèle restreint (avec les contraintes) au modèle non contraint (le modèle saturé). Le chapitre 2 fournit les différentes statistiques de test et leur interprétation. La dernière colonne donne la probabilité que la distribution du χ^2 théorique dépasse la valeur de la statistique (1 dans ce cas). Cette expression est la p-value associée à la statistique de test : comme plus haut, on rejettéra l'hypothèse nulle (ici, que le modèle contraint n'est pas différent du modèle saturé) si cette p-value est inférieure au niveau de risque choisi (0,05 par exemple).

Au vu des résultats du modèle saturé, les paramètres μ_{23} et μ_{123} ne sont pas significativement différents de 0. On précise donc ces résultats en imposant la restriction : $\mu_{23} = \mu_{123} = 0$ et en réestimant le modèle. Les nouveaux résultats d'estimation figurent ci-dessous. Les commandes utilisées pour restreindre le modèle sont très simples ; il suffit de ne prendre qu'une partie de la matrice de spécification (enlever les deux dernières colonnes) et limiter de même le vecteur des noms des paramètres :

```
matdes=matdes[,1 : 4];
lbl=lbl[1 : 4];
```

PARAMETER ESTIMATES :

Parameter		Coeff.	Std. Error	z-value	p> z
mu0	1	3.729601	0.063717	58.53	0.00
mu1	2	0.210322	0.053454	3.93	0.00
mu2	3	-0.085625	0.055542	-1.54	0.12
mu3	4	0.322108	0.053163	6.06	0.00
mu12	5	0.324095	0.055542	5.84	0.00
mu13	6	-0.236527	0.062462	-3.79	0.00

TABLE OF OBSERVED, EXPECTED COUNTS, RESIDUALS AND WEIGHTS :

Cell	Index			OBS N	%	EXP N	%	RESIDUAL	STD RES	CELL WEIGHT
	A	B	C							
1	0	0	0	32.00	0.08	29.09	0.07	2.911	0.540	1.00
2	0	0	1	86.00	0.21	88.91	0.22	-2.911	-0.309	1.00
3	0	1	0	11.00	0.03	12.82	0.03	-1.819	-0.508	1.00

3	0	1	0	11.00	0.03	12.82	0.03	-1.819	-0.508	1.00
4	1	0	0	35.00	0.09	37.18	0.09	-2.184	-0.358	1.00
5	1	1	0	61.00	0.15	59.91	0.15	1.092	0.141	1.00
6	1	0	1	73.00	0.18	70.82	0.17	2.184	0.260	1.00
7	0	1	1	41.00	0.10	39.18	0.10	1.819	0.291	1.00
8	1	1	1	70.00	0.17	71.09	0.17	-1.092	-0.130	1.00

Number of cases : 409.00

GOODNESS OF FIT STATISTICS (tolerance of 0.0000 achieved) :

Likelihood Ratio	Chi-square :	0.9669	df : 2	Prob :	0.617
Pearson's	Chi-square :	0.9614	df : 2	Prob :	0.618

On voit que les restrictions imposées permettent d'améliorer significativement les estimations des paramètres. Les statistiques de Student sont nettement supérieures à celles du modèle saturé, avec des différences raisonnables entre effectifs observés et prédicts. Les statistiques de test du dernier tableau indiquent que les deux contraintes ne sont pas rejetées, la probabilité de dépasser la valeur de la statistique de test étant supérieure à 0,05.

VIII. Le χ^2 minimum

On dispose souvent, en pratique, de données dites *groupées*, caractérisées par le fait que les individus peuvent être classés en plusieurs catégories selon un ensemble de variables qualitatives. Pour chacune de ces catégories, les caractéristiques des individus sont identiques et l'on observe une variable binaire par individu, ce qui permet de calculer la proportion, dans chaque cellule, de cas pour lesquels la variable binaire vaut 1 à partir de l'effectif de la catégorie (Ashford et Sowden, 1970 ; Amemiya, 1985). Ce modèle est donc très proche du modèle log-linéaire, car il revient à expliquer la probabilité, approchée par la fréquence empirique, que la variable binaire principale vaut 1, mais cette fois en fonction de variables prédéterminées x . Nous prenons le cas simple d'une seule variable binaire, la généralisation au cas de plusieurs variables dichotomiques étant aisée (voir Maddala, 1983).

Soit j l'indice de la catégorie (cellule), $j = 1, 2, \dots, m$, et x_j , le vecteur $1 \times K$ des caractéristiques spécifiques aux individus de la cellule. On dispose de N_j observations sur cette dernière, et on note $y_{ij} = \{0, 1\}$ la variable binaire d'intérêt, pour l'individu i dans la cellule j . Sur l'effectif N_j , l'événement $\{y_{ij} = 1\}$ se produit pour m_j individus. Pour les $N_j - m_j$ autres, on a $y_{ij} = 0$. La fréquence

empirique (ou proportion) de $y_{ij} = 1$ pour la catégorie j est donc :

$$\hat{P}_j = \frac{m_j}{N_j} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij} \quad j = 1, \dots, m$$

On suppose alors que la probabilité théorique, notée P_j , dépend linéairement du vecteur x_j via un vecteur de paramètres β de dimension $(K \times 1)$. Cette probabilité théorique peut être reliée à la fréquence empirique par :

$$P_j = x_j \beta$$

soit :

$$\hat{P}_j = x_j \beta + u_j$$

où :

$$u_j = \hat{P}_j - P_j$$

Pour N_j assez grand, on a $\hat{P}_j \approx P_j$ et $E(u_j) \approx 0$. La variance de la proportion vaut $Var(\hat{P}_j) = P_j(1 - P_j)/N_j$, qui peut être estimée par $\hat{P}_j(1 - \hat{P}_j)/N_j$. La méthode du χ^2 consiste à estimer les paramètres β par une méthode de moindres carrés, en régressant \hat{P}_j sur x_j . Comme l'erreur u_j du modèle est hétérosclélastique (de variance non constante) par construction, on utilise une méthode robuste à l'hétérosclélasticité. (Voir le chapitre 2 sur un exposé de ces méthodes.) Le plus simple est de recourir aux *moindres carrés pondérés*, consistant à minimiser la somme des carrés suivante⁵ :

$$\frac{1}{m} \sum_{j=1}^m w_j (\hat{P}_j - x_j \beta)^2$$

où w_j est le poids égal à l'inverse de la variance de u_j . On a donc :

$$w_j = \left[\frac{N_j}{\hat{P}_j(1 - \hat{P}_j)} \right]$$

Ce modèle est le plus souvent utilisé sous forme log-linéaire ou non linéaire. Dans le cas du modèle log-linéaire, on prend :

$$\log \hat{P}_j = x_j \beta + u_j \quad \text{avec } u_j = \log \hat{P}_j - \log P_j$$

5. L'appellation de χ^2 minimum vient du fait que le critère à minimiser, évalué en $\hat{\beta}$, suit une loi du χ^2 .

Par un développement limité autour de P_j , on a :

$$\log \hat{P}_j = \log P_j + (\hat{P}_j - P_j) \frac{1}{P_j} + \dots$$

et donc :

$$u_j \approx \frac{(\hat{P}_j - P_j)}{P_j}$$

Une forme de modèle non linéaire également très utilisée est :

$$\log \left(\frac{P_j}{1 - P_j} \right) = x_j \beta$$

qui revient à poser la forme suivante pour la probabilité :

$$P_j = \frac{1}{1 + \exp(-x_j \beta)}$$

(Voir l'exercice 4 pour le calcul des variances.)

Berkson (1953) a proposé une méthode générale d'estimation de modèles avec la méthode du χ^2 minimum. Le modèle général est $P_j = F(x_j \beta)$, où $F(\cdot)$ est une fonction continue, comprise entre 0 et 1, dont les dérivées première et seconde existent, et $f(x) = \partial F(x)/\partial x > 0, \forall x$. Avec ces hypothèses, on peut inverser la relation et effectuer un développement limité de $F^{-1}(\hat{P}_j)$ autour de P_j :

$$F^{-1}(\hat{P}_j) = x_j \beta + \frac{\partial F^{-1}}{\partial P_j}|_{P_j} (\hat{P}_j - P_j) = x_j \beta + v_j + w_j$$

où P_j^* est entre \hat{P}_j et P_j , $v_j = (\hat{P}_j - P_j)/f[F^{-1}(P_j)]$

et :

$$w_j = (\hat{P}_j - P_j)/[f[F^{-1}(\hat{P}_j)] - f[F^{-1}(P_j)]]$$

Le terme w_j peut être ignoré pour des N_j grands, et la variance de v_j vaut :

$$Var(v_j) = \sigma_j^2 = \frac{P_j(1 - P_j)}{N_j f^2[F^{-1}(P_j)]}$$

L'estimateur du χ^2 minimum est alors :

$$\hat{\beta} = \left(\sum_{j=1}^m \frac{x_j' x_j}{\hat{\sigma}_j^2} \right)^{-1} \times \left(\sum_{j=1}^m \frac{x_j' F^{-1}(\hat{P}_j)}{\hat{\sigma}_j^2} \right)$$

où $\hat{\sigma}_j^2$ est la variance calculée en remplaçant P_j par \hat{P}_j . (Voir Amemiya, 1985, pour les propriétés asymptotiques de cet estimateur.)

IX. Modèle à probabilité linéaire

On cherche maintenant à estimer la valeur de la variable binaire par une méthode de moindres carrés sur des données non groupées. Le modèle est, pour un échantillon de taille N :

$$y_i = x_i\beta + u_i \quad i = 1, 2, \dots, N$$

où y_i est une variable binaire $\{0, 1\}$, x_i un vecteur $(1 \times K)$ de variables explicatives pour l'individu i associées au vecteur de paramètres β ($K \times 1$). L'hypothèse derrière ce modèle est que l'espérance conditionnelle $E(y_i|x_i) = x_i\beta$, et donc la fréquence ou probabilité conditionnelle, est une fonction linéaire des x_i . Ce modèle, appelé *modèle à probabilité linéaire* ou *modèle LP (Linear Probability)*, possède certaines propriétés le rendant peu adapté à la modélisation d'une probabilité.

Tout d'abord, notons que rien dans le modèle ne contraint la probabilité estimée à être comprise entre 0 et 1. Cette propriété est gênante lorsque le modèle LP est utilisé pour réaliser des prévisions.

Une autre propriété de ce modèle est que les résidus de la régression ne peuvent prendre que deux valeurs. Si $y_i = 1$, alors $u_i = 1 - x_i\beta$ avec probabilité $x_i\beta$, car d'après le modèle, $Prob(y_i = 1|x_i) = E(y_i = 1|x_i) = x_i\beta$. Si $y_i = 0$, on a $u_i = -x_i\beta$ avec la probabilité complémentaire, $1 - x_i\beta$. Le modèle à probabilité linéaire représente un processus binomial dont les probabilités sont modifiées à chaque tirage (c'est-à-dire pour chaque individu de l'échantillon). Le cadre d'analyse des moindres carrés dans lequel les résidus de la régression sont des variables aléatoires continues se prête mal, par conséquent, à la modélisation d'une variable dépendante binaire. Comme $E(u_i) = 0$ par définition, la variance de u_i est :

$$\begin{aligned} Var(u_i) &= (u_i^2|y_i = 1) \times Prob(y_i = 1) + (u_i^2|y_i = 0) \times Prob(y_i = 0) \\ &= (1 - x_i\beta)^2 \times (x_i\beta) + (-x_i\beta)^2 \times (1 - x_i\beta) \\ &= x_i\beta(1 - x_i\beta) \end{aligned}$$

Le modèle est donc hétéroscélastique par construction, la variance de l'erreur dépendant de x_i . Une méthode permettant de corriger l'hétéroscéasticité doit être utilisée, comme dans le cas des modèles pour données groupées de la section précédente.

À retenir

On a proposé dans ce chapitre une introduction aux variables qualitatives sous forme de classification de ces variables selon leur nature (dichotomique ou polytomique, ordonnée, séquentielle ou non ordonnée, catégorielle). Ce chapitre a ensuite introduit des éléments d'analyse statistique permettant d'évaluer l'indépendance des modalités associées à des variables qualitatives (tableaux de contingence), et l'influence conjointe de ces variables dans des modèles de probabilité (modèles log-linéaires). L'approche descriptive sert souvent de première analyse avant d'envisager des modèles dans lesquels les variables qualitatives sont expliquées. Dans le cas des données groupées, on utilise des méthodes de χ^2 minimum pour expliquer les proportions des modalités en fonction de caractéristiques identiques dans chaque cellule. On a introduit enfin, dans le cas général, le modèle à probabilité linéaire, caractérisé par des inconvénients que les modèles des chapitres suivants vont pallier.

Exercices

Exercice n° 1 : MCO et variable indicatrice

On considère un échantillon de 200 observations formées de couples (x_i, y_i) où x_i est une variable dichotomique (indicatrice). Le modèle envisagé est $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, où $E(\varepsilon_i) = 0$. Si X est une matrice formée d'une colonne de 1 et de x , les matrices $(X'X)^{-1}$ et $(X'Y)$ sont respectivement :

$$(X'X)^{-1} = \begin{bmatrix} 0,008\,928 & -0,008\,928 \\ -0,008\,928 & 0,020\,292 \end{bmatrix} \quad \text{et } (X'Y) = \begin{bmatrix} 441,744\,25 \\ 218,213\,07 \end{bmatrix}$$

- 1) Calculer le vecteur des paramètres estimés $\hat{\beta}$ par moindres carrés ordinaires. Si les statistiques de Student pour $\hat{\beta}_0$ et $\hat{\beta}_1$ sont respectivement 65,69 et 10,56 que vaut la variance des résidus notée σ^2 ?
- 2) À partir de ces estimations et du fait que $E(x_i y_i) = 1,0910$ calculer :
 - a) la moyenne de y_i sur le sous-échantillon où $x_i = 0$;
 - b) la moyenne des x_i sur l'ensemble de l'échantillon.

Exercice n° 2 : dépollution industrielle

Soit une population de 297 industriels équipés d'un dispositif de dépollution de leurs effluents. On observe δ , le taux de dépollution obtenu (entre 0 et 1), comme le rapport entre la pollution retirée et la pollution brute. Les effectifs des industriels sont représentés dans le tableau de contingence suivant, en fonction de leur branche d'activité et de l'intervalle dans lequel se situe δ :

δ	Secteur					Total
	Agroalimentaire (1)	Lait (2)	Chimie (3)	Métaux (4)	Papier (5)	
< 0,25	6	7	28	8	11	60
[0,25 ; 0,5]	11	11	24	4	8	58
[0,5 ; 0,75]	26	20	15	7	52	120
> 0,75	7	3	36	11	2	59
Total	50	41	103	30	73	297

1) Construire le tableau des fréquences empiriques. Calculer les profils-lignes et les profils-colonnes. Quel est alors le secteur le plus important, en termes relatifs, dans la catégorie $\delta > 0,75$? Quelle est la catégorie pour δ la plus importante, toujours en termes relatifs, pour le secteur agroalimentaire?

2) Calculer la statistique de test du χ^2 . Quel est le nombre de degrés de liberté du test? L'appartenance à un secteur d'activité donné a-t-elle une influence sur les performances en dépollution?

Exercice n° 3

Reprendre l'exercice précédent, avec le nouveau tableau de contingence où maintenant on a regroupé l'agroalimentaire avec le lait (1+2), les métaux avec la chimie (3+4), et pour les taux de dépollution, on considère trois catégories :

δ	Secteur			Total
	1+2	3+4	5	
< 0,5	70	12	36	118
[0,5 ; 0,75]	35	7	78	120
> 0,75	39	11	9	59
Total	144	30	123	297

1) Recalculer les profils-lignes, les profils-colonnes, ainsi que la statistique de test du χ^2 . Les résultats changent-ils par rapport à l'exercice précédent?

2) Considérons un regroupement supplémentaire où $\delta \leq 0,5$ ou $\delta > 0,5$; pour les secteurs, on regroupe 1+2 d'une part, 3+4+5 d'autre part. Le nouveau tableau de contingence est le suivant :

δ	Secteur		Total
	1+2	3+4+5	
$\delta \leq 0,5$	81	97	178
$\delta > 0,5$	93	26	119
Total	174	123	297

Écrire le tableau des fréquences empiriques et le modèle log-linéaire correspondant, avec les deux variables dichotomiques.

3) Résoudre le système, en exprimant les effets principal, marginaux et croisés en fonction des moyennes des logarithmes des fréquences empiriques. Les effets seront notés comme

dans le texte, $\mu_0, \mu_1, \mu_2, \mu_{12}$ et on utilisera les notations suivantes :

$$\bar{P}_{..} = \frac{1}{4} \sum_{j=0,1} \sum_{k=0,1} \log(P_{jk}), \bar{P}_{1.} = \frac{1}{2} \sum_{k=0,1} \log(P_{1k}), \bar{P}_{.1} = \frac{1}{2} \sum_{j=0,1} \log(P_{j1})$$

et ainsi de suite. La position d'un point (.) indique que l'on prend la moyenne sur les modalités de la variable ainsi repérée. De même, un 1 indique que l'on ne conserve pour la variable que le cas où elle vaut 1.

Exercice n° 4 : modèles pour données groupées

On considère les deux variantes du modèle à la base de la méthode du χ^2 minimum : $\log P_j = x_j \beta$ (modèle A) et $\log(P_j/(1 - P_j)) = x_j \beta$ (modèle B).

- 1) Pour le modèle A, calculer l'expression de la variance des résidus et donner la forme du poids optimal correspondant. Utiliser un développement limité comme dans le texte. Attention : dans le calcul de la variance, seul \hat{P}_j est aléatoire, et P_j est pris comme une constante.
- 2) Même question qu'en 1) pour le modèle B.



2. Méthodes d'estimation

Ce chapitre présente les rappels théoriques nécessaires à la compréhension des méthodes économétriques d'estimation et de test utilisées dans les autres chapitres. Nous verrons que la méthode du maximum de vraisemblance permet de mener facilement une inférence statistique sur des modèles très généraux. Une application est le modèle de comptage de Poisson. Les estimateurs obtenus en maximisant la vraisemblance sont efficaces, alors que ceux basés sur la méthode des moments sont en général robustes. Le chapitre décrit les propriétés et les techniques d'estimation utilisées pour les deux classes d'estimateurs, et se termine par un aperçu des algorithmes de maximisation numérique.



Objectifs du chapitre : détailler le principe de l'estimation par maximum de vraisemblance et le comparer avec les estimateurs des moments ; dériver les statistiques de test usuelles ; décrire les algorithmes de maximisation numérique ; proposer comme application aux variables discrètes les modèles de comptage (Poisson).

Concepts clés étudiés : vraisemblance d'un échantillon, maximisation de la vraisemblance, estimateurs efficaces, modèles de comptage, méthode des moments, estimateurs robustes, tests d'hypothèses, maximisation numérique.

I. Le maximum de vraisemblance

Cette section traite de la vraisemblance d'un échantillon et décrit les estimateurs maximisant celle-ci comme efficaces. La dérivation des statistiques de test usuelles est exposée dans ce contexte.

A. Définition de la vraisemblance

La vraisemblance en économétrie est définie comme la probabilité d'observer un échantillon, étant donné les paramètres du processus ayant engendré les données. Soit un échantillon de taille N constitué de paires $\{y_i, x_i\}$, $i = 1, 2, \dots, N$, où y_i est une variable dépendante (scalaire) et x_i un vecteur $1 \times K^1$. Si les couples $\{y_i, x_i\}$ sont indépendamment distribués, la probabilité de l'échantillon est le produit des probabilités associées à chaque couple si le processus ayant engendré les données est à valeurs discrètes, ou le produit des densités si ce processus est à valeurs continues. Soit $f(y_i, x_i; \theta)$ la probabilité (ou la densité) conjointe associée au couple $\{y_i, x_i\}$, où $f(\cdot)$ est une fonction croissante et à valeurs positives, et θ un vecteur $G \times 1$ de paramètres. La vraisemblance est donnée par :

$$L(y, x, \theta) = \prod_{i=1}^N f(y_i, x_i; \theta)$$

On peut aussi décomposer la densité conjointe en le produit de la densité conditionnelle de y_i sachant x_i fois la densité marginale de x_i pour obtenir une écriture alternative : $L = \prod_{i=1}^N f(y_i|x_i; \theta)h(x_i)$. Le problème est alors le suivant : étant donné la forme fonctionnelle de $f(\cdot)$ et les N observations sur $\{y_i, x_i\}$, quelles valeurs des paramètres rendent l'observation de l'échantillon la plus vraisemblable ? Intuitivement, puisque l'on a effectivement observé ce dernier, c'est que sa probabilité d'apparition était forte. Si on maximise la vraisemblance associée à l'échantillon, on peut remonter au processus générateur des données. Comme ce dernier n'est défini que par la forme fonctionnelle adoptée et le vecteur de paramètres θ , il suffit d'estimer ce dernier. C'est le principe de la maximisation de la vraisemblance. Deux conditions sont nécessaires pour que ce principe ait un sens en pratique. Tout d'abord, il faut évidemment bien spécifier la forme fonctionnelle $f(\cdot)$. Ensuite, il faut s'assurer qu'une seule valeur de θ maximise la vraisemblance. Cette dernière condition est facilement obtenue par des conditions de régularité (Amemiya, 1985 ; Davidson et McKinnon, 1993). En revanche, rien n'indique *a priori* que la condition relative au bon choix de la fonction $f(\cdot)$ est satisfaite en pratique. Ainsi, il est fort possible que le vrai processus générateur des données ne corresponde pas à celui choisi par l'économétrie. Le θ estimé pourra alors être entaché d'un *biais de spécification*, et des méthodes alternatives devront être recherchées (voir les estimateurs des moments, plus loin dans ce chapitre). Notons qu'il est relativement équivalent de spécifier la distribution de la variable y_i ou du résidu, du moment qu'une relation déterministe existe entre les deux.

1. La généralisation au cas vectoriel de y est immédiate.

Soit ε_i le terme d'erreur du modèle, ce dernier peut s'écrire de façon générale comme $g(y_i, x_i, \varepsilon_i; \theta) = 0$, soit $y_i = g^{-1}(x_i, \varepsilon_i; \theta)$. Si ε_i admet une fonction de densité $f(\varepsilon_i)$, alors y_i admettra une densité de la forme :

$$f(y_i) = f[g^{-1}(\varepsilon_i)] \times \left| \frac{d\varepsilon_i}{dy_i} \right|$$

où $|d\varepsilon_i/dy_i|$ est le Jacobien de la transformation de ε_i en y_i . Ainsi, estimer un modèle de régression par maximum de vraisemblance revient à estimer les paramètres de la densité conditionnelle $f(y_i|x_i; \theta)$. Si $y_i = x_i\beta + \varepsilon_i$, et l'on suppose que ε_i suit une distribution normale $N(0, \sigma^2_\varepsilon)$, alors $y_i \sim N(x_i\beta, \sigma^2_\varepsilon)$. Les paramètres β et σ^2 fourniront toute l'information nécessaire pour l'inférence statistique sur le modèle.

B. Déivation et propriétés de l'estimateur

L'estimateur du maximum de vraisemblance (EMV, ou MLE, *Maximum Likelihood Estimator*) est défini par la solution du problème :

$$\max_{\theta} L(y, x, \theta) \equiv \max_{\theta} \log L(x, y, \theta) \Leftrightarrow \frac{\partial \log L(x, y, \theta)}{\partial \theta} = 0$$

Sous des hypothèses de régularité (voir Amemiya, 1985 ; Greene, 1997), l'estimateur EMV $\hat{\theta}$ a les propriétés suivantes :

- convergence : $\text{plim } \hat{\theta} = \theta$;
- invariance à une transformation paramétrique : si $\theta^* = t(\theta)$, alors $\hat{\theta}^* = t(\hat{\theta})$;
- efficacité asymptotique : $\hat{\theta}$ atteint la borne inférieure de Cramér-Rao ;
- normalité asymptotique : $\hat{\theta} \sim N(\theta, [-E\partial^2 \log L/\partial \theta \partial \theta']^{-1})$.

Notons que la variance de la condition du premier ordre donnée plus haut est :

$$\begin{aligned} \text{Var}\left(\frac{\partial \log L(x, y, \theta)}{\partial \theta}\right) &= E\left(\frac{\partial \log L(x, y, \theta)}{\partial \theta} \times \frac{\partial \log L(x, y, \theta)}{\partial \theta'}\right) \\ &= -E\left(\frac{\partial^2 \log L(x, y, \theta)}{\partial \theta \partial \theta'}\right) \end{aligned}$$

Soit $g(\theta)$ ($G \times 1$) et $H(\theta)$ ($G \times G$) le gradient et le Hessien associés au problème : $g(\theta) = \partial \log L/\partial \theta$, $H(\theta) = \partial^2 \log L/\partial \theta \partial \theta'$. On note θ_0 la vraie valeur du paramètre θ . De la condition du premier ordre du problème plus haut : $g(\theta) = 0$, on effectue un développement limité :

$$g(\hat{\theta}) = g(\theta_0) + H(\theta_0) \times (\hat{\theta} - \theta_0) + \text{termes d'ordre supérieur},$$

ou encore $(\hat{\theta} - \theta_0) = -[H(\theta_0)]^{-1} g(\hat{\theta})$, en négligeant les termes d'ordre supérieur², et puisque $g(\hat{\theta}) = 0$ par hypothèse. On peut encore écrire :

$$\sqrt{N}(\hat{\theta} - \theta_0) = -\left[\frac{1}{N}H(\theta_0)\right]^{-1}\left[\sqrt{N}g(\theta_0)\right]$$

où $\bar{g}(\theta)$ est un vecteur ($G \times 1$) formé des $(1/N) \sum_i \partial \log L(x_i, y_i, \theta_0) / \partial \theta$. En appliquant un théorème central-limite à $\sqrt{N}\bar{g}(\theta_0)$ et une loi des grands nombres pour $H(\theta_0)/N$, on montre que l'expression $-[\frac{1}{N}H(\theta)]^{-1} [\sqrt{N}\bar{g}(\theta_0)]$ admet une distribution normale de moyenne 0 et de matrice de variance-covariance $[-EH(\theta)/N]^{-1} \times [-EH(\theta)/N] \times [-EH(\theta)/N]^{-1}$, soit encore :

$$\boxed{\sqrt{N}(\hat{\theta} - \theta_0) \sim N(0, -[EH(\theta)/N]^{-1})}$$

ce qui montre la propriété de normalité asymptotique de l'estimateur EMV. La matrice de variance-covariance de $\hat{\theta}$ est souvent notée $I(\theta_0)^{-1}$ où $I(\theta_0) = -E(\partial^2 \log L(x, y, \theta_0) / \partial \theta \partial \theta')$ est la *matrice d'information*. En pratique, il est souvent impossible de caractériser cette expression à cause de l'opérateur espérance. On approche donc la matrice de variance-covariance par $[-1/N \sum_i \partial^2 \log L(x_i, y_i, \hat{\theta}) / \partial \theta \partial \theta']^{-1}$, c'est-à-dire en négligeant l'espérance et en remplaçant θ_0 par son estimateur, $\hat{\theta}$. Si les dérivées secondes de la log-vraisemblance sont difficiles à obtenir, on utilise la forme plus simple :

$$\boxed{Var(\hat{\theta}) = \left[\frac{1}{N} \sum_{i=1}^N \frac{\partial \log L(x_i, y_i, \hat{\theta})}{\partial \theta} \frac{\partial \log L(x_i, y_i, \hat{\theta})}{\partial \theta'} \right]^{-1}}$$

Cette écriture utilise le fait que la matrice d'information est égale à l'espérance du produit croisé des composantes du gradient (voir la variance de la condition du premier ordre plus haut, Berndt *et al.* 1974).

C. Vraisemblance concentrée

Il est possible de simplifier le problème de maximisation de la vraisemblance de la façon suivante. Supposons que l'on puisse décomposer le vecteur de paramètres en $\theta = (\theta_1, \theta_2)$, de sorte que :

$$\frac{\partial \log L}{\partial \theta} = 0 \Leftrightarrow \theta_2 = \psi(\theta_1)$$

2. Voir Greene (1997) ou Davidson et McKinnon (1993) pour une justification rigoureuse.

En d'autres termes, le système de G équations à G inconnues peut se résoudre partiellement par élimination des θ_2 . La nouvelle log-vraisemblance s'écrit alors : $\log L^* = \log L(\theta_1, \theta_2) = \log L(\theta_1, \psi(\theta_1))$, qui peut être maximisée par rapport au sous-vecteur θ_1 . Considérons l'exemple d'une vraisemblance déduite d'une distribution normale : $\varepsilon_i = y_i - x_i\beta \sim N(0, \sigma^2)$. Pour un échantillon de taille N , on a :

$$\log L = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \varepsilon_i^2$$

et de la condition du premier ordre $\partial \log L / \partial \sigma^2 = 0$, on a $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - x_i\beta)^2$, que l'on reporte dans l'expression de la log-vraisemblance. Celle-ci s'écrit maintenant en fonction de β seulement :

$$\log L^* = -\frac{N}{2} - \frac{N \log(2\pi)}{2} - \frac{N}{2} \log \left[\frac{1}{N} \sum_i^N (y_i - x_i\beta)^2 \right]$$

D. Inférence statistique

1) Test de spécification

Nous avons vu que la matrice d'information $I(\theta)$ admet deux écritures : $I(\theta) = -E(\partial^2 \log L / \partial \theta \partial \theta')$ et $I(\theta) = E(\partial \log L / \partial \theta \times \partial \log L / \partial \theta')$. Or, cette équivalence n'est vraie que si le modèle est bien spécifié (voir Amemiya, 1985). White (1982) propose un test de spécification du modèle, basé sur l'hypothèse nulle $E(\partial^2 \log L / \partial \theta \partial \theta') + E(\partial \log L / \partial \theta \times \partial \log L / \partial \theta') = 0$. Ce test est relativement complexe à mettre en oeuvre, aussi le résultat (voir plus haut) est surtout utilisé pour calculer une matrice de variance-covariance robuste à la mauvaise spécification du modèle. Sous l'hypothèse nulle que le modèle est bien spécifié, l'estimateur EMV est asymptotiquement normal, de moyenne θ et de matrice de variance-covariance $I(\theta)^{-1} = [-EH(\theta)]^{-1}$. En revanche, si le modèle est mal spécifié, on a toujours la variance-covariance qui vaut :

$$[-E\partial^2 \log L / \partial \theta \partial \theta']^{-1} \times [E(\partial \log L / \partial \theta \times \partial \log L / \partial \theta')] \times [-E\partial^2 \log L / \partial \theta \partial \theta']^{-1}$$

Cette dernière expression permet de fournir une inférence statistique correcte (en termes d'écart types des paramètres), même si le modèle est mal spécifié. Puisqu'elle reste valide si le modèle est bien spécifié, utiliser la matrice de variance-covariance robuste ci-dessus est donc recommandé³.

3. Cette procédure est similaire à celle utilisée dans l'estimation des modèles de régression avec hétéroscédasticité, et décrite à la section 2.III.C.

2) Critères d'information

Les critères d'information fournissent une mesure de la quantité d'information donnée par le modèle. Ils sont obtenus par combinaison entre la valeur de la log-vraisemblance, le nombre d'observations et de paramètres. Les critères d'Akaike, Schwarz et Hannan-Quinn sont donnés par :

Critère	Expression
Akaike (AIC)	$-2 \log(L)/N + 2G/N$
Schwarz (SC)	$-2 \log(L)/N + G \log(N)/N$
Hannan-Quinn (HQ)	$-2 \log(L)/N + 2G \log[\log(N)]/N$

où N est le nombre d'observations, G le nombre de paramètres du modèle. On préférera le modèle fournissant un critère minimal. Notons encore que ces critères dépendent de l'unité de mesure de la variable dépendante. Ils ne peuvent donc être utilisés pour comparer deux modèles où l'expression de la variable dépendante serait différente, par exemple y et $\log(y)$.

3) Tests d'hypothèses

Supposons que l'on cherche à évaluer la pertinence d'une hypothèse sur les paramètres estimés par maximum de vraisemblance. L'hypothèse nulle est transcrise en un ensemble de q restrictions $R(\theta) = 0$, où $R(\theta)$ est un vecteur de dimension $(q \times 1)^4$. Soit $\hat{\theta}$ le vecteur des paramètres estimés sous la contrainte ci-dessus. On cherche à vérifier l'hypothèse nulle H_0 en construisant une *statistique de test*, qui est une fonction aléatoire des données et des paramètres estimés, et en choisissant un niveau de risque de première espèce α^5 . Une fois démontrée la convergence asymptotique de la statistique de test *sous l'hypothèse nulle* H_0 vers une distribution connue, il suffit de comparer la valeur de la statistique avec la valeur théorique tabulée.

Les trois statistiques de test usuelles dans le contexte du maximum de vraisemblance sont le rapport des vraisemblances (LRT, *Likelihood Ratio Test*), le test de Wald et le test du multiplicateur de Lagrange (LM, *Lagrange Multiplier*). Ces trois statistiques sont asymptotiquement équivalentes, mais ont des comportements différents en petit échantillon. Toutes suivent asymptotiquement une distribution χ^2 avec comme degrés de liberté le nombre de contraintes imposées (q). La statistique du rapport des vraisemblances est :

$$2[\log L(\hat{\theta}) - \log L(\bar{\theta})]$$

4. Si l'on veut imposer par exemple des contraintes linéaires homogènes (nullité d'un paramètre, égalité de deux paramètres, etc.), on adoptera $R(\theta) = \bar{R}\theta$, où \bar{R} est une matrice $(q \times G)$.
5. Le risque de première espèce α est la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie. Le risque de deuxième espèce, notée β , mesure au contraire la probabilité de ne pas rejeter l'hypothèse nulle alors qu'elle est fausse. Les statisticiens étant des personnes prudentes, on ne parle jamais d'acceptation d'une hypothèse, mais plutôt de son non-rejet.

L'intuition de ce test est que, si H_0 est vraie, la probabilité d'observer l'échantillon sous la contrainte doit être très proche de la vraisemblance non-contrainte. La statistique de test de Wald est :

$$R(\hat{\theta})' \left\{ \text{Var}[R(\hat{\theta})] \right\}^{-1} R(\hat{\theta})$$

et s'interprète par la mesure de l'écart à 0 de la contrainte $R(\hat{\theta})$. On suppose que cette dernière est asymptotiquement normalement distribuée, de moyenne 0 et de variance-covariance $(\partial R(\hat{\theta})/\partial\theta)\text{var}(\hat{\theta})(\partial R(\hat{\theta})/\partial\theta)'$. Un résultat important en statistique est que, pour tout vecteur x ($K \times 1$) normalement distribué, de moyenne μ et de matrice de variance-covariance Σ , $(x - \mu)'\Sigma^{-1}(x - \mu)$ suit une distribution $\chi^2(K)$. Par conséquent, sous l'hypothèse nulle, $R(\hat{\theta})$ sera centrée sur 0 et le test de Wald évalue les déviations par rapport à la « moyenne » de $R(\theta)$. Un aspect intéressant de ce test est que, contrairement au rapport des vraisemblances, il ne nécessite que le calcul de l'estimateur sous la contrainte. Cela peut être utile en pratique si le modèle non-contraint est trop complexe. Le dernier test est celui du multiplicateur de Lagrange, dont l'expression est :

$$\left(\frac{\partial \log L(x, y, \bar{\theta})}{\partial \theta} \right)' [I(\bar{\theta})]^{-1} \left(\frac{\partial \log L(x, y, \bar{\theta})}{\partial \theta} \right)$$

L'idée sous-jacente de ce test est que l'imposition de la contrainte $R(\theta) = 0$ revient à maximiser la fonction objectif suivante (le Lagrangien du problème) :

$$\log L^* = \log L(\theta) + \lambda'R(\theta)$$

où λ est un vecteur $q \times 1$ de multiplicateurs de Lagrange associés aux lignes de R . De la maximisation de $\log L^*$ par rapport à θ on obtient :

$$-\frac{\partial R(\hat{\theta})}{\partial \theta}' \lambda = \frac{\partial \log L(\theta)}{\partial \theta}$$

Si H_0 est vraie, alors les multiplicateurs de Lagrange associés à $R(\theta) = 0$ seront petits et donc le gradient $\partial \log L(\theta)/\partial \theta$ également. Comme on a vu plus haut que la variance de ce dernier terme est la matrice d'information $I(\theta)$, il est aisé de construire la statistique de test à partir de la normalité asymptotique du gradient. Le test du multiplicateur de Lagrange est donc très proche dans son esprit de celui de Wald. Il ne requiert lui aussi que le calcul du modèle sous l'hypothèse nulle.

Un résultat très utile en pratique concerne la statistique LM. Celle-ci peut s'écrire en fonction de la matrice $G(\theta)$ ($N \times K$) contenant les dérivées de $\log L(\theta)$ pour chaque observation :

$$LM = e_N' G(\hat{\theta}) \left[G(\hat{\theta})' G(\hat{\theta}) \right]^{-1} G(\hat{\theta})' e_N$$

où $G_{ij} = \partial \log L(x_i, y_i, \hat{\theta}) / \partial \theta_j$, $\forall i = 1, 2, \dots, N$, $j = 1, 2, \dots, K$, et e_N est un vecteur ($N \times 1$) de 1. On a utilisé le résultat plus haut pour les écritures alternatives de la matrice d'information. La statistique de test vaut alors NR^2 , où R^2 est le coefficient de détermination de la régression de e_N sur les colonnes de $G(\hat{\theta})$. Cette dernière matrice est obtenue après convergence de l'estimateur du maximum de vraisemblance sous la contrainte.

E. Exemple : régression hétéroscléastique

Illustrons l'application du maximum de vraisemblance par le modèle de régression hétéroscléastique suivant :

$$y_i = x_i \beta + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \sigma_i^2 = \text{var}(\varepsilon_i) = \sigma_0^2 z_i^2$$

avec $i = 1, 2, \dots, N$, où σ_0^2 est une constante et z_i une variable exogène. La log-vraisemblance s'écrit :

$$\log L = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_0^2) - \frac{1}{2} \sum_i^N \log(z_i^2) - \frac{1}{2\sigma_0^2} \sum_i^N \frac{\varepsilon_i^2}{z_i^2}$$

Les conditions du premier ordre de la maximisation de $\log L$ sont :

$$\begin{cases} \partial \log L / \partial \beta = \sum_i^N x'_i \varepsilon_i / (\sigma_0^2 z_i^2) = 0 \\ \partial \log L / \partial \sigma_0^2 = -N / (2\sigma_0^2) + 1 / (2\sigma_0^4) \sum_i^N \varepsilon_i^2 / (z_i^2) = 0 \end{cases}$$

La matrice d'information est donnée par :

$$I(\theta) = -E\left(\frac{\partial^2 \log L}{\partial \theta \partial \theta'}\right) = \begin{bmatrix} (1/\sigma_0^2) \sum_i^N x'_i x_i / z_i^2 & 0 \\ 0 & N/2\sigma_0^4 \end{bmatrix}$$

où $\theta = (\beta, \sigma_0^2)'$. Notons que cette matrice est diagonale par blocs ; la covariance entre β et σ_0^2 , qui serait calculée à partir de $I(\theta)^{-1}$, est nulle. Par conséquent, il est possible d'estimer β et σ_0^2 séparément, ce qui est fait par exemple avec la technique de la vraisemblance concentrée vue plus haut.

II. Les modèles de comptage

On présente ici un bref aperçu de l'utilisation des méthodes décrites dans ce chapitre, pour les modèles de comptage. Dans ces derniers, la variable dépendante est à valeurs discrètes, mais non catégorielle (voir le chapitre 1) :

nombre d'accidents dans un département, nombre d'années d'études effectuées, etc. On a choisi de traiter ces modèles dans ce chapitre, car la nature des variables de comptage est différente de celles étudiées dans le reste de l'ouvrage (variables catégorielles). La spécification la plus courante est le modèle de Poisson, définissant la probabilité d'occurrence d'un événement issu d'un processus de comptage⁶.

A. Modèle de Poisson

Soit y_i la variable dépendante de comptage, $y_i = 0, 1, \dots$. La probabilité donnée par la distribution de Poisson que y_i soit égale à un entier r quelconque est :

$$\text{Prob}[y_i = r] = \exp(-\lambda) \frac{\lambda^r}{r!}$$

où λ est le paramètre de la distribution, tel que $E(y_i) = \text{var}(y_i) = \lambda$. Pour introduire des variables explicatives dans le modèle, on conditionne λ en imposant la forme log-linéaire suivante :

$$\log \lambda_i = x_i \beta \quad i = 1, 2, \dots, N$$

où x_i est un vecteur ($1 \times K$) associé au vecteur de paramètres β ($K \times 1$). Le paramètre associé au terme constant, $x_{i1} = 1 \forall i$, est noté β_1 . Cette spécification revient à supposer que les variables expliquées y_1, y_2, \dots, y_N sont distribuées selon des lois de Poisson indépendantes, de paramètres $\lambda_1, \lambda_2, \dots, \lambda_N$ respectivement :

$$\boxed{\text{Prob}[y_i = r] = \exp[-\lambda_i(x_i)] \frac{\lambda_i(x_i)^r}{r!}}$$

L'espérance conditionnelle de la variable dépendante est $E(y_i|x_i) = \lambda_i = \exp(x_i \beta)$, et les effets marginaux sont : $\partial E(y_i|x_i)/\partial x_{ij} = \lambda_i \beta_j$. Le modèle de comptage de Poisson peut *a priori* être estimé par une méthode des moments ou par moindres carrés non linéaires, (voir Greene 1997, chap.10). La vraisemblance du modèle de Poisson est :

$$L = \prod_{i=1}^N \left[\exp(-\lambda_i) \frac{\lambda_i^{y_i}}{y_i!} \right] = \prod_{i=1}^N \exp[-\lambda_i + y_i x_i \beta - \log(y_i!)]$$

$$\Leftrightarrow \log L = -\sum_{i=1}^N \lambda_i + \beta_1 \sum_{i=1}^N y_i + \sum_{j=2}^K \beta_j \sum_{i=1}^N y_i x_{ij} - \sum_{i=1}^N \log(y_i!)$$

6. Voir Cameron et Trivedi (1986) pour une exposition générale du modèle.

On utilise les moments empiriques : $\mu_j = \sum_{i=1}^N x_{ij} y_i$, $j = 1, 2, \dots, K$. En particulier, on a $\mu_1 = \sum_{i=1}^N y_i$. En égalisant à 0 les dérivées de $\log L$ par rapport à β , on a $\mu_1 = \sum_{i=1}^N \hat{\lambda}_i$, $\mu_j = \sum_{i=1}^N x_{ij} \hat{\lambda}_i$, $j = 2, 3, \dots, K$, où $\hat{\lambda}_i = \exp(\beta_1 + \sum_{j=2}^K \beta_j x_{ij})$. On résout ce système non linéaire de K équations à K inconnues itérativement, en utilisant la matrice des dérivées secondes :

$$\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^N x_{ij} x_{ik} \lambda_i \quad j, k = 1, 2, \dots, K$$

B. Extensions

En raison de l'hypothèse parfois peu réaliste que $E(y_i|x_i) = \text{var}(y_i|x_i)$ dans le modèle de Poisson, certains auteurs ont d'abord proposé un test de cette restriction, puis des spécifications moins contraignantes (Hausman *et al.*, 1984 ; McCullagh et Nelder, 1983 ; Cameron et Trivedi, 1986 ; Gurmu, 1991). Une procédure de test de la surdispersion (la variance est supérieure à la moyenne) basée sur des conditions de moments est décrite ci-après. (Voir la section 2.III.D sur la méthode des moments utilisée ici.) L'hypothèse nulle $E(y_i|x_i) = \text{var}(y_i|x_i)$ se transcrit en les conditions d'orthogonalité suivantes :

$$E \left\{ W_i' \left[(y_i - E(y_i|x_i))^2 - E(y_i|x_i) \right] \right\} = E \left\{ W_i' \left[(y_i - \lambda_i)^2 - \lambda_i \right] \right\} = 0$$

où W_i est un vecteur ($1 \times L$) d'instruments (on peut adopter $W_i = (x_{i1}, x_{i2}, \dots, x_{iK})$). On forme alors une statistique de test à partir des paramètres estimés par maximum de vraisemblance, $\hat{\beta}$:

$$\left(\sum_i^N W_i' (\varepsilon_i^2 - \hat{\lambda}_i) \right)' \Omega^{-1} \left(\sum_i^N W_i' (\varepsilon_i^2 - \hat{\lambda}_i) \right) \sim \chi^2(L - K)$$

où :

$$\Omega = \left[\sum_i^N W_i' W_i (\varepsilon_i^2 - \hat{\lambda}_i)^2 - \sum_i^N W_i' x_i \varepsilon_i (\varepsilon_i^2 - \hat{\lambda}_i) \right. \\ \left. \left(\sum_i^N x_i' x_i \varepsilon_i^2 \right)^{-1} \sum_i^N x_i' W_i \varepsilon_i (\varepsilon_i^2 - \hat{\lambda}_i) \right]$$

et :

$$\varepsilon_i = y_i - \hat{\lambda}_i$$

Si l'hypothèse sous-jacente au modèle de Poisson est rejetée par ce test, on peut utiliser une autre spécification, celle du modèle de régression de la loi

binomiale négative (ou modèle NegBin, *Negative Binomial Model*). Ce modèle généralise la forme log-linéaire utilisée ci-dessus de la façon suivante :

$$\log \nu_i = \log \lambda_i + \log u_i = x_i \beta + v_i$$

où v_i est une variable aléatoire captant l'hétérogénéité non observable dans les observations. Conditionnellement à x_i et u_i , la distribution de y_i est toujours une loi de Poisson :

$$Prob(y_i = r) = \exp(-\lambda_i u_i) \frac{(\lambda_i u_i)^r}{r!}$$

mais la densité de y_i sachant x_i est obtenue en prenant l'espérance de l'expression ci-dessus par rapport à la densité de u_i :

$$f(y_i|x_i) = \int_0^{+\infty} \exp(-\lambda_i u_i) \frac{(\lambda_i u_i)^r}{r!} g(u_i) du_i$$

Si l'on suppose que u_i suit une loi Gamma de moyenne 1 (la moyenne de cette loi n'est pas identifiée si le modèle contient un terme constant), on a :

$$f(y_i|x_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} z_i^{y_i} (1 - z_i)^\theta$$

où $z_i = \lambda_i / (\lambda_i + \theta)$ et θ est le paramètre de la distribution Gamma. L'espérance conditionnelle est toujours λ_i , mais la variance vaut $\lambda_i(1 + \lambda_i/\theta)$. Le modèle NegBin peut être estimé par maximum de vraisemblance et la spécification de Poisson facilement testée par l'hypothèse nulle $H_0 : \theta = 0$.

C. Application : choix d'irrigation

Le modèle de comptage de Poisson est appliqué à l'estimation du nombre de tours d'eau (nombre de passages) d'agriculteurs irriguants. L'échantillon est composé de 457 producteurs pour lesquels on observe :

- *NPAR* : nombre de tours d'eau sur la parcelle ;
- *OBJREND* : objectif de rendement (en quintaux/ha) ;
- *TEMMOY* : température moyenne de mai à octobre (en °C) ;
- *PLUMOY* : précipitations moyennes de mai à octobre (en mm) ;
- *VTEMMOY* : variance des températures de mai à octobre (en °C) ;
- *CEREAL* : 1, si une céréale a été plantée l'année précédente ;
- *RESIND* : 1, si le producteur dispose d'un dispositif d'irrigation individuel.

La probabilité associée au nombre de tours est estimée par le modèle de Poisson, sur le logiciel GAUSS (procédure Poisson). Les résultats sont les suivants :

Variable	Paramètre	Écart type	<i>t</i> de Student
Constante	-2,556 410	0,516 468	-4,949 794
<i>OBJREND</i>	0,021 754	0,002 478	8,777 816
<i>TEMMOY</i>	0,137 823	0,026 759	5,150 439
<i>PLUMOY</i>	-0,019 913	0,005 088	-3,913 654
<i>VTEMMOY</i>	-0,006 904	0,006 175	-1,118 075
<i>CEREAL</i>	-0,084 164	0,048 678	-1,728 996
<i>RESIND</i>	0,248 482	0,058 565	4,242 860

La log-vraisemblance du modèle est 2 562,8184. Il apparaît que l'objectif de rendement influence positivement le nombre de passages, ainsi que la température moyenne sur la période. En revanche, on confirme bien qu'une pluviométrie limitée entraîne un nombre de tours plus important. La variance des températures joue aussi à la baisse sur *NPAR*, mais le paramètre n'est pas significatif. Enfin, le fait de posséder un dispositif d'irrigation individuel a un impact positif sur le nombre de tours d'eau.

Pour vérifier si la contrainte d'égalité entre espérance et variance de la variable dépendante est réaliste dans ce cas, on estime le modèle NegBin (procédure NegBin dans GAUSS) sur les mêmes observations, avec les mêmes variables explicatives. Les résultats sont :

Variable	Paramètre	Écart type	<i>t</i> de Student
Constante	-2,062 971	0,493 394	-4,181 180
<i>OBJREND</i>	0,020 706	0,002 387	8,675 720
<i>TEMMOY</i>	0,116 464	0,025 074	4,644 746
<i>PLUMOY</i>	-0,018 436	0,004 835	-3,812 791
<i>VTEMMOY</i>	-0,006 768	0,005 221	-1,296 330
<i>CEREAL</i>	-0,055 289	0,051 862	-1,066 080
<i>RESIND</i>	0,204 842	0,052 852	3,875 789
γ	-0,420 784	0,242 818	-1,732 922

La log-vraisemblance est dans ce cas 2 605,4734. La procédure NegBin de GAUSS estime un paramètre supplémentaire γ , tel que : $Var(Y) = E(Y) \times [1 + \exp(\gamma)]$. On constate que les paramètres estimés ne sont pas très différents de ceux obtenus avec le modèle de Poisson. Pour tester l'hypothèse que le modèle de Poisson est correct, on calcule la statistique de test du rapport des vraisemblances. En effet, le modèle de Poisson introduit une restriction sur la valeur du paramètre γ . La statistique de test LR suit un χ^2 à 1 degré de liberté, et vaut $2 \times (2605,4734 - 2562,8184) = 85,31$. La contrainte correspondant à la spécification de Poisson est donc rejetée.

III. Estimateurs des moments

Bien que leur utilisation dans les modèles à variables qualitatives soit moins répandue que celle par maximum de vraisemblance, les estimateurs des moments sont parfois des substituts intéressants. Le grand intérêt de ces estimateurs est qu'ils sont en général robustes à la mauvaise spécification du modèle, puisque leur dérivation ne requiert pas d'hypothèses distributionnelles particulières sur les résidus. Cette différence par rapport au maximum de vraisemblance, qui n'est pas convergent si ces hypothèses sont mises en défaut, est l'une des raisons du succès de cette classe d'estimateurs. Nous présentons d'abord l'estimateur par variable instrumentale comme introduction aux GMM.

