

# Estimation des modèles probit polytomiques : un survol des techniques

## The Estimation of Multinomial Probit Models: A Survey of the Techniques

Denis Bolduc and Mustapha Kaci

Volume 69, Number 3, septembre 1993

Symposium sur l'économie des transports

URI: <https://id.erudit.org/iderudit/602113ar>

DOI: <https://doi.org/10.7202/602113ar>

[See table of contents](#)

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)

1710-3991 (digital)

[Explore this journal](#)

Cite this article

Bolduc, D. & Kaci, M. (1993). Estimation des modèles probit polytomiques : un survol des techniques. *L'Actualité économique*, 69(3), 161–191.  
<https://doi.org/10.7202/602113ar>

Article abstract

The Multinomial Probit (MNP) model provides the most general framework to allow for interdependent alternatives in discrete choice analysis. The primary impediment to this methodology is related to the dimensionality of the response probabilities which are multifold normal integrals of about the size of the choice set. During the last two decades, numerous researches have been devoted to develop practical methodologies to replace these hard to compute choice probabilities in the estimation process. The main objective of this paper is to survey the major and the most important of these techniques.

## ESTIMATION DES MODÈLES PROBIT POLYTOMIQUES : UN SURVOL DES TECHNIQUES\*

Denis BOLDUC  
Mustapha KACI  
*Département d'économie  
Université Laval*

RÉSUMÉ — Parce qu'il admet des structures très générales d'interdépendance entre les modalités, le probit polytomique (MNP) fournit une des formes les plus intéressantes pour modéliser les choix discrets qui découlent d'une maximisation d'utilité aléatoire. L'obstacle majeur et bien connu dans l'estimation de ce type de modèle tient à la complexité que prennent les calculs lorsque le nombre de modalités considérées est élevé. Cette situation est due essentiellement à la présence d'intégrales normales multidimensionnelles qui définissent les probabilités de sélection. Au cours des deux dernières décennies, de nombreux efforts ont été effectués visant à produire des méthodes qui permettent de contourner les difficultés de calcul liées à l'estimation des modèles probit polytomiques. L'objectif de ce texte consiste à produire un survol critique des principales méthodes mises de l'avant jusqu'à maintenant pour rendre opérationnel le cadre MNP. Nous espérons qu'il éclairera les praticiens de ces modèles quant au choix de technique d'estimation à favoriser au cours des prochaines années.

ABSTRACT — *The Estimation of Multinomial Probit Models : A Survey of the Techniques.* The Multinomial Probit (MNP) model provides the most general framework to allow for inter-dependent alternatives in discrete choice analysis. The primary impediment to this methodology is related to the dimensionality of the response probabilities which are multifold normal integrals of about the size of the choice set. During the last two decades, numerous researches have been devoted to develop practical methodologies to replace these hard to compute choice probabilities in the estimation process. The main objective of this paper is to survey the major and the most important of these techniques.

---

\* Article présenté dans le cadre du 31<sup>e</sup> Congrès annuel de la Société canadienne de science économique, du 15 au 17 mai, Ste-Foy, Québec, 1991. Nous tenons à remercier Pierre Fortin ainsi que les arbitres anonymes pour leurs commentaires judicieux. Le premier auteur désire remercier le Fonds FCAR nouveaux chercheurs ainsi que le Conseil de recherche en sciences humaines du Canada pour leur support financier.

## INTRODUCTION

Les modèles à choix discrets sont utilisés pour analyser les situations dans lesquelles un individu doit faire un choix parmi un ensemble fini et exhaustif de modalités mutuellement exclusives. Ces modèles calculent la probabilité qu'un individu sélectionne une modalité particulière parmi cet ensemble, étant donné les observations. Ils diffèrent entre eux selon la forme fonctionnelle employée pour lier les observations à la probabilité. La forme fonctionnelle est habituellement déduite d'un cadre de maximisation d'utilité aléatoire (RUM), lequel fournit un critère fondamental pour une analyse opérationnelle des problèmes de choix.

Deux grandes classes de modèles sont généralement retenues. La première est dite classe à valeurs extrêmes généralisées (GEV) dont la principale caractéristique est que les probabilités de choix ont une forme explicite ne comportant pas d'intégrales. Le membre le plus utilisé de cette classe est le logit polytomique (MNL) qui doit sa grande popularité à la simplicité des calculs rattachés à son utilisation. En revanche, cette simplicité a toutefois un coût, celui causé par une perte de généralité due à des hypothèses restrictives qui imposent l'indépendance entre les modalités. Un autre membre de la classe GEV qui, pour sa part, admet certaines formes d'interdépendances, est le logit emboîté (NMNL). Toutefois, les formes d'interdépendances admises sont souvent trop restrictives. Dans les modèles à choix discrets, le fait d'ignorer ou de mal spécifier la structure d'interdépendance entre les modalités conduit au problème d'estimation non convergente des paramètres. (voir Horowitz, 1980 pour plus de détails). L'utilisation du MNL ou du NMNL peut donc être non pertinente dans des applications où les modalités risquent d'être fortement interdépendantes. En outre, si l'on désire utiliser un modèle de choix à des fins de prévision, il sera important de déterminer la structure d'erreurs qui donne le meilleur ajustement aux données, plutôt que de supposer faussement une structure particulière.

Dans la seconde classe, on retrouve le modèle probit polytomique (MNP) lequel permet un degré considérable de flexibilité concernant la prise en considération des interdépendances entre les modalités. L'hypothèse faite par le MNP est que ces interdépendances originent de la structure de corrélation entre les erreurs spécifiques aux modalités. Pour plusieurs applications des modèles de choix polytomiques, la possibilité d'admettre des modalités interdépendantes est une nécessité. Par exemple, en économie des transports, on peut considérer un modèle simultané qui explique à la fois le choix de localisation de résidence, de lieu de travail et du mode de transport à favoriser pour les déplacements au travail. Une source supplémentaire d'interdépendance entre les modalités peut aussi provenir du phénomène d'hétérogénéité entre individus. Divers aspects reliés à la formulation et à l'utilisation des modèles probit polytomiques sont discutés abondamment dans Daganzo (1979). La principale difficulté liée à l'utilisation du modèle MNP vient du calcul des probabilités de choix qui sont exprimées sous forme d'intégrales normales multidimensionnelles. Lorsque le

nombre de modalités est supérieur à 5, les calculs deviennent trop exigeants en temps CPU, même sur des ordinateurs super-performants qui permettent le traitement vectoriel et/ou parallèle.

Au cours des deux dernières décennies, de nombreux efforts ont été déployés afin de produire des méthodes permettant l'utilisation du cadre probit polytomique même lorsque le nombre de modalités est élevé. Certaines des recherches ont abouti à la formulation de méthodes pratiques et assez précises d'approximation et de simulation pour remplacer les probabilités de choix. Un autre courant de recherche essaie de mettre au point des spécifications qui conservent la souplesse du modèle MNP conventionnel mais qui conduisent à d'importantes simplifications de calcul. Ces diverses méthodologies sont nombreuses et certaines possèdent des propriétés statistiques plus intéressantes que d'autres. Notre objectif est de rassembler dans un même document et d'analyser de façon critique les récents développements concernant l'estimation pratique des modèles probit polytomiques basés sur un nombre élevé de modalités. En s'efforçant de conserver une notation simple et en évitant d'entrer trop dans les détails pour ne pas compliquer la discussion, le présent texte veut éclairer le praticien de ces modèles dans la sélection des techniques d'estimation à favoriser dans les prochaines années.

L'exposé comprend deux grandes sections. Dans la section 1, on présente divers aspects reliés à la formulation et à l'estimation du modèle probit polytomique conventionnel. Les méthodes d'estimation considérées sont dans l'ordre : la méthode du maximum de vraisemblance, la méthode des moments généralisés et la méthode du score. La discussion permettra de mettre en lumière les problèmes majeurs liés à l'estimation des modèles MNP. La section 2 traite des diverses solutions apportées à ces problèmes. Les techniques de calculs pour les probabilités de choix sont classées dans les catégories suivantes : 1) Intégration numérique (Hausman et Wise, 1978 ; Owen, 1956). Il est bien connu que cette technique bien qu'étant la plus précise n'est applicable de façon pratique qu'avec moins de 5 modalités. 2) Approximation numérique (Kamakura, 1988 ; Langdon, 1984 ; Daganzo, 1979 ; Clark, 1961). Ces méthodologies introduisent généralement des biais dont l'importance dépend de la structure d'interdépendance présente entre les modalités. De plus, les biais sont quelquefois voués à ne pas disparaître même lorsqu'on accroît la taille de l'échantillon. 3) Simulateurs stochastiques (Börsch-Supan et Hajivassiliou, 1990 ; Hajivassiliou, McFadden et Ruud, 1990 ; McFadden, 1989 ; Stern, 1987 ; Albright, Lerman et Manski, 1977). Cette classe comprend : le simulateur de fréquence d'Albright, Lerman et Manski (1977) dont le principal inconvénient est la discontinuité par rapport aux paramètres du modèle. Les autres simulateurs considérés sont lisses, dû à leur continuité par rapport aux paramètres. La propriété de continuité n'est pas essentielle du point de vue de l'induction statistique. Toutefois, elle simplifie de beaucoup la recherche itérative des estimateurs qui maximisent la fonction objectif retenue. Finalement, la dernière section donne un bref résumé des principaux travaux reliés à l'estimation des modèles MNP tout en concluant l'exposé.

## 1. LE MODÈLE PROBIT POLYTOMIQUE

## 1.1 Formulation

Dans le modèle probit polytomique (MNP), les probabilités de choix entre modalités prennent la forme d'intégrales normales multidimensionnelles de dimension égale au nombre de modalités  $J$ . Après modification et transformation, on peut ramener le problème à des intégrales de dimension  $J - 2$ . Dès que ,  $J \geq 5$  cela crée de sérieux problèmes lors de l'estimation des paramètres par des méthodes conventionnelles d'estimation.

Pour bien saisir la portée du problème, considérons le modèle statique suivant où un individu  $n, n = 1, \dots, N$  choisit la modalité  $i, i = 1, \dots, J$  lui procurant la plus grande utilité.  $J$  représente le nombre total de modalités dans l'ensemble de sélection  $C = \{1, \dots, J\}$ , où l'on suppose pour simplifier la notation, que chaque individu  $n$  fait face à un même ensemble  $C$ . On appelle alors modèle probit polytomique le modèle de la forme :

$$y_{in} = \begin{cases} 1 & \text{si } U_{in} \geq U_{jn} \text{ pour } j = 1, \dots, J \\ 0 & \text{sinon,} \end{cases}$$

et

$$U_{in} = X_{in}\beta + \varepsilon_{in} \quad (1)$$

où  $y_{in}$  désigne le choix observé,  $U_{in}$  est une variable aléatoire non observable représentant l'utilité de la modalité  $i$  telle que perçue par l'individu  $n$ ,  $X_{in}$  est un vecteur ( $1 \times K$ ) de variables explicatives caractérisant la modalité  $i$  et l'individu  $n$ ,  $\beta$  est un vecteur ( $K \times 1$ ) de paramètres de goûts pouvant varier entre individus pour tenir compte du problème d'hétérogénéité (pour simplifier, on suppose  $\beta$  fixe) et  $\varepsilon_{in}$  est un terme aléatoire d'erreur de loi normale de moyenne nulle qui admet la corrélation avec les erreurs des autres modalités auquel est confronté l'individu.

