HAL
open science

L'OPTION "ZERO" EN ANALYSE CONJOINTE:Une nouvelle spécification de l'indécision et du refus. Application au marché de la Vidéo à la Demande

Silva Ohannessian

To cite this version:

Silva Ohannessian. L'OPTION "ZERO" EN ANALYSE CONJOINTE:Une nouvelle spécification de l'indécision et du refus. Application au marché de la Vidéo à la Demande. domain_other. Conservatoire national des arts et metiers - CNAM, 2008. Français. NNT : . tel-00265910

HAL Id: tel-00265910

https://theses.hal.science/tel-00265910v1

Submitted on 20 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS
PARIS

Thèse

Pour obtenir le grade de

DOCTEUR DU CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS

Spécialité : Sciences de gestion

Silva OHANNESSIAN

L'OPTION "ZÉRO" EN ANALYSE CONJOINTE : Une nouvelle spécification de l'indécision et du refus. Application au marché de la Vidéo à la Demande

Soutenue publiquement le 10 mars 2008

Jury :

M. Gilbert SAPORTA	Professeur, CNAM	Directeur de thèse
M. Pierre CAZES	Professeur, Université Paris-Dauphine	Rapporteur
M. Giuseppe GIORDANO	Professeur, Université de Salerne	Rapporteur
M. Jean-Paul AIMETTI	Professeur, CNAM	Examinateur
M. Oliviero MARCHESE	Directeur général, IPSOS Observer	Examinateur
M. Alain MONFORT	Professeur, CNAM	Examinateur

L'OPTION "ZÉRO" EN ANALYSE CONJOINTE :
Une nouvelle spécification de l'indécision et du refus.
Application au marché de la Vidéo à la Demande

Résumé

Cette thèse étudie l'option "zéro" dans l'analyse conjointe. Cette dernière s'assimile à un non choix des scénarios présentés lors d'enquêtes. Ce non choix est représenté par deux concepts distincts qui sont le refus et le conflit. Le premier concerne l'inappréciation des produits, alors que le second traite de l'incertitude dans les préférences et les choix. Une nouvelle modélisation économétrique de ce non choix est proposée dans ce travail, en supposant un mélange de spécification des comparaisons d'utilité et des seuils de réaction. Ce mélange permet de n'associer des utilités qu'aux produits, et de les comparer aux seuils de l'option "zéro". Ces comparaisons déterminent, ainsi, les situations de non choix, sans pour autant postuler une utilité au refus et au conflit.
Une étude sur le marché de la Vidéo à la Demande y est réalisée. Les résultats sont obtenus en utilisant l'approche bayésienne sur les modèles individuels et le maximum de vraisemblance sur le modèle agrégé. Les estimations sont cohérentes avec la réalité et la significativité des valeurs du refus et du conflit démontre leur importance dans la prise de décision.

Mots-clés : spécification de l'option "zéro", inappréciation des produits, indécision dans le choix, séparation et non convergence, approche bayésienne, simulations de Monte Carlo, parts de marché

"ZERO" OPTION IN CONJOINT ANALYSIS :
A new specification of the indecision and the refusal.
Application to the Video on Demand market

Abstract

This thesis undertakes a study about the "zero" option in conjoint analysis. The "zero" option relates to the no choice of products presented to individuals within the frame of a survey. This no choice embeds two distinct concepts, the refusal and the conflict. The first represents the inappreciation of products while the second is defined by the preference and choice uncertainty. This work proposes a new econometric specification of the no choice by assuming a mix of utilities maximisation and ordered response models. This mix only associates utilities with products and compares them to the "zero" option thresholds. These comparisons lead to no choice situations without linking utilities to refusal and conflict.
A study on the Video on Demand market has been conducted. The results are obtained by applying a bayesian approach in the case of individual models, and the maximum likelihood in the case of aggregate models. The estimates fit the reality and the significance of the refusal and the conflict demonstrates the importance of these variables in decision making.

Keywords : "zero" option specification, inappreciation of the products, indecision in the choice, separation and non convergence, bayesian approach, Monte Carlo simulations, market shares

Remerciement

Je tiens à remercier en premier lieu et tout particulièrement le directeur de ma thèse, le Professeur Gilbert Saporta du CNAM de Paris, qui a toujours cru en mon travail, et qui a ainsi permis son aboutissement.

Je tiens également à exprimer ma reconnaissance aux rapporteurs, les Professeurs Pierre Cazes et Giuseppe Giordano, pour avoir accepté d'étudier attentivement mon travail, ainsi qu'aux autres membres du jury pour l'intérêt qu'ils ont montré en acceptant de participer à ma soutenance de thèse.

Une attention toute particulière se porte également à Olivier Monti qui a réalisé la quasi-totalité des travaux de programmation sur Matlab et TSP-GiveWin de ma thèse. Sans son aide précieuse, le résultat de cette étude n'aurait pu être tel qu'il est présenté dans ce doctorat.

Je souhaite également remercier le Professeur Alain Monfort du CNAM de Paris qui m'a apporté une aide précieuse d'un point de vue économétrique, le Professeur Gérard Antille, de l'Université de Genève, qui m'a permis de faire passer mon questionnaire à ses étudiants de première année en Sciences Sociales, le Professeur Per-Olov Johansson, de Stockholm School of Economics and University of Oslo, qui m'a suggéré, au début de mon doctorat, une référence très utile dans le cadre de mon sujet.

Finalement, je remercie toute ma famille et Olivier Monti du soutien moral qu'ils m'ont apporté tout au long de mon doctorat. Je remercie également Philippe Posta pour son aide et ses suggestions dans cette thèse.

Table des matières

Introduction **1**

1 Notation-Classement versus Choix **7**

Introduction 7

1.1 Préférences par la notation ou le classement 9

1.2 Préférences par le choix 12

Conclusion 19

2 Les plans d'expérience **21**

Introduction 21

2.1 Plans optimaux pour modèles linéaires 22

2.2 Exemple de plans factoriels complets et de plans orthogonaux fractionnaires 27

2.3 Plans en carré latin 34

2.4 Plans en carré gréco-latin 35

2.5 Plans en bloc incomplets équilibrés 36

2.6 Plans en carré de Youden 39

2.7 Plans D-optimaux 40

2.8 Plans pour modèles non linéaires 42

2.9 Plans en bloc incomplets équilibrés associés aux plans D-optimaux 42

2.10 Problèmes liés au choix des paires 44

Conclusion 45

3 Modèle de choix de l'analyse conjointe et l'option "zéro" **47**

Introduction 47

3.1 Modèles multinomiaux logit : formulation générale 48

3.1.1 Probabilités du modèle logit : source 48

3.1.2 Probabilités du modèle logit : démonstration 51

3.1.3 La fonction de vraisemblance du modèle logit 54

3.1.4 Résolution par la méthode du gradient 56

3.2 Le modèle MNL de l'analyse conjointe et l'option "zéro" 57

3.2.1 Le modèle MNL 57
3.2.2 La fonction de vraisemblance 60
3.2.3 Formulation matricielle 61
3.2.4 L'option "zéro" dans la littérature 64
3.3 Les raisons psychologiques du non choix 67
3.4 Le modèle de l'option "zéro" de l'analyse conjointe 69
3.4.1 Les modèles censurés 70
3.4.2 Nouvelle spécification de l'option "zéro" 72
3.4.3 Représentation graphique du modèle 72
3.4.4 Les probabilités de choix 76
3.4.5 Les probabilités de non choix 80
3.4.6 La fonction de vraisemblance 84
3.4.7 Identification des paramètres structurels 85
3.4.8 Modèle agrégé de l'option "zéro" 85
3.5 Cas spéciaux 86
3.5.1 Modèle où y=1,2,3 88
3.5.2 Modèle où y=1,2,4 90
3.5.3 Modèle où y=1,3,4, respectivement y=2,3,4 93
3.5.4 Modèle où y=1,2 97
3.5.5 Modèle où y=1,3, respectivement y=2,3 99
3.5.6 Modèle où y=1,4, respectivement y=2,4 103
3.5.7 Modèle où y=3,4 107
Conclusion 109

4 Notions de convergence et de séparation dans la méthode du maximum de vraisemblance 113
Introduction 113
4.1 Notions de convergence et de séparation 114
4.2 L'influence de la séparation sur les tests statistiques et les intervalles de confiance 120
4.3 Solutions 121
4.3.1 Modification des variables ou des données 122
4.3.2 L'inférence exacte 122
4.3.3 L'approche bayésienne 124
4.3.4 La vraisemblance pénalisée ou l'*a priori* de Jeffreys 128
Conclusion 132

5 Application de l'option "zéro" sur le marché de la VoD **133**
Introduction . . . 133
5.1 La Vidéo à la Demande (VoD) . . . 134
5.1.1 Gestion numérique des droits (Digital Rights Management DRM) . 137
5.1.2 Intérêt de l'analyse de la Vidéo à la Demande . . . 138
5.1.3 Présentation de certains sites de VoD . . . 142
5.1.4 Les autres acteurs de la VoD . . . 150
5.1.5 Présentation de la VoD proposée par les FAI . . . 152
5.1.6 Les motivations et les raisons du choix des attributs de la VoD . . . 153
5.2 Provenance de l'échantillon . . . 162
5.3 Le plan d'expérience adapté à l'application . . . 163
5.4 Mise en oeuvre de l'interview . . . 172
5.4.1 Le questionnaire . . . 172
5.4.2 L'interview . . . 174
5.4.3 Résultat de l'interview et données retenues . . . 174
5.5 Application du modèle de l'option "zéro" avec le refus et le conflit . . . 176
5.5.1 Le maximum de vraisemblance . . . 176
5.5.2 Vérification de la séparation . . . 176
5.5.3 L'approche bayésienne et les simulations de Monte Carlo dans l'option "zéro" . . . 179
5.6 Résultats . . . 183
5.6.1 Modèle y=1,2 . . . 184
5.6.2 Modèle y=1,2,3,4 . . . 187
5.6.3 Modèle y=1,2,3 . . . 190
5.6.4 Modèle y=1,2,4 . . . 192
5.6.5 Modèle y=1,3 . . . 194
5.6.6 Le modèle agrégé . . . 195
5.7 Les parts de marché . . . 198
5.7.1 Les probabilités de choix . . . 200
5.7.2 Les parts de marché . . . 204
Conclusion . . . 207

Conclusion et perspectives **211**

Annexes **217**
Annexe 1 . . . 217
Annexe 2 . . . 218
Annexe 3 . . . 219

Annexe 4 221
Annexe 5 225
Annexe 6 237
Annexe 7 238

Bibliographie **239**

Introduction

L'analyse conjointe est une méthode de calcul d'utilités partielles basée sur les réponses des individus par rapport à des combinaisons d'attributs dits de décision. Cette méthode nécessite la mise en oeuvre d'un questionnaire, l'utilisation de plans d'expérience et la détermination des utilités partielles des attributs par une procédure adaptée d'estimation. C'est pourquoi elle se définit souvent comme une méthode d'étude intégrée. De plus, contrairement à la microéconomie qui traite des préférences agrégées des consommateurs, elle s'intéresse aux préférences individuelles. Elle agrège, par la suite, ces préférences dans le cadre d'étude de parts de marché, par exemple.

A l'origine, l'analyse conjointe a été développée par le psychologue Luce et le statisticien Tukey (1964) dans les années 60. Elle est un terme générique adopté par Green et Srinivasan (1978). Au départ, elle était essentiellement appliquée en marketing. Son introduction dans ce domaine est apparue dans les années 70 (Green et Rao 1971 ; Green et Wind 1973) avant de voir son application commerciale se développer dans les années 80. En 1982 et en 1986, Cattin et Wittink ont publié des applications de cette méthodologie. Aujourd'hui, les techniques de l'analyse conjointe se répandent dans de nombreux domaines, tels que la géographie, le transport, l'organisation urbaine, la sociologie, etc., bien qu'elles étaient déjà bien connues et appliquées en psychologie avant même leur introduction en marketing (Anderson 1962, 1970 ; Hoffman 1960 ; Hoffman, Slovic et Rorer 1968).

L'analyse conjointe est comparable à un processus qui intègre la prise de décision souvent complexe du consommateur [1]. Ce processus se décrit, généralement, par des étapes d'intégration d'information, d'appréciation, de comparaison et d'évaluation des produits afin d'en décider les aspects importants et pouvoir juger et comparer sur ces critères le produit à choisir parmi tous les produits compétitifs. Ce concept comportemental de l'analyse conjointe provient de la théorie de l'intégration d'information (IIT) en psychologie (Hammond 1955 ; Slovic et Lichtenstein 1971 ; Anderson 1970, 1981, 1982) pour les

[1] L'analyse conjointe a ses origines théoriques dans la littérature psychologique du processus d'information et de la prise de décision complexe.

modèles descriptifs du développement d'information dans le jugement (Louviere 1988 a, b). De plus, il provient également de la théorie de la consommation de Lancaster (1966) qui considère les biens économiques comme des groupes de caractéristiques ou d'attributs et qui suggère que les préférences pour ces biens peuvent être décomposées en des préférences séparables des caractéristiques ou des avantages des composantes. En d'autres termes, cette théorie suppose qu'un produit peut se décomposer en plusieurs caractéristiques ou attributs et que les préférences des consommateurs pour un produit dépendent de la façon dont les attributs sont valorisés par les consommateurs et se combinent pour déterminer sa préférence globale. L'objectif principal de l'analyse conjointe est, donc, de décomposer l'utilité d'un produit (ou plus généralement d'un objet ou d'une offre) pour calculer les utilités partielles des attributs.

Une autre source essentielle de la théorie comportementale de l'analyse conjointe est la théorie de l'utilité aléatoire (RUT) qui forme la base de plusieurs modèles et plusieurs théories de jugement du consommateur et de la prise de décision en psychologie et en économie. Cette théorie fut, en premier, proposée par Thurstone (1927) comme une manière de modéliser les jugements dominants dans un contexte de comparaison par paires. L'idée de base de Thurstone, bien que désormais élargie, est restée relativement inchangée. En réalité, la théorie de l'utilité aléatoire suggère que les consommateurs choisissent (achètent, etc.) les alternatives qu'ils préfèrent selon les contraintes auxquelles ils sont soumis. Cependant, d'un point de vue du chercheur, les consommateurs ne choisissent pas forcément le produit qu'ils apprécient le plus. En effet, certains de leurs choix changent selon les opportunités du marché. Ces changements dans le comportement sont représentés dans la théorie RUT par une composante aléatoire dans la fonction d'utilité du consommateur. La théorie de l'utilité aléatoire fournit donc une structure générale de compréhension et de modélisation du comportement du marché des consommateurs dans beaucoup de circonstances.

En France, le terme "trade-off", qui signifie compromis en anglais, est utilisé comme un synonyme de l'analyse conjointe. La méthode dite "trade-off" englobe la catégorie des modèles compensatoires ou de compromis. Cette catégorie suppose qu'une note positive sur un attribut peut être compensée par une note moins bonne sur un autre. En d'autres termes, le choix du consommateur suit une logique de compensation, où l'avantage retiré par un attribut est compensé par les autres attributs qu'il est amené à rejeter. L'objet de l'analyse conjointe représente donc l'analyse des compromis qui mène à un choix. En effet, dans l'analyse conjointe, on présente généralement des produits qui obligent le consommateur à faire des compromis entre plusieurs attributs avantageux. De plus, puisque le modèle "trade-off" ressort du domaine particulier de la mesure des phénomènes psycholo-

giques qu'on identifie sous le terme de "mesure conjointe", la méthode regroupe finalement l'ensemble des procédures métriques qui visent à apprécier l'effet simultané de plusieurs variables indépendantes sur une même variable dépendante.

Dans une acception restreinte, l'approche "trade-off" correspond également à un mode d'évaluation des produits, qui consiste à comparer deux à deux les attributs représentés généralement par une matrice croisant les deux ensembles de modalités (niveaux d'attributs). Les modes d'évaluation, tels que les classements ou les notations de préférences sur une échelle fixée au préalable, sont les plus fréquemment mentionnés dans la littérature. Il existe d'autres méthodes de collectes, à savoir la méthodes des profils complets (Green et Rao 1971), la méthode de la matrice d'équilibre (Johnson 1974), la méthode de comparaison par paires, ACA ("Adaptive Conjoint Analysis")[2], ou l'analyse conjointe hybride (Green, Goldberg et Montemayos 1981 ; Green 1984). Un autre mode d'évaluation des scénarios, qui apparut dans les années '80, fut les préférences par un choix unique de produits. En effet, l'article de 1983 de Louviere et Woodworth intégrait le choix discret à l'analyse conjointe. Cette approche a été utilisée avec succès dans des domaines, tels que la conception de nouveaux produits (Louviere et Woodworth 1983), l'optimisation d'un produit (Green et Krieger 1993) et l'évaluation des ensembles compétitifs (Mahagan, Green et Goldberg 1982 ; Louviere et Woodworth 1983). Son concept a l'avantage de mieux représenter la situation réelle du marché. En effet, le choix d'un produit parmi plusieurs alternatives ressemble davantage au processus d'achat réel des consommateurs que le classement ou la notation. De plus, il existe des modèles logit multinomiaux introduisant l'option de non choix (non achat) ou de décision de retarder son choix, dite option "zéro", qui permet de rendre aux consommateurs l'expérience d'achat encore plus réelle. Toutefois, cette dernière n'a pas été énormément abordée dans la littérature actuelle.

La finalité de l'analyse conjointe est la détermination des utilités partielles qui composent l'utilité globale. Ces utilités partielles s'estiment différemment selon le mode de collecte des données. Green et Srinivasan (1978) classifient les méthodes d'estimation des utilités des attributs en trois catégories. Dans le cas d'échelle de mesure d'intervalle, l'estimation du modèle se fait par l'analyse de variance. Elle permet de définir la forme de l'utilité globale par rapport aux utilités partielles, à savoir une forme additive, ou s'il existe des interactions entre les attributs. Les variables métriques sont, généralement, recueillies en demandant aux consommateurs de noter sur une échelle fixe les différents produits ("rating"). Dans le cas d'échelle de mesure ordinale, telle qu'un classement ("ranking"), l'utilisation d'algorithmes, tels que MONANOVA (Kruskal 1965) pour un modèle additif, LINMAP (Srinivasan et Shocker 1973 a/b ; Pekelman et Sen 1974) pour un modèle

[2]ACA, une méthode créée en 1985 par Johnson, est une marque de la société Sawtooth Software.

avec interactions, ou des méthodes d'analyse des préférences, telles que PREFMAP (Carroll 1972) pour un modèle avec interactions, est appropriée. Néanmoins, malgré la nature ordinale des classements, l'utilisation des moindres carrés ordinaires (OLS) convient également. En effet, les comparaisons des estimateurs obtenus par MONANOVA et ceux obtenus par les OLS [3] semblent fournir des résultats presque identiques (Green et Srinivasan 1978). Wittink et Cattin (1981) montrèrent également qu'il n'y avait pas beaucoup de différences entre la méthode OLS, MONANOVA et LINMAP. Toutefois, la méthode OLS ne peut pas être utilisée pour estimer les effets d'interactions à partir de données de classement si les modèles additifs sont inappropriés. Dans le cas de données issues de la comparaison par paires ou d'appréciation par un oui ou par un non d'un produit, le modèle logit binaire ou le modèle probit binaire peuvent être appliqués.

Les modèles logit multinomiaux (MNL) (Louviere et Woodworth 1983 ; Elrod, Louviere et Davey 1992) sont, quant à eux, utilisés lorsque la variable dépendante est une variable de choix parmi plusieurs alternatives. Ces modèles MNL dans le cas d'estimation agrégée ont certaines limites. En effet, ils supposent l'homogénéité des consommateurs. C'est pourquoi, dans les années 1994-1995, les modèles de classe latente ont été développés (Kamakura, Wedel et Agrawal 1994 ; De Sarbo, Ramaswamy et Cohen 1995) afin d'intégrer l'hétérogénéité des consommateurs. Les modèles hiérarchiques de Bayes (HB) prennent également en compte l'hétérogénéité des consommateurs. Ces modèles se décrivent en deux niveaux, dont le plus élevé émet l'hypothèse d'une distribution normale des utilités partielles avec une espérance de a et une matrice de variance-covariance de Σ ($N(a;\Sigma)$). Le niveau plus bas suppose, quant à lui, que pour une utilité partielle donnée, la probabilité de choisir un produit ou de noter les marques d'une certaine manière est gouvernée par un modèle particulier, tel que le modèle logit multinomial ou celui de la régression linéaire. Le principe général de cette méthode est d'"emprunter" l'information des autres répondants pour calculer des résultats individuels relativement stables des observations multiples (par exemple, tester plusieurs produits). Les modèles HB ont certains avantages par rapport à ceux de classe latente et ceux de régression. En effet, contrairement aux modèles HB, les résultats obtenus par les classes latentes sont sujets à un minimum local. De plus, dans l'estimation par les classes latentes, l'analyste doit, comme dans toutes les méthodes de classification, choisir le nombre de segments appropriés. Un désavantage des modèles de régression par rapport aux modèles HB est l'apparition fréquente de colinéarités. Les résultats d'estimations obtenus par les modèles hiérarchique de Bayes sont alors plus précis.

[3] En régressant les classements par rapport à une matrice de variables indicatrices représentant les niveaux d'attributs

Une autre hypothèse forte des modèles MNL est la propriété IIA ("Independence of Irrelevant Alternative"), qui suppose un rapport des probabilités $\frac{P_h}{P_l}$ indépendant des autres probabilités restantes. Cette propriété est issue des hypothèses d'indépendance des erreurs entre les alternatives, les ensembles de choix, les individus et de variance des erreurs constante (IID "Identiquement et Indépendamment Distribué"). Selon plusieurs auteurs, cette propriété peut limitée l'utilisation du modèle MNL dans les applications marketing (Srinivasan 1980 ; Kamakura et Srivastava 1984 ; Louviere et Timmermans 1990). Le modèle multinomial probit (MNP) permet de s'affranchir de cette hypothèse, malgré une complexité accrue du traitement informatique. De plus, il convient encore de mentionner les modèles de valeur extrême hétéroscédastique HEV qui ne suppose pas une distribution IID des erreurs comme les modèles logit multinomiaux.

A la lumière de ces éléments, il apparaît clairement que chaque méthode a ses avantages et ses inconvénients et qu'il n'est *a priori* pas possible de statuer l'une par rapport à l'autre. Néanmoins, de nouveaux éléments modifiant certains des modèles de l'analyse conjointe ont récemment été introduits. Ils n'ont, cependant, pas énormément été abordés dans la littérature. Il existe, parmi ces nouveautés, l'introduction de l'option "zéro" dans les modèles logit multinomiaux. Cette option permet à l'individu de retarder son choix ou de ne pas sélectionner un produit. Sa mise en oeuvre n'est pas clairement définie dans la littérature. L'objet de cette thèse est donc d'approfondir ce domaine de l'option "zéro" en décrivant sa méthode d'application. De plus, nous ajoutons une composante supplémentaire à l'option "zéro" de la littérature, à savoir l'indécision du consommateur dans le choix des produits similaires. En effet, l'option "zéro" de la littérature ne considère que les non choix liés à l'inappréciation des produits. Or, dans la littérature psychologique, lorsque des produits sont très similaires, il arrive que le consommateur n'en sélectionne aucun. La raison d'un tel comportement est l'incertitude dans le choix des produits. Par conséquent, le consommateur préfère ne pas sélectionner de produits tant qu'il n'est pas sûr de son choix. La littérature psychologique fait remarquer que cet aspect du non choix n'existe pas dans les modèles économétriques des préférences des consommateurs. C'est pourquoi, nous le spécifions dans notre nouveau modèle de l'option "zéro".

La thèse se décompose en cinq chapitres. Elle aborde, premièrement, les préférences selon le mode d'évaluation classique en comparaison au mode d'évaluation par le choix. Deuxièmement, elle donne un aperçu des différents plans d'expérience utilisés dans la détermination des scénarios ou des produits, utiles dans des études en analyse conjointe. Troisièmement, elle décrit l'analyse conjointe des préférences par choix unique et elle introduit l'option "zéro" dans ces modèles. Quatrièmement, elle aborde les problèmes liés à la séparation des données et à la non convergence du modèle estimé par le maximum de

vraisemblance (MV). Finalement, une application sur le marché de la Vidéo à la Demande (VoD) de notre nouvelle spécification de l'option "zéro" est effectuée sur un échantillon d'étudiants universitaires suisses de première année. Cette étude a pour but de constater l'efficacité de notre modèle, qui intègre aussi bien le non choix de la littérature, c'est-à-dire l'inappréciation des produits, que le non choix associé à une indécision dans les préférences du consommateur.

Chapitre 1

Notation-Classement versus Choix

Introduction

L'analyse conjointe peut se définir de deux manières selon le mode de recueil de données envisagé. La première, notation-classement ("rating-ranking"), modélise les préférences du consommateur par un modèle linéaire. La variable dépendante, représentant les préférences, est soit ordinale (classement ou "ranking") soit métrique (notation ou "rating").

La deuxième approche, le choix, se caractérise par un modèle non linéaire et une variable dépendante qualitative de choix (variable nominale). Cette variable est définie par un ou plusieurs choix uniques par individu. Dans le cas d'un choix unique, nous présentons à l'individu tous les scénarios envisagés et nous lui demandons d'en choisir un. Dans le cas de plusieurs choix par individu, nous groupons les scénarios en plusieurs paires, triplets, etc. et nous demandons à l'individu de faire un choix par paires (comparaison par paires), par triplets, etc. Le résultat est donc plusieurs choix binaires ou multiples selon le groupage défini. Dans ce chapitre, lorsque nous abordons la méthode de choix, nous nous référons essentiellement au choix dans le cas d'estimation agrégée. La raison provient du fait que la littérature traitant ce sujet est plus fréquemment axée sur le cas agrégé.

Ces deux approches se distinguent également par leur mode d'estimation. La première approche utilise généralement la méthode des moindres carrés ordinaires pour estimer les paramètres du modèle de l'analyse conjointe. La deuxième approche utilise la méthode d'estimation du maximum de vraisemblance. Le maximum de vraisemblance s'applique à un modèle probit binaire ou multinomial, selon le groupage, lorsque l'erreur du modèle non linéaire est supposée normale, et à un modèle logit binaire ou multinomial lorsque l'erreur suit une loi exponentielle.

L'information fournie lors du recueil des données diffère d'une approche à l'autre. La notation est le mode de recueil théoriquement le plus riche en information. En effet, une variable dépendante d'intervalle renseigne davantage sur les goûts des individus, car elle fournit l'ordre et le niveau des préférences. Quant au classement, il est moins riche que la notation, car il ne fournit que l'ordre. Toutefois, il est plus riche que le choix unique ou par paires, par triplets, etc. (choix groupé). Entre le choix groupé et non groupé, l'apport d'information n'est pas similaire. En effet, la comparaison par paires, triplets, etc. fournit un ordre partiel en ne présentant, généralement, pas toutes les paires, triplets, etc. de combinaisons des produits aux consommateurs, alors que le choix unique ne fournit aucun ordre. De plus, compte tenu du faible degré d'information associé au choix unique, l'estimation des paramètres associés à ce type de données est généralement sous forme agrégée. Elle peuvent toutefois être sous forme individuelle. Quant aux autres types de recueil de données, leurs estimations sont de type individuel pour être ensuite agrégées.

Un autre point qui distingue l'approche de la notation à celle du classement est la possibilité d'obtenir des ex-aequo dans le cas de la notation. En effet, le consommateur peut noter de la même manière des produits, alors qu'avec le classement, il a l'obligation d'ordonner les scénarios. Un autre inconvénient de l'approche par la notation est le fait que les individus n'utilisent pas toujours la même amplitude pour noter les scénarios. En effet, certains basent leur notation sur une échelle de mesure comprise entre 1 et 10, alors que pour d'autres, cette mesure se situe entre 1 et 50, et ce malgré une échelle de mesure fixe variant entre 1 et 100, par exemple. Par conséquent, l'agrégation de ce type de résultats ne peut être envisageable, compte tenu des échelles de mesure différentes. L'agrégation pourrait être, cependant, considérée à condition que les résultats soient normalisés. Ainsi, nous pourrions obtenir des moyennes de préférences basées sur l'approche de la notation.

Dans ces deux approches, les scénarios présentés aux consommateurs peuvent être soit complets, soit partiels. Les scénarios complets consistent à définir les produits en fonction de tous ces attributs. Les méthodes de notation et de classement de la première approche et celles de la comparaison par paires, par triplets, etc. et de choix unique de la deuxième approche utilisent ce principe de scénarios complets. Les scénarios partiels n'affichent, quant à eux, pas tous les attributs dans leurs produits. Les méthodes de "trade-off" de la première approche et celles de ACA de la deuxième approche présentent leur scénario en ne les définissant qu'en fonction de certains attributs. ACA est plus complexe que la méthode du "trade-off" et est brièvement abordée dans la suite du travail.

Dans ce chapitre, nous décrivons ces approches de notation-classement et de choix de l'analyse conjointe. Une section présente brièvement les méthodes que nous associons à

l'approche notation-classement. Une autre section décrit l'approche de choix dans le cas d'un modèle agrégé.

1.1 Préférences par la notation ou le classement

L'approche notation-classement se présente sous de nombreuses variantes dans la littérature. Green et Srinivasan (1978) développèrent une procédure commune à ces méthodologies. Elle se résume en sept étapes. Les différentes étapes et décisions leur étant associées ne peuvent être réalisées indépendamment l'une de l'autre.

La première étape sélectionne la fonction de préférence. Il existe trois modèles de préférence, à savoir le modèle des utilités partielles, le modèle du vecteur idéal et le modèle du point idéal. Le modèle du vecteur idéal admet que les utilités partielles sont liées linéairement aux niveaux d'attributs. Le modèle du point idéal suppose, quant à lui, l'existence d'un point dit "idéal" (ou maximum). Cette hypothèse se traduit par une forme quadratique entre les utilités partielles et les niveaux d'attributs. Ces deux modèles sont des cas particuliers du modèle d'utilités partielles. En effet, le modèle d'utilités partielles ne suppose pas de processus fonctionnel spécifique et les niveaux d'attributs ne peuvent qu'être interpolés si le niveau d'échelle est métrique. Le modèle d'utilités partielles est principalement utilisé comme un modèle de préférences pour l'analyse conjointe. Il est formellement identique à celui du vecteur idéal, lorsque les attributs sont qualitatifs. Une règle de composition des utilités partielles, obtenues par la fonction de préférence, est exigée pour déterminer l'utilité globale, à savoir les modèles compensatoires et non compensatoires. Les modèles non compensatoires ne seront pas abordés dans cette thèse.

La deuxième étape consiste à sélectionner une méthode de collecte des données parmi plusieurs techniques, telles que les profils complets (ou l'évaluation de facteurs multiples (MFE)), l'évaluation par un facteur de l'approche "self-explicated", celle par deux facteurs de l'approche "trade-off", les comparaisons par paires et ACA. La méthode ACA est une combinaison de deux méthodes. La première (évaluation de l'attribut seul) demande aux répondants de classer ou de noter selon ses préférences les niveaux par attribut et les attributs eux-mêmes. Des attributs naturellement ordonnés, tel qu'un prix, ne sont ni classés ni notés par les consommateurs. Le but de cette première approche est donc de définir un ordre sur les niveaux et les attributs et d'estimer les utilités partielles dites initiales. La deuxième méthode, quant à elle, raffine ces estimations initiales en utilisant le principe des comparaisons par paires. Ce principe permet aux consommateurs d'évaluer plusieurs ensembles de deux produits qui sont composés de tous les attributs. Dans ACA, par contre, les produits évalués ne sont pas toujours composés de tous les attributs,

mais varient durant l'interview adaptative [1]. Cette étape de ACA amène les individus à augmenter leur attention sur des attributs moins importants, car comparer des différences d'attribut est généralement plus facile.
Dans l'évaluation par un facteur de l'approche self-explicated, les individus évaluent les niveaux des attributs de manière individuelle en les notant sur une échelle fixe ou en les classant. Ensuite, une pondération de chaque attribut est déterminée grâce à l'utilisation d'échelle continue ou par l'affectation de points. Cette méthode self-explicated donne donc une bonne introduction des niveaux d'attributs et permet de considérer plus d'attributs. Finalement, l'utilité d'un produit est la combinaison pondérée de cette évaluation des niveaux. Cette approche ne calcule donc pas les utilités partielles. Elle estime directement l'utilité globale. De plus, l'approche self-explicated est en contradiction avec l'idée de l'analyse conjointe, qui est de mesurer conjointement l'effet de variables indépendantes par des compromis entre les facteurs ou attributs.
Quant à la méthode des profils complets, elle évalue les produits ou scénarios définis par tous ses attributs. La méthode par le classement, par la notation sur une échelle fixe, par les comparaisons par paires et par l'appréciation des profils par un "oui, ce produit me plaît" ou un "non, il ne me plaît pas" font partie de l'approche des profils complets. Par contre, la méthode par deux facteurs de l'approche "trade-off" (ou l'évaluation par deux facteurs (TFE)) décrit les profils à travers deux attributs et fait donc partie des profils partiels. Néanmoins, la méthode des comparaisons par paires est également applicable avec des profils partiels, comme c'est le cas dans ACA.