A. L'estimateur par variable instrumentale

Soit le modèle sous forme matricielle :

$$Y = X\beta + \varepsilon$$

où X est une matrice ($N \times K$) de variables explicatives, β un vecteur ($1 \times K$) de paramètres, ε est un vecteur ($N \times 1$) de résidus de moyenne 0 et de variance-covariance $\sigma^2 I_N$, où I_N est la matrice identité ($N \times N$) (les éléments de la diagonale principale, dans l'axe Nord-Ouest/Sud-Est, sont égaux à 1, tous les autres éléments sont nuls). Pour les MCO, l'hypothèse indispensable assurant l'absence de biais est : $E(\varepsilon|X) = 0$. Si cette condition n'est pas vérifiée, il est néanmoins possible de trouver des variables auxiliaires W , appelées *instruments*, tels que $E(\varepsilon|W) = 0$. Cette écriture signifie que le résidu n'est corrélé avec aucune fonction des W , et en pratique elle est remplacée par la condition moins forte $E(W'\varepsilon) = 0$. Si W est de dimension $N \times L$ avec $L = K$ et si $W'X$ est de rang plein, on remplace la condition $E(W'\varepsilon) = 0$ par son équivalent empirique :

$$\frac{1}{N} \sum_{i=1}^N W_i' \varepsilon_i = \frac{1}{N} \sum_{i=1}^N W_i'(Y_i - X_i \beta) = 0 \Leftrightarrow \beta = \frac{\left[\frac{1}{N} \sum_{i=1}^N W_i' Y_i \right]}{\left[\frac{1}{N} \sum_{i=1}^N W_i' X_i \right]}$$

où W_i et X_i sont des vecteurs ($1 \times L$) et ($1 \times K$) respectivement, soit sous forme matricielle :

$$\boxed{\beta = (W'X)^{-1} W'Y}$$

L'estimateur ci-dessus est appelé *estimateur des moindres carrés indirects* (EMCI, *Indirect Least Squares Estimator*). La condition $E(W'\varepsilon) = 0$ est un système de *conditions d'identification* (ou d'orthogonalité) de $K = L$ équations permettant d'estimer K paramètres. Ces conditions sont aussi appelées *conditions de moment*.

Supposons maintenant que la matrice des instruments W admette $L > K$ composantes. Il n'est plus possible d'inverser la matrice $W'X$, car elle n'est plus carrée. En revanche, si l'on prémultiplie l'égalité $W'Y = W'X\beta$ par $X'W(W'W)^{-1}$, on a :

$$X'W(W'W)^{-1}W'Y = X'W(W'W)^{-1}W'X\beta$$

$$\Leftrightarrow \beta = [X'W(W'W)^{-1}W'X]^{-1} [X'W(W'W)^{-1}W'Y]$$

c'est-à-dire un vecteur ($K \times 1$). Cet estimateur est appelé *estimateur de la variable instrumentale* (EVI, *Instrumental Variable Estimator*) ou encore *estimateur des doubles moindres carrés* (EDMC, *Two-stage Least Squares Estimator*). La matrice $W(W'W)^{-1}W'$ est symétrique et semi-définie positive, et est appelée *matrice de projection*, notée P_W ⁷. La philosophie de la méthode décrite ci-dessus est intéressante, car elle consiste à estimer un vecteur de paramètres à partir d'un ensemble de conditions d'identification, de façon équivalente à la minimisation d'un critère. Dans le dernier cas où $L > K$, le critère est $\varepsilon'P_W\varepsilon$, et le système admet L conditions de moment pour K paramètres : on a $L - K$ contraintes suridentifiantes⁸.

Un aspect très important en pratique est de vérifier la validité des instruments W choisis, c'est-à-dire qu'ils ne soient pas corrélés avec le terme d'erreur ε . Un test simple revient à régresser les résidus EVI sur les instruments W . La statistique de test est alors simplement : $NR^2 \sim \chi^2(L - K)$ (voir Davidson et McKinnon, 1993).

B. L'estimateur GMM

L'idée de base de l'estimation par GMM est de ne spécifier que la forme paramétrique de certains moments (en général l'espérance), et d'utiliser ces moments pour construire des conditions d'identification. Aucune hypothèse sur la distribution des ε n'est nécessaire. En ce sens, l'estimateur GMM est robuste à la mauvaise spécification du modèle, et la procédure de test de cette dernière est remarquablement simple (voir plus bas). En revanche, l'estimateur GMM est en général moins efficace que celui du maximum de vraisemblance.

L'estimateur de la méthode des moments généralisés (GMM, *Generalized Method of Moments*) se base sur les L conditions, pour un modèle très général (linéaire ou non, voir Hansen, 1982) : $E(W'\varepsilon) = 0$ où $\varepsilon = h(Y, X; \theta)$ et θ est

7. Prémultiplier un vecteur (ou une matrice) par P_W revient à calculer une prédiction de ce vecteur (ou de chaque colonne de cette matrice) par une combinaison linéaire des W .

8. Certains auteurs utilisent la terminologie de variable instrumentale pour le cas $L = K$ et celui des doubles carrés lorsque $L > K$.

un vecteur ($K \times 1$) de paramètres. Ces conditions sont approchées par leur équivalent empirique :

$$\frac{1}{N} \sum_i^N W_i' \varepsilon_i = G(Y, X, W; \theta) \quad (\text{un vecteur } L \times 1)$$

et l'estimateur GMM de θ est le minimant du critère suivant :

$$G'(Y, X, W; \theta) A G(Y, X, W; \theta)$$

où A est une matrice de pondération semi-définie positive ($L \times L$).

On montre que le critère GMM produit un vecteur de paramètres estimés convergent et la variance des estimateurs est :

$$\text{Var}(\hat{\theta}) = \frac{1}{N} (D' A D)^{-1} D' A \Phi A D (D' A D)^{-1}$$

où :

$$D = \text{plim} \frac{1}{N} \sum_{i=1}^N W_i' \left(\frac{\partial \varepsilon_i}{\partial \theta} \right) \quad (\text{une matrice } L \times K)$$

et :

$$\Phi = \left\{ \text{plim} \frac{1}{N-1} \sum_{i=1}^N W_i' \varepsilon_i \varepsilon_i' W_i \right\}$$

une matrice de dimension $L \times L$. Un estimateur GMM est efficace si et seulement si la matrice de pondération vaut $A = (\Phi/N)^{-1}$, qui n'est autre que l'inverse de la matrice de variance-covariance de $G(Y, X, W; \theta)$ ⁹. La variance de l'estimateur est alors $\text{var}(\hat{\theta}) = (1/N)(D' \Phi^{-1} D)^{-1}$. L'utilité de la pondération dans le critère est de normaliser les résidus pour éviter qu'une observation ait une influence trop grande sur le critère. On voit aisément que si les résidus sont homoscédastiques, l'estimateur GMM est équivalent à l'estimateur EVI. En revanche, si la variance des erreurs n'est pas constante, l'estimateur GMM diffère, en ce sens que la matrice Φ^{-1} effectue une pondération des conditions de moment.

Prenons par exemple la condition d'identification $G = (1/N)(Y - X\beta)'W$. On suppose à présent que ε n'est plus homoscédastique, mais admet une matrice de variance-covariance Ω diagonale, d'élément typique $E(Y_i - X_i\beta)^2$ et de dimension ($N \times N$). Dans ce cas :

$$D = -\text{plim} \frac{1}{N} W' X$$

9. Si une variable Z admet une moyenne μ_Z et une variance σ_Z^2 calculées sur un échantillon de taille N , alors $\text{var}(\mu_Z) = \sigma_Z^2/N$.

$$\Phi = \text{plim} \left(\frac{1}{N-1} \sum_i (Y_i - X_i \beta)^2 W_i' W_i \right) = \text{plim} \left(\frac{1}{N-1} W' \Omega W \right)$$

L'estimateur est alors $\hat{\beta} = [X' W \Phi^{-1} W' X]^{-1} [X' W \Phi^{-1} W' Y]$ et la matrice de variance-covariance asymptotique est donnée par :

$$\boxed{\text{var}(\hat{\theta}) = [X' W (W' \Omega W)^{-1} W' X]^{-1}}$$

Dans le cas particulier où $\Omega = \sigma^2 I_N$, on retrouve bien l'expression de l'estimateur EVI vu plus haut.

On montre que l'estimateur GMM est asymptotiquement normal :

$$\boxed{\sqrt{N}(\hat{\theta} - \theta) \sim N\left(0, [D' \Phi^{-1} D]^{-1}\right)}$$

Pour le voir, on remarque que le critère GMM peut s'écrire :

$$\begin{aligned} & \left(\frac{1}{N} \sum_i^N \varepsilon_i W_i \right) \left[\frac{\Phi}{N} \right]^{-1} \left(\frac{1}{N} \sum_i^N W_i' \varepsilon_i \right) \\ &= \left(\frac{1}{\sqrt{N}} \sum_i^N \varepsilon_i W_i \right) \Phi^{-1} \left(\frac{1}{\sqrt{N}} \sum_i^N W_i' \varepsilon_i \right) \end{aligned}$$

Or, sous la condition d'identification, la variable aléatoire $W_i' \varepsilon_i$ est de moyenne nulle. Par conséquent, $\sum_i^N W_i' \varepsilon_i$ est d'ordre \sqrt{N} ($O(\sqrt{N})$) et un théorème central-limite est utilisé pour déduire la normalité de $(1/\sqrt{N}) \sum_i^N W_i' \varepsilon_i$. Ainsi, l'expression du critère GMM ci-dessus est bien une forme quadratique d'une variable aléatoire normale, ramenée à sa variance ($\Phi = \text{var}((1/\sqrt{N}) \sum_i^N W_i' \varepsilon_i)$) (voir Davidson et McKinnon, 1993). Le seul point délicat pour une mise en œuvre de l'estimateur est le calcul de la matrice de variance-covariance Φ . Une méthode en deux étapes est suggérée par Hansen (1982) :

- estimer les paramètres θ par variable instrumentale ($\Phi = I_L$) ;
- calculer un estimateur de Φ : $\hat{\Phi} = (1/N) \sum_i^N W_i' \hat{\varepsilon}_i \hat{\varepsilon}_i' W_i$;
- remplacer Φ par $\hat{\Phi}$ dans le critère GMM et obtenir les estimateurs $\hat{\theta}$ de deuxième étape.

L'estimateur GMM basé sur cette méthode en deux étapes est robuste à toute forme d'hétérosécédasticité, ce qui peut constituer un avantage important par rapport à l'estimateur EVI. On a vu qu'un estimateur GMM est efficace si le critère à minimiser est pondéré par la matrice de variance-covariance des conditions d'identification. Le choix optimal des instruments est $W = \Omega^{-1} J$, où $J(\theta)$ est la matrice des dérivées premières de la fonction $h(Y, X; \theta)$ par rapport à θ .

C. Exemple : les moindres carrés généralisés

Soit la condition d'identification $X'(Y - X\beta) = 0$ et ε de variance-covariance Ω , avec $\Omega \neq \sigma^2 I_N$. Les résidus sont hétéroscédastiques et, d'après le résultat ci-dessus, on doit prendre $W = \Omega^{-1}X$:

$$\Leftrightarrow X'\Omega^{-1}(Y - X\beta) = 0 \Leftrightarrow X'\Omega^{-1}Y = X'\Omega^{-1}X\beta$$

$$\Leftrightarrow \hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$$

Cette expression correspond à l'estimateur des moindres carrés généralisés (MCG, *Generalized Least Squares*). Si la matrice Ω est inconnue, on peut l'estimer dans une première étape à partir des résidus MCO. Dans le cas particulier où Ω est diagonale, ce qui correspond à une forme d'hétéroscédasticité plus simple pour laquelle les covariances $E(\varepsilon_i \varepsilon_j)$ sont nulles pour tout $i \neq j$, on utilise la procédure des moindres carrés pondérés (MCP, *Weighted Least Squares*). Cette méthode revient à estimer dans un premier temps les écarts types $\sigma_i = \sqrt{var(\varepsilon_i)}$, $\forall i$. On pondère ensuite chaque observation en divisant toutes les variables de l'équation (*y compris le terme constant*) par σ_i . Il est facile de montrer que les MCP correspondent aux MCG dans ce cas. Même si les MCO sont convergents dans un modèle avec hétéroscédasticité, les écarts types sont biaisés et l'inférence statistique (t de Student, etc.) n'est pas correcte. On retrouve ici le résultat d'optimalité des MCG dans le modèle linéaire, par rapport aux MCO qui supposent une homoscédasticité des résidus.

Une littérature importante a été consacrée à la détection et la correction de l'hétéroscédasticité dans les modèles de régression. On a vu à la section 2.I.D. qu'une procédure de correction existe pour les modèles mal spécifiés, estimés par maximum de vraisemblance. La procédure analogue pour les moindres carrés est fournie par White (1980). Sous des conditions très générales et une variance de l'erreur $E(\varepsilon_i^2) = \sigma_i^2$, $i = 1, 2, \dots, N$, l'estimateur de White, robuste à toute forme d'hétéroscédasticité, est donné par :

$$var(\hat{\beta}) = \frac{1}{N} \left(\frac{1}{N} \sum_i^N x_i' x_i \right)^{-1} \left(\frac{1}{N} \sum_i^N \hat{\varepsilon}_i^2 x_i' x_i \right) \left(\frac{1}{N} \sum_i^N x_i' x_i \right)^{-1}$$

où $\hat{\varepsilon}_i$ est le résidu estimé dans une première étape par MCO. Le test de l'hétéroscédasticité proposé par White (1980) est alors NR^2 , où R^2 est le coefficient de détermination de la régression des $\hat{\varepsilon}_i^2$ sur une constante et toutes les combinaisons des produits croisés des X . Cette statistique de test, du type multiplicateur de Lagrange, est distribuée comme un $\chi^2(L)$, où L est le nombre de variables explicatives dans la régression de $\hat{\varepsilon}_i^2$ (*y compris le terme constant*).

D. Inférence statistique

1) Test de Hansen

Un test fondamental est celui de la pertinence des conditions d'identification. Comme on a vu que $(1/\sqrt{N}) \sum_i^N W_i' e_i$ suit une loi normale, le critère GMM suit une distribution χ^2 avec comme degrés de liberté le nombre de contraintes suridentifiantes ($L - K$) (Hansen, 1982). Il est alors évident qu'un tel test de spécification n'a de sens que si le modèle est suridentifié. La statistique de test de Hansen sert à vérifier la spécification du modèle. Sous l'hypothèse nulle $E(W'e) = 0$:

$$\left(\sum_i^N \hat{e}_i W_i \right) \left[\sum_i^N W_i' \hat{e}_i \hat{e}_i' W_i \right]^{-1} \left(\sum_i^N W_i' \hat{e}_i \right) \sim \chi^2(L - K)$$

où \hat{e}_i est le résidu de deuxième étape : $\hat{e}_i = h(Y_i, X_i; \hat{\theta})$ et \hat{e}_i le résidu de première étape (par exemple obtenu avec la méthode de la variable instrumentale) utilisé pour le calcul de la variance-covariance. Si la statistique de Hansen conduit à un rejet de l'hypothèse nulle, il faut respécifier le modèle ($h(Y, X; \theta)$) ou bien modifier la liste des instruments.

2) Test d'Hausman

Il est également possible de tester la spécification d'un modèle à l'aide du test d'Hausman, dont le principe est le suivant. On souhaite évaluer la validité des conditions de moments conduisant à l'identification d'un vecteur de paramètres, qui sera convergent uniquement si l'hypothèse nulle est vraie. Supposons qu'il existe un autre estimateur, moins efficace sous l'hypothèse nulle, mais toujours convergent. Le test d'Hausman consiste à comparer les deux estimateurs qui devraient être très proches sous l'hypothèse nulle (Hausman, 1978). Là encore, on utilise le résultat de convergence de la forme quadratique d'une variable normale centrée et ramenée à sa variance vers un χ^2 . Soit $\hat{\theta}$ le vecteur de paramètres estimés avec les instruments W , $\hat{\theta}_0$ les paramètres estimés avec les instruments W^* , où W englobe les composantes de W^* . Si l'hypothèse nulle $E(W'e) = 0$ est vraie, $\hat{\theta}$ sera plus efficace que $\hat{\theta}_0$ car il est calculé avec plus d'instruments. On aura donc : $var(\hat{\theta}) < var(\hat{\theta}_0)$. La statistique de test d'Hausman s'écrit :

$$(\hat{\theta} - \hat{\theta}_0)' [var(\hat{\theta}_0) - var(\hat{\theta})]^{-1} (\hat{\theta} - \hat{\theta}_0)$$

sous l'hypothèse nulle : $E(W'e) = 0$. Ce test est utilisable même si les deux vecteurs n'ont pas le même nombre de composantes. Il suffit de calculer les différences dans l'expression du test ci-dessus pour les sous-vecteurs et les sous-matrices de composantes communes à $\hat{\theta}$ et $\hat{\theta}_0$.

3) Tests d'hypothèses

La transposition des statistiques de test du rapport des vraisemblances, de Wald et du multiplicateur de Lagrange au cas des GMM ne pose pas de difficultés particulières. (Voir dans ce chapitre la section consacrée à ces tests dans le contexte du maximum de vraisemblance.) On teste un ensemble de q contraintes $R(\theta) = 0$ et l'on pose :

$$C(\theta) = \frac{\partial R(\theta)}{\partial \theta} \quad (\text{une matrice } q \times K)$$

La statistique de test de Wald est, en notant que $\text{var}[R(\theta)] = C(\theta)\text{var}(\theta)C'(\theta)$:

$$NR'(\hat{\theta}) \left[C(\hat{\theta})(D'\Phi^{-1}D)^{-1}C'(\hat{\theta}) \right]^{-1} R(\hat{\theta})$$

évaluée en $\hat{\theta}$, la valeur estimée des paramètres par GMM.

Le test du multiplicateur de Lagrange se base sur le gradient de la fonction objectif, qui est ici $(\partial G(Y, X, W; \theta)/\partial \theta)[\Phi/N]^{-1}G(Y, X, W; \theta)$. Sous l'hypothèse nulle, la statistique de test est :

$$LM = NG'(Y, X, W; \theta)\Phi^{-1}(\partial G(Y, X, W; \theta)/\partial \theta)(D'\Phi^{-1}D)^{-1} \\ \times (\partial G(Y, X, W; \theta)/\partial \theta')\Phi^{-1}G(Y, X, W; \theta).$$

Enfin, l'équivalent du test LR est ici simplement la différence entre les critères GMM constraint et libre. Si $\bar{\theta}$ est le vecteur de paramètres estimés sous la contrainte $R(\theta) = 0$, la statistique du rapport des vraisemblances est :

$$LR = N \times \left[G'(Y, X, W; \bar{\theta})\Phi(\bar{\theta})^{-1}G'(Y, X, W; \bar{\theta}) \right. \\ \left. - G'(Y, X, W; \hat{\theta})\Phi(\hat{\theta})^{-1}G'(Y, X, W; \hat{\theta}) \right]$$

Notons enfin que pour cette dernière statistique de test, la matrice Φ doit être calculée d'après $\bar{\theta}$ puis $\hat{\theta}$, c'est-à-dire que la contrainte $R(\theta) = 0$ doit être imposée dès la première étape d'estimation par GMM.

E. Application : valeurs foncières

L'exemple proposé porte sur un échantillon très utilisé dans la littérature empirique, celui des valeurs foncières dans la région de Boston. Cet échantillon est constitué de 506 observations sur la médiane, pour chaque collectivité locale, des valeurs foncières des habitations (*MEDV*). On cherche à expliquer ces valeurs médianes en fonction de caractéristiques de chaque commune : *CRIM* (taux de criminalité), *ZN* (proportion de la surface résidentielle destinée

à des lotissements de taille supérieure à 2 300 m²), *RAD* (indice d'accessibilité aux voies rapides), *TAX* (taux de la taxe foncière, exprimé en 10 000 US \$) et *AGE* (proportion des habitations construites avant 1940).

Les autres variables de l'échantillon sont :

- *LSTAT* : pourcentage des habitants avec un revenu modeste ;
- *B* : proportion de Noirs dans la commune ;
- *RM* : nombre moyen de pièces par logement ;
- *INDUS* : proportion de terrain industriel et commercial ;
- *PTRATIO* : rapport élèves-enseignants sur la commune.

La variable dépendante, *MEDV*, est mesurée en 10 000 US \$. Le modèle est estimé par MCO, puis par variable instrumentale et GMM, avec trois jeux d'instruments :

- $W_1 = (1, LSTAT, ZN, B, RM, INDUS, PTRATIO, AGE)$;
- $W_2 = (1, LSTAT, ZN, B, RM, INDUS, RAD)$;
- $W_3 = (1, LSTAT, ZN, B, RM, INDUS, AGE)$.

Le nombre de variables explicatives (hors terme constant) étant de 5, tous les modèles sont suridentifiés. La première spécification (W_1) contient l'ensemble d'instruments le plus complet. Le deuxième (W_2) contient *RAD* mais élimine *PTRATIO* et *AGE*, alors que le troisième (W_3) substitue *AGE* à *RAD* dans W_2 . Il est donc possible de vérifier la validité des instruments *PTRATIO*, *AGE* et *RAD*.

Le tableau suivant donne les résultats des estimations, avec les *t* de Student entre parenthèses, pour les trois spécifications.

Variable	MCO	(1) EVI	(1) GMM	(2) EVI	(2) GMM	(3) EVI	(3) GMM
Constant	33,0982	21,4299	22,4653	21,4299	9,8802	21,4299	39,5900
	27,6696	(3,9758)	(4,3991)	(2,2289)	(1,1580)	(2,9621)	(3,2392)
<i>CRIM</i>	-0,2407	-4,7510	-1,7241	-4,7510	-4,4723	-4,7510	-4,6419
	-2,8225	(-4,7147)	(-2,6913)	(-4,4986)	(-4,1063)	(-3,5358)	(-3,1910)
<i>ZN</i>	0,0883	4,9851	0,0971	4,9851	0,3551	4,9851	0,2303
	0,1404	(2,9459)	(3,9633)	(2,5593)	(3,5505)	(2,4828)	(3,6234)
<i>RAD</i>	0,4420	4,5012	0,3480	4,5012	3,9051	4,5012	6,7024
	2,5337	(2,8806)	(0,5627)	(3,9853)	(4,1665)	(2,8328)	(2,6716)
<i>TAX</i>	-0,0341	-6,8175	-0,0095	-6,8175	-0,1159	-6,8175	-0,1767
	-0,0658	(-2,0112)	(-0,4482)	(-2,2768)	(-3,5230)	(-2,3558)	(-2,3776)
<i>AGE</i>	-0,0143	-0,8873	0,0684	-0,8873	0,5030	-0,8873	0,0764
	0,0895	(-1,9655)	(2,4200)	(-1,7963)	(2,6674)	(-0,9597)	(1,4914)
χ^2		23,42	21,23	19,93	2,93	12,14	0,0026
p-value		(0,0006)	(0,00002)	(0,002)	(0,086)	(0,0588)	(0,9585)

Les statistiques du χ^2 des colonnes EVI correspondent au test d'Hausman MCO contre EVI, sous l'hypothèse nulle d'exogénéité des variables explicatives. Les statistiques des colonnes GMM proviennent du test de spécification d'Hansen : elles indiquent la pertinence des conditions d'identification à la base de l'estimateur GMM. Au vu des tests d'Hausman, on constate que seule la spécification (3) pour les instruments ne conduit pas à un rejet de l'hypothèse d'exogénéité. On en déduit que le choix des instruments est capital pour juger de la pertinence de l'estimateur MCO. C'est l'un des inconvénients de l'estimateur EVI : l'inférence sur le modèle MCO est très sensible à l'hypothèse alternative (le modèle EVI). En ce qui concerne l'estimateur GMM, on voit que la spécification du modèle (1) est rejetée par le test d'Hansen, ce que l'on interprète comme la nature endogène de certains instruments. Comme l'instrument *PTRATIO* ne figure que dans le modèle (1), on en déduit que cette variable est très certainement invalide en tant qu'instrument. Une remarque importante concerne la grande variabilité des estimations obtenues des paramètres en fonction des différentes méthodes et du choix des instruments. Alors que *CRIM* et *TAX* influencent la variable dépendante à la baisse et les variables *ZN* et *RAD* à la hausse dans tous les cas, l'effet de *AGE* est négatif avec l'EVI et positif avec l'estimateur GMM. Les variables *ZN* et *AGE* ont des paramètres estimés très différents selon que la méthode est EVI ou GMM.

IV. Algorithmes d'optimisation

Les estimateurs par maximum de vraisemblance (et la méthode des moments pour les modèles non linéaires) nécessitent l'optimisation numérique d'un critère, la solution pour l'estimateur n'étant pas analytique. Cette section présente les méthodes numériques couramment utilisées dans la maximisation de fonctions, telles qu'elles sont mises en oeuvre dans les logiciels économétriques. On considère le problème général d'obtenir un vecteur de paramètres θ maximisant une fonction $F(\theta)$. Cette fonction peut s'écrire comme une somme de contributions individuelles, $F(\theta) = \sum_i^N f_i(y_i; x_i; \theta)$ comme pour le maximum de vraisemblance ou les moindres carrés (dans ce dernier cas, on maximise $-\sum_i^N f_i$). Trois classes d'algorithmes importants sont : le simplexe, le « recuit » simulé (*simulated annealing*) et les méthodes de gradient. Les deux premières sont utilisées dans des cas très particuliers, par exemple lorsque le critère à maximiser n'est pas dérivable ou admet de nombreux maxima locaux. Les méthodes de gradient étant les plus courantes et les plus efficaces pour la maximisation de vraisemblances ou de formes quadratiques, on se concentrera donc sur celles-ci. (Voir Mead et Nelder, 1965 ; Press *et al.*, 1986 ; Ferrier, Goffe et Rodgers, 1994, pour les méthodes du simplexe et du « recuit » simulé.)

A. Position du problème

Les algorithmes de maximisation sont utilisés lorsque la condition du premier ordre $\partial F(\theta)/\partial\theta = 0$ n'admet pas de solution analytique : θ doit être déduit par tâtonnement ou par un algorithme itératif. Un cas important en économétrie, pour lequel un tel algorithme n'est pas nécessaire, est le critère quadratique en θ : $F(\theta) = a + b\theta + \theta'c\theta$, où a est un scalaire, b un vecteur $1 \times K$ et c une matrice $(K \times K)$ semi-définie négative¹⁰. Si par exemple $F = -\sum_i^N (y_i - x_i\beta)^2$, on a $\theta = \beta$, $a = -\sum_i^N y_i^2$, $b = 2\sum_i^N y_i x_i$ et $c = x_i' x_i$. La solution de ce problème quadratique est $\theta = c^{-1}b$, soit ici $\beta = [\sum_i^N x_i' x_i]^{-1} [\sum_i^N x_i' y_i]$. Un algorithme de maximisation utilise plusieurs éléments :

- des valeurs initiales θ_0 pour amorcer le processus itératif ;
- une règle de passage d'un vecteur θ au suivant ;
- une règle d'arrêt si convergence.

De façon générale, plus le critère est de forme complexe, plus le choix des valeurs initiales est important. Si $F(\theta)$ est globalement concave, c'est-à-dire $\partial^2 F / \partial \theta \partial \theta' < 0$ pour toute valeur de θ , alors l'algorithme convergera vers le θ optimal, quelle que soit la valeur de départ. Mais même dans ce cas relativement favorable, la convergence peut être extrêmement longue si les valeurs de départ sont éloignées de l'optimum, et/ou si l'algorithme n'est pas adapté. Les algorithmes usuels se distinguent essentiellement par leur règle de passage ; on dispose actuellement de procédures optimisées et donc rapides. Enfin, la règle d'arrêt est généralement du type : arrêter le processus itératif si la variation de θ ou du critère entre l'itération actuelle et la précédente est inférieure à une valeur seuil (par exemple 0.000 01).

La règle de passage est définie de la façon suivante : partant des valeurs initiales θ_0 , trouver le prochain vecteur de paramètres θ_1 tel que $F(\theta_1) \geq F(\theta_0)$, et ainsi de suite :

$$\theta_k = \theta_{k-1} + \lambda_{k-1} D_{k-1}$$

pour l'itération k , où λ_{k-1} est le *pas* à l'itération $k-1$, D_{k-1} la *direction*. D_{k-1} indique quelle orientation les composantes du nouveau vecteur devront prendre, λ_{k-1} indique l'amplitude du saut dans cette orientation. Prenons le cas simple d'un critère avec un seul paramètre ($K = 1$). Puisque l'on cherche un maximum, un gradient positif indique que l'on se trouve sur le « flanc gauche » de la colline à franchir, il faut donc prendre $\theta_1 > \theta_0$. En ce qui concerne le pas, il faut trouver la valeur λ_0 telle que le gradient $\partial F(\theta_0 + \lambda_0 D_0) / \partial \lambda_0 \approx 0$. Bien sûr, dans le cas général où $K > 1$, la solution n'est pas immédiate et la recherche du pas (*line search*) doit se faire à chaque itération. La majorité des problèmes de convergence avec les algorithmes de gradient proviennent de la

10. Bien sûr, dans le cas de la minimisation de F , $-c$ est semi-définie positive.

difficulté d'évaluer le Hessien. Un bon test est de vérifier si les éléments de la diagonale du Hessien sont à la même échelle. Si tel n'est pas le cas, l'algorithme aura des difficultés à appliquer la règle de passage. Normaliser le problème en ramenant les variables à une même échelle peut parfois le résoudre. Il existe souvent des façons plus simples d'écrire un critère à maximiser, en effectuant par exemple une transformation des paramètres ou des variables. Enfin, si la concavité globale du problème ne peut être démontrée, il est indispensable de mettre en oeuvre l'algorithme à partir de plusieurs valeurs de départ éloignées les unes des autres, afin de vérifier que le maximum trouvé n'est pas seulement local.

B. Méthodes de gradient

La méthode la plus simple est celle du *Steepest Ascent* (ascension la plus raide), où la direction est égale au gradient. Si la fonction est concave au point θ_{k-1} , cette direction correspond à la progression la plus grande possible de F (d'après le théorème de l'enveloppe). Le pas est calculé facilement par :

$$\lambda_{k-1} = -g(\theta_{k-1})'g(\theta_{k-1})/g'(\theta_{k-1})H(\theta_{k-1})g(\theta_{k-1})$$

où $g(\theta_{k-1})$ est le gradient de F au point θ_{k-1} , et la règle de passage est :

$$\theta_k = \theta_{k-1} - g(\theta_{k-1}) \times g'(\theta_{k-1})'g(\theta_{k-1})/g'(\theta_{k-1})H(\theta_{k-1})g(\theta_{k-1})$$

Cette méthode nécessite le calcul du Hessien à chaque itération, ce qui la rend souvent lente. De plus, il est possible que $H(\theta)$ ne soit pas semi-défini négatif (en particulier si l'on se trouve loin du maximum), auquel cas l'algorithme peut diverger.

Une méthode plus rapide est celle dite de Newton (ou Newton-Raphson), qui est recommandée si le critère est globalement concave et qui permet d'atteindre le maximum en une seule itération si le critère est quadratique. En partant d'un développement limité de la condition du premier ordre autour d'un point, on trouve :

$$\begin{aligned}\theta_{k-1} : g(\theta_k) &= 0 = g(\theta_{k-1}) + H(\theta_{k-1})(\theta_k - \theta_{k-1}) \\ \Leftrightarrow \theta_k &= \theta_{k-1} - H(\theta_{k-1})g(\theta_{k-1})\end{aligned}$$

Comme la méthode de Newton est très efficace pour des critères pouvant être facilement approchés par des formes quadratiques et pour une région paramétrique proche du maximum, on utilise parfois un algorithme plus robuste pour localiser cette région, puis l'algorithme de Newton pour préciser le maximum.

Pour pallier les inconvénients de la méthode de Newton (pas de 1, calcul du Hessien qui de plus peut être non semi-défini négatif), on utilise des variantes

de la méthode, dites *à métrique variable* : les procédures de Davidon-Fletcher-Powell (DFP) et de Broyden-Fletcher-Goldfarb-Shanno (BFGS). Ces deux méthodes construisent le Hessian, qui n'a donc pas besoin d'être calculé, à partir d'une matrice semi-définie négative (Fletcher, 1980). Pour l'algorithme DFP, par exemple, on part d'un pas égal à 1 et un Hessian égal à moins la matrice identité. La direction à l'étape k vaut $D_k = J_k g_k$ où :

$$J_k = J_{k-1} + \frac{\lambda_{k-1}^2 J_{k-1} g_{k-1} g'_{k-1} J'_{k-1}}{g'_{k-1} J'_{k-1} \Delta g_{k-1}} + \frac{J_{k-1} \Delta g_{k-1} \Delta g'_{k-1} J_{k-1}}{\Delta g'_{k-1} J_{k-1} \Delta g_{k-1}}$$

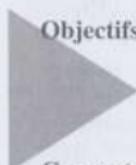
avec $g_k = g(\theta_k)$ et $\Delta g_k = g_k - g_{k-1}$. Le terme $-H(\theta_k)^{-1}$ est approché par J_k après plusieurs itérations, et cette approximation est constamment actualisée. Comme la somme des deux derniers termes dans J_k est définie positive, le Hessian sera toujours défini négatif. Notons enfin que le pas fait l'objet d'une recherche automatique à chaque itération, comme cela est décrit précédemment.

À retenir

La méthode du maximum de vraisemblance est utilisée dans la plupart des modèles à variables qualitatives. Les propriétés d'efficacité et de normalité asymptotique de son estimateur rendent l'inférence statistique particulièrement intéressante. En revanche, ces propriétés dépendent de la validité des hypothèses faites sur la distribution des erreurs. Une autre classe d'estimateurs, basés sur la méthode des moments (variable instrumentale, GMM), possède au contraire l'avantage de s'affranchir de telles hypothèses, ce qui rend ces estimateurs robustes à la mauvaise spécification du modèle. On a vu dans ce chapitre comment utiliser en pratique les deux classes d'estimateurs, et la façon de conduire des tests d'hypothèses et de spécification. Comme la méthode du maximum de vraisemblance nécessite l'optimisation d'une fonction des paramètres, des algorithmes numériques doivent être utilisés. La plupart utilisent des méthodes de gradient, qui sont d'autant plus performantes que le critère à maximiser est régulier et admet peu d'*extrema locaux*.

3. Modèles à variable qualitative binaire

Nous avons vu que la méthode des moindres carrés se prête mal, en général, à l'analyse des modèles à variable dépendante binaire. Ce chapitre introduit une méthode adéquate de modélisation basée sur l'utilisation d'une variable latente. On expose l'approche générale et les deux modèles les plus couramment utilisés (Logit et Probit). Les techniques d'inférence (tests d'ajustement et de significativité globale) sont détaillées pour ces modèles, et des extensions récentes du modèle binaire (modèle bivarié, méthodes non paramétriques et données de panel) sont présentées.



Objectifs du chapitre : présenter la modélisation de variables dépendantes binaires à l'aide d'un modèle à variable latente, dans un cadre d'estimation par maximum de vraisemblance ; détailler les méthodes d'estimation et de test pour les modèles binaires Logit et Probit usuels.

Concepts clés étudiés : variable dépendante qualitative binaire, estimation par maximum de vraisemblance, critères d'ajustement et de choix de modèles, modèles bivariés, méthodes non paramétriques, modèles avec données de panel.

I. Variable binaire et variable latente

Afin de pallier les inconvénients de la méthode des moindres carrés appliquée à une variable dépendante qualitative (hétéroscédasticité, nature continue du résidu inadaptée, valeurs prédictives en dehors de l'intervalle $[0, 1]$), une solution consiste à considérer la réalisation de la variable dépendante binaire comme

provenant d'une certaine *règle de décision*. Cette règle est un mécanisme associant les variables explicatives x_i à l'observation de l'événement $\{y_i = 0\}$ ou $\{y_i = 1\}$. L'intuition est la suivante. Supposons que la réalisation de $\{y_i = 1\}$ est plutôt associée à des valeurs élevées des x_i , celle de $\{y_i = 0\}$ à des valeurs faibles des x_i . Il doit alors exister une certaine valeur seuil dépendant de la combinaison linéaire $x_i\beta$, où β est un vecteur de paramètres, au-delà de laquelle la proportion des $\{y_i = 1\}$ l'emporte sur celle des $\{y_i = 0\}$. Notons c cette valeur seuil. Il est également raisonnable de supposer que notre règle de décision n'est pas déterministe, c'est-à-dire que pour certaines observations y_i peut être nul alors que les valeurs de x_i sont élevées. Ce caractère non déterministe peut être intégré en ajoutant à notre combinaison linéaire un terme aléatoire noté u_i . Notre règle de décision est alors : la proportion des $\{y_i = 1\}$ est « élevée » pour $x_i\beta + u_i > c$, et « faible » pour $x_i\beta + u_i \leq c$. On passe ainsi de la règle déterministe qui serait : $\{y_i = 1\}$ si $x_i\beta > c$, à une règle aléatoire (ou probabiliste) $Prob(y_i = 1) = Prob(x_i\beta + u_i > c)$.

De façon plus rigoureuse, le modèle s'écrit en considérant une variable y_i^* représentant la tolérance du système, telle que :

$$\begin{cases} y_i = 1 & \text{si } y_i^* > c \\ y_i = 0 & \text{si } y_i^* \leq c \end{cases}$$

où la variable $y_i^* = x_i\beta + u_i$ est aléatoire du fait de la présence du terme u_i . Cette variable est appelée *variable latente* (inobservable), et le modèle ci-dessus est le *modèle sous-jacent*. La règle de décision probabiliste devient alors :

$$\begin{cases} Prob(y_i = 1) = Prob(x_i\beta + u_i > c) = 1 - Prob(u_i < c - x_i\beta) \\ Prob(y_i = 0) = Prob(x_i\beta + u_i \leq c) = Prob(u_i \leq c - x_i\beta). \end{cases}$$

Plusieurs remarques sont de rigueur. Tout d'abord, puisque la variable latente est aléatoire, la règle de décision est nécessairement probabiliste. Ainsi, le modèle ne détermine pas exactement la réalisation de l'événement $\{y_i = 1\}$ ou $\{y_i = 0\}$, mais fournit plutôt une mesure théorique de la proportion d'observations pour lesquelles cet événement s'est réalisé. Ensuite, puisque cette mesure théorique (la probabilité) est croissante dans son argument, la probabilité que $y_i = 1$ sera croissante en les composantes de x_i dont les paramètres associés sont positifs, et décroissante pour celles dont les paramètres sont négatifs. On note également que la valeur seuil c dans ce modèle est identique pour toutes les observations. Or, il est possible d'envisager le cas où le seuil associé à une observation particulière dans le système dépend des caractéristiques de cette observation. Nous reviendrons sur ce point plus loin. Si c est un scalaire et si le vecteur de variables explicatives contient un terme constant, les deux ne peuvent être distingués, et l'on peut alors fixer arbitrairement la valeur seuil à 0, ainsi d'ailleurs que l'espérance de u_i . Enfin, peut-être le plus important est de bien noter que l'écriture probabiliste de notre règle de décision dépend exclusivement de la distribution statistique de la seule variable

aléatoire du système, c'est-à-dire u_i . Par conséquent, dès que l'on impose une loi particulière à ce terme aléatoire, les probabilités de $\{y_i = 1\}$ et $\{y_i = 0\}$ pourront être calculées en faisant référence à cette loi¹.

Prenons quelques exemples pour illustrer la démarche basée sur la variable latente.

► Exemple 3.1 : choix de tarification

Considérons un ménage devant choisir entre deux modes de tarification pour sa consommation électrique. Le premier mode comporte une partie fixe (abonnement) A_1 faible, et un prix au kWh p_1 relativement élevé. En revanche, le second mode de tarification est destiné aux gros consommateurs, avec une partie fixe A_2 importante et un moindre coût unitaire au kWh p_2 : on a $A_1 < A_2$ et $p_1 > p_2$. Soit Q_i la quantité d'électricité consommée (en kWh) du ménage. Le ménage choisira le premier mode de tarification si :

$$A_1 + p_1 Q_i < A_2 + p_2 Q_i \quad \text{soit} \quad Q_i > (A_1 - A_2)/(p_2 - p_1)$$

Si Q_i est inconnue de l'économètre, on peut supposer que la consommation réelle d'électricité dépend de caractéristiques du ménage, tel le nombre de personnes constituant le foyer, le mode de chauffage, le nombre d'appareils électroménagers, etc. Si x_i est le vecteur de ces caractéristiques, on peut écrire $Q_i = x_i \beta + u_i$, avec u_i une erreur de moyenne 0. En notant $\gamma = (A_1 - A_2)/(p_2 - p_1)$, le ménage choisira le premier mode de tarification si $Q_i > \gamma$, soit :

$$\text{Prob}(\text{Choix 1er mode}) = \text{Prob}(u_i > \gamma - x_i \beta) = 1 - \text{Prob}(u_i \leq \gamma - x_i \beta)$$

Dans cet exemple encore, la variable latente admet une interprétation directe, à savoir le niveau (inobservable) de consommation électrique du ménage.

► Exemple 3.2 : choix d'une filière d'enseignement

On observe un échantillon d'étudiants dont les uns ont opté pour une filière d'enseignement technologique et les autres une filière générale. Les variables constituant l'échantillon sont y_i (1 si filière technologique, 0 sinon) et un vecteur de caractéristiques noté x_i : revenu et niveau d'éducation des parents, sexe, lieu de résidence. Le modèle est le suivant :

$$y_i^* = x_i \beta + u_i$$

$$\text{Prob}(y_i = 1) = \text{Prob}(y_i^* > c) = 1 - \text{Prob}(u_i < -x_i^* \beta)$$

1. Goldberger, 1964.

où x_i^* est constitué des composantes du vecteur original x_i , à la différence du terme constant qui intègre maintenant le seuil c . Dans ce modèle, la probabilité qu'un étudiant choisisse la filière technologique est croissante en les composantes de x_i dont les paramètres associés sont positifs, et décroissante pour celles dont les paramètres sont négatifs. Mais l'interprétation de la variable latente y_i est peu claire, en dehors d'une mesure de la propension des étudiants à se diriger vers les filières technologiques en fonction de leur milieu social d'origine, lieu de résidence, etc.

L'intérêt de cet exemple est de montrer que, contrairement au précédent, il n'est pas toujours possible de donner un sens économique à la variable latente. Celle-ci n'est alors qu'un artifice destiné à modéliser la réalisation d'un événement en reliant la probabilité de ce dernier à un ensemble de variables explicatives.

II. Modèles binaires Logit et Probit

Nous avons vu à la section précédente que l'introduction d'une variable latente continue permet de modéliser de façon pertinente la réalisation d'un événement, en faisant référence à sa probabilité. Pour calculer cette probabilité et estimer les paramètres du modèle, il est nécessaire de spécifier une distribution statistique pour u_i . Les deux lois statistiques les plus couramment utilisées dans la pratique sont la loi logistique et la loi de Gauss (distribution normale)². Ces deux distributions donnent alors les modèles qualitatifs binaires appelés respectivement Logit et Probit³.

Notons $F(\cdot)$ la fonction de répartition issue de la distribution statistique du terme d'erreur u_i , et $f(\cdot)$ la fonction de densité associée. Comme la valeur seuil peut être normalisée à 0, le modèle s'écrit de façon générale :

$$\begin{cases} \text{Prob}(y_i = 1) = \text{Prob}(u_i > -x_i\beta) = 1 - F(-x_i\beta) \\ \text{Prob}(y_i = 0) = \text{Prob}(u_i \leq -x_i\beta) = F(-x_i\beta) \end{cases}$$

On fait l'hypothèse de symétrie de la densité de u_i autour de 0 :

$$f(x_i\beta) = f(-x_i\beta)$$

2. Voir Amemiya, 1981, pour une présentation d'ensemble de ces modèles binaires.

3. Le préfixe exprime l'origine : le Logit provient de la logistique, et l'on y ajoute le suffixe *it*. Pour la loi normale, il aurait donc été plus correct d'utiliser le terme Normit au lieu de Probit.

et donc :

$$F(x_i\beta) = 1 - F(-x_i\beta)$$

ce qui permet de considérer les valeurs observées de y_i comme les réalisations d'un processus binomial avec probabilité $F(x_i\beta)$ pour $y_i = 1$, et $1 - F(x_i\beta)$ pour $y_i = 0$. À la différence des processus binomiaux usuels, les probabilités varient à chaque « expérience », dans la mesure où elles dépendent de x_i . On aura donc $Prob(y_i = 1) = 1 - F(-x_i\beta) = F(x_i\beta)$ et $Prob(y_i = 0) = F(-x_i\beta) = 1 - F(x_i\beta)$. À partir de ces égalités, le modèle peut s'écrire de façon plus compacte, en introduisant une variable, notée δ_i , prenant la valeur +1 si $y_i = 1$ et -1 sinon. Cette nouvelle variable sera construite comme $\delta_i = 2y_i - 1$. On a alors $Prob(y_i) = F(\delta_i x_i \beta) = 1 - F(-x_i \beta)$ si $y_i = 1$, et $Prob(y_i) = F(-x_i \beta)$ si $y_i = 0$.

Notons qu'un autre modèle est parfois utilisé, le modèle Gompit (de la loi de Gompertz). Contrairement aux Probit et Logit, cette distribution n'est pas symétrique, et la probabilité associée à $\{y_i = 1\}$ est :

$$Prob(y_i = 1) = \exp(-e^{-x_i\beta})$$

A. Le modèle Probit

On fait ici l'hypothèse que l'erreur u_i est une variable suivant une loi normale de moyenne 0 et de variance σ^2 . Les fonctions de densité et de répartition associées à une variable normale $N(0, 1)$ sont respectivement :

$$\phi(X) = \frac{1}{\sqrt{2\pi}} e^{-X^2/2} \quad \text{et} \quad \Phi(X) = \int_{-\infty}^X \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Pour utiliser ces expressions avec u_i , qui est de variance σ^2 , on doit « réduire » l'argument de la probabilité que $y_i = 1$, en divisant par l'écart type σ :

$$Prob(y_i = 1) = Prob(u_i > -x_i\beta) = Prob(u_i < x_i\beta)$$

$$\boxed{\text{(par la propriété de symétrie)} = Prob\left(\frac{u_i}{\sigma} < \frac{x_i\beta}{\sigma}\right) = \Phi\left(\frac{x_i\beta}{\sigma}\right)}$$

Un problème important apparaît à ce stade. Nous voyons que l'écart type σ du terme aléatoire u_i figure toujours en association avec le vecteur de paramètres β dans l'argument de la fonction de répartition normale $\Phi(\cdot)$. En d'autres termes, il n'est pas possible d'identifier ce paramètre séparément de β . Une solution couramment adoptée consiste alors à normaliser l'écart type σ à 1, ce que nous effectuerons par la suite.

Estimation du modèle Probit

En faisant l'hypothèse que les observations sont indépendantes et identiquement distribuées, la vraisemblance du modèle s'écrit comme le produit des probabilités associées aux réalisations de y_i . Notons N_1 et N_0 respectivement l'ensemble des observations pour lesquelles $\{y_i = 1\}$ et $\{y_i = 0\}$. Cette vraisemblance s'écrit pour le Probit, en séparant les observations :

$$L(y, x, \beta) = \prod_{N_1} [\Phi(x_i \beta)] \prod_{N_0} [1 - \Phi(x_i \beta)]$$

ou de façon plus compacte, en considérant l'ensemble des N observations :

$$L(y, x, \beta) = \prod_{i=1}^N [\Phi(x_i \beta)]^{y_i} [1 - \Phi(x_i \beta)]^{1-y_i}$$

soit enfin à l'aide de la nouvelle variable δ_i :

$$L(y, x, \beta) = \prod_{i=1}^N \Phi(\delta_i x_i \beta)$$

Le gradient de la log-vraisemblance est :

$$\begin{aligned} G(\beta) &= \frac{\partial \log L}{\partial \beta} = \sum_i^N \frac{y_i \phi(x_i \beta) x'_i}{\Phi(x_i \beta)} + \frac{(y_i - 1)}{1 - \Phi(x_i \beta)} \phi(x_i \beta) x'_i \\ &= \sum_{i=1}^N \frac{[y_i - \Phi(x_i \beta)] \phi(x_i \beta) x'_i}{\Phi(x_i \beta) (1 - \Phi(x_i \beta))} \end{aligned}$$

En égalisant le gradient à 0 et en notant $w_i = \phi(x_i \beta) / [\Phi(x_i \beta)(1 - \Phi(x_i \beta))]$, on a :

$$\sum_{i=1}^N [y_i - \Phi(x_i \beta)] x'_i w_i = 0$$

L'écriture ci-dessus est comparable à l'équation d'identification d'un modèle de régression non linéaire (voir le chapitre 2). En effet, elle constitue l'équivalent empirique de la condition $E[(x'_i w_i) v_i] = 0$, où v_i est le résidu dans le modèle non linéaire $y_i = \Phi(x_i \beta) + v_i$, et w_i est une variable de pondération.

La présence de ce facteur de pondération w_i rend l'équation d'identification non linéaire, et le vecteur de paramètres ne peut être estimé directement. Par conséquent, un algorithme de maximisation numérique de la log-vraisemblance est nécessaire. Les algorithmes usuels se fondent sur l'utilisation de deux expressions : le gradient et le Hessien associés à la log-vraisemblance. Ce dernier s'écrit :

$$H(\beta) = - \sum_{i=1}^N \frac{\phi^2(x_i\beta)x_i'x_i}{\Phi(x_i\beta)[1-\Phi(x_i\beta)]}$$

Voir le chapitre 2 pour un bref aperçu des méthodes numériques d'optimisation.

B. Le modèle Logit

La distribution logistique à l'origine du modèle Logit admet comme fonctions de répartition et de densité les expressions suivantes :

$$\Lambda(x_i\beta) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} \quad \lambda(x_i\beta) = \frac{\exp(x_i\beta)}{[1+\exp(x_i\beta)]^2}$$

Ces expressions sont plus faciles à manipuler que les fonctions $\phi(.)$ et $\Phi(.)$ de la loi normale. On a ainsi :

$$1 - \Lambda(x_i\beta) = \Lambda(-x_i\beta) = \frac{\exp(-x_i\beta)}{1+\exp(-x_i\beta)} = \frac{1}{1+\exp(x_i\beta)}$$

Remarquons que la probabilité associée à la loi logistique peut être inversée. Si l'on note p_i la probabilité que $y_i = 1$, on a alors la représentation suivante :

$$\log\left(\frac{p_i}{1-p_i}\right) = x_i\beta$$

et l'on vérifie bien que la probabilité que $y_i = 1$ est une fonction croissante de la combinaison linéaire $x_i\beta$.