*Modèle en notation vectorielle*

Pour un individu  $n$  donné, lorsque l'équation (1) est écrite en notation vectorielle, on obtient alors :

$$y_n = [y_{1n}, \dots, y_{Jn}]'$$

$$U_n = X_n\beta + \varepsilon_n ; \varepsilon_n \sim N(0, \Sigma_\varepsilon), \quad (2)$$

où  $y_n$ ,  $U_n$  et  $\varepsilon_n$  sont des vecteurs d'ordre ( $J \times 1$ ),  $X_n$  est une matrice ( $J \times K$ ) et  $\Sigma_\varepsilon$  est une matrice ( $J \times J$ ) de variances et covariances des erreurs. Dans le vecteur  $y_n$ , on retrouve un 1 à la position correspondant à la modalité choisie et des 0 ailleurs.

*Autre écriture du modèle (2)*

Un peu plus loin, nous emploierons une notation tout à fait équivalente à cette dernière. Cette notation qui exploite la décomposée de Cholesky de la matrice  $\Sigma_\varepsilon$  s'écrit :

$$U_n = X_n\beta + \Gamma\xi_n ; \xi_n \sim N(0, I_J), \quad (3)$$

où  $\Gamma$  est une matrice ( $J \times J$ ) triangulaire par le bas de Cholesky telle que  $\Sigma_\varepsilon = \Gamma\Gamma'$ . D'après la structure des équations (1)-(3), il est utile de rappeler que dans le cadre du probit polytomique, l'interdépendance entre les modalités est captée via la corrélation entre les erreurs du modèle. Comme il a déjà été mentionné, l'approche suppose que l'individu  $n$  choisit la modalité lui procurant l'utilité maximale. Pour représenter cela, les techniques de choix discrets associent le choix observé à celui qui a la plus grande probabilité de réalisation.

*Forme des probabilités de choix*

Dans une situation de choix à  $J$  modalités, l'individu  $n$  choisit la modalité 1, par exemple, si  $\{U_{1n} \geq U_{jn}, \forall j \in C, j \neq 1\}$ . La probabilité qui correspond à cet événement a la forme suivante :

$$\begin{aligned} P_n(1|C) &= \text{pr}(U_{1n} \geq U_{jn}, \forall j \in C, j \neq 1), \\ &= \text{pr}(\varepsilon_{jn} \leq (X_{1n} - X_{jn})\beta + \varepsilon_{1n}, \forall j \in C, j \neq 1), \\ &= \int_{\varepsilon_{1n}=-\infty}^{\infty} \int_{\varepsilon_{2n}=-\infty}^{(X_{1n}-X_{2n})\beta+\varepsilon_{1n}} \dots \int_{\varepsilon_{Jn}=-\infty}^{(X_{1n}-X_{Jn})\beta+\varepsilon_{1n}} n(\varepsilon_{1n}, \dots, \varepsilon_{Jn}; \Sigma_\varepsilon) \\ & \quad d\varepsilon_{Jn} \dots d\varepsilon_{1n}, \end{aligned} \quad (4)$$

où  $n(e; A)$  désigne une densité conjointe normale multidimensionnelle avec moyenne nulle et matrice de variances et covariances  $A$ . Sans aucune restriction sur  $\Sigma_\varepsilon$ , la dimension de l'intégrale multiple intervenant dans l'expression de cette probabilité est égale au nombre de modalités possibles  $J$ . Cette dimension  $J$  risque de devenir trop grande dans certaines situations de choix intéressantes, par exemple dans les modèles de choix de destination. En général, à l'aide de la théorie de probabilités conditionnelles et d'un changement de variables, on peut réduire de deux crans la dimension de cette intégrale, c'est-à-dire  $J - 2$ . Comme on le verra un peu plus loin, dès que  $J \geq 5$  et que la matrice de variances-covariances des termes d'erreurs  $\Sigma_\varepsilon$  est maintenue générale, on peut alors se heurter à deux obstacles majeurs sur le plan opérationnel au niveau de l'estimation classique des paramètres. Comme il vient d'être constaté, le premier obstacle est directement relié au calcul des intégrales de dimension  $J$ . Il existe d'ailleurs

beaucoup de littérature sur ce sujet (voir la section 2.1). Quant au deuxième, il survient dès lors qu'une structure générale de covariances des erreurs est postulée. Dans ce cas, le nombre de paramètres de nuisance à estimer, intervenant dans  $\Sigma_\varepsilon$ , augmente très rapidement avec le nombre de modalités considérées. Plus précisément, avec  $J$  modalités, on a  $J(J - 1)/2$  paramètres de covariances à estimer dans le modèle MNP non contraint<sup>1</sup>. Ceci occasionne donc une énorme complexité de calculs numériques qui augmente beaucoup plus rapidement que linéairement avec le nombre  $J^2$ . Cette situation est essentiellement due au coût d'évaluation du vecteur du score, c'est-à-dire le vecteur de dérivées premières de la log-vraisemblance du modèle MNP général par rapport aux composantes de cet ensemble de paramètres. En outre, dans un tel contexte, les algorithmes numériques non linéaires seront incapables de délivrer, à un coût relativement raisonnable, des estimateurs d'un tel vecteur de paramètres à grande dimension.

Afin de réaliser l'estimation à un coût raisonnable dans des ensembles de choix de grande dimension, une démarche habituellement suivie, qui s'avère des plus attrayantes, mais en même temps très dévastatrice dans la plupart des situations de choix au niveau des propriétés statistiques des estimations, consiste à ignorer la structure de covariance en employant par exemple, le logit polytomique (MNL). Dans ce cas, la probabilité de choix de la modalité 1, donnée à l'équation (4), se réduit simplement à un rapport de deux expressions exponentielles :

$$P_n(1|C) = \Lambda_n(1|C) = \frac{e^{X_{in}\beta}}{\sum_{j=1}^J e^{X_{jn}\beta}}. \quad (5)$$

Cette forme s'obtient en supposant que les termes d'erreurs  $\varepsilon_{in}$ ,  $i = 1, \dots, J$  sont distribués de loi Gumbel de façon indépendante et identique (i.i.d.).

Bien qu'une telle spécification conduise à des calculs relativement simples, on peut rencontrer des difficultés au niveau de la description de certains comportements, car elle suppose une hypothèse peu désirable dans la plupart des applications : l'*indépendance par rapport aux alternatives non pertinentes* (I.I.A.). Or dans bien des cas, cette hypothèse est violée. La formulation MNL produit alors des estimations non convergentes des paramètres et des probabilités, ce qui explique pourquoi plusieurs auteurs proposent des modèles qui éliminent la propriété d'indépendance de la spécification logit. Parmi eux, on retrouve le modèle exponentiel négatif (Daganzo, 1979), le modèle Dogit (Gaudry et Dagenais, 1979) et le modèle à valeurs extrêmes généralisé (McFadden, 1981)

- 
1. Sans oublier, bien sûr, le nombre de paramètres  $\beta$  à estimer.
  2. En plus de celles déjà reliées à la présence des intégrales.

comportant comme cas spécial le modèle logit *emboîté*. Ces formulations alternatives ont l'avantage de produire des probabilités de choix qui possèdent une forme explicite. Elles admettent toutefois des structures de dépendance entre les modalités qui s'avèrent souvent contraignantes. Si les interdépendances sont en réalité plus générales que celles admises, ces méthodologies produiront des estimations non convergentes.

À notre avis, il est nettement préférable de conserver l'approche MNP pour modéliser le choix entre modalités interdépendantes. De plus, comparée aux autres modèles de choix, la formulation probit offre bien d'autres avantages (voir aussi à ce sujet, Van Lierop, 1986 et Sheffi *et al.*, 1982). Elle permet entre autres le traitement du phénomène d'hétérogénéité entre les individus. Le MNP s'avère aussi d'une grande utilité pour traiter les observations discrètes avec données de panel et longitudinales. Il possède également une bonne propriété d'agrégation qui apparaît importante pour un grand nombre d'applications.

En résumé, bien que la formulation conventionnelle du modèle MNP soit, du moins en théorie, bien adaptée à l'analyse des choix discrets, en pratique toutefois, son analyse pose encore beaucoup de difficultés de nature numérique dans l'estimation de ses paramètres par des méthodes classiques, en particulier par la méthode du maximum de vraisemblance. Ces méthodes sont brièvement revues ci-dessous dans le but de mettre en lumière ces difficultés qui peuvent se manifester.

### 1.2 Méthodes d'estimation

La méthodologie d'estimation la plus populaire en choix discrets est sans l'ombre d'un doute celle du maximum de vraisemblance. Pour fin de notation, appelons  $\delta$  le vecteur conjoint des paramètres du modèle. Il incorpore,  $\beta$ , le vecteur des coefficients de la partie déterministe du modèle et  $\psi$  le vecteur colonne composé de tous les paramètres qui forment la structure de covariance des erreurs. Ils peuvent désigner soit les éléments de  $\Sigma_\epsilon$  ou encore ceux de la matrice triangulaire de Cholesky,  $\Gamma$ . Donc, nous posons  $\delta = (\beta', \psi)'$ . Une fois mis en évidence les paramètres identifiables de ce vecteur, il s'ensuit que la contribution de l'observation  $n$  à la fonction de vraisemblance du modèle est notée  $f(y_n | X_n; \delta)$  qui s'écrit :

$$f(y_n | X_n; \delta) = \prod_{i=1}^J P_n(i|C)^{y_{in}}. \quad (6)$$

Cette expression retient uniquement la probabilité associée à la modalité choisie par l'individu  $n$ . Pour sa part, la fonction de vraisemblance de l'échantillon s'écrit :

$$L_N(y | X; \delta) = \prod_{n=1}^N f(y_n | X_n; \delta) = \prod_{n=1}^N \prod_{i=1}^J P_n(i|C)^{y_{in}}. \quad (7)$$



Le principe du maximum de vraisemblance (MV) consiste à retenir la valeur  $\hat{\delta}_{MV}$  de  $\delta$  qui maximise l'équation (7). Habituellement, il est plus simple de maximiser le logarithme naturel de la vraisemblance. Dans ce cas, la log-vraisemblance associée à l'estimation du modèle MNP s'écrit :

$$L_N = \sum_{n=1}^N \ln f(y_n | X_n; \delta) = \sum_{n=1}^N \sum_{i=1}^J y_{in} \ln P_n(i | C). \quad (8)$$

Le score de la vraisemblance, défini comme étant la dérivée de la log-vraisemblance par rapport aux paramètres, s'écrit :

$$\frac{\partial L_N(\delta)}{\partial \delta} = \sum_{n=1}^N \frac{\partial \ln f(y_n | X_n; \delta)}{\partial \delta} = \sum_{n=1}^N \sum_{i=1}^J \frac{y_{in}}{P_n(i | C)} \frac{\delta P_n(i | C)}{\partial \delta}. \quad (9)$$

Afin d'établir le lien entre cette méthode et les deux autres méthodes décrites par la suite, il est intéressant de réécrire le score à l'équation (9) sous une forme équivalente :

$$\frac{\partial L_N(\delta)}{\partial \delta} = \sum_{n=1}^N \sum_{i=1}^J \left\{ \frac{1}{P_n(i | C)} \frac{\partial P_n(i | C)}{\partial \delta} \right\} [y_{in} - P_n(i | C)]. \quad (10)$$

Ce résultat est issu de la soustraction de l'équation (9) du vecteur nul que produit la dérivée de l'identité  $\sum_{i=1}^J P_n(i | C) = 1, \forall n$  par rapport à  $\delta$ . L'expression (10) donne à la méthode du maximum de vraisemblance une interprétation de méthode de moments car elle exprime le score comme une condition d'orthogonalité entre une erreur  $y_{in} - P_n(i | C)$  de moyenne nulle et un instrument ici représenté par  $\partial \ln P_n(i | C) / \partial \delta$ .