La méthode des profils complets a connu des améliorations durant le milieu des années '80. La nouvelle approche de ce type de collecte de données, appelée l'analyse conjointe hybride, combine une évaluation directe de l'importance des attributs individuels par les consommateurs et une interview classique de l'analyse conjointe sur les combinaisons des attributs choisis.

Dans la troisième étape, il faut choisir entre le plan d'expérience des profils complets ou celui des profils réduits. Le nombre de combinaisons des niveaux d'attributs n'augmentant pas de manière linéaire, le total des produits évalués par les individus peut donc être très élevé. D'un point de vue empirique, 30 scénarios semblent représenter une limite maximale (Green et Srinivasan 1978). Toutefois, dans la pratique, la qualité des résultats est correcte seulement si la limite est fixée à 16 scénarios. Si le nombre de combinaisons des niveaux d'attributs n'est pas trop élevé, les plans factoriels complets conviennent. En

[1] ACA est une interview interactive assistée par ordinateur et spécifique à chaque individu. Elle crée les questions en sa basant sur les réponses et elle estime les utilités partielles à l'aide d'un algorithme itératif.

effet, ils ont l'avantage de conserver tous les scénarios théoriquement concevables. Les plans d'expérience réduits, ou plans factoriels fractionnaires, permettent de diminuer le nombre de scénarios, tout en conservant l'orthogonalité des attributs.

La quatrième étape définit la manière de présenter les scénarios. Elle peut être verbale ou visuelle. Le mode verbal présente les scénarios sur une feuille décrivant le produit à partir de mots clefs, de phrases descriptives ou des deux à la fois. Le mode visuel utilise une représentation graphique (un dessin ou une photo) ou une représentation physique (un prototype ou un produit).

La cinquième étape détermine la manière d'interviewer les consommateurs. Les entretiens sont principalement des face à faces entre le répondant et l'enquêteur, des enquêtes par mail ou par téléphone ou des questionnaires informatiques.

La sixième étape correspond au mode d'évaluation des produits. Ces modes se partagent en une catégorie d'échelle métrique et non métrique. La mesure métrique est essentiellement [2] une note ("rating") attribuée aux produits sur une échelle fixe. L'échelle non métrique correspond, quant à elle, aux classements ("ranking"), aux comparaisons par paires et aux appréciations du type "ce produit me plaît ou ne me plaît pas". Pour les classements, l'interviewé doit classer par ordre de préférence les différents produits sélectionnés par un plan d'expérience adapté. La comparaison par paires consiste, quant à elle, à comparer deux à deux des profils décrits par tous les attributs. Ces paires de profils sont fournies à l'aide des plans d'expérience. Cette approche permet d'attirer l'attention du consommateur sur les différences de niveaux d'attributs de deux scénarios. Par conséquent, en évaluant ces différences, l'individu ne considère plus le niveau de manière individuelle. Cette méthode tend finalement à réduire l'importance individuelle des niveaux. L'évaluation des profils par un "oui, ce produit me plaît" ou un "non, il ne me plaît pas" consiste à donner une valeur binaire à chaque produit.

La dernière étape estime les utilités partielles selon le mode d'évaluation sélectionné. Selon une mesure métrique, l'estimation se fera par les moindres carrés ordinaires (OLS) ou une analyse de variance (ANOVA). Pour une mesure non métrique de classement, les procédures d'estimation seront des algorithmes tels que PREFMAP, une analyse des préférences développée par Carroll (1972), LINMAP, programmé par Srinivasan et Shocker (1973a, 1977) et POLYCON, introduit par Young (1972). Il existe également une approche non métrique MONANOVA qui a été développée par Kruskal (1965) dans le cadre de l'analyse conjointe. Les méthodes des moindres carrés ordinaires (OLS) fonctionnent également,

[2] Il existe d'autres mesures métriques. Toutefois, elles ne seront pas abordées dans la thèse.

malgré la nature ordinale de la variable dépendante. Toutefois, dans la classe des variables d'intervalle et de classement, les techniques d'estimation les plus fréquemment utilisées sont MONANOVA, LINMAP et OLS. Les utilités partielles découlant des comparaisons par paires ou d'appréciation par un oui ou par un non sont, quant à elles, estimées par les modèles logit et probit binaires.

1.2 Préférences par le choix

L'approche de l'analyse conjointe par le choix est une méthode récente (années '80). Elle intègre le processus de choix discret dans l'évaluation des produits. Sa méthodologie consiste à regrouper plusieurs produits dans un certain nombre de sous-ensembles de choix. L'individu doit, ensuite, choisir un produit dans chacun de ces sous-ensembles. Il y a généralement entre 10 et 20 sous-ensembles auxquels le répondant est soumis. Etant donné que la littérature traite essentiellement du choix dans le cas agrégé, les commentaires relatifs à cette section ne concerneront donc que la situation d'estimation sous forme agrégée.
Une option supplémentaire, à savoir l'option "aucun produit ne me satisfait", est parfois introduite dans les sous-ensembles de choix. Cette option permet de rendre les situations de choix des consommateurs plus réalistes. Cependant, l'estimation des utilités partielles avec cette option supplémentaire n'est pas clairement définie dans la littérature.

Une des différences par rapport à l'analyse conjointe notation-classement se situe au niveau des plans d'expérience. En effet, l'analyse conjointe basée sur le choix doit concevoir deux plans d'expérience. Le premier limite le nombre de scénarios ou de produits, alors que le second permet de déterminer les produits par sous-ensembles et les individus par groupes de consommateurs [3]. Cette procédure est identique pour la méthode des comparaisons par paires. De plus, l'analyse conjointe par le choix a l'avantage de calculer aisément les effets d'interaction fréquemment rencontrés dans les marchés. En effet, ces effets d'interaction se traduisent, par exemple, par un effet prix amenant à un effet qualité pour les consommateurs. Dans ce cas, les deux attributs sont liés et l'effet d'interaction doit être pris en considération dans la spécification du modèle.

Il existe d'autres effets modélisables de l'analyse conjointe par le choix. Ces effets sont l'effet spécifique à une alternative et l'effet croisé. Ces effets ne peuvent pas être calculés avec la méthode notation-classement, contrairement à l'effet d'interaction. L'effet spécifique à une alternative, appelé également "l'effet de la marque" peut être, d'une

[3] Les données de choix étant agrégées, il faut donc déterminer à quels groupes d'individus sont soumis les différents sous-ensembles de choix.

part, défini comme une interaction entre une alternative et une marque. En effet, lorsqu'un produit a une utilité plus forte avec une certaine marque, l'utilité est dite spécifique à la marque. D'autre part, l'effet spécifique à l'alternative se rencontre également lorsqu'il faut faire varier les niveaux d'attributs d'un produit par rapport à un autre. Par exemple, pour des voitures de luxe, les niveaux de prix sont évidemment au-dessus de ceux des voitures économiques. Une autre définition des effets spécifiques aux alternatives concerne les différences des attributs d'une alternative à l'autre. Par exemple, l'attribut "temps d'attente" d'un bus n'existe pas pour une voiture. Le dernier effet mentionné est l'effet croisé. Il suppose qu'un nouveau produit n'obtient pas les parts de marché des autres concurrents de manière proportionnelle. L'inconvénient de ces deux derniers effets est de compliquer le plan d'expérience. Effectivement, l'effet spécifique à l'alternative ajoute un peu de complexité aux plans, alors que l'effet croisé en ajoute bien plus.

L'analyse conjointe par le choix emploie comme méthode d'estimation des utilités partielles le logit multinomial (MNL). En réalité, les spécifications des modèles multinomiaux sont dérivées de la maximisation de l'utilité aléatoire (RUT). Cette maximisation s'effectue à partir des modèles probabilistes à choix discret. Ces modèles sont décrits ci-dessous. En effet, supposons qu'un consommateur i soit confronté à plusieurs produits j de la même catégorie, dont sa consommation est exclusive (c'est-à-dire qu'un seul produit doit être consommé). Chaque produit a une utilité différente selon ses attributs. Traditionnellement, le consommateur maximise son utilité sous une contrainte budgétaire, ou plus simplement il décide du produit à consommer. Toutefois, en raison d'une perception imparfaite de la qualité des différents produits ou parce que le problème de maximisation est trop complexe pour fournir un niveau exact d'utilité attaché à chaque produit, on suppose que l'utilité est une fonction aléatoire des attributs (Luce, 1959 ; Anderson, Palma et Thisse, 1992). Par conséquent, nous pouvons écrire l'utilité indirecte du produit j en fonction du vecteur x_j des attributs comme suit :

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

ε_j est la variable aléatoire qui représente les erreurs de perception et/ou d'optimisation de la part du consommateur i [4] et qui est supposée être indépendante de x_j. $v(.)$ est une fonction continue et déterministe. Elle dépend du vecteur des attributs et elle doit spécifier la manière dont ces attributs se combinent pour influencer le choix. En d'autres mots, il faut déterminer une "règle de combinaison" ou, plus généralement, une fonction d'utilité qui exprime officiellement la relation entre les variables explicatives et le comportement

[4] L'indice i du consommateur n'a pas été introduit dans l'équation de l'utilité afin de simplifier les notations.

de choix. Cette relation est généralement une fonction linéaire entre les paramètres β et les variables explicatives ou attributs x_j :

$$v_j = \beta x'_j$$

Cette formulation suppose que les choix des consommateurs varient uniquement selon les différences entre les attributs des produits. Cependant, les choix diffèrent également d'un individu à l'autre. C'est pourquoi nous incluons dans l'ensemble des variables explicatives celles qui décrivent les différences des consommateurs individuels, à savoir les contraintes financières ou de temps ou les mesures socio-démographiques. En nommant z' de telles mesures et γ le vecteur des coefficients associé, nous pouvons réécrire l'expression ci-dessus comme suit :

$$v_j = \beta x'_j + \gamma z'$$

L'utilité indirecte du produit j devient :

$$U_j = v(x_j, z) + \varepsilon_j = \beta x'_j + \gamma z' + \varepsilon_j$$

Le problème du consommateur se transpose dorénavant en un univers aléatoire. Il doit choisir le produit j lui procurant la plus grande utilité indirecte aléatoire $U(x_j, z, \varepsilon_j)$. Formellement, cette définition se traduit par la règle de décision suivante :

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

où y_j est une variable latente égale à 1 lorsque le produit j est choisi et 0 sinon.

La probabilité découlant de cette maximisation s'écrit de la manière suivante :

$$P(y_j = 1) = P(U_j > U_1, \ldots, U_j > U_m) = P(v_j + \varepsilon_j > v_1 + \varepsilon_1, \ldots, v_j + \varepsilon_j > v_m + \varepsilon_m)$$

où v_j est l'abréviation de $v(x_j, z)$.

En réarrangeant les termes aléatoires et déterministes, la probabilité de choisir le produit j devient :

$$P(\varepsilon_1 < v_j - v_1 + \varepsilon_j, \ldots, \varepsilon_{j-1} < v_j - v_{j-1} + \varepsilon_j, \varepsilon_{j+1} < v_j - v_{j+1} + \varepsilon_j, \ldots, \varepsilon_m < v_j - v_m + \varepsilon_j)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{v_j - v_1 + \varepsilon_j} \dots \int_{-\infty}^{v_j - v_{j-1} + \varepsilon_j} \int_{-\infty}^{v_j - v_{j+1} + \varepsilon_j} \dots \int_{-\infty}^{v_j - v_m + \varepsilon_j} f(\varepsilon_1, \dots, \varepsilon_j, \dots, \varepsilon_m)\, d\varepsilon_1 \dots d\varepsilon_j \dots d\varepsilon_m$$

En supposant que les termes d'erreur $\varepsilon_1, \dots, \varepsilon_m$ sont indépendamment et identiquement distribués (iid), l'intégrale peut s'écrire :

$$P(y_j = 1) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{v_j - v_1 + \varepsilon_j} f(\varepsilon_1) \dots \int_{-\infty}^{v_j - v_{j-1} + \varepsilon_j} f(\varepsilon_{j-1}) \int_{-\infty}^{v_j - v_{j+1} + \varepsilon_j} f(\varepsilon_{j+1}) \dots \right.$$

$$\left. \int_{-\infty}^{v_j - v_m + \varepsilon_j} f(\varepsilon_m)\; d\varepsilon_1 \dots d\varepsilon_{j-1} d\varepsilon_{j+1} \dots d\varepsilon_m \right] f(\varepsilon_j) d\varepsilon_j$$

Cette expression est très difficile, voir 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 de Gumbel (McFadden 1974), le problème admet une solution analytique très simple. Réécrivons de manière plus générale ou plus simplifiée l'expression de la probabilité de choisir le produit j :

$$P(y_j = 1) = P(\varepsilon_k < \varepsilon_j + v_j - v_k) \qquad \forall j \neq k \quad (k = 1, \dots, (j-1), (j+1), \dots, m)$$

$$= \int_{-\infty}^{+\infty} \prod_{k \neq j} F(\varepsilon_j + v_j - v_k) f(\varepsilon_j) d\varepsilon_j$$

La loi de Gumbel ou des valeurs extrêmes de type I est définie de la manière suivante :

$$\begin{cases} F(\varepsilon_j) = exp\ [-exp\ (-\varepsilon_j)] \\ f(\varepsilon_j) = exp\ [-\varepsilon_j - exp\ (-\varepsilon_j)] \end{cases}$$

La probabilité de choisir le produit j devient donc :

$$P(y_j = 1) = \frac{exp\ [v_j]}{\sum_{k=1}^{m} exp\ [v_k]}$$

On retrouve ainsi l'expression du logit multinomial ou du logit conditionnel si l'hypothèse

de linéarité de la fonction $v(.)$ par rapport aux variables x_j est vérifiée. La variable dépendante de ce type de modèle est généralement une variable de choix. Elle s'obtient, dans les modèles d'analyse conjointe, en effectuant des plans d'expérience auprès des consommateurs. En effet, les individus révèlent leurs utilités par des choix [5] qu'ils font en réponse à une expérience de préférence énoncée. Par "préférence énoncée", nous voulons dire que les individus expriment leurs préférences dans un marché hypothétique dans lequel ni les choix réels correspondants, ni leurs conséquences se produisent forcément. Bien que le marché soit peut-être hypothétique, la théorie de l'utilité aléatoire suggère néanmoins que les consommateurs devraient essayer de maximiser leur utilité sous différentes contraintes. La maximisation de l'utilité est donc la base de l'analyse conjointe pour comprendre et modéliser le comportement de marché dans beaucoup de circonstances.
Il existe également des modèles combinant les préférences énoncées aux préférences révélées. Par "préférences révélées", nous associons les choix ou les goûts des consommateurs provenant du marché "réel" ou actuel. De ce fait, dans les données de préférences révélées, il n'apparaît que des alternatives ou des produits déjà existants sur le marché. Par contre, dans les données de préférences énoncées, il est possible de modéliser des produits ou des attributs nouveaux. Toutefois, ces données énoncées offrent une fiabilité plus faible que celle issue des données révélées [6]. Les données de type énoncé et révélé sont généralement combinées pour combler les désavantages de chacune.

L'application des modèles logit multinomiaux dans une étude par l'analyse conjointe a toutefois certaines limites contraignantes. En effet, comme il a été mentionné précédemment, ce type d'estimation suppose que les erreurs sont identiquement et indépendamment distribuées (*iid*). De plus, de cette hypothèse découle l'hypothèse IIA ("Independence of Irrelevant Alternatives" : indépendance des alternatives non pertinentes). Ces aspects négatifs rendent les modèles logit multinomiaux indésirables dans certaines applications, malgré leur expression simple et intuitive des fonctions de densité, de répartition et des probabilités. C'est pourquoi, il existe une continuité de développement sur des formes multinomiales alternatives en économétrie des variables qualitatives (le logit multinomial emboîté, le logit multinomial hiérarchisé et le probit multinomial (MNP)).

Le modèle MNP permet de s'affranchir de l'hypothèse IIA. Il suppose que les erreurs du modèle à utilité aléatoire (RUT) suivent une distribution normale multivariée (Thurstone 1927) avec un vecteur d'espérance nulle et une matrice de variance-covariance composée de covariances non nulles. Les probabilités qui découlent de cette spécification sont

[5] par des jugements dans l'analyse conjointe classique

[6] Dans les données de préférences énoncées, le consommateur est dans une marché hypothétique et il n'agit ou ne choisit pas toujours de la même manière que dans le marché réel.

des intégrales multiples. Toutefois, 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 dont le nombre de modalités ou alternatives ne dépasse pas 4. Par conséquent, 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. Un solution consiste à recourir à des méthodes de simulation numérique, telles que l'estimateur par la méthode des moments simulés (MSM) (McFadden 1989) ou d'autres méthodes de simulation des probabilités conjointes issues de la loi normale n'ayant pas les désavantages de l'estimateur MSM (Stern 1992 ; Hajivassiliou et Ruud 1994 ; Geweke, Keane et Runkle 1994). L'estimateur GHK (Geweke, Hajivassiliou et Keane : Hajivassiliou, McFadden et Ruud 1992) fait partie de ceux qui n'ont pas ces inconvénients.

Une autre limite des modèles multinomiaux logit est l'hypothèse d'homogénéité des consommateurs. Une alternative à ces modèles est le modèle des classes latentes (DeSarbo, Ramaswamy et Cohen 1995) qui intègre l'hétérogénéité des consommateurs. En effet, ces modèles segmentent les individus en des classes les plus homogènes possibles. L'hétérogénéité est donc inclue entre les classes. Néanmoins, il reste l'hypothèse d'homogénéité à l'intérieur des classes. La méthodologie permettant d'enlever cette hypothèse est décrite ci-dessous. Elle conduit finalement à un modèle individuel.

Reprenons les modèles logit multinomiaux. Ces derniers n'estimant généralement pas les utilités partielles de manière individuelle, agrègent les données en supposant que tous les individus sont similaires (hypothèse d'homogénéité). Les résultats par cette méthode de choix donnent donc des valeurs d'utilités communes ou des moyennes de ces utilités. Les modèles de classe latente résolvent ce problème d'homogénéité entre les individus en créant des segments de consommateurs homogènes. Toutefois, comme il a été mentionné précédemment, il reste encore de l'hétérogéneité dans les segments d'individus. Cette contrainte est palliée par l'estimation des utilités de manière individuelle. En effet, Johnson (1997) utilise une méthode modifiée de l'approche des classes latentes qui permet d'estimer les utilités individuelles et qui évite l'hypothèse d'homogénéité dans les classes. La procédure de cette méthode est au début identique à l'approche des classes latentes. Les modèles de classe latente supposent que chaque individu appartient à une seule classe. Lorsqu'ils sont appliqués à des données de choix, ils estiment l'ensemble des utilités de chaque classe, en tenant compte de la probabilité d'appartenir à cette classe. On obtient donc par l'approche des classes latentes les utilités des niveaux d'attribut par groupe et les probabilités estimées d'appartenance à chaque groupe homogène. L'ex-

tension proposée par Johnson (1997) est de calculer les utilités par individu à partir de ces données. Effectivement, l'estimation des utilités individuelles s'effectue en utilisant les probabilités d'appartenance de chaque classe comme des pondérations appliquées aux utilités de chaque groupe [7].
Une meilleure manière de calculer les probabilités d'appartenance à un groupe est la méthode des pondérations individuelles non contraintes. Cette méthode estime les pondérations individuelles en utilisant une régression logit multinomiale. En effet, une première étape consiste à effectuer une analyse par les classes latentes traditionnelles ou par d'autres méthodes de "clustering" pour obtenir les estimations des utilités partielles par groupe. Ensuite, une seconde étape est d'utiliser ces utilités groupées comme variables indépendantes [8] dans une régression propre à chaque individu [9]. Nous obtenons, ainsi, les pondérations qui, lorsqu'elles sont utilisées pour associer les utilités des groupes, produisent la combinaison pondérée des utilités qui correspond le mieux aux données de choix individuel en utilisant le critère du maximum de vraisemblance.

Une autre approche d'estimation des utilités individuelles à partir des données de choix, appelée ICE (Individual Choice Estimation) [10] et introduite également par Johnson (1997), n'exige pas au préalable l'application de la méthode des classes latentes. En réalité, l'algorithme ICE était à l'origine développé comme une extension de l'analyse des classes latentes (cf. ci-dessus) afin d'obtenir de meilleures prédictions en reconnaissant davantage l'hétérogénéité dans les préférences. Néanmoins, il s'avère que cet algorithme est plus général et qu'on peut obtenir des résultats similaires sans avoir besoin des classes latentes au préalable.

Les différentes méthodes proposées dans ce chapitre peuvent également être comparées d'une autre manière, à savoir selon le point de vue des interviewés lors d'un sondage. En effet, Furlan et Corradetti (2004) présentent une comparaison des méthodes de ACA, de CBC [11] et des profils complets selon le point de vue des interviewés. Pour ce faire, ils utilisent une analyse discriminante.

[7] Par exemple, si nous avons deux classes d'individu, l'utilité partielle du niveau d'attribut de l'individu i, dont sa probabilité d'appartenance au groupe 1 est de 0.51 et celle au groupe 2 de 0.49, sera égale à la somme des produits de l'utilité du groupe 1 avec sa probabilité d'appartenance de 0.51 et de l'utilité du groupe 2 avec sa probabilité d'appartenance de 0.49.

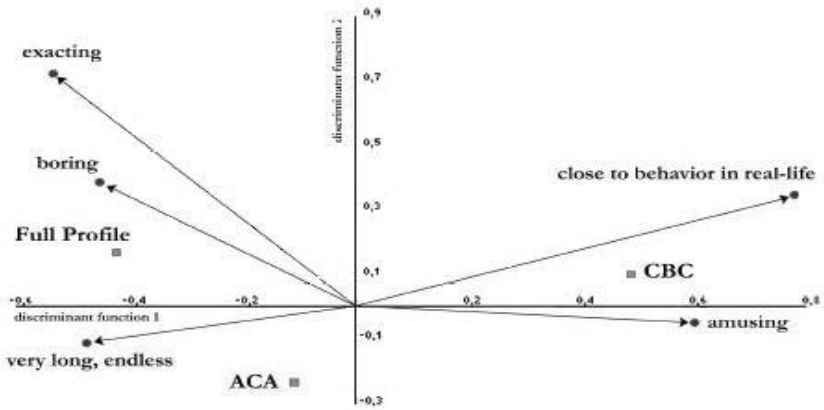
[8] Les variables indépendantes sont les sommes d'utilités partielles pour chaque alternative de choix.

[9] La variable dépendante est le choix observé de l'individu.

[10] ICE est une marque de la société Sawtooth Software.

[11] CBC est une marque de la société Sawtooth Software.

Le résultat de cette analyse est illustré à l'aide de la représentation graphique suivante :



D'après cette illustration, la méthode CBC est plus appréciée par les interviewés et s'assimile plus à la réalité. Quant aux autres méthodes, elles sont plus proches de l'ennui que de l'amusement.

Conclusion

Il existe plusieurs différences entre la méthode notation-classement et celle du choix. L'approche notation-classement est tout de même plus fréquemment abordée et présente dans la littérature de l'analyse conjointe. De plus, un grand nombre de packages ou de softwares informatiques, offrant la possibilité de compiler les données issues de l'analyse conjointe notation-classement, sont disponibles. L'approche du choix est une méthode plus récente et donc moins développée dans le domaine informatique. Elle décrit toutefois de manière plus réaliste le comportement d'achat du consommateur sur le marché. Dans ce chapitre, elle n'a été abordée que dans le cas d'estimation du modèle sous forme agrégée. Elle peut également être utilisée de manière individuelle, permettant ainsi d'analyser les préférences individuelles des consommateurs, telles que l'approche notation-classement le fait. Dans le cas agrégé de l'approche conjointe par le choix, les effets spéciaux, tels que les effets croisés et les effets spécifiques à l'alternative, peuvent, contrairement au cas individuel, être calculés. Les degrés de liberté associés au cas agrégé permettent, en effet, ce type de calcul. La méthode déposée par le logiciel Sawtooth et appelée ICE estime également les utilités individuelles en utilisant l'approche par le choix.

Telle que la littérature le montre, l'analyse conjointe par la notation ou le classement a souvent été utilisée dans des cas pratiques. Toutefois, nous remarquons que l'analyse

conjointe par le choix s'est également bien développée rendant ainsi son utilisation plus fréquente.

Chapitre 2

Les plans d'expérience

Introduction

L'analyse conjointe évalue les préférence des individus sur des produits constitués d'attributs à plusieurs modalités. La détermination de ces produits dépend donc du choix des attributs et de leurs niveaux. En d'autres termes, ils sont le résultat d'une combinaison des niveaux des attributs. Or, selon le nombre d'attributs et de niveaux, le nombre total de combinaisons résultant peut très vite atteindre des valeurs astronomiques. Par exemple, si le produit à analyser est caractérisé par 5 attributs à 4 niveaux, le nombre total de combinaisons représentant les différents produits est de $4^5 = 1024$. Tester 1024 produits par individu n'est pas concevable. Les plans d'expérience permettent la réduction de ce nombre. Dans notre application, nous réduisons le nombre de paires de produits à l'aide des plans D-optimaux. D'autres plans permettent de réduire le nombre de combinaisons. Ils seront abordés dans ce chapitre.

Dans ce chapitre, nous abordons d'abord les plans optimaux pour les modèles linéaires. Puis, nous illustrons, par des exemples, les plans factoriels complets et les plans orthogonaux réduits, tels que les plans factoriels fractionnaires, les carrés latins et les gréco-latins. Ensuite, nous définissons un autre type de plan réduit, à savoir les plans en bloc incomplets équilibrés et les plans en carré de Youden. Finalement, nous décrivons les plans D-optimaux, les plans pour les modèles non linéaires et les plans en blocs incomplets équilibrés associés aux plans D-optimaux.

2.1 Plans optimaux pour modèles linéaires

Le modèle linéaire de la régression s'écrit de la manière suivante :

$$y = X\beta + \varepsilon \qquad \text{(forme matricielle)}$$

Lorsque le modèle n'a qu'une variable explicative x_1 dans la matrice X, le modèle de régression est appelé le modèle linéaire simple : $y = \beta_0 + \beta_1 x_1 + \varepsilon$. Lorsqu'il a plusieurs régresseurs $x_1, \ldots, x_r$, il est appelé modèle linéaire multiple : $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_r x_r + \varepsilon$.

Les plans d'expérience dans le cadre d'une régression linéaire ont pour but de trouver le contenu de la matrice X apportant une estimation optimale des paramètres. Cette dernière se déduit de la matrice d'expérience. Cette matrice d'expérience contient uniquement les facteurs ou variables explicatives $x_1, \ldots, x_r$, alors que la matrice X contient la matrice d'expérience, les interactions entre ces facteurs, une colonne de 1 lorsque le modèle est défini avec une constante, et éventuellement les effets associés au degré supérieur ou égal à deux des facteurs. Les variables explicatives de la matrice d'expérience peuvent être quantitatives et/ou qualitatives. Elles sont définies en accord avec le domaine expérimental. Ce dernier permet de définir les observations ou les modalités des variables explicatives prises en compte dans l'expérience. Par exemple, un chimiste s'intéressant à l'effet de la température (x_1) et de la pression (x_2) sur une variable expliquée ou réponse y, choisit différentes combinaisons des niveaux de ces variables. Il décide, par exemple, de tester les températures 20 et 100 °C et les pressions 5 et 10 bar. La matrice expérimentale est constituée de ces combinaisons supposées réalisables :

	x_1	x_2
1	20	5
2	20	10
3	100	5
4	100	10

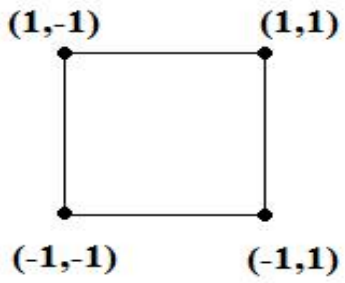
Pour chaque réalisation, nous avons une variables réponse y :

	y
1	y_1
2	y_2
3	y_3
4	y_4

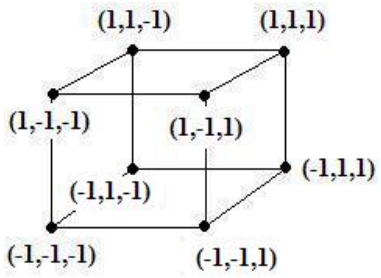
Les variables explicatives de notre exemple étant sous la forme quantitative, elles sont généralement transformées en -1 pour le minimum et en 1 pour le maximum. Dans le

cas d'une variable qualitative, il n'y a pas de transformation étant donné la nature de la variable.

Ces combinaisons de températures et de pressions peuvent être illustrées à l'aide d'un carré où les extrémités représentent les situations où la température et la pression sont soit au maximum de la valeur fixée par le chercheur soit au minimum. L'intérieur, quant à lui, illustre les situations dans lesquelles les valeurs des variables sont comprises entre le minimum et le maximum. Les plans factoriels ne s'intéressent qu'aux situations aux extrémités.



En rajoutant une variables explicatives, notre carré devient un cube :



Dès que le plan d'expérience est réalisé, c'est-à-dire que les valeurs de la variable dépendante y sont récoltées, l'analyse des effets de la température et de la pression s'effectue à l'aide de l'estimation du modèle :

$$y = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 TP + \varepsilon,$$

où T représente la température, P la pression, TP leur interaction et ε l'erreur du modèle.

Sous forme matricielle, le modèle devient :

$$\begin{pmatrix} y \\ \boxed{\begin{matrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{matrix}} \end{pmatrix} = \begin{pmatrix} & x_1 & x_2 & x_1x_2 \\ 1 & 20 & 5 & 100 \\ 1 & 20 & 10 & 200 \\ 1 & 100 & 5 & 500 \\ 1 & 100 & 10 & 1000 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

où la matrice X du modèle de notre exemple est constituée d'un vecteur de 1, de la matrice d'expérience (en encadré) et du vecteur des interactions.

Si nous remplaçons les minimums par la valeur -1 et les maximums par la valeur 1, la forme matricielle du modèle se transforme de la manière suivante :

$$\begin{pmatrix} y \\ \boxed{\begin{matrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{matrix}} \end{pmatrix} = \begin{pmatrix} & x_1 & x_2 & x_1x_2 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

En estimant ce modèle par les moindres carrés ordinaires ($\hat{\beta} = (X'X)^{-1}X'y$), nous obtenons les effets principaux des variables explicatives et les effets d'interactions. La matrice $(X'X)$ est diagonale si le plan d'expérience est orthogonal. Les plans complets, c'est-à-dire ceux contenant toutes les combinaisons, sont orthogonaux. Par contre, les plans réduits ne sont pas toujours orthogonaux. Ces derniers sont utiles lorsque le nombre total de combinaisons des facteurs est élevé. Certains plans d'expérience, tels que les plans factoriels fractionnaires, les carrés latins, les plans gréco-latin, les plans de Youden, etc. peuvent réduire ce nombre. La différence entre certains de ces plans est la contrainte sur le nombre de facteurs. La conséquence découlant de cette réduction est l'impossibilité de calcul de certains effets associés aux facteurs.