Estimation du modèle Logit

L'estimation du modèle Logit repose aussi sur la maximisation de la log-vraisemblance. La vraisemblance s'écrit :

$$L(y, x, \beta) = \prod_{i=1}^N \left[\frac{1}{1+\exp(x_i\beta)} \right]^{1-y_i} \left[\frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} \right]^{y_i}$$

soit la log-vraisemblance :

$$\begin{aligned} \log L &= \sum_{i=1}^N \left\{ (1-y_i) \log \left[(1+\exp(x_i\beta))^{-1} \right] + y_i(x_i\beta) - y_i \log(1+\exp(x_i\beta)) \right\} \\ &= - \sum_{i=1}^N \{ \log [1+\exp(x_i\beta)] - y_i x_i \beta \} \end{aligned}$$

Les conditions du premier ordre sont :

$$G(\beta) = \frac{\partial \log L}{\partial \beta} = \sum_{i=1}^N y_i x'_i - \sum_{i=1}^N \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)} x'_i = 0$$

$$\Leftrightarrow \sum_{i=1}^N [y_i - \Lambda(x_i \beta)] x'_i = 0$$

Le Hessien est :

$$H(\beta) = - \sum_{i=1}^N \frac{\exp(x_i \beta)}{[1 + \exp(x_i \beta)]^2} x'_i x_i$$

Comme dans le cas du modèle Probit, les conditions du premier ordre s'écrivent sous la forme d'une condition d'identification : au point solution pour β , l'écart entre la variable dépendante y_i et son espérance conditionnelle ne doit pas être corrélé avec le vecteur des variables explicatives x_i . On note qu'il n'y a pas dans ce cas de variable de pondération.

Après convergence, les valeurs des probabilités estimées sont alors calculées en remplaçant β par son estimation $\hat{\beta}$:

$$\hat{p}_i = \frac{\exp(x_i \hat{\beta})}{1 + \exp(x_i \hat{\beta})}$$

et de la condition du premier ordre, on a :

$$\sum_{i=1}^N x_i y_i = \sum_{i=1}^N \hat{p}_i x_i$$

Supposons que le modèle contient un terme constant, par exemple $x_{i0} = 1$, $\forall i$. De l'expression ci-dessus, il vient :

$$\sum_{i=1}^N y_i = \sum_{i=1}^N \hat{p}_i$$

c'est-à-dire qu'en divisant par N , la fréquence estimée est égale à la fréquence observée. Si par ailleurs le modèle contient une variable indicatrice identifiant un sous-groupe dans l'échantillon (sexe de la personne par exemple), la fréquence estimée est égale à la fréquence théorique sur le sous-groupe pour lequel la variable indicatrice vaut 1⁴.

4. Voir Maddala, 1983.

C. Comparaison entre le Probit et le Logit

Dans la pratique, les modèles Probit et Logit sont très similaires en termes d'ajustement statistique. En effet, les distributions normale et logistique font partie de la même famille de lois exponentielles⁵. On note des différences uniquement dans le cas de très grands échantillons, car le comportement de ces deux distributions de probabilité ne diffère qu'aux extrémités du support (faibles et fortes valeurs de la fonction de répartition F). Pour la loi normale, les probabilités extrêmes sont moins pondérées, la fonction de répartition tendant plus vite vers 0 ou vers 1.

On a vu que, dans le cas du modèle Probit, il n'est pas possible d'estimer séparément le vecteur de paramètres β et l'écart type σ , car ils apparaissent comme un rapport. Outre la possibilité de normaliser σ à 1, il est également possible de contraindre l'un des paramètres de β à 1, par exemple $\beta_1 = 1$. La probabilité que $y_i = 1$ s'écrira :

$$\Phi [(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + x_{ij} + \cdots + \beta_K x_{iK})/\sigma]$$

et σ sera alors obtenu comme l'inverse du paramètre estimé $\hat{\beta}_j$. Dans le cas de la distribution logistique, ce problème d'identification se pose aussi, le modèle Logit imposant la variance de l'erreur du modèle à $\pi^2/3$. Il n'est donc pas possible non plus d'identifier l'écart type de u_i .

Une différence importante entre les deux modèles repose ainsi sur le fait que, les normalisations n'étant pas les mêmes, les valeurs estimées des paramètres ne sont pas directement comparables. Pour ce faire, il convient d'utiliser le facteur $\pi/\sqrt{3}$: on multiplie les estimateurs du Probit par $\pi/\sqrt{3}$ pour comparer avec le Logit, ou l'on multiplie ceux du Logit par $\sqrt{3}/\pi$ pour comparer avec le Probit⁶. Amemiya (1981) préconise l'emploi d'un facteur différent, partant du constat que le rapport des densités au point 0, $\phi(0)/\lambda(0)$ vaut à peu près 1,6. Si, en effet, $x_i\beta$ est proche de 0 en moyenne sur l'échantillon, et que les modèles Probit et Logit prédisent les mêmes effets sur p_i d'une variation de l'un des x_i , alors les coefficients du Logit doivent être environ 1,6 fois ceux du Probit. Cette approximation fonctionnera d'autant mieux que la valeur moyenne de p_i est proche de 0,5 (Davidson et McKinnon, 1993).

Pour finir avec les comparaisons des valeurs numériques des paramètres estimés, il faut mentionner la règle permettant de comparer les estimateurs du modèle à probabilité linéaire LP, vu au chapitre 1, avec ceux du Logit⁷ :

$$\begin{cases} \hat{\beta}_{LP} \approx 0,25\hat{\beta}_{Logit} & (\text{sauf le terme constant}) \\ \hat{\beta}_{LP} \approx 0,25\hat{\beta}_{Logit} + 0,5 & (\text{terme constant}) \end{cases}$$

5. Gouriéroux, 1989.

6. $\pi/\sqrt{3}$ vaut à peu près 1,81.

7. Maddala, 1983.

III. Inférence dans les modèles binaires

Nous présentons dans cette section des éléments d'inférence statistique à partir de l'estimation des modèles Probit et Logit décrits plus haut. Nous précisons également quelques utilisations et interprétations possibles à partir des probabilités estimées.

A. Interprétation des paramètres estimés

Dans les deux modèles étudiés ci-dessus, la variance de l'erreur du modèle n'est pas identifiable : elle est normalisée à 1 pour le Probit et à $\pi^2/3$ pour le Logit. Par conséquent, la valeur numérique des paramètres estimés n'a pas vraiment d'intérêt en soi, dans la mesure où ils ne correspondent aux paramètres de l'équation de la variable latente qu'à une constante multiplicative près. De plus, on a vu que le seuil n'était pas non plus identifiable, car il se confond au terme constant dans x_i . Ainsi, dans ces modèles binaires, les paramètres β ne sont identifiés qu'à une constante additive (c) et une constante multiplicative ($1/\sigma$) près.

La seule information réellement utilisable est le signe des paramètres, indiquant si la variable associée influence la probabilité à la hausse ou à la baisse. Il est cependant possible, dans le cas de certaines variables explicatives qualitatives, d'utiliser la valeur des paramètres estimés pour classer les catégories. C'est le cas, par exemple, des variables indicatrices, dont les paramètres associés peuvent servir à classer les catégories en fonction de leur influence sur la probabilité.

De façon plus générale, on peut étudier l'impact des variables x_i sur la probabilité que $y_i = 1$, en calculant les effets marginaux.

B. Calcul des effets marginaux

Il est important en pratique d'avoir une idée de la sensibilité de la probabilité de l'événement $\{y_i = 1\}$ ou $\{y_i = 0\}$ par rapport à des variations dans les variables explicatives x_i . Cette sensibilité est généralement fournie par le calcul de l'*effet marginal*, c'est-à-dire la dérivée de la probabilité estimée par rapport aux composantes de x_i . Il peut sembler plus satisfaisant de considérer l'*élasticité* de la probabilité par rapport à x_i , celle-ci étant indépendante des unités de mesure. Alors que dans le modèle à probabilité linéaire LP, cette élasticité est proportionnelle au rapport entre la composante de x_i considérée : (x_{ik}) et

$Prob(y_i = 1)$:

$$\begin{aligned}\varepsilon_{Prob,x_{ik}} &= (\partial Prob(y_i = 1) / \partial x_{ik}) \times (x_{ik} / Prob(y_i = 1)) \\ &= \frac{\beta_k x_{ik}}{Prob(y_i = 1)} = \frac{\beta_k x_{ik}}{x_i \beta}\end{aligned}$$

dans les modèles Probit et Logit, on a les formules suivantes :

$$\varepsilon_{Prob,x_{ik}} = \frac{\partial \Phi(x_i \beta)}{\partial x_{ik}} \times \frac{x_{ik}}{\Phi(x_i \beta)} = \beta_k x_{ik} \frac{\phi(x_i \beta)}{\Phi(x_i \beta)} \quad (\text{Probit})$$

$$\varepsilon_{Prob,x_{ik}} = \frac{\partial \Lambda(x_i \beta)}{\partial x_{ik}} \times \frac{x_{ik}}{\Lambda(x_i \beta)} = \frac{x_{ik} \beta_k}{[1 + \exp(x_i \beta)]} \quad (\text{Logit})$$

Pour ces deux modèles, l'élasticité de la probabilité de l'événement $\{y_i = 1\}$ par rapport à une composante x_{ik} est une fonction non linéaire des autres composantes du vecteur x_i . Cela permet en particulier d'analyser l'influence de variables explicatives annexes sur la sensibilité du modèle à l'évolution d'une variable particulière.

Ces formules fournissent des mesures individuelles, qu'il est utile de compléter par des élasticités au point moyen de l'échantillon, afin de répondre à la question : quel est l'impact moyen (dans l'échantillon) de la variation d'une composante x_k ? Les expressions des élasticités ci-dessus doivent être modifiées de la façon suivante : on remplace les valeurs individuelles des x_i par les moyennes empiriques de ces composantes, calculées sur toutes les observations.

C. Test d'ajustement global

Dans les modèles de régression linéaire avec variable dépendante continue, il est usuel de tester l'hypothèse que les variables explicatives n'ont aucune influence sur les variations de la variable expliquée. Cette hypothèse revenant à tester la nullité globale des paramètres de pente, on utilise un test de Fisher. Dans le cas des modèles estimés par maximum de vraisemblance, un test analogue est employé, fondé sur le rapport des vraisemblances (voir le chapitre 2). Soit, par exemple, le modèle défini en termes de l'espérance conditionnelle de la variable binaire par rapport aux explicatives x_i :

$$E(y_i|x_i) = F(\beta_1 + x_2 \beta_2)$$

où $F(\cdot)$ est la fonction de répartition, x_2 est un vecteur $1 \times K$ de variables explicatives, sans le terme constant, et β_2 le vecteur $K \times 1$ des paramètres associés. Notons $\hat{\beta}_1, \hat{\beta}_2$ les paramètres estimés par maximum de vraisemblance. Sous l'hypothèse nulle $H_0 : \beta_2 = 0$, le modèle devient :

$$E(y_i|x_i) = F(\beta_1) = E(y_i)$$

et l'espérance conditionnelle de y_i est alors égale à l'espérance non conditionnelle, qui peut être estimée par la moyenne empirique $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$. Soit β_1 le paramètre associé à la constante et estimé sous l'hypothèse nulle. On a $\hat{y} = F(\hat{\beta}_1)$, et la log-vraisemblance du modèle contraint s'écrit alors :

$$\log L(\hat{\beta}_1, \beta_2 = 0) = N\bar{y}\log(\bar{y}) + N(1 - \bar{y})\log(1 - \bar{y})$$

Si $\log L(\hat{\beta}_1, \hat{\beta}_2)$ est la log-vraisemblance du modèle libre, la statistique de test du rapport des vraisemblances vaut :

$$2 [\log L(\hat{\beta}_1, \hat{\beta}_2) - \log L(\hat{\beta}_1, \beta_2 = 0)]$$

et est distribuée comme un $\chi^2(K)$ sous l'hypothèse nulle.

La statistique de test LR est intimement liée à la mesure de la qualité de l'ajustement du modèle. L'analogie du R^2 de la régression par moindres carrés est, dans le cas des modèles à variables qualitatives, le *pseudo-R²*, appelé aussi R^2 de McFadden, qui est calculé ainsi :

$$R^2 = 1 - \frac{\log L(\hat{\beta}_1, \hat{\beta}_2)}{\log L(\hat{\beta}_1, \beta_2 = 0)}$$

La log-vraisemblance $\log L(\hat{\beta}_1, \hat{\beta}_2)$ est constituée d'une somme de termes négatifs ou nuls ($\log F(\beta_1 + x_{i2}\beta_2) \leq 0$ et $y_i = 0$ ou 1). Le rapport ci-dessus est formé de deux termes négatifs ou nuls, le numérateur étant inférieur *en valeur absolue* au dénominateur (la log-vraisemblance d'un modèle contraint est toujours inférieure à celle du modèle libre). La statistique du pseudo- R^2 est par conséquent bien comprise entre 0 et 1. Les deux cas extrêmes sont :

- si le modèle contraint est correct, $L(\hat{\beta}_1, \hat{\beta}_2) = L(\hat{\beta}_1, \beta_2 = 0)$ et la statistique vaut 0 (les variations de y_i ne sont pas expliquées par les x_{ij});
- si le modèle non-contraint ajuste parfaitement les données, $\log L(\hat{\beta}_1, \hat{\beta}_2) = 0$ et le pseudo- R^2 vaut 1.

D. Critères de choix de modèles

Supposons que l'on cherche à comparer deux modèles concurrents estimés à partir d'un même échantillon et que ces modèles ne soient pas emboîtés. En d'autres termes, il n'est pas possible d'exprimer le second modèle comme l'analogie du premier pour certaines valeurs des paramètres. Deux cas sont envisageables :

- La forme fonctionnelle utilisée dans la spécification de la variable latente est différente ; pour l'un des modèles on suppose par exemple que $y_i^* = x_i\beta + u_i$, alors que dans le second modèle on adopte $y_i^* = \log(x_i)\beta + u_i$.

– La fonction de répartition est différente ; on veut par exemple comparer un modèle Probit à un modèle Logit.

Il n'est pas toujours possible dans la pratique de conclure à la supériorité d'un modèle sur un autre. Un modèle peut sembler plus satisfaisant en termes de significativité globale des paramètres, mais être moins performant en termes de prédition. Nous présentons une série de statistiques permettant de juger de la performance d'un modèle, en recommandant toutefois d'utiliser l'ensemble des statistiques de significativité, d'ajustement ou de prédition dans la comparaison entre deux modèles alternatifs, et de faire son choix selon l'objectif souhaité.

1) Nombre de prédictions fausses

On calcule $\sum_i^N (y_i - \hat{y}_i)$, où $\hat{y}_i = 1$ si $F(x_i \hat{\beta}) \geq 1/2$, $\hat{y}_i = 0$ sinon.

Lorsqu'un événement $y_i = 1$ se produit, un individu avec une probabilité estimée de 0,49 et un autre avec une probabilité estimée de 0,0001 sont identiquement pénalisés. Le désavantage de cette statistique est que la plupart des modèles seront acceptables au vu de ce critère, si la probabilité de l'événement ($y_i = 1$) est soit très forte, soit très faible.

2) Étude graphique des résidus

Une analyse des résidus du modèle peut également fournir des informations sur les performances du modèle en matière d'ajustement. Les trois types de résidus estimés, \hat{u}_i , sont :

$$\text{Résidu brut : } \hat{u}_i = y_i - F(x_i \hat{\beta})$$

$$\text{Résidu standardisé : } \hat{u}_i = \frac{y_i - F(x_i \hat{\beta})}{\sqrt{F(x_i \hat{\beta})(1 - F(x_i \hat{\beta}))}}$$

$$\text{Résidu généralisé : } \hat{u}_i = \frac{[y_i - F(x_i \hat{\beta})]f(x_i \hat{\beta})}{F(x_i \hat{\beta})(1 - F(x_i \hat{\beta}))}$$

Les résidus standardisés normalisent u_i par une estimation de leur écart type théorique. Les résidus généralisés proviennent directement des conditions du premier ordre (gradient égal à 0). Il est facile de voir que l'on peut calculer les composantes du gradient en multipliant les résidus généralisés par les différentes composantes de x_i ⁸.

8. Voir Chesher *et al.*, 1985.

3) Somme des carrés des résidus pondérée

$$\sum_i^N \frac{(y_i - F(x_i\hat{\beta}))^2}{F(x_i\beta)(1 - F(x_i\beta))}$$

Avantage : il est raisonnable d'attacher un poids plus fort à l'erreur de prédiction d'une observation avec une variance plus faible, car cette observation est plus simple à prédire qu'une observation avec une forte variance. La somme des carrés pondérée par la variance en tient compte, en normalisant les carrés des résidus par leur variance théorique estimée.

4) Critères d'information

Les critères d'information fournissent une mesure de la quantité d'information donnée par le modèle. Les critères d'Akaike, Schwarz et Hannan-Quinn ont été présentés au chapitre 2.

E. Le problème des variables omises

Considérons le modèle :

$$y_i^* = x_{i1}\beta_1 + x_{i2}\beta_2 + u_i \quad i = 1, 2, \dots, N,$$

où x_{i1} et x_{i2} sont des vecteurs $1 \times K_1$ et $1 \times K_2$, respectivement. On estime une version réduite du modèle où $\beta_2 = 0$, c'est-à-dire que les composantes de x_{i2} n'apparaissent pas dans la liste des variables explicatives. L'hypothèse $H_0 : \beta_2 = 0$ est un test de variables omises et il est bien connu que les estimations des paramètres du modèle restreint seront biaisées si les variables omises x_{i2} sont (non parfaitement) corrélées avec les variables incluses x_{i1} . Trois statistiques de test peuvent être utilisées : le multiplicateur de Lagrange (LM), la statistique de Wald et le rapport des vraisemblances (LR). Comme les expressions générales de ces tests ont été données au chapitre 2, nous ne commenterons que brièvement leur application aux modèles qualitatifs binaires.

Le test de Wald procède de la façon suivante. Le modèle complet (libre) est estimé par maximum de vraisemblance, ce qui fournit les vecteurs de paramètres $\hat{\beta}_1$ et $\hat{\beta}_2$, ainsi que la matrice de variance-covariance associée, $\hat{\Sigma}$. Si $\hat{\Sigma}_2$ est le bloc $K_2 \times K_2$ de $\hat{\Sigma}$ associé à $\hat{\beta}_2$, la statistique de test de Wald est $\hat{\beta}_2'\hat{\Sigma}_2^{-1}\hat{\beta}_2$, qui suit un $\chi^2(K_2)$ sous l'hypothèse nulle H_0 .

Pour le test du multiplicateur de Lagrange, on montre⁹ que la statistique de test LM est numériquement équivalente à NR^2 , où R^2 est le coefficient de

9. Voir Blundell, 1987.

détermination de la régression de :

$$w_i[y_i - F(x_i\beta)] \text{ sur } w_i f(x_i\beta) \times x_i$$

où

$$w_i = [F(x_i\beta)(1 - F(x_i\beta))]^{-1/2}$$

F. Test de l'hétéroscédasticité

On considère une hétéroscédasticité de forme générale¹⁰ :

$$\text{var}(\mu_i) = [\exp(z_i\gamma)]^2$$

où z_i est un vecteur de dimension $1 \times G$. L'hypothèse nulle d'homoscédasticité est $H_0 : \gamma = 0$. Cette spécification est compatible avec le modèle Probit seulement. Notons que les paramètres (β et γ) ne sont pas tous identifiables, pour la même raison que dans le Probit homoscédastique. On doit donc imposer que γ ne contient pas de terme constant. Les conditions du premier ordre de la maximisation de la log-vraisemblance sont :

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^N w_i \exp^{-z_i\gamma} x'_i$$

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^N w_i \exp^{-z_i\gamma} z'_i (-x_i\beta)$$

où :

$$w_i = \frac{f(x_i\beta)[y_i - F(x_i\beta)]}{F(x_i\beta)(1 - F(x_i\beta))}$$

La statistique de test de LM est aisément calculée sous l'hypothèse nulle H_0 . Il faut remarquer toutefois que ce test permet difficilement de discriminer l'existence d'hétéroscédasticité et la mauvaise spécification du modèle. En effet, les termes $x_i\beta$ et $\exp(z_i\gamma)$ apparaissent toujours sous forme de rapport dans la log-vraisemblance. Par exemple, la probabilité que $y_i = 1$ peut s'écrire :

$$F\left[\frac{x_i\beta}{\exp(z_i\gamma)}\right] = F(x_i\beta \exp^{-z_i\gamma}) = F\left[\sum_{k=1}^K x_{ik}\beta_k \exp\left(-\sum_{g=1}^G z_{ig}\gamma_g\right)\right]$$

Si le vrai modèle contient des termes croisés entre x_i et z_i , le test risque de détecter l'existence de variables omises (les z_i) plutôt que l'hétéroscédasticité.

10. Cette spécification est due à Harvey, 1976.

IV. Application : choix du mode de climatisation

Nous utilisons les estimateurs Probit et Logit présentés plus haut pour estimer la probabilité d'acquisition d'un climatiseur domestique pour un échantillon de ménages. Les variables observées sont les suivantes :

- *CENTRAL* : variable indicatrice, 1 si le ménage est équipé d'un climatiseur central (fixe) ;
- *PORTABLE* : variable indicatrice, 1 si le ménage est équipé d'un climatiseur portable (mobile) ;
- *REVENU* : revenu annuel du ménage (en log) ;
- *NBFOYER* : nombre de personnes du ménage ;
- *TAILLE* : surface de l'habitation ;
- *MAISON* : variable indicatrice, 1 si le ménage réside en logement individuel.

Le nombre d'observations est 137. Le revenu est exprimé en logarithme afin de réduire la dispersion de cette variable.

Le premier problème consiste à étudier l'impact des variables socio-économiques relatives au ménage sur la probabilité de posséder ou non un climatiseur central. Il sera possible d'en déduire le marché potentiel pour ce bien, à partir de la moyenne de ces variables dans la population totale.

Le tableau suivant donne les estimations du modèle par Probit binaire, où la variable *CENTRAL* est expliquée par les cinq autres. L'incorporation de *PORTABLE* dans les variables explicatives est motivée par le souhait d'analyser l'existence d'une substitution éventuelle entre les deux types de climatiseurs. Le logiciel utilisé est GAUSS, procédure Probit. Les commandes de création d'une table GAUSS et d'estimation Probit sont :

```
© La matrice data contient les variables central, portable, nbfoyer, taille,
revenu, maison ©
vnames={ "central", "portable", "nbfoyer", "taille", "revenu", "maison" } ; © Noms
des variables ©
call saved(data,"aircond",vnames);
© Enregistrement de la table "aircond" ©
iv={ nbfoyer, taille, revenu, maison, portable } ; © Variables explicatives ©
dv={ central } ; © Variable dépendante ©
vtype=1; © Variable numérique ©
qrstat=1; © Pour afficher les statistiques descriptives ©
qfit=1; © Pour les statistiques d'ajustement ©
title="Estimation Probit";
qrcatnm={ Y=0, Y=1 } ; © Nom des catégories ©
call probit("aircond",dv,iv); © Commande principale ©
```

PROBIT Version 3.1.5

Data Set : aircond

CASES PROCESSED BY PROBIT : 1

137 cases were kept out of 137 in file. 2

DEPENDENT CATEGORIES ARE DESIGNATED AS : 3

0 - $Y = 0$
1 - $Y = 1$

DISTRIBUTION AMONG OUTCOME CATEGORIES FOR central 4

	$Y = 0$	$Y = 1$
PROPORTION	0.6788	0.3212

DESCRIPTIVE STATISTICS (N=137) : 5

	Mean 6	Std. Dev. 7	Minimum 8	Maximum 9
nbfoyer	3.1752	1.6781	1.0000	11.0000
taille	1367.6788	659.0316	216.0000	4200.0000
revenu	9.3176	0.7383	7.2079	11.1664
maison	0.8467	0.3603	0.0000	1.0000
portable	0.3942	0.4887	0.0000	1.0000

Iteration : 1 2 3 4 10

ESTIMATES FROM PROBIT ANALYSIS OF VARIABLE : central 11

Convergence after 4 Newton-Raphson iterations. 12

Tolerance of 0.0000029 achieved after 0.00 minutes. 13

MEASURES OF FIT : 14

Variable	Comparison 15	Probit Estimate	Std. Error	t-value	p> t
CONSTANT	0/1	-7.13757	2.2930	-3.11	0.002
NBFOYER	0/1	-0.04309	0.0871	-0.49	0.621
TAILLE	0/1	0.00061	0.0003	2.13	0.033
REVENU	0/1	0.69431	0.2590	2.68	0.007
MAISON	0/1	-0.30394	0.3836	-0.79	0.428
PORTABLE	0/1	-1.03542	0.2959	-3.50	0.000

Likelihood Ratio Chi-square : 52.8737 16

with 5 d.f., prob = 0.000 17

-2 Log Likelihood for full model : 119.1289 18

-2 Log Likelihood for restricted model : 172.0026 19

Percent Correctly Predicted : 81.7518 20

TABLE OF OBSERVED AND PREDICTED OUTCOMES : [2]

Observed	Predicted			Total
	$Y = 0$	$Y = 1$		
$Y = 0$	85	8		93
$Y = 1$	17	27		44
Total	102	35		137

Le logiciel rappelle le nom du fichier GAUSS (*Data Set* : *aircond*) et le nombre d'observations traitées [1] et [2]. Il indique ensuite la désignation adoptée [3] et fournit en [4] les proportions empiriques de 0 et 1 de la variable dépendante (0,6788 ; 0,3212). Les statistiques descriptives des variables explicatives sont données en [5] : [6] donne la moyenne, [7] l'écart type, [8] et [9] le minimum et le maximum. Après affichage des itérations [10], le logiciel présente les résultats en [11] indiquant d'abord le nombre d'itérations [12], la tolérance et le temps écoulé avant convergence [13]. GAUSS donne ensuite le tableau des paramètres estimés. Les colonnes indiquent successivement le nom de la variable, le mode de comparaison [14] (0 contre 1), la valeur estimée du paramètre, son écart type estimé, la statistique de Student ($t - value$) et la probabilité de dépasser la valeur absolue de la statistique de Student (la *p-value*, $p > |t|$).

Le tableau suivant [15] fournit les mesures d'ajustement du modèle estimé : statistique de test du rapport des vraisemblances [16] et degrés de liberté du test [17], valeurs des log-vraisemblances (multipliées par -2) libre [18] et contrainte [19]. L'hypothèse nulle est que tous les paramètres sont nuls excepté la constante (voir la section 3.III.C). Une indication importante est le pourcentage de prévisions correctes [20]. Ce dernier peut être calculé en étudiant le tableau des effectifs observés et prévus [21], où les effectifs prévus (*Predicted*) sont en colonne, les effectifs observés (*Observed*) en ligne. Le nombre de prédictions fausses s'obtient donc en additionnant les valeurs hors-diagonale, et en comparant avec le nombre d'observations totales. On a $(17 + 8)/137 = 0,1824$ soit $1 - 0,8176$ où 0,8176 est le pourcentage d'observations correctement prédictes.

Le tableau ci-dessous donne les estimations du modèle Logit binaire, obtenues avec le logiciel EVIEWS.

Une fois le fichier de travail créé (*Workfile*), les commandes consistent à entrer l'équation à estimer dans la boîte de dialogue *Equation Selection* : *central c nbfoyer taille revenu maison portable* (*c* est pour le terme constant) et à sélectionner *Logit*.

Dependent Variable : CENTRAL [1]
 Method : ML - Binary Logit [2]
 Date : 05/06/99 Time : 10:27
 Sample : 1 137 [3]
 Included observations : 137 [4]

Convergence achieved after 7 iterations [5]

Covariance matrix computed using second derivatives [6]

Variable [7]	Coefficient [8]	Std. Error [9]	z-Statistic [10]	Prob. [11]
C	-12.04714	4.156458	-2.898415	0.0038
NBFOYER	-0.064585	0.149259	-0.432709	0.6652
TAILLE	0.001099	0.000512	2.149067	0.0316
REVENU	1.158715	0.464816	2.492847	0.0127
MAISON	-0.494055	0.684706	-0.721558	0.4706
PORTABLE	-1.806437	0.535372	-3.374169	0.0007

EVIEWS rappelle le nom de la variable dépendante [1], le modèle (Logit binaire estimé par maximum de vraisemblance [2]), l'étendue de l'échantillon et sa taille ([3], [4]). Le logiciel indique ensuite le nombre d'itérations avant convergence [5] et la méthode de calcul de la matrice de variance-covariance ([6], matrice calculée avec les expressions des dérivées secondes, voir le chapitre 2). Le tableau des paramètres estimés fournit le nom de la variable [7], son coefficient estimé [8], l'écart type [9], la statistique de Student [10] et la p-value [11].

Mean dependent var [12]	0.321168	S.D. dependent var [13]	0.468639
S.E. of regression [14]	0.382572	Akaike info criterion [15]	0.957688
Sum squared resid [16]	19.17337	Schwarz criterion [17]	1.085571
Log likelihood [18]	-59.60166	Hannan-Quinn criter. [19]	1.009657
Restr. log likelihood [20]	-86.00129	Avg. log likelihood [21]	-0.435049
LR statistic (5 df) [22]	52.79926	McFadden R-squared [23]	0.306968
Probability(LR stat) [24]	3.70E-10		

Obs with Dep=0 93 Total obs 137 [25]

Obs with Dep=1 44

EVIEWS indique ensuite les statistiques sur la variable dépendante : moyenne [12] et écart type [13]. On vérifie bien que la proportion empirique de *central* = 1 est de 32,11 %. Les statistiques d'ajustement du modèle sont : écart type de la variable dépendante estimée [14], critères d'information d'Akaike [15], de Schwarz [17] et de Hannan-Quinn [19], somme des carrés des résidus [16]. Suivent la valeur de la log-vraisemblance totale [18], puis divisée le nombre d'observations [21], la log-vraisemblance contrainte [20] sous l'hypothèse de nullité de tous les paramètres sauf la constante, ce qui permet de calculer le test du rapport des vraisemblances avec 5 degrés de liberté [22] et la p-value associée [24]. Le coefficient R^2 de McFadden est donné en [23].

Enfin, le logiciel rapporte que 93 observations ont $Y = 1$ et 44 ont $Y = 0$, sur un total de 137 observations ([25]).

V. Extensions

Les modèles à variables qualitatives binaires ont connu des développements théoriques récents, portant sur les méthodes d'estimation et sur un certain nombre de généralisations. Nous présentons brièvement ici les modèles bivariés comme cas particulier des systèmes d'équations simultanées avec variables qualitatives, et donnons un aperçu de quelques méthodes alternatives d'estimation, basées sur des techniques non paramétriques. Nous terminons cette section par le cas des modèles avec données de panel.

A. Les modèles binaires bivariés

Il est intéressant dans la pratique de pouvoir estimer la probabilité de deux événements simultanés. Par exemple, un consommateur se voit proposer deux biens complémentaires et l'on cherche la probabilité qu'il les acquière tous les deux ou bien qu'il achète le second sachant qu'il détient déjà le premier. Un autre exemple peut être la probabilité qu'une entreprise industrielle soit présente simultanément sur deux marchés à l'export ou sur un seul.

Le modèle Probit se généralise aisément au cas de deux équations, permettant d'envisager des modèles à équations simultanées analogues à ceux de la régression linéaire. Considérons par exemple le modèle suivant :

$$\begin{cases} y_{i1}^* = x_{i1}\beta_1 + u_{i1} \\ y_{i2}^* = x_{i2}\beta_2 + u_{i2} \end{cases}$$

où x_{i1} est un vecteur $1 \times K_1$, x_{i2} un vecteur $1 \times K_2$. Les deux variables latentes y_{i1}^* et y_{i2}^* sont expliquées par des combinaisons linéaires de x_{i1} et x_{i2} respectivement, et l'on observe les variables y_{i1} et y_{i2} telles que : $y_{i1} = 1$ si $y_{i1}^* > 0$, 0 sinon ; $y_{i2} = 1$ si $y_{i2}^* > 0$, 0 sinon.

Quatre situations sont alors possibles : $(y_{i1} = 1, y_{i2} = 1)$, $(y_{i1} = 1, y_{i2} = 0)$, $(y_{i1} = 0, y_{i2} = 1)$ et $(y_{i1} = 0, y_{i2} = 0)$. Ce modèle bivarié permet d'estimer les probabilités de deux événements simultanés, une fois que la loi conjointe des termes u_{i1} et u_{i2} est spécifiée. Pour le Probit bivarié, on suppose que ces derniers sont normalement et conjointement distribués, avec :

$$E(u_{i1}) = E(u_{i2}) = 0, \quad Var(u_{i1}) = Var(u_{i2}) = 1, \quad Cov(u_{i1}, u_{i2}) = \rho$$

où $\rho \in [0, 1]$ est le coefficient de corrélation entre u_{i1} et u_{i2} . Notons que les variances sont normalisées à 1, comme dans le Probit univarié. La fonction

de répartition associée à la loi normale bivariée avec variances égales à 1, notée $\Phi_2(\bar{u}_1, \bar{u}_2, \rho)$, est :

$$\boxed{\Phi_2(\bar{u}_1, \bar{u}_2, \rho) = \text{Prob}(u_{11} < \bar{u}_1, u_{12} < \bar{u}_2) = \int_{-\infty}^{\bar{u}_1} \int_{-\infty}^{\bar{u}_2} \phi_2(u_1, u_2, \rho) du_1 du_2}$$

où $\phi_2(u_1, u_2, \rho)$ est la densité de la loi normale bivariée définie par :

$$\phi_2(u_1, u_2, \rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp \left\{ \frac{-1/2(u_1^2 + u_2^2 - 2\rho u_1 u_2)}{(1-\rho^2)} \right\}$$

La fonction de vraisemblance pourrait être construite en calculant les probabilités associées aux quatre cas plus haut. Il existe cependant une construction plus compacte, utilisant comme précédemment une nouvelle variable $\delta_{ij}, j = 1, 2$ telle que $\delta_{ij} = 2y_{ij} - 1$; on a alors $\delta_{ij} = +1$ si $y_{ij} = 1$ et $\delta_{ij} = -1$ si $y_{ij} = 0, j = 1, 2$. La log-vraisemblance s'écrit :

$$\log L = \sum_{i=1}^N \log \Phi_2(\delta_{i1}x_{i1}\beta_1, \delta_{i2}x_{i2}\beta_2, \delta_{i1}\delta_{i2}\rho)$$

Notons la façon dont le coefficient de corrélation apparaît dans l'écriture des probabilités. Si $(y_{i1} = 0, y_{i2} = 1)$ ou $(y_{i1} = 1, y_{i2} = 0)$, le terme croisé dans la densité évaluée en $(\delta_{i1}x_{i1}\beta_1, \delta_{i2}x_{i2}\beta_2)$ n'aura pas le signe correct. Pour récupérer le bon signe, ρ est multiplié par $\delta_{i1}\delta_{i2} = -1$ dans l'un des deux cas ci-dessus.

La vraisemblance est ensuite maximisée numériquement par rapport aux paramètres (β_1, β_2, ρ) . Il est aisément vérifiable que, dans le cas $\rho = 0$, les conditions du premier ordre de maximisation de la log-vraisemblance par rapport à $\beta_j, j = 1, 2$ sont identiques à celles du cas univarié. Si les termes u_{11} et u_{12} ne sont pas corrélés, la densité bivariée ϕ_2 est égale au produit des densités marginales, et on a $\Phi_2(\delta_{i1}x_{i1}\beta_1, \delta_{i2}x_{i2}\beta_2, \delta_{i1}\delta_{i2}\rho) = \Phi(\delta_{i1}x_{i1}\beta_1) \times \Phi(\delta_{i2}x_{i2}\beta_2)$, où $\Phi(\cdot)$ est la fonction de répartition de la loi normale univariée. Pour tester l'hypothèse nulle $H_0 : \rho = 0$, on peut utiliser une statistique de test LM (Kiefer, 1982), ou, plus simplement, recourir à la statistique du t de Student, construite à partir de la matrice de variance-covariance des estimateurs par maximum de vraisemblance.

Le calcul des effets marginaux est compliqué par le fait que certaines variables explicatives peuvent être communes à x_{i1} et x_{i2} . Voir Greene (1997) pour l'expression des effets marginaux relatifs à des variables apparaissant éventuellement dans les deux équations.

B. Application

Revenons à notre exemple de choix du mode de climatisation. Certains ménages possédant les deux types de climatiseurs, on considère maintenant la probabilité conjointe associée à la détention d'un climatiseur central et portable, à l'aide d'un modèle Probit bivarié. On présente ci-dessous les résultats d'estimation à l'aide du logiciel LIMDEP où l'on utilise les mêmes variables explicatives pour les deux types de climatiseurs :

CONSTANTE, NBFOYER, TAILLE, REVENU, MAISON.

Les commandes sous LIMDEP permettant de lire le fichier et d'estimer le modèle bivarié sont :

```
read ;file=c :\dunod\exemples\aircond.dat ;nobs=137 ;nvar=6 ;
names=nbfoy,taille,revenu,central,portable,maison$  
bivariate ;lhs=central,portable ;rhi=one,nbfoy,taille,revenu,maison ;
rh2=one,nbfoy,taille,revenu,maison$
```

Le logiciel fournit le nombre d'observations [1] et [2], indique que les estimations du Probit bivarié sont obtenues par le maximum de vraisemblance en information complète (FIML) [3] et [4]. La ligne [5] rappelle le nom des variables dépendantes en abrégé, le nombre d'itérations [6], la valeur de la log-vraisemblance [7]. Le tableau des estimations donne le nom de la variable [8], le paramètre estimé [9], son écart type [10], la statistique de Student [11], la p-value associée [12], et la moyenne empirique de la variable explicative [13].

Les paramètres estimés sont donnés successivement pour les deux équations *central* ([14]) puis *portable* ([15]), et l'estimation du coefficient de corrélation entre les deux équations (ρ) est donnée en [16].

Le logiciel fournit enfin en [17] le tableau croisé des observations observées et prédites (entre parenthèses). L'équation pour *portable* est en colonne, celle pour *central* en ligne, et [18] rappelle que les nombres entre parenthèses indiquent les effectifs prédits (*Count of Fitted Values*). On voit ainsi que 6 ménages ont à la fois un climatiseur central et portable, et que le nombre estimé de combinaisons (1, 1) est de 0. De même, 38 ménages n'ont pas de portable mais un central, et le modèle en prédit 40.

```
LIMDEP Estimation Results Run log line 10 Page 2
Current sample contains 137 observations [1]
FIML Estimates of Bivariate Probit Model [2]
Maximum Likelihood Estimates [3]
Dependent variable CENPOR [4]
Number of observations 137 [5]
Iterations completed 16 [6]
Log likelihood function -147.8441 [7]
```

Variable	Coefficient	Standard Error	$z=b/\text{se}$	$P Z> z $	Mean of X
Index equation for CENTRAL [14]					
Constant	-7.007258	2.4574	-2.851	0.00435	
TAILLE	0.6947615E-03	0.29248E-03	2.375	0.01753	1368.
NBFOYER	-0.9421566E-02	0.11096	-0.085	0.93233	3.175
REVENU	0.6188307	0.28167	2.197	0.02802	9.318
MAISON	-0.2775645	0.35782	-0.776	0.43792	0.8467
Index equation for PORTABLE [15]					
Constant	0.3857295	1.6011	0.241	0.80962	
TAILLE	-0.4367700E-03	0.24490E-03	-1.783	0.07452	1368.
NBFOYER	-0.8802204E-01	0.81040E-01	-1.086	0.27741	3.175
REVENU	0.1871548E-01	0.18334	0.102	0.91869	9.318
MAISON	0.3389939E-01	0.33216	0.102	0.91871	0.8467
Disturbance correlation and std. dev. [16]					
RHO(1,2)	-0.5648072	0.13857	-4.076	0.00005	

Joint Frequency Table : Columns = PORTABLE [17]

Rows = CENTRAL

(N) = Count of Fitted Values [18]

	0	1	TOTAL
0	45 (39)	48 (58)	93 (97)
1	38 (40)	6 (0)	44 (40)
TOTAL	83 (79)	54 (58)	137 (137)

C. Méthodes d'estimation non paramétriques

Nous avons vu au chapitre 2 que l'estimateur du maximum de vraisemblance était efficace au sens de Cramer-Rao, si le modèle était bien spécifié. En particulier, la distribution que l'on impose doit correspondre à la loi qui a engendré les observations. Comme il est difficile dans la pratique de s'assurer d'une telle condition, on peut se demander si les méthodes décrites plus haut, basées sur la distribution normale ou logistique, conduisent à des estimations cohérentes des paramètres β . Une possibilité intéressante et qui a connu récemment de nombreux développements consiste à faire abstraction de toute distribution statistique paramétrique pour les termes d'erreur : ce sont les méthodes d'estimation non paramétriques. Nous en présenterons brièvement deux : la méthode du *score maximum* et une méthode semi-paramétrique.

1) La méthode du score maximum

Cette méthode consiste à maximiser par rapport à β un critère constitué du nombre de fois où $x_i\beta > 0$ lorsque $y_i = 1$, et du nombre de fois où $x_i\beta < 0$

lorsque $y_i = 0$:

$$\max_{\beta} \frac{1}{N} \sum_{i=1}^N [\mathbb{I}(y_i = 1) \times \mathbb{I}(x_i\beta > 0) + \mathbb{I}(y_i = 0) \times \mathbb{I}(x_i\beta < 0)]$$

où $\mathbb{I}(.)$ est la fonction indicatrice.

La motivation de la méthode est la suivante. On peut faire l'hypothèse que la vraie distribution (inconnue) F est de médiane 0 (c'est-à-dire $F(0) = 0.5$). Pour des valeurs faibles de u_i par rapport à $x_i\beta$, on aura $Prob(y_i = 1) \approx Prob(x_i\beta > 0)$ et $Prob(y_i = 0) \approx Prob(x_i\beta \leq 0)$. Comme l'événement à modéliser n'admet que deux issues (0 ou 1), il est équivalent de considérer une prédiction correcte comme : $\{y_i = 1 \text{ quand } x_i\beta > 0\}$ ou $\{y_i = 0 \text{ quand } x_i\beta \leq 0\}$, ou encore : le signe de la prédiction $x_i\beta$ est le même que celui de la variable $\delta_i(y_i)$ utilisée précédemment¹¹.

Une interprétation équivalente de la méthode du score maximum est donc qu'elle compare le *signe* de la prédiction, c'est-à-dire de $x_i\beta$, avec celui de la nouvelle variable δ_i . Une autre écriture possible du problème est :

$$\max_{\beta} \frac{1}{N} \sum_{i=1}^N [\delta_i sgn(x_i\beta)]$$

où $sgn(x_i\beta) = +1$ si $x_i\beta > 0$, $sgn(x_i\beta) = -1$ sinon.

Notons que la méthode du score maximum n'est pas basée sur une quelque vraisemblance, et que le critère à maximiser n'est pas continûment dérivable. Par conséquent, le calcul des écarts types des paramètres estimés ne peut se faire avec les méthodes usuelles. Une possibilité consiste à calculer les variances des paramètres β par la méthode du *bootstrap*, voir Greene (1997).

L'information fournie par la méthode du score maximum est limitée, et l'estimateur est généralement très inefficace par rapport au maximum de vraisemblance. Il n'est pas possible, par exemple, de déduire analytiquement les effets marginaux à partir des estimations, puisque la fonction F reste inconnue. De plus, ne pas imposer de distribution *a priori* ne signifie pas que le score maximum sera plus satisfaisant en termes de prédiction.

2) Estimation semi-paramétrique

Ces inconvénients ont conduit certains auteurs à proposer une méthode d'estimation basée sur des procédures semi-paramétriques¹². L'idée de la méthode est de séparer le modèle économétrique en deux : une partie paramétrique constituée de $x_i\beta$, et une partie non paramétrique constituée de la fonction de répartition $F(.)$, d'où l'appellation de modèle *semi-paramétrique*.

11. Rappelons que $\delta_i = 2y_i - 1$.

12. Voir Klein et Spady, 1993.

Considérons tout d'abord la fonction $F(x_i\beta)$ qui n'est autre que l'espérance conditionnelle de y_i sachant x_i . Cette dernière peut s'écrire :

$$E(y_i|x_i) = \int y_i [f(y_i, x_i)/f(x_i)] dy_i = r(x_i)$$

où $r(\cdot)$ est appelée la *fonction de lien*.

La probabilité associée à y_i en tout point x_i^0 peut être estimée par la moyenne pondérée :

$$E(y_i|x_i^0) = \frac{\sum_{j=1}^N w_j(x_j^0) y_j}{\sum_{j=1}^N w_j(x_j^0)}$$

où :

$$w_j(x_j^0) = K((x_j - x_j^0)/h)$$

$K(\cdot)$ est le noyau, h la fenêtre. Cet opérateur fournit une mesure de la distance entre le point considéré et n'importe quel autre point de l'échantillon (x_i). Le noyau (kernel) est une fonction continue, symétrique autour de 0, intégrant à 1, et nulle pour de grandes valeurs de son argument¹³. La fenêtre sert à calibrer cette distance, en pénalisant plus ou moins les poids éloignés de x_i^0 . Si h est petit, d'après les propriétés du noyau ci-dessus, l'opérateur w_j privilégie les points proches de x_i^0 . Le choix de la fenêtre h est effectué selon un arbitrage entre le biais de l'estimateur non paramétrique et sa variance. Une fenêtre « petite » fournira un biais plus faible mais un estimateur moins efficace (de variance plus grande), alors qu'une fenêtre plus large s'approchera de l'estimation par moindres carrés linéaires (dans laquelle tous les points sont pris en compte).

Ainsi, toute la fonction de répartition F peut être reconstruite à partir de la formule ci-dessus, pour chaque point x_i , $i = 1, 2, \dots, N$. Dans ce cas, la méthode s'applique très facilement, en deux étapes.

Première étape : obtenir un estimateur de β et construire un indice $z_i = x_i\beta$. Cet estimateur peut être celui obtenu par la méthode du score maximum.

Deuxième étape : estimer la fonction de lien au voisinage de $z_i^0 = x_i^0\beta$ par :

$$r(z_i^0) = E(y_i|x_i^0\beta) = \frac{\sum_{j=1}^N w_j(z_j^0) y_j}{\sum_{j=1}^N w_j(z_j^0)}$$

Il est bien sûr possible d'employer une méthode itérative consistant à estimer alternativement β et la fonction F :

- partir d'une valeur donnée $z_i = x_i\beta$;
- calculer la fonction de répartition estimée $\hat{F}(z_i)$ en tout point $i = 1, 2, \dots, N$ par la méthode du noyau ;

13. Un noyau couramment utilisé est le noyau gaussien : $K(X) = (2\pi)^{-1/2} \exp(-X^2/2)$.

– maximiser la log-vraisemblance

$$\sum_i [y_i \log \hat{F}(x_i \beta) + (1 - y_i) \log[1 - \hat{F}(x_i \beta)]]$$

par rapport à β ; on note $\hat{\beta}$ l'estimateur obtenu ;

– remplacer \bar{z}_i par la nouvelle valeur $x_i \hat{\beta}$ et retourner à la deuxième étape.

D. Modèles avec données de panel

Nous étudions maintenant l'utilisation des modèles univariés Probit et Logit dans le cas de données de panel. Rappelons que dans un échantillon en panel, les variables expliquée et explicatives sont observées pour un même individu à différentes dates successives. La variable latente et les variables explicatives s'écrivent à présent en fonction de deux indices i et t , où i est l'indice de l'individu, t celui de la période :

$$y_{it}^* = x_{it} \beta + u_{it} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T$$

Comme les individus sont présents plusieurs fois dans l'échantillon, il est raisonnable de supposer que le résidu du modèle peut se décomposer en deux :

$$u_{it} = \alpha_i + \varepsilon_{it}$$

où α_i est un terme aléatoire propre à l'individu, et ε_{it} un terme d'erreur indépendamment et identiquement distribué entre les individus et les périodes.

Les deux modèles à variables qualitatives les plus courants dans ce contexte sont le Probit avec effets aléatoires et le Logit avec effets fixes. Dans le cas du modèle Probit, on suppose $E(\alpha_i) = E(\varepsilon_{it}) = 0 \forall i, t$ et $var(\alpha_i) = \sigma_a^2$, et la variance de ε_{it} est normalisée à 1. En raison de l'existence du terme individuel α_i , les termes u_{it} sont corrélés entre eux pour un même individu. On suppose donc que les erreurs $(u_{i1}, u_{i2}, \dots, u_{iT})$ sont conjointement normalement distribuées, ce qui implique de modéliser la probabilité jointe associée aux réalisations des variables binaires observées $(y_{i1}, y_{i2}, \dots, y_{iT})$. On a, de façon analogue au modèle binaire usuel :

$$Prob(y_{it} = 1) = Prob(y_{it}^* > 0) = Prob(u_{it} > -x_{it} \beta) \quad \forall i, t$$

et pour le groupe d'observations $(y_{i1}, y_{i2}, \dots, y_{iT})$:

$$Prob(y_{i1}, \dots, y_{iT}) = \int_{-\infty}^{\delta_{i1} x_{i1} \beta} \dots \int_{-\infty}^{\delta_{iT} x_{iT} \beta} \phi_T(u_{i1}, \dots, u_{iT}) du_{i1} \dots du_{iT}$$

où, comme précédemment, $\delta_{it} = 2y_{it} - 1$. $\phi_T(\cdot)$ est la densité de la loi T -variée (normale) du groupe i .

L'évaluation d'une telle intégrale étant pratiquement impossible pour $T > 3$, une solution consiste alors à réduire le problème à une seule intégrale.

Butler et Moffitt (1982) montrent que la contribution à la vraisemblance du groupe i est :

$$L(y_{i1}, y_{i2}, \dots, y_{iT}) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\alpha_i^2} \left[\prod_{t=1}^T \Phi \left(\delta_{it} x_{it} \beta + \delta_{it} \alpha_i - \frac{\sqrt{2\rho}}{\sqrt{1-\rho}} \right) \right] d\alpha_i$$

Cette contribution peut être évaluée par intégration numérique et la log-vraisemblance totale est calculée en prenant la somme des logarithmes de $L(y_{i1}, y_{i2}, \dots, y_{iT})$ pour $i = 1, 2, \dots, N$.

Contrairement au Probit, le modèle Logit admet une représentation satisfaisante avec effets fixes. Les probabilités associées aux variables latentes positives s'écrivent alors en fonction des effets individuels α_i :

$$\text{Prob}(y_{it}^* > 0) = \text{Prob}(y_{it} = 1) = \frac{\exp^{\alpha_i \beta + \alpha_i}}{1 + \exp^{\alpha_i \beta + \alpha_i}}$$

Les termes d'hétérogénéité individuelle α_i sont des paramètres de nuisance dont le nombre s'accroît avec le nombre d'individus dans l'échantillon.