De façon générale, l'estimation du modèle probit polytomique par la méthode des moments généralisés (MG) vise à trouver  $\hat{\delta}_{MG}$  qui rend l'équation :

$$s(\delta) = \sum_{n=1}^N \sum_{i=1}^J W_{in} [y_{in} - P_n(i | C)], \quad (11)$$

aussi près de zéro que possible.  $W_{in}$  désigne un instrument qui a la propriété d'être orthogonal avec l'erreur  $y_{in} - P_n(i | C)$ . Dans le cas où cet instrument est assimilé à  $\partial \ln P_n(i | C) / \partial \delta$ , l'estimateur des moments généralisés est équivalent à l'estimateur du maximum de vraisemblance. L'instrument est alors dit *idéal*. Produire l'estimateur des moments généralisés consiste à résoudre le problème suivant :

$$\text{Min}_{\delta} s(\delta)' D_N s(\delta), \quad (12)$$

où  $D_N$  désigne une matrice de direction convenablement choisie. Par exemple, dans McFadden (1989),  $D_N$  est une matrice identité. Un estimateur des moments

généralisés est dit optimal si  $D_N$  est posé comme étant égal à une matrice qui converge vers la matrice d'information du score :

$$D_o = V^{-1}(s(\delta_o)) = E[s(\delta_o)s(\delta_o)']^{-1}, \quad (13)$$

où  $\delta_o$  désigne la vraie valeur du vecteur  $\delta$ . De façon pratique,  $D_N$  est calculé comme étant l'expression empirique de (13) évaluée à une certaine valeur  $\bar{\delta}$  convergente de  $\delta$ .

La dernière méthode d'estimation traitée ici est la méthode du score (MS) telle que suggérée par Ruud (1986). Son expression suit l'application d'un résultat fondamental qui dit que le score du modèle observable peut toujours s'écrire comme l'espérance mathématique conditionnelle du score du modèle latent, sachant le vecteur  $y_n$  observé (Pour plus de détails, consulter Gouriéroux et Monfort, 1989 : 424-425). Soit  $f^*(U_n | X_n; \delta)$  la densité du modèle latent de l'équation (2). Ce résultat nous permet d'écrire :

$$\frac{\partial L_N(\delta)}{\partial \delta} = \sum_{n=1}^N E \left[ \frac{\partial \ln f^*(U_n | X_n; \delta)}{\partial \delta} \middle| y_n \right]. \quad (14)$$

Lorsque le modèle latent appartient à la famille exponentielle, le score en (14) prend aussi une forme qui exprime une condition d'orthogonalité entre une erreur et un instrument. L'erreur est dite généralisée. Tout comme pour la méthode des moments généralisés, la fonction objectif de la méthode du score produisant un estimateur de  $\hat{\delta}_{MS}$  de  $\delta$  se trouve à l'équation (12) où dans ce cas,  $s(\delta)$  représente l'équation (14). Afin de saisir la différence fondamentale entre ces deux méthodes, considérons le modèle probit dichotomique suivant :

$$y_n = \begin{cases} 1 & \text{si } U_n \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

où

$$U_n = X_n \beta + \varepsilon_n, \quad \varepsilon_n \sim N(0, 1),$$

et où  $U_n$  est une variable aléatoire univariée. Le score du maximum de vraisemblance exprimé sous la forme de l'estimateur de moments généralisés avec poids idéal s'écrit :

$$\frac{\partial L_N(\beta)}{\partial \beta} = \sum_{n=1}^N \frac{\varphi(X_n \beta) X_n'}{\Phi(X_n \beta) [1 - \Phi(X_n \beta)]} [y_n - \Phi(X_n \beta)], \quad (15)$$

où  $\varphi(\cdot)$  et  $\Phi(\cdot)$  désignent respectivement la densité de probabilité et la fonction de distribution cumulative de la loi normale standardisée. D'un autre côté, la méthodologie du score considérera l'expression équivalente suivante :

$$\begin{aligned}\frac{\partial L_N(\beta)}{\partial \beta} &= \sum_{n=1}^N E[X'_n(U_n - X_n\beta) | y_n] \\ &= \sum_{n=1}^N X'_n [E(U_n | y_n) - X_n\beta],\end{aligned}\quad (16)$$

où

$$E(U_n | y_n = 1) = E(U_n | U_n \geq 0) = X_n\beta + \frac{\varphi(X_n\beta)}{\Phi(X_n\beta)}$$

et

$$E(U_n | y_n = 0) = E(U_n | U_n < 0) = X_n\beta - \frac{\varphi(X_n\beta)}{1 - \Phi(X_n\beta)}.$$

Cette même expression peut finalement s'écrire comme suit :

$$\frac{\partial L_N(\beta)}{\partial \beta} = \sum_{n=1}^N X'_n \tilde{\varepsilon}_n, \quad (17)$$

où

$$\tilde{\varepsilon}_n = E(U_n | y_n) - X_n\beta = \frac{\varphi(X_n\beta)}{\Phi(X_n\beta)[1 - \Phi(X_n\beta)]} [y_n - \Phi(X_n\beta)],$$

est appelé erreur généralisée. Bien que ce ne soit généralement pas le cas dans cet exemple, l'équation estimante associée à la méthode du score est identique analytiquement à celle associée à la méthode des moments généralisés. L'interprétation est toutefois différente. À l'équation (15), à la fois l'erreur  $y_n - \Phi(X_n\beta)$  et l'instrument dépendent tous deux des probabilités de choix. Dans l'autre cas, celui présenté à l'équation (17), seule l'erreur généralisée dépend des probabilités. L'instrument ne dépend pas de paramètres inconnus car il est formé uniquement de variables exogènes  $X_n$ .

Ce que l'on doit retenir de ces trois méthodologies conventionnelles d'estimation, pour le modèle probit polytomique, c'est que les fonctions objectifs et les scores associés dépendent tous d'une façon ou d'une autre des probabilités multidimensionnelles de choix. Dès que les applications contiennent plus de 5 modalités interdépendantes, le modèle probit polytomique ne peut être estimé de façon pratique par ces méthodologies. La prochaine section présente les principales solutions aux problèmes d'estimation du probit polytomique lorsque le nombre de modalités possibles est élevé.

## 2. LES SOLUTIONS AUX PROBLÈMES DE CALCUL

On a vu dans la section précédente, que le fait de tenir compte de la structure générale de covariances des erreurs du modèle de choix cause de sérieux problèmes majeurs sur le plan opérationnel de l'estimation des paramètres lors-

que le nombre de modalités  $J$  est supérieur à 5. Le premier est directement relié au calcul des intégrales de dimension  $J$ . L'autre existe du fait que le nombre de paramètres à estimer dans la structure de corrélation des erreurs augmente très rapidement avec le nombre  $J$  de modalités possibles. Le manque de popularité du cadre MNP dans les applications empiriques est sans aucun doute directement relié à ces deux principales difficultés. Les deux prochaines sous-sections présentent les solutions apportées aux deux problèmes qui viennent d'être soulevés.

### 2.1 Solutions au problème de dimension des intégrales

Dans cette section, nous présentons les techniques de calculs suggérées jusqu'à maintenant pour calibrer les probabilités de choix. Il existe d'ailleurs beaucoup de littérature sur ce sujet. En général, elles peuvent se classer dans les catégories suivantes :

- limiter le problème à cinq modalités ou postuler une structure à facteurs et utiliser l'intégration numérique ;
- remplacer les probabilités de choix par des approximations numériques ;
- remplacer les probabilités de choix par des simulateurs.

Comme on le verra, l'approche qui utilise les simulateurs est de loin celle que l'on doit favoriser. Pour uniformiser la notation et pour clarifier la présentation des méthodologies, nous écrivons maintenant le modèle original en déviation par rapport à l'utilité de la modalité  $i$  choisie par l'individu  $n$ . Cette opération est légitime en ce sens qu'elle ne modifie en rien la problématique de départ. De plus, le choix de la modalité à prendre comme référence dans la déviation n'a aucune conséquence sur l'estimation. (Pour plus de détails, consulter Bolduc, 1991). Il est finalement bien connu que parmi les paramètres  $\delta = (\beta', \psi)'$ , ne sont estimables que ceux pouvant être déduits de façon unique à partir des paramètres de la forme en déviation.

#### *Le modèle de choix en déviation*

Lorsqu'il est exprimé en terme de déviation par rapport à l'utilité de la modalité choisie par l'individu  $n$ , le modèle probit polytomique de l'équation (2) s'écrit :

$$U_{jn} - U_{in} = (X_{jn} - X_{in})\beta + \varepsilon_{jn} - \varepsilon_{in} \quad j \in C, j \neq i. \quad (18)$$

Pour simplifier, nous utilisons pour cette formulation du modèle la notation évidente suivante :

$$Z_{ln} = V_{ln} + \eta_{ln}, \quad l = 1, \dots, m, \quad (19)$$

où  $m = J - 1$  et où  $V_{ln}$  représente la partie déterministe du modèle en déviation. Le choix observé est lié à la différence d'utilité  $Z_{ln}$  par :

$$d_{ln} = \begin{cases} 1 & \text{si } Z_{ln} \leq 0, l = 1, \dots, m \\ 0 & \text{sinon,} \end{cases} \quad (20)$$

En notation vectorielle, on obtient :

$$d_n = [d_{1n}, \dots, d_{mn}]',$$

$$Z_n = V_n + \eta_n, \quad \eta_n \sim N(0, \Omega), \quad (21)$$

où  $Z_n$ ,  $V_n$  et  $\eta_n$  sont de dimension  $(m \times 1)$  et où  $\Omega$  est de dimension  $(m \times m)$ . À l'aide de la décomposée de Cholesky, on écrit aussi :

$$Z_n = V_n + S w_n, \quad w_n \sim N(0, I_m), \quad (22)$$

où  $S$  est une matrice triangulaire par le bas de dimension  $(m \times m)$ , telle que  $SS' = \Omega$ . Lorsque cela deviendra évident par le contexte même, nous omettrons l'indice  $n$  représentant l'individu. Nous écrirons alors :

$$Z = V + S w, \quad (23)$$

où les vecteurs  $Z$ ,  $V$  et  $w$  sont de dimension  $(m \times 1)$  et  $S$  est de dimension  $(m \times m)$  avec :

$$S = \begin{bmatrix} S_{11} & 0 & \dots & 0 \\ S_{21} & S_{22} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{bmatrix}. \quad (24)$$