Soit la régression suivante : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$. Les plans d'experience permettent de définir le contenu de la matrice X de la régression $y = X\beta + \varepsilon$ à l'aide de la matrice d'expérience. Supposons que le nombre n d'essais est fixé. Il faut désormais définir un critère permettant d'obtenir une matrice X offrant une estimation optimale des paramètres de la régression. Nous savons que ces paramètres sont optimaux lorsque leur variance est minimale. Les paramètres d'un modèle linéaire sont généralement estimés par les moindres carrés ordinaires. Sous certains conditions (la matrice X est de rang complet, les erreurs sont centrées, non corrélées et de variance égale à σ^2), ces estimateurs sont

BLUE (Best Linear Unbiased Estimator), c'est-à-dire qu'ils sont à variance minimale. La matrice des variance-covariances des paramètres estimés par les moindres carrés linéaires $\hat{\beta}$ étant égale à $\sigma^2(X'X)^{-1}$, cette dernière est donc minimale, au sens du déterminant minimal, lorsque $(X'X)^{-1}$ l'est aussi. $(X'X)^{-1}$ est minimal lorsque $(X'X)$ est maximale. Par conséquent, un des critères d'optimalité est de maximiser le déterminant de la matrice $(X'X)$. Ce critère, qui est très souvent utilisé, s'appelle la D-optimalité. Il permet de déterminer les variables explicatives, ainsi que leurs interactions devant figurer dans la matrice X.
Un autre critère, la A-optimalité, fonctionne sur le même principe énoncé précédemment, à savoir la minimisation des variances des paramètres. Le critère A-optimal consiste à minimiser la trace de la matrice de variance-covariance, qui est égale à la somme des variances des paramètres.
Dans le modèle linéaire, une matrice orthogonale est forcément optimale pour tous les critères de type A, D, etc. C'est pourquoi, la recherche de plans orthogonaux a été très développée.

Dans les plans d'expérience, il existe une condition liée au nombre minimum d'expériences pour estimer tous les effets principaux. L'équation définissant ce nombre minimum d'expérience est la suivante :

$$\text{Nombre minimum d'expériences} > \sum_{i=1}^{\text{nb attributs}} m_i - \text{nb attributs}$$

Soit, par exemple, 4 attributs composés de 2 modalités et 3 de 3 modalités. Le nombre total de modalités ($\sum_i m_i$) est égal à $4 \times 2 + 3 \times 3 = 17$. L'équation ci-dessus est donc la suivante : $17 - 7 = 10$. Il faut donc au moins 11 essais ou expériences pour estimer les 10 effets principaux, étant donné qu'il faut enlever une modalité par attribut pour obtenir une matrice X de rang complet. Le $11^{\text{ème}}$ essai sert à calculer la somme au carré de l'erreur. Dans l'analyse conjointe, les modèles ne comportent pas les interactions. Si les effets d'interactions intéressent le chercheur, le nombre minimum cité ci-dessus n'est pas suffisant. En effet, il correspond au nombre minimum de combinaisons à faire tester pour connaître l'effet principal de chaque facteur.

Les plans factoriels fractionnaires orthogonaux réduisent le nombre d'essais. Cette réduction dépend du nombre de facteurs et de modalités. Il n'existe pas toujours des plans orthogonaux associés à chacun de ces nombres. L'ouvrage de Kuhfeld (2005) fournit une liste non exhaustive de plans orthogonaux réduits selon le nombre de facteurs et de modalités :

n_r	m^f	n_r	m^f	n_r	m^f	n_r	m^f
4	2^3	16	2^{15}	21	$3^1 7^1$	28	2^{27}
6	$2^1 3^1$		$2^{12} 4^1$	22	$2^1 11^1$		$2^{12} 7^1$
8	2^7		$2^9 4^2$	24	2^{23}		$2^2 14^1$
	$2^4 4^1$		$2^8 8^1$		$2^{20} 4^1$		$4^1 7^1$
9	3^4		$2^6 4^3$		$2^{16} 3^1$		
10	$2^1 5^1$		$2^3 4^4$		$2^{14} 6^1$		
12	$2^1 1$		4^5		$2^{13} 3^1 4^1$		
	$2^4 3^1$	18	$2^1 3^7$		$2^{12} 12^1$		
	$2^2 6^1$		$2^1 9^1$		$2^{11} 4^1 6^1$		
	$3^1 4^1$		$3^6 6^1$		$3^1 8^1$		
14	$2^1 7^1$	20	2^{19}	25	5^6		
15	$3^1 5^1$		$2^8 5^1$	26	$2^1 13^1$		
			$2^2 10^1$	27	3^{13}		
			$4^1 5^1$		$3^9 9^1$		

où n_r correspond au nombre d'essais réduits, f au nombre maximal de facteurs et m au nombre de modalités.

L'utilisation des plans optimaux, mais non-orthogonaux (A-optimalité, D-optimalité, G-optimalité (basé sur la minimisation de la variance de l'erreur standard de prédiction σ du modèle linéaire), intervient lorsque le nombre n souhaité d'essais, compte tenu des modalités et des facteurs, ne figure pas dans la liste ci-dessus ou simplement ne correspond pas à un plan orthogonal existant. Par conséquent, le plan découlant de ces critères n'aura que l'optimalité. Toutefois, malgré le fait qu'il ne soit pas dans la liste ci-dessus, il peut, tout de même, être orthogonal. Il suffit que ce soit un plan orthogonal n'ayant pas été trouvé par les chercheurs.

2.2 Exemple de plans factoriels complets et de plans orthogonaux fractionnaires

Reprenons le modèle de régression à trois variables explicatives x_1, x_2, x_3. La matrice d'expérience s'écrit comme suit :

obs	A	B	C
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Nous avons donc un plan factoriel complet à 8 essais étant donné que nous avons 3 facteurs à 2 modalités ($2^3 = 8$).

Etant donné que le modèle de régression s'écrit : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, il faut donc ajouter une colonne de 1 à la matrice d'expérience :

	A	B	C
1	-1	-1	-1
1	1	-1	-1
1	-1	1	-1
1	1	1	-1
1	-1	-1	1
1	1	-1	1
1	-1	1	1
1	1	1	1

Avec nos 8 essais, nous avons à calculer la constante, la variance du modèle et les trois effets principaux associés à x_1, x_2 et x_3. Nous pouvons encore calculer 3 effets supplémentaires, à savoir les interactions AB, AC et BC.
Par conséquent, la matrice du modèle devient comme suit :

	A	B	C	AB	AC	BC
1	-1	-1	-1	1	1	1
1	1	-1	-1	-1	-1	1
1	-1	1	-1	-1	1	-1
1	1	1	-1	1	-1	-1
1	-1	-1	1	1	-1	-1
1	1	-1	1	-1	1	-1
1	-1	1	1	-1	-1	1
1	1	1	1	1	1	1

Illustrons les plans factoriels complets à l'aide d'un exemple sur la mode. Soit le produit, un jeans, défini selon sa forme (Slim, Bootcut), sa marque (Gap, H&M), son prix (75,

100, 115) et sa couleur (Claire, Foncée). Le plan complet suivant présente toutes les combinaisons des modalités des attributs :

Obs	A	B	C	D
1	1	1	1	1
2	1	1	1	2
3	1	1	2	1
4	1	1	2	2
5	1	1	3	1
6	1	1	3	2
7	1	2	1	1
8	1	2	1	2
9	1	2	2	1
10	1	2	2	2
11	1	2	3	1
12	1	2	3	2
13	2	1	1	1
14	2	1	1	2
15	2	1	2	1
16	2	1	2	2
17	2	1	3	1
18	2	1	3	2
19	2	2	1	1
20	2	2	1	2
21	2	2	2	1
22	2	2	2	2
23	2	2	3	1
24	2	2	3	2

ou

Produit	Forme	Marque	Prix	Couleur
1	Slim	Gap	75	Claire
2	Slim	Gap	75	Foncée
3	Slim	Gap	100	Claire
4	Slim	Gap	100	Foncée
5	Slim	Gap	115	Claire
6	Slim	Gap	115	Foncée
7	Slim	H&M	75	Claire
8	Slim	H&M	75	Foncée
9	Slim	H&M	100	Claire
10	Slim	H&M	100	Foncée
11	Slim	H&M	115	Claire
12	Slim	H&M	115	Foncée
13	Bootcut	Gap	75	Claire
14	Bootcut	Gap	75	Foncée
15	Bootcut	Gap	100	Claire
16	Bootcut	Gap	100	Foncée
17	Bootcut	Gap	115	Claire
18	Bootcut	Gap	115	Foncée
19	Bootcut	H&M	75	Claire
20	Bootcut	H&M	75	Foncée
21	Bootcut	H&M	100	Claire
22	Bootcut	H&M	100	Foncée
23	Bootcut	H&M	115	Claire
24	Bootcut	H&M	115	Foncée

L'analyse du plan dépend du mode de recueil de la variable dépendante y du modèle. Il se partage comme suit :
- Classement - Notation
- Choix unique (CBC "Choice Based Conjoint")
- Choix par paires, par triplets, etc

Dans le classement, les individus sont amenés à classer les produits dans un ordre croissant de préférence. Dans la notation, ils doivent noter chaque produit sur une échelle définie au préalable. Les modèles associés à ce type de recueil de données sont les modèles linéaires estimés par les moindres carrés ordinaires. Contrairement au classement, il peut y avoir des valeurs ex-aequo dans la notation. Dans notre exemple, nous proposons 24 produits. Classer dans un ordre croissant de préférence ou noter 24 jeans peut amener à des biais, compte tenu du nombre élevé de produits. En effet, l'individu classe correctement un nombre restreint de produits. Une solution serait de demander à l'individu de partager en groupes ou en blocs les 24 produits, tel qu'un des groupes correspondrait aux produits les plus appréciés et l'autre à ceux les moins appréciés. Ensuite, il pourrait classer ou noter les produits de chaque sous-groupe. Cependant, contrairement à la notation, dans le classement, il est difficile de regrouper ces blocs et de fusionner ces classements. La raison est que l'échelle de mesure n'est pas forcément la même d'un groupe à l'autre.

Une autre solution à ce nombre élevé de produits est les plans d'expérience. Ils permettent de réduire le nombre de produits à classer ou à noter. Selon la liste non exhaustive des plans orthogonaux réduits figurant dans l'ouvrage de Kuhfeld (2005), les 24 jeans de notre exemple pourraient être diminués à 12 produits. En effet, la combinaison $2^4 3^1$, où 4 dans l'expression 2^4 correspond au nombre maximum de facteurs à 2 modalités, donne un plan réduit orthogonal à 12 produits. Notre plan complet comprenant $2^3 3^1$ peut être réduit à 12 produits à l'aide d'un plan factoriel fractionnaire.

Un des plans factoriels fractionnaires à 12 produits est le suivant (SAS) :

Obs	Forme	Marque	Couleur	Prix
1	2	2	2	2
2	2	2	1	3
3	2	2	1	1
4	2	1	2	3
5	2	1	2	2
6	2	1	1	1
7	1	2	2	3
8	1	2	2	1
9	1	2	1	2
10	1	1	2	1
11	1	1	1	3
12	1	1	1	2

ou

Obs	Forme	Marque	Couleur	Prix
1	Bootcut	H&M	Foncée	100
2	Bootcut	H&M	Claire	115
3	Bootcut	H&M	Claire	75
4	Bootcut	Gap	Foncée	115
5	Bootcut	Gap	Foncée	100
6	Bootcut	Gap	Claire	75
7	Slim	H&M	Foncée	115
8	Slim	H&M	Foncée	75
9	Slim	H&M	Claire	100
10	Slim	Gap	Foncée	75
11	Slim	Gap	Claire	115
12	Slim	Gap	Claire	100

Ce plan est orthogonal et a l'avantage d'être également équilibré. Le terme équilibré signifie que chaque modalité par attribut apparaît le même nombre de fois dans le plan. Certains plans peuvent être orthogonaux sans être équilibrés, à savoir un plan où une des modalités apparaît plus que les autres.

Le nombre de produits peut également être obtenu à l'aide des plans D-optimaux. La procédure associée à ces plans est décrite en détail dans le paragraphe 4.7. Soit l'exemple du jeans où le plan réduit comporte 12 produits. Ce plan est orthogonal et optimal. Supposons que le nombre de jeans réduit soit encore trop élevé. Nous aimerions en réalité proposer uniquement 9 jeans à classer ou à noter. Selon la liste de Kuhfeld (2005), il n'y a pas de plans orthogonaux réduits à 9 produits avec notre configuration d'attributs et de modalités. Par conséquent, en utilisant un plan D-optimal, le plan réduit résultant sera

optimal mais pas orthogonal. Voici un plan réduit des jeans à l'aide du plan D-optimal (SAS) :

Obs	Forme	Marque	Couleur	Prix
1	2	2	2	2
2	2	2	1	1
3	2	1	2	3
4	2	1	1	3
5	2	1	1	2
6	1	2	2	2
7	1	2	1	3
8	1	1	2	1
9	1	1	1	2

ou

Obs	Forme	Marque	Couleur	Prix
1	Bootcut	H&M	Foncée	100
2	Bootcut	H&M	Claire	75
3	Bootcut	Gap	Foncée	115
4	Bootcut	Gap	Claire	115
5	Bootcut	Gap	Claire	100
6	Slim	H&M	Foncée	100
7	Slim	H&M	Claire	115
8	Slim	Gap	Foncée	75
9	Slim	Gap	Claire	100

Dès que les produits réduits sont déterminés, l'individu peut les classer ou les noter. Ainsi, nous obtenons la variable dépendante de notre modèle de régression, composée de chiffre allant de 1 à 9 avec le classement, si le plan réduit est à 9 produits, ou de 9 valeurs comprises, par exemple, entre 1 et 100 avec la notation, si l'échelle varie de 1 à 100. Le modèle étant le suivant :

$$y = \alpha_1 \times \text{Forme}_{\text{bootcut}} + \alpha_2 \times \text{Marque}_{\text{H\&M}} + \alpha_3 \times \text{Couleur}_{\text{Foncée}} + \alpha_4 \times \text{Prix}_{115} + \alpha_5 \times \text{Prix}_{100}$$

où par exemple,

$$y = \begin{pmatrix} 7 \\ 6 \\ 1 \\ 2 \\ 9 \\ 5 \\ 4 \\ 8 \\ 3 \end{pmatrix}$$

il suffit donc d'effectuer les moindres carrés ordinaires sur ce modèle pour obtenir les valeurs des α et leur significativité. Dès que nous avons obtenu les valeurs de α, nous pouvons calculer les utilités de toutes les modalités des attributs en utilisant la condition

de compensation exprimant que la somme des utilités d'un attribut est égale à 0.

Les résultats du modèle d'analyse conjointe se trouvent dans le tableau ci-dessous (SAS) :

Utilities Table Based on the Usual Degrees of Freedom

Label	Utility	Standard Error	Importance (% Utility Range)	Variable
Intercept	5.050	0.136		Intercept
Forme 1 Slim	2.400	0.132	56.14%	Class.Forme1
Forme 2 Bootcut	-2.400	0.132		Class.Forme2
Marque 1 Gap	1.100	0.132	25.73%	Class.Marque1
Marque 2 H&M	-1.100	0.132		Class.Marque2
Couleur 1 Claire	0.600	0.132	14.03%	Class.Couleur1
Couleur 2 Foncée	-0.600	0.132		Class.Couleur2
Prix 1 75	-0.050	0.208	4.09%	Class.Prix1
Prix 2 100	0.200	0.176		Class.Prix2
Prix 3 115	-0.150	0.192		Class.Prix3

Les importances se calculent à l'aide de l'équation suivante :

$$\text{Importance d'un attribut } i = \frac{\text{maximum utilité } i - \text{minimum utilité } i}{\sum_i (\text{maximum utilité } i - \text{minimum utilité } i)}$$

Les résultats sont dans le tableau suivant :

Attributs	Utilités extrêmes	Etendue de l'utilité	Importance Relative
Forme	-2.400 2.400	4.8	4.8/8.55= 56.14%
Marque	1.100 -1.100	2.2	2.2/8.55= 25.73%
Couleur	0.600 -0.600	1.2	1.2/8.55= 14.03%
Prix	-0.050 0.200 -0.150	0.35	0.35/8.55= 4.09%
Somme		8.55	

L'analyse de variance montre que le plan n'est pas orthogonal. En effet, les sommes des carrés de type I[1] et III[2] sont différentes :

R-Square	Coeff Var	Root MSE	y Mean
0.992	7.745	0.387	5.000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Forme	1	45.000	45.000	300.00	0.000
Marque	1	11.363	11.363	75.76	0.003
Couleur	1	2.969	2.969	19.80	0.021
Prix	2	0.216	0.108	0.72	0.554

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Forme	1	49.371	49.371	329.14	0.000
Marque	1	10.371	10.371	69.14	0.003
Couleur	1	3.085	3.085	20.57	0.020
Prix	2	0.216	0.108	0.72	0.554

Dans le choix unique, les individus doivent choisir un produit parmi un nombre fixe. Dans cette situation, le nombre fixe ne doit pas être trop élevé. Soit l'exemple des 24 jeans. L'individu choisira difficilement un produit préféré parmi 24 choix. Il faut donc réduire ce nombre. Cette réduction s'effectue de la même manière que précédemment, à savoir avec un plan factoriel fractionnaire ou un plan D-optimal. Le modèle n'est toutefois pas linéaire. L'estimation s'effectue avec les modèles multinomiaux de choix. La variable dépendante y est, quant à elle, constituée de 1 pour le choix et de 0 sinon.

Dans le choix par paires, par triplets, etc., le modèle n'est également pas linéaire. L'estimation s'effectue également avec les modèles multinomiaux de choix. Toutefois, il faut également utiliser les plans d'expérience afin d'obtenir un nombre de paires, de triplets, etc. par blocs ou par choix pas trop élevé. En effet, soit les 24 jeans du plan complet. Si nous voulons que l'individu fasse tous les choix par paires, cela donne $\frac{24\times 23}{2} = 276$ paires où il faut faire un choix. Ce nombre étant trop élevé, il faut donc réduire le nombre de paires par choix. Les plans d'expérience permettant cette réduction sont les plans en bloc incomplet équilibré. Ils sont abordés dans la suite.

[1]Selon la documentation du logiciel SAS, "*les sommes des carrés (sums of squares (SS)) de type I sont une amélioration incrémentielle des sommes des carrés des erreurs puisque chaque effet est ajouté au modèle. Elles peuvent être calculées en adaptant le modèle dans les étapes et en enregistrant la différence dans les sommes des carrés des erreurs de chaque étape.*" De plus, elles dépendent de l'ordre d'introduction des variables dans le modèle.

[2]Selon la documentation du logiciel SAS, "*les sommes des carrés de type III sont calculées en construisant une matrice d'hypothèses L et en calculant ensuite le SS associé aux hypothèses $L\beta = 0$.*" Contrairement aux sommes des carrés de type I, elles ne dépendent pas de l'ordre d'introduction des variables dans le modèle.

Les plans D-optimaux peuvent également être appliqués dans la réduction du nombre de paires. Etant donné que notre modèle n'est pas linéaire, les plans D-optimaux appliqués à ce type de modèle diffèrent en théorie des plans D-optimaux utilisés dans le cas des modèles linéaires. Cependant, par souci de simplicité, nous utilisons tout de même les plans D-optimaux adaptés aux modèles linéaires pour ces modèles non linéaires. L'application de ces plans dans le cadre de modèles non linéaires sera décrite en détail par la suite.

D'autres plans d'expérience peuvent réduire le nombre de combinaisons d'attributs. Il s'agit des plans en carré latin et gréco-latin.

2.3 Plans en carré latin

Les plans en carré latin sont des plans orthogonaux réduits pour trois facteurs ayant le même nombre k de niveaux. L'application de ces plans permet de réduire les k^3 combinaisons totales en k^2.

Par exemple, nous désirons étudier l'effet de 3 facteurs à 3 niveaux. Le nombre total de combinaisons s'élève à $3^3 = 27$ produits. Le plan en carré latin 3×3 ne considère que $3^2 = 9$ produits. Il existe de nombreux carrés latins présentant les 9 produits au lieu des 27. En voici un :

facteurs	f_2		
f_1	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

Dans cette illustration du carré latin, les lignes correspondent aux niveaux (1,2,3) du facteur 1, les colonnes aux niveaux (1,2,3) du facteur 2 et l'intérieur aux niveaux (A,B,C) du facteur 3.

Le carré latin 3×3 présenté ci-dessus est appelé un carré latin standard. Les carrés latins standards ont la caractéristique d'avoir sur la première ligne et sur la première colonne la suite A B C. Plusieurs ouvrages présentant les plans en carré latin standards ou non standards existent dans la littérature. En annexe 1, nous reprenons un des tableaux de l'ouvrage de Montgomery (1976) précisant le nombre total de plans en carré latin selon les niveaux des facteurs.

Afin d'illustrer ces plans réduits, nous reprenons l'exemple des jeans avec certaines modifications. En effet, nous n'avons plus que trois facteurs, à savoir la forme, la couleur et le

prix. Ces facteurs ont tous 3 modalités. Les modalités de la forme sont slim, bootcut et évasée, celles de la couleur sont claire, foncée et blanche et celle du prix sont 75, 100 et 115. Le nombre total de produits est 27 jeans. Avec les plans en carré latin, nous n'en avons plus que 9 :

facteurs	**Couleur**		
Forme	Claire	Foncée	Blanche
Slim	75	100	115
Bootcut	100	115	75
Evasée	115	75	100

Ces plans en carré latin sont orthogonaux. Les 9 produits sont par la suite soumis à un classement, une notation ou un choix unique, fournissant ainsi la valeur de la variable y. Cette valeur permet d'estimer le modèle linéaire par les moindres carrés linéaires et le modèle non linéaire par le probit ou le logit.

2.4 Plans en carré gréco-latin

Les gréco-latins sont des plans d'expérience fractionnaires à 4 facteurs (attributs) dont le nombre de niveaux est identique. Soit, par exemple, un plan pour facteurs à 4 niveaux. Un plan gréco-latin possible est donné ci-dessous :

facteurs	f_2			
f_1	1	2	3	4
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

Dans cette illustration du gréco-latin, les lignes correspondent aux niveaux (1,2,3,4) du facteur 1, les colonnes aux niveaux (1,2,3,4) du facteur 2, les lettres latines aux niveaux (A,B,C,D) du facteur 3 et les lettres grecques aux niveaux (α,β,γ,δ) du facteur 4.

Avec l'exemple sur la mode, où les 4 variables sont la forme (slim, bootcut, taille haute, taille basse), la marque (Gap, H&M, Diesel, Le Temps Des Cerises), la couleur (claire, foncée, blanc et brute) et le prix (75, 100, 115, 175), le plan gréco-latin réduisant les $4^4 = 256$ produits à $4 \times 4 = 16$ produits est le suivant :

facteurs	**Marque**			
Forme	Gap	H&M	Diesel	Cerises
Slim	Claire 75	Foncée 100	Blanche 115	Brute 175
Bootcut	Foncée 175	Claire 115	Brute 100	Blanche 75
Taille haute	Blanche 100	Brute 75	Claire 175	Foncée 115
Taille basse	Brute 115	Blanche 175	Foncée 75	Claire 100

Les plans gréco-latin sont orthogonaux. Les 16 produits sont par la suite soumis à un classement, une notation ou un choix unique, fournissant ainsi la valeur de la variable y. Dès que nous avons obtenu cette valeur de y, nous pouvons procéder à l'estimation du modèle linéaire par les moindres carrés linéaires ou du modèle non linéaire par le probit ou le logit.

2.5 Plans en bloc incomplets équilibrés

Les plans en bloc incomplets sont des plans en bloc randomisés dont certaines données sont manquantes. Les plans en bloc randomisés partagent les unités en blocs ou en groupes homogènes afin d'appliquer les modalités d'un facteur dans chaque bloc. Toutefois, le nombre d'exécutions associées à ces modalités par bloc peut devenir très élevé. Dans ces situations, il est donc plus convenable d'utiliser des plans réduits, tels que ceux en bloc incomplets qui ne considèrent pas tous les niveaux dans chaque bloc. Ces plans s'accompagnent généralement de la notion d'équilibre. Les plans en bloc incomplets équilibrés sont des plans dans lesquels les paires de modalités apparaissent un nombre de fois λ dans chaque bloc. L'interprétation de cette définition est que l'importance de chaque comparaison de niveaux est similaire.

Ce type de plan d'expérience est intéressant, par exemple, lorsque le chercheur, désirant tester les modalités des facteurs sur des supports différents, ne peut effectuer cette démarche sur tous les supports compte tenu de la taille de ces derniers et de la quantité des combinaisons de modalités de facteurs. Par conséquent, étant donné la taille du support, il n'y aura qu'un nombre réduit de combinaisons de modalités de facteurs par support. Toutefois, les combinaisons figurant sur chaque support varieront et ce afin d'obtenir l'analyse de toutes ces combinaisons. Dans l'analyse conjointe, un exemple d'application de ces plans en bloc incomplet équilibré est l'analyse des préférences moyennes d'un grand nombre de produits départagés dans plusieurs groupes d'individus. Les combinaisons de modalités des facteurs, ou les produits, ne sont pas toutes testées par chaque consommateur. En effet, les con so mma teurs sont groupés de manière homogène et ils sont interrogés sur leur préférence vis-à-vis d'un sous-ensemble de produits. Par la suite, chaque groupe ayant testé son sous-ensemble de produits, le chercheur peut analyser l'effet moyen de tous les produits pour tous les consommateurs. L'intérêt de ces plans semble donc établi compte tenu de ces exemples. Il est désormais intéressant de décrire leur procédure d'application.

Les plans en bloc incomplets équilibrés sont définis en fonction de plusieurs paramètres. Ces paramètres sont les suivants :
a : le nombre de niveaux du facteur

k : le nombre de niveaux par bloc
r : le nombre de fois que chaque niveau apparaît $\Rightarrow$ le nombre de blocs où le traitement x apparaît
λ : le nombre de fois que chaque paire apparaît par bloc
b : le nombre de blocs

Le plan en bloc incomplet est équilibré, lorsque la valeur de λ est la même dans chaque bloc. C'est pourquoi, nous n'indiçons pas la valeur de λ.

Le plan en bloc incomplet équilibré est une méthode permettant de disposer k niveaux d'un facteur ($k < a$ le nombre total de niveaux) dans b blocs. Ce nombre b de blocs dépend du nombre a de modalités et du nombre k de niveaux que l'on désire avoir par bloc :

$$b = \begin{pmatrix} a \\ k \end{pmatrix} = \frac{a!}{(a-k)!k!}$$

Lorsque les k niveaux par bloc sont déterminés, le nombre total N d'observations est égal à :

$$N = ar = bk$$

Quant au paramètre λ, il est déterminé à l'aide de l'équation suivante :

$$\lambda = \frac{r(k-1)}{a-1} \quad \text{où } \lambda \text{ est un entier relatif}$$

Si $a = b$, le plan est dit symétrique.
L'ouvrage de Federer (1955) présente un tableau des différents plans résolvables, à résoudre ou non résolvables. Ce tableau se trouve en annexe 2.

Soit un exemple de choix avec 10 paires de jeans présentées en paires, en triplets, etc. Les 10 paires de jeans découlent des combinaisons entre les attributs forme (slim, bootcut, taille haute, taille basse, évasée) et marque (Gap, H&M). Les $2^1 \times 5^1 = 10$ jeans sont les suivants :

Obs	Forme	Marque
1	Slim	Gap
2	Bootcut	Gap
3	Taille haute	Gap
4	Taille basse	Gap
5	Evasée	Gap
6	Slim	H&M
7	Bootcut	H&M
8	Taille haute	H&M
9	Taille basse	H&M
10	Evasée	H&M

Les 10 jeans représentent les niveaux du facteur. Nous voulons créer 15 blocs dans lesquels nous ne proposons que 4 jeans :

a=10 jeans

k=4 jeans par bloc

b=15 blocs

Par conséquent, $N = bk = 60$. Comme $N = ar \Rightarrow r = \frac{N}{a} = \frac{60}{10} = 6$ et $\lambda = \frac{r(k-1)}{a-1} = \frac{6(4-1)}{9} = 2$ un entier relatif.

Le plan en bloc incomplet équilibré, correspondant au 15 blocs à 4 jeans ou produits, se trouve dans l'ouvrage de Dagnelie (2003) :

1	10	6	8
7	3	8	1
9	1	4	10
2	7	10	4
5	4	8	3
7	5	6	4
7	2	8	9
4	2	1	3
8	6	4	9
7	9	5	1
3	10	9	5
6	5	2	1
10	3	6	7
2	6	3	9
10	8	5	2

facteur	**Blocs**														
f_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	y	y	y	-	-	-	-	y	-	y	-	y	-	-	-
2	-	-	-	y	-	-	y	y	-	-	-	y	-	y	y
3	-	y	-	-	y	-	-	y	-	-	y	-	y	y	-
4	-	-	y	y	y	y	-	y	y	-	-	-	-	-	-
5	-	-	-	-	y	y	-	-	-	y	y	y	-	-	y
6	y	-	-	-	-	y	-	-	y	-	-	y	y	y	-
7	-	y	-	y	-	y	y	-	-	y	-	-	y	-	-
8	y	y	-	-	y	-	y	-	y	-	-	-	-	-	y
9	-	-	y	-	-	-	y	-	y	y	y	-	-	y	-
10	y	-	y	y	-	-	-	-	-	-	y	-	y	-	y

où $y = \begin{cases} 1 & \text{si l'individu a choisi le produit a dans le bloc b} \\ 0 & \text{sinon} \end{cases}$

L'estimation s'effectue à l'aide des modèles multinomiaux de choix.

Tel que cet exemple l'illustre, l'existence des plans en blocs incomplets équilibrés dépend du nombre de niveaux des attributs, ainsi que du nombre de blocs et d'observations totales. Il se peut donc que pour certaines valeurs de ces niveaux, de ces blocs et de ces observations, il n'y ait aucun plan en blocs incomplets équilibrés. Kessels, Goos et Vandebroek (2007) proposent comme alternative à ces plans en blocs incomplets équilibrés, les plans D-optimaux. L'algorithme des plans D-optimaux, introduit dans leur article, permet aussi bien de déterminer, dans le cas d'analyse conjointe par la notation, le nombre optimal d'interviewés que le nombre optimal de scénarios par interviewé. Pour comprendre cette procédure, il faut toutefois connaître le principe général des plans D-optimaux, qui est décrit un peu plus loin.

2.6 Plans en carré de Youden

Les plans en carré de Youden sont, comme l'énonce Montgomery (1976) dans son ouvrage, *des carrés latins "incomplets"* dont le nombre des niveaux des 3 facteurs peut être différent. Les carrés de Youden sont également définis dans l'ouvrage de Montgomery (1976) comme étant des carrés latins dont une colonne, une ligne ou une diagonale a été enlevée. Néanmoins, l'élimination de plusieurs colonnes, lignes ou diagonales n'amène pas forcément au carré de Youden. De plus, une élimination arbitraire de plusieurs colonnes, lignes ou diagonales peut détruire l'équilibrage du plan résultant. En résumé, le carré de Youden est un plan incomplet symétrique équilibré dans lequel chaque lettre latine apparaît exactement un fois dans chaque colonne. Il existe de nombreux ouvrages présentant ces carrés de Youden.

Reprenons l'exemple des jeans avec comme attribut la forme (slim, bootcut, taille haute, taille basse, évasée), la couleur (claire, foncée, blanc, brute) et le prix (75, 100, 115, 150, 175). Le nombre de combinaisons totales s'élève à $5^2 \times 4^1 = 100$ produits. De l'ouvrage de Montgomery (1976), nous tirons la structure du plan de Youden que nous adaptons à notre exemple :

facteurs	**Couleur**			
Forme	Claire	Foncée	Blanc	Brute
Slim	75	100	115	150
Bootcut	100	115	150	175
Taille haute	115	150	175	75
Taille basse	150	175	75	100
Evasée	175	75	100	115

Ce plan est considéré comme un plan incomplet équilibré avec a=b=5, r=k=4 et λ=3. Les 20 produits sont par la suite soumis à un classement, une notation ou un choix unique, fournissant ainsi la valeur de la variable y. Cette valeur permet d'estimer le modèle linéaire par les moindres carrés linéaires et le modèle non linéaire par le probit ou le logit. Cet exemple avec 20 produits n'est toutefois pas réaliste. En effet, dans une application, 20 produits à classer, à noter, etc. est trop élevé. C'est pourquoi, il faut considérer cet exemple comme une illustration théorique des plans en carré de Youden.

2.7 Plans D-optimaux

Rechercher des plans D-optimaux consiste à réduire le nombre de combinaisons totales à l'aide du critère de maximisation du déterminant de la matrice de variance-covariance des paramètres. Ces plans peuvent ne pas être orthogonaux. En effet, les plans D-optimaux non orthogonaux signifient que pour certains nombres d'essais, de facteurs et de modalités fixés, il n'existe aucun plan orthogonal.