Une solution (Chamberlain, 1980) consiste à éliminer l'effet individuel en maximisant la log-vraisemblance *conditionnelle* suivante, le conditionnement étant effectué par rapport à cette statistique suffisante :

$$\log L = \sum_{i=1}^N \log \left[\text{Prob} \left(y_{i1}, y_{i2}, \dots, y_{iT} \mid \sum_{t=1}^T y_{it} \right) \right]$$

Prenons par exemple le cas $T = 2$. Puisque :

$$\text{Prob}(y_{i1} = 0, y_{i2} = 0 \mid y_{i1} + y_{i2} = 0) = \text{Prob}(y_{i1} = 1, y_{i2} = 1 \mid y_{i1} + y_{i2} = 2) = 1$$

les termes de ce type n'interviennent pas dans la log-vraisemblance car $\log(1) = 0$, mais les termes tels que $y_{i1} + y_{i2} = 1$ jouent un rôle. En utilisant le modèle Logit, on voit alors que les effets individuels peuvent se simplifier dans l'écriture des probabilités :

$$\text{Prob}(y_{i1} = 0, y_{i2} = 1 \mid y_{i1} + y_{i2} = 1) = \frac{\exp^{\alpha_1 \beta}}{\exp^{\alpha_1 \beta} + \exp^{\alpha_2 \beta}}$$

$$\text{Prob}(y_{i1} = 1, y_{i2} = 0 \mid y_{i1} + y_{i2} = 1) = \frac{\exp^{\alpha_2 \beta}}{\exp^{\alpha_1 \beta} + \exp^{\alpha_2 \beta}}$$

La log-vraisemblance est alors maximisée avec les techniques usuelles. L'hypothèse d'existence d'effets individuels peut être évaluée par la procédure

du test de Hausman (voir le chapitre 2), en considérant le fait que l'estimateur est convergent sous l'hypothèse nulle ou l'hypothèse alternative. En revanche, le modèle Logit estimé sans référence aux effets individuels ne sera convergent que sous l'hypothèse nulle ; dans ce dernier cas, il sera aussi plus efficace que l'estimateur de Chamberlain.

À retenir

La manière correcte de modéliser les événements associés à une variable dépendante qualitative consiste à spécifier un modèle sous-jacent, à variable latente. Les deux modèles les plus utilisés sont le Probit et le Logit, basés sur la distribution normale et logistique. La méthode d'estimation est le maximum de vraisemblance, à partir de laquelle des critères d'ajustement et de choix de modèles sont aisément obtenus. Ces deux modèles donnent des résultats sensiblement équivalents dans la pratique, et peuvent être généralisés aux données de panel et à un cadre bivarié. Les méthodes d'estimation non paramétriques constituent une alternative intéressante permettant de s'affranchir de toute hypothèse distributionnelle.

Exercices

Exercice n° 1 : modèles Logit et Probit binaires

- 1) Reprenons les estimations des modèles Probit (par GAUSS) et Logit (par EVIEWS) du modèle de choix de climatiseur (section 3.IV). Pour comparer les valeurs estimées des paramètres par les deux procédures, utiliser les règles de conversion données dans le texte. Les valeurs numériques obtenues sont-elles sensiblement les mêmes ?
- 2) À partir des paramètres du Logit, calculer la probabilité qu'un ménage de 3 personnes, en maison individuelle, de revenu 10, de taille d'habitation 1360 et ne possédant pas de climatiseur portable, achète un climatiseur central.
- 3) Calculer l'effet marginal associé à la variable *REVENU*, au point moyen de l'échantillon, à partir des estimations Logit. Utiliser pour cela les moyennes des variables explicatives dans le tableau des statistiques descriptives.
- 4) Toujours avec les résultats du Logit, calculer l'effet marginal associé à la variable *PORTABLE*, au point moyen de l'échantillon. Que pouvez-vous dire de la substitution entre climatiseur central et climatiseur portable ? Si l'on cherche à évaluer l'impact du revenu sur l'effet marginal de la variable *PORTABLE*, quelle mesure peut-on proposer ?

Exercice n° 2 : décision d'abonnement

L'échantillon suivant se rapporte à 15 personnes interrogées sur leur volonté ou non de s'abonner à un journal. Les variables sont : *SEXE* (M pour homme, F pour femme), *AGE* et *ABON*, égale à 1 si la personne désire s'abonner, 0 sinon.

Obs.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
SEXÉ	F	M	M	F	F	F	M	M	F	F	F	F	M	F	F
AGE	35	44	45	47	51	47	54	47	35	34	48	56	46	59	46
ABON	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0

- 1) Construire une variable binaire (0, 1) à partir de la variable *SEXÉ*. On notera par convention cette variable *FEMME*, avec *FEMME* = 1 si *SEXÉ* = F et 0 sinon. Calculer la proportion de personnes acceptant l'abonnement, et la comparer avec la proportion de femmes dans l'échantillon. Tester par une statistique χ^2 (voir le chapitre 1) si le sexe de la personne intervient significativement dans la décision.
- 2) Estimer la probabilité de s'abonner par la procédure Probit, avec *FEMME* et *AGE* comme variables explicatives. Quelles sont les variables significatives ? Calculer la statistique d'ajustement global du pseudo-*R*² à partir des log-vraisemblances données par votre logiciel. Calculer le pourcentage de prédictions fausses.
- 3) Estimer les mêmes paramètres par moindres carrés ordinaires, avec le modèle à probabilité linéaire (LP) du chapitre 1. Comparer les résultats avec les estimations par Probit. Pour quelle valeur de la variable *AGE* a-t-on une probabilité d'abonnement supérieure à 1 avec ce modèle LP ? Et une probabilité négative ?
- 4) Supposons que l'on cherche à présent l'âge moyen noté \bar{AGE} pour lequel la probabilité d'abonnement est de 40 %, pour les femmes uniquement. Le modèle sous-jacent s'écrit :

$$y_i^* = \gamma_0 + \gamma_1 AGE_i + \gamma_2 FEMME_i + u_i$$

avec $var(u_i) = 1$.

Calculer alors la valeur de la variable \bar{AGE} telle que $Prob(y_i^* > 0) = 0,4$. Indication : utiliser une table des valeurs de la loi normale pour trouver une valeur approchée.

Exercice n° 3 : accès au crédit

On dispose d'un échantillon de 100 ménages ayant eu accès à un crédit auprès d'un organisme financier. Parmi ces 100 ménages, 29 ont disposé d'un crédit à un taux avantageux, et 71 à un taux élevé. L'objectif est de repérer les déterminants des ménages ayant conduit à cette répartition. La variable expliquée est *AVANTAGE*, égale à 1 si le crédit est à un taux avantageux, 0 sinon.

L'enquête comprend également des informations sur la recherche préalable de crédit par les ménages, auprès de différents organismes financiers.

Les variables explicatives disponibles sont :

REVENU : revenu du ménage ;

REJET : indicatrice égale à 1 si le ménage s'est vu auparavant refuser ce crédit lors de sa recherche de financement ;

AVANT : indicatrice égale à 1 si le ménage a eu recours dans le passé à l'organisme de crédit qui vient de l'accepter ;

MONTANT : montant du crédit demandé et accepté ;

LOCAL : indicatrice égale à 1 si le ménage réside dans la région.

Une estimation par le modèle Logit simple sur toutes les variables explicatives donne les résultats suivants :

Tableau 3.1 — Modèle complet, estimation Logit

Variable	Paramètre estimé	Écart type	t de Student	p-value
Constante	-1,4971	0,7583	-1,97	0,048
<i>REVENU</i>	0,3943	0,2372	1,66	0,096
<i>REJET</i>	-0,5083	0,5302	-0,96	0,338
<i>AVANT</i>	1,5921	0,5142	3,10	0,002
<i>MONTANT</i>	0,4146	0,1602	2,59	0,010
<i>LOCAL</i>	0,6484	0,6279	1,03	0,302

Log-vraisemblance du modèle complet : -47,8768
 Log-vraisemblance du modèle restreint : -60,2151

Observé/Prédit	1	0	Total
1	11	18	29
0	9	62	71
Total	20	80	100

- 1) Interpréter les estimations obtenues (signe des paramètres, significativité, ajustement global du modèle). Calculer le pourcentage de prédictions fausses.
- 2) On désire à présent tester deux modèles concurrents. Le premier, dénoté modèle A, explique la probabilité d'obtenir un crédit à taux avantageux (*AVANTAGE* = 1) en fonction de variables relatives à l'expérience passée du ménage en matière de demande de crédit (*AVANT* et *REJET*). Le second modèle, dénoté modèle B, se base au contraire sur les caractéristiques du ménage et les variables relatives au prêt demandé (*MONTANT*, *REVENU*, *LOCAL*).

Les tableaux 3.2 et 3.3 ci-dessous donnent les résultats des estimations pour les modèles A et B respectivement.

Tableau 3.2 — Modèle A : historique de la demande de crédit, estimation Logit

Variable	Paramètre estimé	Écart type	t de Student	p-value
Constante	0,3303	0,3824	0,86	0,388
<i>AVANT</i>	1,4953	0,4778	3,13	0,002
<i>REJET</i>	-0,6794	0,4925	-1,38	0,168

Log-vraisemblance du modèle complet : -53,3374
 Log-vraisemblance du modèle restreint : -60,2151

Observé/Prédit	1	0	Total
1	9	20	29
0	7	64	71
Total	16	84	100

Tableau 3.3 — Modèle B : Variables du ménage relatives au crédit,
estimation Logit

Variable	Paramètre estimé	Écart type	<i>t</i> de Student	p-value
Constante	-0,9072	0,6043	-1,50	0,133
<i>MONTANT</i>	0,3868	0,1461	2,65	0,008
<i>REVENU</i>	0,4774	0,2250	2,12	0,034
<i>LOCAL</i>	0,4444	0,5788	0,77	0,443
Log-vraisemblance du modèle complet : -53,8374				
Log-vraisemblance du modèle restreint : -60,2151				

Observé/Prédit	1	0	Total
1	10	19	29
0	3	68	71
Total	13	87	100

- 1) Proposer un test du modèle A, puis du modèle B, en vous basant sur les indications données dans les tableaux. Calculer la statistique de test pour chacun des deux modèles, et conclure quant à l'acceptation ou au rejet de chacun des modèles.
- 2) Pouvez-vous exhiber un modèle qui soit à la fois non-rejeté par votre test et qui prédisse mieux vos données ?
- 3) En vous basant sur les estimations du tableau 3.3 (modèle B), calculer la probabilité qu'un ménage, dont le revenu est de 2,5 et habitant la région, obtienne un crédit dont le montant est 3 avec un taux avantageux.

4. Les modèles multinomiaux

Ce chapitre généralise les méthodes d'estimation basées sur la vraisemblance au cas des variables qualitatives multinomiales (polytomiques). On présente tout d'abord les modèles ordonnés et séquentiels, avant de développer les modèles non ordonnés qui constituent la catégorie la plus utilisée en pratique. Le chapitre détaille les méthodes d'estimation du Logit multinomial et du Logit conditionnel, et donne une interprétation économique des modèles multinomiaux en présentant le concept d'utilité aléatoire. Comme une propriété fondamentale des Logit multinomiaux est l'indépendance des états non pertinents (IIA), on consacre une partie importante du chapitre aux conséquences de cette propriété et à des extensions permettant dans certains cas de résoudre les problèmes liés à cette propriété.

Objectifs du chapitre : détailler les différents modèles multinomiaux (ordonnés, séquentiels, non ordonnés) ; présenter le Logit multinomial et le Logit conditionnel ; justifier l'utilisation de ces modèles par la notion de maximisation d'une fonction d'utilité aléatoire ; proposer des estimateurs plus robustes, Logit hiérarchisé et Probit multinomial, avec les techniques d'estimation associées.

Concepts clés étudiés : variable dépendante multinomiale, estimation par maximum de vraisemblance, modèles à choix discrets, modèles hiérarchisés, indépendance des états non pertinents, Logit et Probit multinomiaux.

I. Introduction

On a vu au chapitre 1 que les variables qualitatives polytomiques sont couramment utilisées dans les enquêtes, essentiellement en raison du besoin de condenser les informations sous une forme de codage pratique pour l'enquêteur. Il existe trois grandes catégories de modèles multinomiaux, qui se distinguent par la façon de modéliser le processus aléatoire ayant engendré la variable qualitative et/ou par le choix du codage de la variable. Ces trois catégories sont les modèles ordonnés, les modèles séquentiels et les modèles non ordonnés. Dans tous les cas, on peut définir un modèle multinomial de façon générale comme suit. Pour un individu i , $i = 1, \dots, N$, la variable à expliquer y_i peut prendre $m_i + 1$ modalités : $0, 1, 2, \dots, m_i$. On suppose que les modalités j sont *mutuellement exclusives* pour chaque individu i , c'est-à-dire que : $\sum_{j=0}^{m_i+1} \text{Prob}(y_i = j) = 1, \forall i$.

La probabilité associée à chaque réponse est :

$$\text{Prob}(y_i = j) = F_{ij}(x, \beta) \quad i = 1, \dots, N \quad j = 0, 1, 2, \dots, m_i$$

où F_{ij} est la fonction de répartition utilisée pour le calcul de la probabilité que l'individu i choisisse la modalité j , fonction des variables x et du vecteur de paramètres β .

→ Remarques

Les variables explicatives x peuvent dépendre de l'individu i , de la modalité j , ou des deux. De même, les paramètres β peuvent dépendre de j ou non.

$\text{Prob}(y_i = 0) = F_{i0}$ n'a pas besoin en fait d'être spécifiée, car cette probabilité est égale à $1 - \sum_{j=1}^{m_i} F_{ij}(x, \beta)$.

On autorise un nombre de modalités différent pour chaque individu. Dans beaucoup d'applications en effet, les individus font face à des ensembles de choix différents : certaines personnes ont accès à un ensemble restreint de marchés ou de biens publics. Un exemple classique est le cas du choix de mode de transport, où la modalité « chemin de fer » n'est pas incluse dans l'ensemble des choix des individus en dehors de la zone desservie par le train.

On définit maintenant les variables binaires y_{ij} en nombre $\sum_{i=1}^N (m_i + 1)$ comme suit :

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

On peut écrire la vraisemblance du modèle comme le produit des probabilités associées aux différentes modalités, ceci pour tous les individus :

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

L'écriture des différentes probabilités en fonction des variables explicatives et des paramètres dépendra alors d'une règle de décision faisant intervenir, comme précédemment, une ou plusieurs variables latentes.

II. Les modèles ordonnés

Ce type de modèle est très proche des spécifications binaires du chapitre 3. Les modèles ordonnés sont utilisés quand les valeurs prises par la variable multinomiale correspondent à des *intervalles* dans lesquels va se trouver *une seule* variable latente (inobservable) continue¹.

A. Exemple : enquête de consommation

Prenons l'exemple d'une enquête de consommation dans laquelle la question posée concerne l'achat éventuel d'un bien, en fonction de variables tels le revenu du foyer et sa composition. La variable dépendante *ACHAT* est codée comme suit :

$$\left\{ \begin{array}{lll} ACHAT & = & 1 \quad \text{si « achat au prix } p \text{ »} \\ " & = & 2 \quad \text{si « peut-être »} \\ " & = & 3 \quad \text{si « pas d'achat au prix } p \text{ »} \end{array} \right.$$

Supposons qu'une variable latente y_i^* représente le consentement du ménage à payer pour le bien, celui-ci étant croissant avec le revenu et le nombre de personnes du foyer. Si $ACHAT = 1$, le consentement à payer est supérieur au prix p ; si $ACHAT = 3$, il est inférieur. En revanche, une zone d'indécision existe entre ces deux extrêmes, lorsque $ACHAT = 2$. Une façon de modéliser le problème est de considérer des intervalles dans lesquels la variable latente

1. Voir Dahm, Gurlan et Lee, 1960 ; Deacon et Shapiro, 1975.

y_i^* peut se trouver, et qui détermineront la réalisation de la variable observée $ACHAT$:

$$\begin{cases} ACHAT_i = 1 & \text{si } y_i^* - p \geq c_1 \\ ACHAT_i = 2 & \text{si } c_1 > y_i^* - p \geq c_2 \\ ACHAT_i = 3 & \text{si } c_2 > y_i^* - p \end{cases}$$

où c_1 et c_2 sont des bornes à estimer, définissant les frontières des intervalles. Comme au chapitre 3, on spécifie la variable latente comme une combinaison linéaire des variables explicatives x : $y_i^* = x_i\beta + u_i$, où u_i admet une fonction de répartition $F(\cdot)$. Les probabilités associées aux trois réalisations de $ACHAT$ sont alors :

$$\begin{cases} Prob(ACHAT_i = 1) & = Prob(x_i\beta + u_i - p \geq c_1) = 1 - F(p + c_1 - x_i\beta) \\ Prob(ACHAT_i = 2) & = Prob(y_i^* - p < c_1) - Prob(y_i^* - p < c_2) \\ & = F(p + c_1 - x_i\beta) - F(p + c_2 - x_i\beta) \\ Prob(ACHAT_i = 3) & = Prob(c_2 > y_i^* - p) = F(p + c_2 - x_i\beta) \end{cases}$$

La vraisemblance du modèle s'écrit alors :

$$L = \prod_{i=1}^N [1 - F(p + c_1 - x_i\beta)]^{y_{ij}} \times [F(p + c_1 - x_i\beta) - F(p + c_2 - x_i\beta)]^{y_{i2}} \times [F(p + c_2 - x_i\beta)]^{y_{i3}}$$

où $y_{ij} = 1$ si $ACHAT_i = j$, 0 sinon, $j = 1, 2, 3$.

Notons que l'on peut généraliser le modèle en utilisant différentes fonctions de x pour les différents seuils :

$$\begin{cases} Prob(ACHAT_i = 1) & = 1 - F(p + c_1 - x_i\beta_1) \\ Prob(ACHAT_i = 2) & = F(p + c_1 - x_i\beta_1) - F(p + c_2 - x_i\beta_2) \\ Prob(ACHAT_i = 3) & = F(p + c_2 - x_i\beta_2) \end{cases}$$

B. Formulation courante d'un modèle ordonné

Soit une mesure de probabilité P dépendant de β et de x , et une suite finie d'intervalles successifs $\{S_j\}, j = 0, 1, \dots, m$ sur la droite des réels, telle que : $\bigcup_j S_j = R$.

Le modèle ordonné est alors défini par : $Prob[y = j|x, \beta] = P(S_j)$.

Considérons une variable aléatoire y^* qui détermine la réalisation de y par la règle suivante :

$$y = j \quad \text{si et seulement si } c_j < y^* < c_{j+1} \quad j = 0, 1, \dots, m$$

Si la mesure de probabilité P est associée à une densité symétrique et une fonction de répartition F , on a alors :

$$\boxed{\text{Prob}[y=j|x, \beta, c] = F(c_{j+1} - x\beta) - F(c_j - x\beta)}$$

$$j = 0, 1, 2, \dots, m \quad c_0 = -\infty \quad c_j \leq c_{j+1} \quad c_{m+1} = +\infty$$

Notons que les conditions sur les bornes c_0 et c_{m+1} assurent que la somme des probabilités sur tous les intervalles vaut 1.

III. Application

Reprendons notre modèle de choix de climatiseur du chapitre 3, section IV. Supposons que les ménages aient une certaine utilité de la climatisation, en fonction de leurs caractéristiques (revenu, taille du foyer, etc). Le désir de la climatisation constitue notre variable latente y_i^* , que l'on peut supposer croissante soit avec le type de climatiseur, soit avec leur nombre. Si la première hypothèse est vraie, alors la nature du climatiseur importe au ménage ; si en revanche, la seconde est vraie, il lui importe peu entre un climatiseur portable ou central. Étudions d'abord la première hypothèse. La variable y_i^* sera minimale lorsque le ménage ne possède pas de climatiseur, puis supérieure s'il possède un climatiseur portable, puis central, et enfin les deux. L'hypothèse implicite est alors que la puissance de la climatisation est supérieure avec un système central par rapport à un système portable. La variable qualitative observée, y_i , est codée comme suit :

$$\begin{cases} y_i = 1 & \text{si aucun climatiseur,} \\ y_i = 2 & \text{si climatiseur portable,} \\ y_i = 3 & \text{si climatiseur central,} \\ y_i = 4 & \text{si les deux.} \end{cases}$$

Les paramètres associés aux variables explicatives de y_i^* peuvent être estimés conjointement aux bornes $a_j, j = 1, 2, 3$ délimitant les quatre régions ci-dessus, avec un modèle Logit ou Probit ordonné. Les résultats d'estimation du Logit ordonné avec le logiciel GAUSS sont donnés ci-dessous (procédure ORDERED). Comme les sorties du logiciel sont comparables à celles des procédures GAUSS pour les modèles binaires, vues plus haut, on ne rappellera pas la signification des termes employés par le logiciel.

=====

ORDERED Version 3.1.5

=====

Data Set : aircond

CASES PROCESSED BY THIS PROGRAM :

137 cases were kept out of 137 in file.

DEPENDENT CATEGORIES ARE DESIGNATED AS :

1 - AUCUN 2 - PORTABLE 3 - CENTRAL 4 - DEUX

DISTRIBUTION AMONG ORDINAL GROUPS FOR VARIABLE y :

PROPORTION	AUCUN	PORTABLE	CENTRAL	DEUX
	0.3285	0.3504	0.2774	0.0438

ITERATION : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

ESTIMATES FROM ORDINAL LOGIT ANALYSIS OF VARIABLE : y

Convergence after 17 iterations.

Tolerance of 0.0000 achieved in 0.01 minutes.

Variable	Logit Estimate	Std. Error	t-value	p> t
TAILLE	-0.09183	0.1186	-0.77	0.440
NBFOYER	0.00091	0.0004	2.52	0.013
REVENU	0.73251	0.3262	2.25	0.026
MAISON	-0.31030	0.4756	-0.65	0.515
Constant	Estimate	Std. Error	t-value	p> t
Alpha ₁	6.60960	2.8450	2.32	0.022
Alpha ₂	8.36990	2.8963	2.89	0.005
Alpha ₃	11.13545	2.9735	3.74	0.000

MEASURES OF FIT :

Likelihood Ratio Chi-square : 32.8066

with 4 d.f., prob= 0.0000

-2 Log Likelihood for full model : 303.0756

-2 Log likelihood for restricted model : 335.8822

Percent Correctly Predicted : 44.5255

TABLE OF OBSERVED AND PREDICTED OUTCOMES :

Observed	Predicted				Total
	AUCUN	PORTABLE	CENTRAL	DEUX	
AUCUN	17	23	5	0	45
PORTABLE	20	24	4	0	48
CENTRAL	4	13	20	1	38
DEUX	1	2	3	0	6
Total	42	62	32	1	137

Notons que l'ordonnée à l'origine n'est pas identifiable avec les modèles ordonnés, elle n'est donc pas reportée. Comme les bornes α_j sont significativement différentes de 0, nous concluons que le type de climatiseur importe au ménage. Étudions à présent la deuxième hypothèse selon laquelle le nombre de climatiseurs importe. On redéfinit la variable observée comme suit :

$$\begin{cases} y_i = 1 & \text{si aucun climatiseur,} \\ y_i = 2 & \text{si un climatiseur,} \\ y_i = 3 & \text{si deux climatiseurs.} \end{cases}$$

Les résultats avec le modèle Probit ordonné sont :

DEPENDENT CATEGORIES ARE DESIGNATED AS :

1 - AUCUN 2 - UN 3 - DEUX

DISTRIBUTION AMONG ORDINAL GROUPS FOR VARIABLE y:

	AUCUN	UN	DEUX
PROPORTION	0.3285	0.6277	0.0438

ITERATION : 1 2 3 4 5 6 7 8 9 10 11 12 13

ESTIMATES FROM ORDINAL PROBIT ANALYSIS OF VARIABLE : y

Convergence after 13 iterations.

Tolerance of 0.0000 achieved in 0.01 minutes.

Variable	Probit Estimate	Std. Error	t-value	p> t
TAILLE	-0.08144	0.0741	-1.10	0.273
NBFOYER	0.00023	0.0002	1.00	0.319
REVENU	0.39290	0.1978	1.99	0.049
MAISON	-0.15117	0.3105	-0.49	0.627
Constant	Estimate	Std. Error	t-value	p> t
Alpha ₁	3.11194	1.7495	1.78	0.078
Alpha ₂	5.41398	1.8267	2.96	0.004

MEASURES OF FIT :

Likelihood Ratio Chi-square : 13.8757

with 4 d.f., prob= 0.0077

-2 Log Likelihood for full model : 203.9506

-2 Log likelihood for restricted model : 217.8263

Percent Correctly Predicted : 70.0730

TABLE OF OBSERVED AND PREDICTED OUTCOMES :

Observed	Predicted				Total
	AUCUN	UN	DEUX		
AUCUN	13	32	0		45
UN	3	83	0		86
DEUX	0	6	0		6
Total	16	121	0		137

On voit à présent que la borne α_1 , délimitant la région « aucun » de la région « un climatiseur », n'est pas significativement différente de 0, contrairement à la borne α_2 séparant la région « un » de « deux climatiseurs ». Par conséquent, ce modèle indique seulement que les ménages effectueraient leur choix entre les modalités : « moins de deux » contre « deux climatiseurs ». Le premier modèle est donc préférable, dans la mesure où la distinction entre 0, 1 ou 2 climatiseurs est nette, en fonction de la nature de l'équipement.

IV. Les modèles séquentiels

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 événement* à chaque étape. Soit T le nombre d'étapes et $Y_i, i = 1, 2, \dots, N$ une variable polytomique dont les valeurs possibles sont $\{1, 2, \dots, T\}$. La valeur de Y_i indique donc l'étape à laquelle le processus s'est arrêté. On peut écrire la probabilité de s'arrêter à l'étape t comme une fonction $F_t(x_i, \beta), t = 1, 2, \dots, T$, où le conditionnement par t indique que les variables x_i ou les paramètres β peuvent dépendre de l'étape. On a alors :

$$\text{Prob}(Y_i = j) = \prod_{s=1}^{j-1} [1 - F_s(x_i, \beta)] \times F_j(x_i, \beta)$$

► Exemple 4.1

On s'intéresse à la probabilité qu'un étudiant obtienne une maîtrise universitaire, en fonction de variables comme son revenu, celui de ses parents, le choix de ses options, etc. La variable dépendante est notée Y_i , avec $Y_i = 1$ si l'individu n'a pas obtenu le baccalauréat ; $Y_i = 2$ si l'individu a le baccalauréat mais pas le DEUG ; $Y_i = 3$ si l'individu a le DEUG mais pas la licence ; $Y_i = 4$ si l'individu a achevé la licence mais pas la maîtrise. La probabilité associée à l'obtention de la licence est par exemple :

$$\text{Prob}(Y_i = 4) = [1 - F_1(x_i, \beta)][1 - F_2(x_i, \beta)][1 - F_3(x_i, \beta)]F_4(x_i, \beta)$$

La probabilité que $Y_i = 1$ est calculée sur tout l'échantillon, en le divisant en deux groupes : les étudiants ayant obtenu le baccalauréat et les autres. On utilise ensuite le sous-échantillon « avec baccalauréat » divisé en 2 groupes : DEUG ou non pour calculer $F_2(x_i, \beta)$, et ainsi de suite.

V. Modèles non ordonnés : le Logit multinomial

Nous étudions à présent la classe la plus importante des modèles multinomiaux, celle des modèles non ordonnés. Commençons d'abord par le plus courant, le Logit multinomial, qui constitue une généralisation assez triviale *a priori* du modèle Logit binaire.

A. Généralisation du Logit binaire au cas multinomial

Dans le cas du Logit simple du chapitre précédent, nous avions $\text{Prob}(y_i = 1) = \exp(x_i\beta)/[1 + \exp(x_i\beta)]$. Du fait de la contrainte $\text{Prob}(y_i = 0) + \text{Prob}(y_i = 1) = 1$, un seul vecteur de paramètres β était suffisant pour expliquer les deux probabilités élémentaires, qui ne variaient qu'en fonction des caractéristiques x_i . Le passage au cas multinomial pose alors le problème suivant : si la fonction de répartition $F(\cdot)$ est donnée, comment expliquer des différences entre plus de deux probabilités si β est constant ? La solution est évidemment : soit considérer un vecteur de paramètres β différent selon chaque alternative, soit autoriser les variables explicatives x à varier en fonction des modalités. Dans le cas binaire (cf. chap. 3), si l'on introduit deux vecteurs β_0 et β_1 , on peut normaliser les probabilités que $y_i = 1$ et $y_i = 0$ pour obtenir :

$$\text{Prob}(y_i = j) = \exp[x_i(\beta_j - \beta_0)] / \sum_{k=0,1} \exp[x_i(\beta_k - \beta_0)] \quad j = 0, 1$$

et les probabilités du modèle original sont inchangées si $\beta = \beta_1 - \beta_0$. Le rapport des probabilités $\text{Prob}(y_i = 1)/\text{Prob}(y_i = 0)$ vaut $\exp(x_i\beta) = \exp[x_i(\beta_1 - \beta_0)] = \exp(x_i\beta_1)/\exp(x_i\beta_0)$. En d'autres termes, le rapport des probabilités dans le modèle Logit simple est le rapport de deux fonctions dont les paramètres sont indicés par la modalité. On peut alors se demander si une généralisation du Logit au cas multinomial est possible directement à partir de l'équation pour $\text{Prob}(y_i = j)$ développée ci-dessus. La réponse est oui, et nous examinons maintenant comment obtenir cette généralisation.

Soit $m + 1$ modalités et $(p_0, p_1, p_2, \dots, p_m)$ les probabilités associées. La catégorie 0 est prise comme référence, et l'on pose :

$$\frac{p_1}{p_1 + p_0} = H(x_i\beta_1), \quad \frac{p_2}{p_2 + p_0} = H(x_i\beta_2), \quad \dots, \quad \frac{p_m}{p_m + p_0} = H(x_i\beta_m)$$

où $H(\cdot)$ est une fonction continue et croissante. Comme :

$$\frac{p_j + p_0}{p_j} = H^{-1}(x_i\beta_j)$$

on a :

$$\frac{p_0}{p_j} = \frac{1}{H(x_i\beta_j)} - 1 \quad j = 1, 2, \dots, m$$

et donc :

$$\frac{p_j}{p_0} = \frac{H(x_i\beta_j)}{1 - H(x_i\beta_j)} = G(x_i\beta_j) \quad j = 1, 2, \dots, m$$

où $G(\cdot)$ est une fonction croissante de $x_i\beta_j$. Calculons maintenant l'expression de p_0 :

$$\sum_{j=1}^m \frac{p_j}{p_0} = \frac{1}{p_0} \sum_{j=1}^m p_j = \frac{1}{p_0} (1 - p_0) = \frac{1}{p_0} - 1$$

$$\Leftrightarrow \frac{1}{p_0} - 1 = \sum_{j=1}^m G(x_i\beta_j) \quad j = 1, 2, \dots, m \quad \Leftrightarrow \quad p_0 = \left[1 + \sum_{j=1}^m G(x_i\beta_j) \right]^{-1}$$

$$\Leftrightarrow p_j = \frac{G(x_i\beta_j)}{1 + \sum_{k=1}^m G(x_i\beta_k)} = \frac{G(x_i\beta_j)}{\sum_{k=0}^m G(x_i\beta_k)} \quad \text{si } G(x_i\beta_0) = 1$$

Cette dernière condition s'obtient simplement si on pose maintenant $G(x_i\beta_j) = \exp(x_i\beta_j)$, de sorte que $G(x_i\beta_0) = 1 \Leftrightarrow \beta_0 = 0$. On obtient alors une expression de la probabilité compatible avec celle du Logit binaire :

$$p_j = \text{Prob}(y_i = j) = \frac{\exp(x_i\beta_j)}{(1 + \sum_{k=1}^m \exp(x_i\beta_k))} \quad j = 1, 2, \dots, m$$

Le rapport entre deux probabilités p_j et p_k , $j, k = 1, 2, \dots, m$, vaut $\exp[x_i(\beta_j - \beta_k)]$ et le vecteur de paramètres β_0 est normalisé à 0. Une remarque importante est que ce rapport est indépendant des alternatives autres que j et k . Nous reviendrons en détail sur ce point à la section 4.VII.B.

B. Logit multinomial : spécification et estimation

La généralisation ci-dessus part du postulat que les paramètres du modèle dépendent des alternatives, mais que les variables explicatives ne dépendent que des individus. Les probabilités s'expriment comme des rapports d'exponentielles, plusieurs propriétés en découlent. Tout d'abord, le rapport entre deux probabilités est invariant à une translation dans les vecteurs des paramètres : $\text{Prob}(y_i = j)/\text{Prob}(y_i = k) = \exp(x_i\beta_j)/\exp(x_i\beta_k) = \exp[x_i(\beta_j - \beta_k)]/\exp[x_i(\beta_k - \beta_k)]$. Ce rapport est par contre modifié si les variables explicatives x_i varient (d'un facteur Δ) : $\exp(\Delta x_i\beta_j)/\exp(\Delta x_i\beta_k) \neq \exp(x_i\beta_j)/\exp(x_i\beta_k)$.

Deuxièmement, une remarque importante concerne la signification des paramètres estimés du modèle. Si l'on part de l'expression de la probabilité : $p_j = \text{Prob}(y_i = j) = \exp(x_i\beta_j)/[\sum_{k=0}^m \exp(x_i\beta_k)]$ sans imposer la contrainte

$\sum_{k=0}^m p_k = 1$ (c'est-à-dire sans imposer $\beta_0 = 0$), on voit que cette contrainte s'obtient de façon équivalente en divisant, en haut et en bas, le rapport ci-dessus par $\exp(x_i\beta_0)$. On a en effet :

$$p_j = \text{Prob}(y_i = j) = \frac{\exp[x_i(\beta_j - \beta_0)]}{\sum_{k=0}^m \exp(x_i(\beta_k - \beta_0))} = \frac{\exp(x_i\beta_j^*)}{1 + \sum_{k=1}^m \exp(x_i\beta_k^*)}$$

avec $\beta_j^* = \beta_j - \beta_0 \forall j = 1, 2, \dots, m$.

On voit donc que la contrainte $\sum_{k=0}^m p_k = 1$ revient à normaliser les paramètres du modèle, qui sont maintenant les différences entre les paramètres originaux β et le vecteur de paramètres de la modalité de référence, β_0 . Il est important de garder à l'esprit l'interprétation des paramètres estimés comme des écarts au référentiel (modalité 0). Une conséquence importante de ce résultat est que les paramètres du modèle ne peuvent être estimés que pour les m modalités, sans le cas de référence. L'estimation du modèle Logit multinomial s'effectue alors en maximisant la log-vraisemblance par rapport aux vecteurs de paramètres $(\beta_1, \beta_2, \dots, \beta_m)$:

$$\log L = \sum_{i=1}^N \{y_{i1} \log p_{i1} + y_{i2} \log p_{i2} + \dots + y_{im} \log p_{im} + y_{i0} \log p_{i0}\}$$

où $p_{ij} = \frac{\exp(x_i\beta_j)}{1 + \sum_{k=1}^m \exp(x_i\beta_k)}$, $j = 1, 2, \dots, m$ et $p_{i0} = \frac{1}{1 + \sum_{k=1}^m \exp(x_i\beta_k)}$

La log-vraisemblance s'écrit encore :

$$\log L = \sum_{i=1}^N \sum_{k=1}^m y_{ik} x_i \beta_k - (m+1) \sum_{i=1}^N \log \left(1 + \sum_{k=1}^m \exp(x_i \beta_k) \right)$$

car $\beta_0 = 0$ par convention. Il est aisément de montrer que :

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^N (y_{ik} - p_{ik}) x'_i$$

et que :

$$\frac{\partial^2 \log L}{\partial \beta_k \partial \beta_l} = - \sum_{i=1}^N p_{ik} (\mathbb{I}_{k,l} - p_{il}) x'_i x_i$$

où $\mathbb{I}_{k,l} = 1$ si $k = l$, 0 sinon.

Les effets marginaux des variables x_i sont calculés comme suit :

$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij} \left[\beta_j - \sum_{k=0}^m p_{ik} \beta_k \right]$$

et dépendent de façon non linéaire de tous les paramètres β_j intervenant dans les probabilités p_{ij} . Considérons à présent le test de significativité globale des

paramètres au moyen, par exemple, de la statistique du rapport des vraisemblances. Si tous les paramètres du modèle sont nuls excepté le terme constant, noté β_j pour la modalité j , la probabilité contrainte est :

$$p_{ij}(\bar{\beta}_j) = \text{Prob}(y_i = j) = \frac{\exp(\beta_j)}{1 + m \exp(\beta_j)}$$

qui ne dépend plus de i , et peut donc être remplacée par son équivalent empirique, à savoir la proportion d'individus choisissant la modalité j . Notons $\bar{p}_j = n_j/N$ cette proportion, où n_j est le nombre d'individus représentés dans la modalité j . La log-vraisemblance contrainte s'écrit :

$$\log \bar{L} = \sum_{i=1}^N \sum_{j=0}^m y_{ij} \log \bar{p}_j = \sum_{j=0}^m n_j \log \bar{p}_j$$

C. Le Logit conditionnel de McFadden

La généralisation du modèle Logit au cas multinomial se fait, comme on l'a vu, en considérant des paramètres différents, en fonction des alternatives, de la forme fonctionnelle de la probabilité ainsi que des variables explicatives restant constantes sur les alternatives. Il existe cependant une autre possibilité, celle à la base du Logit conditionnel de McFadden : considérer un vecteur de paramètres constants et autoriser les variables explicatives à dépendre des modalités (McFadden, 1973, 1981). Le modèle s'écrit :

$$\begin{aligned} \text{Pr}(y_i = j) &= \frac{\exp(x_{ij}\beta)}{\sum_{k=0}^m \exp(x_{ik}\beta)} \\ &= \frac{\exp(x_{ij}^*\beta)}{1 + \sum_{k=1}^m \exp(x_{ik}^*\beta)} \quad j = 1, 2, \dots, m \end{aligned}$$

avec $x_{ij}^* = x_{ij} - x_{i0}$, et le rapport des probabilités est :

$$\frac{\text{Pr}(y_i = j)}{\text{Pr}(y_i = l)} = \frac{\exp(x_{ij}^*\beta)}{\exp(x_{il}^*\beta)} = \frac{\exp(x_{ij}\beta)}{\exp(x_{il}\beta)} = \exp[(x_{ij} - x_{il})\beta] \quad \forall j, l = 1, 2, \dots, m$$

qui, comme dans le cas du Logit multinomial, est indépendant des autres modalités.

Pour le calcul des effets marginaux, on s'intéresse à la variation estimée de la probabilité qu'un individu i choisisse la modalité j , lorsque la k -ème

variable explicative associée à une modalité quelconque (l) varie. En indiquant les paramètres par rapport aux composantes du vecteur x_{ij} , on a :

$$p_{ij} = \frac{\exp(\sum_{k=1}^K x_{ijk}\beta_k)}{1 + \sum_{h=1}^m \exp(\sum_{k=1}^K x_{ihk}\beta_k)}$$

et l'effet marginal $\partial p_{ij} / \partial x_{ik}$ est :

$$\begin{cases} \beta_k p_{ij}(1 - p_{ij}) & \text{si } j = l, \\ -\beta_k p_{ij}(1 - p_{il}) & \text{si } j \neq l \end{cases}$$

Alors qu'avec le Logit multinomial, le vecteur de paramètres β était interprété comme la différence entre les paramètres originaux et ceux du cas de référence (β_0), pour le Logit conditionnel, les variables x_{ij} sont normalisées de sorte que $x_{ij}^* = x_{ij} - x_{i0}$. Les paramètres estimés doivent donc s'interpréter comme associés aux différences des variables de chaque modalité par rapport aux variables du cas de référence. Ceci provient comme précédemment de la contrainte $\sum_{j=0}^m p_{ij} = 1 \forall i$.

Cette spécification est très utile lorsque l'on souhaite prédire la probabilité d'une nouvelle modalité (virtuelle) en fonction de variables explicatives simulées. Prenons l'exemple d'une collectivité locale envisageant la mise en place d'un nouveau mode de transport public, le métro, en plus des modes existants (le bus). Pour évaluer la probabilité que les administrés adoptent ce nouveau mode de transport, on conduit d'abord une enquête sur le choix du mode de transport, dont le but est de calculer la probabilité qu'un individu choisisse le bus (modalité 1), la voiture (modalité 2) ou le vélo (modalité 3). Le nombre de modalités possibles est ici $m + 1 = 4$, le mode de référence 0 étant les autres modes (marche, rollers, auto-stop, etc.). Les variables explicatives sont le temps de transport moyen du domicile au lieu de travail pour le mode j , x_{1ij} , et le coût au kilomètre de ce mode, x_{2ij} , pour l'individu i . Le modèle donne alors la probabilité qu'un individu caractérisé par les variables $(x_{1i1}, x_{1i2}, x_{1i3})$ et $(x_{2i1}, x_{2i2}, x_{2i3})$ choisisse le mode de transport j , $j = 1, 2, 3$. La probabilité que l'individu i choisisse le bus est par exemple :

$$\text{Prob}(i \text{ choisit le bus}) = \frac{\exp(\beta_0 + \beta_1 x_{1i1} + \beta_2 x_{2i1})}{1 + \sum_{k=1}^3 \exp(\beta_0 + \beta_1 x_{1ik} + \beta_2 x_{2ik})}$$

Comme le mode de transport « métro » (modalité 4) est encore inexistant, les variables x_{1i4} et x_{2i4} sont bien sûr inconnues, mais elles peuvent être simulées à partir d'une estimation préalable du temps de transport par métro et du coût au kilomètre. En remplaçant x_{1i4} et x_{2i4} par leurs estimations et le vecteur β par son estimation préalable sur les trois modes de transport préexistants, on peut calculer la probabilité que les individus choisissent le métro plutôt que les autres modes de transport.

D. Un modèle Logit plus général

Le choix entre le Logit multinomial et le Logit conditionnel dépend essentiellement du type de données dont l'économètre dispose. Si les données sont majoritairement individuelles (propres aux individus), on utilisera le Logit multinomial. Dans le cas plus rare où les variables dépendent des modalités, on préférera le Logit conditionnel. Notons que dans la mesure où les paramètres du Logit conditionnel sont indépendants des modalités, il n'est pas possible *a priori* d'incorporer dans le modèle des variables propres aux individus, car les paramètres associés ne seraient pas identifiables. Si Z_i est un vecteur de variables individuelles, on aurait en effet :

$$\begin{aligned} \text{Prob}(y_i = j) &= \frac{\exp(x_{ij}\beta + Z_i\gamma)}{\sum_{k=0}^m \exp(x_{ik}\beta + Z_i\gamma)} = \frac{\exp(x_{ij}\beta) \exp(Z_i\gamma)}{\sum_{k=0}^m \exp(x_{ik}\beta) \exp(Z_i\gamma)} \\ &= \frac{\exp(x_{ij}\beta)}{\sum_{k=0}^m \exp(x_{ik}\beta)} = \frac{\exp(x_{ij}^*\beta)}{1 + \sum_{k=1}^m \exp(x_{ik}^*\beta)} \end{aligned}$$

Une formulation plus générale du modèle Logit peut alors être proposée, qui contient à la fois le Logit multinomial et le Logit conditionnel comme des cas particuliers. On note comme auparavant Z_i le vecteur de caractéristiques propres à l'individu, x_{ij} le vecteur de variables dépendant à la fois de l'individu et de la modalité. On pose alors :

$$Pr[y_i = j] = \frac{\exp(x_{ij}\beta + Z_i\alpha_j)}{\sum_{k=0}^m \exp(x_{ik}\beta + Z_i\alpha_k)}$$

et en divisant par $\exp(x_{i0}) \times \exp(Z_i\alpha_0)$ au numérateur et au dénominateur :

$$\begin{aligned} Pr[y_i = j] &= \frac{\exp[(x_{ij} - x_{i0})\beta + Z_i(\alpha_j - \alpha_0)]}{\sum_{k=0}^m \exp[(x_{ik} - x_{i0})\beta + Z_i(\alpha_k - \alpha_0)]} \\ &= \frac{\exp(x_{ij}^*\beta + Z_i\alpha_j^*)}{1 + \sum_{k=1}^m \exp(x_{ik}^*\beta + Z_i\alpha_k^*)} \end{aligned}$$

Les paramètres de ce modèle s'interprètent de la façon suivante : ceux associés aux variables spécifiques aux individus (α_j^*) constituent la différence entre les paramètres originaux (non contraints) et ceux de la modalité de référence 0, alors que les paramètres β sont associés aux différences entre les variables dépendant des modalités et la valeur des variables dans le cas de référence.

On peut écrire ce modèle général sous la forme compacte d'un Logit conditionnel, comme suit :

$$v_{ij} = \exp(\omega_{ij}\delta)$$

où :

$$\begin{aligned}\omega_{i1} &= (x_{i1} - x_{i0}, Z_i, 0_z, \dots, 0_z), \omega_{i2} = (x_{i2} - x_{i0}, 0_z, Z_i, 0_z, \dots, 0_z), \\ \omega_{i3} &= (x_{i3} - x_{i0}, 0_z, 0_z, Z_i, 0_z, \dots, 0_z), \dots, \omega_{im} = (x_{im} - x_{i0}, 0_z, \dots, 0_z, Z_i) \\ \delta &= (\beta, \alpha_1 - \alpha_0, \alpha_2 - \alpha_0, \dots, \alpha_m - \alpha_0)'\end{aligned}$$

et où 0_z est un vecteur de 0 de la même dimension que le vecteur Z_i .

VI. Application : mode de transport

Nous étudions l'application du modèle Logit multinomial au choix du mode de transport. Un échantillon de ménages comporte les variables suivantes :

- *MODE* : 1 si transport public, 2 si voiture, 3 si autre ;
- *NBVOIT* : nombre de voitures du ménage ;
- *REV* : revenu annuel du ménage ;
- *KMS* : distance en kilomètres entre le lieu de résidence et le travail ;
- *VILLE* : 1 si le ménage réside dans une petite ou moyenne ville.

Les variables sont toutes spécifiques au ménage, et nous prenons le cas *MODE* = 1 comme référence, c'est-à-dire les transports publics. Le modèle est estimé par maximum de vraisemblance avec le logiciel EVIEWS. Comme ce dernier ne contient pas de procédures dédiées aux modèles multinomiaux, le problème de maximisation de la log-vraisemblance doit être écrit avec le langage EVIEWS. Le programme est le suivant :

```
' Chargement de la base de données MNL1
load "mnl1"
' déclaration du vecteur des paramètres
coef(5) b2
coef(5) b3
logl mlogit
mlogit.append @logl logl1
mlogit.append xb2 = b2(1)+b2(2)*rev+b2(3)*kms+b2(4)*nbvoit+b2(5)*ville
mlogit.append xb3 = b3(1)+b3(2)*rev+b3(3)*kms+b3(4)*nbvoit+b3(5)*ville
mlogit.append denom = 1+exp(xb2)+exp(xb3)
mlogit.append pr1 = 1/denom
mlogit.append pr2 = exp(xb2)/denom
mlogit.append pr3 = exp(xb3)/denom
mlogit.append logli = (1-dd2-dd3)*log(pr1)+dd2*log(pr2)+dd3*log(pr3)
' Dérivées analytiques
```

```

for !i=2 to 3
mlogit.append @deriv b!i(1) grad!i1 b!i(2) grad!i2 b!i(3) grad!i3
b!i(4) grad!i4 b!i(5) grad!i5
mlogit.append grad!i1 = dd!i-pr!i
mlogit.append grad!i2 = grad!i1*rev
mlogit.append grad!i3 = grad!i1*kms
mlogit.append grad!i4 = grad!i1*nboit
mlogit.append grad!i5 = grad!i1*ville
next
' Calculer les valeurs initiales à partir de Logit binomiaux
equation eq2.binary(d=1) dd2 c rev kms nbvoit ville
b2 = eq2.@coefs
equation eq3.binary(d=1) dd3 c rev kms nbvoit ville
b3 = eq3.@coefs
' Estimation Max. de Vraisemblance et affichage résultats
mlogit.ml(d)
show mlogit.output

```

Les résultats d'estimation sont les suivants :

LogL : MLOGIT

Method : Maximum Likelihood (Marquardt) 1

Date : 08/12/99 Time : 22:49

Sample : 1 3177 2

Included observations : 3177 3

Evaluation order : By observation

Initial Values : B2(1)=-1.75460, B2(2)=-0.12985, B2(3)=0.03170, B2(4)=-0.19328, B2(5)=0.27205, B3(1)=-0.03489, B3(2)=0.09943, B3(3)=-0.03222, B3(4)=0.90479, B3(5)=0.22929 4

Convergence achieved after 14 iterations 5

	Coefficient	Std. Error	z-Statistic	Prob.
B2(1)	-1.893495	0.189999	-9.965832	0.0000
B2(2)	-0.215066	0.090071	-2.387740	0.0170
B2(3)	0.015705	0.007431	2.113526	0.0346
B2(4)	2.139926	0.131128	16.31937	0.0000
B2(5)	1.277326	0.197324	6.473239	0.0000
B3(1)	-0.640715	0.175233	-3.656364	0.0003
B3(2)	-0.077147	0.076865	-1.003669	0.3155
B3(3)	-0.018325	0.007215	-2.539952	0.0111
B3(4)	2.527233	0.125531	20.13239	0.0000
B3(5)	1.148959	0.184664	6.221890	0.0000
Log likelihood		-1823.102	Akaike info criterion	1.153983
Avg. log likelihood		-0.573844	Schwarz criterion	1.173069
Number of Coefs. 6	10		Hannan-Quinn criter.	1.160828

Le logiciel rappelle en 1 la méthode d'estimation : maximum de vraisemblance avec algorithme d'optimisation de Marquardt, l'étendue et la dimension de l'échantillon 2 et 3. Les valeurs initiales utilisées pour amorcer l'algorithme de maximisation de la log-vraisemblance sont données en 4, et 5 indique le nombre d'itérations avant convergence. Les tableaux des paramètres estimés et des statistiques d'ajustement sont comparables dans leur présentation à ceux vus plus haut pour le modèle Logit binaire. 6 indique simplement le nombre de coefficients estimés.

Les résultats d'estimation s'interprètent de la façon suivante. Les paramètres estimés $BJ(K)$ donnent l'impact de la variable explicative K sur la probabilité de choisir le mode J relativement au mode de référence, les transports publics dans notre cas. Ainsi, les coefficients $B2(3)$ et $B3(3)$ associés à la variable KMS (voir le programme plus haut) sont respectivement positif et négatif, ce qui indique que plus la distance entre le domicile et le lieu de travail est grande, plus les individus seront enclins à préférer la voiture aux transports publics, mais ils utiliseront moins les autres modes (marche, vélo) que les transports publics. De même, le nombre de véhicules du ménage a un effet positif sur la probabilité d'utiliser la voiture relativement aux transports publics, et sur celle d'utiliser un autre mode ($B2(4)$ et $B3(4)$ sont positifs). Le fait d'habiter une petite ville (paramètres $B2(5)$ et $B3(5)$) influence de façon positive la probabilité de choisir la voiture ou un autre mode par rapport aux transports publics, ce qui s'explique aisément : la marche ou le vélo sont des moyens de déplacement plus adaptés aux petites agglomérations, et les transports publics sont souvent moins développés dans ces dernières par rapport aux grandes villes.