La probabilité de la modalité choisie se calcule comme suit :

$$P(i|C) = \text{pr}(Z \leq 0) = \text{pr}(Z_1 \leq 0, Z_2 \leq 0, \dots, Z_m \leq 0), \quad (25)$$

où  $Z \leq 0$  représente un ensemble de  $m$  inégalités qui s'écrivent :

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ Z_m \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ V_m \end{bmatrix} + \begin{bmatrix} S_{11} & 0 & \dots & 0 \\ S_{21} & S_{22} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ w_m \end{bmatrix} \leq 0.$$

Ou, sous forme développée :

$$S_{11} w_1 \leq -V_1 \quad \Rightarrow \quad w_1 \leq -\frac{V_1}{S_{11}} = W_1,$$

$$S_{21} w_1 + S_{22} w_2 \leq -V_2 \quad \Rightarrow \quad w_2 \leq \frac{-V_2 - S_{21} w_1}{S_{22}} = W_2(w_1),$$

$$S_{m1} w_1 + \dots + S_{mm} w_m \leq -V_m \quad \Rightarrow \quad w_m \leq W_m(w_1, w_2, \dots, w_{m-1}). \quad (26)$$

Avec cette notation en main, nous pouvons donc écrire la probabilité de choix de l'équation (25) comme suit :

$$P(i|C) = \text{pr}(Z \leq 0) \\ = \text{pr}(w_1 \leq W_1, w_2 \leq W_2(w_1), \dots, w_m \leq W_m(w_1, \dots, w_{m-1})). \quad (27)$$

Passons maintenant aux solutions partielles proprement dites.

### 2.1.1 *Se limiter aux applications à 5 modalités et moins et utiliser l'intégration numérique*

Utilisant le fait que  $w \sim N(0, I_m)$ , la probabilité de la modalité  $i$  choisie s'écrit :

$$P(i|C) = \int_{w_1=-\infty}^{W_1} \varphi(w_1) \int_{w_2=-\infty}^{W_2(w_1)} \varphi(w_2) \dots \int_{w_m=-\infty}^{W_m(w_1, \dots, w_{m-1})} \varphi(w_m) dw_1 dw_2 \dots dw_m, \quad (28)$$

ce qui peut s'écrire également :

$$P(i|C) = \int_{w_1=-\infty}^{W_1} \varphi(w_1) \int_{w_2=-\infty}^{W_2(w_1)} \varphi(w_2) \dots \int_{w_{m-1}=-\infty}^{W_{m-1}(w_1, \dots, w_{m-2})} \Phi[W_m(w_1, \dots, w_{m-1})] \\ \varphi(w_{m-1}) dw_1 \dots dw_{m-1}, \quad (29)$$

une intégrale de dimension  $J - 2$ . Pour  $J > 5$ , cette intégrale devient impossible à calculer dans un temps raisonnable. Owen (1956) a été le premier à utiliser l'intégration numérique pour évaluer une distribution normale cumulative bivariable. Hausman and Wise (1978) ont étendu les travaux d'Owen à l'évaluation d'une normale trivariée. Une méthode de tabulation a été suggérée par Sparmann (1980), élaborant ainsi un programme qui emploie une tabulation d'une fonction log-vraisemblance et des approximations de séries de Taylor d'ordre 2 dans les calculs des dérivées du TNP (probit trichotomique). Avec cette méthode, le temps de calcul sur ordinateur se compare avantageusement à la vitesse de calibrage d'un TNL (logit trichotomique). Finalement, la méthode d'intégration numérique se révèle être la technique la plus précise pour calculer les probabilités de choix ainsi que les intégrales présentes dans le score. Elle est d'ailleurs souvent utilisée comme mesure de degré de précision pour évaluer les propriétés d'autres méthodes d'approximations et de simulations dans des situations de choix comportant au plus 5 modalités. Elle perd toutefois son intérêt dans les applications comportant plus de 5 modalités.

### 2.1.2 *Réduction de la dimension des intégrales : l'approche à facteurs*

Une façon alternative de réduire la complexité de calcul des intégrales est d'utiliser une approche à facteurs. Un modèle basé sur celle-ci peut se formuler comme suit :

$$Z = V + R\zeta, \quad (30)$$

où  $\zeta$  est un vecteur ( $Q \times 1$ ) et  $R$  est une matrice ( $m \times Q$ ) affectant les facteurs aux modalités. Ainsi, la dimension de l'intégrale diminue de  $m$  à  $Q$  en passant du modèle général à une spécification plus contraignante qui elle, peut être évaluée à un coût raisonnable, même si  $m > 4$ .

Il est bon de souligner que le modèle (30) a été initialement suggéré par Butler et Moffit (1982). Plus précisément, ces auteurs postulent une structure analytique à un facteur pour les erreurs. Dans leur formulation, MNP à un facteur, le terme d'erreur contient deux composantes distribuées indépendamment selon des lois normales. La première présente une forme analytique factorielle ayant une part unitaire alors que la seconde composante possède une propriété i.i.d.. Dans un tel cas, il est possible de réduire le problème d'intégrales multidimensionnelles à un problème unidimensionnel; ce qui peut être résolu par des techniques existantes. Par exemple, le calcul de l'intégrale simple ainsi dérivée pourrait s'effectuer directement par quadrature gaussienne. En revanche, cette approche se montre très restrictive puisqu'elle permet uniquement des structures d'erreurs équirélinées. Par la suite, elle a été généralisée afin d'incorporer plus de facteurs et admettre des parts arbitraires. En fait, le modèle défini ici peut être vu comme une extension d'une telle approche.

Cette pratique inspirée de l'analyse factorielle présente donc des avantages tout comme des désavantages :

**Avantages:** (i) Offre une solution probit approximative en réduisant à  $Q$  la dimension des intégrales. (ii) Pour des modèles à très grand nombre de modalités, la méthodologie à facteurs avec probabilités de choix calculées à l'aide de l'intégration numérique, n'est intéressante que si  $Q \leq 4$ , c'est-à-dire si on peut classer les modalités en un maximum de 4 regroupements de modalités homogènes. Ceci produit une structure de corrélation des erreurs en blocs où l'on affecte un facteur par bloc de modalités. Si la vraie structure des erreurs est sensiblement similaire, la formulation à facteurs captera la majorité des effets de corrélation. (iii) Pour des modèles ayant un nombre relativement élevé de modalités, le nombre de paramètres  $\psi$  à estimer dans la structure des erreurs de la formulation à facteurs sera moindre que dans la formulation standard de l'équation (22).

**Désavantages:** (i) Avec peu de facteurs, la structure de variance-covariance des erreurs captée est quelquefois restrictive. Ex : équirélination, emboîtement, .... (ii) L'interprétation n'est pas toujours évidente. (iii) À cause de la dimension ( $m \times Q$ ) de la matrice  $R$ , le nombre de paramètres à estimer dans la structure de covariances des erreurs augmente aussi rapidement avec  $m$  et  $Q$ .

Il paraît donc important de conserver la dimension des intégrales définissant les probabilités d'une façon intacte et ainsi chercher d'autres méthodes d'approximations. Ceci fera l'objet des paragraphes suivants.

### 2.1.3 Conserver la dimension intacte et chercher à approximer les intégrales

#### 2.1.3.1 Approximation numérique

L'approche d'approximation numérique consiste à remplacer les probabilités de choix dans les fonctions objectifs (8) et (12) relatives aux méthodologies d'estimation MV, MG et MS, par des fonctions relativement faciles à calculer. Sous des conditions appropriées, ces fonctions approchent assez bien les vraies valeurs des probabilités. Évidemment, des ajustements conséquents de ces substitutions doivent aussi être exécutés dans les scores. Globalement, ces méthodologies introduisent des biais dont l'importance dépend de la structure d'interdépendance entre les modalités. De plus, les biais ne disparaissent pas nécessairement lorsqu'on accroît la taille de l'échantillon. Deux approches s'affirment dans cette classe :

##### a) Approximation de Clark

Cette approximation de la probabilité de choix  $P(i|C)$  est basée sur un résultat de Clark (1961) qui dit que sous certaines conditions,  $\max(U_1, U_2)$ , où  $U_1$  et  $U_2$  sont des variables aléatoires normales standardisées, suit approximativement une normale. Daganzo (1979) développe des estimateurs MV approchés du MNP via cette approximation.

**Avantages:** (i) Nécessite seulement l'évaluation d'intégrales normales univariées. Par conséquent, elle est intéressante pour sa rapidité d'application. (ii) Son applicabilité n'est pas limitée par le nombre de modalités mises en jeu.

**Désavantages:** (i) La précision ne s'améliore pas en augmentant la taille de l'échantillon. (ii) La méthode se révèle inadéquate dans des situations où les composantes aléatoires des utilités ont des variances très différentes et/ou comporte des corrélations négatives (voir Horowitz, Sparmann and Daganzo, 1981). (iii) Sacrifie une certaine précision pour atteindre une épargne notable en temps de calcul sur ordinateur.

##### b) Approximation de Mendell-Elston (M-E)

Celle-ci consiste à approcher des lois conditionnelles qui dépendent de moments provenant de la normale tronquée par des lois normales conditionnelles non tronquées. La probabilité de choix  $P(i|C)$  est alors approchée par un ensemble d'équations récursives comportant seulement des intégrales normales à une dimension.

Pour illustrer la technique, considérons un cas à  $J = 3$  modalités. Utilisant l'équation (25) comme point de départ,

$$\begin{aligned} P(i|C) &= P(Z_1 \leq 0, Z_2 \leq 0) \\ &= P(Z_1 \leq 0 | Z_2 \leq 0)P(Z_2 \leq 0) = P(Z_2 \leq 0 | Z_1 \leq 0)P(Z_1 \leq 0). \end{aligned}$$



Choisissons arbitrairement la seconde forme conditionnelle. Grâce aux équations (26) - (27), nous pouvons écrire :

$$P(ilC) \approx P\left(w_2 \leq \frac{-V_2 - S_{12}w_1}{S_{22}} \mid w_1 \leq -\frac{V_1}{S_{11}}\right) P\left(w_1 \leq -\frac{V_1}{S_{11}}\right)$$

$$\equiv P(w_2 \leq W_2(w_1) \mid w_1 \leq W_1) P(w_1 \leq W_1), \quad (31)$$

sachant que  $w_1 \sim N(0, 1)$ . En se basant sur un modèle conditionnel de deuxième ordre pour  $w_2$ , nous pouvons déduire facilement les deux premiers moments  $\mu_{211}$  et  $\sigma_{211}^2$  de la distribution de  $w_2 \mid (w_1 \leq W_1)$ . Pearson (1903) suggère d'approcher la distribution de  $w_2 \mid (w_1 \leq W_1)$  par une normale. Ceci implique alors que (31) peut s'écrire :

$$P(ilC) \approx \Phi(W_{211}) \Phi(W_1), \quad (32)$$

$$\text{où } W_{211} = \frac{W_2(w_1) - \mu_{211}}{\sigma_{211}}.$$

Mendell et Elston ont étendu l'équation (32) aux problèmes de grande dimension. Pour  $m=5$ , par exemple, les applications récursives les amènent à l'approximation suivante  $P(ilC) \approx \Phi(W_{41321})\Phi(W_{3121})\Phi(W_{211})\Phi(W_1)$ . Pour plus de détails sur les aspects de calcul reliés à cette technique, voir Bolduc et Kaci (1989A). Sur le choix quant à la séquence de conditionnement à retenir, consulter Langdon (1984).