Tel que nous l'avons énoncé précédemment, les plans D-optimaux sont basés sur l'idée d'obtenir une variance minimale des paramètres du modèle des moindres carrés. En effet, dans une régression ayant comme modèle $y = Xb + \epsilon$ et comme variance des paramètres b, $V(b) = \sigma^2(X'X)^{-1}$, les plans D-optimaux minimisent cette variance $V(b)$ à travers $(X'X)^{-1}$ ou en maximisant $(X'X)$. Dans l'analyse conjointe, la matrice X considérée ci-dessus n'est rien d'autre que la matrice comportant les caractéristiques des produits. Cette matrice comporte donc que des 0 et que des 1. Toutefois, cette matrice étant singulière, il est d'usage d'enlever une modalité, appelée modalité de référence, pour chaque facteur ou attribut. Les plans D-optimaux sélectionnent plusieurs lignes ou essais, fixés au préalable, de X et calculent le déterminant de la sous-matrice de $(X'X)$. Le déterminant le plus élevé donne les combinaisons du plan réduit.

Un des algorithmes, l'algorithme d'échange, appliqué à la recherche de plans D-optimaux ne calcule pas, en réalité, tous les sous-ensembles de la matrice X. Pour appliquer cet algorithme, il faut d'abord fixer le nombre d'essais n contenu dans la matrice réduite. Ensuite, l'algorithme choisit au hasard une sous-matrice. Puis, il échange une des valeurs ou lignes de cette sous-matrice avec une autre et recalcule le déterminant de cette nouvelle sous-matrice. Si ce déterminant est supérieur à l'autre sous-matrice, l'échange se maintient. Si, par contre, il est inférieur, l'échange ne se maintient pas. Etant donné que cet algorithme n'est pas forcément convergent, il continue ces échanges et s'arrête au nombre d'itérations fixées. Le résultat n'atteint donc pas toujours l'optimum du critère. C'est pourquoi l'algorithme est généralement effectué plusieurs fois sur le plan complet. Par la

suite, nous pouvons sélectionner le plan ayant l'efficacité la plus élevée. Cette efficacité est calculée par la formule suivante : $\frac{|X'X|^{\frac{1}{p}}}{n}$. Elle est formulée de sorte à avoir une valeur de 1, lorsque le plan est orthogonal. En effet, le déterminant de la matrice $(X'X)$ d'un plan orthogonal est égal à n^p. En remplaçant cette valeur dans la formule de l'efficacité, nous obtenons bien 1. L'algorithme de recherche des plans D-optimaux peut également fournir un plan réduit orthogonal. Dans cette situation, le résultat de l'algorithme, s'il converge, est un plan orthogonal optimal.

Reprenons l'exemple sur la mode. Le produit, le jeans, est décrit selon sa forme (slim, bootcut), sa marque (Gap, H&M), son prix (75, 100, 115) et sa couleur (claire, foncée). Le plan complet des 24 produits est le suivant :

Produit	Forme	Marque	Prix	Couleur
1	Slim	Gap	75	Claire
2	Slim	Gap	75	Foncée
3	Slim	Gap	100	Claire
4	Slim	Gap	100	Foncée
5	Slim	Gap	115	Claire
6	Slim	Gap	115	Foncée
7	Slim	H&M	75	Claire
8	Slim	H&M	75	Foncée
9	Slim	H&M	100	Claire
10	Slim	H&M	100	Foncée
11	Slim	H&M	115	Claire
12	Slim	H&M	115	Foncée
13	Bootcut	Gap	75	Claire
14	Bootcut	Gap	75	Foncée
15	Bootcut	Gap	100	Claire
16	Bootcut	Gap	100	Foncée
17	Bootcut	Gap	115	Claire
18	Bootcut	Gap	115	Foncée
19	Bootcut	H&M	75	Claire
20	Bootcut	H&M	75	Foncée
21	Bootcut	H&M	100	Claire
22	Bootcut	H&M	100	Foncée
23	Bootcut	H&M	115	Claire
24	Bootcut	H&M	115	Foncée

Dans l'analyse conjointe, nous présentons ses choix de jeans par paires. Avec 24 produits, nous avons $\frac{24\times23}{2} = 276$ paires de jeans.
L'individu ne pouvant donner ces préférence sur 276 choix, nous optons pour une réduction à 10 choix. Nous avons 5 paramètres à estimer. Par conséquent, 10 choix suffisent pour estimer ces paramètres et l'erreur du modèle. Un plan orthogonal réduit du plan complet factoriel $2^3 \times 3$ à 10 essais n'existant pas, nous utilisons l'algorithme des plans D-optimaux

décrits dans la section 2.7. Le plan D-optimal à 10 essais résultant est le suivant :

obs	paire 1	paire 2
1	8	13
2	7	14
3	7	13
4	5	19
5	6	20
6	2	13
7	1	14
8	2	7
9	1	8
10	1	3
D-Efficiency	91.7071	

L'efficacité est donnée à la fin du tableau.

2.8 Plans pour modèles non linéaires

Les plans décrit précédemment sont utilisables lorsque le modèle décrivant la situation est supposé linéaire. Lorsque le modèle n'est pas linéaire, tel que pour les modèles logit, probit, etc., les plans associés à ces situations deviennent plus complexes. Il existe une théorie sur des plans applicables aux modèles non linéaires, ainsi que divers travaux (J.P. Gauchi). Cependant, ces résultats sont difficilement applicables.

En théorie, il ne faudrait pas appliquer les plans associés aux modèles linéaires sur les modèles non linéaires. Toutefois, dans la pratique, l'application de plans linéaires sur des modèles non linéaires est fréquente. La raison est que ces plans fournissent des résultats corrects avec les modèles non linéaires, sans être toutefois optimaux. Avec les plans pour modèles non linéaires, il n'existe pas un plan optimal pour toutes les valeurs considérées. L'optimalité est en réalité locale. Il faut donc connaître à l'avance les coefficients. Dans le cas contraire, l'application des plans linéaires aux modèles non linéaires est suffisante.

2.9 Plans en bloc incomplets équilibrés associés aux plans D-optimaux

Dans l'analyse conjointe, il faut d'abord choisir les attributs et leurs niveaux faisant partie du plan complet. Ensuite, les combinaisons totales, issues de l'étape précédente, sont soit mises en paires, en triplets, etc. pour être réduites à l'aide des plans D-optimaux, tel que cela est illustré dans la section 2.7, soit directement réduites à l'aide des plans D-optimaux, puis groupées dans des sous-ensembles de paires, de triplets, etc. grâce aux plans en bloc incomplets équilibrés. Lors de la réduction directe des combinaisons totales

à l'aide des plans D-optimaux, il faut connaître à l'avance le nombre final de combinaisons souhaitées. Ce nombre final doit s'accorder avec le nombre de sous-ensembles du plan en bloc incomplet équilibré.

Nous allons illustrer l'utilisation simultanée des plans D-optimaux et des plans en bloc incomplets équilibrés avec l'exemple qui suit. Supposons que le nombre de blocs b que l'interviewé peut supporter soit entre 10 et 15. De plus, nous n'aimerions que $k = 2$ produits par bloc. Pour quelle valeur de b existe-t-il une valeur de a, le nombre de combinaisons total ? Si $b = 15$, alors

$$N = ar = bk = 15.2 = 30$$

Or, de l'équation

$$\lambda = \frac{r(k-1)}{a-1} = \frac{r}{a-1},$$

les seules valeurs de r amenant l'λ à être un entier positif sont des multiples α de $a - 1$. Par conséquent, de l'équation précédente, nous tirons

$$30 = ar = \alpha a(a-1)$$

Si $\alpha = 1$, alors $30 = a(a-1) = a^2 - a \;\Rightarrow a^2 - a - 30 = 0$. La solution de ce polynôme de degré 2 est :

$$a = \frac{1 \pm \sqrt{1 + 4 \cdot 30}}{2} = \frac{1 \pm 11}{2} = \begin{cases} -5 \\ 6 \end{cases}$$

Etant donné que -5 est négatif, la seule valeur de a est 6. Il faut donc sélectionner 6 produits par le biais du plan D-optimal pour avoir ensuite 2 produits dans les 15 blocs.

En reprenant l'exemple des 24 jeans, il nous faut d'abord trouver un plan en bloc incomplet équilibré avec $a = 6$, $k = 2$, $b = 15$, $r = \frac{bk}{a} = \frac{15 \cdot 2}{6} = 5$ et $\lambda = 1$. Selon l'annexe 2, ce plan existe.
Ensuite, il suffit de sélectionner 6 produits sur les 24 avec un plan D-optimal et de les introduire dans le plan en bloc incomplet équilibré.

Pour certaines valeurs de b, il n'y a pas de combinaisons réduites possibles. Il faut donc dans ces cas modifier le nombre de blocs. Il se peut également que la valeur a issue de celle de b soit trop faible. Dans ce cas-là, il faut également changer la valeur de b. Cette procédure est donc basée sur le tâtonnement.

Une interrogation relative aux plans D-optimaux et au principe de compromis de l'analyse

conjointe est l'assurance de combinaisons dont le choix n'est pas évident. En effet, l'algorithme des D-optimaux l'assure-t-il ? La réponse est oui car le résultat de cet algorithme est une sous-matrice ayant le déterminant le plus élevé. Par conséquent, un tel résultat ne peut être obtenu qu'à condition que les niveaux des attributs soient très éloignés les uns des autres.

Le plan résultant par cette méthode de choix, c'est-à-dire l'association des plans D-optimaux et des plans en blocs incomplets équilibrés, comporte certaines contraintes liées à la variabilité des scénarios proposés dans un sondage. En effet, pour satisfaire un nombre de blocs ou de questions adéquates, les plans en blocs incomplets équilibrés demandent une réduction trop élevée des produits. La conséquence est que les produits et les paires proposés aux individus ne sont pas très variables et, selon le nombre de paires existantes, ils peuvent ne pas refléter la variété des paires initiales. Ainsi, les personnes interrogées lors d'un sondage peuvent ne pas se rendre compte qu'il existe une multitude de paires et de produits. De plus, elles sont amenées à faire des choix entre des paires et des produits très similaires. En réalité, ce type de méthodologie est plus adapté aux tests sensoriels. Ces derniers préfèrent comparer un nombre restreint de produits, étant donné que cette restriction diminue les coûts de fabrication tout en permettant d'effectuer une analyse sur tous les attributs et leurs modalités.
Par exemple, nous souhaitons analyser l'effet d'une sauce, ayant comme attributs, la présence ou non de la moutarde, du soja, de la mayonnaise, du piment ou du vinaigre. Le nombre d'attributs dichotomique est de 5. Le nombre de combinaison totale est de 32 et le nombre de paires de 496. Il est évident que faire goûter 496 paires est inimaginable. Par contre, l'utilisation d'un plan en blocs incomplets équilibrés et d'un plan D-optimal offre la possibilité d'évaluer l'importance des attributs et de leurs modalités en ne préparant qu'une quantité restreinte de produits. En effet, le plan en blocs incomplets équilibrés restreindra le nombre de produits. Ce ne seront que ces derniers qui seront utilisés pour le test sensoriel.

2.10 Problèmes liés au choix des paires

Dans la comparaison par paires, certaines peuvent s'avérer inutiles d'un point de vue informationnel. En effet, elles peuvent ne fournir aucun renseignement supplémentaire dans le calcul des effets principaux des attributs ou être trop simples pour être présentées lors d'un sondage. Ces paires appelées inutiles sont celles dites évidentes, celles qui ne comparent qu'un attribut ou celles dont leur ligne est identique en valeur absolue à une autre ligne de la matrice.
Les paires évidentes sont celles qui comparent des produits identiques en certains points,

sauf en ce qui concerne des attributs avec des modalités ordonnées. Soit l'exemple d'un produit, tel qu'un site de téléchargement de musique, où tous les niveaux des attributs sont identiques à l'exception de la quantité de musique proposée. Par exemple, l'un propose 10'000 chansons, alors que l'autre en offre 5'000. Cette situation de choix est clairement évidente. En effet, étant donné que la variable, nombre de chansons, est ordonnée, il semble évident que le produit sélectionné sera celui dont la quantité est la plus élevée.
Les paires, dont les produits ne varient qu'en fonction d'un attribut, ne sont pas d'un grand intérêt. En effet, elles ne renseignent que sur l'attribut en question, alors que nous sommes intéressés à connaître l'effet des variations simultanées de plusieurs attributs.

Les paires, dont le vecteur des différences des modalités des attributs est identique en valeur absolue à d'autres vecteurs de la matrice d'expérience, apportent la même information à l'analyse. Par conséquent, il est inutile de toutes les garder dans le modèle. Il suffit d'en sélectionner une qui représentera le contenu informationnel des autres paires. Cette diminution des paires est très utile et facilite ainsi l'algorithme de sélection du plan D-optimal. Elle s'appuie sur l'article de Benammou, Saporta et Swissi (2007), "*Une procédure de réduction du nombre de paires en analyse conjointe*". Cette méthodologie permet de diminuer considérablement le nombre de ces paires. Dans quelques exemples, il est même inutile d'appliquer les plans D-optimaux pour réduire la quantité de paires. En effet, si nous reprenons les chiffres de l'article, 4 variables dichotomiques amènent à un nombre de combinaisons de 16 et un nombre de paires de 120. L'application de la méthode de Benammou, Saporta et Swissi (2007) réduit les 120 paires à 15. Dans l'article, cette valeur est davantage diminuée en appliquant la réduction des paires qui ne comparent qu'un attribut. Au final, le nombre de paires s'élève à 11. L'évaluation de ces 11 paires, par rapport aux 120 initiales, devient donc concevable pour l'individu. En reprenant les termes de l'article, nous pouvons conclure que "*la réduction du nombre de paires dépasse les 90% dès que le nombre de variables dépasse 3*".

Conclusion

L'analyse conjointe est une méthodologie complète permettant la collecte de données, la modélisation des préférences et l'analyse des résultats. La quantité de données varie en fonction du nombre de niveaux d'attributs à analyser. Les plans d'expérience permettent de réduire ce nombre lorsque ce dernier devient trop élevé. Toutefois, l'aspect des tests de significativité des facteurs des plans d'expérience n'est pas pris en compte dans l'analyse conjointe. Seul la construction de ces plans est utile dans ce cadre d'analyse.

Chapitre 3

Modèle de choix de l'analyse conjointe et l'option "zéro"

Introduction

L'analyse conjointe par le choix utilise le modèle MNL pour estimer les utilités partielles. Cependant, cette spécification est légèrement modifiée par rapport à celle présentée dans les ouvrages sur les variables dépendantes qualitatives. En effet, les modèles à variables dépendantes qualitatives supposent un choix unique par personne. Or, dans l'analyse conjointe, il est demandé aux consommateurs de choisir un produit dans plusieurs sous-ensembles d'alternatives. Il en résulte que l'individu fait plusieurs choix uniques, c'est-à-dire un seul choix par sous-ensembles de produits. L'ajout de l'option "zéro" est également spécifiée dans la littérature par un modèle MNL. En effet, certains auteurs modélisent l'option "zéro" en ajoutant une ligne de 0 à la matrice des attributs et/ou une constante dans le modèle MNL. Toutefois, cette spécification n'est pas toujours adéquate pour des raisons qui seront abordées dans ce chapitre. Par conséquent, dans la thèse, nous proposons une nouvelle spécification de l'option "zéro" qui s'inspire des modèles quantitatifs censurés Tobit. Cette nouvelle spécification est construite sur le même principe que les modèles Tobit, à savoir l'idée qu'à partir d'un certain seuil la réponse du consommateur change. De plus, Tversky et Shafir (1992) et Dhar (1997) discutent des raisons psychologiques du non choix sans pour autant les modéliser. Dans ce chapitre, nous reprenons ces aspects psychologiques en les introduisant dans notre nouvelle spécification de l'option "zéro".

Ce chapitre aborde les modèles multinomiaux logit (MNL) dans une formulation générale et dans une formulation adaptée aux données de choix de l'analyse conjointe. Les raisons psychologiques du non choix sont également introduites dans ce chapitre. Elles permettent de mieux comprendre l'importance de leur modélisation dans une spécification de non

choix. Un nouveau modèle de l'option "zéro" est ensuite développé en tenant compte des notions et des principes formulés dans les sections précédentes et suivantes.

3.1 Modèles multinomiaux logit : formulation générale

3.1.1 Probabilités du modèle logit : source

Une variable dépendante qualitative à plusieurs modalités est modélisée dans la littérature économétrique par des modèles multinomiaux.
Supposons que cette variable ait $H+1$ catégories ou modalités. Á ces $H+1$ modalités, nous associons $H+1$ probabilités $P_0, P_1, \ldots, P_H$ de choisir la catégorie $h = 0, \ldots, H$. Dans l'ouvrage "*Limited-Dependent and Qualitative Variables in Econometrics*, Maddala (1983) exprime les probabilités P_h sous une forme binaire et suppose une linéarité entre les paramètres γ_h et les variables explicatives x :

$$\frac{P_1}{P_1 + P_0} = F(\gamma_1' x)$$

$$\frac{P_2}{P_2 + P_0} = F(\gamma_2' x)$$

$$\vdots$$

$$\frac{P_H}{P_H + P_0} = F(\gamma_H' x)$$

où P_0 est la variable dite de référence.

Ces équations permettent d'écrire le rapport entre les probabilités P_h, $h = 1, \ldots, H$ et P_0 de la manière suivante :

$$\frac{P_h}{P_0} = \frac{F(\gamma_h' x)}{1 - F(\gamma_h' x)} = G(\gamma_h' x) \quad (h = 1, \ldots, H)^1$$

[1] Car $\frac{F(\gamma_h' x)}{1-F(\gamma_h' x)} = \frac{\frac{P_h}{P_h+P_0}}{1-\frac{P_h}{P_h+P_0}} = \frac{\frac{P_h}{P_h+P_0}}{\frac{P_h+P_0-P_h}{P_h+P_0}} = \frac{P_h}{P_0}$

En sommant sous $l = 1, \ldots, H$ ce rapport, l'équation devient :

$$\sum_{l=1}^{H} \frac{P_l}{P_0} = \frac{1}{P_0} \sum_{l=1}^{H} P_l = \frac{1}{P_0} \left[\left(\underbrace{\sum_{l=1}^{H} P_l + P_0}_{=1} \right) - P_0 \right] = \frac{1}{P_0}(1 - P_0) = \frac{1 - P_0}{P_0} = \frac{1}{P_0} - 1$$

En réarrangeant cette équation, la probabilité de la modalité de référence P_0 est :

$$P_0 = \left[1 + \sum_{l=1}^{H} \frac{P_l}{P_0} \right]^{-1} = \left[1 + \sum_{l=1}^{H} G(\gamma_l' x) \right]^{-1}$$

De l'équation $\frac{P_h}{P_0} = G(\gamma_h' x)$, la probabilité P_h de choisir la modalité h s'écrit :

$$P_h = P_0 G(\gamma_h' x) = \frac{G(\gamma_h' x)}{1 + \sum_{l=1}^{H} G(\gamma_l' x)}$$

Les probabilités P_h, $h = 0, \ldots, H$ et les variables explicatives des modèles multinomiaux varient d'un individu i à l'autre. Il faut donc les indicer par i. Ces probabilités deviennent :

$$P_{hi} = \frac{G(\gamma_h' x_i)}{1 + \sum_{l=1}^{H} G(\gamma_l' x_i)}, \quad h = 1, \ldots, H$$

$$P_{0i} = \left[1 + \sum_{l=1}^{H} G(\gamma_l' x_i) \right]^{-1}$$

De cette formulation des modèles multinomiaux, on peut spécifier les modèles logit multinomiaux qui supposent que la fonction cumulative $G(\cdot)$ s'exprime comme une fonction exponentielle $\exp(\cdot)$. Les probabilités deviennent :

$$P_{hi} = \frac{\exp(\gamma_h' x_i)}{1 + \sum_{l=1}^{H} \exp(\gamma_l' x_i)}, \quad h = 1, \ldots, H$$

$$P_{0i} = \left[1 + \sum_{l=1}^{H} \exp(\gamma_l' x_i) \right]^{-1}$$

Il existe une formulation équivalente et générale de ces probabilités. Il suffit de poser γ_h égale à $(\beta_h - \beta_0)$, $h = 0, \ldots, H$:

$$P_{hi} = \frac{\exp((\beta_h - \beta_0)' x_i)}{1 + \sum_{l=1}^{H} \exp((\beta_l - \beta_0)' x_i)} = \frac{\exp((\beta_h - \beta_0)' x_i)}{\underbrace{\exp((\beta_0 - \beta_0)' x_i)}_{=1} + \sum_{l=1}^{H} \exp((\beta_l - \beta_0)' x_i)}$$

$$= \frac{\exp((\beta_h - \beta_0)'x_i)}{\sum_{l=0}^{H} \exp(\beta_l - \beta_0)'x_i)} = \frac{\exp(-\beta_0' x_i)\exp(\beta_h' x_i)}{\exp(-\beta_0' x_i)\sum_{l=0}^{H}\exp(\beta_l' x_i)} = \frac{\exp(\beta_h' x_i)}{\sum_{l=0}^{H}\exp(\beta_l' x_i)}, \quad h = 1, \ldots, H$$

et

$$P_{0i} = \frac{1}{1 + \sum_{l=1}^{H}\exp((\beta_l - \beta_0)'x_i)} = \frac{\exp((\beta_0 - \beta_0)'x_i)}{\underbrace{\exp((\beta_0 - \beta_0)'x_i)}_{=1} + \sum_{l=1}^{H}\exp((\beta_l - \beta_0)'x_i)}$$

$$= \frac{\exp((\beta_0 - \beta_0)'x_i)}{\sum_{l=0}^{H}\exp((\beta_l - \beta_0)'x_i)} = \frac{\exp(-\beta_0' x_i)\exp(\beta_0' x_i)}{\exp(-\beta_0' x_i)\sum_{l=0}^{H}\exp(\beta_l' x_i)} = \frac{\exp(\beta_0' x_i)}{\sum_{l=0}^{H}\exp(\beta_l' x_i)}$$

Donc

$$P_{hi} = \frac{\exp(\beta_h' x_i)}{\sum_{l=0}^{H}\exp(\beta_l' x_i)}, \quad h = 0, \ldots, H$$

Cette spécification est identique à celle avec les γ_h si on pose $\beta_0 = 0$.

Cette expression des probabilités suppose que les paramètres dépendent des modalités h de la variable dépendante y_i de l'individu i et que les variables explicatives dépendent, quant à elles, seulement des individus i. Cette approche se nomme "modèle logit généralisé" [2]. Toutefois, cette formulation peut être modifiée en émettant l'hypothèse que les variables explicatives dépendent aussi bien des individus i que des modalités h et que les paramètres sont constants par rapport aux individus et aux modalités. Le modèle devient le modèle logit conditionnel et les probabilités de choix h pour l'individu i s'écrivent :

$$P_{hi} = \frac{\exp(\beta'(x_{hi} - x_{0i}))}{\sum_{l=0}^{H}\exp(\beta'(x_{li} - x_{0i}))} = \frac{\exp(-\beta' x_{0i})\exp(\beta' x_{li})}{\sum_{l=0}^{H}\exp(-\beta' x_{0i})\exp(\beta' x_{li})}$$

$$= \frac{\exp(\beta' x_{hi})}{\sum_{l=0}^{H}\exp(\beta' x_{li})}, \quad h = 0, \ldots, H$$

Il faut donc changer les x_{hi} de sorte que x_{0i} soit égal à 0 [3].

Algébriquement, ces deux spécifications sont identiques. En effet, il suffit d'écrire dans le modèle logit généralisé,

$$\beta_h' = [0', \ldots, \beta', \ldots, 0'] \quad \text{et} \quad x_i = [x_{0i}, \ldots, x_{hi}, \ldots, x_{Hi}],$$

[2] Le modèle logit généralisé est généralement confondu dans la littérature avec le modèle logit multinomial (MNL).

[3] Cependant, les β n'ont pas besoin d'être normalisés.

ou, respectivement, dans le modèle logit conditionnel,

$$\beta' = [\beta'_0, \ldots, \beta'_h, \ldots, \beta'_H] \quad \text{et} \quad x_{hi} = [0, \ldots, x_i, \ldots, 0]$$

pour obtenir l'égalité $\beta' x_{hi} = \beta'_h x_i$.

Une combinaison de ces deux spécifications existe et se nomme "modèle logit mixte". Les probabilités associées à ce modèle s'écrivent de la manière suivante :

$$P_{hi} = \frac{\exp(\beta'_h x_i + \beta' x_{hi})}{\sum_{l=0}^{H} \exp(\beta'_h x_i + \beta' x_{hi})}, \quad h = 0, \ldots, H$$

où $\beta_0 = 0$ et en modifiant les x_{hi} tels que $x_{0i} = 0$.

3.1.2 Probabilités du modèle logit : démonstration

Dans les modèles multinomiaux, la variable expliquée qualitative peut être une variable ordonnée ou non ordonnée. Les modèles MNL de l'analyse conjointe utilisent uniquement des variables dépendantes non ordonnées. La raison est qu'une variable modélisant le choix entre plusieurs alternatives (ou modalités) est généralement non ordonnée. Les modèles logit multinomiaux des variables non ordonnées comprennent plusieurs types de spécifications mentionnées ci-dessus, à savoir le modèle logit généralisé, le modèle logit conditionnel et le modèle logit mixte. L'aspect commun de ces modèles est de supposer que les erreurs ε_{hi} associées aux utilités des alternatives sont indépendamment et identiquement distribuées selon une loi appartenant à la classe des doubles exponentielles négatives, par exemple la loi de Gumbel (ou loi de valeur extrême de type I). Le modèle logit peut également être obtenu en spécifiant une loi de Weibull. Les fonctions de densité et de répartition de la loi de Gumbel se définissent de la manière suivante :

$$\text{Fonction de densité} \quad f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} \exp[-e^{-\mu(\varepsilon-\eta)}] = \mu e^{-\mu(\varepsilon-\eta)} F(\varepsilon), \quad \mu > 0$$

$$\text{Fonction de répartition} \quad F(\varepsilon) = \exp[-e^{-\mu(\varepsilon-\eta)}]$$

où η est un paramètre de localisation et μ est un paramètre d'échelle positive.

Les probabilités associées au modèle multinomial à variable dépendante non ordonnée découlent de ces distributions et de la logique des comparaisons d'utilités qui décrit l'alternative choisie comme étant celle ayant l'utilité la plus élevée. Cette utilité s'écrit :

$$U_{hi} = v_{hi} + \varepsilon_{hi}, \qquad h = 0, \ldots, H$$

La logique des comparaisons d'utilités se formalise de la manière suivante :

$$\text{Choix de la modalité } h \quad \Longleftrightarrow \quad U_{hi} > U_{li} \ \forall \ l \neq h \quad \text{ou} \quad U_{hi} = \max_{l=0,\ldots,H}(U_{li}),$$

Par conséquent, en appliquant cette logique et en supposant que les erreurs ε_{hi} suivent une loi de Gumbel, la probabilité que l'individu i choisisse l'alternative h s'écrit donc :

$$P_{hi} = \frac{\exp(v_{hi})}{\sum_{l=0}^{H} \exp(v_{li})}$$

La démonstration de ce résultat découle des propriétés des distributions de Gumbel. Ces propriétés sont les suivantes :

1. Le mode vaut η.
2. La moyenne est $\eta + \gamma/\mu$, où γ est la constante d'Euler (~ 0.577).
3. La variance vaut $\pi^2/6\mu^2$.
4. Si ε est distribué selon un loi de Gumbel de paramètres (η, μ) et V et $\alpha > 0$ sont des scalaires quelconques constants, alors $\alpha\varepsilon + V$ est distribué selon un loi de Gumbel de paramètres $(\alpha\eta + V, \mu/\alpha)$.
5. Si ε_1 et ε_2 sont des variables indépendantes de la distribution de Gumbel de paramètres (η_1, μ) et (η_2, μ) respectivement, alors $\varepsilon^* = \varepsilon_1 - \varepsilon_2$ est distribué par une loi logistique :

$$F(\varepsilon^*) = \frac{1}{1 + e^{\mu(\eta_2 - \eta_1 - \varepsilon^*)}}$$

6. Si ε_1 et ε_2 sont indépendants et suivent une distribution de Gumbel de paramètres (η_1, μ) et (η_2, μ) respectivement, alors

$$\max(\varepsilon_1, \varepsilon_2)$$

suit une loi de Gumbel de paramètres

$$\left(\frac{1}{\mu}\ln(e^{\mu\eta_1} + e^{\mu\eta_2}), \mu\right)$$

7. Le corollaire de la propriété 6 est que si $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$ sont des variables aléatoires indépendantes distribuées selon une loi de Gumbel de paramètres $(\eta_1, \mu), (\eta_2, \mu), \ldots,$ (η_J, μ) respectivement, alors $\max(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$ suit une loi de Gumbel de paramètres

$$\left(\frac{1}{\mu} \ln \sum_{j=1}^{J} e^{\mu \eta_j}, \mu\right)$$

De ces propriétés des distributions de Gumbel, on peut déduire le modèle logit multinomial. Cette démonstration, tiré du livre *Discrete Choice Analysis* de Ben-Akiva et Lerman (1985), est une variante de celle de Domencich et McFadden (1975). Dans cette démonstration, le paramètre de localisation η est supposé être égal à 0 pour toutes les erreurs ε_{hi}. Selon la logique des comparaisons d'utilités, la probabilité de choisir l'alternative 0 est la suivante :

$$P_{0i} = \left[U_{0i} \geq \max_{l=1,\ldots,H} U_{li}\right] = \left[v_{0i} + \varepsilon_{0i} \geq \max_{l=1,\ldots,H} (v_{li} + \varepsilon_{li})\right]$$

Soit

$$U_{0i}^* = \max_{l=1,\ldots,H} U_{li} = \max_{l=1,\ldots,H} (v_{li} + \varepsilon_{li})$$

De la propriété 7, U_{0i}^* suit une distribution de Gumbel de paramètres

$$\left(\frac{1}{\mu} \ln \sum_{l=1}^{H} e^{\mu v_{li}}, \mu\right)$$

En utilisant la propriété 4, $U_{0i}^* = v_{0i}^* + \varepsilon_{0i}^*$ suit une loi de Gumbel dont le paramètre de localisation vaut

$$v_{0i}^* = \frac{1}{\mu} \ln \sum_{l=1}^{H} e^{\mu v_{li}}$$

et ε_{0i}^* est également distribué selon une loi de Gumbel de paramètre $(0, \mu)$.

Finalement, la probabilité de choisir l'alternative 0

$$P_{0i} = [v_{0i} + \varepsilon_{0i} \geq v_{0i}^* + \varepsilon_{0i}^*] = [(v_{0i}^* + \varepsilon_{0i}^*) - (v_{0i} + \varepsilon_{0i}) \leq 0]$$

se résout en appliquant la propriété 5 des lois de Gumbel. En effet,

$$P_{0i} = \frac{1}{1 + e^{\mu(v_{0i}^* - v_{0i})}} = \frac{e^{\mu v_{0i}}}{e^{\mu v_{0i}} + e^{\mu v_{0i}^*}}$$

En remplaçant

$$v_{0i}^* = \frac{1}{\mu} \ln \sum_{l=1}^{H} e^{\mu v_{li}}$$

la probabilité de choisir l'alternative 0 devient

$$P_{0i} = \frac{e^{\mu v_{0i}}}{e^{\mu v_{0i}} + e^{\mu \frac{1}{\mu} \ln \sum_{l=1}^{H} e^{\mu v_{li}}}} = \frac{e^{\mu v_{0i}}}{\sum_{l=0}^{H} e^{\mu v_{li}}}$$

La procédure est identique pour calculer les probabilités P_{hi} de choisir l'alternative h pour $h = 1, \ldots, H$. De plus, le paramètre μ, qui apparaît dans l'équation de probabilité, n'est pas identifiable. Par conséquent, il est généralement supposé être égal à 1.