VII. Modèles de choix probabilistes et hypothèse IIA

Les modèles Logit multinomiaux sont très populaires dans les applications économétriques en raison de leur expression simple et intuitive. En revanche, ils possèdent des propriétés qui peuvent les rendre indésirables dans un certain nombre d'applications. Cet aspect négatif des modèles est à la base du développement continu de formes multinomiales alternatives en économétrie des variables qualitatives. Il est important, pour bien comprendre les désavantages du Logit multinomial, de se pencher au préalable sur les modèles théoriques de choix probabiliste. On verra que les spécifications des modèles multinomiaux peuvent être dérivées de façon très simple de la maximisation de l'utilité aléatoire.

A. Modèles probabilistes à choix discret

Considérons une population d'agents économiques (consommateurs par exemple) rationnels, faisant face à un certain nombre de biens dont la consommation est exclusive (un seul produit doit être consommé). Chaque bien procure un niveau d'utilité différent (en fonction notamment des attributs des biens), et le consommateur cherche de façon traditionnelle à maximiser son utilité sous une contrainte budgétaire, ce qui revient ici à décider lequel des biens consommer. Cependant, en raison d'une perception imparfaite de la qualité des différents biens, ou parce que le problème de maximisation est trop complexe pour fournir un niveau exact d'utilité attaché à chaque bien, on suppose que l'utilité est une fonction aléatoire des attributs (Luce, 1959 ; Anderson, Palma et Thisse, 1992).

Soit un ensemble de $m + 1$ biens, dont les attributs sont regroupés dans un vecteur de variables observables $x_j, j = 0, 1, \dots, m$. On note U_j l'utilité indirecte associée à la consommation du bien j , qui est fonction des attributs par la relation :

$$U_j = v(x_j) + \varepsilon_j$$

où ε_j est une variable aléatoire représentant les erreurs de perception et/ou d'optimisation de la part du consommateur, supposées indépendantes de x_j , et $v(\cdot)$ est une fonction continue et déterministe. Le problème du consommateur se transpose alors en univers aléatoire : choisir le bien j lui procurant la plus grande utilité indirecte aléatoire, $U(x_j, \varepsilon_j)$. On introduit une variable indicatrice y_j , égale à 1 si le bien j est choisi, et 0 sinon. La règle de décision est alors :

$$\begin{cases} y_j &= 1 \text{ si } U_j = \max(U_0, U_1, \dots, U_j, \dots, U_m), \\ &= 0 \text{ sinon.} \end{cases}$$

Si par exemple le bien 1 est choisi, on a :

$$\begin{aligned} Pr(y_1 = 1) &= Pr(U_1 > U_0, U_1 > U_2, \dots, U_1 > U_m) \\ &= Pr(v(x_1) + \varepsilon_1 > v(x_0) + \varepsilon_0, v(x_1) + \varepsilon_1 > v(x_2) + \varepsilon_2, \dots, v(x_1) + \varepsilon_1 > v(x_m) + \varepsilon_m) \\ &= Pr(\varepsilon_0 < v(x_1) - v(x_0) + \varepsilon_1, \varepsilon_2 < v(x_1) - v(x_2) + \varepsilon_1, \dots, \varepsilon_m < v(x_1) - v(x_m) + \varepsilon_1). \end{aligned}$$

On suppose que les termes $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_m$ sont indépendamment et identiquement distribués avec une fonction de répartition $F(\cdot)$ sur l'ensemble des réels, et une densité $f(\cdot)$. La probabilité ci-dessus s'écrit :

$$\begin{aligned} &\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{v(x_1) - v(x_0) + \varepsilon_1} f(\varepsilon_0) \int_{-\infty}^{v(x_1) - v(x_2) + \varepsilon_1} f(\varepsilon_2) \cdots \int_{-\infty}^{v(x_1) - v(x_m) + \varepsilon_1} f(\varepsilon_m) \right. \\ &\quad \times f(\varepsilon_0) d\varepsilon_0 d\varepsilon_2 \dots d\varepsilon_m \left. \right] f(\varepsilon_1) d\varepsilon_1 \end{aligned}$$

Cette expression est très difficile ou impossible à calculer, même numériquement, pour la plupart des distributions usuelles dès que le nombre d'alternatives (m) est supérieur à 3. En revanche, si l'on adopte une distribution particulière, le problème admet une solution analytique très simple. Dans le cas général, on peut écrire :

$$\begin{aligned} Pr[y_j = 1] &= Pr[\varepsilon_k < \varepsilon_j + v(x_j) - v(x_k)] \quad \forall j \neq k \\ &= \int_{-\infty}^{+\infty} \prod_{k \neq j} F(\varepsilon_j + v(x_j) - v(x_k)) f(\varepsilon_j) d\varepsilon_j \end{aligned}$$

Si l'on adopte la loi des valeurs extrêmes de type I, définie par :

$$\begin{cases} F(\varepsilon_j) &= \exp[-\exp(-\varepsilon_j)] \\ f(\varepsilon_j) &= \exp[-\varepsilon_j - \exp(-\varepsilon_j)] \end{cases}$$

alors la probabilité est :

$$\frac{\exp[v(x_j)]}{\sum_{k=0}^m \exp[v(x_k)]}$$

On retrouve ainsi l'expression du Logit multinomial ou du Logit conditionnel en faisant l'hypothèse de linéarité de la fonction $v(\cdot)$ par rapport aux variables x .

B. L'hypothèse IIA

L'expression très simple des probabilités dans le cas du Logit multinomial est rendue possible par le choix de la distribution des valeurs extrêmes de type I, avec l'hypothèse simplificatrice que les termes aléatoires ε_j sont indépendamment et identiquement distribués. Ainsi, si ces termes reflètent des erreurs de mesure dans les attributs des biens, ces erreurs sont non corrélées entre les différents biens et ont une même variance. De telles hypothèses conduisent le modèle Logit à posséder une propriété très importante, appelée IIA (*Independence of Irrelevant Alternatives* : indépendance des états non pertinents). Cette propriété s'énonce comme suit : le rapport des probabilités associées au choix entre deux modalités est indépendant des autres modalités. Ajouter ou éliminer une tierce modalité, ou bien modifier les caractéristiques d'une modalité déjà incluse, ne change pas le rapport entre ces probabilités. Cette propriété peut être acceptée si les modalités sont peu comparables, mais elle devient discutable lorsque le Logit est utilisé pour modéliser des choix très similaires.

Reprenons l'exemple du mode de transport présenté ci-dessus. Supposons qu'avant l'introduction du métro, les individus se répartissent comme suit entre les quatre modes : bus (modalité 1, 30 %), voiture (modalité 2, 50 %), vélo

(modalité 3, 10 %) et autres (modalité 0, 10 %). Ces proportions reflètent les probabilités initiales, dont le rapport pour le bus par rapport aux autres modes est $30/70 = 42,85\%$. D'après la propriété IIA, ce rapport ne devrait pas être modifié par l'apparition du métro. Or, si ce dernier possède un réseau semblable à celui du bus et est tarifé de façon identique, le bus et le métro se partageront à égalité le marché du transport en commun, de sorte que le rapport bus/métro = 1. On aurait alors sous cette propriété :

$$Prob(\text{métro}/\text{autres}) = 0,4285$$

et :

$$\begin{aligned} & Prob(\text{bus}) + Prob(\text{métro}) + Prob(\text{autres}) = 1 \\ \Leftrightarrow & \frac{Prob(\text{bus})}{Prob(\text{autres})} + \frac{Prob(\text{métro})}{Prob(\text{autres})} + \frac{Prob(\text{autres})}{Prob(\text{autres})} = \frac{1}{Prob(\text{autres})} \\ \Leftrightarrow & Prob(\text{autres}) = \frac{1}{1 + 2 \times 3/7} = 0,5384 \end{aligned}$$

On voit que les autres modes de transport ont une part en baisse d'environ 16 % en passant de 70 % à 53,84 % du total, alors que le bus a perdu 7 % de parts de marché (en passant de 0,3 à $1/2(1 - 0,5384) = 0,2308$). Le métro a immédiatement conquis 23,08 % du marché, dont 7 % au détriment du bus et 16,15 % au détriment des autres modes. Ce résultat ne semble pas compatible avec l'hypothèse que le bus et le métro offrent les mêmes services au même prix. On aurait dû s'attendre à une part inchangée des autres modes de transport (à 70 %), les transports en commun se partageant le reste à 50-50, soit 15 % chacun.

Cet exemple illustre, de façon peut-être un peu caricaturale, la non-pertinence de la propriété IIA lorsque des modalités sont très similaires, ce qui est le cas du bus et du métro qui offrent les mêmes services selon notre hypothèse. Le nouveau mode de transport (métro) est en effet un substitut du bus plutôt que de la voiture. Mais en raison de l'hypothèse d'indépendance entre les termes aléatoires ε_j , une utilité forte pour le bus implique très certainement une forte utilité aussi pour le métro. Donc la probabilité de préférer les autres modes à la fois au bus *et* au métro, calculée sous l'hypothèse d'indépendance des erreurs, sous-estime la vraie probabilité d'utiliser ces autres modes, car on ignore le fait que l'événement « les autres modes sont préférés au bus » rend l'événement « les autres modes sont préférés au métro » plus vraisemblable. En d'autres termes, on néglige la corrélation entre les erreurs ε_j présentes dans les utilités indirectes associées au bus et au métro.

Un autre aspect restrictif du Logit multinomial, dû à la propriété IIA, est le suivant. Soit le modèle multinomial général introduit plus haut, avec des variables spécifiques aux individus (z_i) et des variables propres aux individus et aux modalités (x_{ij}), notons :

$$v_{ij} = x_{ij}\beta + z_i\gamma_j, \quad v_{ik} = x_{ik}\beta + z_i\gamma_k, \quad \forall i = 1, 2, \dots, N.$$

$$\text{On a alors : } \frac{\partial \log p_{ij}}{\partial \log v_{ij}} = (1 - p_{ij})v_{ij} \quad \text{et} \quad \frac{\partial \log p_{ik}}{\partial \log v_{ik}} = -p_{ik}v_{ik}$$

c'est-à-dire que les élasticités croisées des probabilités par rapport à v_{ik} sont identiques pour tout j . La conséquence est alors que si la modalité k est éliminée de l'ensemble des choix, la probabilité de choisir une modalité j est égale à l'ancienne probabilité (quand k était incluse), multipliée par un facteur commun à toutes les modalités $j, j \neq k$. Le rapport entre la nouvelle et l'ancienne probabilité est égal à $\sum_j \exp(v_{ij}) / \sum_{j \neq k} \exp(v_{ij})$.

C. Un test de la propriété IIA

La propriété IIA provient de l'hypothèse d'indépendance (et d'homoscédastité) des termes d'erreur entre les différentes modalités². Par conséquent, cette propriété est mise en défaut si le rapport des probabilités p_{ij}/p_{ik} n'est pas indépendant des autres modalités $l, l \neq j, k$. Une façon simple de tester la propriété IIA est alors d'estimer le modèle en restreignant l'ensemble des choix, et de comparer les paramètres obtenus avec ceux du modèle complet. Si la propriété IIA est valide, inclure ou exclure des modalités tierces ne devrait pas modifier de façon significative les paramètres estimés. Si par contre les paramètres sont sensiblement différents, le rapport des probabilités dépendra des autres modalités et la propriété sera rejetée. La solution sera alors à rechercher dans une autre spécification du modèle ou une modification de la typologie des choix. Le test proposé (Hausman et McFadden, 1984) est basé sur la statistique de test de Hausman (1978), comparant un estimateur convergent, mais non efficace, à un estimateur convergent (et efficace) seulement sous l'hypothèse nulle. Dans notre cas, les paramètres estimés du modèle complet, notés β_C , seront convergents et efficaces sous l'hypothèse nulle H_0 : la propriété IIA est valide, et non convergents sinon. Les paramètres du modèle restreint, notés $\hat{\beta}$, seront toujours convergents mais non efficaces³. Par conséquent, la différence entre les matrices de variances-covariances de $\hat{\beta}$ et β_C sera une matrice semi-définie positive. La statistique de test est :

$$HT = (\hat{\beta} - \beta_C)' \left[Var(\hat{\beta}) - Var(\beta_C) \right]^{-1} (\hat{\beta} - \beta_C)$$

qui suit un χ^2 à K degrés de liberté, où K est la dimension du vecteur de paramètres.

2. En revanche, le caractère indépendamment et identiquement distribué (i.i.d.) d'un modèle n'implique pas nécessairement la propriété IIA. De plus, il est possible d'obtenir le Logit multinomial sans erreurs i.i.d. (voir Anderson, Palma et Thisse, 1992 ; Ben Akiva et Lerman, 1979).

3. Cette perte d'efficacité provient de ce que la variance de l'estimateur augmente du fait de la perte d'information occasionnée par l'élimination de certaines modalités.

D. Application numérique

On donne ici un exemple de test de la propriété IIA, tiré du logiciel EVIEWS, avec des données simulées. Il y a 3 modalités et 2 variables explicatives : *STORE* et *DISTRIB*. Les commandes EVIEWS sont les suivantes.

```
load "mlogit"
coef(2) a1 = 0.1
coef(2) b1 = 0.1
series stores = d1*stores1+d2*stores2+d3*stores3
series dist = d1*dist1+d2*dist2+d3*dist3
logl l11
l11.append logl logl1
l11.append xb1 = a1(1)*stores1+a1(2)*dist1
l11.append xb2 = a1(1)*stores2+a1(2)*dist2+b1(1)*inc
l11.append xb3 = a1(1)*stores3+a1(2)*dist3+b1(2)*inc
l11.append denom = exp(xb1)+exp(xb2)+exp(xb3)
l11.append logl1 = d1*xb1+d2*xb2+d3*xb3-log(denom)
l11.append deriv a1(1) grada1 a1(2) grada2 b1(1) gradb2 b1(2) gradb3
l11.append grada1 = stores-(stores1*exp(xb1)
+stores2*exp(xb2)+stores3*exp(xb3))/denom
l11.append grada2 = dist-(dist1*exp(xb1)+dist2*exp(xb2)
+dist3*exp(xb3))/denom
l11.append gradb2 = (d2-exp(xb2)/denom)*inc
l11.append gradb3 = (d3-exp(xb3)/denom)*inc
l11.ml(d)
show l11.output
```

La sortie du logiciel donnée ci-dessous est comparable aux applications précédentes avec EVIEWS, pour le Logit binaire et le Logit multinomial.

LogL : LL1

Method : Maximum Likelihood (Marquardt)

Date : 08/16/99 Time : 11 :23

Sample : 1 1000

Included observations : 1000

Evaluation order : By observation

Initial Values : A1(1)=0.10000, A1(2)=0.10000, B1(1)=0.10000,

B1(2)=0.10000

Convergence achieved after 10 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
A1(1)	0.672530	0.044925	14.97000	0.0000
A1(2)	-0.362920	0.025836	-14.04694	0.0000
B1(1)	0.667741	0.038313	17.42865	0.0000
B1(2)	0.772680	0.067733	11.40781	0.0000
Log likelihood	-494.5753		Akaike info criterion	0.997151
Avg. log likelihood	-0.494575		Schwarz criterion	1.016782
Number of Coefs.		4	Hannan-Quinn criter.	1.004612

On fournit ensuite à EVIEWS les instructions permettant de présenter le tableau croisé des effectifs observés et prédis, et de calculer les statistiques de contingence vues au chapitre 1.

```
' Calcul des proba. prédictes
series y2hat = exp(xb2)/denom
series y3hat = exp(xb3)/denom
series y1hat = 1-y2hat-y3hat
' Calcul de y prédit (yhat)
series y = (d2=1)*2+3*(d3=1)+1*(d2=0 and d3=0)
series yhat = 2*(y2hat>y3hat and y2hat>y1hat) + 3*(y3hat>y2hat and
y3hat>y1hat)
+ 1*(y1hat>=y3hat and y1hat>=y2hat)
' Table des valeurs prédictes
group predict y yhat
show predict.freq
```

La sortie ci-dessous rappelle le nombre de modalités pour la variable dépendante [1] et sa prédiction [2], calcule le nombre total de cellules [3] et présente les mesures d'association [4] : coefficient Phi [5], de Cramer [6] et de contingence [7]. Le tableau suivant fournit les statistiques de χ^2 (Pearson) [8] et le coefficient G [9], avec les p-values associées [10]. La ligne [11] avertit que les effectifs sont inférieurs à 5 dans 33 % des cas, ce qui doit inciter à la prudence dans l'utilisation de ces statistiques.

Tabulation of Y and YHAT

Date : 08/16/99 Time : 11 :23

Sample : 1 1000

Included observations : 1000

Tabulation Summary

Variable Categories

Y 3 [1]

YHAT 3 [2]

Product of Categories 9 [3]

Measures of Association Value [4]

Phi Coefficient 0.638274 [5]

Cramer's V 0.451328 [6]

Contingency Coefficient 0.538021 [7]

Test Statistics	df	Value	Prob [10]
Pearson X2 [8]	4	407.3933	0.0000
Likelihood Ratio G2 [9]	4	381.6082	0.0000

WARNING : Expected value is less than 5 in 33.33% of cells (3 of 9). [11]

Le dernier tableau contient les effectifs observés et prédis dans les 9 cas considérés.

	YHAT			
Count	1	2	3	Total
Y	1	317	100	0
	2	94	453	1
	3	5	27	3
Total	416	580	4	1000

On estime enfin une version restreinte du modèle en éliminant la modalité 3. Les commandes EVIEWS sont :

```
' Estimation modèle restreint : modalité 3 éliminée
coef(2) a2 = 0.1
coef(1) b2 = 0.1
logl 112
112.append logl logl2
112.append xb12 = a2(1)*stores1+a2(2)*dist1
112.append xb22 = a2(1)*stores2+a2(2)*dist2+b2(1)*inc
112.append denom2 = exp(xb12)+exp(xb22)
112.append logl2 = d1*xb12+d2*xb22-log(denom2)
112.append deriv a2(1) grada12 a2(2) grada22 b2(1) gradb22
112.append grada12 = d1*stores1+d2*stores2-(stores1*exp(xb12)
+stores2*exp(xb22))/denom2
112.append grada22 = d1*dist1+d2*dist2-(dist1*exp(xb12)
+dist2*exp(xb22))/denom2
112.append gradb22 = (d2-exp(xb22)/denom2)*inc
smpl all if d3=0
112.ml(d)
show 112.output
```

LogL : 11.2

Method : Maximum Likelihood (Marquardt)

Date : 08/16/99 Time : 11 :23

Sample : 1 1000 IF D3=0

Included observations : 965

Evaluation order : By observation

Initial Values : A2(1)=0.10000, A2(2)=0.10000, B2(1)=0.10000

Convergence achieved after 9 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
A2(1)	0.625877	0.051413	12.17352	0.0000
A2(2)	-0.342854	0.027998	-12.24569	0.0000
B2(1)	0.633954	0.042565	14.89384	0.0000
Log likelihood		-390.4249	Akaike info criterion	0.815388
Avg. log likelihood		-0.404585	Schwarz criterion	0.830535
Number of Coefs.		3	Hannan-Quinn criter.	0.821155

La statistique de test d'Hausman est calculée en utilisant les possibilités d'écriture matricielle du logiciel :

```
' Test d'Hausman pour propriété IIA
matrix var_1 = l11.coefcov
matrix(3,3) var_1
matrix var_2 = l12.coefcov
for !i=1 to 2
matrix(3,1) coef_!i
coef_!i(1,1) = a!i(1)
coef_!i(2,1) = a!i(2)
coef_!i(3,1) = b!i(1)
next
matrix hs = transpose(coef_2-coef_1)*inverse(var_2-var_1)
*(coef_2-coef_1)' p-value [12]
matrix hs_pval = 1-cchisq(hs[1],3) [13]
show hs
7.214020 [14]
show hs_pval
0.065380 [15]
```

L'expression du test est donnée en [12], celle de la p-value en [13]. Les résultats sont respectivement indiqués en [14] et [15]. La p-value associée au test étant de 0,0653, l'hypothèse nulle n'est pas rejetée. Par conséquent, la propriété IIA est valide pour ces données.

VIII. Modèles multinomiaux alternatifs

Afin de s'affranchir de la propriété IIA, plusieurs alternatives au modèle Logit multinomial (et Logit conditionnel) ont été proposées. Les premières conservent la spécification Logit, mais effectuent des regroupements entre modalités similaires, alors que la dernière, le Probit multinomial, adopte l'hypothèse de normalité et autorise la corrélation entre les niveaux d'utilité.

A. Le Logit multinomial emboîté

Le Logit multinomial emboîté (appelé parfois *Nested Logit*) s'utilise dans le cas simple où, sur trois modalités, deux sont très similaires. L'utilité indirecte

de l'agent i est :

$$U_{ij} = \mu_{ij} + \varepsilon_{ij} \quad j = 0, 1, 2$$

où μ_{ij} est une fonction déterministe des caractéristiques de l'agent i et des attributs de la modalité j . Notons ε_{i1} , ε_{i2} et ε_{i0} les termes aléatoires associés aux modalités. McFadden propose d'écrire la distribution conjointe bivariée suivante entre les modes similaires (1 et 2) :

$$F(\varepsilon_{i1}, \varepsilon_{i2}) = \exp \left\{ - \left[\exp(-\frac{\varepsilon_{i1}}{\rho}) + \exp(-\frac{\varepsilon_{i2}}{\rho}) \right]^\rho \right\} \quad 0 \leq \rho \leq 1$$

Cette distribution est la loi des valeurs extrêmes de type II. Le coefficient de corrélation entre ε_{i1} et ε_{i2} est $1 - \rho^2$. Si $\rho = 1$, on a indépendance, $F(\varepsilon_{i1}, \varepsilon_{i2})$ devient le produit de deux lois à valeurs extrêmes de type I, et l'on obtient alors le Logit multinomial. Pour la troisième modalité indépendante des deux autres, on suppose que ε_{i0} admet la distribution à valeurs extrêmes de type I : $F(\varepsilon_{i0}) = \exp[-\exp(-\varepsilon_{i0})]$.

Sous ces hypothèses, on a :

$$\Pr[y_i = 0] = \exp(\mu_{i0}) \left\{ \exp(\mu_{i0}) + \left[\exp(\frac{\mu_{i1}}{\rho}) + \exp(\frac{\mu_{i2}}{\rho}) \right]^\rho \right\}^{-1}$$

$$\Pr[y_i = 1 | y_i \neq 0] = \exp(\frac{\mu_{i1}}{\rho}) \left[\exp(\frac{\mu_{i1}}{\rho}) + \exp(\frac{\mu_{i2}}{\rho}) \right]^{-1}$$

et :

$$\Pr[y_i = 2 | y_i \neq 0] = \exp(\frac{\mu_{i2}}{\rho}) \left[\exp(\frac{\mu_{i1}}{\rho}) + \exp(\frac{\mu_{i2}}{\rho}) \right]^{-1}$$

Ces probabilités dépendent de $\mu_{i2} - \mu_{i0}$, $\mu_{i1} - \mu_{i0}$ et ρ . Le choix entre les deux modalités similaires est expliqué par un Logit binaire, celui entre l'alternative tierce et les deux autres l'est selon un Logit un peu particulier faisant intervenir une moyenne pondérée de $\exp(\mu_{i1})$ et de $\exp(\mu_{i2})$.

Si μ_{ij} dépend de caractéristiques x_i , le modèle s'apparente au Logit conditionnel, alors qu'il est compatible avec lui si μ_{ij} est une fonction de variables caractérisant simultanément l'individu et la modalité (x_{ij}).

B. Le Logit multinomial hiérarchisé

Ce modèle est aussi appelé « Logit multinomial séquentiel », sa structure s'apparentant à la fois au Logit multinomial, de par ses hypothèses distributionnelles, et aux modèles séquentiels de la section 4.IV, du fait aussi de la structure

imbriquée des choix le long d'un arbre de décision. Encore une fois, le terme « séquentiel » ne signifie pas que les décisions sont toujours prises suivant une chronologie bien définie ; ce modèle est compatible avec des séquences de décisions simultanées (Ben Akiva et Lerman, 1979).

► Exemple 4.2 : choix du logement

Considérons un ménage désireux d'acquérir un logement dans une certaine région. Deux types de choix sont envisagés : le choix du lieu de résidence et celui du logement. Le ménage peut choisir entre C localités, $j = 1, 2, \dots, C$ et N_j logements $k = 1, 2, \dots, N_j$ dans chaque localité. L'utilité indirecte dépend alors des attributs de la localité j et du logement k :

$$U_{jk} = \mu_{jk} + \varepsilon_{jk}$$

où l'indice de l'individu i est omis pour simplifier les notations. Les attributs de la commune j sont, par exemple, la facilité d'accès au réseau routier, l'existence de transports publics, la qualité des services publics locaux, les écoles, et les attributs associés au logement sont la surface, le prix, la taille du jardin, etc. Si l'on suppose que ε_{jk} suit une loi à valeur extrême de type I, la probabilité de choisir le k -ème logement dans la j -ème commune est :

$$p_{jk} = \frac{\exp(\mu_{jk})}{\sum_{m=1}^C \sum_{n=1}^{N_m} \exp(\mu_{mn})}$$

On pose $\mu_{jk} = X_{jk}\beta + Y_j\alpha$, où Y_j est le vecteur des attributs ne variant qu'avec la localité. On écrit la probabilité : $p_{jk} = p_k|j \times p_j$ de la probabilité conditionnelle :

$$p_k|j = \frac{p_{jk}}{p_j} = \frac{Pr(j \text{ et } k)}{Pr(j)}$$

On peut alors écrire :

$$p_{jk} = \frac{\exp(\mu_{jk})}{\sum_{l=1}^{N_j} \exp(\mu_{jl})} \times \sum_{l=1}^{N_j} \exp(\mu_{jl}) / \sum_{m=1}^C \sum_{n=1}^{N_m} \exp(\mu_{mn})$$

que l'on peut donc décomposer en :

$$p_k|j = \frac{\exp(\mu_{jk})}{\sum_{l=1}^{N_j} \exp(\mu_{jl})} = \frac{\exp(X_{jk}\beta) \exp(Y_j\alpha)}{\sum_{l=1}^{N_j} \exp(X_{jl}\beta) \exp(Y_j\alpha)}$$

et :

$$p_j = \frac{\sum_{l=1}^{N_j} \exp(\mu_{jl})}{\sum_{m=1}^C \sum_{n=1}^{N_m} \exp(\mu_{mn})} = \frac{\exp(Y_j\alpha) \sum_{l=1}^{N_j} \exp(X_{jl}\beta)}{\sum_{m=1}^C \exp(Y_m\alpha) \sum_{n=1}^{N_m} \exp(X_{mn}\beta)}$$

En définissant une « variable d'inclusion » :

$$I_j = \log \left(\sum_{l=1}^{N_j} \exp(X_{jl}\beta) \right)$$

on écrit :

$$p_k|j = \frac{\exp(X_{jk}\beta)}{\exp(I_j)} \quad \text{et} \quad p_j = \frac{\exp(Y_j\alpha + I_j)}{\sum_{m=1}^C \exp(Y_m\alpha + I_m)}$$

Deux méthodes d'estimation sont possibles. La première consiste à écrire la log-vraisemblance comme suit :

$$\log L = \sum_{i=1}^N \sum_j \sum_k y_{ijk} \log [p_{ikj} \times p_{ij}]$$

soit une méthode d'information complète intégrant l'ensemble des choix. Une autre méthode, plus simple, utilise le fait que β et γ peuvent être estimés séquentiellement. La méthode d'estimation est :

- 1) Estimer β à partir de $p_k|j = \exp(X_{jk}) / \sum_{l=1}^{N_j} \exp(X_{jl}\beta)$, comme un Logit conditionnel ;
- 2) Calculer les valeurs d'inclusion I_j puis estimer α à partir de p_j ci-dessus, toujours à l'aide de la procédure du Logit conditionnel.

→ **Remarque**

La variable d'inclusion s'interprète comme une moyenne sur les logements d'une même localité : $Y_j\alpha + I_j$ représente donc une espérance d'utilité sur la commune j . Il n'est pas indispensable d'avoir un coefficient de 1 associé aux valeurs d'inclusion. Si l'on écrit :

$$p_j = \frac{\exp(Y_j\alpha + \sigma I_j)}{\sum_{m=1}^C \exp(Y_m\alpha + \sigma I_m)}$$

on voit que l'on obtient le modèle Logit conditionnel de départ en imposant $\sigma = 1$, et I_j est traitée de la même façon qu'un attribut Y_j , mais avec un coefficient de 1. En revanche, si cette restriction n'est pas imposée, on obtient le modèle Logit hiérarchisé, et le coefficient σ mesure la similitude des choix observés au niveau inférieur de la structure. Dans ce dernier cas, on estime simultanément les paramètres α et σ à l'étape 2) ci-dessus.

On a vu que dans cet exemple, le modèle est du type Logit conditionnel : les attributs diffèrent mais les paramètres (β) sont identiques pour toutes les alternatives. Donc, il n'est pas indispensable d'estimer β à partir de toutes les localités, une seule suffit *a priori*.

Il est aisément de généraliser le modèle à plus de deux niveaux (branches) (voir Maddala, 1983). Prenons par exemple le cas de trois niveaux de choix, avec I, j, k . L'utilité indirecte s'écrit : $U_{ijk} = \mu_{ijk} = X_{ijk}\beta + Y_{ij}\alpha + Z_{i}\gamma + \varepsilon_{ijk}$ et la probabilité associée au choix (I, j, k) est $P_{ijk} = P_{k|I,j} \times P_{j|I} \times P_I$.

La probabilité conditionnelle $P_{k|I,j}$ n'impliquera que les paramètres β : $P_{k|I,j} = \exp(X_{ijk}\beta) / \sum_n \exp(X_{ijn}\beta)$ et l'on définit les valeurs d'inclusion : $I_{ij} = \log [\sum_n \exp(X_{ijn}\beta)]$. On a alors : $P_{j|I} = \exp(Y_{ij}\alpha + I_{ij}) / \sum_m \exp(Y_{im}\alpha + I_{im})$. On définit un autre ensemble de valeurs d'inclusion :

$$J_I = \log \left(\sum_m \sum_n \exp(X_{lm}\beta + Y_{lm}\alpha) \right) = \log \left(\sum_m \exp(Y_{lm}\alpha + I_{lm}) \right)$$

et l'on a : $P_I = \exp(Z_I\gamma + J_I) / \sum_h \exp(Z_h\gamma + J_h)$.

C. Application : transport aérien ou terrestre

On applique le modèle Logit multinomial hiérarchisé au problème de transport suivant. Quatre modes de transport existent entre Sydney et Melbourne en Australie : voiture, car, avion, train. Le premier niveau est défini par le type de transport : 1) air, ou 2) terre, le second niveau reprenant les quatres modes :

Voyage	Air	[Avion
	Terre	[Bus Train Voiture

Le choix d'un mode particulier dépend d'une constante propre au mode et d'une mesure du coût du voyage, GC . Le choix entre Air et Terre dépend de deux variables : le revenu du ménage $HINCA$ et le nombre de voyageurs $PSIZEA$. Pour assurer l'identification et éviter la possible colinéarité dans le modèle, on contraint à 0 la contrainte associée au bus, et l'on multiplie les variables $HINCA$ et $PSIZEA$ par une variable indicatrice égale à 1 si le mode Air est choisi. Enfin, $MODE$ est la variable polytymique du choix. (Voir la documentation de LIMDEP à propos de ce modèle.)

Les commandes pour estimer ce modèle par une procédure Logit multinomial hiérarchisé sont, avec le logiciel LIMDEP :

```

Sample; 1 - 840 $ Nlogit; lhs=mode; start=logit;
effects : gc(air,train,bus,car);
choices=Air,Train,Bus,Car/.14,.13,.29,.44;
tree = travel(fly(Air),land(Train,Bus,Car));
model :U(Air) = BA + Bz*gc/
U(Train) = BT + BZ*gc/
U(Bus) = BZ*gc/
U(Car) = BC + BZ*gc/
U(fly,land) = A1*hinca + A2*psizea $

```

Les résultats d'estimation sont donnés ci-dessous. Le logiciel rappelle l'étendue [1] et la taille de l'échantillon [2], indique que les valeurs initiales pour maximiser la log-vraisemblance sont tirées du modèle non hiérarchisé [3], que la méthode est le maximum de vraisemblance [4], avec la variable dépendante *Choice* [5] et 210 observations [6]. LIMDEP donne le nombre d'itérations avant convergence [7], la valeur de la log-vraisemblance [8] et celle du modèle pour *Choice* [9]. Trois statistiques sont fournies en [10] : la valeur de la log-vraisemblance, le coefficient R^2 et le coefficient R^2 ajusté, cela pour les modèles restreints sans coefficients [11] et avec constantes associées aux choix seulement [12]. Le test du χ^2 sous l'hypothèse d'absence de hiérarchie est donné en [13] avec sa p-value [14]. Le logiciel présente ensuite la valeur de la log-vraisemblance pour le modèle restreint à *Branch* [15], indique que les variables ont été corrigées du biais de sélection [16], que les données correspondent à des choix individuels [17], et que le nombre d'observations pour ce modèle restreint est de 210 [18]. Les paramètres estimés figurent dans le tableau final, séparés en deux parties : ceux du modèle entre les alternatives [19] et entre les branches [20].

| LIMDEP Estimation Results Run log line 4 Page 1

| Sample was reset : SAMPLE 1-840 [1]

| Current sample contains 840 observations. [2]

| Start values obtained from nonnested model [3]

| Maximum Likelihood Estimates [4]

| Dependent variable Choice [5]

| Number of observations 210 [6]

| Iterations completed 15 [7]

| Log likelihood function -241.1614 [8]

| Log-L for Choice model = -241.1614 [9]

| R2=1-LogL/LogL* Log-L fncc R-sqrd RsqAdj [10]

| No coefficients -312.5500 0.22841 0.22348 [11]

| Constants only -283.7588 0.15012 0.14469 [12]

| Chi-squared[1] = 85.19478 [13]

- | Significance for chi-squared = 1.00000 [14]
| Log-L for Branch model = -121.7766 [15]
| Vars. corrected for choice based sampling [16]
| Response data are given as ind. choice. [17]
| Number of obs.= 210 , skipped 0 bad obs. [18]

Variable	Coefficient	Standard Error	z=b/s.e.	P	[Z >z]	Mean of X
Model for Choice Among Alternatives [19]						
BA	-1.320796	0.47634	-2.773	0.00556		
BZ	-0.3029683E-01	0.72756E-02	-4.164	0.00003		
BT	-0.2915457	0.25975	-1.122	0.26169		
BC	-0.1630691	0.17443	-0.935	0.34984		
Model for Choice Among Branches [20]						
A1	0.1344577E-01	0.82205E-01	0.164	0.87007		
A2	-0.7758111	0.16501	-4.702	0.00000		

D. Le Probit multinomial

Ce modèle permet de s'affranchir de l'hypothèse IIA présente dans le Logit multinomial. On suppose que les termes d'erreurs dans le modèle à utilité aléatoire suivent une distribution normale multivariée. Considérons les trois alternatives (1,2,3) avec les utilités indirectes associées (on élimine l'indice i de l'individu pour simplifier) :

$$U_j = \mu_j + \varepsilon_j \quad j = 1, 2, 3$$

Les termes aléatoires ($\varepsilon_1, \varepsilon_2, \varepsilon_3$) suivent une loi normale trivariée avec comme moyenne un vecteur de 0 et une matrice de variances-covariances Σ :

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

Considérons la probabilité que la première alternative soit choisie. On a :

$$Prob(U_1 > U_2, U_1 > U_3) = Prob(\varepsilon_2 - \varepsilon_1 < \mu_1 - \mu_2, \varepsilon_3 - \varepsilon_1 < \mu_1 - \mu_3)$$

On pose :

$$\eta_{21} = \varepsilon_2 - \varepsilon_1, \eta_{31} = \varepsilon_3 - \varepsilon_1, \mu_{12} = \mu_1 - \mu_2 \quad \text{et} \quad \mu_{13} = \mu_1 - \mu_3$$

η_{21} et η_{31} suivent alors une distribution normale bivariée de matrice de variances-covariances Σ_1 :

$$\Sigma_1 = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} \\ \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix}$$

et la probabilité que l'alternative 1 soit choisie est alors :

$$p_1 = \int_{-\infty}^{\mu_{12}} \int_{-\infty}^{\mu_{31}} f(\eta_{21}, \eta_{31}) d\eta_{21} d\eta_{31}$$

où $f(\eta_{21}, \eta_{31})$ est la densité de la loi normale bivariée, avec comme moyenne un vecteur de 0 et une matrice de variances-covariances Σ_1 . Les probabilités p_2 et p_3 peuvent être calculées de façon similaire, et la vraisemblance constituée du produit de ces probabilités est maximisée avec les techniques habituelles. Les paramètres à estimer sont ceux associés aux caractéristiques x dans les fonctions μ_j , ainsi que les variances-covariances dans la matrice Σ .

On voit que la probabilité associée à chaque alternative s'exprime comme une intégrale multiple de dimension $m - 1$. Comme dans la pratique, les algorithmes d'intégration numérique fournissent des valeurs satisfaisantes pour des intégrales triples au maximum, le Probit multinomial est donc bien adapté aux modèles où le nombre de modalités ne dépasse pas 4.

IX. Extensions

On a vu que le Probit multinomial ne peut être appliqué aisément qu'à un petit nombre d'alternatives. Si tel n'est pas le cas, il est très difficile d'évaluer numériquement les probabilités des différents choix, qui prennent la forme d'intégrales multiples. Une solution consiste à recourir à des méthodes de simulation numérique.

A. Estimateur MSM de McFadden

L'estimateur par la *méthode des moments simulés* (MSM, McFadden, 1989) repose sur le principe que la condition du premier ordre dans la maximisation de la log-vraisemblance peut être utilisée comme une condition d'identification des paramètres. Ainsi, la méthode des GMM vue au chapitre 2 peut en principe être employée, au lieu du maximum de vraisemblance. Comme le modèle Probit est difficile à traiter avec un nombre important de modalités, la méthode évite d'écrire directement les probabilités associées aux choix comme des intégrales multiples, mais consiste à simuler les niveaux d'utilité indirecte à partir d'une loi normale multivariée, et à comparer ces niveaux pour construire des fréquences empiriques. La démarche est la suivante.

Notons $p_{ij}(b)$ la probabilité que l'individu i choisisse le mode j , en fonction des paramètres b . Ces derniers comprennent les paramètres structurels de

l'utilité indirecte pour les différentes modalités, ainsi que ceux de la distribution des termes ε_{ij} . La condition du premier ordre de maximisation de la log-vraisemblance (divisée par N) est :

$$\frac{1}{N} \sum_{i=1}^N \sum_{j=0}^m y_{ij} \frac{\partial p_{ij}(b)}{\partial b} \frac{1}{p_{ij}(b)} = 0$$

Puisque $\sum_{j=0}^m p_{ij} = 1 \forall i$, on a :

$$\sum_{j=0}^m \frac{\partial p_{ij}(b)}{\partial b} = \sum_{j=0}^m p_{ij}(b) \frac{\partial p_{ij}(b)}{\partial b} \frac{1}{p_{ij}(b)} = 0$$

que l'on remplace dans la condition du premier ordre pour obtenir :

$$\boxed{\frac{1}{N} \sum_{i=1}^N \sum_{j=0}^m [y_{ij} - p_{ij}(b)] \frac{\partial \log p_{ij}(b)}{\partial b} = 0}$$

Cette dernière expression peut être vue comme une condition de moments, avec $[y_{ij} - p_{ij}(b)]$ le « résidu » et $\partial \log p_{ij}(b)/\partial b$ le vecteur d'« instruments ». Notons que dans ce cas, les instruments dépendent des paramètres b . Soit l'utilité indirecte $U_{ij} = x_{ij}\beta + \varepsilon_{ij}$, où, dans le cas du Probit multinomial, les ε_{ij} sont distribués selon une loi normale. On écrit le modèle sous une forme plus compacte avec des vecteurs y_i, ε_i, p_i de dimension $(m+1)$: par exemple $y_i = (y_{i0}, \dots, y_{im})'$. La condition du premier ordre est :

$$\frac{1}{N} \sum_{i=1}^N z'_i (y_i - p_i) = 0$$

où :

$$z_i = \frac{\partial \log p_i(b)}{\partial b}$$

et ε_i suit une distribution normale $N(0, \Omega)$, où Ω est une matrice de dimension $(m+1) \times (m+1)$. On s'autorise des corrélations entre les différentes modalités, c'est-à-dire que Ω contient des éléments hors-diagonale non nuls.

La première étape consiste à engendrer des valeurs simulées u_i à partir d'une loi normale $N(0, I_{m+1})$. Notons S le nombre de tirages aléatoires et $u_i(s)$ un tirage particulier de dimension $m+1$. On conservera ces valeurs tout au long de l'algorithme. Des valeurs initiales sont choisies pour le vecteur de paramètres b , à savoir β et Ω . On obtient ensuite des valeurs simulées $\eta_i(s) = (\eta_{i0}(s), \dots, \eta_{im}(s))'$ à partir de $u_i(s)$ par :

$$\eta_i(s) = A u_i(s) \quad s = 1, 2, \dots, S$$

où A est une matrice obtenue par la décomposition : $AA' = \Omega$, et des valeurs de l'utilité indirecte :

$$U_{ij}(s) = x_{ij}\beta + \eta_{ij}(s) \quad s = 1, 2, \dots, S$$

La fréquence estimée que i choisisse j est :

$$f_{ij}(b) = \frac{1}{S} \sum_{s=1}^S \mathbb{I}[U_{ij}(s) > U_{ik}(s), k \neq j]$$

et constitue un estimateur de $p_{ij}(b)$. La dernière étape consiste à remplacer $p_i(b)$ par $f_i(b)$, obtenu en empilant les $f_{ij}(b)$ ci-dessus, et à utiliser la formule itérative issue des GMM :

$$\begin{aligned} b_1 = b_0 + & \left[\frac{1}{N} \sum_{i=1}^N z'_i(b_0)[y_i - p_i(b_0)][y_i - p_i(b_0)]' z_i(b_0) \right]^{-1} \\ & \times \frac{1}{N} \sum_{i=1}^N z'_i(b_0)[y_i - p_i(b_0)] \end{aligned}$$

Les nouvelles valeurs des paramètres b_1 remplacent b_0 , de nouveaux $\eta_i(s)$, $U_{ij}(s)$ sont calculés, et ainsi de suite jusqu'à convergence, toujours avec les valeurs $u_i(s)$. L'estimateur MSM est convergent et asymptotiquement normal :

$$\sqrt{N}(\hat{b} - b) = N \left(0, \left(1 + \frac{1}{S} \right) \Sigma \right)$$

où Σ est la matrice de variances-covariances de l'estimateur du maximum de vraisemblance :

$$\Sigma = \left[\frac{1}{N} \sum_{i=1}^N z'_i(b)[y_i - p_i(b)][y_i - p_i(b)]' z_i(b) \right]^{-1}$$

Si $S \rightarrow +\infty$, l'estimateur MSM est aussi efficace que celui du maximum de vraisemblance.

B. L'estimateur GHK

La méthode des moments simulés, ci-dessus, a plusieurs inconvénients :

- 1) elle peut nécessiter un nombre de tirages aléatoires très important pour estimer avec précision une faible probabilité ;
- 2) la fréquence empirique f_{ij} n'est pas une fonction dérivable à cause de l'indicateur $\mathbb{I}[U(s_{ij}) > U(s_{ik})]$;

3) la fréquence f_{ij} peut être nulle, ce qui rend périlleux le calcul de l'instrument $z_i = \partial \log p_i(b) / \partial b$.

D'autres méthodes de simulation des probabilités conjointes issues de la loi normale, qui n'ont pas ces inconvénients, ont alors été proposées (voir Stern, 1992; Hajivassiliou et Ruud, 1994; Geweke, Keane et Runkle, 1994). Nous présentons une des techniques les plus utilisées et efficaces, l'estimateur GHK (d'après Geweke, Hajivassiliou et Keane : voir Hajivassiliou, McFadden et Ruud, 1992).

Considérons un vecteur Y_i^* de dimension $(m \times 1)$ suivant une loi $N(\mu_i^*, \Omega)$. Ce vecteur représente les termes aléatoires ε_{ij} associés aux choix possibles pour l'individu i . L'événement $y_{ij} = 1$ est écrit de façon générale $a_i^* \leq M Y_i^* < b_i^*$ où $a_i^*, b_i^* \in R$ sont les bornes dépendant des caractéristiques de l'individu i . On définit $\mu_i = M\mu_i^*$, les bornes centrées $a_i = a_i^* - M\mu_i^*$ et $b_i = b_i^* - M\mu_i^*$, et la matrice de variances-covariances $\Sigma = M\Omega M' = LL'$, où L est la matrice triangulaire inférieure de la décomposition de Choleski. Pour tout vecteur e et toute matrice A , $e_{<j}$ est le sous-vecteur constitué des $j - 1$ éléments de e , et $A_{j,<j}$ est le vecteur des premiers $j - 1$ éléments de la j -ème ligne de A .

Un tirage aléatoire e à partir d'une loi $N(0, 1)$ tronquée par la condition $a_i \leq e < b_i$ peut être obtenu par :

$$e = \Phi^{-1} [\Phi(a_i) + U \times (\Phi(b_i) - \Phi(a_i))] \quad U \sim U(0, 1)$$

où Φ est la fonction de répartition de la loi $N(0, 1)$. On engendre un vecteur aléatoire $e_i = (e_{1s}, e_{2s}, \dots, e_{ms})$ de dimension m avec la formule récursive suivante :

$$\begin{cases} e_{1s} \sim N(0, 1) \text{ tel que } a_{11} \leq l_{11} e_{1s} < b_{11} \\ e_{2s} \sim N(0, 1) \text{ tel que } a_{22} \leq l_{21} e_{1s} + l_{22} e_{2s} < b_{22} \\ \cdots & \cdots \\ e_{ms} \sim N(0, 1) \text{ tel que } a_{mm} \leq l_{m1} e_{1s} + \cdots + l_{mm} e_{ms} < b_{mm} \end{cases}$$

Les probabilités associées à chaque élément e_{js} sont :

$$\begin{cases} Q_{1s} & = \text{Prob}(a_{11}/l_{11} \leq e_{1s} < b_{11}/l_{11}) \\ \cdots & \cdots \\ Q_j(e_{1s}, e_{2s}, \dots, e_{j-1s}) & = \text{Prob}[(a_{jj} - L_{j,<j} \times e_{<j})/l_{jj} \leq e_{js} \\ & < (b_{jj} - L_{j,<j} \times e_{<j})/l_{jj}] \end{cases}$$

Avec S tirages aléatoires pour chaque individu, on obtient les vecteurs $e_{is} = (e_{i1s}, e_{i2s}, \dots, e_{i(j-1)s})$, $i = 1, 2, \dots, N$ et les probabilités Q_{ij} , $i = 1, 2, \dots, N$; $j = 1, 2, \dots, m$. La probabilité de l'événement ($y_{ij} = 1$) $\equiv (a_i^* \leq M Y_i^* < b_i^*)$ est estimée par :

$$\boxed{\frac{1}{S} \sum_{s=1}^S \prod_{j=1}^m Q_{ij}(e_{i1s}, e_{i2s}, \dots, e_{i(j-1)s})}$$

ce qui fournit la contribution de l'individu i à la vraisemblance. Il reste à maximiser la somme des logarithmes de ces contributions sur tous les individus.

À retenir

Il existe trois catégories de modèles multinomiaux : ordonnés, séquentiels et non ordonnés. Dans la dernière, les modèles Logit multinomiaux sont les plus utilisés et sont une extension directe du Logit binaire, du chapitre 3, au cas multinomial (polytomique). Si ces Logit multinomiaux sont simples à utiliser, ils posent des problèmes de cohérence en raison de la propriété peu réaliste de l'indépendance des états non pertinents (IIA). Pour pallier cet inconvénient, des modèles alternatifs sont développés : Logit hiérarchisé et Probit multinomial. Ces derniers s'affranchissent dans une certaine mesure des inconvénients précités, mais requièrent des techniques d'estimation plus complexes (méthodes de simulation). Les modèles multinomiaux possèdent une motivation intéressante en termes d'utilité aléatoire, ce qui les rend très utiles dans les problèmes de choix discrets.

Exercices

Exercice n° 1 : modèles automobiles

Un constructeur automobile produit trois modèles de véhicules : deux portes, cinq portes, et monospace. Afin de cibler les publics intéressés par ces modèles, il commande une enquête auprès de ménages ayant acquis l'un de ces trois modèles. L'échantillon porte sur 30 ménages, les variables sont *TYPE*, égal à 1 si deux portes, 2 si cinq portes et 3 si monospace ; *AGE*, âge du chef de famille l'année de l'achat ; *ENFPLUS2*, égal à 1 si le ménage a plus de 2 enfants. Les données sont les suivantes :

Obs.	AGE	ENFPLUS2	TYPE	Obs.	AGE	ENFPLUS2	TYPE
1	34,00	0,000	1,000	16	28,00	1,000	3,000
2	36,00	1,000	3,000	17	27,00	0,000	1,000
3	23,00	0,000	1,000	18	27,00	1,000	2,000
4	29,00	1,000	2,000	19	37,00	1,000	3,000
5	34,00	0,000	1,000	20	31,00	0,000	2,000
6	28,00	1,000	3,000	21	29,00	0,000	1,000
7	26,00	0,000	1,000	22	38,00	0,000	3,000
8	37,00	1,000	3,000	23	37,00	1,000	2,000
9	27,00	0,000	1,000	24	37,00	1,000	1,000
10	41,00	1,000	3,000	25	36,00	1,000	2,000
11	35,00	1,000	3,000	26	27,00	0,000	3,000
12	23,00	1,000	2,000	27	26,00	0,000	1,000
13	28,00	0,000	1,000	28	23,00	1,000	1,000
14	25,00	0,000	1,000	29	29,00	1,000	2,000
15	30,00	0,000	2,000	30	40,00	1,000	3,000

1) Estimer par un Logit multinomial le modèle de choix de ces ménages, où $TYPE$ est la variable dépendante, AGE et $ENFPLUS2$ les variables explicatives. On prendra le cas $TYPE = 3$ comme cas de référence.

2) Éliminer le cas $TYPE = 2$ de l'échantillon et estimer le modèle par un Logit binaire simple. Commenter les résultats et comparer avec ceux obtenus en 1).