La méthodologie Mendell-Elston a été testée par Kamakura (1988), Bolduc et Kaci (1989A) et Mahmassani et Lam (1990). De façon générale, la méthode M-E semble plus performante que la méthode de Clark au point de vue numérique. Toutefois, lorsque le nombre de modalités est supérieur à 10, la méthodologie de Clark est plus économique au niveau du temps de calcul.

**Avantages:** (i) Cette méthodologie est aussi simple à utiliser que celle de Clark. (ii) L'approche M-E fournit généralement une meilleure approximation de probabilités que celle de Clark ; voir Kamakura (1988). (iii) Demeure utilisable même si le nombre de modalités est très élevé.

**Désavantages:** (i) Cette méthode est plus exigeante que celle de Clark en terme de temps d'ordinateur. Pour un problème à quinze modalités, Kamakura (1988) rapporte que la méthode M-E est plus lente que celle de Clark par un facteur de 5. Des résultats similaires ont été rapportés par Mahmassani et Lam (1990). (ii) Peu de choses sont connues sur la précision du résultat de Pearson (1903). (iii) Les probabilités estimées ne possèdent pas la propriété d'additivité, c'est-à-dire qu'elles ne se somment pas à 1.

La critique la plus sérieuse faite au sujet des méthodologies d'approximations numériques est qu'il est difficile d'établir des résultats statistiques concer-

nant les estimateurs qui découlent de leur utilisation. Un résultat important est que la qualité de l'approximation n'augmente pas si on accroît la taille de l'échantillon. Nous passons maintenant à une seconde classe de techniques pour approcher des probabilités de choix. Comme nous le verrons, les méthodologies basées sur le principe de simulation stochastique ont la plupart des caractéristiques nécessaires pour produire des estimateurs à bonnes propriétés asymptotiques.

### 2.1.3.2 *Simulation stochastique*

La solution la plus prometteuse concernant l'évaluation des probabilités de choix probit est sans aucun doute celle basée sur l'emploi des simulateurs stochastiques. L'idée de base consiste à remplacer les probabilités dans les fonctions objectifs (8) et (12) ainsi que les fonctions des probabilités intervenant dans les scores (10), (11) ou (14) par des moyennes empiriques de fonctions calculées à partir de simulations stochastiques faites sur le modèle latent de l'équation (1). Ces moyennes empiriques sont qualifiées de simulateurs. Albright, Lerman et Manski (1977) ont été parmi les premiers à suggérer cette méthodologie. De récentes techniques proposées notamment dans McFadden (1989) sont un raffinement de cette démarche. L'emploi de simulateurs dans les trois procédures conventionnelles d'estimations déjà discutées produit respectivement : 1) la méthode du maximum de vraisemblance simulée (MVS), 2) la méthode des moments généralisés simulés (MGS) et 3) la méthode du score simulé (MSS). La principale caractéristique des simulateurs plus sophistiqués est qu'ils sont continus dans les paramètres, ce qui permet d'utiliser les algorithmes conventionnels d'optimisation dans l'estimation sans affecter son comportement numérique.

Les méthodes d'estimation des modèles MNP, basées sur les simulateurs stochastiques, sont ici évoquées seulement pour leurs liens avec les méthodes classiques<sup>3</sup>. Les références concernant la méthode MGS sont issues de McFadden (1989), celles de la méthode MVS sont entre autres issues de Börsch-Supan et Hajivassiliou (1990) alors que se trouvent dans Hajivassiliou et McFadden (1990) celles de la méthode MSS. Pour ce qui est des propriétés asymptotiques de ces estimateurs, il est possible, sous certaines conditions, non seulement d'en prouver la convergence et la normalité asymptotique, mais aussi d'en établir les vitesses de convergence et d'évaluer l'importance d'éventuels biais asymptotiques. On peut trouver, dans Hajivassiliou et McFadden (1990), une discussion des propriétés des estimateurs de type MSS dans le cas des modèles à variables latentes limitées comportant, bien sûr comme cas spécial, le modèle MNP, et dans McFadden (1989) on retrouvera un travail du même genre

---

3. C'est en effet, tout récemment, que la simulation stochastique s'est avérée utile dans les procédures d'estimation classiques de modèles MNP (voir par exemple Gouriéroux et Monfort, 1991 et pour des références supplémentaires voir Hajivassiliou et McFadden, 1990).

concernant l'estimateur MGS dans le cadre du modèle MNP<sup>4</sup>. Quant aux propriétés de l'estimateur obtenu par la procédure MVS<sup>5</sup>, elles ont été démontrées dans un cadre assez général, plus précisément dans des modèles avec hétérogénéité (voir Gouriéroux et Monfort, 1991).

Sur la base des propriétés asymptotiques, nous pouvons donc utiliser indifféremment l'une ou l'autre des trois méthodes d'estimation basées sur les simulateurs que nous venons d'évoquer. Le seul critère qui, logiquement, permet la mise en application de telle ou telle méthode est le coût en temps de calculs numériques requis, ou en d'autres termes, la facilité d'emploi avec laquelle la méthode peut être réalisée dans différentes situations de choix particulières. Jusqu'ici, vu de cet angle, aucune de ces procédures n'a dominé l'autre d'une façon évidente. Par contre, la plupart des applications empiriques récentes ont favorisé l'utilisation de la méthode MGS (McFadden, 1989; Pakes et Pollard, 1989; Hajivassiliou et McFadden, 1989; Berkovec et Stern, 1989; Berry, 1989; Hotz et Sanders, 1990) alors que la méthode MVS est demeurée nettement impopulaire parce qu'elle introduit un biais dû à la simulation lorsque le nombre de tirages est limité. D'ailleurs à notre connaissance, il n'existait que deux études empiriques utilisant la méthode MVS dans la littérature économétrique, soit celle de Börsch-Supan *et al.* (1990) et celle de Ben-Akiva et Bolduc (1991). Mais, il est intéressant de noter que dans ces études, il a été montré que la méthode MVS était tout de même la moins coûteuse à mettre en application, contrairement à la méthode MGS. En ce qui concerne la méthode MSS, son application semble elle aussi relativement facile d'emploi. L'application déjà disponible dans Hajivassiliou et McFadden (1990) semble très prometteuse même si plus d'expériences pratiques s'avèrent encore nécessaires. Cette méthode a de plus le mérite, du point de vue du praticien, d'être applicable dans un bon nombre de modèles latents autres que le modèle MNP, en particulier dans des modèles tobit et des modèles de déséquilibre.

Pour donner un aperçu de l'approche dite à simulateurs, nous présentons d'abord le **simulateur de fréquence** suggéré par Albright, Lerman et Manski. Par la suite, nous passons à la description des simulateurs plus sophistiqués. Notre survol se veut avant tout simple. Une présentation critique beaucoup plus détaillée que la nôtre concernant les simulateurs stochastiques se retrouve dans Hajivassiliou, McFadden et Ruud (1991). La compréhension de cet article n'est toutefois pas très facile pour le lecteur non familier avec cette littérature.

#### a) *Simulateur de fréquence (Albright, Lerman et Manski, 1977)*

Utilisant le processus de l'équation (3) qui engendre les observations, tout en omettant l'indice propre à l'individu afin d'en simplifier la notation, nous obtenons :

---

4. Notons également que Pakes et Pollard (1989) ont dérivé des propriétés asymptotiques pour une large classe d'estimateurs fondés sur des simulations, dont l'estimateur MGS de McFadden est justement un cas particulier.

5. En plus de celles de l'estimateur obtenu de la méthode de MGS.

$$U = X\beta + \Gamma\xi, \quad \xi \sim N(0, I_J). \quad (33)$$

Étant donné  $\beta$  et  $\Gamma$ , tirer d'abord des vecteurs  $\xi_r$ ,  $r = 1, \dots, R$  de loi normale standardisée. Calculer le nombre de fois  $R_i$  que  $U_{in} > U_{jn} \quad \forall j \neq i$ . Utiliser alors la fréquence de chaque événement pour évaluer les probabilités de choix :

$$P(i|C) \approx \frac{R_i}{R}, \quad \forall i \in C. \quad (34)$$

Remarques : (i) Dans certains cas, pour un nombre de tirages donné  $R$ , le nombre de succès enregistré peut être zéro. La log-vraisemblance vaut alors  $-\infty$ . Afin de contourner ce problème Albright, Lerman et Manski (1977) suggèrent de remplacer l'équation (34) par  $P(i|C) \approx (R_i + 1) / (R + J)$ . Le biais introduit par cette expression est faible pour  $P(i|C)$  élevée mais il peut être important pour de faibles probabilités  $P(i|C) \approx 0$ . (ii) L'erreur relative à la méthode de simulation telle que décrite dans leur article est inversement proportionnelle à  $\sqrt{R_i}$ .

**Avantages:** (i) La réalisation de cette procédure est simple. (ii) On peut contrôler la qualité de l'approximation en augmentant le nombre de tirages. (iii) La méthodologie est applicable à des problèmes contenant un nombre très élevé de modalités.

**Désavantages:** (i) Cette méthode exige un nombre très élevé de tirages pour approcher les probabilités faibles avec un niveau de précision acceptable. Le prix à payer est une augmentation notable du temps de calcul sur ordinateur. (ii) L'inconvénient le plus important est que ce simulateur est sujet à des discontinuités par rapport aux paramètres. Son usage dans des algorithmes conventionnels peut causer de sérieuses difficultés numériques en utilisant les méthodes itératives conventionnelles d'optimisation.