Cette probabilité diffère selon que le modèle utilisé est un modèle logit généralisé, un logit conditionnel ou un logit mixte. En effet, les formes de l'utilité U_{hi} ou de v_{hi} ne sont pas identiques d'un modèle à l'autre. Dans le modèle logit généralisé, l'utilité dépend uniquement des caractéristiques x_i de l'individu i et les paramètres associés β_h diffèrent d'une modalité à l'autre :

$$v_{hi} = \beta_h' x_i$$

Dans le modèle logit conditionnel, les variables explicatives sont fonction de l'individu i et des modalités h de la variable y_i, alors que les paramètres associés sont constants :

$$v_{hi} = \beta' x_{hi}$$

Le modèle logit mixte, quant à lui, combine les spécifications du modèle logit généralisé et du modèle logit conditionnel :

$$v_{hi} = \beta_h' x_i + \beta' x_{hi}$$

3.1.3 La fonction de vraisemblance du modèle logit

Les paramètres β_h' et/ou β' des modèles logit multinomiaux sont estimés par la méthode du maximum de vraisemblance. La fonction de vraisemblance du modèle multinomial s'écrit en fonction des probabilités P_{hi} et des variables latentes y_{hi} :

$$L = \prod_{i=1}^{N} P_{0i}^{y_{0i}} P_{1i}^{y_{1i}} \ldots P_{Hi}^{y_{Hi}} = \prod_{i=1}^{N} \prod_{h=0}^{H} P_{hi}^{y_{hi}}$$

où

$$y_{hi} = \begin{cases} 1 & \text{si la modalité } h \text{ est choisie} \\ 0 & \text{sinon} \end{cases}$$

La méthode du maximum de vraisemblance maximise le logarithme de cette fonction de vraisemblance :

$$\max \log L = \sum_{i=1}^{N} \sum_{h=0}^{H} y_{hi} \log P_{hi}$$

Dans les modèles logit généralisés, où $v_{hi} = \beta'_h x_i$, les équations d'estimation des paramètres β'_h sont :

$$\sum_{i=1}^{N} (y_{hi} - P_{hi}) x_i = 0 \quad h = 0, \ldots, H$$

Ces équations découlent de :

$$\frac{\partial \log L}{\partial \beta_h} = 0$$

En effet,

$$\frac{\partial \log L}{\partial \beta_h} = \sum_{i=1}^{N} \left[y_{hi} \underbrace{\frac{\partial \log P_{hi}}{\partial \beta_h}}_{= \frac{\frac{\partial P_{hi}}{\partial \beta_h}}{P_{hi}}} + \sum_{l=0 \ l \neq h}^{H} y_{li} \underbrace{\frac{\partial \log P_{li}}{\partial \beta_h}}_{= \frac{\frac{\partial P_{li}}{\partial \beta_h}}{P_{li}}} \right] = 0$$

où

$$\frac{\partial P_{hi}}{\partial \beta_h} = P_{hi}(1 - P_{hi}) x_i, \quad \frac{\partial P_{hi}}{\partial \beta_l} = -P_{hi} P_{li} x_i \quad \text{et} \quad \frac{\partial P_{0i}}{\partial \beta_h} = -P_{hi} P_{0i} x_i$$

avec

$$P_{hi} = \frac{\exp(\beta'_h x_i)}{1 + \sum_{l=1}^{H} \exp(\beta'_h x_i)} \qquad (h = 1, \ldots, H) \quad et \quad P_{0i} = \frac{1}{1 + \sum_{l=1}^{H} \exp(\beta'_l x_i)} \quad \text{(référence)}$$

Donc

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^{N} \left[\frac{y_{hi}}{P_{hi}} P_{hi}(1 - P_{hi}) + \sum_{l=0 \ l \neq h}^{H} \frac{y_{li}}{P_{li}} (-P_{li} P_{hi}) \right] x_i$$

$$= \sum_{i=1}^{N} \left[y_{hi} - P_{hi} \underbrace{\left(\sum_{l=0}^{H} y_{li} \right)}_{=1} \right] x_i = \sum_{i=1}^{N} (y_{hi} - P_{hi}) x_i = 0$$

3.1.4 Résolution par la méthode du gradient

Les équations découlant de la maximisation du logarithme de la fonction de vraisemblance n'étant pas résolvables analytiquement, on utilise une méthode numérique, comme celle du gradient ou de Newton-Raphson, décrite dans "*Stated Choice Methods : Analysis and Application*" par Louviere, Hensher et Swait (2000). Le concept général d'une technique de recherche de gradient pour la maximisation d'une fonction $\log L(\theta)$ ($\theta = (\beta_0, \ldots, \beta_H)$) est de partir d'une solution initiale $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_r^0) = \{(\beta_0, \ldots, \beta_H)_1^0, \ldots, (\beta_0, \ldots, \beta_H)_r^0\}$ et par une série d'itérations se déplacer vers d'autres solutions, $\theta^1, \theta^2, \ldots,$ de manière à toujours améliorer (augmenter) la valeur de $\log L(\theta)$. Le changement d'itération est donné par :

$$\theta^{p+1} = \theta^p + d^p t,$$

où $\theta^p = (\theta_1^p, \theta_2^p, \ldots, \theta_r^p)$ et $d^p = (d_1^p, d_2^p, \ldots, d_r^p)$ est un vecteur de direction choisi de telle sorte que la valeur de $\log L(\theta)$ augmente lorsque la solution θ bouge dans la direction de θ^p et t est la taille du pas (scalaire) définissant la grandeur du mouvement dans la direction d^p.

$$\theta^{p+1} = (\theta_1^p, \theta_2^p, \ldots, \theta_r^p) + (d_1^p, d_2^p, \ldots, d_r^p) t$$

La taille du pas optimal peut être trouvée en résolvant l'équation suivante :

$$\frac{\partial}{\partial t} \log L(\theta^p + d^p t) = 0$$

Une convergence rapide est souvent trouvée par l'utilisation de la méthode de Newton-Raphson dans laquelle d^p est déterminé par la solution de l'équation matrice-vecteur

$$\left[\frac{\partial^2 \log L(\theta)}{\partial \theta_l \partial \theta_k} \right] [d_k^p] = \left[-\frac{\partial \log L(\theta)}{\partial \theta_l} \right]$$

où $k = 1, 2, \ldots, \quad l = 1, 2, \ldots, r$, et les dérivées premières et deuxièmes de $\log L(\theta)$ sont évaluées à θ^p. Dans le cas où $\log L(\theta)$ a une forme quadratique, par exemple, la méthode de Newton-Raphson converge en une seule itération avec une taille du pas de 1.

En ce qui concerne le modèle logit conditionnel et le modèle logit mixte, le processus est identique malgré une différence dans les dérivées partielles. Nous ne développerons donc pas le processus d'estimation des paramètres de ces modèles dans la thèse.

L'analyse conjointe par le choix estime des utilités partielles (β) en s'inspirant de ces modèles multinomiaux. Son application se rapproche davantage des modèles logit conditionnels que des modèles logit généralisés. Cependant, la littérature sur l'analyse conjointe par le choix dit utiliser les modèles logit multinomiaux (MNL) dans ses estimations. Or, ces modèles ne doivent pas être associés aux modèles logit généralisés. Ils ramènent, en réalité, à l'appellation générale des modèles multinomiaux avec une spécification logit. Dans la section suivante, nous définissons ces modèles logit multinomiaux (MNL) dans le cadre de l'analyse conjointe par le choix.

3.2 Le modèle MNL de l'analyse conjointe et l'option "zéro"

3.2.1 Le modèle MNL

L'analyse conjointe par le choix étudie les comportement déclarés des consommateurs en spécifiant un modèle logit multinomial MNL. Cependant, cette approche MNL de l'analyse conjointe diffère sur certains points de celle présentée dans la section précédente. En effet, le modèle logit multinomial MNL, décrit précédemment, suppose que les consommateurs font un seul choix dans un ensemble unique d'alternatives. Le modèle MNL de l'analyse conjointe doit, par contre, intégrer plusieurs choix uniques par ensembles ou sous-ensembles d'alternatives. Plus précisément, dans l'analyse conjointe par le choix, plusieurs sous-ensembles regroupant un certain nombre d'alternatives (deux ou plus) sont présentés aux consommateurs qui doivent choisir un produit dans chacun de ces sous-ensembles. Par conséquent, il y a autant de choix par individu que de sous-ensembles. De plus, ces sous-ensembles sont créés grâce aux plans dits d'équilibre incomplets. Les profils présentés dans les sous-ensembles sont, quant à eux, déterminés par des méthodes similaires aux plans d'expérience de l'approche classique [4]. La méthode par le choix utilise donc deux types de plans d'expérience par opposition à la méthode classique.
De plus, un plan d'équilibre incomplet supplémentaire doit être ajouté à l'étude si le modèle estime des utilités partielles agrégées. Ce plan permet de déterminer les groupes d'individus par ensembles de sous-ensembles de produits. En effet, dans le modèle agrégé, tous les individus ne donnent pas leurs choix pour les mêmes sous-ensembles de choix. Par

[4] Les plans d'équilibre incomplets et les méthodes de plans d'expérience de l'approche classique de l'analyse conjointe ont été décrits dans un chapitre 2 des plans d'expérience.

conséquent, les plans d'équilibre incomplet déterminent différents groupes d'individus et les associent à un ensemble de sous-ensembles de produits qui est différent d'un groupe à l'autre. Le modèle individuel, quant à lui, propose des sous-ensembles d'alternatives similaires aux individus.

Décrivons désormais les modèles logit multinomiaux (MNL) applicables à l'analyse conjointe par le choix. Cette approche est déduite des modèles logit conditionnels. Cependant, les variables explicatives du modèle de l'analyse conjointe ne dépendent pas des mêmes caractéristiques que celles du modèle logit conditionnel. En effet, dans l'approche par l'analyse conjointe, elles sont fonctions des sous-ensembles d'alternatives et des alternatives, alors que dans les modèles logit conditionnels, elles dépendent des individus et des caractéristiques des modalités. Par conséquent, l'utilité de l'alternative ou du produit $h = 0, \ldots, H$ du sous-ensemble $s = 1, \ldots, S$ de l'analyse conjointe par le choix s'écrit :

$$U_{hs} = v_{hs} + \varepsilon_{hs} = \beta' x_{hs} + \varepsilon_{hs}$$

où x_{hs} sont les caractéristiques ou les niveaux d'attributs du produit h du sous-ensemble s et ε_{hs} est l'erreur aléatoire suivant une distribution de Gumbel de paramètres $(0, 1)$. Dans la suite de la thèse, nous supposerons qu'il y a le même nombre h d'alternatives par sous-ensemble s [5]. De plus, nous posons l'hypothèse d'indépendance entre les réponses ou les choix par sous-ensemble d'alternatives provenant du même individu.

L'utilité formulée ci-dessus n'est valable que pour le modèle conjoint de choix individuel. Dans l'analyse conjointe agrégée, il faut ajouter à l'utilité U_{hs} et aux variables explicatives x_{hs} l'indice $i = 1, \ldots, N$ [6] représentant les individus répondant aux différents ensembles de sous-ensembles d'alternatives. L'utilité du modèle de choix agrégé devient :

$$U_{hsi} = v_{hsi} + \varepsilon_{hsi} = \beta' x_{hsi} + \varepsilon_{hsi}$$

dont la distribution des erreurs ε_{hsi} est identique à celle des erreurs ε_{hs} du modèle individuel.

En utilisant le principe de comparaison d'utilités dans la modélisation du choix par sous-ensemble et l'hypothèse de distribution des erreurs ε_{hs}, respectivement des erreurs ε_{hsi},

[5] Cependant, la modélisation pour un nombre différent d'alternatives par sous-ensemble est identique à celle présentée ici.

[6] Il serait plus correct d'indicer x_{hs} selon l'ensemble de sous-ensembles d'alternatives auquel l'individu est soumis. En effet, les valeurs des niveaux d'attributs des consommateurs ayant le même sous-ensemble ne varient pas d'un individu à l'autre. Par contre, ces valeurs changent lorsque l'ensemble des sous-ensembles varie. Nous conserverons, toutefois, cette notation avec l'indice i afin de ne pas alourdir les notations.

selon une loi de Gumbel, les probabilités de choisir le produit h dans le sous-ensemble s du modèle individuel, respectivement du modèle agrégé, se formulent de la manière suivante :

$$P_{hs} = \frac{exp(\beta' x_{hs})}{\sum_{l=0}^{H} \exp(\beta' x_{ls})}, \quad h = 0, \ldots, H \quad \text{et} \quad s = 1, \ldots, S$$

$$P_{hsi} = \frac{exp(\beta' x_{hsi})}{\sum_{l=0}^{H} \exp(\beta' x_{lsi})}, \quad h = 0, \ldots, H, \quad s = 1, \ldots, S \quad \text{et} \quad i = 1, \ldots, N$$

Dans la formulation générale des modèles logit multinomiaux, il existe une contrainte sur les probabilités P_{hi} :

$$\sum_{l=0}^{H} P_{li} = 1$$

Dans le modèle logit multinomial individuel de l'analyse conjointe, il y a autant de contraintes que de sous-ensembles :

$$\sum_{l=0}^{H} P_{ls} = 1, \quad s = 1, \ldots, S$$

dont certaines sont identiques pour les individus ayant le même ensemble de sous-ensembles d'alternatives.

Dans le modèle logit multinomial agrégé de l'analyse conjointe, il y a $s \times i$ contraintes :

$$\sum_{l=0}^{H} P_{lsi} = 1, \quad s = 1, \ldots, S \quad \text{et} \quad i = 1, \ldots, N$$

De ces contraintes découlent l'hypothèse qu'une variable explicative de chaque sous-ensemble $s = 1, \ldots, S$ doit être égale 0 ($x_{0s} = 0$) dans le modèle individuel. En effet,

$$P_{hs} = \frac{exp(\beta' x_{hs})}{\sum_{l=0}^{H} \exp(\beta' x_{ls})} = \frac{exp(\beta'(x_{hs} - x_{0s}))}{\sum_{l=0}^{H} \exp(\beta'(x_{ls} - x_{0s}))}, \quad h = 0, \ldots, H \quad \text{et} \quad s = 1, \ldots, S$$

Par conséquent, les variables explicatives d'un sous-ensemble s du modèle de l'analyse conjointe sont les différences des caractéristiques.

Dans le modèle agrégé, les probabilités P_{hsi} considèrent les différences de niveaux d'attributs par sous-ensemble s et par individu i :

$$P_{hsi} = \frac{exp(\beta' x_{hsi})}{\sum_{l=0}^{H} \exp(\beta' x_{lsi})} = \frac{exp(\beta'(x_{hsi} - x_{0si}))}{\sum_{l=0}^{H} \exp(\beta'(x_{lsi} - x_{0si}))},$$

$$h = 0, \ldots, H \quad s = 1, \ldots, S \quad \text{et} \quad i = 1, \ldots, N$$

3.2.2 La fonction de vraisemblance

L'estimation des β ou des utilités partielles de ces modèles logit multinomiaux individuels, respectivement agrégés, s'effectue par la méthode du maximum de vraisemblance. La fonction de vraisemblance L du modèle individuel est la suivante :

$$L = \prod_{s=1}^{S} \prod_{h=0}^{H} P_{hs}^{y_{hs}}$$

avec

$$y_{hs} = \begin{cases} 1 & \text{lorsque le produit } h \text{ du sous-ensemble } s \text{ est choisi} \\ 0 & \text{sinon} \end{cases}$$

La fonction de vraisemblance l du modèle agrégé s'écrit :

$$L = \prod_{s=1}^{S} \prod_{h=0}^{H} \prod_{i=1}^{N} P_{hsi}^{y_{hsi}}$$

avec

$$y_{hsi} = \begin{cases} 1 & \text{lorsqu'un individu } i \text{ choisit le produit } h \text{ du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

Les log-vraisemblances correspondantes sont :

$$\log L = \sum_{s=1}^{S} \sum_{h=0}^{H} y_{hs} \log P_{hs}$$

et

$$\log L = \sum_{s=1}^{S} \sum_{h=0}^{H} \sum_{i=1}^{N} y_{hsi} \log P_{hsi}$$

De $\frac{\partial \log L}{\partial \beta} = 0$, on déduit les équations de résolution des paramètres β :

$$\frac{\partial \log L}{\partial \beta} = \sum_{s=1}^{S}\sum_{h=0}^{H}\sum_{i=1}^{N} y_{hsi}\frac{\partial \log P_{hsi}}{\partial \beta} = \sum_{s=1}^{S}\sum_{h=0}^{H}\sum_{i=1}^{N} y_{hsi}\frac{\partial P_{hsi}}{\partial \beta}\frac{1}{P_{hsi}}$$

où

$$\frac{\partial P_{hsi}}{\partial \beta} = P_{hsi} \times \frac{\sum_{l=0}^{H}(x_{hsi} - x_{lsi})\exp(\beta' x_{lsi})}{\sum_{l=0}^{H} exp(\beta' x_{lsi})}$$

D'où

$$\frac{\partial \log L}{\partial \beta} = \sum_{s=1}^{S}\sum_{h=0}^{H}\sum_{i=1}^{N} y_{hsi}\frac{\sum_{l=0}^{H}(x_{hsi} - x_{lsi})\exp(\beta' x_{lsi})}{\sum_{l=0}^{H} exp(\beta' x_{lsi})} = 0$$

Ces équations ne sont pas résolvables analytiquement et il faut, par conséquent, utiliser un algorithme du type de celui de la méthode du gradient.

3.2.3 Formulation matricielle

L'expression de la log-vraisemblance peut être écrite matriciellement en créant $H+1$ vecteurs $y_0, \ldots, y_H$ $((s \times i) \times 1)$ représentant les choix h par sous-ensemble s des individus i et $P_0, \ldots, P_H$ correspondant aux probabilités pour le choix h du sous-ensemble s de l'individu i. En effet, les vecteurs des choix s'écrivent [7] :

$$y_0 = [y_{0si}]_{s=1,\ldots,S \ et \ i=1,\ldots,N} = \begin{pmatrix} y_{011} \\ \vdots \\ y_{0SN} \end{pmatrix}$$

$$\vdots$$

$$y_H = [y_{Hsi}]_{s=1,\ldots,S \ et \ i=1,\ldots,N} = \begin{pmatrix} y_{H11} \\ \vdots \\ y_{HSN} \end{pmatrix}$$ [8]

[7] Pour le modèle individuel, il suffit de poser $i = 1$

[8] Ici les données sont rangées ou ordonnées selon l'alternative h. En annexe 3, elles sont ordonnées selon la modalité h, l'individu i et le sous-ensemble s.

Les vecteurs des probabilités sont :

$$P_0 = [P_{0si}]_{s=1,\ldots,S \ \ et \ \ i=1,\ldots,N} = \begin{pmatrix} P_{011} \\ \vdots \\ P_{0SN} \end{pmatrix}$$

$$\vdots$$

$$P_H = [P_{Hsi}]_{s=1,\ldots,S \ \ et \ \ i=1,\ldots,N} = \begin{pmatrix} P_{H11} \\ \vdots \\ P_{HSN} \end{pmatrix}$$

La log-vraisemblance s'écrit en fonction de ces vecteurs de la manière suivante :

$$\log L = y_0'(\log P_0) + \ldots + y_H'(\log P_H) = y'(\log P)^9$$

où

$$y = \begin{pmatrix} y_0 \\ \vdots \\ y_H \end{pmatrix} \qquad (\log P)' = (\log P_0 \ldots \log P_H)$$

Les probabilités peuvent s'écrire en fonction des matrices $x_0, \ldots, x_H$ $((s \times i) \times K)$ des K variables explicatives de $H+1$ alternatives. Ces matrices se formulent comme suit :

$$x_0 = \begin{pmatrix} x_{011}^1 & \ldots & x_{011}^K \\ \vdots & \ddots & \vdots \\ x_{0SN}^1 & \ldots & x_{0SN}^K \end{pmatrix} = \begin{pmatrix} x_{011} \\ \vdots \\ x_{0SN} \end{pmatrix}$$

$$\vdots$$

$$x_H = \begin{pmatrix} x_{H11}^1 & \ldots & x_{H11}^K \\ \vdots & \ddots & \vdots \\ x_{HSN}^1 & \ldots & x_{HSN}^K \end{pmatrix} = \begin{pmatrix} x_{H11} \\ \vdots \\ x_{HSN} \end{pmatrix}^{10}$$

[9] $(\log P_h)' = \begin{pmatrix} \log P_{h11} & \ldots & \log P_{hSN} \end{pmatrix}$

[10] La matrice des variables explicatives X $((H+1) \times S \times N)$ du modèle est donnée par $X = \begin{pmatrix} x_0 \\ \vdots \\ x_H \end{pmatrix}$.
On remarque que les éléments de cette matrice ont été classés en fonction des modalités h. En annexe 3, on présente cette matrice ordonnée selon la modalité h, le sous-ensemble s et l'individu i.

Les probabilités en fonction de ces matrices sont :

$$P_0 = \frac{\exp((x_0 - x_0)\beta)}{\sum_{h=0}^{H} \exp((x_h - x_0)\beta)} = \frac{1}{\sum_{h=0}^{H} \exp((x_h - x_0)\beta)}$$

$$\vdots$$

$$P_H = \frac{\exp((x_H - x_0)\beta)}{\sum_{h=0}^{H} \exp((x_H - x_0)\beta)}$$

où le produit 0 est l'alternative de référence et $\beta' = (\beta_1 \dots \beta_K)$.

Matriciellement, ces expressions s'écrivent comme suit :

$$P_0 = \left(diag \left[\sum_{h=0}^{H} \exp((x_h - x_0)\beta) \right] \right)^{-1} \exp((x_0 - x_0)\beta) = diag \left(\left[\sum_{h=0}^{H} \exp((x_h - x_0)\beta) \right] \right)^{-1} \iota$$

$$\vdots$$

$$P_H = \left(diag \left[\sum_{h=0}^{H} \exp((x_H - x_0)\beta) \right] \right)^{-1} \exp((x_H - x_0)\beta)$$

où

$$\exp((x_h - x_0)\beta) = \exp \left[\begin{pmatrix} (x_{h11}^1 - x_{011}^1) & \dots & (x_{h11}^K - x_{011}^K) \\ \vdots & \ddots & \vdots \\ (x_{hSN}^1 - x_{0SN}^1) & \dots & (x_{hSN}^K - x_{0SN}^K) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \right]$$

$$= \exp \begin{pmatrix} (x_{h11}^1 - x_{011}^1)\beta_1 + \dots + (x_{h11}^K - x_{011}^K)\beta_K \\ \dots \\ (x_{hSN}^1 - x_{0SN}^1)\beta_1 + \dots + (x_{hSN}^K - x_{0SN}^K)\beta_K \end{pmatrix}$$

$$= \begin{pmatrix} \exp\left(((x_{h11}^1 - x_{011}^1)\beta_1 + \dots + (x_{h11}^K - x_{011}^K)\beta_K \right) \\ \dots \\ \exp\left((x_{hSN}^1 - x_{0SN}^1)\beta_1 + \dots + (x_{hSN}^K - x_{0SN}^K)\beta_K \right) \end{pmatrix}$$ [11]

[11] L'exponentielle est appliquée à chaque élément de la matrice.

et

$$\left(diag\left[\sum_{h=0}^{H}\exp((x_h - x_0)\beta)\right]\right)^{-1}$$

$$= \left(diag\begin{bmatrix} \sum_{h=0}^{H}\left(\exp\left((x_{h11}^1 - x_{011}^1)\beta_1 + \ldots + (x_{h11}^K - x_{011}^K)\beta_K\right)\right) \\ \ldots \\ \sum_{h=0}^{H}\left(\exp\left((x_{hSN}^1 - x_{0SN}^1)\beta_1 + \ldots + (x_{hSN}^K - x_{0SN}^K)\beta_K\right)\right) \end{bmatrix}\right)^{-1}$$

$$= \begin{pmatrix} \frac{1}{\sum_{h=0}^{H}\left(\exp\left((x_{h11}^1 - x_{011}^1)\beta_1 + \ldots + (x_{h11}^K - x_{011}^K)\beta_K\right)\right)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sum_{h=0}^{H}\left(\exp\left((x_{hSN}^1 - x_{0SN}^1)\beta_1 + \ldots + (x_{hSN}^K - x_{0SN}^K)\beta_K\right)\right)} \end{pmatrix}$$

Il est utile de remarquer que le modèle agrégé dispose d'un plus grand nombre d'observations par alternative $(s \times i)$ que le modèle individuel s. Par conséquent, le modèle agrégé peut modéliser les effets principaux et tous les effets spéciaux, tels que les effets d'interaction, les effets croisés et les effets spécifiques à l'alternative, alors que le modèle individuel ne tient compte que des effets principaux. Cependant, le modèle agrégé ne fournit que des préférences agrégées, alors qu'avec le modèle individuel, on peut "connaître les goûts d'un consommateur". L'avantage du modèle agrégé, toutefois, est l'estimation directe des parts de marché. Dans les modèles individuels de l'analyse conjointe classique ou par le choix, les parts de marché du produit analysé ne s'obtiennent pas directement dans les résultats de l'estimation des utilités partielles individuelles. Il faut les calculer en supposant une spécification pour cette probabilité ou proportion d'achat.

3.2.4 L'option "zéro" dans la littérature

Á la formulation du modèle MNL de l'analyse conjointe présentée ci-dessus, Elrod, Louviere et Krishnakumar (1992) ont rajouté une option de non-choix dite option "zéro". Cette option permet aux consommateurs de ne pas choisir de produits lorsqu'ils préfèrent attendre une meilleure offre (par exemple une promotion), lorsqu'ils préfèrent se renseigner d'avantage sur un produit pour faire de meilleurs choix ou lorsqu'ils n'apprécient aucun des produits proposés. Dans leur article, les auteurs proposent de modéliser ce non-choix en introduisant une constante dans le vecteur des paramètres β et en posant à 0 tous les niveaux des attributs de l'option "zéro", y compris la constante.

Matriciellement, les variables explicatives [12] du modèle de Elrod, Louviere et Krishnakumar (1992) peuvent s'écrire ainsi :

$$x_0 = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}^{13}$$

$$x_1 = \begin{pmatrix} 1 & x_{111}^2 & \dots & x_{111}^K \\ \vdots & \ddots & \vdots & \\ 1 & x_{1SN}^2 & \dots & x_{1SN}^K \end{pmatrix}$$

$$\vdots$$

$$x_H = \begin{pmatrix} 1 & x_{H11}^2 & \dots & x_{H11}^K \\ \vdots & \ddots & \vdots & \\ 1 & x_{HSN}^2 & \dots & x_{HSN}^K \end{pmatrix}$$

Elrod, Louviere et Krishnakumar (1992) proposent également dans leur article d'autres modèles intégrant les différents effets spéciaux. En effet, dans un de leurs modèles, ils ajoutent les effets d'interaction. Ceci implique une augmentation des matrices x_h du modèle de base (sans effets spéciaux) de $\sum_{k=2}^{K} \begin{pmatrix} K \\ k \end{pmatrix} = \sum_{k=2}^{K} \frac{K!}{k!(K-k)!}$ colonnes et également une augmentation du vecteur β de $\sum_{k=2}^{K} \frac{K!}{k!(K-k)!}$ lignes.

Les auteurs présentent également un modèle à effets croisés. Ces effets sont ajoutés à la matrice X présentée en annexe 3 du modèle MNL avec l'option "zéro" en créant plusieurs matrices identiques à X où l'on échange les lignes correspondant aux produits 1 et 2 des sous-ensembles $s = 1, \dots, S$, puis les lignes correspondant aux produits 1 et 3 des sous-ensembles $s = 1, \dots, S$ jusqu'aux lignes correspondant aux produits 1 et H des sous-ensembles $s = 1, \dots, S$, ensuite les lignes correspondant aux produits 2 et 3 des sous-ensembles $s = 1, \dots, S$ jusqu'aux lignes correspondant aux produits 2 et H des sous-ensembles $s = 1, \dots, S$ et ainsi de suite jusqu'aux lignes correspondant aux produits $(H-1)$ et H des sous-ensembles $s = 1, \dots, S$ [14]. Au total, on rajoute $\frac{H(H-1)}{2}$ matrices X_{hl} [15], $h, l = 1, \dots, H$ et $l > h$, à la matrice X ordonnée selon les alternatives h, les sous-ensembles s et les individus i et qui est formulée en annexe 3. Finalement, Elrod,

[12] Soit x_0 l'option "zéro" et x_h^1 la constante du modèle.
[13] On remarque que la constante est posée à 0 pour l'option "zéro".
[14] Les lignes correspondant à l'option "zéro" (x_0) ne sont pas interverties avec les produits.
[15] On échange la ligne correspondant au produit h avec celle correspondant au produit l.

Louviere et Krishnakumar (1992) décrivent un modèle incluant les effets d'interaction et les effets croisés.

Le modèle avec les effets croisés permet de tester l'hypothèse IIA des modèles MNL. En effet, cette hypothèse n'est pas respectée si les effets croisés sont significatifs. De plus, la grandeur et le signe de ces termes croisés indiquent l'attribut qui contribue au non-respect.

Dans "*Modeling Conjoint Choice Experiments with the Probit Model*", Haaijer (1999) présente plusieurs modèles avec l'option "zéro". Il décrit cette option comme une alternative de base permettant de mettre à la même échelle les utilités des divers sous-ensembles de choix. Il présente d'autres alternatives de base ayant cette fonctionnalité, telles qu'un profil régulier qui est constant sur tous les sous-ensembles de choix ou une marque courante. Les différentes spécifications intégrant l'option "zéro" sont le modèle MNL ayant une série de 0 pour décrire les valeurs des attributs de l'option "zéro", le modèle MNL avec des valeurs 0 pour l'option "zéro" et l'ajout d'un terme constant c_{nc} pour l'option de non choix et un modèle logit emboîté. L'auteur fait remarquer qu'une valeur de 0 pour un attribut linéaire, tel qu'un prix, suppose la gratuité du produit. Par conséquent, dans les résultats d'estimation, il apparaît un biais sur les variables linéaires liées à cette formulation de gratuité du prix. Nous pensons que ce ce biais persiste si les données sont codées par des variables indicatrices, car la valeur 0 dans un codage correspond généralement à la modalité de référence. Haaijer (1999) conclut en disant que l'ajout d'une constante c_{nc} pour l'option de non-choix améliore l'ajustement du modèle MNL de non choix et réduit le biais des attributs linéaires. Il propose une autre spécification pour l'option "zéro" qui permet de ne pas avoir de biais lié à un codage linéaire des variables. Cette spécification est le modèle logit emboîté qui ne considère plus l'option "zéro" comme une des alternatives. En effet, il considère un niveau lié au non choix et un autre lié au choix des produits. Par conséquent, le biais des paramètres linéaires n'apparaît plus car le modèle emboîté ne traite pas les 0 du non choix comme des niveaux.

Dans sa thèse, Haaijer (1999) n'estime les utilités partielles qu'avec une spécification MNL car il dit que ce modèle suffit pour les conclusions qu'il veut obtenir. Toutefois, il fait remarquer que le modèle MNL suppose l'indépendance des observations d'un individu. Il dit que pour pallier à ce problème, il faut une spécification multinomiale probit (MNP). Cette approche est développée dans un autre chapitre de sa thèse mais elle n'intègre pas l'option "zéro". De plus, il semble que l'ajout de l'option "zéro" ne respecte pas la propriété IIA des modèles MNL. En effet, avec l'option "zéro", les utilités des produits peuvent être corrélées. Par conséquent, l'alternative du modèle MNP devrait apporter de meilleurs résultats. Toutefois, son application étant plus complexe, c'est pourquoi, on

utilise plus fréquemment le modèle MNL.

Haaijer (1999) aborde également dans son texte les aspects avantageux à modéliser l'option "zéro". En effet, l'option de non choix rend plus réaliste la décision de choix et mène à de meilleures prédictions de pénétration du marché. Toutefois, elle permet à l'individu d'éviter des choix difficiles et ne fournit aucune information sur les préférences des produits. Dans la prochaine section de la thèse, nous abordons les différentes raisons psychologiques à ne pas choisir d'alternatives lorsque cette option est disponible ou à choisir aléatoirement un produit.