Exercice n° 2 : test IIA

On estime un modèle Logit multinomial avec 4 modalités et 4 variables explicatives, x_1 à x_4 , qui ne dépendent que des individus. Le vecteur des paramètres estimés est (y compris le terme constant) :

$$\hat{\beta}_1 = (-1,8934; -0,2150; 0,0257; -0,1399; 0,3773)'$$

et la matrice de variances-covariances est :

$$\hat{\Sigma}_1 = \begin{bmatrix} 0,016\,157 & -0,002\,531 & -0,000\,106 & -0,003\,072 & -0,004\,683 \\ -0,002\,531 & 0,002\,222 & -0,000\,0117 & -0,001\,072 & 0,000\,638 \\ -0,000\,106 & -0,000\,0117 & 0,000\,0114 & -0,000\,0165 & -0,000\,0298 \\ -0,003\,072 & -0,001\,072 & -0,000\,0165 & 0,003\,910 & -0,001\,482 \\ -0,004\,683 & 0,000\,638 & -0,000\,0298 & -0,001\,482 & 0,011\,160 \end{bmatrix}$$

Le modèle est ensuite estimé en éliminant la quatrième modalité. Les paramètres estimés du modèle restreint sont :

$$\hat{\beta}_2 = (-1,7545; -0,1298; 0,0317; -0,1932; 0,2720)'$$

et la nouvelle matrice de variances-covariances est :

$$\hat{\Sigma}_2 = \begin{bmatrix} 0,036\,099 & -0,007\,796 & -0,000\,524 & -0,004\,934 & -0,009\,482 \\ -0,007\,796 & 0,008\,113 & -0,000\,0308 & -0,005\,784 & 0,000\,834 \\ -0,000\,524 & -0,000\,0308 & 0,000\,0552 & -0,000\,0713 & -0,000\,210 \\ -0,004\,934 & -0,005\,784 & -0,000\,0713 & 0,017\,195 & -0,002\,187 \\ -0,009\,482 & 0,000\,834 & -0,000\,210 & -0,002\,187 & 0,038\,937 \end{bmatrix}$$

1) Calculer les statistiques de Student associées aux paramètres dans les deux cas (modèle 1 complet, modèle 2 restreint). Les estimateurs du modèle complet sont-ils plus efficaces ?

2) Tester la propriété IIA à l'aide du test de spécification d'Hausman. Cette propriété est-elle vérifiée ?

Exercice n° 3 : enquête de consommation

On conduit une enquête de consommation portant sur un bien (bien A). Il n'existe qu'un seul bien de ce type sur le marché et le consommateur peut l'acheter ou non. Les variables observées sont : y_i , où $y_i = 1$ si le consommateur achète le bien, 0 sinon ; R_i le revenu du consommateur ; I_i une variable indicatrice de résidence ($I_i = 1$ si habitation urbaine, 0 sinon).

1) On souhaite estimer la probabilité associée au choix du bien A par un modèle basé sur la distribution logistique. Écrire la probabilité d'achat du bien, et la log-vraisemblance associée, en fonction des paramètres et des variables observables. Indication : on notera β le vecteur des paramètres.

2) Une première estimation par maximum de vraisemblance donne les résultats suivants :

Variable	Estimation	Écart type
Constante	-0,50	0,34
R	0,25	0,04
I	0,36	0,16

Les paramètres estimés sont-ils significatifs ? La log-vraisemblance du modèle est égale à -0,5612. Écrire la log-vraisemblance dans le cas où les paramètres associés à R et I sont nuls. Indication : on notera β_0 le paramètre constant du modèle. La valeur numérique de la log-vraisemblance de ce modèle contraint est -3,7762. Calculer le test de significativité globale des paramètres associés à R et I. L'hypothèse de nullité de ces paramètres est-elle validée ? Calculer la probabilité qu'un consommateur, dont le revenu R est égal à 2,50 et qui réside en zone urbaine, achète le bien A.

3) On suppose à présent qu'un second bien (bien B) est introduit sur le marché. Un consommateur ne peut acheter qu'un seul de ces deux biens ou n'en achète aucun. Pour estimer la probabilité d'achat des biens (A ou B), quel modèle multinomial peut-on utiliser ? Décrire les avantages et les inconvénients associés au modèle proposé.

4) Les résultats d'estimation par Logit sont donnés ci-dessous. *Le cas « pas d'achat » est pris comme référence.*

Variable	Choix A	Choix B
Constante	0,21	0,30
R	-0,12	-0,34
I	0,46	0,28

Calculer la probabilité qu'un consommateur de revenu 2,5 et résidant en zone urbaine achète le bien A, puis celle qu'il achète le bien B. Comparer la valeur numérique de la probabilité d'achat du bien A avec celle obtenue en 2). Étant donné la caractéristique fondamentale des modèles Logit multinomiaux (hypothèse IIA), que pouvez-vous conclure ?

5. Les modèles à variable dépendante limitée

Ce chapitre introduit des modèles de régression dans lesquels la variable à expliquer n'est observée que sur un intervalle. Selon que les variables explicatives sont disponibles ou non, on parle de modèles tronqués ou censurés, et les méthodes d'estimation utilisent à la fois des aspects de la régression linéaire simple et des résultats antérieurs sur les variables expliquées qualitatives. Un cas important est celui de la troncature auxiliaire, dans lequel l'affectation des individus de l'échantillon à différents régimes dépend du signe d'une variable auxiliaire. Bien que plus sophistiqués, les modèles étudiés dans ce chapitre ne font appel qu'à des techniques déjà étudiées dans les chapitres 2 et 3.

Objectifs du chapitre : présenter les modèles tronqués et censurés, introduire le modèle Tobit simple, puis les modèles Tobit généralisés ; détailler les modèles de sélection endogène et à régimes.

Concepts clés étudiés : variable dépendante limitée, échantillons censurés et tronqués, modèles Tobit, modèles avec sélection endogène, modèles de déséquilibre.

I. Introduction

Nous considérons dans ce chapitre des modèles pour lesquels la variable dépendante est continue, mais n'est observable que sur un certain intervalle. Nous verrons que ces modèles sont pratiquement tous des extensions d'un modèle appelé Tobit et qu'ils contiennent comme ingrédient essentiel l'espérance de la

variable dépendante continue, conditionnellement à la réalisation d'un événement discret. Dans la mesure où, contrairement aux chapitres précédents, nous expliquons à présent la valeur d'une variable continue, une partie de la modélisation fera référence à la régression linéaire. Et comme la variable dépendante n'est observable que sur un certain intervalle, la probabilité de se situer dans ce dernier devra être prise en compte, avec des notions similaires à celles utilisées dans les modèles qualitatifs. Prenons quelques exemples des modèles que ce chapitre permettra de traiter.

► Exemple 5.1

Des ménages sont interrogés sur leur disposition à payer pour la préservation d'un site naturel à proximité de leur lieu de résidence. Ces enquêtes font appel à l'approche de l'évaluation contingente consistant à construire une fonction de demande pour un bien naturel non marchand¹. La variable continue que l'on cherche à expliquer est le montant que les ménages déclarent, en fonction de leurs caractéristiques sociodémographiques : âge, revenu du foyer, éducation, distance par rapport au site, etc. Or, dans ce genre d'enquêtes, beaucoup de ménages reportent un montant nul pour leur disposition à payer. Le problème qui se pose à l'économètre est de fournir la meilleure estimation de la moyenne de la disposition à payer, ainsi que la fonction reliant cette dernière aux caractéristiques sociodémographiques, en tenant compte du fait que, pour une certaine part de la population enquêtée, la disposition à payer est égale à 0.

► Exemple 5.2

On s'intéresse à l'activité d'une population de salariés à temps partiel, et on fait l'hypothèse que les personnes enquêtées acceptent de travailler si leur taux de salaire dépasse un certain niveau, leur coût d'opportunité du travail. Le taux de salaire observé dépend de variables tels l'expérience professionnelle, le niveau de qualification, le nombre d'heures travaillées, le secteur d'activité, etc. Le coût d'opportunité du travail est différent selon les individus et dépend de la distance du lieu de résidence au lieu de travail, du nombre d'enfants (notamment en bas âge), du niveau des allocations de chômage, du salaire du conjoint, etc. L'économètre cherche à expliquer les variations du taux de salaire entre les différents individus de l'échantillon,

1. Les sites naturels dits récréatifs ne sont pas marchands dans la mesure où leur accès est libre : la fréquentation du site a un impact positif sur l'utilité du ménage, pour un prix nul, hormis le coût de transport. On généralise souvent cet exemple aux biens environnementaux comme la qualité de l'eau ou de l'air, la préservation d'une espèce animale ou végétale menacée, etc. Ces biens ne faisant pas l'objet d'un échange, il est très difficile de leur associer une valeur monétaire sans recourir à des enquêtes.

alors que cette variable n'est positive que pour les personnes en activité lors de l'enquête. Les autres individus ont une valeur nulle pour cette variable, mais les caractéristiques mentionnées plus haut sont néanmoins observables. Il est alors nécessaire d'expliquer le taux de salaire en fonction des caractéristiques relatives au secteur d'activité et à la productivité de l'individu (formation, expérience, etc.), compte tenu du fait que les valeurs positives de ce taux indiquent que le salaire obtenu est supérieur au coût d'opportunité du travail. Ce dernier est inconnu, mais dépend de variables observables pour chaque personne.

► Exemple 5.3

Une enquête porte sur les entreprises industrielles et l'économètre s'intéresse à la stratégie d'endettement de chacune. Or, le service statistique ayant mis à disposition les données a éliminé de l'échantillon les entreprises dont l'endettement était inférieur à un certain pourcentage de leur résultat d'exploitation. Pour ces dernières que le service statistique jugeait peut-être peu intéressantes ou peu représentatives, aucune variable n'est observée. La question qui se pose alors est de savoir comment estimer un modèle décrivant la stratégie d'endettement des entreprises, sachant que l'échantillon disponible est le résultat d'un filtrage sur une variable d'intérêt et que la variable « endettement » n'est observable que si elle est supérieure à une certaine borne.

Avant de présenter les modèles Tobit et les modèles de sélection permettant de traiter les différents cas illustrés par ces exemples, nous donnons deux définitions importantes.

II. Modèles censurés et tronqués

On considère un ensemble de couples $\{y_i^*, x_i^*\}$, où y_i^* est engendré par un processus aléatoire tel que $E(y_i^* | x_i^*) = x_i^* \beta$, x_i^* est un vecteur $1 \times K$ de variables pré-déterminées et β est un vecteur $(K \times 1)$ de paramètres. Un échantillon de taille N est constitué à partir de ces couples mais la variable y_i^* n'est pas toujours disponible. On observe à la place la variable y_i qui est reliée à la variable originale y_i^* par la règle suivante :

$$\begin{cases} y_i &= y_i^* \quad \text{si } y_i^* > c_i \\ &c_i \quad \text{sinon} \end{cases}$$

où c_i est une constante qui peut être commune aux N observations. L'exemple 5.1 donné plus haut s'applique si y_i^* est la disposition à payer, y_i le montant reporté lors de l'enquête, $c_i = 0 \forall i$ et x_i est le vecteur des caractéristiques du ménage. Si le véritable consentement à payer est nul ou négatif, la réponse sera $y_i = 0$, et $y_i = y_i^*$ dans le cas contraire. Notons que la règle ci-dessus peut provenir d'une décision du ménage de répondre 0 si $y_i^* \leq 0$, ou bien de l'enquêteur qui aurait eu comme instruction de reporter 0 en cas de réponse nulle ou « aberrante ».

Si le vecteur x_i est observé pour tout i , que y_i^* soit égal à c_i ou non, on a un échantillon censuré : seule la variable dépendante n'est observée que sur un certain intervalle (ici : $[0, +\infty]$).

Supposons à présent que les observations telles que $y_i^* \leq 0$ sont éliminées de l'échantillon, soit parce que les ménages avec $y_i^* \leq 0$ ont refusé de répondre à l'enquête dans son ensemble, soit parce que l'enquêteur les a éliminées. Si l'on n'observe ni y_i^* ni x_i dans ce cas, on parle alors d'un échantillon tronqué.

Exemple

La loi normale : on adopte une distribution normale pour y_i^* , de moyenne $x_i\beta$ et de variance σ^2 . Considérons tout d'abord le cas d'un échantillon censuré. La probabilité d'avoir $y_i = c_i$ est :

$$\begin{aligned} \text{Prob}(y_i = c_i) &= \text{Prob}(y_i^* \leq c_i) = \text{Prob}\left(\frac{y_i^* - x_i\beta}{\sigma} \leq \frac{c_i - x_i\beta}{\sigma}\right) \\ &= \Phi\left(\frac{c_i - x_i\beta}{\sigma}\right) \end{aligned}$$

Comme la variable y_i^* est inobservée dans ce cas, on modélise la probabilité de l'événement qu'elle appartienne à un certain intervalle par le biais de la fonction de répartition de la loi normale, $\Phi(\cdot)$. En revanche, dans le cas où $y_i > c_i$, y_i^* est observée et l'on sait qu'elle admet une distribution continue. La probabilité associée à y_i^* lorsque $y_i > c_i$ est alors modélisée avec la fonction de densité de la loi normale $\phi(\cdot)$:

$$\text{Prob}(y_i > c_i) = \text{Prob}(y_i = y_i^*) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right] = \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right)$$

Le terme $\frac{1}{\sigma}$ vient de la formule de la densité pour une fonction de variable : $f(y_i) = f(u_i) \times |du_i/dy_i|$ et dans notre cas $u_i = (y_i - x_i\beta)/\sigma$ (voir le chapitre 2).

On suppose à présent que l'échantillon est tronqué : on ne dispose plus que des observations pour lesquelles $y_i^* > c_i$. Dans ce cas, la probabilité d'observer y_i^* est égale à la densité associée à $y_i = y_i^*$, conditionnellement au fait que

$y_i^* > c_i$. On a l'expression suivante pour la densité conditionnelle :

$$f(y_i^* | y_i^* > c_i) = \frac{1}{\sigma} \phi\left(\frac{y_i^* - x_i \beta}{\sigma}\right) / \left[1 - \Phi\left(\frac{c_i - x_i \beta}{\sigma}\right)\right]$$

Le terme $[1 - \Phi]$ est la constante de normalisation assurant que :

$$\int_{c_i}^{+\infty} f(y_i^* | y_i^* > c_i) = 1$$

En pratique, les échantillons disponibles peuvent être à la fois tronqués et censurés, doublement tronqués, ou doublement censurés. Considérons un échantillon tronqué au niveau c_{1i} et censuré au niveau c_{2i} , avec $c_{2i} < c_{1i}$. L'échantillon ne contient que les observations pour lesquelles $y_i^* \leq c_{1i}$ et parmi celles-ci $y_i = y_i^*$ seulement si $y_i^* > c_{2i}$. Pour les observations telles que $y_i^* \leq c_{2i}$ on enregistre $y_i = c_{2i}$. La probabilité associée au cas $y_i^* > c_{2i}$ est :

$$\Phi\left(\frac{c_{1i} - x_i \beta}{\sigma}\right)^{-1} \times \frac{1}{\sigma} \phi\left(\frac{y_i - x_i \beta}{\sigma}\right)$$

et celle associée à l'événement $y_i^* \leq c_{2i}$ est :

$$\Phi\left(\frac{c_{1i} - x_i \beta}{\sigma}\right)^{-1} \times \Phi\left(\frac{c_{2i} - x_i \beta}{\sigma}\right)$$

Dans les deux cas, le premier terme prend en compte la troncature de l'échantillon au niveau c_{1i} .

Pour calculer l'espérance d'une variable dépendante normale censurée, on utilise le résultat suivant :

Résultat

Espérance d'une variable normale censurée : soit $y^* \sim N(\mu, \sigma^2)$ et $y = c$ si $y^* \leq c$, $y = y^*$ sinon. L'espérance de y est alors :

$$E(y) = \Phi\left(\frac{c - \mu}{\sigma}\right) \times c + \left[1 - \Phi\left(\frac{c - \mu}{\sigma}\right)\right] \times \left[\mu + \sigma \frac{\phi[(c - \mu)/\sigma]}{1 - \Phi[(c - \mu)/\sigma]}\right]$$

On a en effet :

$$E(y) = Prob(y = c) \times E(y|y = c) + Prob(y > c) \times E(y|y > c)$$

avec :

$$Prob(y = c) = Prob(y^* \leq c) = \Phi\left(\frac{c - \mu}{\sigma}\right), E(y|y = c) = c$$

$$Prob(y > c) = 1 - \Phi\left(\frac{c - \mu}{\sigma}\right)$$

et :

$$E(y|y > c) = E(y^*|y^* > c) = \mu + \sigma \frac{\phi[(c - \mu)/\sigma]}{1 - \Phi[(c - \mu)/\sigma]}$$

On donne maintenant l'expression de l'espérance conditionnelle d'une variable normale tronquée :

Résultat

Espérance conditionnelle d'une variable normale tronquée : soit $y \sim N(\mu, \sigma^2)$ et y n'est observée que si $y > c$. L'espérance conditionnelle de y est :

$$E(y|y > c) = \mu + \sigma \frac{\phi[(c - \mu)/\sigma]}{1 - \Phi[(c - \mu)/\sigma]}$$

Si y n'est observée que si $y \leq c$, l'espérance conditionnelle est :

$$E(y|y \leq c) = \mu - \sigma \frac{\phi[(c - \mu)/\sigma]}{\Phi[(c - \mu)/\sigma]}$$

L'espérance non conditionnelle de y est alors :

$$E(y) = E(y|y > c) \times Prob(y > c) + E(y|y \leq c) \times Prob(y \leq c)$$

et dans le cas particulier où $c = 0$, on a bien :

$$\begin{aligned} E(y) &= \left[\mu + \sigma \frac{\phi(-\mu/\sigma)}{1 - \Phi(-\mu/\sigma)} \right] \\ &\quad \times [1 - \Phi(-\mu/\sigma)] + \left[\mu - \sigma \frac{\phi(-\mu/\sigma)}{\Phi(-\mu/\sigma)} \right] \times \Phi(-\mu/\sigma) \\ &= \mu \end{aligned}$$

Le terme ϕ/Φ est appelé dans la littérature « le rapport inverse de Mills » (*Inverse Mills Ratio*).

III. Le modèle Tobit simple

Nous présentons dans cette section le modèle Tobit, d'après Tobin (1958), dans sa version la plus simple. Dans les modèles à variable qualitative binaire du chapitre 3, nous avions défini une variable latente y_i^* dont le signe déterminait la valeur de la variable dépendante qualitative y_i . Le modèle Tobit simple fait la même hypothèse sur la spécification de la variable latente et suppose, de

plus, que ε_i est normalement distribuée : $\varepsilon_i \sim N(0, \sigma^2)$. Une hypothèse supplémentaire est que, par rapport aux modèles censurés plus haut, la borne c_i est identique pour tous les individus. Par conséquent, si le vecteur x_i contient un terme constant, celui-ci est indissociable de la borne. Le modèle est :

$$\begin{cases} y_i^* = x_i\beta + \varepsilon_i \\ y_i = y_i^* \text{ si } y_i^* > 0 \\ \quad 0 \text{ sinon} \end{cases}$$

Posons N_0 le nombre d'observations pour lesquelles $y_i = 0$, N_1 le nombre d'observations pour lesquelles $y_i > 0$. Par la suite, on utilisera les symboles \sum_0 et \sum_1 pour indiquer une sommation sur les observations telles que $y_i = 0$ et $y_i > 0$ respectivement.

Pour les N_0 observations telles que $y_i = 0$, on a :

$$Pr(y_i = 0) = Pr(\varepsilon_i < -x_i\beta) = 1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)$$

et pour les N_1 observations avec $y_i > 0$:

$$Pr(y_i > 0) \times f(y_i | y_i > 0) = \Phi\left(\frac{x_i\beta}{\sigma}\right) \frac{f(\varepsilon_i)}{\Phi(x_i\beta/\sigma)} = f(\varepsilon_i) = \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right)$$

La log-vraisemblance de l'échantillon s'écrit en combinant les deux probabilités :

$$\log L = \sum_0 \log \left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right) \right] + \sum_1 \log \left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \sum_1 \frac{(y_i - x_i\beta)^2}{2\sigma^2}$$

A. Comparaison avec les modèles tronqués

La log-vraisemblance d'un modèle tronqué en c_i s'écrit :

$$\begin{aligned} \log L = & -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - x_i\beta)^2 \\ & - \sum_{i=1}^N \log \left[1 - \Phi\left(\frac{c_i - x_i\beta}{\sigma}\right) \right] \end{aligned}$$

et l'on voit que son expression est très proche de celle du modèle Tobit, avec la borne c fixée à 0. La seule différence réside dans la façon dont la probabilité $1 - \Phi[(c_i - x_i\beta)/\sigma]$ est prise en compte. Dans le modèle tronqué, cette

probabilité est associée à toutes les observations, alors que pour le Tobit, elle ne concerne que les N_0 observations pour lesquelles $y_i = 0$. D'après le résultat plus haut sur l'espérance d'une variable censurée, l'espérance de la variable censurée y_i lorsque $c = 0$ est égale à $[x_i\beta + \sigma\phi(-x_i\beta/\sigma)]/[1 - \Phi(x_i\beta/\sigma)]$ multiplié par la probabilité que $y > 0$, c'est-à-dire $[1 - \Phi(x_i\beta/\sigma)]$. Par conséquent, l'espérance conditionnelle de y_i vaut :

$$E(y_i|x_i, y_i > 0) = \frac{E(y_i|x_i)}{Prob(y_i > 0)} = x_i\beta + \sigma \frac{\phi(-x_i\beta/\sigma)}{1 - \Phi(-x_i\beta/\sigma)}$$

et l'on retrouve bien l'espérance conditionnelle d'une variable tronquée. Notons enfin qu'en vertu de la symétrie de la densité normale, cette espérance conditionnelle peut aussi s'écrire plus simplement :

$$E(y_i|x_i, y_i > 0) = x_i\beta + \sigma \frac{\phi(x_i\beta/\sigma)}{\Phi(x_i\beta/\sigma)}$$

B. Estimation par maximum de vraisemblance

Notons pour simplifier $f_i = (1/\sigma)\phi(y_i - x_i\beta)$, la fonction de densité associée à ε_i et $\Phi_i = \Phi(-x_i\beta/\sigma)$.

Les conditions du premier ordre associées à la maximisation de la log-vraisemblance par rapport aux paramètres β et σ^2 sont :

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= - \sum_0 \frac{f_i x'_i}{1 - F_i} + \frac{1}{\sigma^2} \sum_1 (y_i - x_i\beta)x'_i = 0 \\ \frac{\partial \log L}{\partial \sigma^2} &= \frac{1}{2\sigma^2} \sum_0 \frac{x_i\beta f_i}{1 - F_i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_1 (y_i - x_i\beta)^2 = 0 \end{aligned}$$

En transposant $\partial \log L / \partial \beta$, en le multipliant par $\beta / 2\sigma^2$, puis en l'ajoutant à $\partial \log L / \partial \sigma^2$, on a :

$$\begin{aligned} \frac{1}{2\sigma^4} \sum_1 (y_i - x_i\beta)x_i\beta - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_1 (y_i - x_i\beta)^2 &= 0 \\ \Leftrightarrow \frac{1}{\sigma^2} \sum_1 \varepsilon_i x_i\beta + \frac{1}{\sigma^2} \sum_1 \varepsilon_i^2 &= N_1 \\ \Leftrightarrow \frac{1}{\sigma^2} \sum_1 \varepsilon_i (x_i\beta + \varepsilon_i) &= N_1 \Leftrightarrow \sigma^2 = \frac{1}{N_1} \sum_1 (y_i - x_i\beta)y_i \end{aligned}$$

L'estimateur par maximum de vraisemblance de la variance des résidus est donc la variance empirique des termes d'erreurs sur la partie de l'échantillon

contenant les y_i strictement positifs. La dérivée $\partial \log L / \partial \beta$ est, en multipliant par σ :

$$\begin{aligned} & -\sum_0 \frac{\sigma f_i x'_i}{1 - F_i} + \frac{1}{\sigma} \sum_1 (y_i - x_i \beta) x'_i = 0 \\ \Leftrightarrow & -\sum_0 \frac{x'_i \phi_i}{1 - \Phi_i} + \frac{1}{\sigma} \sum_1 (y_i - x_i \beta) x'_i = 0 \end{aligned}$$

et en utilisant une écriture matricielle plus compacte :

$$\frac{1}{\sigma} X'_1 Y_1 - \frac{1}{\sigma} X'_1 X_1 \beta = X'_0 \gamma_0 \Leftrightarrow \frac{1}{\sigma} (X'_1 X_1)^{-1} X'_1 Y_1 - \frac{1}{\sigma} \beta = (X'_1 X_1)^{-1} X'_0 \gamma_0$$

où :

$$\gamma_0 = \frac{\phi_i}{1 - \Phi_i} = \frac{\sigma f_i}{1 - F_i}$$

L'expression de l'estimateur par maximum de vraisemblance de β est finalement (Olsen, 1980) :

$$\boxed{\beta = (X'_1 X_1)^{-1} X'_1 Y_1 - \sigma (X'_1 X_1)^{-1} X'_0 \gamma_0}$$

soit $\beta = \beta_{MCO} - \sigma (X'_1 X_1)^{-1} X'_0 \gamma$. Cette formule illustre le biais des moindres carrés ordinaires dans le contexte du modèle à variable dépendante censurée (Heckman, 1979). En pratique, les estimateurs par maximum de vraisemblance sont souvent supérieurs en valeur absolue à ceux des MCO (voir Goldberger, 1981 ; Cheung et Goldberger, 1984). De plus, on constate que les estimateurs MLE sont très proches des estimateurs MCO divisés par la proportion N_1/N .

On peut montrer que le modèle Tobit peut être reparamétrisé en posant $\alpha = \beta/\sigma$ et $h = 1/\sigma$, et que cette nouvelle spécification conduit à une log-vraisemblance globalement concave en α et h (Olsen, 1978). Une conséquence intéressante est alors qu'un algorithme itératif de maximisation de $\log L$ permettra toujours de converger vers le maximum global. La log-vraisemblance du modèle d'Olsen s'écrit :

$$\log L = \sum_0 \log [1 - \Phi(x_i \alpha)] + N_1 \log h - \frac{1}{2} \sum_1 (hy_i - x_i \alpha)^2$$

et le Hessien est :

$$\left[\begin{array}{cc} \sum_0 [\phi_i / (1 - \Phi_i)] (x_i \alpha - \phi_i / (1 - \Phi_i)) x'_i x_i & 0 \\ 0 & -N_1 / h^2 \end{array} \right]$$

où maintenant $\phi_i = \phi(x_i \alpha)$ et $\Phi_i = \Phi(x_i \alpha)$. Comme $x'_i x_i$ est une matrice définie positive et que le terme $(x_i \alpha - \phi_i / (1 - \Phi_i))$ est toujours négatif, le Hessien est une matrice définie négative, d'où le résultat.

C. Estimation en deux étapes

Les algorithmes actuels d'optimisation ont accompli de tels progrès en vitesse et précision que l'estimation par maximum de vraisemblance semble la plus adaptée. Puisque l'estimateur MLE est efficace et que la log-vraisemblance du modèle reparamétrisé admet un maximum global, il semble inutile de rechercher des méthodes alternatives d'estimation. Nous présentons toutefois la procédure d'Heckman (1976), dans la mesure où cet estimateur est encore très utilisé et que la méthode proposée illustre bien l'utilisation des techniques pour variables qualitatives binaires vues au chapitre 3. De plus, la procédure en deux étapes d'Heckman sera utilisée dans les sections suivantes pour estimer les modèles de sélection.

L'espérance conditionnelle de y_i , sachant que $y_i > 0$, est :

$$E(y_i|x_i, y_i > 0) = x_i\beta + E(\varepsilon_i|\varepsilon_i > -x_i\beta) = x_i\beta + \sigma \frac{\phi(x_i\beta/\sigma)}{\Phi(x_i\beta/\sigma)}$$

que l'on peut séparer en une fonction linéaire $x_i\beta$ et une fonction non linéaire $\sigma\phi_i/\Phi_i$ des variables x_i . Comme dans le cas d'un modèle censuré les observations sont disponibles sur les individus dont la variable dépendante est nulle, il est possible de calculer la probabilité de censure séparément de β . Définissons la variable indicatrice suivante : $I_i = 1$ si $y_i > 0$, $I_i = 0$ sinon. La log-vraisemblance associée à ce modèle est :

$$\log L = \sum_{i=1}^N \left\{ I_i \log \left[\Phi \left(\frac{x_i\beta}{\sigma} \right) \right] + (1 - I_i) \log \left[1 - \Phi \left(\frac{x_i\beta}{\sigma} \right) \right] \right\}$$

qui correspond au modèle Probit simple du chapitre 3. En imposant $\sigma = 1$ les estimateurs du maximum de vraisemblance de α sont convergents, où $\alpha = \beta/\sigma$. Ces derniers permettent d'estimer $\hat{\phi}_i = \phi(x_i\hat{\alpha})$ et $\hat{\Phi}_i = \Phi(x_i\hat{\alpha})$ et fournissent donc un estimateur de première étape convergent du rapport inverse de Mills. Ce dernier est remplacé dans l'équation du Tobit et les paramètres β et σ sont estimés par MCO sur le sous-échantillon tel que $y_i > 0$, σ étant le coefficient associé à $\hat{\phi}_i/\hat{\Phi}_i$:

$$y_i = E(y_i|x_i, y_i > 0) + v_i = x_i\beta + \sigma \frac{\hat{\phi}_i}{\hat{\Phi}_i} + v_i$$

Le résidu v_i est de moyenne nulle, mais sa variance n'est pas constante. On montre que :

$$Var(v_i) = \sigma^2 - \sigma^2 \frac{x_i\beta}{\sigma} \frac{\phi(x_i\beta/\sigma)}{\Phi(x_i\beta/\sigma)} - \sigma^2 \frac{\phi^2(x_i\beta/\sigma)}{\Phi^2(x_i\beta/\sigma)}$$

Une méthode de correction comme celle des moindres carrés pondérés est applicable (voir le chapitre 2), mais elle doit procéder en deux étapes.

En effet, la variance de y_i dépend des paramètres β qui doivent être préalablement estimés. La première étape fournira les estimations de β/σ et ϕ_i/Φ_i , puis l'expression de la variance sera calculée. Les observations seront alors pondérées par $[Var(y_i)]^{1/2}$ dans la seconde étape. Notons qu'il est aussi possible d'estimer le modèle avec toutes les observations de y_i en utilisant le fait que :

$$\begin{aligned} E(y_i|x_i) &= Pr(y_i > 0) \times E(y_i|x_i, y_i > 0) + Pr(y_i = 0) \times E(y_i|x_i, y_i = 0) \\ &= \Phi_i(x_i\beta + \sigma\frac{\phi_i}{\Phi_i}) + (1 - \Phi_i) \times 0 = \Phi_i x_i\beta + \sigma\phi_i \end{aligned}$$

L'équation de régression linéaire est alors :

$$y_i = E(y_i|x_i) + u_i = \hat{\Phi}_i x_i\beta + \sigma\hat{\phi}_i + u_i$$

Il est difficile de conclure à la supériorité de cette procédure, avec les données complètes, sur la première possibilité, avec les données censurées. La différence entre les deux dépend des observations et des paramètres, et n'est ni toujours définie positive ni toujours définie négative.

D. Effets marginaux avec le modèle Tobit

Les effets marginaux dans un modèle de régression censuré correspondent à des prévisions sur une variable continue, lorsque les variables explicatives x_i sont modifiées. Il y a plusieurs prédictions possibles pour le Tobit, selon que l'on s'intéresse à la distribution de la variable censurée ou à celle de la variable latente. Les différentes espérances que l'on peut facilement calculer sont :

- variable latente : $E(y_i^*|x_i)$;
- variable dépendante sans censure : $E(y_i|x_i)$;
- variable dépendante avec censure : $E(y_i|x_i, y_i > 0)$.

On a vu plus haut que :

$$\left\{ \begin{array}{l} E(y_i|x_i) = \Phi_i x_i\beta + \sigma\phi_i \\ E(y_i|x_i, y_i > 0) = E(y_i^*|y_i^* > 0) = E(y_i|y_i^* > 0) = x_i\beta + \sigma\phi_i/\Phi_i \\ E(y_i^*|x_i) = x_i\beta \end{array} \right.$$

L'effet marginal correspondant à la sensibilité de la variable latente y_i^* par rapport à une variable x_{ik} , $k = 1, 2, \dots, K$ est simplement $\partial E(y_i^*|x_i)/\partial x_{ik} = \beta_k$.

Pour la variable dépendante non censurée, on montre que :

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = \Phi_i(x_i\beta/\sigma)\beta_k$$

Enfin, l'effet marginal correspondant à la variable dépendante censurée est donné par :

$$\frac{\partial E(y_i|x_i, y_i^* > 0)}{\partial x_{ik}} = \beta_k \left[1 - \frac{x_i \beta}{\sigma} \frac{\phi_i}{\Phi_i} - \left(\frac{\phi_i}{\Phi_i} \right)^2 \right]$$

IV. Tests de spécification

Comme on l'a vu au chapitre 2, les estimateurs du maximum de vraisemblance sont convergents et efficaces si les hypothèses distributionnelles du modèle sont vérifiées par les données. Cela implique en particulier qu'une inférence exacte à partir du modèle ne sera possible que si *tous* les moments de la distribution des erreurs sont correctement spécifiés. Dans le cas du modèle Tobit, il existe différents tests de spécification relatifs à l'hypothèse d'homoscédasticité, de normalité des résidus et d'égalité des paramètres entre l'équation de censure et celle de l'espérance conditionnelle de la variable dépendante.

A. Hétéroscédasticité

La façon la plus simple de vérifier l'hypothèse d'homoscédasticité dans le modèle Tobit est de spécifier la variance du résidu en fonction de variables supplémentaires z_i comme suit :

$$Var(\varepsilon_i) = \sigma_i^2 = \sigma_0^2 \times \exp(z_i \gamma)$$

où z_i est un vecteur de dimension $(1 \times K_2)$ (x_i est de dimension $1 \times K$). L'hypothèse nulle d'homoscédasticité est $H_0 : \gamma = 0$ que l'on peut tester à l'aide de la statistique du multiplicateur de Lagrange. Pour cela, il faut évaluer le gradient de la log-vraisemblance sous H_0 :

$$\begin{cases} \partial \log L / \partial \beta = \sum_{i=1}^N x_i^T [z_i \varepsilon_i / \sigma^2 - (1 - z_i) \lambda_i / \sigma] \\ \partial \log L / \partial \sigma = \sum_{i=1}^N z_i (\varepsilon_i^2 / \sigma^2 - 1) / 2\sigma^2 + (1 - z_i) \lambda_i x_i \beta / 2\sigma^3 \\ \partial \log L / \partial \gamma = \sum_{i=1}^N \sigma^2 z_i^T [z_i (\varepsilon_i^2 / \sigma^2 - 1) / 2\sigma^2 + (1 - z_i) \lambda_i x_i \beta / 2\sigma^3] \end{cases}$$

où :

$$z_i = 1 \quad \text{si } y_i > 0, \quad 0 \text{ sinon,} \quad \text{et } \lambda_i = \frac{\phi(-x_i \beta / \sigma)}{\Phi(-x_i \beta / \sigma)}$$

Soit $G(\beta, \sigma, \gamma)$ la matrice $(N \times (K + K_2 + 1))$ formée à partir des dérivées ci-dessus (en omettant l'opérateur somme). Cette matrice est évaluée sous H_0 .

La statistique de test LM est alors donnée par $LM = e'_N G[G'G]^{-1}G'e_N$, où e_N est un vecteur ($N \times 1$) de 1 (voir le chapitre 2). Cette statistique vaut NR^2 , où R^2 est le coefficient de détermination de la régression d'une colonne de 1 sur la matrice G .

B. Normalité

La difficulté réside ici dans le fait qu'une distribution concurrente de la loi normale conduit à des procédures d'estimation souvent très compliquées (voir Kalbfleisch et Prentice, 1980; Duncan, 1983; Fernandez, 1986). Il est par contre possible de tester simplement l'hypothèse de normalité dans le cas du Tobit, en employant la procédure de test d'Hausman. Si l'hypothèse nulle de normalité est valide, les estimateurs du maximum de vraisemblance seront convergents et efficaces. Ils peuvent donc être comparés à un estimateur convergent sous l'hypothèse alternative, mais non efficace sous l'hypothèse nulle. Un tel estimateur est, par exemple, celui proposé par Powell (1984) : l'estimateur LAD (*Least Absolute Deviations*, moindres valeurs absolues). Cet estimateur de la médiane est robuste, convergent quelle que soit la distribution des erreurs (Melenberg et Van Soest, 1996). Une autre possibilité est de considérer une distribution normale, dont l'argument n'est plus seulement la valeur du seuil de censure, mais un polynôme de cette valeur. La procédure de Ruud (1984) revient à utiliser :

$$Prob\left(\varepsilon_i \leq \frac{x_i\beta}{\sigma}\right) = \Phi\left(\sum_{l=0}^L \mu_l \left(\frac{x_i\beta}{\sigma}\right)^l\right)$$

avec $\mu_1 = 1$. L'hypothèse nulle de normalité s'obtient avec $\mu_l = 0, \forall l, l \neq 1$. Si le modèle contient un terme constant, μ_0 peut de plus être normalisé à 0.

Une procédure de test originale repose enfin sur l'idée que, si les erreurs sont normalement distribuées, on doit satisfaire certaines conditions de moments (Pagan et Vella, 1989). Par exemple, les moments impairs sont tous nuls, et le moment d'ordre 4 vaut $(3\sigma^4)^2$. Pour les observations où y_i est observé ($y_i > 0$), les conditions de moment s'écrivent :

$$\begin{cases} E(m_{1i}) = E\{w_i(y_i - x_i\beta)\} = 0 \\ E(m_{2i}) = E\{w_i[(y_i - x_i\beta)^2 - \sigma^2]\} = 0 \\ E(m_{3i}) = E\{w_i[(y_i - x_i\beta)^3]\} = 0 \\ E(m_{4i}) = E\{w_i[(y_i - x_i\beta)^4 - 3\sigma^4]\} = 0 \end{cases}$$

où w_i est un vecteur ($1 \times L$) d'instruments (voir le chapitre 2 sur la méthode des moments généralisés).

2. Ce moment capte l'aplatissement de la distribution (*kurtosis*). Une mesure non paramétrique usuelle d'aplatissement est la différence entre le moment centré d'ordre 4 et 3 fois la variance au carré.

Pour les observations censurées, les conditions de moment sont plus complexes car y_i est remplacé par l'espérance conditionnelle de y_i , sachant que $y_i = 0$. On a :

$$\begin{cases} E(m_{1i}) = E\{w_i(Ey_i - x_i\beta | y_i = 0)\} = -w_i\sigma\lambda_i = 0 \\ E(m_{2i}) = E\{w_i[(Ey_i - x_i\beta)^2 - \sigma^2 | y_i = 0]\} = \sigma w_i\lambda_i x_i\beta = 0 \\ E(m_{3i}) = E\{w_i[(Ey_i - x_i\beta)^3 | y_i = 0]\} = -\sigma^3[2 + (x_i\beta/\sigma)^2]\lambda_i = 0 \\ E(m_{4i}) = E\{w_i[(Ey_i - x_i\beta)^4 - 3\sigma^4 | y_i = 0]\} = \sigma^4[3 + (x_i\beta/\sigma)^2](x_i\beta/\sigma)\lambda_i = 0 \end{cases}$$

Ces conditions de moment sont alors utilisées pour calculer un test de spécification à la Hansen.

C. Égalité des paramètres

On peut soulever l'objection que, dans le modèle Tobit simple, les paramètres dans la probabilité de censure et ceux de l'espérance conditionnelle $E(y_i|x_i)$ sont les mêmes. Par conséquent, $\Phi(\cdot)$ étant croissante dans ses arguments, une variable jouant à la hausse sur la probabilité de non-censure augmente parallèlement l'espérance conditionnelle (Fin et Schmidt, 1984). Cet effet pouvant être indésirable dans certaines modélisations, on peut généraliser le modèle Tobit en considérant des paramètres différents, β_1 dans l'espérance conditionnelle et β_2 dans la probabilité que $y_i > 0$. Il est bien sûr aisément de tester la spécification du modèle Tobit simple en prenant comme hypothèse nulle $H_0 : \beta_1 = \beta_2$. Le test peut par exemple reposer sur la statistique du rapport des vraisemblances. Notons qu'une procédure d'estimation en deux étapes est toujours envisageable dans le cas général, car la première étape (Probit) estime les probabilités de censure en fonction des variables x_i et ne contraint nullement les paramètres à être identiques.

V. Application : dépenses de restaurant

Nous appliquons la procédure Tobit au problème d'estimation de la demande pour les repas pris au restaurant. *A priori*, cette demande devrait dépendre positivement de variables sociodémographiques comme le revenu, le nombre de voitures du ménage, une distance limitée par rapport au centre-ville, et négativement de variables telles la taille de la famille, l'âge de l'enfant le plus jeune. Un échantillon de 5 360 ménages est constitué, dans lequel on observe les variables suivantes :

- *RESTO* : dépenses annuelles en restaurant (en 100 US \$) ;
- *REVENU* : revenu annuel du foyer (en 100 US \$) ;
- *AGJUNIOR* : âge de l'enfant le plus jeune ;
- *NBVOIT* : nombre de voitures du foyer ;
- *TAILLE* : nombre de personnes du foyer ;
- *VILLE* : indicatrice, 1 si le ménage réside dans une petite agglomération, 0 sinon ;
- *DISTANCE* : distance (en km) par rapport au centre-ville.

Dans l'échantillon, 4 373 ménages ont des dépenses positives pour le budget « restaurant » et 987 des dépenses nulles. Comme on observe les variables explicatives dans tous les cas, on a bien un échantillon censuré. Les statistiques descriptives pour l'échantillon figurent dans le tableau suivant.

Variable	Moyenne	Écart type
<i>RESTO</i>	6.2288	7.4529
<i>REVENU</i>	170,7524	180,0022
<i>AGJUNIOR</i>	3,1744	4,8059
<i>NBVOIT</i>	1,4337	1,0528
<i>TAILLE</i>	2,8798	1,5399
<i>VILLE</i>	0,5080	0,4999
<i>DISTANCE</i>	7,5059	11,5331

Le modèle Tobit est estimé avec le logiciel SAS, par la procédure LIFEREG. Les commandes sont les suivantes :

```
data toto; set resto;
revenu=revenu/100;
resto=resto/100;
if resto=0 then lower=.;
else lower=resto;
proc lifereg data=toto;
model (lower,resto) = revenu agjunior nbvoit taille ville distance /d=normal;
run;
```

Les instructions `if resto=0 then lower=.; else lower=resto;` servent à spécifier la valeur de la borne de censure (0 ici). L'option `d=normal` indique que l'on spécifie la distribution normale pour les erreurs. Les résultats d'estimation de la procédure LIFEREG sont les suivants.

The SAS System

14:27 Friday, August 20, 1999

Lifereg Procedure

Data Set =WORK.TOTO

Dependent Variable=LOWER

Dependent Variable=RESTO

Noncensored Values= 4373 [1] Right Censored Values= 0
 Left Censored Values= 987 [2] Interval Censored Values= 0

Log Likelihood for NORMAL -36157.76358 [3]

Lifereg Procedure

Variable	DF	Estimate	Std. Err	ChiSquare	Pr>Chi	Label/Value
[4]	[5]	[6]	[7]	[8]	[9]	[10]
INTERCPT	1	2.934824	0.278574	110.9895	0.0001	Intercept
REVENU	1	0.01446748	0.000689	440.5235	0.0001	
AGJUNIOR	1	-0.0546407	0.025928	4.441113	0.0351	
NBVOIT	1	1.88239439	0.121764	238.9938	0.0001	
TAILLE	1	-0.8058984	0.084618	90.70529	0.0001	
VILLE	1	-1.2557751	0.226559	30.7228	0.0001	
DISTANCE	1	0.03001962	0.009943	9.115528	0.0025	
SCALE [11]	1	7.86247619	0.086177			

La sortie du logiciel indique le nombre d'observations non-censurées [1] et censurées [2], et fournit la valeur de la log-vraisemblance du modèle [3]. Le tableau des paramètres estimés donne successivement le nom de la variable [4], le nombre de degrés de liberté [5] (1 ici), le coefficient estimé [6], son écart type [7], la statistique de test du χ^2 associée à la nullité de ce paramètre [8], la p-value du test [9], et un commentaire éventuel sur la variable [10].

Le dernier paramètre [11] (SCALE) est l'écart type estimé de la variable dépendante limitée. La valeur calculée de cet écart type sur le sous-ensemble des observations telles que $RESTO > 0$ est 7.5778. On constate que la dépense en restaurant est bien croissante avec le revenu et le nombre de voitures. Elle est décroissante avec l'âge du plus jeune enfant, la taille du ménage et le fait de résider dans une petite commune. Un résultat moins intuitif est que la distance au centre-ville joue de façon positive sur cette dépense. Il est instructif de comparer les résultats du Tobit avec ceux des moindres carrés ordinaires. Les résultats de l'estimation par MCO sont les suivants, avec la procédure REG de SAS. Le logiciel rappelle le nom de la variable dépendante [12] et présente une analyse de la variance du modèle [13]. Celle-ci décompose la somme des carrés totale (*C Total*) en deux : celle issue du modèle (*Model*) et celle des erreurs (*Error*). Les deux colonnes [14] et [15] indiquent respectivement la somme des carrés et leur moyenne, et indique la valeur du test de significativité globale des paramètres de Fisher [16] ainsi que la p-value associée [17]. Le tableau des paramètres estimés fournit tout d'abord quelques statistiques d'ajustement : erreur quadratique moyenne (*Root Mean Squared Error*) [18], coefficient R^2 [19] et R^2 ajusté [20], moyenne de la variable dépendante [21] et coefficient de variation [22]. Le logiciel indique le nom de la variable explicative [23], le nombre de degrés de liberté (1 ici) [24], le coefficient estimé [25], son écart type [26], le *t* de Student [27] et la p-value associée au test [28].

Dependent Variable : RESTO

Analysis of Variance

Source	DF	[14] Sum of		[15] Mean	F Value [16]	Prob>F [17]
		Squares	Square			
Model	6	47936.64130	7989.44022		171.252	0.0001
Error	5353	249733.55238	46.65301			
C Total	5359	297670.19367				

Root MSE [18]	6.83030	R-square [19]	0.1610
Dep Mean [20]	6.22882	Adj R-sq [21]	0.1601
C.V. [22]	109.65637		

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0 :	Prob > T
		Estimate	Error	Parameter=0	
INTERCEP	1	4.619269	0.23131943	19.969	0.0001
REVENU	1	0.012142	0.00059248	20.493	0.0001
AGJUNIOR	1	-0.035560	0.02173968	-1.636	0.1020
NBVOIT	1	1.289087	0.10259768	12.564	0.0001
TAILLE	1	-0.600173	0.06985363	-8.592	0.0001
VILLE	1	-1.127687	0.19101604	-5.904	0.0001
DISTANCE	1	0.013618	0.00852712	1.597	0.1103

VI. Extensions du modèle Tobit

On présente dans cette section des extensions du modèle Tobit simple, qui peuvent être classées en trois catégories. La première extension consiste à spécifier un modèle Tobit avec plus d'un seuil de censure. La deuxième catégorie, les Tobit généralisés, contient des modèles censurés ou tronqués dans lesquels le fait qu'une première variable observée soit nulle ou égale à l'expression de la variable latente associée dépend du signe d'une seconde variable latente. La troisième catégorie concerne des modèles censurés ou tronqués caractérisés par une borne de censure ou de troncature aléatoire. Une variable sera par exemple observable ou non nulle si la variable latente associée est supérieure à une seconde variable latente.

A. Modèles Tobit à censure multiple

Dans certaines applications, la variable dépendante est censurée à la fois par le haut et par le bas. Par exemple, sur un marché de cotation, il peut exister des limites inférieure et supérieure aux cours, au-delà desquelles la valeur du cours est soit égale au seuil plancher, soit plafonnée. Un autre exemple est le cas d'agents économiques ne réalisant des transactions que si la variation d'une valeur fondamentale est suffisante. Si les coûts de transaction sont trop importants, un investisseur sera par exemple amené à modifier sa position (acheter ou vendre des titres) seulement lorsque la variation du rendement du titre dépasse un certain niveau en valeur absolue.

Le modèle général avec double censure est le suivant :

$$\begin{cases} y_i^* &= x_i\beta + \varepsilon_i \\ y_i &= c_{1i} \quad \text{si} \quad y_i^* \leq c_{1i} \\ & y_i^* \quad \text{si} \quad c_{1i} < y_i^* \leq c_{2i} \\ & c_{2i} \quad \text{si} \quad y_i^* > c_{2i} \end{cases}$$

où c_{1i} et c_{2i} , $c_{1i} < c_{2i}$, sont les bornes de censure inférieure et supérieure. Elles peuvent correspondre à des limites fixées par un régulateur de marché, des contraintes d'approvisionnement, etc. La variable y_i^* peut alors représenter le niveau de prix virtuel lié aux fondamentaux, le niveau de transaction désiré, etc. On suppose que $\varepsilon \sim N(0, \sigma^2)$.

Les probabilités associées aux événements $y_i = c_{1i}$ et $y_i = c_{2i}$, respectivement, sont $\Phi[(c_{1i} - x_i\beta)/\sigma]$ et $1 - \Phi[(c_{2i} - x_i\beta)/\sigma]$. Comme dans le modèle Tobit simple, la probabilité d'une valeur de $y_i = y_i^*$ est simplement la densité $\phi(\varepsilon/\sigma)/\sigma$. La vraisemblance de ce modèle s'écrit :

$$L(\beta, \sigma^2 | y_i, x_i, c_{1i}, c_{2i}) = \prod_{y_i=c_{1i}} \Phi\left(\frac{c_{1i} - x_i\beta}{\sigma}\right) \prod_{y_i=c_{2i}} \left[1 - \Phi\left(\frac{c_{2i} - x_i\beta}{\sigma}\right)\right] \prod_{y_i=y_i^*} \frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right)$$

En notant :

$$\Phi_{1i} = \Phi\left(\frac{c_{1i} - x_i\beta}{\sigma}\right) \quad \Phi_{2i} = \Phi\left(\frac{c_{2i} - x_i\beta}{\sigma}\right)$$

L'espérance conditionnelle de la variable dépendante limitée est

$$\begin{aligned} E(y_i | x_i, c_{1i} < y_i^* \leq c_{2i}) &= x_i\beta + E(\varepsilon_i | c_{1i} - x_i\beta < \varepsilon_i \leq c_{2i} - x_i\beta) \\ &= x_i\beta + \sigma \frac{\phi_{1i} - \phi_{2i}}{\Phi_{2i} - \Phi_{1i}} \end{aligned}$$

L'espérance de la variable non censurée est :

$$\begin{aligned} E(y_i | x_i) &= Pr(y_i = c_{1i}) \times c_{1i} + Pr(y_i = c_{2i}) \times c_{2i} + Pr(c_{1i} < y_i^* \leq c_{2i}) \\ &\quad \times E(y_i | x_i, c_{1i} < y_i^* \leq c_{2i}) \\ &= \Phi_{1i}c_{1i} + (1 - \Phi_{2i})c_{2i} + x_i\beta(\Phi_{2i} - \Phi_{1i}) + \sigma(\phi_{1i} - \phi_{2i}) \end{aligned}$$

B. Exemple : les modèles à friction

Ces modèles sont utilisés dans des applications où la variable dépendante ne répond qu'à des fortes variations (ou des fortes valeurs) des variables explicatives (marchés, demandes d'actifs, distribution de dividendes, etc.).