Pour relier ce simulateur à ceux plus performants, donnons-lui une formulation alternative basée sur la représentation en différence fournie aux équations (18)-(20). Omettant l'indice  $n$ , la probabilité de la modalité choisie se calcule comme suit :

$$\begin{aligned} P(i|C) &= \text{pr}(Z \leq 0) = \int 1[Z \leq 0] \quad n(Z - V, \Omega) dZ \\ &= \int 1[V + Sw \leq 0] \quad n(w, I_m) dw, \end{aligned} \quad (35)$$

où  $1[\cdot]$  représente la fonction indicatrice qui prendra la valeur 1 lorsque la condition entre crochets sera satisfaite, sinon la valeur 0 et  $n(u, A)$  désigne la densité normale multivariée de moyenne nulle et de matrice de variances et covariances  $A$ . La notation précédente représente une intégrale de dimension  $m = J - 1$  dont l'interprétation est celle d'une espérance mathématique :  $P(i|C) = E(1[V + Sw \leq 0])$ . Le simulateur de fréquence est tout simplement l'expression d'échantillon de cette espérance. Notons que la fonction indicatrice a un comportement à palier par rapport à des changements marginaux dans la valeur des paramètres. La prochaine classe de simulateurs décrite ci-dessous

remplace la fonction indicatrice par une fonction continue par rapport aux paramètres du modèle (*simulateurs lisses*). Nous présentons d'abord des simulateurs lisses qui ont la propriété de produire des estimations sans biais des probabilités de choix. Par la suite, nous considérons une sous-classe de simulateurs dont le biais ne disparaît qu'asymptotiquement.

b) *Simulateurs lisses sans biais*

*Simulateur d'échantillonnage par importance (McFadden, 1989)*

L'objectif ici est d'évaluer l'intégrale de l'équation (35). L'approche d'échantillonnage par importance remplace cette expression par une intégrale plus simple à calculer dont le domaine d'intégration coïncide avec celui de (35). Soit  $\gamma(\bullet)$  une densité ayant comme support l'orthant non positif. Par exemple, prenons  $\gamma(\bullet)$  comme désignant une densité conjointe de variables exponentielles indépendantes. La solution d'échantillonnage par importance *Monte Carlo importance sampling* consiste à réécrire l'équation (35) de la façon suivante :

$$P(i|C) = \int h(Z)\gamma(Z)dZ, \quad (36)$$

où

$$h(Z) = \frac{n(Z - V, \Omega)}{\gamma(Z)}.$$

Cette expression est une espérance mathématique prise par rapport à une distribution exponentielle indépendante<sup>6</sup>. Utilisons l'indice  $r$ ,  $r=1, \dots, R$  pour désigner les tirages. Étant donné  $\beta$  et  $\Gamma$ , la probabilité de choix de la modalité  $i$  est approchée par :

$$P(i|C) = E\{h(Z)\} \approx \sum_{r=1}^R \frac{h(Z_r)}{R}. \quad (37)$$

L'éventuel inconvénient de ce simulateur est qu'il ne respecte pas la propriété d'additivité, c'est-à-dire que  $\sum_i P(i|C) \neq 1$ . Ceci s'explique du fait que pour chaque modalité  $i$ , d'après l'équation (37), on tire un vecteur aléatoire  $Z_r$  différent. Un simulateur qui satisfait la propriété d'additivité requiert que l'on base les calculs de probabilité de chaque modalité  $i$  sur un même vecteur  $Z_r$ .

*Le simulateur récursif conditionnel (Börsch-Supan et Hajivassiliou, 1990)*

Rappelant les équations (22) et (26) et conditionnant sur  $w_1$ , la variable aléatoire  $w_2$  est alors distribuée comme une variable normale standardisée tronquée à droite avec point de troncature  $W_2(w_1)$ . De la même façon, en conditionnant sur  $w_1$  et  $w_2$ , la variable  $w_3$  est distribuée comme une normale standardisée tronquée à droite avec point de troncature  $W_3(w_1, w_2)$ , et ainsi de suite... Étant donné  $\beta$  et  $\Gamma$ ; pour évaluer  $P(i|C)$ , cette méthodologie suggère de tirer  $R$  vecteurs nor-

6. Des variables aléatoires provenant de cette distribution peuvent être tirées en prenant le logarithme naturel de nombres aléatoires uniformes  $(0,1)$   $[U(0, 1)]$ .

maux tronqués<sup>7</sup>  $w_r$ , à composantes indépendantes, calculées de façon récursive avec points de troncature à droite telles que définies à l'équation (26). La probabilité de choix de la modalité  $i$  peut alors être approchée par

$$P(i|C) = \frac{1}{R} \sum_{r=1}^R P_r(i|C) \quad (38)$$

où

$$P_r(i|C) = \Phi(W_{1,r})\Phi(W_{2,r}(w_{1,r})|w_{1,r})\dots\Phi(W_{m,r}(w_{1,r},\dots)|w_{1,r},\dots),$$

qui s'exprime comme un produit d'intégrales normales univariées. Börsch-Supan et Hajivassiliou (1990) ainsi que Hajivassiliou et McFadden (1990) démontrent entre autres que ce simulateur de probabilités de choix est sans biais. Les expériences empiriques pour tester ce simulateur ont donné des résultats des plus encourageants. Toutefois, les probabilités calculées ne satisfont pas la propriété d'additivité.

#### *Le simulateur économique de Deák (McFadden, 1989)*

Ce simulateur que l'on doit à Deák (1980) et qui est décrit dans McFadden (1989) est extrêmement économique en temps de calcul. Il utilise un résultat connu selon lequel un vecteur de variables normales standardisées peut être exprimé comme le produit d'un vecteur  $s$  de composantes tirées de la sphère unitaire dans  $\mathfrak{R}^l$  par  $\lambda$ , la racine carrée d'une variable aléatoire de loi  $\chi^2$  indépendante du vecteur  $s$ . Ceci permet donc de réécrire l'équation (3) comme suit :

$$U = X\beta + \lambda\Gamma s. \quad (39)$$

Pour un vecteur  $s$  donné, cette expression projette l'espace  $\mathfrak{R}^l$  des probabilités de choix sur le domaine univarié  $\mathfrak{R}^+$   $[0, \infty]$  de  $\lambda$ . Ensuite, on peut toujours scinder le domaine  $[0, \infty]$ , en intervalles  $[\lambda_j, \lambda_{j+1}]$ ,  $j=0, \dots, m$  où à chaque intervalle est associé une composante de  $U_n$  maximale. Certains des intervalles peuvent être dégénérés (vides). Conditionnellement à un  $s$  donné, supposons que  $U_1$  soit maximal sur l'intervalle  $[\lambda_3, \lambda_4]$ , la probabilité conditionnelle de choix pour la modalité 1 serait donc calculée comme suit :

$$P(1|C, s) = H_f(\lambda_4^2) - H_f(\lambda_3^2), \quad (40)$$

la différence entre deux valeurs de la distribution cumulative de la variable  $\lambda$  de loi  $\chi^2$  univariées. Utilisant une notation plus générale, la probabilité de choix non conditionnelle pourra alors se calculer de la façon suivante :

7. Pour simuler des variables normales standard tronquées à droite avec le point de troncature  $T$ , on utilise la formule d'inversion suivante:

$$\int_{-\infty}^x \frac{\Phi(s)}{\Phi(T)} ds = p \Rightarrow x = \Phi^{-1}(p\Phi(T)), \text{ où } p \text{ est } U(0,1).$$

$$P(i|C) = E_s[H_j(\lambda_{j+1}^2) - H_j(\lambda_j^2)] \quad j = 0, \dots, m.$$

Utilisant  $R$  tirages du vecteur  $s$  ( $s_r$ ,  $r = 1, \dots, R$ ), le simulateur est calculé comme la moyenne empirique suivante :

$$P(i|C) = \frac{\sum_{r=1}^R P(i|C, s_r)}{R}.$$

Parce qu'il est très rapide et peu coûteux, ce simulateur est très attrayant. Il peut être employé pour des problèmes ayant un nombre très élevé de modalités. Comme les probabilités conditionnelles de toutes modalités  $i \in C$  utilisent un même vecteur  $s_r$ , la propriété d'additivité est satisfaite automatiquement. Toutefois, cette technique peut produire des intervalles vides, ce qui introduit des discontinuités pouvant nuire au comportement des algorithmes conventionnels d'optimisation.

### c) *Simulateurs lisses de type noyau*

Pour produire un simulateur lisse, McFadden (1989) ajoute au modèle de l'équation (3) un terme de simulation asymptotiquement absent. Le modèle MNP ainsi modifié s'écrit :

$$\tilde{U} = X\beta + \Gamma\xi + b_N\mu, \quad b_N \xrightarrow{N \rightarrow \infty} 0, \quad (41)$$

où  $\mu$  est un vecteur ( $J \times 1$ ) de variables aléatoires indépendantes de loi donnée. Nous présentons deux versions, celle basée sur  $\mu$  avec composantes de loi Gumbell i.i.d. et celle basée sur  $\mu$  de composantes normales standardisées indépendantes. Celles-ci produisent respectivement des simulateurs à noyau logistique et des simulateurs à noyau normal.

#### *Noyau logistique*

Conditionnellement à  $\xi$ , si  $\mu$  suit une loi Gumbel standardisée i.i.d.,

$$P(i|C, \xi) = \Lambda(i|\xi) = \frac{e^{(X_i\beta + \Gamma_i\xi)/b_N}}{\sum e^{(X_j\beta + \Gamma_j\xi)/b_N}}, \quad (42)$$

où  $X_i$  et  $\Gamma_i$  dénotent la ligne  $i$  de  $X$  et  $\Gamma$ , respectivement. La probabilité de choix de la modalité  $i$  s'obtient en intégrant sur  $\xi$ . On aura donc :

$$P(i|C) = \int_{\xi} \Lambda(i|\xi) n(\xi, I_j) d\xi = E_{\xi}[\Lambda(i|\xi)].$$

Le simulateur lisse de type noyau logistique pour  $P(i|C)$  est calculé à partir d'une moyenne empirique  $\sum_{r=1}^R \Lambda(i|\xi_r) / R$ . La propriété d'additivité sera satisfaite si un même tirage est utilisé pour l'ensemble des modalités  $i \in C$ .

*Noyau normal standard*

Le second simulateur découle de l'hypothèse que  $\mu$  suit une  $N(0, I_j)$ . Utilisant la notation du modèle exprimé en différence par rapport à la modalité choisie, cette approche produit la relation suivante

$$\tilde{Z} = V + Sw + b_N(\mu_{-i} - \mu_i), \quad (43)$$

où  $\mu_i$  est un vecteur ( $m \times 1$ ) qui exclut la composante  $\mu_i$ . Les probabilités de choix peuvent alors être exprimées à l'aide de la récursion suivante : d'abord conditionnellement à  $w$  et  $\mu_i$ , l'équation (43) permet d'écrire :

$$P_n(\tilde{Z} \leq 0 | C, w, \mu_i) = \prod_{l=1}^m \Phi \left( \mu_i - \frac{V_l + S_l w}{b_N} \right), \quad (44)$$

où  $V_l$  est la composante  $l$  de  $V$  et  $S_l$  représente la ligne  $l$  de  $S$ . Intégrant ensuite sur  $w$  et  $\mu_i$ , on obtient la probabilité non conditionnelle suivante

$$\begin{aligned} P(i|C) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \prod_{l=1}^m \Phi \left( \mu_i - \frac{V_l + S_l w}{b_N} \right) \phi(\mu_i) d\mu_i \right] n(w, I_m) dw, \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{l=1}^m \Phi \left( \mu_i - \frac{V_l + S_l w}{b_N} \right) n \left( \begin{matrix} w \\ \mu_i \end{matrix}, I_{m+1} \right) dw d\mu_i, \end{aligned} \quad (45)$$

l'espérance de la probabilité conditionnelle  $P(Z \leq 0 | C, w, \mu_i)$ . Le simulateur est calculé comme une moyenne empirique  $\sum_{r=1}^R P(i|C, w_r, \mu_{i,r}) / R$  prise sur les  $R$  tirages.