3.3 Les raisons psychologiques du non choix

Dans l'article de Dhar (1997), une des causes de non choix fréquemment mentionnée est que le consommateur a plus tendance à ne pas choisir lorsque la différence d'intérêt des alternatives d'un sous-ensemble est petite que lorsqu'elle est grande. Dhar (1997) montre dans son article que ce résultat a été observé par d'autres auteurs. En effet, selon Beattie et Barlas (1992), Festinger (1964) et Janis et Mann (1977), un refus défensif est une réponse probable à des choix difficiles. Tversky et Shafir (1992) disent qu'un non choix est plus probable à partir de sous-ensemble de choix où aucune des alternatives n'est dominante qu'à partir de sous-ensembles avec une alternative clairement supérieure. Kuhl (1986) et Sjoberg (1980) ont remarqué qu'il est difficile pour le répondant de maintenir son intention d'agir lorsqu'il y a des désirs ou des tentations compétitifs. Montgomery (1989) suggère que le répondant peut abandonner ou reporter son choix s'il échoue dans la tentative de trouver une structure de dominance pour une alternative prometteuse. Scholnick et Wing (1988) prétend qu'une situation de décision offrant beaucoup d'alternatives acceptables, dont aucune ne peut clairement être vérifiée comme étant la meilleure, peut créer un sentiment de confusion menant à une réticence de faire une action. Dans des études récentes, Baron et Ritov (1994), Ritov et Baron (1990) et Spranca, Minsk et Baron (1991) trouvent un biais systématique vers l'inaction dans la prise de décision du consommateur.

Le principe du non choix basé sur la différence d'intérêt ne coincide pas avec la théorie rationnelle de recherche citée dans l'article de Dhar (1997) et définie tel que l'option "zéro" doit être choisie lorsqu'aucune des alternatives n'est perçue comme attrayante, ou lorsqu'il y a des avantages pour des recherches plus approfondies (Karni et Schwarz 1977). Par contre, elle correspond à la recherche psychologique dans le domaine des processus pré-décisionnels qui suggère que les consommateurs peuvent ne pas décider de choisir pour éviter de faire des compromis difficiles (Tversky et Shafir (1992)). Tversky et Shafir (1992) définissent ce compromis difficile en se basant sur la notion de conflit. En effet, se-

lon les auteurs, un conflit apparaît lorsqu'un individu n'arrive pas à faire des compromis. Il en résulte que certaines décisions importantes ou insignifiantes deviennent difficiles. Les auteurs mentionnent également dans leur article la complexité de résolution du conflit due à l'incertitude des conséquences de l'action d'un individu et à la gêne provoquée par l'anticipation de la dissonance et du regret.

Tversky et Shafir (1992) s'intéressent également aux différences de théorie lorsque le conflit est ou n'est pas pris en considération. Ils montrent que la prise en compte du conflit n'est pas en accord avec la théorie de la maximisation d'utilité ou la théorie rationnelle du choix. Cette constatation est due au fait que la maximisation d'utilité ne suppose pas que l'option "zéro" puisse être due à une difficulté dans le choix. Cette théorie de maximisation prétend que le consommateur choisira le produit ayant l'utilité la plus élevée en ne supposant aucun lien entre la décision du consommateur et le conflit dû au compromis. Or, les auteurs affirment que le conflit influence l'état psychologique du consommateur et de ce fait son choix. Dans la littérature psychologique, ce comportement de non choix s'explique par le fait qu'un individu préfère les conséquences de l'inaction aux conséquences de l'action. En effet, le consommateur incertain préfère ne pas choisir qu'assumer les conséquences du choix, telles que le regret d'achat du produit. En outre, une des conséquences de l'inaction, c'est-à-dire du non choix, peut être la non disponibilité du produit. La théorie psychologique prétend donc que l'individu préfère risquer de ne plus obtenir le produit que regretter son achat. La maximisation d'utilité pense, quant à elle, que ce conflit n'influence pas le non choix car l'individu sélectionne l'option "aucune alternative" que lorsque les produits lui déplaisent. Toutefois, dans leur article, Tversky et Shafir (1992) démontrent le contraire. En effet, ils montrent par une application que la théorie rationnelle du choix n'est pas respectée lorsqu'il y a conflit. Dans leur application, ils présentent aux individus des paires d'alternatives et une option supplémentaire permettant de retarder son choix. Ils montrent par cet exemple que la proportion d'individus choisissant l'option supplémentaire augmente lorsque le conflit s'accroît. Ils font remarquer que cette situation est inversée avec la maximisation d'utilité. Par conséquent, dans la pratique, le principe de maximisation d'utilité n'est pas respecté dans certaines situations.

Le fondement de l'analyse conjointe réside dans le compromis. Dans l'approche par le choix, les utilités partielles ou les importances des attributs sont estimées selon le principe de comparaison d'utilité, c'est-à dire celui de la maximisation d'utilité. Partant de l'article de Tversky et Shafir (1992), ce principe de comparaison d'utilité devient donc inadéquate lorsque l'option "zéro" est prise en considération. Or, le non choix modélisé par Elrod, Louviere et Krishnakumar (1992) et par Haaijer (1999) qui ajoutent une série de 0 et/ou une constante au modèle MNL, ainsi que le non choix spécifié par Haaijer (1999) à l'aide

des modèles emboîtés, se base sur cette théorie de maximisation d'utilité. C'est pourquoi, dans le section suivante, nous développons l'option "zéro" avec un modèle qui ne suppose pas une utilité au non choix et n'associe une utilité qu'aux produits. Cette procédure permet également d'introduire le degré de conflit influençant l'option "zéro".

3.4 Le modèle de l'option "zéro" de l'analyse conjointe

La spécification économétrique de l'option "zéro" n'est pas énormément dépeinte dans la littérature sur l'analyse conjointe. Le logiciel CBC de Sawtooth Software estime le non choix en modifiant la méthode CBC (conjoint based-choice), qui correspond dans la littérature à l'approche conjointe par le choix. En effet, Sawtooth propose le non choix comme une option de la méthode CBC. De plus, comme l'approche CBC utilise les modèles logit multinomiaux MNL, on peut conclure que la spécification du CBC avec l'option "zéro" est un modèle MNL. Toutefois, les modifications économétriques de l'ajout du non choix au modèle CBC ne sont pas entièrement spécifiées. Par conséquent, il n'est pas possible de connaître le processus d'estimation employé par la méthode CBC lorsque l'option "zéro" est intégrée.

Quelques auteurs, cités précédemment, ont néanmoins décrit la modélisation de l'ajout de l'option "zéro" dans les modèles de choix MNL. En effet, Elrod, Louviere et Krishnakumar (1992) spécifient cette option en associant dans les modèles MNL une ligne de 0 à l'option de non choix et en ajoutant une constante à toutes les alternatives du modèle, à l'exception de l'option "zéro". Cependant, cette modélisation du non choix amène un biais comme le fait remarquer Haaijer (1999) dans sa thèse. Haaijer (1999) prétend également que l'ajout d'une constante au modèle améliore le modèle MNL de non choix. Il propose une autre spécification pour l'option "zéro" qui permet de ne pas avoir de biais lié à un codage linéaire des variables. Cette spécification est le modèle logit emboîté qui considère deux niveaux, à savoir celui lié au non choix et celui de la continuité du choix.

D'un point de vue statistique et économique, il n'y a pas ou très peu d'étude comparative entre l'ajout de l'option "zéro" et le simple modèle de choix estimé par une spécification MNL. Il est donc intéressant d'appliquer ces modèles à un même échantillon pour mesurer l'apport statistique et économique du non choix. Dans notre thèse, nous n'effectuons pas cette étude comparative. Nous préférons nous concentrer sur l'application, par un exemple concret, de notre spécification de l'option "zéro" et la vérification de l'efficacité de ce nouveau modèle. La raison est que la spécification de l'option "zéro" décrite dans la littérature ne semble pas satisfaisante. En effet, elle amène des biais lorsque les variables explicatives sont codées de manière linéaire. Le modèle logit emboîté proposé par Haaijer

(1999) semble préférable. Toutefois, il suppose encore un choix entre l'option "zéro" et la non option "zéro". Par conséquent, les variables explicatives du non choix dépendent encore des attributs et du sous-ensemble auquel ils appartiennent. Nous proposons donc, dans cette section, un nouveau modèle de l'option "zéro", qui n'exige pas plusieurs étapes dans sa spécification et qui intègre les concepts psychologiques influençant le non choix.

Notre spécification de l'option "zéro" s'inspire du principe des modèles quantitatifs censurés Tobit. Nous modifions ces modèles Tobit pour qu'ils soient adaptés aux modèles à variable dépendante qualitative. Cette modification est facile car les modèles Tobit sont inspirés des modèles qualitatifs à variables dépendantes dichotomiques. Cette nouvelle spécification est ensuite adaptée à l'analyse conjointe.

3.4.1 Les modèles censurés

Pour décrire la spécification du modèle censuré à variables qualitatives de l'analyse conjointe, il faut introduire les modèles censurés quantitatifs. Selon Maddala (1983), un échantillon censuré apparaît lorsque les valeurs de la variable dépendante quantitative y_i ne sont enregistrées que pour des montants supérieurs à une constante δ, par exemple et valent y_i^*. Par contre, pour des valeurs inférieures à δ, la variable dépendante est posée à δ. Formellement, dans l'ouvrage de Maddala (1983), y_i est spécifiée de la manière suivante :

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > \delta \\ \delta & \text{sinon} \end{cases}$$

De ce modèle, il apparaît que l'information pouvant être fournie lorsque $y_i = \delta$ est la probabilité suivante :

$$P(y_i = \delta) = P(y_i^* \leq \delta)$$

Si y_i^* est supposé être linéaire, tel que $y_i^* = \beta' x_i + u_i$, le modèle devient donc :

$$y_i = \begin{cases} \beta' x_i + u_i & \text{si } \beta' x_i + u_i > \delta \\ \delta & \text{sinon} \end{cases}$$

ou

$$y_i = \max(y_i^*, \delta) = \max(\beta' x_i + u_i, \delta)$$

La probabilité $P(y_i = \delta)$ se modifie, quant à elle, de la manière suivante :

$$P(y_i = \delta) = P(y_i^* \leq \delta) = P(\beta' x_i + u_i \leq \delta) = P(u_i \leq \delta - \beta' x_i) = \int_{-\infty}^{\delta - \beta' x_i} f(u_i) du_i$$

où $f(u_i)$ est la fonction de densité de l'erreur u_i du modèle $y_i^* = \beta' x_i + u_i$.

Par contre, les observations $y_i = y_i^*$ se modélisent à travers la fonction de densité $f(u_i)$ de l'erreur u_i.

L'estimation des paramètres inconnus d'un modèle censuré exige donc la connaissance de la fonction de densité $f(u_i)$. En effet, la fonction de vraisemblance de ces modèles dépend de cette fonction de densité :

$$L = \prod_{y_i^* > \delta} f(u_i) \prod_{y_i^* \leq \delta} \left(\int_{-\infty}^{\delta - \beta' x_i} f(u_i) du_i \right)$$

Il existe un cas particulier des modèles censurés, appelé modèle Tobit simple, qui suppose une valeur de 0 au paramètre δ. Ce modèle Tobit simple s'écrit comme suit :

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

Il est, toutefois, possible d'obtenir cette spécification à partir des modèles censurés sans pour autant poser la valeur de δ à 0. En effet, il suffit d'utiliser la variable $y_i - \delta$ au lieu de y_i. De ce fait, y_i^* devient $y_i^* - \delta = -\delta + \beta' x_i + u_i$ et δ devient $\delta - \delta = 0$. Formellement, le modèle censuré modifié de manière à obtenir un modèle Tobit simple s'écrit :

$$y_i - \delta = \begin{cases} y_i^* - \delta & \text{si } y_i^* - \delta > 0 \\ 0 & \text{sinon} \end{cases} \Longrightarrow z_i = \begin{cases} z_i^* & \text{si } z_i^* > 0 \\ 0 & \text{sinon} \end{cases}$$

Lorsque la variable dépendante est qualitative, les modèles censurés deviennent des modèles dichotomiques, qui sont en réalité la source d'inspiration des modèles censurés. Ces modèles dichotomiques supposent qu'il y a une variable dépendante latente, $y_i^* = \beta' x_i + u_i$, non observable. Cependant, il existe une variable binaire observable y_i qui est définie comme suit :

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

La différence entre ce modèle dichotomique et le modèle censuré est la présence d'une valeur discrète lorsque $y_i^* > 0$. C'est pourquoi, le modèle dichotomique n'utilise que les probabilités liées à $y_i = 1$ et $y_i = 0$, alors que le modèle censuré considère la probabilité liée à $y_i = 0$ et la fonction de densité $f(u_i)$ liée à $y_i = y_i^*$.

3.4.2 Nouvelle spécification de l'option "zéro"

Les modèles de non choix présentés dans la littérature utilisent le principe de maximisation d'utilité pour estimer les utilités partielles. En effet, Elrod, Louviere et Krishnakumar (1992) modélisent le non choix comme une autre alternative dont les attributs sont nuls et ils déterminent le choix entre les produits et l'option "zéro" en comparant leur utilité. Haaijer (1999) considère approximativement le même modèle qu'Elrod, Louviere et Krishnakumar (1992) en modifiant certains aspects. Il propose également d'estimer le non choix par un modèle logit emboîté en deux étapes afin de ne pas supposer l'option "zéro" comme une autre alternative. Notre spécification n'utilise pas le principe de maximisation d'utilité, elle ne considère pas l'option "zéro" comme une autre alternative et elle est formulée en une étape. En réalité, elle s'inspire des modèles censurés à variable dépendante qualitative (modèle Tobit) qui suppose un changement de la variable expliquée à partir d'un certain seuil. Toutefois, une comparaison entre les utilités subsiste. Cette comparaison n'intervient qu'entre les utilités des produits, car l'option "zéro" n'est pas décrite en fonction d'une utilité.

Le non choix décrit dans la littérature ne considère que les situations dû au refus, c'est-à-dire à l'inapréciation des produits. Dans notre nouvelle spécification de l'option "zéro", nous tenons compte du non choix décrit par Tversky et Shafir (1992), et appelé conflit.

Cette section présente un nouvelle façon de définir et de modéliser l'option "zéro". En effet, cette nouvelle spécification utilise le principe de maximisation d'utilités pour comparer les produits et celui des modèles censurés pour modéliser le non choix. De plus, il y intègre le non choix dû au conflit. Il est uniquement décrit dans le cas où il y a deux produits et l'option "zéro" due au refus et au conflit. Quatre alternatives sont ainsi présentées à l'individu, à savoir le produit h, le produit l, le refus et le conflit.

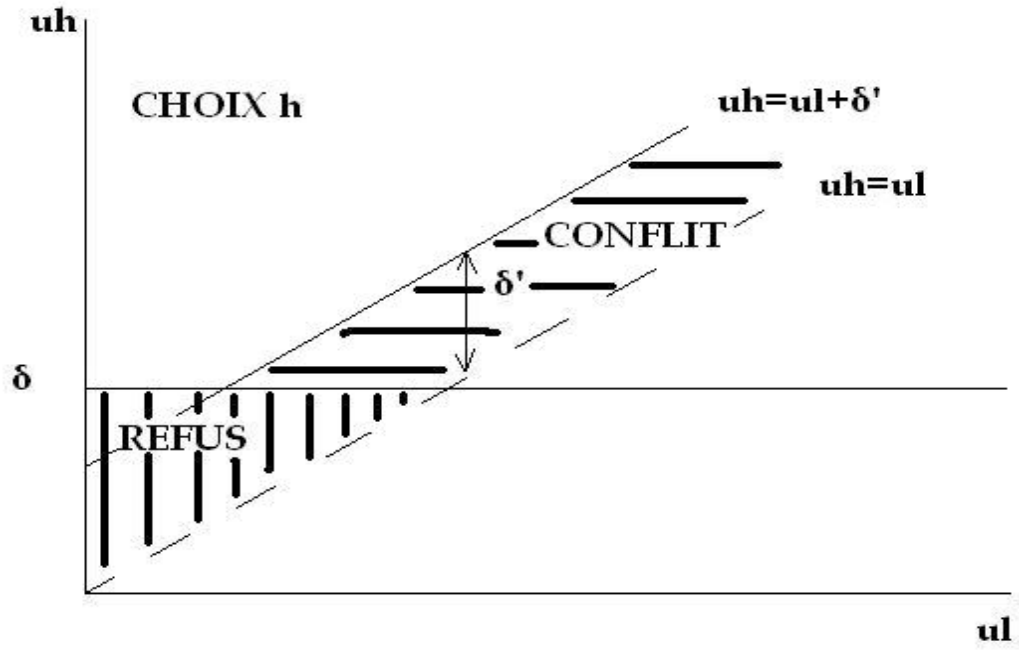
Dans la suite de la thèse, nous présentons une illustration graphique de notre spécification de l'option "zéro", puis nous décrivons les calculs des probabilités associées au choix h, au choix l, au refus et au conflit. Ensuite, nous formulons l'estimation des paramètres du modèle par la méthode de maximum de vraisemblance et finalement, nous abordons l'identification des paramètres structurels.

3.4.3 Représentation graphique du modèle

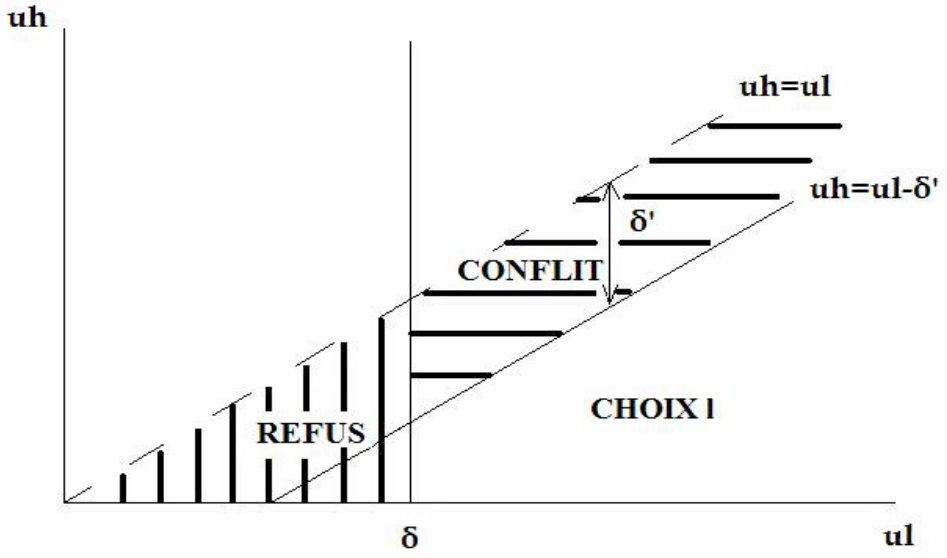
Nous avons supposé que l'individu choisit entre quatre alternatives, comprenant l'option "zéro" due au refus et au conflit. En effet, partant de l'article de Tversky et Shafir (1992), nous savons qu'en plus du refus, il existe également des situations de non choix dû au

conflit. Ce conflit se manifeste lorsqu'il y a une similitude d'intérêt entre deux produits rendant le choix du consommateur difficile. Cette similitude se traduit en économie par des utilités de produits proches. Par conséquent, dans cette situation de conflit, le consommateur choisit l'option "zéro", car ce dernier préfère, par exemple, effectuer une recherche plus approfondie sur les caractéristiques des produits proposés, afin de se décider avec certitude.

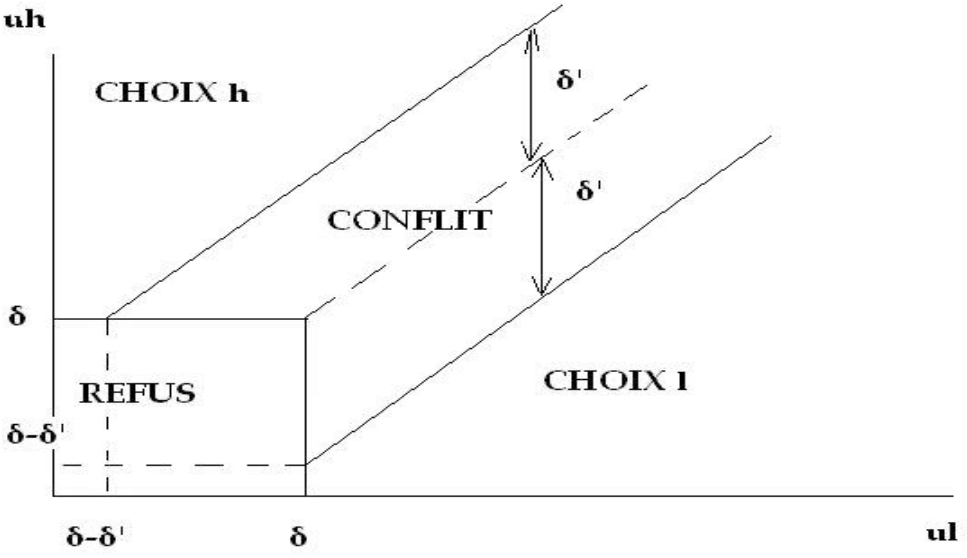
Supposons qu'une différence de δ' entre les utilités des deux biens suffise pour que l'individu ne se trouve pas en situation de conflit. Par conséquent, dans un graphique à deux dimensions où l'axe des abscisses représente l'utilité du bien l et l'axe des ordonnées l'utilité du bien h, l'individu ne choisira le bien h que lorsque son utilité sera supérieure à celle du bien l augmentée de δ'. Il faut également un autre seuil δ représentant le niveau d'utilité suffisant à intéresser le consommateur au produit h. Par conséquent, nous supposons qu'une utilité en dessous de δ se traduit par un refus d'achat du consommateur du bien h.
Graphiquement, ces hypothèses se dessinent de la manière suivante :



Ce graphique ne s'intéresse qu'aux limites de choix du produit h. En réalité, il est possible de tracer par symétrie un graphe représentant les contraintes liées au bien l :



En réunissant ces deux illustrations, nous obtenons le graphique suivant :



Cette illustration du choix et du non choix présente différentes zones reflétant le comportement du consommateur sur le marché. En effet, une des zones représentant le non

intérêt des produits suggère des utilités de produits faibles, c'est-à-dire inférieures au seuil δ. Cette situation est illustrée dans le graphique par le terme de **REFUS**. Les situations de non choix liées à la difficulté de prise de décision sont également pris en considération dans le graphe. Elles sont décrites par une faible différence d'utilités des biens, à savoir une différence en valeur absolue se trouvant en-dessous de δ'. Toutefois, ces choix difficiles n'interviennent pas lorsque les utilités sont faibles ou inférieures à δ (situations de refus). Ils s'associent à la zone **CONFLIT** du graphique.
Le principe de maximisation d'utilité ne tient pas compte de ces conflits engendrés par la difficulté de choix. Toutefois, Tversky et Shafir (1992) montrent les biais liés à la non considération du conflit. C'est pourquoi, notre spécification de l'option "zéro" considère les deux types de non choix, qui sont donc le refus et le conflit. En outre, aucune utilité ne définit ces non choix. En réalité, l'option "zéro" est déterminée à travers les utilités des produits. En effet, les délimitations des zones ne sont qu'en fonction des utilités des biens.
Le graphique illustre également les situations de choix du produit l "**CHOIX** l" et du produit h "**CHOIX** h". Le choix l intervient lorsque la différence d'utilité [16] est inférieure ou égale à $(-\delta')$ et lorsque l'utilité du bien l est supérieure ou égale à δ. Le choix h intervient, quant à lui, lorsque la différence d'utilité est supérieure ou égale à δ' et lorsque l'utilité du bien h est supérieure ou égale à δ.

La modélisation de l'option "zéro" considère donc quatre modalités à expliquer. La variable dépendante qualitative y de notre spécification se formule comme suit :

$$y = \begin{cases} 1 & \text{Choix du bien } h \\ 2 & \text{Choix du bien } l \\ 3 & \text{Refus} \\ 4 & \text{Conflit} \end{cases}$$

ou

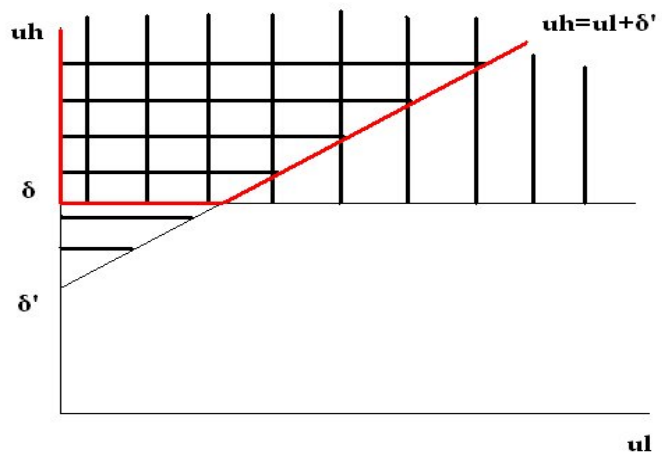
$$y_1 = \begin{cases} 1 & \text{Choix } h \\ 0 & \text{sinon} \end{cases} \quad y_2 = \begin{cases} 1 & \text{Choix } l \\ 0 & \text{sinon} \end{cases} \quad y_3 = \begin{cases} 1 & \text{Refus} \\ 0 & \text{sinon} \end{cases} \quad y_4 = \begin{cases} 1 & \text{Conflit} \\ 0 & \text{sinon} \end{cases}$$

Dans la suite de cette sous-section, nous déterminons les probabilités associées aux modalités de la variable dépendante y.

[16] La différence d'utilité mentionnée dans le texte est la suivante : $\triangle u = u_h - u_l$.

3.4.4 Les probabilités de choix

Selon l'illustration précédente, le choix du deuxième produit suppose une utilité u_h supérieure à δ et une différence d'utilité $(u_h - u_l)$ supérieure à δ' ou $u_h > u_l + \delta'$. Graphiquement, nous cherchons l'intersection entre les deux zones suivantes :



Par conséquent, la probabilité qu'un consommateur choisisse le bien h est modélisée par :

$$P(\text{Choix h}) = P(y = 1) = P(u_h \geq \delta \cap u_h \geq u_l + \delta') = P(u_h \geq \delta, u_h \geq u_l + \delta')$$

La littérature de l'analyse conjointe par le choix exprime l'utilité d'un bien sous une forme linéaire. Dans notre spécification du non choix, nous reprenons cette hypothèse sur les utilités des produits. L'utilité u_h s'écrit donc de la manière suivante : $u_h = \beta' x_h + \varepsilon_h$. Cette linéarité s'explique par l'utilisation des modèles compensatoires dans l'analyse conjointe, à savoir qu'une utilité négative est compensée par une utilité positive. C'est pourquoi, nous postulons une forme additive de l'utilité.
La probabilité $P(y = 1)$ de choisir le produit h, avec l'hypothèse de linéarité, se formule donc comme suit :

$$P(y = 1) = P(u_h \geq \delta, u_h \geq u_l + \delta') = P(\beta' x_h + \varepsilon_h \geq \delta, \beta' x_h + \varepsilon_h \geq \beta' x_l + \varepsilon_l + \delta')$$

$$= P(\varepsilon_h \geq \delta - \beta' x_h, \varepsilon_h - \varepsilon_l \geq \beta' x_l - \beta' x_h + \delta')$$

Cette probabilité se résout à condition de connaître la loi jointe entre ε_h et $(\varepsilon_h - \varepsilon_l)$. Supposons une loi normale bivariée et une corrélation nulle entre les erreurs ε_h $(Cov(\varepsilon_h, \varepsilon_l) = E(\varepsilon_h \varepsilon_l) = 0,\ h \neq l)$:

$$\begin{pmatrix} \varepsilon_h \\ \varepsilon_h - \varepsilon_l \end{pmatrix} = \begin{pmatrix} \varepsilon_h \\ \varepsilon^* \end{pmatrix} \sim N(0; \Sigma_h) \quad \text{où} \quad \Sigma_h = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix}$$

Ce résultat se déduit des distributions des erreurs $\varepsilon_h \sim N(0, \sigma^2)$. En effet, $(\varepsilon_h - \varepsilon_l) = \varepsilon^* \sim N(0, 2\sigma^2)$, car $V(\varepsilon^*) = V(\varepsilon_h) + V(\varepsilon_l) - 2Cov(\varepsilon_h, \varepsilon_l) = \sigma^2 + \sigma^2 - 0$.

La covariance entre ε_h et ε^* se déduit, quant à elle, de la manière suivante :

$$Cov(\varepsilon_h, \varepsilon^*) = E(\varepsilon_h, \varepsilon^*) = E(\varepsilon_h^2) - E(\varepsilon_h \varepsilon_l) = \sigma^2 - 0$$

En divisant les erreurs par leur écart-type respectif, la probabilité de choisir le produit h devient :

$$P(y = 1) = P(\frac{\varepsilon_h}{\sigma} \geq \underbrace{\frac{\delta - \beta' x_h}{\sigma}}_{=\alpha_1}, \frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma} \geq \underbrace{\frac{\beta' x_l - \beta' x_h + \delta'}{\sqrt{2}\sigma}}_{=\alpha_2})$$

$$\text{où} \quad \begin{pmatrix} \frac{\varepsilon_h}{\sigma} \\ \frac{\varepsilon^*}{\sqrt{2}\sigma} \end{pmatrix} \sim N(0; \Sigma_h^*) \quad \text{où} \quad \Sigma_h^* = \begin{pmatrix} 1 & \rho_h \\ \rho_h & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

car $\rho_h = E(\frac{\varepsilon_h}{\sigma} \frac{\varepsilon^*}{\sqrt{2}\sigma}) = \frac{1}{\sqrt{2}\sigma^2}(\sigma^2) = \frac{1}{\sqrt{2}}$.