1) Distribution de dividendes

Prenons l'exemple d'une société par actions qui délivre des dividendes à ses actionnaires. Soit $y_t^*, t = 1, 2, \dots, T$ le paiement de dividendes désiré à la date t , qui dépend de variables comme les bénéfices, la conjoncture, les décisions d'autofinancement, etc. On pose $y_t = x_t\beta + \varepsilon_t$, où $\varepsilon \sim N(0, \sigma^2)$. Comme le mécanisme de versement des dividendes est coûteux, l'entreprise préfère limiter leur distribution dans le temps (au plus une fois par an), mais aussi selon leur montant. On supposera qu'elle n'effectue un versement de dividendes que si le niveau espéré y_t^* est supérieur³ à un seuil c_0 . De plus, on fait l'hypothèse que l'entreprise, afin de conserver une certaine marge de manœuvre financière, limite le montant versé à une valeur seuil notée c_1 . Le modèle est le suivant :

$$\begin{cases} y_t^* &= x_t\beta + \varepsilon_t \\ y_t &= 0 \quad \text{si} \quad y_t^* < c_0 \\ &\quad y_t^* \quad \text{si} \quad c_0 \leq y_t^* < c_1 \\ &\quad c_1 \quad \text{si} \quad c_1 \leq y_t^* \end{cases}$$

La vraisemblance de l'échantillon de taille T est :

$$L(\beta, \sigma^2 | x_i, c_0, c_1) = \prod_{y_t=0} \Phi\left(\frac{c_0 - x_t\beta}{\sigma}\right) \prod_{y_t=y_t^*} \phi\left(\frac{y_t - x_t\beta}{\sigma}\right) \prod_{y_t=c_1} \left[1 - \Phi\left(\frac{c_1 - x_t\beta}{\sigma}\right)\right]$$

2) Modèle de Rosett

On considère un modèle d'investissement dans des actifs financiers, où les coûts de transaction peuvent limiter les transactions par rapport au niveau désiré de l'investisseur. Le modèle suppose que les modifications dans la position de l'investisseur (achat ou vente) pour un titre dépendent des variations de son rendement. Ce dernier est $R_t = (D_t + P_t - P_{t-1})/P_{t-1}$, où P_t est le cours à la date t , D_t le dividende. Soit y_t^* la variation désirée de la position du titre, y_t la variation observée de la position, x_t la variation observée du rendement du titre. En raison des coûts de transaction, l'investisseur n'effectuera de transactions que si les variations de rendement sont assez importantes. La position

3. L'entreprise peut, par exemple, décider de regrouper les dividendes de faible valeur en un seul versement annuel ou tous les deux ans.

réelle ne change donc pas pour de petites variations à la hausse ou à la baisse du rendement, et donc de la position désirée. Le modèle est :

$$\begin{cases} y_t^* = x_t \beta + \varepsilon_t \\ y_t = y_t^* - \alpha_1 & \text{si } y_t^* < \alpha_1 \\ & 0 & \text{si } \alpha_1 \leq y_t^* < \alpha_2 \\ & y_t^* - \alpha_2 & \text{si } \alpha_2 \leq y_t^* \end{cases}$$

où $\alpha_1 < 0$ est l'augmentation désirée et $\alpha_2 > 0$ la hausse désirée, qui sont des constantes indépendantes de la période. Si $\beta > 0$, le modèle indique qu'une variation négative du rendement R_t conduira à une variation désirée négative (vente), et positive (achat) si le rendement est positif.

Dans le premier cas, la variation désirée est négative et inférieure à α_1 . Par conséquent, cette variation étant suffisante en valeur absolue, la transaction se réalise au niveau $y_t = y_t^* - \alpha_1$, il y a vente du titre ($y_t < 0$). Dans le deuxième cas, la variation désirée est insuffisante et il n'y a pas de transaction, ni à la vente ni à la baisse. Dans le dernier cas, la variation souhaitée est positive et il y a transaction si la variation désirée dépasse un certain seuil positif α_2 .

La vraisemblance s'écrit :

$$L = \prod_{y_t^* < \alpha_1} \frac{1}{\sigma} \phi \left(\frac{y_t + \alpha_1 - x_t \beta}{\sigma} \right) \prod_{\alpha_1 \leq y_t^* < \alpha_2} \left[\Phi \left(\frac{\alpha_2 - x_t \beta}{\sigma} \right) - \Phi \left(\frac{\alpha_1 - x_t \beta}{\sigma} \right) \right] \\ \times \prod_{\alpha_2 \leq y_t^*} \frac{1}{\sigma} \phi \left(\frac{y_t + \alpha_2 - x_t \beta}{\sigma} \right)$$

Notons l'écriture particulière de ce modèle : les trois régimes possibles sont définis en fonction de la variation désirée y_t^* qui est inobservable. On l'écrit donc en fonction de la variable observée y_t dans le premier et le troisième termes, où elle vaut $x_t + \varepsilon_t$ moins une constante qui vaut α_1 et α_2 respectivement.

C. Modèles Tobit généralisés

Il existe plusieurs classes de modèles Tobit (voir Amemiya, 1985, chap.10), dont la première est la spécification Tobit simple. Nous nous concentrerons ici sur le Tobit généralisé le plus simple, le Tobit de type 2.

Le modèle Tobit de type 2 est défini par un système composé de deux variables latentes et d'une variable observable :

$$\begin{cases} y_{1i}^* = x_{1i} \beta_1 + \varepsilon_{1i} \\ y_{2i}^* = x_{2i} \beta_2 + \varepsilon_{2i} \\ y_{2i} = y_{2i}^* & \text{si } y_{1i}^* > 0 \\ & = 0 & \text{si } y_{1i}^* \leq 0 \end{cases}$$

x_{1i} et x_{2i} sont des vecteurs de dimension $(1 \times K_1)$ et $(1 \times K_2)$ respectivement, β_1 et β_2 les vecteurs de paramètres associés. Certaines variables explicatives peuvent être communes à x_{1i} et x_{2i} . ε_{1i} et ε_{2i} sont des variables distribuées selon une loi normale bivariée, de variance respective σ_1^2 et σ_2^2 , et de covariance σ_{12} . En ce qui concerne les observations, le modèle implique que seul le signe de y_{1i}^* est observé, et que y_{2i}^* est observé seulement lorsque $y_{1i}^* > 0$. Le modèle Tobit de type 2 peut être uniquement censuré si x_{1i} et x_{2i} sont observables, même si $y_{1i}^* \leq 0$, ou bien tronqué si x_{2i} n'est pas observable lorsque $y_{1i}^* \leq 0$. En notant respectivement Π_0 et Π_1 les produits sur les observations telles que $y_{2i} = 0$ et $y_{2i} \neq 0$, la vraisemblance du modèle s'écrit :

$$L = \prod_0 \text{Prob}(y_{1i}^* \leq 0) \prod_1 [f(y_{2i}|y_{1i}^* > 0) \times \text{Prob}(y_{1i}^* > 0)]$$

Il est possible d'écrire le terme $f(y_{2i}|y_{1i}^* > 0) \times \text{Prob}(y_{1i}^*)$ en fonction d'une intégrale simple de la densité bivariée de y_{1i}^* et y_{2i}^* , notée $f(\dots)$:

$$f(y_{2i}|y_{1i}^* > 0) \times \text{Prob}(y_{1i}^* > 0) = \int_0^{+\infty} f(z, y_{2i}) dz = \int_0^{+\infty} f(z|y_{2i}) f(y_{2i}) dz$$

On utilise le résultat suivant :

Résultat

Loi conditionnelle d'une variable normale : soit (y_1, y_2) un couple de variables normales de moyennes μ_1 et μ_2 , de variances σ_1^2 , σ_2^2 et de covariance σ_{12} . La loi conditionnelle de y_1 , sachant que $y_2 = \bar{y}_2$, est une loi normale dont l'espérance est :

$$E(y_1|y_2 = \bar{y}_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(\bar{y}_2 - \mu_2)$$

et la variance est $\text{var}(y_1|y_2 = \bar{y}_2) = \sigma_1^2 - \sigma_{12}^2/\sigma_2^2$.

Avec ce résultat, la loi conditionnelle de y_{1i}^* étant donné $y_{2i}^* = y_{2i}$ a pour moyenne $x_{1i}\beta_1 + \sigma_{12}/\sigma_2^2(y_{2i} - x_{2i}\beta_2)$, et pour variance $\sigma_1^2 - \sigma_{12}^2/\sigma_2^2$. La vraisemblance s'écrit finalement :

$$\begin{aligned} L = \prod_0 \left[1 - \Phi \left(\frac{x_{1i}\beta_1}{\sigma_1} \right) \right] \prod_1 & \Phi \left(\frac{x_{1i}\beta_1 + \sigma_{12}/\sigma_2^2(y_{2i} - x_{2i}\beta_2)}{\sqrt{\sigma_1^2 - \sigma_{12}^2/\sigma_2^2}} \right) \\ & \times \frac{1}{\sigma_2} \phi \left(\frac{y_{2i} - x_{2i}\beta_2}{\sigma_2} \right) \end{aligned}$$

Il est également possible (Heckman, 1976 ; Amemiya, 1985) d'utiliser une procédure d'estimation en deux étapes, similaire à celle du Tobit simple, à

partir de l'équation suivante :

$$y_{2i} = E(y_{2i}^* | y_{1i}^* > 0) + \varepsilon_{2i} = x_{2i}\beta_2 + \frac{\sigma_{12}}{\sigma_1} \frac{\phi(x_{1i}\beta_1/\sigma_1)}{\Phi(x_{1i}\beta_1/\sigma_1)} + \varepsilon_{2i}$$

où ε_{2i} est un terme d'erreur normal, de moyenne nulle et de variance :

$$Var(\varepsilon_{12}) = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \left[\frac{x_{1i}\beta_1}{\sigma_1} \frac{\phi(x_{1i}\beta_1/\sigma_1)}{\Phi(x_{1i}\beta_1/\sigma_1)} + \frac{\phi^2(x_{1i}\beta_1/\sigma_1)}{\Phi^2(x_{1i}\beta_1/\sigma_1)} \right]$$

Les estimateurs de première étape de β_1/σ_1 sont obtenus à partir d'un modèle Probit simple. La variable dépendante binaire est définie comme $y_i = 1$ si $y_{1i}^* > 0 \Leftrightarrow y_{2i} > 0$, et $y_i = 0$ si $y_{1i}^* \leq 0 \Leftrightarrow y_{2i} = 0$. On calcule le ratio inverse de Mills estimé $\phi(x_{1i}\beta_1/\sigma_1)/\Phi(x_{1i}\beta_1/\sigma_1)$, qui est remplacé dans l'équation ci-dessus, pour les observations censurées seulement, telles que $y_{2i} \neq 0$.

Comme pour le modèle Tobit simple, les résidus des équations de régression sont hétérosclélastiques par construction. Il est donc nécessaire pour obtenir une inférence paramétrique correcte de corriger cette hétéroscléasticité (par la méthode de White, ou de préférence celle des moindres carrés généralisés, la forme exacte de l'hétéroscléasticité étant connue dans ce cas, voir Wales et Woodland, 1980).

D. Modèles à seuils stochastiques

Une autre extension importante du modèle Tobit consiste à adopter comme seuil de censure (ou de troncature) non plus une constante, mais une autre variable latente. Ainsi, une première variable dépendante limitée sera non nulle si la variable latente associée est supérieure à une seconde variable latente. On parle alors de modèles à seuils stochastiques et inobservables. Le modèle est le suivant, pour une seule censure :

$$\begin{cases} y_{1i}^* = x_{1i}\beta_1 + \varepsilon_{1i} \\ y_{2i}^* = x_{2i}\beta_2 + \varepsilon_{2i} \\ y_{1i} = y_{1i}^* & \text{si } y_{1i}^* > y_{2i}^* \\ = 0 & \text{si } y_{1i}^* \leq y_{2i}^* \end{cases}$$

On fait l'hypothèse que ε_{1i} et ε_{2i} sont normalement et conjointement distribués :

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right]$$

et on note $f(\cdot, \cdot)$ la densité conjointe de ces deux variables.

Pour les observations telles que y_{1i} est observé, on sait que $y_{1i} = y_{1i}^* = x_{1i}\beta_1 + \varepsilon_{1i}$ et donc que $y_{1i} > y_{2i}^*$, soit $\varepsilon_{2i} < y_{1i} - x_{1i}\beta_1$.

Pour le second groupe d'observations, on n'observe ni y_{1i} ni y_{2i} et l'on sait que $y_{1i}^* \leq y_{2i}^* \Leftrightarrow x_{1i}\beta_1 + \varepsilon_{1i} \leq x_{2i}\beta_2 + \varepsilon_{2i}$. De l'hypothèse de normalité sur le vecteur des erreurs, on a :

$$(\varepsilon_{1i} - \varepsilon_{2i}) \sim N(0, \sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$$

et :

$$Prob(y_{1i}^* \leq y_{2i}^*) = \Phi\left[\frac{x_{2i}\beta_2 - x_{1i}\beta_1}{\sigma}\right]$$

La vraisemblance du modèle $L(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ est :

$$\prod_{y_{1i} \neq 0} \left\{ \int_{-\infty}^{y_{1i} - x_{2i}\beta_2} f(\varepsilon_{1i}, \varepsilon_{2i}) d\varepsilon_{2i} \right\} \times \prod_{y_{1i}=0} \Phi\left(\frac{x_{2i}\beta_2 - x_{1i}\beta_1}{\sigma}\right)$$

En écrivant $f(\varepsilon_{1i}, \varepsilon_{2i}) = f(\varepsilon_{2i} | \varepsilon_{1i}) \times f(\varepsilon_{1i})$, on utilise le résultat donné plus haut sur l'expression de la loi conditionnelle d'une variable normale :

$$\varepsilon_{2i} | \varepsilon_{1i} \sim N\left[\varepsilon_{1i} \frac{\sigma_{12}}{\sigma_1^2}; \sigma_{2,1}^2 = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right]$$

où $\sigma_{2,1}^2$ dénote la variance conditionnelle de ε_{2i} connaissant ε_{1i} . Il vient :

$$\begin{aligned} \prod_{y_{1i} \neq 0} \left\{ \int_{-\infty}^{y_{1i} - x_{2i}\beta_2} f(\varepsilon_{1i}, \varepsilon_{2i}) d\varepsilon_{2i} \right\} &= \prod_{y_{1i} \neq 0} \left\{ f(\varepsilon_{1i}) \int_{-\infty}^{y_{1i} - x_{2i}\beta_2} f(\varepsilon_{2i} | \varepsilon_{1i}) d\varepsilon_{2i} \right\} \\ &= \prod_{y_{1i} \neq 0} \{f(\varepsilon_{1i}) \times Prob(\varepsilon_{2i} \leq y_{1i} - x_{2i}\beta_2 | \varepsilon_{1i} = y_{1i} - x_{1i}\beta_1)\} \end{aligned}$$

et la vraisemblance de l'échantillon est finalement :

$$\begin{aligned} L &= \prod_{y_{1i} \neq 0} \left\{ \frac{\phi}{\sigma_1} \left(\frac{y_{1i} - x_{1i}\beta_1}{\sigma_1} \right) \times \Phi\left(\frac{y_{1i} - x_{2i}\beta_2 - (\sigma_{12}/\sigma_1^2)(y_{1i} - x_{1i}\beta_1)}{\sigma_{2,1}} \right) \right\} \\ &\quad \times \prod_{y_{1i}=0} \Phi\left(\frac{x_{2i}\beta_2 - x_{1i}\beta_1}{\sigma}\right) \end{aligned}$$

On remarque dans l'écriture de la probabilité conditionnelle du régime $y_{1i} \neq 0$ l'argument de la fonction de répartition $\Phi(\cdot)$: la borne supérieure de l'intégrale est centrée par rapport à la moyenne conditionnelle de ε_{2i} connaissant ε_{1i} ($(\sigma_{12}/\sigma_1^2)[y_{1i} - x_{1i}\beta_1]$), et divisée par l'écart type conditionnel $\sigma_{2,1} = [\sigma_2^2 - (\sigma_{12}^2/\sigma_1^2)]^{1/2}$.

La vraisemblance ci-dessus correspond à un modèle censuré : même si $y_{1i}^* \leq y_{2i}^*$, les variables explicatives x_{1i} et x_{2i} sont toujours observées. Il est possible d'écrire l'équivalent du modèle avec seuil stochastique dans le cas

d'un échantillon tronqué. Supposons que l'échantillon soit tronqué si $y_{1i}^* \leq y_{2i}^*$. La vraisemblance du modèle s'écrit alors :

$$L = \prod_{y_{1i} \neq 0} \int_{-\infty}^{y_{1i} - x_{2i}\beta_2} f(y_{1i} - x_{1i}\beta_1, \varepsilon_{2i}) d\varepsilon_{2i} \times \frac{1}{\text{Prob}(y_{1i}^* > y_{2i}^*)}$$

où :

$$\text{Prob}(y_{1i}^* > y_{2i}^*) = \Phi\left(\frac{x_{1i}\beta_1 - x_{2i}\beta_2}{\sigma}\right)$$

La log-vraisemblance est alors :

$$-N \log(\sigma_1) - \frac{1}{2\sigma_1^2} \sum_{i=1}^N (y_{1i} - x_{1i}\beta_1)^2 + \sum_{i=1}^N \log \Phi(w_i) - \sum_{i=1}^N \log \Phi\left(\frac{x_{1i}\beta_1 - x_{2i}\beta_2}{\sigma}\right)$$

où :

$$w_i = \frac{y_{1i} - x_{2i}\beta_2 - (\sigma_{12}/\sigma_1^2)(y_{1i} - x_{1i}\beta_1)}{\sigma_{2,1}}$$

VII. Modèles de sélection

A. Mécanisme de sélection

Dans les modèles des sections précédentes, la censure ou la troncature s'écrivait comme une condition sur la variable latente correspondant à la variable dépendante. Ainsi, cette dernière n'était observée que lorsqu'elle était inférieure ou supérieure à un certain seuil, lequel pouvait être constant ou stochastique. Nous considérons ici des modèles très utilisés en pratique, dans lesquels le phénomène de censure ou de troncature provient d'une condition sur une variable différente de la variable latente d'origine. Il apparaît que deux types de modèles sont possibles. Le premier, appelé *troncature auxiliaire* (« *Incidental Truncation* »), considère le cas où la troncature est liée à une variable différente de la variable dépendante, cette dernière n'étant modélisée que pour les observations non tronquées. Le second type de modèle, appelé *modèle à régimes* (*Switching Regression*), généralise le premier en supposant que les observations sont sélectionnées dans l'un ou l'autre des deux régimes possibles, en fonction d'une variable de sélection observable. Dans ce cas, la variable dépendante admet une explication différente, et donc, elle a des paramètres différents dans les deux régimes.

B. Troncature auxiliaire

On utilise le résultat suivant, relatif à la sélection par une variable auxiliaire.

Résultat

Distribution d'une variable normale avec sélection : soit y_1 et y_2 deux variables normales conjointement distribuées, de moyenne μ_1 et μ_2 respectivement, et de matrice de variances-covariances

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

La distribution de y_1 conditionnellement à $y_2 > a_2$ est une loi normale dont les moments sont :

$$E[y_1 | y_2 > a_2] = \mu_1 + \frac{\sigma_{12}}{\sigma_2} \frac{\phi[(a_2 - \mu_2)/\sigma_2]}{1 - \Phi[(a_2 - \mu_2)/\sigma_2]}$$

$$\begin{aligned} Var(y_1 | y_2 > a_2) \\ = \sigma_2^2 \left[1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2} \frac{\phi[(a_2 - \mu_2)/\sigma_2]}{1 - \Phi[(a_2 - \mu_2)/\sigma_2]} \left(\frac{\phi[(a_2 - \mu_2)/\sigma_2]}{1 - \Phi[(a_2 - \mu_2)/\sigma_2]} - \frac{a_2 - \mu_2}{\sigma_2} \right) \right] \end{aligned}$$

On a en effet :

$$\begin{aligned} E(y_1 | y_2 > a_2) &= \mu_1 + E(u_1 | u_2 > a_2 - \mu_2) \\ &= \mu_1 + cov(u_1, u_2) \times E(u_2 | u_2 > a_2 - \mu_2) \\ &= \mu_1 + cov(u_1, u_2 / \sigma_2) \times E[u_2 / \sigma_2 | u_2 / \sigma > (a_2 - \mu_2) / \sigma_2] \\ &= \mu_1 + \frac{\sigma_{12}}{\sigma_2} \times \frac{\phi[(\mu_2 - a_2)/\sigma_2]}{\Phi[(\mu_2 - a_2)/\sigma_2]} \\ &= \mu_1 + \frac{\sigma_{12}}{\sigma_2} \times \frac{\phi[(a_2 - \mu_2)/\sigma_2]}{1 - \Phi[(a_2 - \mu_2)/\sigma_2]} \end{aligned}$$

où $u_1 = y_1 - \mu_1$ et $u_2 = y_2 - \mu_2$.

Résultat

Avec le même raisonnement, l'espérance de y_1 conditionnellement à $y_2 < a_2$ est :

$$E[y_1 | y_2 < a_2] = \mu_1 - \frac{\sigma_{12}}{\sigma_2} \frac{\phi[(a_2 - \mu_2)/\sigma_2]}{\Phi[(a_2 - \mu_2)/\sigma_2]}$$

L'application à un modèle de régression linéaire est directe. L'échantillon est tronqué lorsqu'une variable latente $z_i^* \leq 0$, et on observe z_i telle que $z_i = 1$ si $z_i^* > 0$. Le modèle est le suivant :

$$\begin{cases} y_i = x_i\beta + \varepsilon_i \\ z_i^* = w_i\gamma + v_i \\ z_i = 1 \quad \text{si } z_i^* > 0, 0 \text{ sinon} \\ E(y_i|z_i=1) = x_i\beta + (\sigma_{\varepsilon v}/\sigma_v)\phi(-w_i\gamma/\sigma_v)/[1 - \Phi(-w_i\gamma/\sigma_v)] \end{cases}$$

C. Application : dépenses en eau domestique

Nous étudions comme exemple d'application les dépenses en eau du robinet de ménages, en fonction de leur revenu et de certaines variables socio-démographiques. Dans la mesure où la plupart des ménages en habitat collectif ne paient pas directement l'eau consommée, laquelle est facturée aux locataires via les charges, en fonction de la surface habitable, on ne dispose pas du montant payé pour tous les ménages. La variable de dépense en eau, notée $H2O$, est par conséquent tronquée, et n'apparaît positive que pour les individus propriétaires de leur logement recevant une facture individualisée. Le modèle est alors le suivant. La variable de troncature auxiliaire, notée $PAYH2O$, vaut 1 si le ménage paie directement l'eau consommée, et la variable latente associée reflète le choix d'un logement individuel, en fonction de variables tel le revenu, le nombre d'enfants, etc. La variable dépendante limitée $H2O$ n'est positive que pour $PAYH2O = 1$. On explique cette variable par des caractéristiques du ménage et du logement. Les variables sont : *REVENU* (revenu du ménage), *NBENF* (nombre d'enfants), *NBADUL* (nombre d'adultes), *NBVOIT* (nombre de voitures), *DISTANCE* (distance du lieu de travail).

Le logiciel LIMDEP est utilisé pour estimer ce modèle. Les résultats ci-dessous présentent d'abord l'estimation de l'équation de sélection par Probit simple, puis celle du modèle complet (troncature et régression) par maximum de vraisemblance (procédure SELECT).

La sortie ci-dessous [1] fournit les résultats de la première étape, le Probit binaire. Les expressions utilisées sont comparables à celles des autres procédures de LIMDEP vues plus haut. Par exemple, [2] fournit la statistique du rapport des vraisemblances et sa p-value [3]. [4] indique que les résultats du Probit sont conservés pour le modèle de sélection.

Binomial Probit Model	Maximum Likelihood Estimates [1]
Dependent variable PAYH2O	Number of observations 3966
Iterations completed 4	Log likelihood function -2501.087
Restricted log likelihood -2723.437	Chi-squared 444.7000 [2]
Degrees of freedom 4	Significance level 0.0000000 [3]
Results retained for SELECTION model. [4]	

Variable	Coefficient	Standard Error	$z=b/s.e.$	$P[Z >z]$	Mean of X
Index function for probability					
Constant	-0.6384198	0.47281E-01	-13.503	0.00000	
REVENU	0.2190158	0.19260E-01	11.372	0.00000	1.795
NBENF	0.6144335E-01	0.99211E-02	6.193	0.00000	1.571
NBVVOIT	0.2503297	0.23942E-01	10.456	0.00000	1.565
DISTANCE	-0.7789844E-02	0.17117E-02	-4.551	0.00001	9.782

Le tableau des effectifs observés et prédicts est le suivant :

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability

	Predicted		
Actual	0	1	TOTAL
0	1005	753	1758
1	569	1639	2208
TOTAL	1574	2392	3966

Le logiciel donne à présent les estimations par maximum de vraisemblance du modèle complet, à partir des estimations Probit de première étape. [5] indique que les premiers paramètres estimés correspondent à l'équation de sélection (*Probit equation*). Les estimations concernent l'équation de sélection [6] avec la variable dépendante *PAYH2O*, puis l'équation corrigée pour le régime 1 [7] avec la variable dépendante *H2O*. Le paramètre *Sigma(1)* en [8] est l'écart type de la dépense et *Rho(1, 2)* en [9] est le coefficient de corrélation des termes d'erreur de l'équation de troncature et de l'équation de régime.

ML Estimates of Selection Model

Maximum Likelihood Estimates

Dependent variable H₂O

Number of observations 3966

Iterations completed 19

Log likelihood function = -16554.26

First 5 estimates are probit equation

Variable	Coefficient	Standard Error	z=b/s.e.	P Z>z
Selection (probit) equation for PAYH2O [6]				
Constant	-0.6440295	0.43246E-01	-14.892	0.00000
REVENU	0.2198589	0.16759E-01	13.119	0.00000
NBENF	0.6394323E-01	0.83426E-02	7.665	0.00000
NBVOIT	0.2501756	0.20853E-01	11.997	0.00000
DISTANCE	-0.7764667E-02	0.16364E-02	-4.745	0.00000
Corrected regression, Regime 1 [7]				
Constant	141.8859	21.080	6.731	0.00000
REVENU	8.541711	1.4245	5.996	0.00000
NBADUL	23.69294	8.4429	2.806	0.00501
NBENF	7.067588	0.67940	10.403	0.00000
Sigma(1) [8]	141.6663	1.3264	106.808	0.00000
Rho(1,2) [9]	-0.1600767	0.98758E-01	-1.621	0.10504

En comparant les estimations de l'équation de troncature (*PAYH2O*) par Probit seul puis pour le système complet, on voit que la deuxième procédure procure des paramètres plus efficaces. Utiliser l'information complète du modèle apporte plus d'information sur l'équation de troncature. Un salaire élevé ainsi qu'un nombre d'enfants et de voitures plus important a un effet positif et significatif sur le choix d'un logement individuel, dans lequel l'eau est facturée individuellement. En ce qui concerne l'équation de dépense en eau, *H2O*, celle-ci dépend positivement du revenu, du nombre d'adultes et d'enfants du foyer.

D. Application : équation de salaire

Dans cette application, on utilise le modèle avec troncature auxiliaire pour déterminer l'influence des études supérieures sur le revenu salarial, avec un échantillon de 3966 individus. La variable de sélection, *ÉTUSUP*, vaut 1 si la personne a accompli un cycle d'études supérieures, 0 sinon. Elle est expliquée par les variables : *SEXÈ* (1 si homme, 0 si femme), *ENFVILLE* (indicateur, 1 si la personne a grandi dans une ville) et *ÉTUSEC* (un indicateur de la scolarité : le nombre d'années dans le système secondaire). La variable dépendante continue est le logarithme du revenu salarial, *REVENU*, expliqué par *SEXÈ*, *ENFVILLE*, *COMMER* (1 si la personne a une activité commerciale), *TRANSPOR* (1 si activité dans les transports), *MANUFACT* (1 si activité dans le secteur manufacturier) et *ADMINIST* (1 si emploi administratif). Le logiciel LIMDEP estime d'abord l'équation de sélection par un Probit simple, puis l'équation de salaire pour les individus ayant un diplôme supérieur (Régime 1). La présentation des sorties de LIMDEP étant similaire à celle de l'application précédente, on ne rappellera pas la signification des termes utilisés. Les paramètres estimés par maximum de vraisemblance figurent en [1], pour les deux équations : Probit binaire pour l'équation de sélection sur la variable *ÉTUSUP* en [2], équation de régime pour la variable dépendante *REVENU* en [3]. Il est intéressant de constater que le fait d'avoir grandi dans une ville (*ENFVILLE*) a un effet négatif sur le fait de suivre des études supérieures, mais un effet positif sur le revenu. Le logiciel fournit l'écart type du log du salaire dans le régime (Sigma (1), en [4]) et la corrélation entre les erreurs dans l'équation de sélection et l'équation de régime (Rho (1, 2), en [5]). On constate que cette corrélation est significative et négative.

ML Estimates of Selection Model
Dependent variable REVENU
Iterations completed 14
First 4 estimates are probit equation.

Maximum Likelihood Estimates [1]
Number of observations 3966
Log likelihood function -2045.439

Variable	Coefficient	Standard Error	z=b/s.e.	P[Z >z]
Selection (probit) equation for ETUSUP [2]				
Constant	-18.95446	0.71852	-26.380	0.00000
SEXE	-0.1815424E-01	0.13467	-0.135	0.89276
ENFVILLE	-0.2498473	0.10596	-2.358	0.01838
ETUSEC	1.267709	0.47400E-01	26.745	0.00000
Corrected regression, Regime 1 [3]				
Constant	1.368104	0.51532	2.655	0.00793
SEXE	1.407135	0.50880	2.766	0.00568
ENFVILLE	0.4826653	0.18394	2.624	0.00869
COMMER	-0.1949914	0.89402	-0.218	0.82735
TRANSPOR	-0.2071763	1.0899	-0.190	0.84924
MANUFACT	0.2905007	0.23109	1.257	0.20872
ADMINIST	-1.599272	6.5148	-0.245	0.80608
Sigma(1) [4]	2.223809	0.33593E-01	66.198	0.00000
Rho(1,2) [5]	-0.2372725	0.11245	-2.110	0.03486

E. Modèles à régimes

Le signe de la variable latente z_i^* détermine ici l'affectation des observations à l'un ou l'autre des deux régimes. La variable dépendante est expliquée de façon différente selon le régime, elle n'est donc plus réellement « tronquée » pour l'observateur. Cependant, le mécanisme de troncation existe bel et bien, dans la mesure où les observations affectées au premier régime sont perdues pour le second et vice versa. L'interprétation de la variable dépendante n'est plus la même selon que l'individu appartient au premier régime ou au second.

En pratique, nous disposons des variables dépendantes y_{1i}, y_{2i} (une pour chaque régime) et des variables explicatives x_{1i}, x_{2i} (qui peuvent être communes aux deux équations). Deux cas se présentent, selon que l'affectation des individus aux équations est observable ou non.

Régime observable

Dans le premier cas, on dispose en plus d'une variable indiquant dans quel régime l'individu se situe. Les paramètres du modèle peuvent alors être estimés par maximum de vraisemblance ou une procédure en deux étapes, de façon analogue aux modèles Tobit vus plus haut. Si I_i est la variable de sélection, telle que $I_i = 1$ si l'individu appartient au premier régime et $I_i = 0$ sinon, le système s'écrit :

$$\begin{cases} y_{1i} = x_{1i}\beta_1 + \varepsilon_{1i} & \text{si } I_i = 1 \\ y_{2i} = x_{2i}\beta_2 + \varepsilon_{2i} & \text{si } I_i = 0 \end{cases}$$

La variable de sélection I_i est une indicatrice observée que l'on peut supposer être déterminée par le signe d'une variable latente I_i^* :

$$\begin{cases} I_i^* = z_i \gamma + u_i \\ I_i = 1 \quad \text{si} \quad I_i^* > 0 \quad \Leftrightarrow u_i > -z_i \gamma \\ I_i = 0 \quad \text{si} \quad I_i^* \leq 0 \quad \Leftrightarrow u_i \leq -z_i \gamma \end{cases}$$

La première équation de ce système est appelée *équation de sélection*. On fait l'hypothèse que les erreurs $\varepsilon_{1i}, \varepsilon_{2i}$ et u_i suivent une loi normale trivariée, avec $E(\varepsilon_{1i}^2) = \sigma_1^2, E(\varepsilon_{2i}^2) = \sigma_2^2, E(\varepsilon_{1i}\varepsilon_{2i}) = \sigma_{12}, E(\varepsilon_{1i}u_i) = \sigma_{1u}, E(\varepsilon_{2i}u_i) = \sigma_{2u}$.

La vraisemblance du modèle s'écrit :

$$L = \prod_{i=1}^N \left[\int_{-\infty}^{z_i \gamma} f(y_i - x_{1i}\beta_1, u_i) du_i \right]^{I_i} \times \left[\int_{z_i \gamma}^{+\infty} g(y_i - x_{2i}\beta_2, u_i) du_i \right]^{1-I_i}$$

où $f(\cdot, \cdot)$ et $g(\cdot, \cdot)$ sont les densités bivariées (normales) de (ε_{1i}, u_i) d'une part, de (ε_{2i}, u_i) de l'autre. La maximisation de la log-vraisemblance étant particulièrement délicate en pratique, une méthode en deux étapes est le plus souvent utilisée. La variable indicatrice du régime I_i est à la base d'une estimation par un modèle Probit simple des rapports inverses de Mills, $-\phi_i/\Phi_i$ et $\phi_i/(1-\Phi_i)$, à partir de l'équation de sélection. Les espérances conditionnelles de y_{1i} et y_{2i} sont en effet :

$$\begin{cases} E(y_{1i}|I_i^* > 0) = x_{1i}\beta_1 + (\sigma_{1u}/\sigma_u)\phi(-z_i \gamma)/[1 - \Phi(-z_i \gamma)] \\ E(y_{2i}|I_i^* \leq 0) = x_{2i}\beta_2 - (\sigma_{2u}/\sigma_u)\phi(-z_i \gamma)/\Phi(-z_i \gamma) \end{cases}$$

Ces estimateurs convergents sont remplacés dans les deux équations de régimes pour estimer $\beta_1, \beta_2, \sigma_{12}/\sigma_2$ et σ_{12}/σ_1 :

$$\begin{cases} y_{1i} = x_{1i}\beta_1 + (\sigma_{1u}/\sigma_u)\phi(-z_i \gamma)/[1 - \Phi(-z_i \gamma)] + v_{1i} \\ y_{2i} = x_{2i}\beta_2 - (\sigma_{2u}/\sigma_u)\phi(-z_i \gamma)/\Phi(-z_i \gamma) + v_{2i} \end{cases}$$

où v_{1i} et v_{2i} sont des résidus de moyenne nulle et hétéroscélastiques. Une méthode de moindres carrés pondérés sera préférée aux MCO (pour l'estimation des variances des erreurs, voir Greene, 1997). Notons que de façon analogue à la procédure en deux étapes des modèles Tobit, la première équation ne sera estimée que sur les observations telles que $I_i = 1$, la seconde sur celles avec $I_i = 0$. De plus, la variance de l'erreur u_i , tout comme dans le modèle Probit simple, n'est pas identifiable ; par conséquent, elle est normalisée à 1.

F. Application : dépenses en logement

On considère un modèle de dépenses en logement pour 3 311 ménages qui se répartissent en deux catégories : achat ou location de leur habitation. L'équation de sélection détermine le choix des ménages entre ces deux catégories, en fonction des variables : *REVENU* (revenu salarial du chef de famille), *NBENF* (nombre d'enfants), *NBVOIT* (nombre de véhicules du ménage), *ENFVILLE* (indicatrice, 1 si le chef de famille a grandi dans une ville), *TEMPS* (temps moyen de déplacement vers le lieu de travail). Les équations de régime expliquent la dépense en logement (variable *DÉPENSE*) qui est formée soit des dépenses en loyer, soit du paiement des échéances du crédit immobilier. Les variables explicatives des équations de régime sont : *NBPIÈCES* (nombre de pièces du logement), *NBVOIT*, *DISTANCE* (distance entre le lieu de résidence et le lieu de travail). Le logiciel LIMDEP permet d'estimer facilement ce modèle à régimes, avec sélection observable. On suppose ici que le ménage choisit le mode associé au coût le plus faible ; la variable dépendante (*DÉPENSE*) sera alors égale au minimum des dépenses en location et en crédit immobilier. La première étape est le modèle Probit appliqué à l'équation de sélection. La variable de sélection est *LOCATION*, égale à 1 si le ménage loue son habitation et 0 s'il l'achète. Les résultats sont donnés dans la sortie [1]. Le logiciel indique la variable dépendante [2] et le nombre d'observations [3], ainsi que la conservation des résultats pour le modèle de sélection ultérieur [4]. Les paramètres estimés sont fournis pour l'équation de sélection [5] (fonction indicatrice pour la probabilité).

Binomial Probit Model Maximum Likelihood Estimates [1]
 Dependent variable LOCATION [2] Number of observations 3311 [3]
 Iterations completed 5 Log likelihood function -1772.688
 Restricted log likelihood -2293.530 Chi-squared 1041.683
 Degrees of freedom 5 Significance level 0.0000000
 Results retained for SELECTION model. [4]

Variable	Coefficient	Standard Error	$z = b/s.e.$	$P(Z > z)$	Mean of X
Index function for probability					
Constant	1.516341	0.69805E-01	21.723	0.00000	
REVENU	-0.3252701	0.24089E-01	-13.503	0.00000	1.843
NBENF	-0.8417306E-01	0.12098E-01	-6.957	0.00000	1.426
NBVVOIT	-0.5813326	0.32593E-01	-17.836	0.00000	1.539
ENFVILLE	0.1821359	0.50415E-01	3.613	0.00030	0.4005
TEMPS	-0.3446740E-03	0.17999E-03	-1.915	0.05550	155.0

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

	Predicted		
Actual	0	1	TOTAL
0	1261	444	1705
1	399	1207	1606
TOTAL	1660	1651	3311

On remarque que le revenu du chef de famille, le nombre d'enfants et le nombre de voitures jouent à la baisse sur la probabilité de choisir une location, ce qui est conforme à l'intuition. Le fait d'avoir grandi dans une ville agit par contre à la hausse et le temps passé pour se rendre au travail n'est pas significatif. Les équations de régimes sont ensuite estimées par maximum de vraisemblance, avec la variable dépendante *DÉPENSE*. Le régime 1 concerne les ménages en location, le régime 2 ceux achetant leur résidence principale. LIMDEP indique l'utilisation de la procédure de régression avec régimes (6, *Switching Regressions*) et le nom de la variable dépendante (7). Le logiciel mentionne aussi le fait que la variable de sélection est *LOCATION* (8) et la forme du modèle (9). Les paramètres estimés sont donnés pour les deux équations : régime 1 en (10) et régime 2 en (11), *RHS for Regime* (*Right-Hand Side*, côté droit de l'équation) signifiant que les variables explicatives dans chacun des deux régimes sont prises en compte. Le paramètre RHO en (12) est le coefficient de corrélation entre les deux équations. Enfin, les paramètres s(1) et s(0) (13) et (14) sont les écarts types de *DÉPENSE* dans les deux régimes. Dans les deux cas, le nombre de pièces et le revenu ont un effet positif sur les dépenses en logement, et la distance de la résidence au lieu de travail a un effet négatif, ce qui indique que les habitations éloignées des lieux de travail sont moins coûteuses, le nombre de pièces du logement étant bien sûr corrélé positivement avec la dépense. La corrélation RHO entre les résidus des deux équations est significative et positive.

Switching Regressions (6) Maximum Likelihood Estimates
 Dependent variable *DÉPENSE* (7) Number of observations 3311
 Iterations completed 18 Log likelihood function -32416.84
 Sample separation variable is *LOCATION* (8)
DÉPENSE is the minimum of $y^*(1)$ and $y^*(0)$ (9)

Variable	Coefficient	Standard Error	$z=b/s.e.$	$P[Z >z]$	Mean of \bar{X}
RHS for Regime 1 (10)					
Constant	1150.095	361.11	3.185	0.00145	
NBPIECES	403.4512	40.429	9.979	0.00000	4.232
NBVOIT	58.30611	45.531	1.281	0.20034	1.098
DISTANCE	-31.62494	3.9909	-7.924	0.00000	8.271
REVENU	748.8991	21.220	35.292	0.00000	1.314

	RHS for Regime 2 [1]			
Constant	-15.42693	277.76	-0.056	0.95571
NBPIECES	666.6400	71.229	9.359	0.00000
NBVOIT	-47.53532	72.609	-0.655	0.51267
DISTANCE	-17.43542	5.9416	-2.934	0.00334
REVENU	1027.882	65.142	15.779	0.00000
RHO [2]	0.6284515	0.15553	4.041	0.00005
s(1) [3]	2436.113	55.880	43.595	0.00000
s(0) [4]	2781.774	184.18	15.104	0.00000

Régime inobservable

Dans ce cas, l'affectation des individus aux différents régimes est toujours supposée dépendre de la même équation de sélection, mais la variable binaire I_i est inobservable. Ce genre de modèle est à la base d'une recherche importante en économétrie du déséquilibre, dont les modèles les plus simples fournissent de bonnes illustrations des problèmes posés. Deux équations donnent respectivement la demande et l'offre sur le marché d'un produit, et l'on relâche l'hypothèse d'équilibre sur ce marché pour étudier les déterminants de l'offre et de la demande quand le marché est en déséquilibre. Ainsi, la possibilité de rationnement des consommateurs ou d'une offre excédentaire est explicitement prise en compte. Soit p_t le prix du bien sur le marché à la date t , Q_t la quantité finale échangée, x_{1t} et x_{2t} des variables explicatives de la demande et de l'offre respectivement (hormis le prix). Toutes ces variables sont observables. Le modèle de Fair et Jaffee (1972) est :

$$D_t = \alpha_1 p_t + x_{1t} \beta_1 + \varepsilon_{1t} \quad (\text{demande})$$

$$O_t = \alpha_2 p_t + x_{2t} \beta_2 + \varepsilon_{2t} \quad (\text{offre})$$

où ε_{1t} et ε_{2t} suivent une loi normale bivariée de densité $f(\cdot, \cdot)$. Dans un modèle d'équilibre, on a $Q_t = D_t = O_t$. Les équations ci-dessus fournissent un système d'équations simultanées avec deux variables endogènes, Q_t et p_t , et des variables exogènes x_{1t} et x_{2t} . Dans le cas du déséquilibre, l'échange s'effectue au minimum de l'offre et de la demande. La condition $Q_t = D_t = O_t$ est donc remplacée dans le modèle de Fair et Jaffee par :

$$Q_t = \min(D_t, O_t)$$

Si $D_t > O_t$, alors la quantité observée est du côté de l'offre, et si $O_t > D_t$, Q_t est du côté de la demande.

La méthode du maximum de vraisemblance pour estimer le modèle de Fair et Jaffee est la suivante (Maddala et Nelson, 1974 ; Amemiya, 1974 ; Fair et Kelejian, 1974). Tout d'abord, puisque le prix est exogène, on peut l'inclure dans la liste des exogènes (x_{1t}, x_{2t}). Le modèle est le suivant :

$$\begin{cases} D_t &= x_{1t} \beta_1 + \varepsilon_{1t} \\ O_t &= x_{2t} \beta_2 + \varepsilon_{2t} \\ Q_t &= \min(D_t, O_t) \end{cases}$$

Soit $g(D_t, O_t)$ la densité de D_t et O_t dérivée de $f(\dots)$. Considérons d'abord le cas où la période t correspond à une quantité située dans le régime de demande : $Q_t = D_t$ et $O_t > D_t$. La probabilité, notée π_t , que l'observation t soit du côté de la demande peut être calculée par :

$$\pi_t = \text{Prob}(D_t < O_t) = \text{Prob}(\varepsilon_{1t} - \varepsilon_{2t} < x_{2t}\beta_2 - x_{1t}\beta_1)$$

La densité associée à la quantité échangée, conditionnellement au fait que $Q_t = D_t$, est :

$$h(Q_t|Q_t = D_t) = \int_{Q_t}^{+\infty} g(D_t = Q_t, O_t) dO_t / \pi_t$$

qui correspond à l'expression de la densité d'une variable tronquée, donnée précédemment dans ce chapitre⁴. Considérons ensuite le cas dans lequel l'observation t est du côté de l'offre, c'est-à-dire que $Q_t = O_t$ et $D_t > O_t$. La densité conditionnelle de la quantité, étant donné que $Q_t = O_t$, est :

$$h(Q_t|Q_t = O_t) = \int_{Q_t}^{+\infty} g(D_t, O_t = Q_t) dD_t / (1 - \pi_t)$$

Puisque le régime est inobservable, on ne dispose pas d'une indicatrice pouvant servir de base à un modèle Probit de sélection, comme précédemment. En revanche, la densité non conditionnelle peut toujours être calculée et servir de base à l'écriture de la fonction de vraisemblance associée aux observations. La densité non conditionnelle de Q_t est :

$$\begin{aligned} h(Q_t) &= \pi_t \times h(Q_t|Q_t = D_t) + (1 - \pi_t)h(Q_t|Q_t = O_t) \\ &= \int_{Q_t}^{+\infty} g(D_t = Q_t, O_t) dO_t + \int_{Q_t}^{+\infty} g(D_t, O_t = Q_t) dD_t \end{aligned}$$

et la vraisemblance du modèle s'écrit :

$$L = \prod_{t=1}^T h(Q_t)$$

En utilisant les résultats de la sous-section consacrée aux modèles à seuils stochastiques et avec l'hypothèse de normalité :

$$\begin{aligned} \int_{Q_t}^{+\infty} g(D_t = Q_t, O_t) dO_t &= g(D_t = Q_t) \times \int_{Q_t}^{+\infty} g(O_t|D_t = Q_t) dO_t \\ &= \frac{\phi}{\sigma_1} \left(\frac{Q_t - x_{1t}\beta_1}{\sigma_1} \right) \times \text{Prob}(\varepsilon_{2t} > Q_t - x_{2t}\beta_2 | \varepsilon_{1t} = Q_t - x_{1t}\beta_1) \end{aligned}$$

4. Le dénominateur π_t joue le rôle d'une constante d'intégration afin que l'intégrale de cette densité somme bien à 1.

$$= \frac{\phi}{\sigma_1} \left(\frac{Q_t - x_{1t}\beta_1}{\sigma_1} \right) \times \left\{ 1 - \Phi \left[\frac{Q_t - x_{2t}\beta_2 - (\sigma_{12}/\sigma_1^2)(Q_t - x_{1t}\beta_1)}{[\sigma_2^2 - \sigma_{12}^2/\sigma_1^2]^{1/2}} \right] \right\}$$

De même, la contribution à la vraisemblance d'une observation du côté de l'offre s'écrit :

$$\begin{aligned} & \int_{Q_t}^{+\infty} g(D_t | O_t = Q_t) dD_t = g(O_t = Q_t) \times \int_{Q_t}^{+\infty} g(D_t | O_t = Q_t) dD_t \\ &= \frac{\phi}{\sigma_2} \left(\frac{Q_t - x_{2t}\beta_2}{\sigma_2} \right) \times \text{Prob}(e_{1t} > Q_t - x_{1t}\beta_1 | e_{2t} = Q_t - x_{2t}\beta_2) \\ &= \frac{\phi}{\sigma_2} \left(\frac{Q_t - x_{2t}\beta_2}{\sigma_2} \right) \times \left\{ 1 - \Phi \left[\frac{Q_t - x_{1t}\beta_1 - (\sigma_{12}/\sigma_2^2)(Q_t - x_{2t}\beta_2)}{[\sigma_1^2 - \sigma_{12}^2/\sigma_2^2]^{1/2}} \right] \right\} \end{aligned}$$

Voir Maddala (1983) et Gouriéroux (1989) pour une discussion plus approfondie de ces modèles.

À retenir

Il existe de nombreux cas en pratique où la variable dépendante n'est observée que sur un intervalle. On parle d'échantillons censurés ou tronqués, selon que les observations restent disponibles ou sont éliminées de l'échantillon. Le modèle Tobit est le plus utilisé pour estimer les modèles de régression censurés (dépenses en biens durables, enquêtes avec consentement à payer, etc.). La généralisation au cas de plusieurs variables dépendantes limitées est immédiat. Les mécanismes de censure et de troncature sont étendus aux cas où l'observation d'une variable dépendante limitée est fonction du signe d'une variable auxiliaire. Un cas d'application intéressant est alors celui des modèles avec sélection endogène et régimes. Une application importante réside dans l'estimation de modèles de déséquilibre sur un marché.

Exercices

Exercice n°1 : renormalisation du Tobit

Considérons le modèle suivant :

$$\begin{cases} y_i &= x_i\beta + e_i \quad \text{si } y_i > c_i \\ &= c_i \quad \text{sinon} \end{cases}$$

Écrire une reparamétrisation des variables afin de retrouver le modèle Tobit simple vu dans ce chapitre, caractérisé par un seuil égal à 0.

Exercice n°2 : modèle de déséquilibre

On considère un modèle de déséquilibre dans lequel la variation du prix sur le marché est observée : $\Delta_t P_t = P_t - P_{t-1}$. L'hypothèse est que la demande est supérieure à l'offre à la date t si $\Delta_t P_t > 0$, et l'offre est supérieure à la demande si $\Delta_t P_t < 0$. Le modèle s'écrit :

$$\begin{cases} D_t = x_{1t}\beta_1 + \alpha_1 P_t + \varepsilon_{1t} \\ O_t = x_{2t}\beta_2 + \alpha_2 P_t + \varepsilon_{2t} \\ Q_t = \min(D_t, O_t) \end{cases}$$

Écrire la vraisemblance du modèle.

Exercice n°3 : variable normale tronquée

Soit y une variable normale, de moyenne μ et de variance σ^2 . On dispose des informations suivantes sur y :

$$E(y|y > 2,5) = 3,80 \quad \text{et} \quad \text{Prob}(y > 2,5) = 0,40$$

En utilisant la formule de la loi normale tronquée, donnée dans ce chapitre, et le fait que $\text{Prob}(y > 2,5) = 0,40 \Leftrightarrow \text{Prob}(y \leq 2,5) = 0,60 = \Phi(w)$, retrouver les valeurs de μ et σ . Indication : si $\Phi(w) = 0,60$ alors $w = 0,2533$.