**Avantages:** (i) Les simulateurs de type noyau peuvent être employés dans des problèmes ayant un nombre élevé de modalités. (ii) Ces deux méthodes produisent des simulateurs asymptotiquement sans biais qui satisfont la propriété d'additivité lorsqu'un même tirage est utilisé pour tout  $i \in C$ . (iii) En pratique, ces simulateurs produisent une approximation assez précise des probabilités de choix même lorsque  $R$  est faible.

**Désavantages:** Le principal inconvénient des techniques de noyau est la nécessité d'avoir des échantillons de grande taille lorsqu'une grande précision est requise pour évaluer les probabilités.

Les procédures décrites ci-haut présentent toutes plus ou moins des lacunes, que ce soit au niveau du degré de précision, de l'exigence en temps d'exécution de calcul ou de la complexité d'application. À l'aide d'expériences de Monte Carlo, Bolduc et Kaci (1989A), ont comparé les principaux simulateurs proposés par McFadden aux méthodes utilisées auparavant. Certaines méthodes ont été écartées dès le départ dues à leurs lacunes. Généralement, les résultats de simulations indiquent que les simulateurs lisses et ceux de type noyau fournis-



sent de bons résultats. En outre, ils permettent d'estimer de façon assez précise les probabilités de choix même lorsque le nombre de modalités est raisonnablement élevé.

Les simulateurs lisses ci-dessus présentent ainsi la possibilité d'obtenir de nouvelles stratégies d'estimation des modèles MNP. Beaucoup de problèmes d'ordre technique qui, avec les méthodes classiques, furent à l'origine non résolus dans des situations de choix de grande dimension, le sont maintenant. Dans ce contexte, il n'est pas inutile de rappeler que le but de la technique de simulation stochastique est d'approximer les intégrales multidimensionnelles intervenant dans la fonction objectif utilisée dans la méthode d'estimation. L'origine de ces intégrales provient de la définition des probabilités de choix ou de celle des moments conditionnels.

Jusqu'à maintenant, notre discussion s'est limitée à l'évaluation et la simulation des probabilités de choix. Rappelons que dans le processus itératif d'optimisation, ces fonctions entrent d'une façon ou d'une autre tant dans la fonction objectif retenue que dans le score associé. Une direction de recherche non décrite ici vise à simuler de façon directe le score correspondant à l'estimation MV qui peut s'exprimer comme une espérance conditionnelle<sup>8</sup>. Tel est d'ailleurs la démarche suivie par la méthode MSS (Hajivassiliou et McFadden, 1990).

En terminant cette sous-section, il convient de souligner brièvement que, compte tenu des différentes techniques de simulation qui viennent d'être évoquées, un autre axe de recherche fructueux pour résoudre les difficultés d'ordre numérique dans l'estimation de modèles probit polytomiques serait de reconsidérer d'autres aspects dans la spécification conventionnelle du modèle MNP qui conservent la même souplesse et qui conduisent en même temps à d'importantes simplifications de calculs. Une démarche possible et intéressante consiste à prendre comme modèle de départ une version modifiée de l'équation (41) en posant  $b_N$  égal à un. Sous l'hypothèse que  $\mu$  est un vecteur de composante Gumbel i.i.d., les probabilités de choix se calculent alors comme des moyennes des probabilités logistiques conditionnelles. Comme cas particulier de ce modèle, on retrouve le logit polytomique. Il s'obtient lorsque  $\Gamma$  est une matrice nulle. On peut trouver dans Berkovec et Stern (1991) ainsi que dans Ben-Akiva et Bolduc (1991) une utilisation de cette démarche. L'approche de ces derniers est d'ailleurs brièvement décrite à la section suivante puisqu'elle apporte également une solution au second problème mentionné.

## 2.2 Solutions au problème de prolifération des paramètres de nuisance

Rappelons que la seconde complication liée à l'usage du MNP survient lorsqu'une structure générale de covariances des erreurs est postulée. Dans ce

---

8. Pour simplifier les choses, on a préféré ne pas aborder ces méthodes de simulation, qui devront normalement faire partie d'un survol de ce genre. Nous référons le lecteur désireux d'obtenir plus de détails à l'article Hajivassiliou, McFadden et Ruud (1991).

cas, le nombre de paramètres de nuisance à estimer intervenant dans la matrice de variances-covariances des erreurs augmente rapidement avec le nombre de modalités considéré. Afin de résoudre ce problème, Ben-Akiva et Bolduc (1990 et 1991) développent une approche à erreurs autorégressives généralisées. L'intérêt de celle-ci tient au fait que la structure de covariance des erreurs est caractérisée à partir de spécifications parcimonieuses dont le nombre de paramètres est indépendant du nombre de modalités. Plus précisément, dans sa forme la plus complète, le modèle qu'ils proposent se résume ainsi :

$$U_n = X_n \beta + M_n \zeta_n + v_n, \quad \zeta_n \sim N(0, I_J) ; v_n \sim i.i.d. \text{Gumbel}(0, I_J), \quad (46)$$

avec  $M_n = T_n(I - \rho W_n)^{-1}$ , une matrice d'ordre  $(J \times J)$ ,  $T_n$  est une matrice diagonale  $(J \times J)$  d'éléments  $\sigma_i$  représentant les écarts-types des erreurs spécifiques aux modalités.  $\rho$  est un coefficient d'autocorrélation contraint à appartenir au domaine  $(-1 < \rho < 1)$ ,  $W_n$  est la matrice de poids d'ordre  $(J \times J)$  qui décrit la structure de corrélation entre les modalités. Celle-ci dépend d'un faible nombre de paramètres associés à des variables caractérisant la corrélation existante. Conditionnellement à  $\zeta_n$ , l'équation (46) définit un modèle MNL conditionnel. La probabilité de choix  $P_n(i|C)$  non conditionnelle pour la modalité est calculée de la façon suivante :

$$P_n(i|C) = \int_{\zeta} \Lambda(i|\delta, \zeta) n(\zeta, I_J) d\zeta \quad (47)$$

avec

$$\Lambda(i|\delta, \zeta) = \frac{e^{X_{in}\beta + M_{in}\zeta}}{\sum_{j=1}^J e^{X_{jn}\beta + M_{jn}\zeta}}. \quad (48)$$

Dans un modèle à nombre élevé de modalités, on remplace la probabilité exacte par un simulateur exprimant la version empirique de l'espérance donnée à l'équation (47).

**Avantages:** (i) Tel que formulé, le modèle apparaît avantageux puisqu'il s'adapte au modèle MNL et qu'il conserve la même souplesse que le modèle MNP général. (ii) La structure de corrélation entre les erreurs est caractérisée à l'aide des spécifications parcimonieuses dont le nombre de paramètres ne dépend pas de  $J$ . D'où la possibilité de réduction substantielle du nombre de paramètres à estimer. (iii) L'approche demeure opérationnelle même lorsque  $J$  est très élevé. (iv) Un modèle construit dans cet esprit a évidemment des implications importantes pour l'emploi pratique des simulateurs de probabilités de type noyau logistique.

**Désavantages:** (i) Le grand désavantage de cette procédure est le risque d'erreur de spécification qu'elle peut introduire. (ii) La présente spécification est considérée comme étant une approximation d'un modèle plus général. Celle-ci peut entraîner à ce titre une perte de précision.

### Une forme à facteurs du modèle

Aussi intéressante du point de vue pratique, est la possibilité d'introduire des facteurs dans la formulation précédente. En effet, Bolduc et Kaci (1991) ont, pour leur part, suggéré de modifier le processus autorégressif d'erreurs en le faisant dépendre d'un faible nombre de facteurs. Un même facteur s'applique à un regroupement donné de modalités homogènes. Leur modèle se présente ainsi :

$$U_n = X_n\beta + M_n\zeta_n + v_n, \quad \zeta_n \sim N(0, I_Q) \quad (49)$$

avec, cette fois-ci,  $M_n = T_n(I - \rho W_n)^{-1}E$  où  $E$  désigne une matrice d'ordre  $(J \times Q)$  affectant les facteurs aux modalités ( $Q \ll J$ ). La probabilité de choix peut alors se calculer comme suit :

$$P_n(i|C) = \int_{\zeta} \Lambda(i|\delta, \zeta) n(\zeta, I_Q) d\zeta \quad (50)$$

avec

$$\Lambda(i|\delta, \zeta) = \frac{e^{X_{in}\beta + M_{in}\zeta}}{\sum_{j=1}^J e^{X_{jn}\beta + M_{jn}\zeta}} \quad (51)$$

Ainsi, il faut calculer une intégrale à  $Q$  dimensions, ce qui peut se faire pour des valeurs de  $Q$  faibles. Des simplifications de calculs peuvent alors s'introduire dans le calcul de l'intégrale, de façon à éviter l'intégrale à  $J$  dimensions ; ce qui devrait se révéler appréciable dans les applications où  $J$  est très grand. Dans un modèle comportant un faible nombre de regroupement de modalités homogènes, l'approche de type facteurs permet alors d'évaluer les probabilités à l'aide de procédures d'intégration numérique standards.

**Avantages:** (i) Ce type de modèle, simple à manier et basé sur des intégrales exactes, permet d'éviter la présence d'imprécisions due à l'emploi de simulateurs. (ii) L'approche peut être utilisée dans la détermination de valeurs d'initialisation pour des procédures algorithmiques.

**Désavantage:** Le grand désavantage d'une telle formulation est qu'elle ne représente qu'une approximation à un modèle plus général qui devrait contenir un facteur par modalité. Il est à souhaiter qu'un petit nombre de facteurs suffise à tenir compte de la plupart des interdépendances entre les modalités.

### CONCLUSION

En économie de transport notamment, les applications empiriques des modèles de choix discrets ont été, la plupart du temps pour des raisons de commodité, trop souvent limitées au logit polytomique et à des formulations associées, alors que la spécification qui se révèle la plus souple pour tenir compte d'éventuelles interdépendances entre les modalités est sans nul doute le modèle MNP. Mais sous sa forme générale, on s'est heurté à des difficultés de nature

numérique restreignant ainsi son application pratique. En particulier, dû à l'inévitable présence des intégrales multidimensionnelles, l'expression de la vraisemblance associée à l'estimation des paramètres d'un tel modèle s'est montrée non seulement compliquée mais également très difficile à utiliser en pratique.

Cependant, de récents développements en techniques statistiques, combinés à des progrès réalisés en analyse numérique et à des ordinateurs super-performants, sont venus briser certaines barrières. Il existe à présent des techniques qui s'avèrent prometteuses pour estimer les modèles MNP dans des ensembles de choix de grande dimension. L'objectif de cet article a été justement de fournir aux praticiens une synthèse critique de ces quelques travaux consacrés à l'estimation de tels modèles. Ceci a permis d'apprécier à la fois l'intérêt et les limites de certaines approches visant à réaliser d'importantes simplifications de calcul. Les méthodes d'estimation des modèles MNP telles que présentées dans ce texte peuvent être classifiées en deux grandes catégories : les méthodes applicables à des problèmes contenant peu de modalités (au plus 5) et les méthodes pour des problèmes plus larges. La première catégorie comporte des estimateurs fondés sur des méthodes standards (MV, MG, MS). Quant à la seconde, elle englobe des estimateurs fondés sur des méthodes de simulation (MVS, MGS, MSS). Signalons que ces derniers estimateurs jouissent eux aussi de bonnes propriétés asymptotiques, puisqu'il est possible d'en démontrer la convergence et la normalité asymptotique sous certaines conditions.