En réarrangeant la probabilité de choisir le bien h de sorte à l'exprimer en fonction des quantiles de la loi normale bivariée et univariée, nous obtenons le résultat suivant :

$$\begin{aligned} P(y = 1) &= \int_{\alpha_1}^{+\infty} \int_{\alpha_2}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma}\right) \\ &= \int_{\alpha_1}^{+\infty} \int_{\alpha_2}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) \end{aligned}$$

$$= \int_{\alpha_1}^{+\infty} \left[\int_{-\infty}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) - \int_{-\infty}^{\alpha_2} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) \right]$$

$$= \underbrace{\int_{\alpha_1}^{+\infty} \int_{-\infty}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\int_{\alpha_1}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right)} - \int_{\alpha_1}^{+\infty} \int_{-\infty}^{\alpha_2} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)$$

$$= 1 - \underbrace{\int_{-\infty}^{\alpha_1} \varphi\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right)}_{=\Phi(\alpha_1)} - \int_{\alpha_1}^{+\infty} \int_{-\infty}^{\alpha_2} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)$$

$$= 1-\Phi(\alpha_1)-\int_{-\infty}^{\alpha_2} \left[\int_{-\infty}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) - \int_{-\infty}^{\alpha_1} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)\right]$$

$$= 1-\Phi(\alpha_1)-\underbrace{\int_{-\infty}^{\alpha_2} \int_{-\infty}^{+\infty} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\int_{-\infty}^{\alpha_2} \varphi\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)} + \underbrace{\int_{-\infty}^{\alpha_2} \int_{-\infty}^{\alpha_1} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\Phi(\alpha_2, \alpha_1; \rho=\frac{1}{\sqrt{2}})}$$

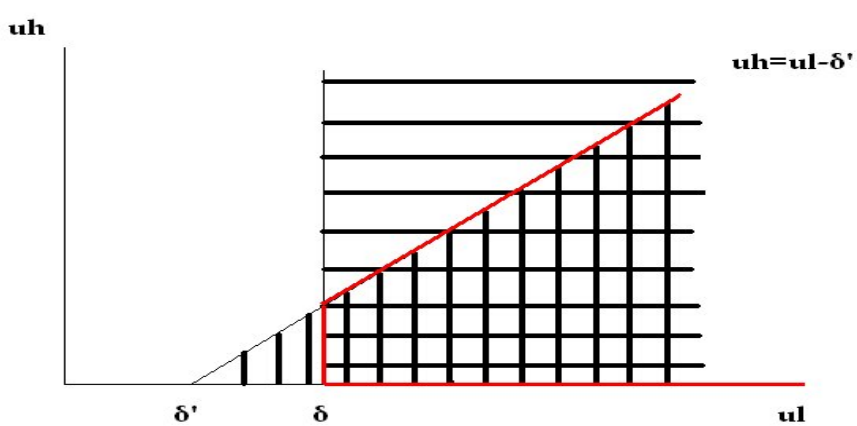
$$= 1 - \Phi(\alpha_1) - \underbrace{\int_{-\infty}^{\alpha_2} \varphi\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\Phi(\alpha_2)} + \Phi(\alpha_2, \alpha_1; \frac{1}{\sqrt{2}})$$

$$P(\text{Choix h}) = P(y = 1) = 1 - \Phi(\alpha_1) - \Phi(\alpha_2) + \Phi(\alpha_2, \alpha_1; \frac{1}{\sqrt{2}})$$

La procédure de détermination de la probabilité de choisir le produit l ($P(y = 2)$) est similaire à celle de choisir le produit h ($P(y = 1)$). En effet, par symétrie, le consommateur choisit le bien l lorsque l'utilité associée à cette alternative est supérieure à un certain seuil δ et lorsque la différence d'utilité des deux produits $(u_l - u_h)$ dépasse la limite de δ'. Formellement, ces conditions s'écrivent :

$$u_l \geq \delta \quad \text{et} \quad u_h \leq u_l - \delta' \Rightarrow u_l - u_h \geq \delta'$$

Graphiquement, ces contraintes se dessinent de la manière suivante :



La probabilité de choisir le produit l s'exprime donc comme suit :

$$P(\text{Choix l}) = P(y=2) = P(u_l \geq \delta \cap u_h \leq u_l - \delta') = P(u_l \geq \delta, u_h \leq u_l - \delta')$$

Etant donné qu'on a supposé la linéarité des utilités ($u_h = \beta' x_h + \varepsilon_h$), la probabilité $P(y=2)$ devient :

$$P(y=2) = P(\beta' x_l + \varepsilon_l \geq \delta, \beta' x_h + \varepsilon_h \leq \beta' x_l + \varepsilon_l - \delta')$$

$$= P(\varepsilon_l \geq \delta - \beta' x_l, \varepsilon_h - \varepsilon_l \leq \beta'(x_l - x_h) - \delta')$$

Il faut connaître la loi jointe entre les erreurs ε_l et $(\varepsilon_h - \varepsilon_l)$. Nous émettons l'hypothèse d'une loi normale bivariée :

$$\begin{pmatrix} \varepsilon_l \\ \varepsilon_h - \varepsilon_l \end{pmatrix} = \begin{pmatrix} \varepsilon_l \\ \varepsilon^* \end{pmatrix} \sim N(0;\Sigma) \quad \text{où} \quad \Sigma_l = \begin{pmatrix} \sigma^2 & -\sigma^2 \\ -\sigma^2 & 2\sigma^2 \end{pmatrix}$$

Ce résultat se déduit, comme précédemment, des distributions des erreurs $\varepsilon_h \sim N(0,\sigma^2)$. En effet, $(\varepsilon_h - \varepsilon_l) = \varepsilon^* \sim N(0, 2\sigma^2)$ et la covariance entre ε_l et ε^* est la suivante :

$$Cov(\varepsilon_l, \varepsilon^*) = E(\varepsilon_l, \varepsilon^*) = E(\varepsilon_l \varepsilon_h) - E(\varepsilon_l^2) = -\sigma^2$$

En divisant les erreurs par leur écart-type respectif, la probabilité que le consommateur choisisse le bien l s'écrit :

$$P(y=2) = P(\frac{\varepsilon_l}{\sigma} \geq \underbrace{\frac{\delta - \beta' x_l}{\sigma}}_{=\alpha_3}, \frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma} \leq \underbrace{\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}}_{=\alpha_4})$$

$$\text{où} \quad \begin{pmatrix} \frac{\varepsilon_l}{\sigma} \\ \frac{\varepsilon^*}{\sqrt{2}\sigma} \end{pmatrix} \sim N(0;\Sigma_l^*) \quad \text{où} \quad \Sigma_l^* = \begin{pmatrix} 1 & \rho_l \\ \rho_l & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

car $\rho_l = E(\frac{\varepsilon_l}{\sigma}\frac{\varepsilon^*}{\sqrt{2}\sigma}) = \frac{1}{\sqrt{2}\sigma^2}(-\sigma^2) = -\frac{1}{\sqrt{2}}$.

La probabilité de choisir le produit l peut être exprimée en fonction des quantiles de la loi normale univariée et bivariée. En effet, cette probabilité se transforme de la manière suivante :

$$P(y=2) = \int_{\alpha_3}^{+\infty} \int_{-\infty}^{\alpha_4} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon_h - \varepsilon_l}{\sqrt{2}\sigma}\right)$$

$$= \int_{\alpha_3}^{+\infty} \int_{-\infty}^{\alpha_4} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)$$

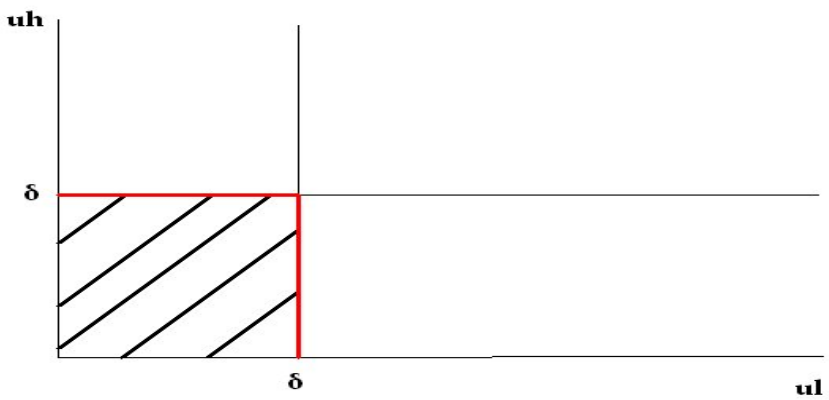
$$= \int_{-\infty}^{\alpha_4} \left[\int_{\infty}^{+\infty} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) - \int_{-\infty}^{\alpha_3} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon^*}{2\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) \right]$$

$$= \underbrace{\int_{-\infty}^{\alpha_4} \int_{\infty}^{+\infty} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\int_{-\infty}^{\alpha_4} \varphi\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)} - \underbrace{\int_{-\infty}^{\alpha_3} \int_{-\infty}^{\alpha_4} \varphi\left(\frac{\varepsilon_l}{\sigma}, \frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\Phi(\alpha_3, \alpha_4; \rho_l = -\frac{1}{\sqrt{2}})}$$

$$P(\text{Choix l}) = P(y = 2) = \underbrace{\int_{-\infty}^{\alpha_4} \varphi\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right) d\left(\frac{\varepsilon^*}{\sqrt{2}\sigma}\right)}_{=\Phi(\alpha_4)} - \Phi(\alpha_3, \alpha_4; -\frac{1}{\sqrt{2}}) = \Phi(\alpha_4) - \Phi(\alpha_3, \alpha_4; -\frac{1}{\sqrt{2}})$$

3.4.5 Les probabilités de non choix

Le non choix dans notre spécification diffère selon qu'il découle d'un non intérêt des biens proposés (refus) ou d'une difficulté de prise de décision due à la similarité des produits (conflit). C'est pourquoi, nous spécifions une probabilité liée au refus et une probabilité liée au conflit. La probabilité du refus découle du graphique suivant :



Le consommateur est dans la situation de refus lorsque les utilités des deux produits sont inférieures à δ. La probabilité associée au refus s'exprime donc de la manière suivante :

$$P(\text{Refus}) = P(y = 3) = P(u_h < \delta, u_l < \delta)$$

Etant donné que précédemment nous avons supposé que les utilités sont linéaires dans les paramètres (les utilités partielles), c'est-à-dire $u_h = \beta' x_h + \varepsilon_h$, la probabilité de ne pas choisir dans la situation de refus s'écrit :

$$P(y=3) = P(\beta' x_h + \varepsilon_h < \delta, \beta' x_l + \varepsilon_l < \delta) = P(\varepsilon_h < \delta - \beta' x_h, \varepsilon_l < \delta - \beta' x_l)$$

Cette probabilité suppose la connaissance de la loi jointe des erreurs ε_h et ε_l, $h \neq l$. Dans notre modèle, nous supposons que les erreurs suivent une loi normale bivariée :

$$\begin{pmatrix} \varepsilon_h \\ \varepsilon_l \end{pmatrix} \sim N(0;\Sigma) \quad \text{où} \quad \Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

En divisant les erreurs par leur écart-type respectif, la probabilité de refus s'exprime de la manière suivante :

$$P(y=3) = P(\frac{\varepsilon_h}{\sigma} < \underbrace{\frac{\delta - \beta' x_h}{\sigma}}_{=\alpha_1}, \frac{\varepsilon_l}{\sigma} < \underbrace{\frac{\delta - \beta' x_l}{\sigma}}_{=\alpha_3})$$

La loi normale bivariée des erreurs réduites s'écrit, quant à elle, comme suit :

$$\begin{pmatrix} \frac{\varepsilon_l}{\sigma} \\ \frac{\varepsilon_h}{\sigma} \end{pmatrix} \sim N(0;\Sigma^*) \quad \text{où} \quad \Sigma^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

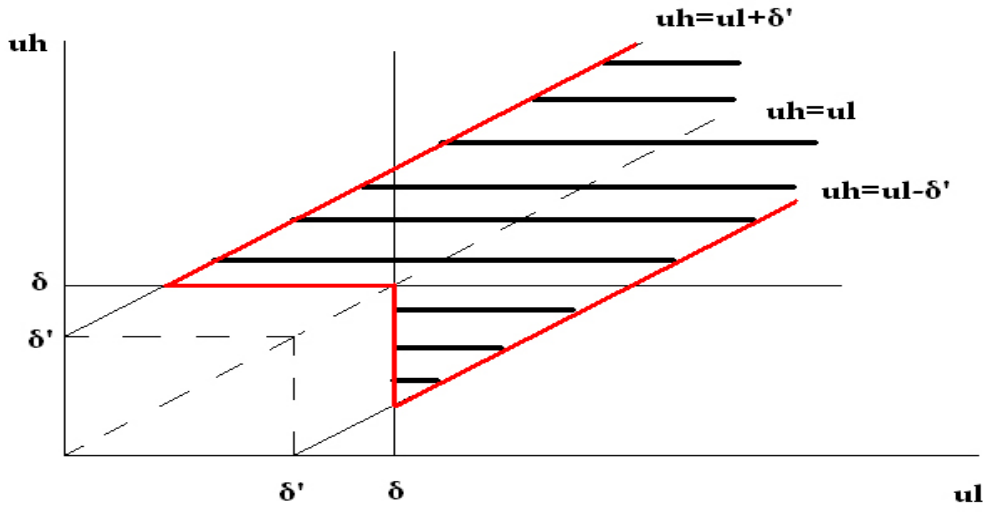
La probabilité de ne pas choisir dans une situation de refus peut se formuler en fonction du quantile d'une loi normale bivariée :

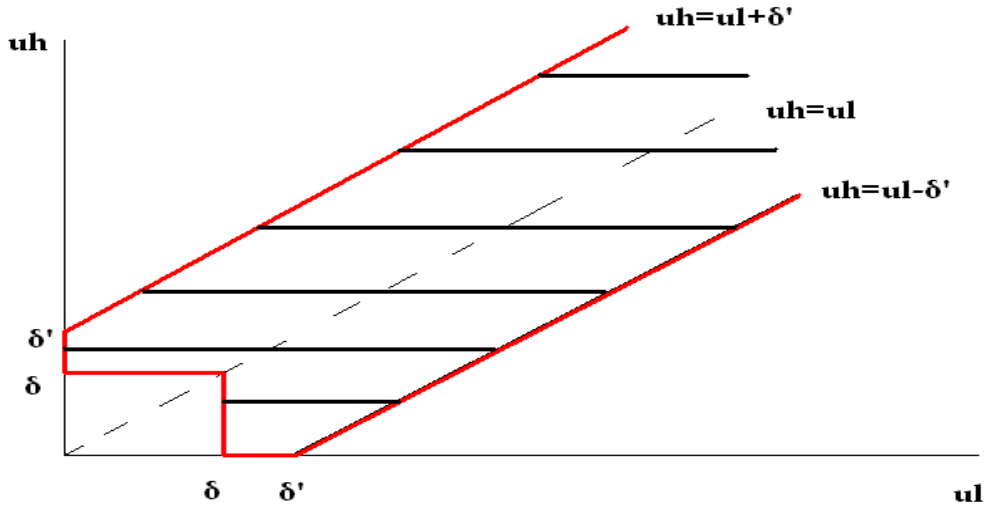
$$P(y=3) = \int_{-\infty}^{\alpha_1} \int_{-\infty}^{\alpha_3} \varphi\left(\frac{\varepsilon_h}{\sigma}, \frac{\varepsilon_l}{\sigma}\right) d\left(\frac{\varepsilon_h}{\sigma}\right) d\left(\frac{\varepsilon_l}{\sigma}\right) = \Phi(\alpha_1, \alpha_3; \rho = 0) = \Phi(\alpha_1)\Phi(\alpha_3)$$

Il y a également le non choix lié au conflit. En effet, selon l'article de Tversky et Shafir (1992), l'individu préfère ne pas acheter de produit lorsque le choix est difficile. Il se trouve donc dans une situation dite de conflit. Les auteurs expliquent ce comportement en se basant sur la théorie psychologique de la préférence des conséquences liées à l'inaction qu'à l'action. Dans les modèles de non choix basés sur la maximisation d'utilité, le conflit n'influence pas l'option "zéro". Cependant, Tversky et Shafir (1992) démontrent dans leur article que le non choix est également lié au conflit et que l'application de la maximisation d'utilité amène à des conclusions inconsistantes avec les résultats de leur application. Ce biais est dû au fait que les modèles de maximisation d'utilité ne tiennent pas compte du conflit et qu'ils considèrent l'option "zéro" comme une autre alternative à laquelle est associée une utilité. Dans notre spécification, il n'existe que des utilités liées

aux produits. Le non choix n'intervient donc que pour certaines valeurs de ces utilités.

Le non choix dans une situation de conflit se déduit des graphiques suivants :





Dans une des représentations, le paramètre δ est supérieur à celui de δ', alors que dans l'autre, c'est l'inverse.

Etant donné que la détermination d'une seule probabilité de conflit à partir de ces illustrations est complexe, nous utilisons donc la propriété de la somme des probabilités égale à 1 pour exprimer la situation de conflit. Cette probabilité de non choix dû à un conflit s'écrit donc de la manière suivante :

$$P(\text{Conflit}) = 1-\underbrace{[1-\Phi(\alpha_1)-\Phi(\alpha_2)+\Phi(\alpha_2,\alpha_1;\frac{1}{\sqrt{2}})]}_{=P(\text{Choix h}))}-\underbrace{[\Phi(\alpha_4)-\Phi(\alpha_3,\alpha_4;-\frac{1}{\sqrt{2}})]}_{=P(\text{Choix l})}-\underbrace{\Phi(\alpha_1)\Phi(\alpha_3)}_{=P(\text{Refus})}$$

$$= \Phi(\alpha_1)+\Phi(\alpha_2)-\Phi(\alpha_2,\alpha_1;\frac{1}{\sqrt{2}})-\Phi(\alpha_4)+\Phi(\alpha_3,\alpha_4;-\frac{1}{\sqrt{2}})-\Phi(\alpha_1)\Phi(\alpha_3)$$

$$P(y=4) = \Phi(\alpha_1)+\Phi(\alpha_2)-\Phi(\alpha_4)-\Phi(\alpha_2,\alpha_1;\frac{1}{\sqrt{2}})-\Phi(\alpha_1)\Phi(\alpha_3)+\Phi(\alpha_3,\alpha_4;-\frac{1}{\sqrt{2}})$$

En résumé, les probabilités de choisir le produit l, le produit h, de ne pas choisir de produit dans une situation de refus et dans une situation de conflit sont les suivantes :

- $P(\text{Choix h}) = P(y=1) = 1-\Phi(\alpha_1)-\Phi(\alpha_2)+\Phi(\alpha_2,\alpha_1;\frac{1}{\sqrt{2}})$
- $P(\text{Choix l}) = P(y=2) = \Phi(\alpha_4)-\Phi(\alpha_3,\alpha_4;-\frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y=3) = \Phi(\alpha_1)\Phi(\alpha_3)$
- $P(\text{Conflit}) = P(y=4) = \begin{cases} \Phi(\alpha_1)+\Phi(\alpha_2)-\Phi(\alpha_4)-\Phi(\alpha_2,\alpha_1;\frac{1}{\sqrt{2}}) \\ -\Phi(\alpha_1)\Phi(\alpha_3)+\Phi(\alpha_3,\alpha_4;-\frac{1}{\sqrt{2}}) \end{cases}$

En remplaçant les α_i par leurs expressions, les probabilités du modèle de l'option "zéro" deviennent :

- $P(\text{Choix h}) = P(y=1) = P_1 = 1-\Phi(\frac{\delta-\beta' x_h}{\sigma})-\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma})+\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma},\frac{\delta-\beta' x_h}{\sigma};\frac{1}{\sqrt{2}})$
- $P(\text{Choix l}) = P(y=2) = P_2 = \Phi(\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma})-\Phi(\frac{\delta-\beta' x_l}{\sigma},\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma};-\frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y=3) = P_3 = \Phi(\frac{\delta-\beta' x_h}{\sigma})\Phi(\frac{\delta-\beta' x_l}{\sigma})$
- $P(\text{Conflit}) = P(y=4) = P_4 = \begin{cases} \Phi(\frac{\delta-\beta' x_h}{\sigma})+\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma})-\Phi(\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}) \\ -\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma},\frac{\delta-\beta' x_h}{\sigma};\frac{1}{\sqrt{2}})-\Phi(\frac{\delta-\beta' x_h}{\sigma})\Phi(\frac{\delta-\beta' x_l}{\sigma}) \\ +\Phi(\frac{\delta-\beta' x_l}{\sigma},\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma};-\frac{1}{\sqrt{2}}) \end{cases}$

En posant $\gamma_1 = \frac{\delta}{\sigma}$, $\gamma_2 = \frac{\beta}{\sigma}$ et $\gamma_3 = \frac{\delta'}{\sqrt{2}\sigma}$, les probabilités de choix et de non choix deviennent :

- $P(\text{Choix h}) = P(y=1) = P_1 = \begin{cases} 1-\Phi(\gamma_1-\gamma_2' x_h)-\Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l-x_h)+\gamma_3) \\ +\Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l-x_h)+\gamma_3,\gamma_1-\gamma_2' x_h;\frac{1}{\sqrt{2}}) \end{cases}$

- $P(\text{Choix l}) = P(y=2) = P_2 = \begin{cases} \Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) - \gamma_3) \\ -\Phi(\gamma_1 - \gamma_2' x_l, \frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) - \gamma_3; \frac{1}{\sqrt{2}}) \end{cases}$
- $P(\text{Refus}) = P(y=3) = P_3 = \Phi(\gamma_1 - \gamma_2' x_h)\Phi(\gamma_1 - \gamma_2' x_l)$
- $P(\text{Conflit}) = P(y=4) = P_4 = \begin{cases} \Phi(\gamma_1 - \gamma_2' x_h) \\ +\Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) + \gamma_3) - \Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) - \gamma_3) \\ -\Phi(\frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) + \gamma_3, \gamma_1 - \gamma_2' x_h; \frac{1}{\sqrt{2}}) \\ -\Phi(\gamma_1 - \gamma_2' x_h)\Phi(\gamma_1 - \gamma_2' x_l) \\ +\Phi(\gamma_1 - \gamma_2' x_l, \frac{\gamma_2'}{\sqrt{2}}(x_l - x_h) - \gamma_3; -\frac{1}{\sqrt{2}}) \end{cases}$

3.4.6 La fonction de vraisemblance

Pour estimer les paramètres du modèle, à savoir les seuils δ et δ', les utilités partielles β, la variance des erreurs σ, il suffit de maximiser le logarithme de la fonction de vraisemblance. La fonction de vraisemblance du modèle de l'option "zéro" pour un individu choisissant à s reprises entre 4 alternatives, à savoir le choix h, le choix l, le refus et le conflit, est la suivante :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 2)^{y_{2s}} P(y_s = 3)^{y_{3s}} P(y_s = 4)^{y_{4s}}$$

où $y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$

où $y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$

où $y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$

où $y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$

La maximisation du logarithme de L s'écrit donc comme suit :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{2s} \log P_{2s} + y_{3s} \log P_{3s} + y_{4s} \log P_{4s} \right)$$

Les équations de résolution de cette maximisation découlent des conditions de premier ordre suivantes :

$$\frac{\partial \log L}{\partial \gamma_1} = 0 \quad ; \quad \frac{\partial \log L}{\partial \gamma_2} = 0 \quad ; \quad \frac{\partial \log L}{\partial \gamma_3} = 0$$

Ces équations fournissent l'estimation des paramètres réduits. Il faut vérifier que les paramètres structurels sont identifiables.

3.4.7 Identification des paramètres structurels

L'estimation par maximum de vraisemblance fournit donc l'estimation des paramètres réduits $\hat{\gamma}_1$, $\hat{\gamma}_2$ et $\hat{\gamma}_3$. Il faut maintenant vérifier si l'estimation des paramètres structurels, (δ, σ, δ', β) peut être déduite de celle des paramètres réduits, c'est-à-dire si les paramètres sont sous-identifiés, juste-identifiés ou sur-identifiés. En réalité, nous constatons que les paramètres structurels sont sous-identifiés. En effet :

$$\hat{\gamma}_1 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\gamma}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\hat{\gamma}_3 = \frac{\delta'}{\sqrt{2}\sigma} \implies \text{inconnues} \quad (\delta', \sqrt{2}\sigma)$$

$$\implies \text{3 équations et 4 inconnues : sous-identification}$$

Etant donné que les paramètres sont sous-identifiés, il faut donc en normaliser un. L'estimation des utilités partielles n'étant pas à un paramètre d'échelle près, nous pouvons donc fixer σ à 1. Avec cette normalisation, nous définissons les utilités partielles à un paramètre d'échelle près. Les paramètres structurels s'expriment donc comme suit :

$$\hat{\delta} = \hat{\gamma}_1$$

$$\hat{\beta} = \hat{\gamma}_2$$

$$\hat{\delta}' = \sqrt{2}\hat{\gamma}_3$$

3.4.8 Modèle agrégé de l'option "zéro"

La spécification décrite dans les sous-sections précédentes est un modèle individuel. En effet, elle estime les utilités partielles d'un consommateur. Supposons que plusieurs individus choisissent parmi les paires de produits et le non choix. Le modèle devient agrégé et amène à de nouvelles hypothèses relatives uniquement aux seuils. En effet, afin de simplifier le modèle, nous conservons l'hypothèse d'indépendance des erreurs d'un produit à l'autre et nous postulons également l'indépendance des erreurs d'un individu à l'autre. Ces paramètres seront modifiés par la suite.

L'utilité associée au modèle agrégé et auquel nous gardons l'hypothèse de linéarité s'écrit :

$$u_{hi} = \beta' x_h + \varepsilon_{hi}, \quad \text{où} \quad \varepsilon_{hi} \sim N(0, \sigma^2)$$

Les probabilités associées aux modalités (choix h, choix l, refus et conflit) de la variable dépendante se modifient en y ajoutant l'indice i aux seuils δ et δ'.

La fonction de vraisemblance s'écrit, quant à elle, de la manière suivante :

$$L = \prod_{i=1}^{N} \prod_{s=1}^{S} P(y_{si} = 1)^{y_{1si}} P(y_{si} = 2)^{y_{2si}} P(y_{si} = 3)^{y_{3si}} P(y_{si} = 4)^{y_{4si}}$$

Les paramètres structurels restent sous-identifiés. En effet :

$$\hat{\gamma}_{1i} = \frac{\delta_i}{\sigma} \implies \text{inconnues} \quad (\delta_i, \sigma)$$

$$\hat{\gamma}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\hat{\gamma}_{3i} = \frac{\delta'_i}{\sqrt{2}\sigma} \implies \text{inconnues} \quad (\delta'_i, \sqrt{2}\sigma)$$

Il faut à nouveau poser σ à 1 et l'expression des variables structurelles devient :

$$\hat{\delta}_i = \hat{\gamma}_{1i}$$

$$\hat{\beta} = \hat{\gamma}_2$$

$$\hat{\delta}_i{}' = \sqrt{2}\hat{\gamma}_{3i}$$

Afin de ne pas avoir un nombre de paramètres inconnus trop élevé et de ce fait, une non convergence de l'estimation par le maximum de vraisemblance du modèle agrégé, nous supposons, dans notre application illustrée dans le chapitre 5, un seuil commun à chaque individu. Avec cette hypothèse, nous pouvons enlever l'indice i des coefficients δ et δ' du modèle agrégé.

3.5 Cas spéciaux

Il existe des situations pour lesquelles l'individu ne choisit pas toutes les options proposées. Dans cette section, nous allons développer les modèles associés à ces cas que l'on nommera "cas spéciaux". Il y a les situations où l'individu ne choisit que trois options parmi les quatre, celle où il ne choisit que deux options parmi les quatre et celle où il n'en choisit qu'une parmi les quatre. La dernière situation ne sera pas considérée et développée,

car elle signifie que l'individu n'a pas répondu sérieusement au questionnaire, étant donné qu'il ne choisit qu'une seule réponse. Il est donc préférable de l'enlever de la base de données, afin d'éviter les biais dus aux réponses incorrectes.

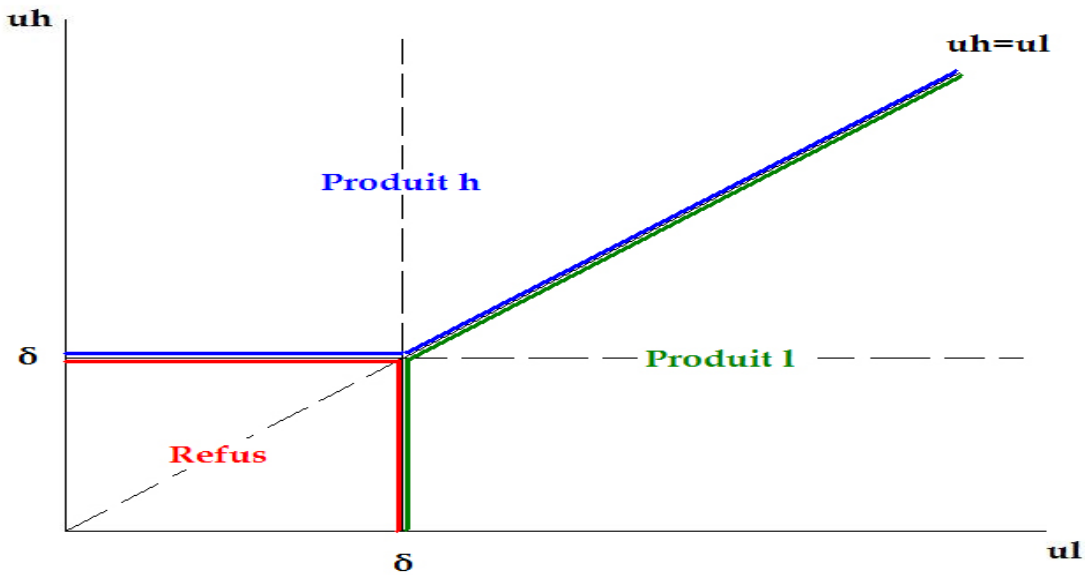
Une des situations, où l'individu ne choisit que trois options parmi les quatre, est celle où ce dernier n'est pas dans l'incertitude lors de ces choix et ne sélectionne donc pas l'option associée au conflit. Dans ce cas, la variable dépendante ne prend plus que les valeurs suivantes : $y = 1, 2, 3$. Une autre situation est celle où le consommateur apprécie tous les produits, mais il a encore certaines incertitudes. Dans ce cas, l'option refus n'est pas sélectionnée et la variable y ne prend que les valeurs $y = 1, 2, 4$. Il se peut également qu'un des produits ne soit pas clairement représentatif des préférences du consommateur. Dans cette situation, l'individu choisit les options refus, conflit et un des produits proposés. Les valeurs de la variable y sont alors, soit $y = 1, 3, 4$, soit $y = 2, 3, 4$. Toutefois, ces situations semblent peu vraisemblables, mais elles seront tout de même développées.

Une des situations pour lesquelles l'individu ne choisit que deux options parmi les quatre est celle où ce dernier n'a pas d'incertitude dans ces choix et il apprécie tous les produits. Dans ce type de cas, l'individu ne sélectionne donc pas les options associées au conflit et au refus. Les valeurs de y sont alors $y = 1, 2$. Cette situation est la plus courante parmi celle où le consommateur ne choisirait que deux options parmi les quatre. Cependant, il existe également d'autres cas, à savoir ceux où l'individu n'apprécie aucun produit et il est incertain ($y = 3, 4$), ceux où, soit un seul produit, soit aucun produit est choisi ($y = 1, 3$, respectivement $y = 2, 3$) et ceux où, soit un seul produit est choisi, soit l'individu est incertain ($y = 1, 4$, respectivement $y = 2, 4$).

La suite de cette section décrit en détails ces différentes situations énumérées précédemment. Les hypothèses supposées dans le cas où $y = 1, 2, 3, 4$ restent valides pour les modèles de cette section "cas spéciaux", à savoir la normalité des erreurs (ε_h, ε_l et ε^*) isolées et jointes, la nullité de la covariance entre les erreurs ε_h et ε_l et la forme linéaire des utilités u_h et u_l. De plus, pour chaque modèle de cette section, nous procédons de la même manière que pour le cas où $y = 1, 2, 3, 4$ dans le développement des probabilités associées à chaque type de réponses sélectionnées par l'interviewé, c'est-à-dire en décrivant ces dernières en fonction des quantiles d'une loi normale. Par conséquent, dans la suite du travail, nous ne développons pas en détail ces probabilités, étant donné que la procédure d'obtention de ces dernières s'assimile à celle du modèle où $y = 1, 2, 3, 4$.

3.5.1 Modèle où y=1,2,3

Dans cette sous-section, l'individu ne se trouve pas dans une situation de conflit. Graphiquement, cette situation s'illustre comme suit :



Selon le graphe, les probabilités associées à chaque cas de figure diffèrent de celles développées dans la section précédente.

Les probabilités de choisir le produit h, l et le refus se formulent de la manière suivante :

- $P(y=h) = P(y=1) = P(u_h > u_l, u_h > \delta) = P(\underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} > \beta'(x_l - x_h), \varepsilon_h > \delta - \beta' x_h)$
- $P(y=l) = P(y=2) = P(u_h < u_l, u_l > \delta) = P(\underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} < \beta'(x_l - x_h), \varepsilon_l > \delta - \beta' x_l)$
- $P(\text{Refus}) = P(y=3) = P(u_h < \delta, u_l < \delta) = P(\varepsilon_h < \delta - \beta' x_h, \varepsilon_l < \delta - \beta' x_l)$

La forme finale de ces dernières s'obtient en procédant de la même manière que pour le modèle où $y = 1, 2, 3, 4$, c'est-à-dire en transformant les intégrales simples et doubles de ces probabilités de sorte qu'elles puissent être ensuite exprimées en fonction des quantiles d'une loi normale univariée ou bivariée. La forme finale de ces probabilités est la suivante :

- $P(y=1) = 1 - \Phi(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}) - \Phi(\frac{\delta - \beta' x_h}{\sigma}) + \Phi(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})$
- $P(y=2) = \Phi(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_l}{\sigma}; -\frac{1}{\sqrt{2}})$
- $P(y=3) = \Phi(\frac{\delta - \beta' x_h}{\sigma})\Phi(\frac{\delta - \beta' x_l}{\sigma})$

Ces résultats auraient également pu être déduites des probabilités du modèle où $y = 1, 2, 3, 4$. En effet, il aurait suffi de poser, dans les probabilités de choisir le produit h, l et le refus, le seuil associé au conflit δ' à 0 :

- $P(\text{Choix h}) = P(y = 1) = P_1 = 1 - \Phi(\frac{\beta'(x_l - x_h) + 0}{\sqrt{2}\sigma}) - \Phi(\frac{\delta - \beta' x_h}{\sigma}) + \Phi(\frac{\beta'(x_l - x_h) + 0}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})$
- $P(\text{Choix l}) = P(y = 2) = P_2 = \Phi(\frac{\beta'(x_l - x_h) - 0}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l - x_h) - 0}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_l}{\sigma}; -\frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y = 3) = P_3 = \Phi(\frac{\delta - \beta' x_h}{\sigma})\Phi(\frac{\delta - \beta' x_l}{\sigma})$

La fonction de vraisemblance d'un individu du modèle où y ne prend que les valeurs 1,2 et 3, s'écrit, quant à elle, de la manière suivante :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 2)^{y_{2s}} P(y_s = 3)^{y_{3s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

Il suffit de maximiser le logarithme de cette fonction de vraisemblance L pour obtenir les paramètres inconnus et réduits du modèle :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{2s} \log P_{2s} + y_{3s} \log P_{3s} \right)$$

Il faut vérifier l'identifiabilité des paramètres pour savoir si l'estimation des paramètres réduits donne directement celle des paramètres structurels.