Exercice n°4 : demande incompressible

On considère une population de consommateurs dotés d'une fonction de demande pour un bien :

$$C_i = a_0 + \alpha_1 P_i + \varepsilon_i \quad i = 1, 2, \dots, N$$

où C_i est la consommation d'unités du bien, P_i le prix unitaire payé par le consommateur et ε_i un terme d'erreur aléatoire de moyenne nulle. Il existe une demande incompressible pour ce bien, notée \bar{C}_i :

$$\bar{C}_i = \beta_0 + \beta_1 A_i + v_i$$

où A_i est l'âge du consommateur et v_i un terme aléatoire de moyenne nulle. Dans l'échantillon des consommateurs, seuls sont observés les niveaux de consommation pour lesquels la consommation C_i est supérieure à la consommation incompressible. On a :

$$\begin{cases} C_i > 0 & \text{si } C_i > \bar{C}_i \\ C_i = 0 & \text{si } C_i \leq \bar{C}_i \end{cases}$$

On suppose que ε_i et v_i suivent une loi normale bivariée, avec $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2$, $\text{var}(v_i) = \sigma_v^2$, $\text{cov}(\varepsilon_i, v_i) = \sigma_{\varepsilon v}$.

1) Écrire le modèle ci-dessus sous la forme de conditions sur les variables aléatoires ε_i et v_i , en fonction des variables observées et des paramètres, et en distinguant les deux cas ($C_i > 0$ et $C_i = 0$).

2) Écrire la log-vraisemblance du modèle. Indication : on notera N_0 le nombre d'observations pour lesquelles $C_i = 0$, N_1 celui pour lesquelles $C_i > 0$.

3) On suppose à présent que toutes les consommations supérieures à un niveau noté C^* sont éliminées de l'échantillon. Néanmoins la règle d'observation de C_i décrite ci-dessus s'applique toujours. Écrire la log-vraisemblance associée à ce modèle.

Exercice n°5 : application numérique

On dispose de l'échantillon suivant :

y	x_1	x_2
0	0,2	2
10	0,8	3
12	2,3	2
13,5	1,8	2
0	0,5	1
9,5	0,7	4
15,5	2,1	5
14	1,9	4
0	1,1	1
0	1,5	2
0	1	2
8,5	1,2	2
11	2	3
14,5	3,9	4
9,5	2,2	2
6,2	0,6	1

- 1) Pour ces 16 observations, estimer les paramètres associés à x_1 et x_2 , ainsi que la constante, avec un modèle Tobit simple. Si on utilise le logiciel SAS (procédure Lifereg), calculer les statistiques de Student.
- 2) Éliminer les observations telles que $y = 0$ de l'échantillon, et estimer le modèle par moindres carrés ordinaires. Comparer les résultats avec ceux en 1. En particulier, la significativité des paramètres est-elle modifiée ?



Corrigés des exercices

Chapitre 1

Exercice n° 1

1) On trouve $\hat{\beta}_0 = 1,995\,68$ et $\hat{\beta}_1 = 0,484\,08$.

Par la formule du t de Student $t = \hat{\beta}/\sqrt{var(\hat{\beta})} = \hat{\beta}/[\sigma^2 diag(X'X)^{-1}]^{1/2}$, on a :

$$1,995\,68/[\sigma * 0,094\,48] = 65,69$$

ou bien : $0,484\,08/[\sigma * 0,142\,44] = 10,56$; soit : $\sigma^2 = (0,321)^2 = 0,1030$.

2) $E(y_i|x_i = 0) = E(\hat{\beta}_0) = 1,995\,68$. Puisque :

$$E(\hat{\beta}_0 + \hat{\beta}_1) = E(y_i|x_i = 1) = E(x_i y_i)/E(x_i) = 1,0910/E(x_i)$$

on déduit que $E(x_i) = 1,0910/(\hat{\beta}_0 + \hat{\beta}_1) = 0,439\,96$.

Exercice n° 2

1) Le tableau suivant donne dans chaque cellule, dans l'ordre et en %, la fréquence empirique, le profil-ligne et le profil-colonne.

Delta	Secteur					Total
	(1)	(2)	(3)	(4)	(5)	
< 0,25	2,02	2,36	9,43	2,69	3,70	20,20
	10,00	11,67	46,67	13,33	18,33	
	12,00	17,07	27,18	26,67	15,07	
[0,25 ; 0,5]	3,70	3,70	8,08	1,35	2,69	19,53
	18,97	18,97	41,38	6,90	13,79	
	22,00	26,83	23,30	13,33	10,96	
[0,5 ; 0,75]	8,75	6,73	5,05	2,36	17,51	40,40
	21,67	16,67	12,50	5,83	43,33	
	52,00	48,78	14,56	23,33	71,23	
> 0,75	2,36	1,01	12,12	3,70	0,67	19,87
	11,86	5,08	61,02	18,64	3,39	
	14,00	7,32	34,95	36,67	2,74	
Total	16,84	13,80	34,68	10,10	24,58	100,00

Pour trouver le secteur le plus important dans le cas $\delta > 0,75$ on cherche le maximum des profils-lignes. C'est donc le secteur (3) (Chimie), avec 62,01 %. La catégorie la plus importante en termes relatifs pour l'agro-alimentaire (secteur (1)) s'obtient en regardant le maximum des profils-colonnes : le cas $\delta \in [0,5 ; 0,75]$ est le plus représenté avec 52 %.

2) La statistique de test χ^2 est 80,9739 avec $(5 - 1)(4 - 1) = 12$ degrés de liberté. On rejette donc l'hypothèse d'indépendance des modalités, les performances en dépollution sont bien liées au secteur industriel.

Exercice n° 3

1) Avec le nouveau regroupement, le tableau est :

Delta	Secteur			Total
	1+2	3+4	5	
< 0,5	11,78	1,35	6,40	19,53
	60,34	6,90	32,76	
	24,31	13,33	15,45	
[0,5 ; 0,75]	11,78	2,36	26,26	40,40
	29,17	5,83	65,00	
	24,31	23,33	63,41	
> 0,75	24,92	6,40	8,75	40,07
	62,18	15,97	21,85	
	51,39	63,33	21,14	
	48,48	10,10	41,41	100,00

La statistique de test χ^2 est 50,4998 avec 4 degrés de liberté. On rejette là encore l'hypothèse d'indépendance des modalités.

2) On a le tableau des fréquences empiriques, profils-lignes et profils-colonnes :

Delta	Secteur		Total
	1+2	3+4+5	
≤ 0,5	27,27	32,66	59,93
	45,51	54,49	
	46,55	78,86	
> 0,5	31,31	8,75	40,07
	78,15	21,85	
	53,45	21,14	
	58,59	41,41	100,00

Le modèle log-linéaire est :

$$\left\{ \begin{array}{lcl} \log P_{00} & = -1,2992 & = \mu_0 - \mu_1 - \mu_2 + \mu_{12} \\ \log P_{01} & = -1,1190 & = \mu_0 - \mu_1 + \mu_2 - \mu_{12} \\ \log P_{10} & = -1,1611 & = \mu_0 + \mu_1 - \mu_2 - \mu_{12} \\ \log P_{11} & = -2,4356 & = \mu_0 + \mu_1 + \mu_2 + \mu_{12} \end{array} \right.$$

où le premier indice est relatif à δ , le second au secteur.

3) Puisque les effets marginaux et croisés sont de moyenne nulle, on obtient en prenant la moyenne des logarithmes des probabilités :

$$\mu_0 = 1/4(\log P_{00} + \log P_{01} + \log P_{10} + \log P_{11}) = \bar{P}_-$$

De plus, en sommant les deux dernières équations, nous obtenons :

$$\mu_1 = 1/2(\log P_{10} + \log P_{11}) - \mu_0 = \bar{P}_{1-} - \bar{P}_-$$

On a aussi, en sommant les deuxième et quatrième équations :

$$\mu_2 = 1/2(\log P_{01} + \log P_{11}) - \mu_0 = \bar{P}_{.1} - \bar{P}_-$$

Enfin, à partir de la dernière par exemple, le terme croisé est :

$$\begin{aligned}\mu_{12} &= \log P_{11} - \mu_0 - \mu_1 - \mu_2 = \log P_{11} - \bar{P}_- - \bar{P}_{1-} + \bar{P}_{.1} - \bar{P}_{.1} + \bar{P}_- \\ &= \log P_{11} - \bar{P}_{1-} - \bar{P}_{.1} + \bar{P}_-\end{aligned}$$

À partir de ces expressions, les valeurs estimées des paramètres sont : $\mu_0 = -1,5037$; $\mu_1 = -0,2946$; $\mu_2 = -0,2736$; $\mu_{12} = 0,7727$.

Exercice n° 4

1) Si $\log P_j$ est la probabilité théorique, on a vu que la fréquence empirique s'écrit $\log \hat{P}_j = x_i \beta + u_j$, où le terme d'erreur est : $u_j = \log P_j - \log \hat{P}_j$. Du développement limité autour de P_j :

$$\log \hat{P}_j = \log P_j + (\hat{P}_j - P_j)1/P_j + \dots$$

on a : $u_j \approx (\hat{P}_j - P_j)/P_j$. La variance de l'erreur est donc approchée par :

$$var(\hat{P}_j)/P_j^2 = P_j(1 - P_j)/N_j \times (1/P_j^2)$$

où l'on remplace P_j par son estimation \hat{P}_j pour obtenir :

$$var(u_j) \approx (1 - \hat{P}_j)/(N_j \hat{P}_j)$$

2) Pour le modèle B, un développement limité de l'expression de $\log(\hat{P}_j/(1 - \hat{P}_j))$ autour de P_j donne :

$$\log(\hat{P}_j/(1 - \hat{P}_j)) \approx \log(P_j/(1 - P_j)) + (\hat{P}_j - P_j) \times [1/P_j + 1/(1 - P_j)] + \dots$$

soit :

$$u_j \approx (\hat{P}_j - P_j) \times [1/(P_j(1 - P_j))]$$

La variance de l'erreur est alors approchée par :

$$[\hat{P}_j(1 - \hat{P}_j)/N_j] \times [1/(\hat{P}_j^2(1 - \hat{P}_j))^2] = 1/[N_j \hat{P}_j(1 - \hat{P}_j)]$$

Chapitre 3

Exercice n° 1

1) Pour comparer les estimations du Probit et du Logit, on multiplie les paramètres Probit par $\pi/\sqrt{3} \approx 1,81$ pour obtenir :

Variable	Probit	Logit	Probit $\times 1,81$
Constante	-7,1375	-12,0471	-12,9188
NBFOYER	-0,0431	-0,0645	-0,0780
TAILLE	0,00061	0,00109	0,00104
REVENU	0,6943	1,1587	1,2566
MAISON	-0,3039	-0,4940	-0,5500
PORTABLE	-1,0354	-1,8064	-1,8740

Au vu des paramètres des colonnes 1 et 3, l'approximation n'est pas très précise. La condition discutée dans le texte pour que cette règle de conversion soit pertinente n'est pas rencontrée, car la proportion des $y_i = 1$ est de 0,32 donc éloignée de 0,5.

2) La probabilité qu'un tel ménage achète un climatiseur central est :

$$Prob(CENTRAL = 1) = \exp(Z)/[1 + \exp(Z)]$$

où :

$$Z = -12,0471 - 0,0645 \times 3 + 0,00109 \times 1360 + 1,1587 \times 10 - 0,4940 \times 1 - 1,8064 \times 0$$

Cette probabilité est égale à $\exp(0,3348)/[1 + \exp(0,3348)] = 0,5829$.

3) L'effet marginal est donné par $\partial Prob(CENTRAL)/\partial REVENU$ au point moyen. Ce dernier est obtenu par la combinaison linéaire des paramètres du Logit multipliés par les moyennes des variables explicatives associées. Ne pas oublier le terme constant, dont la moyenne est évidemment 1. Ce point moyen est $-1,0831$ ce qui donne l'effet marginal :

$$\partial Prob(CENTRAL)/\partial REVENU = \lambda(\bar{x}\beta)\beta_{REVENU} = 0,2189$$

4) L'effet marginal est dans ce cas :

$$\partial Prob(CENTRAL)/\partial PORTABLE = \lambda(\bar{x}\beta) \times \beta_{PORTABLE} = -0,3413$$

Une mesure de l'impact du revenu sur l'effet marginal de portable est la dérivée croisée de la probabilité par rapport à PORTABLE et REVENU. Cette dérivée est égale à :

$$\begin{aligned} \frac{\partial^2 Prob(CENTRAL)}{\partial PORTABLE \partial REVENU} &= \beta_{PORTABLE} \times \frac{\partial [\partial \lambda(\bar{x}\beta)/\partial PORTABLE]}{\partial REVENU} \\ &= \lambda(\bar{x}\beta) [1 - 2\lambda(\bar{x}\beta)] [1 + \exp(\bar{x}\beta)] \times \beta_{PORTABLE} \times \beta_{REVENU} = -0,1954 \end{aligned}$$

Exercice n° 2

1) La proportion de personnes acceptant l'abonnement est de 0,4667 et celle de femmes dans l'échantillon est 0,6667. Le tableau des effectifs est :

		ABON		Total
		0	1	
Femme	0	4	1	5
	1	4	6	10
Total		8	7	15

La statistique de test du χ^2 à 1 degré de liberté vaut 2,1429 soit une p-value de 0,143. L'hypothèse d'indépendance entre sexe et abonnement ne peut être rejetée.

2) Les résultats de l'estimation Probit sont :

Variable	Paramètre	Écart type	t de Student
Constante	7,1334	5,0251	1,42
FEMME	1,3064	0,9130	1,43
AGE	-0,1728	0,1092	-1,58

Aucune variable n'est significative au seuil de 5 %. La log-vraisemblance du modèle est -6,2795 ; celle du modèle contraint est -10,3638. La statistique de test LR est donc $2(10,3638 - 6,2795) = 8,1686$ soit une p-value de 0,017. Le tableau des effectifs observés et prédits est :

Observé	Prédit		Total
	0	1	
0	6	2	8
1	2	5	7
Total	8	7	15

Le pourcentage de prédictions fausses est donc $100 \times 4/15 = 26,66\%$.

3) Les estimations par MCO sont :

Variable	Paramètre	Écart type	t de Student
Constante	1,9659	0,7548	2,6043
FEMME	0,3476	0,2329	1,4920
AGE	-0,0374	0,0154	-2,4166

Avec ce modèle, la constante et le paramètre associé à la variable AGE sont significatifs, mais il ne faut pas oublier que le modèle est hétéroscléastique par construction. Les écarts types des paramètres estimés ne sont donc pas corrects. Le R^2 est de 0,4230 que l'on compare avec le pseudo- R^2 de McFadden pour le Probit, qui vaut $1 - 6,2795/10,3638 = 0,3940$.

La probabilité d'abonnement est supérieure à 1 si $1,9659 + 0,3476 - 0,0374 \text{AGE} > 1 \Leftrightarrow \text{AGE} < 35,12$. La probabilité est négative si $1,9659 + 0,3476 - 0,0374 \text{AGE} < 0$ soit $\text{AGE} > 67,85$.

4) On cherche tout d'abord la valeur Z telle que $\Phi(Z) = 0,4$. Cette valeur, qui peut être approchée à partir de la table de la distribution normale, est $-0,2534$. On doit avoir $Prob(y_i^* > 0) = \Phi(\gamma_0 + \gamma_1 \overline{AGE} + \gamma_2) = 0,4$ soit :

$$\overline{AGE} = [\Phi^{-1}(0,4) - \gamma_0 - \gamma_2]/\gamma_1 = 50,3079$$

Exercice n° 3

1) Seuls les paramètres (excepté la constante) associés à *AVANT* et *MONTANT* sont significatifs. Plus le montant demandé est important, plus le crédit a de chances d'être accepté ; de même si le ménage a déjà eu affaire avec l'organisme de crédit. L'ajustement global du modèle est donné par le test LR : $2[\log L(\beta) - \log L(\bar{\beta})] = 2 \times (60,2151 - 47,8768) = 24,6766$ avec 5 degrés de liberté. La p-value associée est 0,0001 ; le modèle contraint est donc rejeté. Le pseudo- R^2 est égal à $1 - \log L(\beta)/\log L(\bar{\beta}) = 0,2049$. Le pourcentage de prédictions fausses est, d'après le tableau des effectifs observés et prédits, égal à $(9+18)/100 = 27\%$.

2) Pour tester la validité du modèle A, on compare le modèle complet à un modèle restreint où les variables *AVANT* et *REJET* n'apparaîtraient pas. Sous cette hypothèse, la statistique LR est formée par les log-vraisemblances des modèles « complets » dans les tableaux 1 et 3, et est égale à $2 \times (53,8374 - 47,8768) = 11,9212$ avec 2 degrés de liberté, soit une p-value égale à 0,0025. L'hypothèse nulle de nullité de ces variables du modèle A est rejetée, le modèle A n'est donc pas rejeté. Il convient de se souvenir que le test d'un modèle signifie, dans ce contexte, l'étude de la validité de restrictions par rapport à un modèle complet. Ainsi, avoir comparé le modèle complet du tableau 1 avec le modèle dans lequel ne figurent que les variables *AVANT* et *REJET* aurait signifié en réalité tester du modèle B et non du modèle A. Pour le modèle B, on calcule la statistique de test LR associée à l'hypothèse de nullité des paramètres du modèle B : *REVENU*, *MONTANT*, *LOCAL*. Cette statistique est calculée à partir des log-vraisemblances des modèles « complets » des tableaux 1 et 2, et vaut $2(53,3374 - 47,8768) = 10,9212$ avec 3 degrés de liberté, soit une p-value de 0,0121. Le modèle B n'est pas rejeté.

3) La probabilité recherchée est : $Prob(AVANTAGE = 1 | REVENU = 2,5 ; LOCAL = 1 ; MONTANT = 3) = \exp(Z)/[1 + \exp(Z)]$, avec :
 $Z = -0,9072 + 0,3868 \times 3 + 0,4774 \times 2,5 + 0,4444 \times 1 = 1,8911$ soit une probabilité de 0,8688.

Chapitre 4

Exercice n° 1

1) Les résultats du Logit multinomial sont les suivants, le cas 3 étant celui de référence :

Variable	Comparaison	Paramètre	Écart type	<i>t</i> de Student
Constante	1/3	8,9889	3,9489	2,28
	2/3	5,2524	3,3437	1,57
<i>ENFPLUS2</i>	1/3	-2,5248	1,2205	-2,07
	2/3	0,1218	1,2164	0,10
<i>AGE</i>	1/3	-0,2452	0,1249	-1,96
	2/3	-0,1710	0,1034	-1,66

Log-vraisemblance du modèle complet : -23,9662.

Log-vraisemblance du modèle restreint : -32,5556.

Pourcentage d'observations correctement prédites : 70 %.

2) Le modèle Logit pour $TYPE = 1, 2$ donne les résultats suivants :

Variable	Comparaison	Paramètre	Écart type	t de Student
Constante	1/3	7,3495	3,6881	1,99
ENFPLUS2	1/3	-2,4645	1,2091	-2,04
AGE	1/3	-0,1927	0,1166	-1,65

Log-vraisemblance du modèle complet : -8,8833

Log-vraisemblance du modèle restreint : -15,1582

Pourcentage d'observations correctement prédictes : 86,3636 %.

Les paramètres estimés pour la comparaison 1/3 sont assez proches dans les deux cas. On remarque que le second modèle, binaire simple, réalise de meilleures prévisions.

Exercice n° 2

1) Les écarts types et les t de Student des paramètres sont, pour les deux modèles :

Modèle 1		Modèle 2	
Écart type	t de Student	Écart type	t de Student
0,1271	-13,8037	0,1899	-9,9659
0,0471	-2,7547	0,0900	-2,3877
0,0033	9,3896	0,0074	3,4597
0,0625	-3,0910	0,1311	-1,0670
0,1056	2,5752	0,1973	1,9122

Le modèle complet 1 fournit des paramètres estimés plus efficaces, au vu des écarts types estimés.

2) On calcule la statistique de test d'Hausman, basée sur la comparaison des estimateurs du modèle complet 1 et du modèle restreint aux 3 premières modalités :

$$HT = (\hat{\beta}_1 - \hat{\beta}_2)' [var(\hat{\beta}_2) - var(\hat{\beta}_1)]^{-1} (\hat{\beta}_1 - \hat{\beta}_2)$$

qui suit un χ^2 avec 5 degrés de liberté sous l'hypothèse nulle. On trouve $HT = 20,3338$; la p-value est de 0,001 08. La propriété II A est donc rejetée par ce test.

Exercice n° 3

1) La probabilité d'achat du bien est calculée à partir du modèle Logit binaire, car il n'y a que deux possibilités : « achat » ou « non-achat ». La probabilité est :

$$Prob(Y_i = 1) = \Lambda(\beta_0 + \beta_1 R_i + \beta_2 I_i) = \frac{\exp(\beta_0 + \beta_1 R_i + \beta_2 I_i)}{1 + \exp(\beta_0 + \beta_1 R_i + \beta_2 I_i)}$$

et la log-vraisemblance associée est :

$$\begin{aligned} \log L &= \sum_{i=1}^N \left\{ Y_i \log \left(\frac{\exp(\beta_0 + \beta_1 R_i + \beta_2 I_i)}{1 + \exp(\beta_0 + \beta_1 R_i + \beta_2 I_i)} \right) \right. \\ &\quad \left. + (1 - Y_i) \log \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 R_i + \beta_2 I_i)} \right) \right\} \end{aligned}$$

2) Les statistiques de Student sont, dans l'ordre, -1,4705 ; 6,25 ; 2,25. La constante n'est pas significative, les coefficients de R et I le sont. La log-vraisemblance du modèle

contraint par $\beta_1 = \beta_2 = 0$ s'écrit :

$$\begin{aligned}\log L &= \sum_{i=1}^N \left\{ Y_i \log \left(\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right) + (1 - Y_i) \frac{1}{1 + \exp(\beta_0)} \right\} \\ &= \sum_{i=1}^N \{Y_i \beta_0 - \log[1 + \exp(\beta_0)]\}.\end{aligned}$$

Le test de significativité globale des paramètres (β_1 et β_2) est conduit avec la statistique de test LR : $2[\log L(\beta_0, \beta_1, \beta_2) - \log L(\beta_0)] = 2(3,7762 - 0,5612) = 6,43$; un χ^2 avec 2 degrés de liberté. L'hypothèse de nullité des deux paramètres est donc rejetée, la p-value étant égale à 0,0401.

La probabilité qu'un tel individu achète le bien est :

$$\begin{aligned}Prob(Y_i = 1 | R_i = 2,50 ; I_i = 1) &= \frac{\exp(-0,50 + 0,25 \times 2,50 + 0,36 \times 1)}{1 + \exp(-0,50 + 0,25 \times 2,50 + 0,36 \times 1)} \\ &= \frac{\exp(0,485)}{1 + \exp(0,485)} = 0,6189\end{aligned}$$

3) Comme on ne dispose que de caractéristiques des individus, il faut utiliser le Logit multinomial ; les paramètres dépendront des alternatives : achat du bien A, achat du bien B, ou pas d'achat. Avantage : simplicité d'emploi, mais inconvénient : propriété IIA, qui peut ne pas être vérifiée, par exemple si les modalités « achat » (bien A ou B) sont similaires et séparées de la modalité « non-achat ».

4) La probabilité liée à l'achat du bien A est donnée par :

$$\begin{aligned}Prob(A) &= \frac{\exp(0,21 - 0,12 \times 2,5 + 0,46 \times 1)}{1 + \exp(0,21 - 0,12 \times 2,5 + 0,46 \times 1) + \exp(0,30 - 0,34 \times 2,5 + 0,28 \times 1)} \\ &= \frac{1,4477}{3,2110} = 0,4508\end{aligned}$$

et la probabilité de l'achat de B est :

$$\begin{aligned}Prob(B) &= \frac{\exp(0,30 - 0,34 \times 2,5 + 0,28 \times 1)}{1 + \exp(0,21 - 0,12 \times 2,5 + 0,46 \times 1) + \exp(0,30 - 0,34 \times 2,5 + 0,28 \times 1)} \\ &= \frac{0,7633}{3,2110} = 0,2377\end{aligned}$$

Cette probabilité est très éloignée de celle trouvée en 2) (0,6189). La propriété IIA est vraisemblablement mise en défaut.

Chapitre 5

Exercice n° 1

Pour retrouver la forme classique du modèle Tobit simple, il suffit de poser :

$$y_i^* = y_i - c_i, \quad x_i^* = \begin{pmatrix} x_i \\ c_i \end{pmatrix}, \quad \beta^* = \begin{pmatrix} \beta \\ -1 \end{pmatrix}$$

On retrouve bien le modèle ci-dessous avec :

$$\begin{cases} y_i^* = x_i^* \beta + \varepsilon_i & \text{si } y_i^* > 0 \\ y_i^* = 0 & \text{sinon.} \end{cases}$$

Exercice n° 2

Comme ΔP_t est observé, le mécanisme de sélection est observable. Notons μ_t la probabilité que la quantité échangée soit du côté de la demande :

$$\mu_t = \text{Prob}(D_t < O_t) = \text{Prob}(\varepsilon_{1t} - \varepsilon_{2t} < x_{2t}\beta_2 - x_{1t}\beta_1 + P_t(\alpha_2 - \alpha_1))$$

Soit g_{1t} et g_{2t} les densités conditionnelles associées à la demande et à l'offre respectivement :

$$g_{1t} = \int_{O_t}^{+\infty} g(Q_t, O_t) dO_t / \mu_t \quad g_{2t} = \int_{Q_t}^{+\infty} g(D_t, Q_t) dD_t / (1 - \mu_t)$$

où $g(\dots)$ est la densité conjointe de la demande et de l'offre. La vraisemblance s'écrit :

$$L = \prod_{\Delta P_t < 0} g_{1t} \times \mu_t \prod_{\Delta P_t > 0} g_{2t} \times (1 - \mu_t)$$

ou encore :

$$L = \prod_{t=1}^T [g_{1t} \times \mu_t]^{\gamma_t} [g_{2t} \times (1 - \mu_t)]^{1 - \gamma_t}$$

où $y_t = 1$ si $\Delta P_t < 0$, $y_t = 0$ sinon.

Exercice n° 3

Comme $\text{Prob}(y \leq 2,5) = \Phi[(2,5 - \mu)/\sigma] = 0,60$ on a $(2,5 - \mu)/\sigma = 0,2533$ soit $2,5 - \mu = 0,2533\sigma$. De plus, comme $E(y > 2,5) = 3,80$ on a :

$$\mu + \sigma\phi[(2,5 - \mu)/\sigma]/[1 - \Phi((2,5 - \mu)/\sigma)] = 3,80$$

soit encore $3,80 = \mu + \sigma\phi[(2,5 - \mu)/\sigma]/0,40$. La densité de la loi normale $N(0, 1)$ en $(2,5 - \mu)/\sigma = 0,2533$ vaut $\exp(-0,2533^2/2)/\sqrt{2\pi} = 0,3863$. Par conséquent on déduit que $3,80 = \mu + 0,3863\sigma/0,40$. On a donc : $3,80 - 2,5 = 0,7124\sigma \Leftrightarrow \sigma = 1,8248$ et $\mu = 2,0377$.

Exercice n° 4

1) Dans le cas où $C_i \leq \bar{C}_i$, on a :

$$\text{Prob}(C_i \leq \bar{C}_i) = \text{Prob}(\alpha_0 + \alpha_1 P_i + \varepsilon_i \leq \beta_0 + \beta_1 A_i + \nu_i) = \Phi\left(\frac{\beta_0 - \alpha_0 + \beta_1 A_i - \alpha_1 P_i}{\sigma}\right)$$

où $\sigma = [\text{var}(\varepsilon_i - \nu_i)]^{1/2} = [\sigma_\varepsilon^2 + \sigma_\nu^2 - 2\sigma_{\varepsilon\nu}]^{1/2}$. Pour $C_i > \bar{C}_i$, la consommation est observable et on reporte la densité de C_i :

$$f(C_i) = \frac{1}{\sigma} \phi\left(\frac{C_i - \alpha_0 - \alpha_1 P_i}{\sigma}\right)$$

2) La log-vraisemblance du modèle censuré est de la forme étudiée dans le texte :

$$\log L = \sum_{C_i > 0}^{N_1} \log \left[\frac{1}{\sigma_\epsilon} \phi \left(\frac{C_i - \alpha_0 - \alpha_1 P_i}{\sigma_\epsilon} \right) \right] + \sum_{C_i=0}^{N_0} \log \left[\Phi \left(\frac{\beta_0 - \alpha_0 + \beta_1 A_i - \alpha_1 P_i}{\sigma} \right) \right]$$

3) On a maintenant affaire à un modèle censuré et tronqué. On retranche alors à la log-vraisemblance $\log L$ ci-dessus le terme de troncature correspondant à la probabilité que C_i soit inférieure à C^* :

$$\log L^* = \log L - \sum_{i=1}^N \log \left[\Phi \left(\frac{C^* - \alpha_0 - \alpha_1 P_i}{\sigma_\epsilon} \right) \right]$$

Exercice n° 5

1) Les estimations du modèle Tobit simple sont les suivantes, avec le logiciel SAS :

Dependent Variable=LOWER

Dependent Variable=Y

Noncensored Values= 11 Right Censored Values= 0

Left Censored Values= 5 Interval Censored Values= 0

Log Likelihood for NORMAL -37.04239028

Lifereg Procedure

Variable	DF	Estimate	Std. Err	ChiSquare	Pr>Chi	Label/Value
INTERCEPT	1	-5.4307419	3.333769	2.653672	0.1033	Intercept
X1	1	3.24956682	1.552418	4.381608	0.0363	
X2	1	2.94309377	1.193789	6.077896	0.0137	
SCALE	1	4.74258209	1.105041			Normal scale parameter

Les statistiques de Student s'obtiennent en divisant les paramètres par leurs écarts types estimés : -1,6290 ; 2,0932 ; 2,4653 ; 4,2917.

2) Les estimations par MCO du modèle dans lequel les observations nulles de y ont été éliminées sont, toujours avec SAS :

Dependent Variable : Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0 : Parameter=0	Prob > T
INTERCEP	1	4.884334	1.49260211	3.272	0.0113
X1	1	1.589231	0.59243771	2.683	0.0278
X2	1	1.233823	0.45931981	2.686	0.0277

Les résultats d'estimation sont très différents de 1) : la constante devient significative au seuil de 5 %, les paramètres associés à X1 et X2 sont divisés de moitié.

Bibliographie

- AMEMIYA T., « Regression Analysis when the Dependent Variable is truncated Normal », *Econometrica*, n° 41, pp. 997-1016, 1973.
- AMEMIYA T., « A Note on the Fair and Jaffee Model », *Econometrica*, n° 42, pp. 759-762, 1974.
- AMEMIYA T., « The Maximum Likelihood, the Minimum Chi Square and the Non Linear Weighted Least Squares Estimator in the General Qualitative Response Model », *Journal of the American Statistical Association*, n° 71, pp. 347-351, 1976.
- AMEMIYA T., « Qualitative Response Models : A Survey », *Journal of Economic Literature*, n° 19(4), pp. 481-536, 1981.
- AMEMIYA T., « Tobit Models : A Survey », *Journal of Econometrics*, n° 24, pp. 3-63, 1984.
- AMEMIYA T., *Advanced Econometrics*, Cambridge, Harvard University Press, 1985.
- ANDERSON S.P., DE PALMA A., THISSE J.F., *Discrete Choice of Product Differentiation*, Cambridge, MIT Press, 1992.
- ASHFORD J., SOWDEN R., « Multivariate Probit Analysis », *Biometrics*, n° 26, pp. 535-546, 1970.
- BALTAGI B.H., *Econometric Analysis of Panel Data*, New York, John Wiley and Sons, 1995.
- BEN AKIVA M., LERMAN S., « Disaggregate Travel and Mobility Choice Modes and Measures of Accessibility », in *Behavioral Travel Modeling*, Hensher et Stopher (ed.), Londres, Coom Helm, 1979.
- BERA A., JARQUE C., LEE L.F., « Testing the Normality Assumption in Limited Dependent Variable Models », *International Economic Review*, n° 25, pp. 563-578, 1984.
- BERKSON J., « A statistically Precise and relatively Simple Method of Estimating the Bio-Assay with Quantal Response, based on the Logistic Function », *Journal of the American Statistical Association*, n° 48, pp. 565-599, 1953.

- BERNDT E.R., HALL B.H., HALL R.E., HAUSMAN J.A., « Estimation and Inference in Nonlinear Structural Models », *Annals of Economic and Social Measurement*, n° 3, pp. 653-665, 1974.
- BLUNDELL R. (ed.), « Specification Testing in Limited and Discrete Dependent Variable Models », *Journal of Econometrics*, n° 34(1/2), pp. 1-274, 1987.
- BREUSCH T.S., PAGAN A.R., « The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics », *Review of Economic Studies*, n° 47, pp. 239-253, 1980.
- BUTLER J.S., MOFFITT R., « A computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model », *Econometrica*, n° 50, pp. 761-764, 1982.
- CAMERON A., TRIVEDI P., « Econometric Models based on Count Data : Comparisons and Applications of some Estimators and Tests », *Journal of Applied Econometrics*, n° 1, pp. 29-54, 1986.
- CHAMBERLAIN G., « Analysis of Covariance with Qualitative Data », *Review of Economic Studies*, n° 47, pp. 225-238, 1980.
- CHESHER A., LANCASTER T., IRISH M., « On detecting the Failure of Distributional Assumptions », *Annales de l'INSEE*, n° 59/60, pp. 7-44, 1985.
- CHESHER A., IRISH M., « Residual Analysis in the Grouped Data and Censored Normal Linear Model », *Journal of Econometrics*, n° 34, pp. 33-62, 1987.
- CHEUNG C., GOLDBERGER A., « Proportional Projections in Limited Dependent Variable Models », *Econometrica*, n° 52, pp. 531-534, 1984.
- CRAGG J., « Some Statistical Models for Limited Dependent Variables with an Application to the Demand for Durable Goods », *Econometrica*, n° 39, pp. 829-844, 1971.
- DAHM P.A., GURLAND J., LEE I., « Polychotomous Quantal Response in Biological Assay », *Biometrics*, n° 16, pp. 382-398, 1960.
- DAVIDSON R., MACKINNON J.G., « Convenient Specification Tests for Logit and Probit Models », *Journal of Econometrics*, n° 25, pp. 241-262, 1984.
- DAVIDSON R., MACKINNON J.G., *Estimation and Inference in Econometrics*, New York, Oxford University Press, 1993.
- DEACON R., SHAPIRO P., « Private Preference for Collective Goods revealed through Voting and Referenda », *American Economic Review*, n° 65, pp. 943-955, 1975.
- DUNCAN G., « Sample Selectivity as a Proxy Variable Problem : on the Use and Misuse of Gaussian Selectivity Corrections », *Research in Labor Economics*, Supplément n° 2, pp. 333-345, 1983.
- FAIR R., « A Note on the Computation of the Tobit Estimator », *Econometrica*, n° 45, pp. 1723-1727, 1977.

- FAIR R.C., JAFFEE D.M., « Methods of Estimation for Markets in Disequilibrium », *Econometrica*, n° 40, pp. 497-514, 1972.
- FAIR R.C., KELEJIAN H.H., « Methods of Estimation for Markets in Disequilibrium : A Further Study », *Econometrica*, n° 42, pp. 177-190, 1974.
- FERNANDEZ L., « Nonparametric Maximum Likelihood Estimation of Censored Regression Models », *Journal of Econometrics*, n° 32, pp. 35-57, 1986.
- FERRIER G., GOFFE W., RODGERS J., « Global Optimization of Statistical Functions with Simulated Annealing », *Journal of Econometrics*, n° 60, pp. 65-100, 1994.
- FIN T., SCHMIDT P., « A Test of the Tobit Specification against an Alternative Suggested by Cragg », *Review of Economics and Statistics*, n° 66, pp. 174-177, 1984.
- FLANNERY B., PRESS W., TEUKOLSKY S., VETTERLING W., *Numerical Recipes : the Art of Scientific Computing*, Cambridge, Cambridge University Press, 1986.
- FLETCHER R., *Practical Methods of Optimization*, New York, John Wiley and Sons, 1980.
- GEWEKE J., KEANE M., RUNKLE D., « Alternative Computational Approaches to Inference in the Multinomial Probit Model », *Review of Economics and Statistics*, n° 76, pp. 609-632, 1994.
- GOLDBERGER A.S., *Econometric Theory*, New York, John Wiley and Sons, 1964.
- GOLDBERGER A.S., « Linear Regression after Selection », *Journal of Econometrics*, n° 15, pp. 357-366, 1981.
- GOLDFELD S.M., QUANDT R.E., « Estimation in a Disequilibrium Model and the Value of Information », *Journal of Econometrics*, n° 3, pp. 325-348, 1975.
- GOODMAN L., « Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-linear Models Approach in the Analysis of Contingency Tables », *International Economic Review*, n° 54, pp. 243-309, 1986.
- GOODMAN L., KRUSKAL W., *Measures of Association for Cross-classifications*, New York, Springer Verlag, 1979.
- GORIÉROUX C., *Économétrie des variables qualitatives*, Paris, Economica, 1989.
- GORIÉROUX C., MONFORT A., « Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logit Models », *Journal of Econometrics*, n° 17, pp. 83-97, 1981.
- GREENE W.H., « On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model », *Econometrica*, n° 49, pp. 505-513, 1981.

- GREENE W.H., « Estimation of Limited Dependent Variable Models by Ordinary Least Squares and the Method of Moments », *Journal of Econometrics*, n° 21, pp. 195-212, 1983.
- GREENE W.H., *Econometric Analysis*, Londres, Prentice Hall, 1997.
- GURMU S., « Tests for detecting Overdispersion in the Positive Poisson Regression Model », *Journal of Business and Economic Statistics*, n° 9, pp. 215-222, 1991.
- HAJIVASSILIOU V., MCFADDEN D., RUUD P., « Simulation of Multivariate Normal Orthant Probabilities : Methods and Programs », Document de travail 1021, Yale University, Cowles Foundation, 1992.
- HAJIVASSILIOU V., RUUD P.A., « Classical Estimation Methods for LDV Models using Simulation », *Handbook of Econometrics IV*, Amsterdam, North Holland, 1994.
- HANSEN L., « Large Sample Properties of Generalized Method of Moments Estimators », *Econometrica*, n° 50, pp. 1029-1054, 1982.
- HARVEY A., « Estimating Regression Models with Multiplicative Heteroscedasticity », *Econometrica*, n° 44, pp. 461-465, 1976.
- HAUSMAN J., « Specification Tests in Econometrics », *Econometrica*, n° 46, pp. 1251-1271, 1978.
- HAUSMAN J., WISE D., « A Conditional Probit Model for Qualitative Choice : Discrete Decisions, Recognizing Interdependence and Heterogenous Preferences », *Econometrica*, n° 46, pp. 403-426, 1978.
- HAUSMAN J., WISE D., « Discontinuous Budget Constraints and Estimation : the Demand for Housing », *Review of Economic Studies*, n° 67, pp. 75-96, 1980.
- HAUSMAN J., HALL B., GRILICHES Z., « Economic Models for Count Data with an Application to the Patents-R&D Relationship », *Econometrica*, n° 52, pp. 909-938, 1984.
- HAUSMAN J., MCFADDEN D., « A Specification Test for the Multinomial Logit Model », *Econometrica*, n° 52, pp. 1219-1240, 1984.
- HECKMAN J., « The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models », *Annals of Economic and Social Measurement*, n° 5, pp. 475-492, 1976.
- HECKMAN J., « Sample Selection Bias as a Specification Error », *Econometrica*, n° 47, pp. 153-161, 1979.
- JANSSEN J., MARCOTORCHINO J.F., PROTH J.M., *Data Analysis in Real Life Environment*, Amsterdam, North Holland, 1985.
- KALBFLEISCH J., PRENTICE R., *The statistical analysis of failure time data*, Londres, John Wiley and Sons, 1980.

- KIEFER N., « Testing for Independence in Multivariate Probit Models », *Biometrika*, n° 69, pp. 161-166, 1982.
- KLEIN R.W., SPADY R.H., « An Efficient Semiparametric Estimator for Binary Response Models », *Econometrica*, n° 61, pp. 387-421, 1993.
- LANCASTER T., *The Econometric Analysis of Transition Data*, Econometric Society Monographs, N° 17, Cambridge, Cambridge University Press, 1990.
- LEE L., « Some Approaches to the Correction of Selectivity Bias », *Review of Economic Studies*, n° 49, pp. 355-372, 1982.
- LEE M.J., *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, New York, Springer Verlag, 1996.
- LUCE R., *Individual Choice Behavior*, New York, John Wiley and Sons, 1959.
- MADDALA G.S., NELSON F., « Maximum Likelihood Methods for Models of Markets in Disequilibrium », *Econometrica*, n° 42, pp. 1013-1030, 1974.
- MADDALA G.S., *Limited-dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, N° 3, Cambridge, Cambridge University Press, 1983.
- MANSKI C., « The Maximum Score Estimator of the Stochastic Utility Model of Choice », *Journal of Econometrics*, n° 3, pp. 205-228, 1975.
- MANSKI C., LERMAN F., « The Estimation of Choice Probabilities from Choice-based Samples », *Econometrica*, n° 45, pp. 1977-1988, 1977.
- MANSKI C., « Semiparametric Analysis of Discrete Response : Asymptotic Properties of the Maximum Score Estimator », *Journal of Econometrics*, n° 27, pp. 313-333, 1985.
- MCCULLAGH P., NELDER J., *Generalized Linear Models*, Londres, Chapman and Hall, 1983.
- MCFADDEN D., « Conditional Logit Analysis of Qualitative Choice Behavior », in (ed.) ZAREMBKA P., *Frontiers in Econometrics*, New York, Academic Press, 1973.
- MCFADDEN D., « Modeling the Choice of Residential Location », in (ed.) KARLQVIST A. et al., *Spatial Interaction Theory and Planning Models*, Amsterdam, North Holland, 1978.
- MCFADDEN D., « Econometric Models of Probabilistic Choice », in (ed.) MANSKI C., MCFADDEN D., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MIT Press, 1981.
- MCFADDEN D., « A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration », *Econometrica*, n° 57, pp. 995-1026, 1989.

- MEAD R., NELDER J., « A Simplex Method for Function Minimization », *Computer Journal*, n° 7, pp. 308-313, 1965.
- MELENBERG B., VAN SOEST A., « Parametric and Semiparametric Modeling of Vacation Expenditures », *Journal of Applied Econometrics*, n° 11, pp. 59-76, 1996.
- MELINO A., « Testing for Sample Selection Bias », *Review of Economic Studies*, n° 49, pp. 151-153, 1982.
- MORIMUNE K., « Comparisons of Normal and Logistic Models in the Bivariate Dichotomous Analysis », *Econometrica*, n° 47, pp. 957-975, 1979.
- NELSON F.D., « Censored Regression Models with Unobserved Stochastic Censoring Thresholds », *Journal of Econometrics*, n° 6, pp. 309-327, 1977.
- OLSEN R.J., « A Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model », *Econometrica*, n° 46, pp. 1211-1215, 1978.
- OLSEN R.J., « A Least Squares Correction for Selectivity Bias », *Econometrica*, n° 48, pp. 1815-1820, 1980.
- POWELL J., « Least Absolute Deviations Estimation for the Censored Regression Model », *Journal of Econometrics*, n° 25, pp. 303-325, 1984.
- ROSETT R., « A Statistical Model of Friction in Economics », *Econometrica*, n° 27, pp. 263-267, 1959.
- ROSETT R., NELSON F., « Estimation of the Two Limit Probit Regression Model », *Econometrica*, n° 43, pp. 141-146, 1975.
- RUUD P., « Sufficient Conditions for the Consistency of Maximum Likelihood Estimation despite Misspecification of Distribution in Multinomial Discrete Choice Models », *Econometrica*, n° 51, pp. 225-228, 1983.
- RUUD P., « Tests of Specification in Econometrics », *Econometric Reviews*, n° 3, pp. 211-242, 1984.
- SAPORTA G., *Probabilités, analyse des données et statistique*, Paris, Technip, 1990.
- STERN S., « A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models », *Econometrica*, n° 60, pp. 943-952, 1992.
- THURSTONE L., « A Law of Comparative Judgement », *Psychological Review*, n° 34, pp. 273-286, 1927.
- TOBIN J., « Estimation of Relationships for Limited Dependent Variables », *Econometrica*, n° 26, pp. 24-36, 1958.
- WALES T., WOODLAND A., « Sample Selectivity and the Estimation of Labor Supply Functions », *International Economic Review*, n° 21, pp. 437-468, 1980.

WHITE H., « A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity », *Econometrica*, n° 48, pp. 817-838, 1980.

WHITE H., « Maximum Likelihood Estimation of Misspecified Models », *Econometrica*, n° 50, pp. 1-26, 1982.



Index

C

- χ^2 minimum 19
- méthode de Berkson 21
- Coefficients de Pearson et de Cramer 10
- Corrélation
- variable polytomique et continue 5
- Critères d'information 32

D

- Données groupées 19

E

- Échantillon censuré ou tronqué 124
- Étude graphique des résidus 63

G

- GMM
 - critère 40
 - et variable instrumentale 41
 - hétérosclélasticité 42
 - normalité asymptotique 42
 - test d'Hansen 44
 - tests d'hypothèses 45
- Gompit 55

H

- Hétérosclélasticité
 - correction de White 43
 - dans le Probit binaire 65
 - dans le Tobit 131

L

- Logit
 - effets fixes 77
 - Logit binaire
 - définition 57
 - effets marginaux 61
 - estimation 57
 - généralisation au cas multinomial 91
 - Logit conditionnel 94
 - effets marginaux 95
 - interprétation des paramètres 95
 - Logit multinomial
 - effets marginaux 93
 - et Logit conditionnel, comparaison 96
 - formulation 96
 - interprétation des paramètres 93
 - significativité globale des paramètres 94
 - vraisemblance 93
 - Logit multinomial emboîté 107
 - Logit multinomial hiérarchisé 108
 - estimation 110
 - généralisation 111
 - Logit multinomial séquentiel 108
- Loi des valeurs extrêmes
 - type I 101
 - type II 108

M

- Maximum de vraisemblance
 - propriétés de l'estimateur 29

- test de White 31
- tests d'hypothèses 32
- variance de l'estimateur 30
- Mills ratio 126
- Modèle
 - à friction 139
 - à régimes 149
 - régime inobservable 153
 - régime observable 149
 - à seuils stochastiques 142
 - à utilité aléatoire 100
 - de comptage 35
 - test de spécification 37
 - de déséquilibre 153
 - de Rosett 139
 - log-linéaire
 - effets 13
 - estimation 15
 - saturé 14
 - LP 22
 - multinomial
 - formulation 84
 - vraisemblance 85
 - NegBin 37
 - ordonné 87
 - Poisson 35
 - séquentiel 90
- Moindres carrés
 - doubles 40
 - généralisés 43
 - indirects 39
 - pondérés 43
- N
 - Nested Logit 107
 - Nombre de prédictions fausses 63
- O
 - Optimisation numérique
 - DFP, BFGS 50
 - méthodes de gradient 49
 - principe 48
- P
 - Probit
 - effets aléatoires 76
 - Probit binaire
 - effets marginaux 61
 - estimation 56
 - et Logit, comparaison 59
 - Probit bivarié 70
 - effets marginaux 71
 - test LM de corrélation 71
 - Probit multinomial
 - estimateur GHK 116
 - estimateur MSM 114
 - formulation 113
 - Propriété IIA 101
 - et indépendance des erreurs 102
 - test d'Hausman 103
 - Pseudo R^2 62
- R
 - Résidus généralisés 63
- S
 - Score maximum 73
- T
 - τ_b de Goodman et Kruskal 10
 - Tableau de contingence
 - coefficients de contingence 9
 - indépendance des modalités 9
 - test du χ^2 10
 - Test
 - ajustement global 61
 - Hausman 44
 - LM 33
 - LR 32
 - interprétation 34
 - Wald 33
 - Tobit
 - double censure 138
 - effets marginaux 131
 - égalité des paramètres 134
 - estimation en deux étapes 130

- estimation par maximum de vraisemblance 128
 - estimation sur données complètes 131
 - généralisé 140
 - hétéroscédasticité 131, 132
 - normalité des résidus 133
 - simple 126
 - vraisemblance 127
- V**
- Variable
 - catégorielle 3
 - censurée, espérance 125
 - d'inclusion 110
 - dichotomique 2
 - espérance 2
 - indicatrice 4
 - instrumentale 39
 - latente 52
 - multinomiale 2
 - normale
 - loi conditionnelle 141
 - troncature auxiliaire 145
 - omise 64
 - tronquée, espérance 126
- Vraisemblance
- concentrée 30
 - critères d'information 32
 - définition 28



044665-(I)-(1,5) OSB 80°-PUB

STEDI, 1, boulevard Ney, 75018 Paris - Tél. 01.40.38.65.40

Dépôt légal, Imprimeur, n° 6437

Dépôt légal : avril 2000

Imprimé en France



Alban Thomas

ÉCONOMÉTRIE DES VARIABLES QUALITATIVES

MANUEL
ET EXERCICES
CORRIGÉS

Réputés complexes, les modèles à variables qualitatives sont en réalité de plus en plus utilisés parmi l'éventail des outils d'inférence statistique. Leurs applications se révèlent fort diverses, des études de marketing aux bilans commerciaux, en passant par le marché du travail.

Dès lors, cet ouvrage offre au lecteur une classification des variables qualitatives, avant d'en présenter la plupart des utilisations (modèles pour variables binaires, modèles multinomiaux, de régression...).

Il propose en outre :

- de nombreuses applications, sur données réelles, illustrant de manière claire les méthodes exposées et leur mise en œuvre avec des logiciels usuels (SAS, LIMDEP, GAUSS...);
- des échantillons que le lecteur peut utiliser pour vérifier les résultats et développer de nouveaux modèles à sa convenance;
- des exercices corrigés permettant de consolider les acquis au fur et à mesure de la lecture.

ALBAN THOMAS
est directeur de recherche à l'INRA et enseigne à l'université des sciences sociales de Toulouse.

► Deuxième et troisième cycles de sciences économiques et de gestion

► IUP



9 782100 046652

ISBN: 2 10 004665 9
Code 044665

BIBLIOTHÈQUE NATIONALE DE FRANCE



37502 04152148 5

<http://www.dunod.com>



Participant d'une démarche de transmission de fictions ou de savoirs rendus difficiles d'accès par le temps, cette édition numérique redonne vie à une œuvre existant jusqu'alors uniquement sur un support imprimé, conformément à la loi n° 2012-287 du 1^{er} mars 2012 relative à l'exploitation des Livres Indisponibles du XX^e siècle.

Cette édition numérique a été réalisée à partir d'un support physique parfois ancien conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal. Elle peut donc reproduire, au-delà du texte lui-même, des éléments propres à l'exemplaire qui a servi à la numérisation.

Cette édition numérique a été fabriquée par la société FeniXX au format PDF.

La couverture reproduit celle du livre original conservé au sein des collections de la Bibliothèque nationale de France, notamment au titre du dépôt légal.

*

La société FeniXX diffuse cette édition numérique en accord avec l'éditeur du livre original, qui dispose d'une licence exclusive confiée par la Sofia – Société Française des Intérêts des Auteurs de l'Écrit – dans le cadre de la loi n° 2012-287 du 1^{er} mars 2012.

Avec le soutien du