Il convient de rappeler au lecteur que dans le but d'estimer les paramètres, nous disposons donc d'au moins six méthodes d'estimation convergentes et que la meilleure façon de procéder dans les applications des modèles de choix probit dépendra des dimensions relatives aux ensembles de choix et aux échantillons. Quant à l'emploi de simulateurs lisses de probabilités, il est fortement recommandé pour traiter les problèmes impliquant un nombre élevé de modalités possibles. Bref, compte tenu de l'ensemble des principaux résultats théoriques brièvement rappelés ici, quel type d'estimateurs basés sur ces simulateurs devrait être choisi par ceux qui utilisent le cadre MNP dans des problèmes de grande taille ? Pour le moment, on ne peut apporter une réponse définitive, mais quelques observations peuvent toutefois être énumérés. Par exemple, une différence entre les procédures MSS et MGS est que cette dernière exige une simulation additionnelle d'instruments, et parce que ceux-ci seront simulés avec un certain bruit ; il y aura donc une perte dans l'efficacité ; ces instruments étant en général difficilement calculables sur le plan numérique. Quant à la procédure MSS, elle utilise des instruments qui sont des valeurs observées de variables explicatives, mais les simulations associées à cette approche pourront s'avérer plus difficiles à effectuer puisqu'elles exigent des tirages à partir des distributions conditionnelles. Pour plus de détails sur ces problèmes numériques, on pourra se reporter au travail de Hajivassiliou et McFadden (1990) dans le cas des modèles à variables latentes limitées.

En ce qui concerne la méthode MVS, il apparaît évident qu'ici, dans un cadre entièrement paramétrique, cette méthodologie soit la plus attrayante puisqu'elle est asymptotiquement efficace. De ce fait, elle domine les deux autres méthodes. Cependant, son principal inconvénient réside dans la nécessité d'avoir recours à la fois à des échantillons de grande taille et à un grand nombre de tirages lorsque, dans les résultats de simulation, un certain degré de précision est requis car, ne l'oublions pas, c'est seulement dans cette méthode que peut surgir un biais asymptotique dû à la transformation logarithmique non linéaire.

De plus, contrairement à ce qui semble être admis, il existe d'autres aspects de la formulation du modèle MNP qui peuvent amoindrir le coût associé à l'estimation, en autant qu'on ait de bonnes raisons de croire que ces formulations ne feront pas trop violence aux résultats de la recherche empirique; car alors là, l'erreur de spécification encourue pourrait causer des dégâts à l'estimation des paramètres. En particulier, comme on l'a déjà vu, la formulation basée sur l'approche autorégressive généralisée telle que suggérée par Ben-Akiva et Bolduc (1991) est intéressante dans la mesure où les simulateurs lisses de type noyau s'introduisent naturellement dans ce cadre-ci. Elle évite de plus le problème de prolifération des paramètres de nuisance capturant la structure de covariances des erreurs; d'où la possibilité de réduction substantielle du nombre de paramètres à estimer.

Tout compte fait, parmi toutes ces stratégies d'estimation fondées sur des simulateurs lisses, c'est la méthodologie MVS qui, lorsqu'elle utilise les simulateurs lisses de type noyau et est conjuguée à l'approche autorégressive généralisée, devient alors la plus susceptible d'intéresser le praticien sur le plan numérique. Ben-Akiva et Bolduc (1991) ont d'ailleurs discuté du bon comportement numérique d'une telle méthodologie par rapport à celle du MGS dans le contexte d'un modèle de choix entre cinq services téléphoniques. Pour ce qui est de la méthode MSS, les résultats relatifs à cette méthode semblent être très prometteurs (voir Hajivassiliou et McFadden, 1990). Cependant, afin de juger de la pertinence de la spécification MNP fondée sur l'approche autorégressive généralisée, on devra en multiplier les applications.

Finalement, afin de bien saisir la véritable portée pratique des différentes méthodes d'estimation des modèles MNP évoqués dans ce texte, il est souhaitable de conduire d'autres expériences de simulations pour dégager leur domaine d'applicabilité dans des situations concrètes de choix polytomiques.

#### BIBLIOGRAPHIE

ALBRIGHT, R.L., S.R. LERMAN et C.F. MANSKI (1977), « Report on the Development of an Estimation Program for the Multinomial Probit Model », Préparé pour la Federal Highway Administration.

- BEN-AKIVA, M., et D. BOLDUC (1991), «Multinomial Probit with Autoregressive Error Structure», Département d'économique, Université Laval, Cahier 9123.
- BEN-AKIVA, M., et D. BOLDUC (1990), «Multinomial Probit with Taste Variation and Autoregressive Alternatives», 69th Annual Meeting, Transportation Research Board, Washington, Jan. 1990.
- BERKOVEC, J., et S. STERN (1991), «Job Exit Behavior of Older Men», *Econometrica*, Vol. 59, No 1, Janvier: 189-210.
- BERRY, S. (1989), «Estimation of a Model in the Airline Industry», W.P., Yale University.
- BÖRSCH-SUPAN, A., et V. HAJIVASSILIOU (1990), «Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models», W.P., Yale University.
- BÖRSCH-SUPAN, A., V. HAJIVASSILIOU, L.J. KOTLIKOFF et J.N. MORRIS (1990), «Health, Children and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors», NBER, W.P. 3343.
- BOLDUC, D. (1991), «Generalized Autoregressive Errors in the Multinomial Probit Model», à paraître, *Transportation Research B - Methodological*, Vol. 25, No 6.
- BOLDUC, D., et M. KACI (1991), «Multinomial Probit Models with Factor-Based Autoregressive Errors: A Computationally Efficient Estimation Approach», Document de travail, Département d'économique, Université Laval. Papier présenté aux IIèmes Journées des Jeunes Économètres, Toulouse, France, Avril 4-5, 1991.
- BOLDUC, D., et M. KACI (1989A), «Les méthodes d'évaluation des probabilités de choix dans un MNP: un survol de la littérature», Document de travail, Département d'économique, Université Laval.
- BOLDUC, D., et M. KACI (1989B), «On the Estimation of Probabilities in Multinomial Probit Models: An Empirical Comparative Study of Existing Techniques», 6th Annual Meeting of the Canadian Econometric Study Group, Hamilton, Ontario, Oct. 13-14, 1989.
- BUTLER, J.S., et R. MOFFIT (1982), «A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model», *Econometrica*, 50: 761-764.
- CLARK, C. (1961), «The Greatest of a Finite Set of Random Variables», *Operation Research*, 9: 145-162.
- DAGANZO, C. (1979), *Multinomial Probit*, Academic Press, New-York.
- DAGENAIS, M.G., et M. GAUDRY (1979), «The Dogit Model», *Transportation Research B*, 13, 2, décembre 1979: 105-112.
- DEÁK, I. (1980), «Fast Procedures for Generating Stationary Normal Vectors», *Journal of Statistical Computation and Simulation* 10: 225-242.

- GOURIÉROUX, C., et A. MONFORT (1991), « Simulation Based Inference in Models with Heterogeneity », *Annales d'Économie et de Statistique*, 20/21.
- GOURIÉROUX, C., et A. MONFORT (1989), « Statistique et modèles économétriques », *Economica*, Vol. I et II.
- HAJIVASSILIOU, V., et D. MCFADDEN (1990), « The Method of Simulated Scores for the Estimation of LDV Models with Application to External Debt Crises », Yale University, Working Paper.
- HAJIVASSILIOU, V., et D. MCFADDEN (1989), « Country Heterogeneity and External Debt Crises: Estimation by the Method of Simulated Moments », D.P., Yale University.
- HAJIVASSILIOU, V., D. MCFADDEN et P. RUDD (1991), « Simulation Of Multivariate Normal Rectangle Probabilities: Methods and Programs », Working Paper, Yale University.
- HAUSMAN, J.A. (1980), « Les modèles probit de choix qualitatifs », *Cahiers du séminaire d'économétrie*, 21: 11-31.
- HAUSMAN, J.A., et D.A. WISE (1978), « A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences », *Econometrica*, 46: 403-426.
- HOROWITZ, J., J.M. SPARMANN et C.F. DAGANZO (1982), « An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit Model », *Trans. Sci.* 16: 382-401.
- HOROWITZ, J. (1980), « Identification and Diagnosis of Specification Errors in the Multinomial Logit Model », *U.S. Environmental Protection Agency*, Washington, D.C. 20460.
- HOTZ, V.J., et S. SANDERS (1990), « The Estimation of Stochastic Dynamic Discrete Choice Models by the Method of Simulated Moments », D.P., University of Chicago.
- KAMAKURA, W.A. (1988), « The Estimation of Multinomial Probit Models: A New Calibration Algorithm », Mimeo, Vanderbilt University.
- LANGDON, M.G. (1984), « Improved Algorithms for Estimating Choice Probabilities in the Multinomial Probit Model », *Trans. Sci.* 18: 267-299
- MAHMASSENI, H.S., et S. LAM (1990), « Multinomial Probit Estimation: Numerical Experiments On a Supercomputer », 29th Joint National Meeting, TIMS/ORSA, Las-Vegas, 7-9 mai.
- MCFADDEN, D. (1989), « A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration », *Econometrica*, 57, 5: 995-1026.
- MCFADDEN, D. (1981), « Econometric Models of Probabilistic Choice », in MANSKI et MCFADDEN, *Structural Analysis of Discrete Data with Econometric Application*, MIT Press.

- MENDELL, N., et R.C. ELSTON (1974), « Multifactorial Qualitative Traits : Genetic Analysis and Prediction of Recurrence Risks », *Biometrics*, 30: 41-57.
- OWEN, D. (1956), « Tables for Computing Bivariate Normal Probabilities », *Annals of Mathematical Statistics*, 27: 1075-1090.
- PAKES, A., et D. POLLARD (1989), « Simulation and the Asymptotics of Optimization Estimators », *Econometrica*, 57.
- PEARSON, K. (1903), « On the Influence of Natural Selection on the Variability and Correlation of Organs », *Philosophical Transactions of the Royal Society of London*, A200: 1-66.
- RUDD, P. (1986), « On the Method of Simulated Moments for Estimation of Limited Dependent Variable Models », mimeograph, University of California at Berkeley.
- SHEFFI, Y., R. HALL et C.F. DAGANZO (1982), « On the Estimation of the Multinomial Probit Model », *Transportation Research* 16A: 447-456.
- SPARMANN, J. (1980), « Calibration of the Trinomial Model », Ph.D. Dissertation, Department of Civil Engineering, University of California, Berkeley.
- STERN, S. (1987), « A Method of Smoothing Simulated Moments of Probabilities in the Multinomial Probit Models », Working Paper, University of Virginia.
- VAN LIEROP, W. (1986), *Spatial Interaction Modelling and Residential Choice Analysis*, Aldershot, Gower.