Les équations associant les paramètres réduits aux paramètres structurels sont les suivantes :

$$\hat{\eta}_1 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

Etant donné que les paramètres sont sous-identifiés, il faut donc normaliser un des paramètres. Posons $\sigma = 1$. Les paramètres structurels s'expriment donc comme suit :

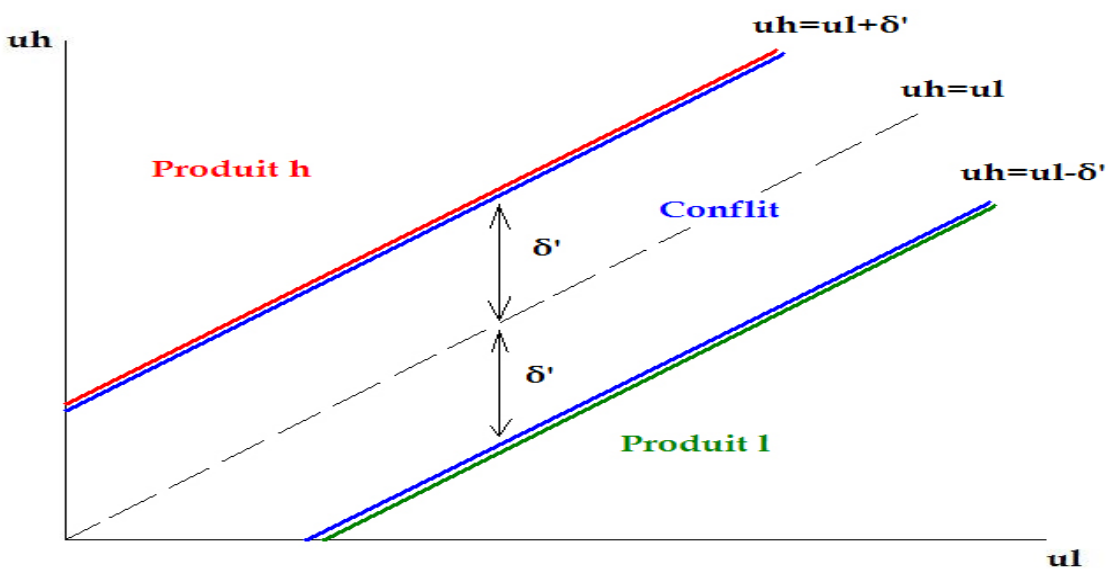
$$\hat{\delta} = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

La maximisation du logarithme de la fonction de vraisemblance donne donc directement l'estimation des paramètres structurels.

3.5.2 Modèle où y=1,2,4

Le modèle où $y = 1, 2, 4$ suppose que l'individu ne choisit pas l'option "refus". Graphiquement, nous illustrons ce modèle de la manière suivante :



A partir de cette représentation graphique, nous pouvons définir les probabilités associées à chaque situation. La probabilité de choisir le produit h se formule comme suit :

$$P(y = 1) = P(y = h) = P(u_h > u_l + \delta') = P(\underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} > \beta'(x_l - x_h) + \delta')$$

En procédant de la même manière que pour le modèle où $y = 1, 2, 3, 4$, nous obtenons finalement l'expression suivante de la probabilité de choisir le produit h :

$$P(y = h) = P(y = 1) = P\left(\frac{\varepsilon^*}{\sqrt{2}\sigma} > \frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}\right) = 1 - \Phi\left(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}\right)$$

Par symétrie, nous pouvons également exprimer la probabilité de choisir le produit l comme suit :

$$P(y=l) = P(y=2) = \Phi\left(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}\right)$$

Le consommateur est dans une situation de conflit lorsque la différence des utilités est comprise entre $-\delta'$ et δ'. La probabilité de conflit se formalise donc de la manière suivante :

$$P(\text{Conflit}) = P(y=4) = P(u_l - \delta' < u_h < u_l + \delta')$$

$$= P(\beta' x_l + \varepsilon_l - \delta' < \beta' x_h + \varepsilon_h < \beta' x_l + \varepsilon_l + \delta')$$

$$= P(\beta'(x_l - x_h) - \delta' < \underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} < \beta'(x_l - x_h) + \delta')$$

Pour obtenir l'expression finale de la probabilité de conflit, nous appliquons la même procédure de résolution que dans le modèle où $y = 1, 2, 3, 4$, à savoir que nous transformons cette probabilité de sorte à obtenir des quantiles d'une loi normale univariée. Cette dernière se formule comme suit :

$$P(y=4) = \Phi\left(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}\right) - \Phi\left(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}\right)$$

En résumé, les probabilités associées au modèle où $y = 1, 2, 4$ s'écrivent :

- $P(\text{Produit } h) = P(y=1) = 1 - \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma})$
- $P(\text{Produit } l) = P(y=2) = \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma})$
- $P(\text{Conflit}) = P(y=4) = \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma})$

Elles aurait également pu être déduites des probabilités du modèle où $y = 1, 2, 3, 4$. En effet, en posant le seuil associé au refus δ à $-\infty$, nous obtenons les mêmes expressions que ci-dessus :

- $P(\text{Choix h}) = P(y=1) = P_1 = 1 - \underbrace{\Phi(\frac{-\infty - \beta' x_h}{\sigma})}_{=0} - \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}) + \underbrace{\Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}, \frac{-\infty - \beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})}_{=0}$

- $P(\text{Choix l}) = P(y=2) = P_2 = \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}) - \underbrace{\Phi(\frac{-\infty - \beta' x_l}{\sigma}, \frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}; -\frac{1}{\sqrt{2}})}_{=0}$

- $P(\text{Conflit}) = P(y=4) = P_4 = \begin{cases} \underbrace{\Phi(\frac{-\infty-\beta' x_h}{\sigma})}_{=0} - \underbrace{\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}, \frac{-\infty-\beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})}_{=0} \\ +\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}) - \underbrace{\Phi(\frac{-\infty-\beta' x_h}{\sigma})\Phi(\frac{-\infty-\beta' x_l}{\sigma})}_{=0} \\ +\underbrace{\Phi(\frac{-\infty-\beta' x_l}{\sigma}, \frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}; -\frac{1}{\sqrt{2}})}_{=0} \end{cases}$

La fonction de vraisemblance de ce modèle est la suivante :

$$L = \prod_{s=1}^{S} P(y_s=1)^{y_{1s}} P(y_s=2)^{y_{2s}} P(y_s=4)^{y_{4s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

La maximisation du logarithme de cette fonction de vraisemblance L permet d'obtenir les paramètres inconnus et réduits du modèle :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{2s} \log P_{2s} + y_{4s} \log P_{4s} \right)$$

L'identification des paramètres n'étant pas directement assurée,

$$\hat{\eta}_1 = \frac{\delta'}{\sigma} \implies \text{inconnues} \quad (\delta', \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

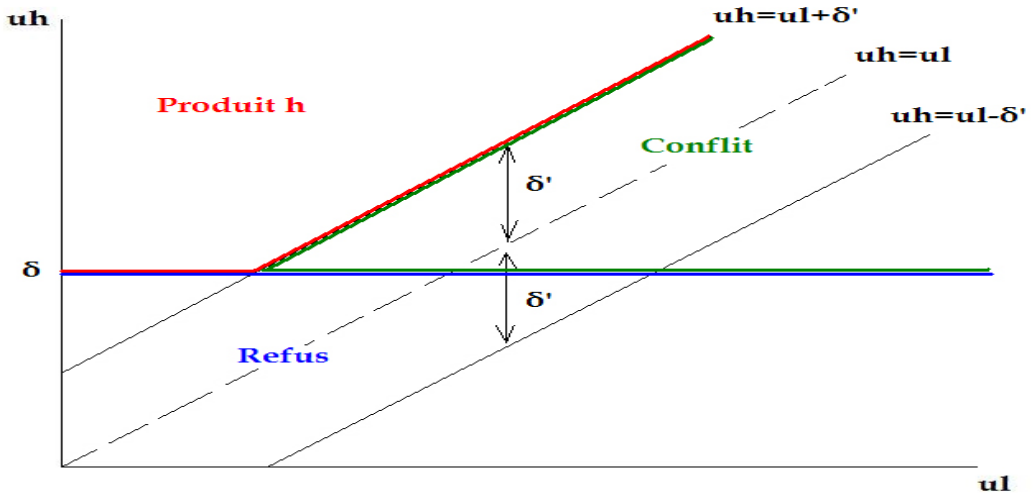
il faut donc en normaliser un, en posant $\sigma = 1$. Avec cette normalisation, l'estimation directe des paramètres structurels est la suivante :

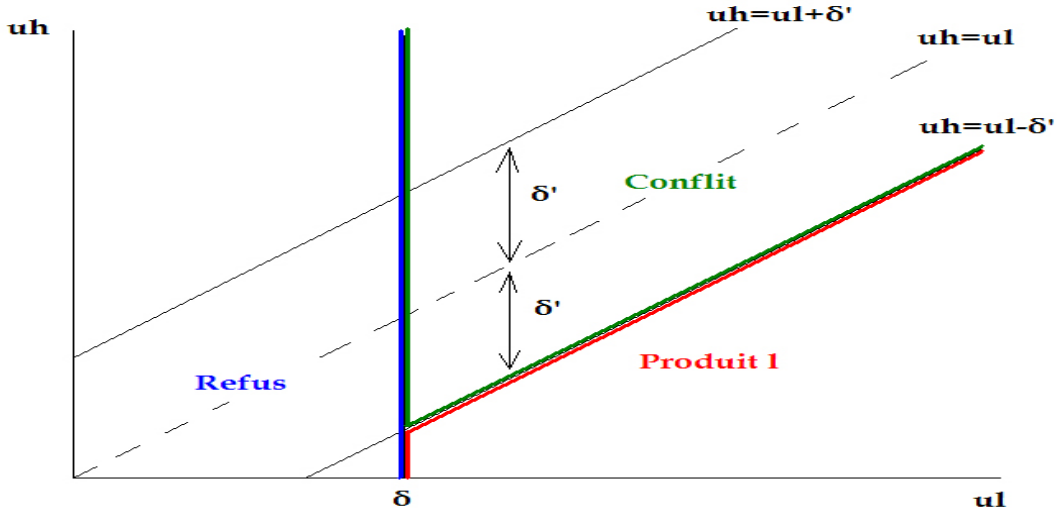
$$\hat{\delta}' = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

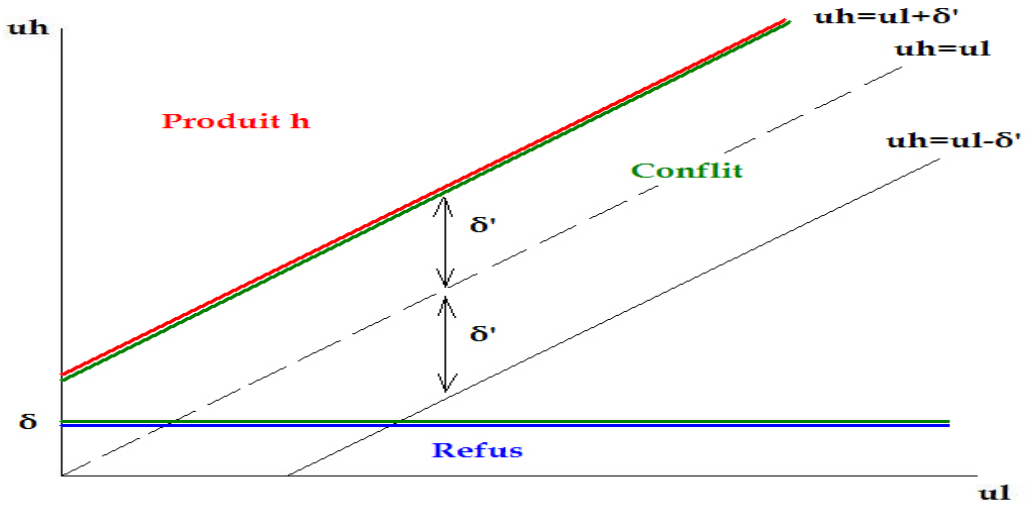
3.5.3 Modèle où y=1,3,4, respectivement y=2,3,4

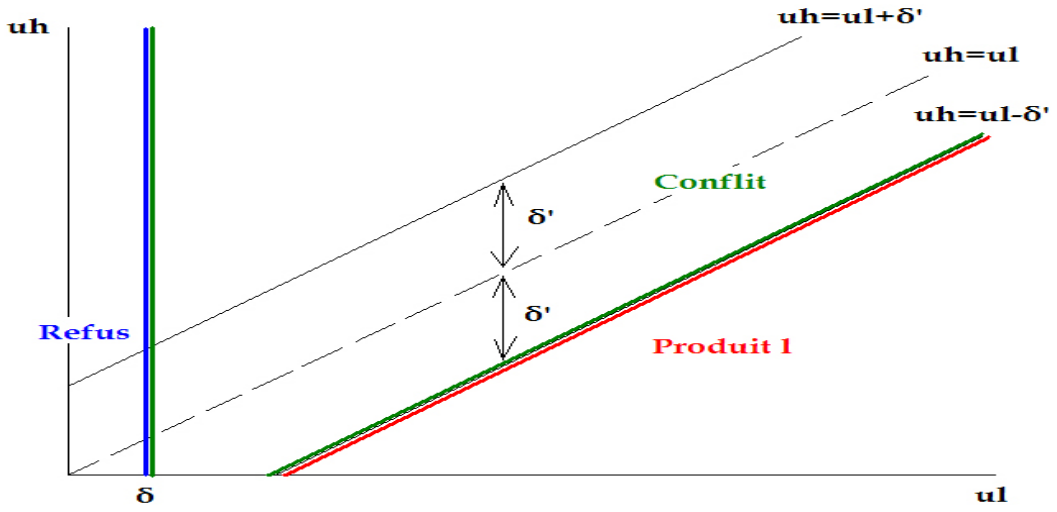
Dans cette sous-section, nous supposons que l'individu ne choisit pas le produit l, respectivement le produit h. Cette situation s'illustre de deux manières selon les valeurs des paramètres δ et δ'. En effet, lorsque $\delta > \delta'$, les graphiques du modèle $y = 1, 3, 4$, respectivement $y = 2, 3, 4$, se dessinent de la manière suivante :





alors que, lorsque $\delta < \delta'$, ils sont représentés de la manière suivante :





Cependant, quelque soit les valeurs de δ et δ', l'expression des probabilités est identique, car nous supposons une étendue des utilités u_h et u_l variant de $-\infty$ à ∞.

Dans cette sous-section, nous ne développons que les probabilités du modèle $y = 1, 3, 4$, étant donné que celles du modèle $y = 2, 3, 4$ sont obtenues en procédant à l'identique. Toutefois, une formulation de ces dernières y est, tout de même, présentée.

Selon les représentations associées au modèle $y = 1, 3, 4$, les probabilités de choisir le produit h, le refus et le conflit se formulent de la manière suivante :

- $P(y = 1) = P(y = h) = P(u_h > u_l + \delta', u_h > \delta) = P(\underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} > \beta'(x_l - x_h) + \delta', \varepsilon_h > \delta - \beta' x_h)$
- $P(\text{Refus}) = P(y = 3) = P(u_h < \delta) = P(\varepsilon_h < \delta - \beta' x_h)$
- $P(y = \text{conflit}) = P(y = 4) = P(u_h < u_l + \delta', u_h > \delta) = P(\underbrace{\varepsilon_h - \varepsilon_l}_{\varepsilon^*} < \beta'(x_l - x_h) + \delta', \varepsilon_h > \delta - \beta' x_h)$

En développant ces dernières, nous obtenons les expressions suivantes du modèle où $y = 1, 3, 4$:

- $P(\text{Produit } h) = P(y = 1) = 1 - \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\delta - \beta' x_h}{\sigma}) + \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y = 3) = \Phi(\frac{\delta - \beta' x_h}{\sigma})$
- $P(\text{Conflit}) = P(y = 4) = \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})$

La fonction de vraisemblance d'un individu de ce modèle s'écrit de la manière suivante :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 3)^{y_{3s}} P(y_s = 4)^{y_{4s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

En maximisant le logarithme de cette fonction de vraisemblance L, nous obtenons les paramètres inconnus et réduits du modèle :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{3s} \log P_{3s} + y_{4s} \log P_{4s} \right)$$

auxquels il faut vérifier l'identifiabilité :

$$\hat{\eta}_1 = \frac{\delta'}{\sigma} \implies \text{inconnues} \quad (\delta', \sigma)$$

$$\hat{\eta}_2 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_3 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{3 équations et 4 inconnues : sous-identification}$$

Les paramètres étant sous-identifiés, il faut donc en normaliser un. En posons $\sigma = 1$, l'estimation directe des paramètres structurels est :

$$\hat{\delta}' = \hat{\eta}_1$$

$$\hat{\delta} = \hat{\eta}_2$$

$$\hat{\beta} = \hat{\eta}_3$$

En procédant de la même manière que précédemment, nous obtenons les probabilités du modèle $y = 2, 3, 4$:

- $P(\text{Produit } l) = P(y = 2) = \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_l}{\sigma}; -\frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y = 3) = \Phi(\frac{\delta - \beta' x_l}{\sigma})$
- $P(\text{Conflit}) = P(y = 4) = 1 - \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\delta - \beta' x_l}{\sigma}) + \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma}, \frac{\delta - \beta' x_l}{\sigma}; -\frac{1}{\sqrt{2}})$

où la fonction de vraisemblance L de ce modèle est la suivante :

$$L = \prod_{s=1}^{S} P(y_s = 2)^{y_{2s}} P(y_s = 3)^{y_{3s}} P(y_s = 4)^{y_{4s}}$$

avec S égal au nombre total de choix effectués par un individu et

$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

et où la maximisation du logarithme de cette fonction L, qui fournit les paramètres inconnus et réduits du modèle, est :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{2s} \log P_{2s} + y_{3s} \log P_{3s} + y_{4s} \log P_{4s} \right)$$

En vérifiant l'identifiabilité des coefficients pour savoir si l'estimation des paramètres réduits donne directement celle des paramètres structurels, nous constatons qu'il y a sous-identification :

$$\hat{\eta}_1 = \frac{\delta'}{\sigma} \implies \text{inconnues} \quad (\delta', \sigma)$$

$$\hat{\eta}_2 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_3 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{3 équations et 4 inconnues : sous-identification}$$

Par conséquent, nous normalisons un des coefficients, à savoir $\sigma = 1$, pour obtenir l'estimation direct des paramètres structurels :

$$\hat{\delta}' = \hat{\eta}_1$$

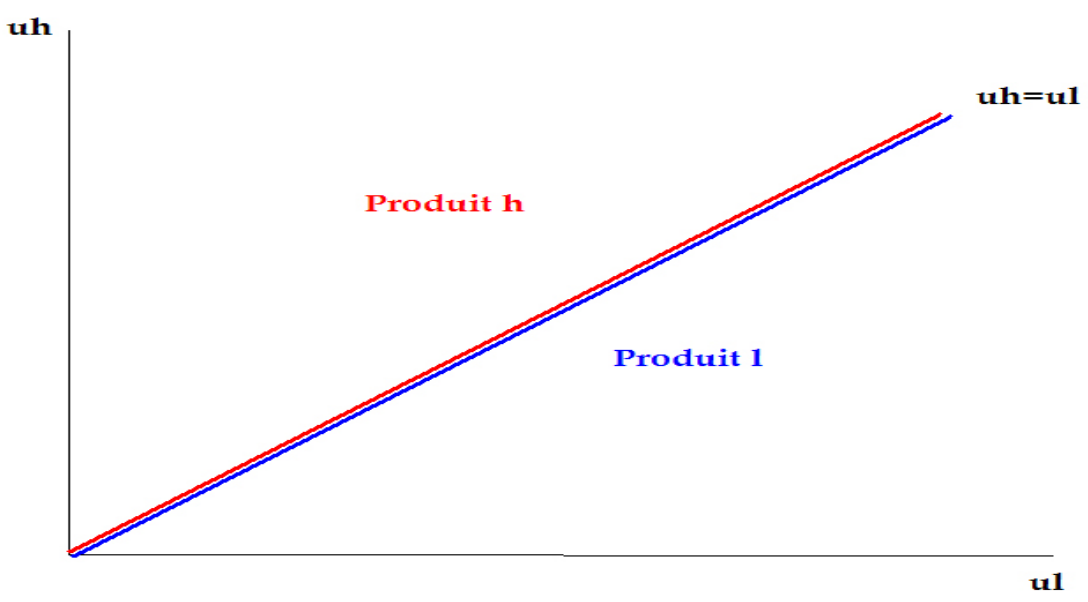
$$\hat{\delta} = \hat{\eta}_2$$

$$\hat{\beta} = \hat{\eta}_3$$

3.5.4 Modèle où y=1,2

Le modèle $y = 1, 2$ suppose que l'individu ne choisit que deux options parmi les quatre. Ces deux options sont le produit h ou le produit l. Ces choix signifient que l'individu ne se trouve pas en situation de conflit et qu'il apprécie tous les produits, étant donné qu'il ne choisit pas l'alternative "refus".

Nous pouvons illustrer ce modèle de la manière suivante :



De cette représentation, nous définissons la probabilité de choix du produits h et du produit l comme suit :

$$P(y = h) = P(y = 1) = P(u_h > u_l) = P(\varepsilon_h - \varepsilon_l > \beta'(x_l - x_h))$$
$$P(y = l) = P(y = 2) = P(u_h < u_l) = P(\varepsilon_h - \varepsilon_l < \beta'(x_l - x_h))$$

Nous transformons ces probabilités de sorte à les formuler en fonction des quantiles d'une loi normale univariée. Leur expression finale est la suivante :

$$P(y = h) = P(y = 1) = 1 - \Phi\left(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}\right)$$
$$P(y = l) = P(y = 2) = \Phi\left(\frac{\beta'(x_l - x_h)}{\sqrt{2}\sigma}\right)$$

Ces probabilités auraient également pu être déduites du modèle où $y = 1, 2, 3, 4$. Il aurait suffi de poser le seuil du conflit δ' à 0 et le seuil du refus δ à $-\infty$ dans les probabilités de choisir le produit h et l.

La fonction de vraisemblance d'un individu du modèle $y = 1, 2$ se formule comme suit :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 2)^{y_{2s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$
$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

La maximisation du logarithme de cette fonction L permet d'obtenir les paramètres inconnus et réduits du modèle :

$$\max \log L \quad \Longleftrightarrow \quad \max\left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{2s} \log P_{2s}\right)$$

Il faut vérifier l'identifiabilité des paramètres pour savoir si l'estimation des paramètres réduits donne directement celle des paramètres structurels :

$$\hat{\eta}_1 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\Longrightarrow \text{ 1 équations et 2 inconnues : sous-identification}$$

Etant donné que les paramètres sont sous-identifiés, il faut donc normaliser une des paramètres. Posons $\sigma = 1$. Les paramètres structurels s'expriment donc comme suit :

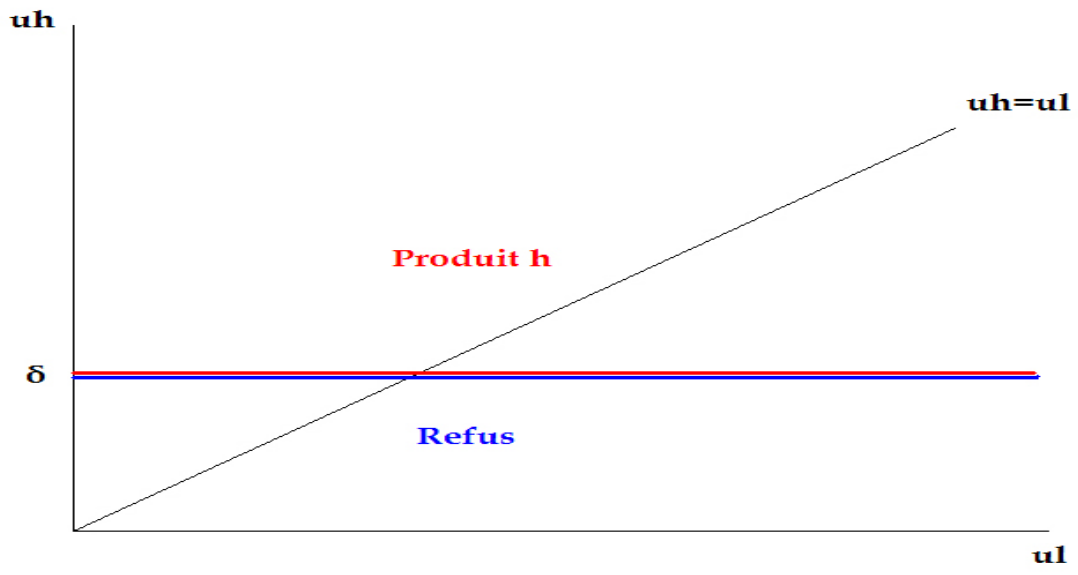
$$\hat{\beta} = \hat{\eta}_1$$

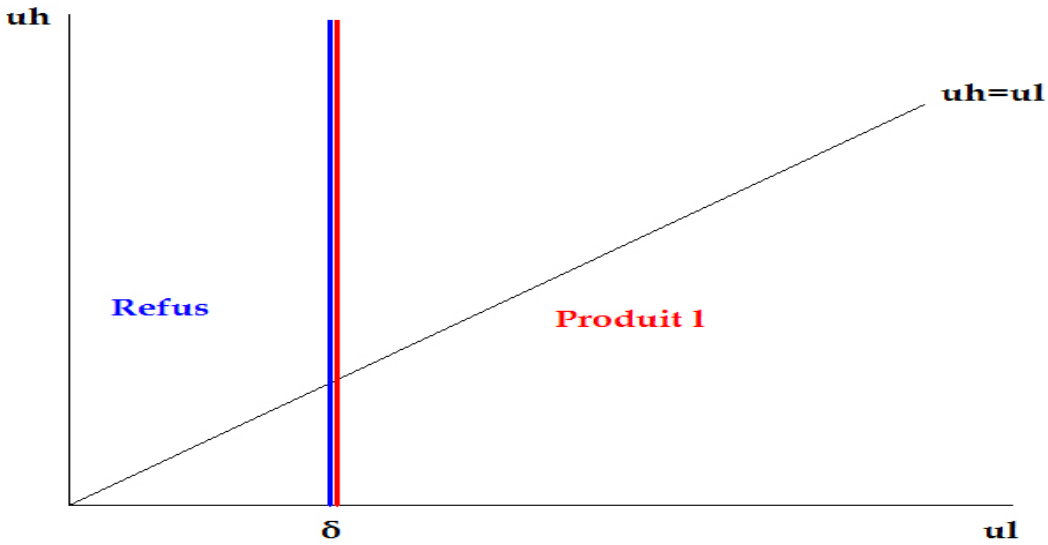
La maximisation du logarithme de la fonction de vraisemblance donne donc directement l'estimation des paramètres structurels.

3.5.5 Modèle où y=1,3, respectivement y=2,3

L'individu du modèle $y = 1, 3$, respectivement $y = 2, 3$, ne choisit que deux options parmi les quatre. Ces deux options sont le produit h, respectivement le produit l, et le refus. Ces choix signifient que le consommateur n'apprécie pas du tout le produit l, respectivement le produit h, et qu'il ne se trouve pas en situation de conflit.

L'illustration graphique du modèle $y = 1, 2$, respectivement $y = 2, 3$, est la suivante :





Selon le graphe du modèle $y = 1, 3$, les probabilités de choisir le produit h et le refus s'écrivent :

$$P(y = h) = P(y = 1) = P(u_h > \delta) = P(\varepsilon_h > \delta - \beta' x_h)$$

$$P(y = \text{Refus}) = P(y = 3) = P(u_h < \delta) = P(\varepsilon_h < \delta - \beta' x_h)$$

Ces probabilités peuvent être formulées en fonction des quantiles de la loi normale univariée :

$$P(y = h) = P(y = 1) = 1 - \Phi\left(\frac{\delta - \beta' x_h}{\sigma}\right)$$

$$P(y = \text{Refus}) = P(y = 3) = \Phi\left(\frac{\delta - \beta' x_h}{\sigma}\right)$$

La fonction de vraisemblance d'un individu du modèle $y = 1, 3$ s'écrit :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 3)^{y_{3s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

La maximisation du logarithme de cette fonction L est :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{3s} \log P_{3s} \right)$$

En vérifiant l'identifiabilité des paramètres :

$$\hat{\eta}_1 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

nous constatons qu'il faut en normaliser un et posons donc $\sigma = 1$ pour obtenir une estimation directe des paramètres structurels :

$$\hat{\delta} = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

En procédant à l'identique pour le modèle $y = 2, 3$, les probabilités de choisir le produit l et le refus sont :

$$P(y = l) = P(y = 2) = 1 - \Phi(\frac{\delta - \beta' x_l}{\sigma})$$

$$P(y = \text{Refus}) = P(y = 3) = \Phi(\frac{\delta - \beta' x_l}{\sigma})$$

où la fonction de vraisemblance d'un individu de ce modèle s'écrit :

$$L = \prod_{s=1}^{S} P(y_s = 2)^{y_{2s}} P(y_s = 3)^{y_{3s}}$$

avec S égale au nombre total de choix effectué par un individu et

$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

Le logarithme de cette fonction de vraisemblance L à maximiser s'écrit :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{2s} \log P_{2s} + y_{3s} \log P_{3s} \right)$$

Il faut vérifier l'identifiabilité des paramètres pour savoir si l'estimation des paramètres réduits donne directement celle des paramètres structurels.

L'identification des paramètres structurels s'effectue à l'aide des équations suivantes :

$$\hat{\eta}_1 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

qui montrent une sous-identification des paramètres et amènent à en normaliser un, à savoir $\sigma = 1$, pour obtenir le résultat direct d'estimation de ces derniers :

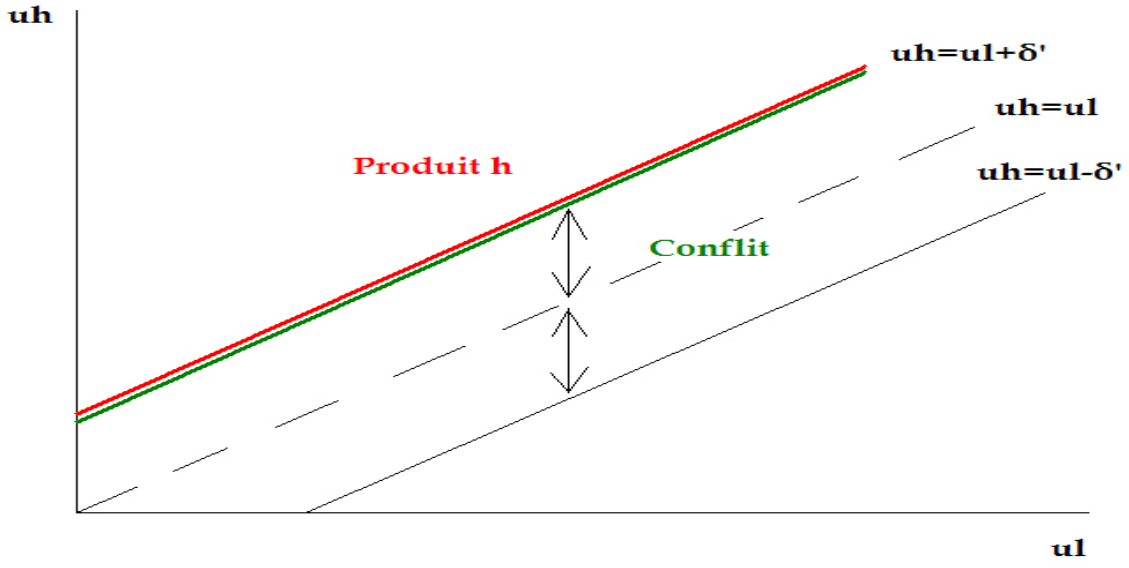
$$\hat{\delta} = \hat{\eta}_1$$

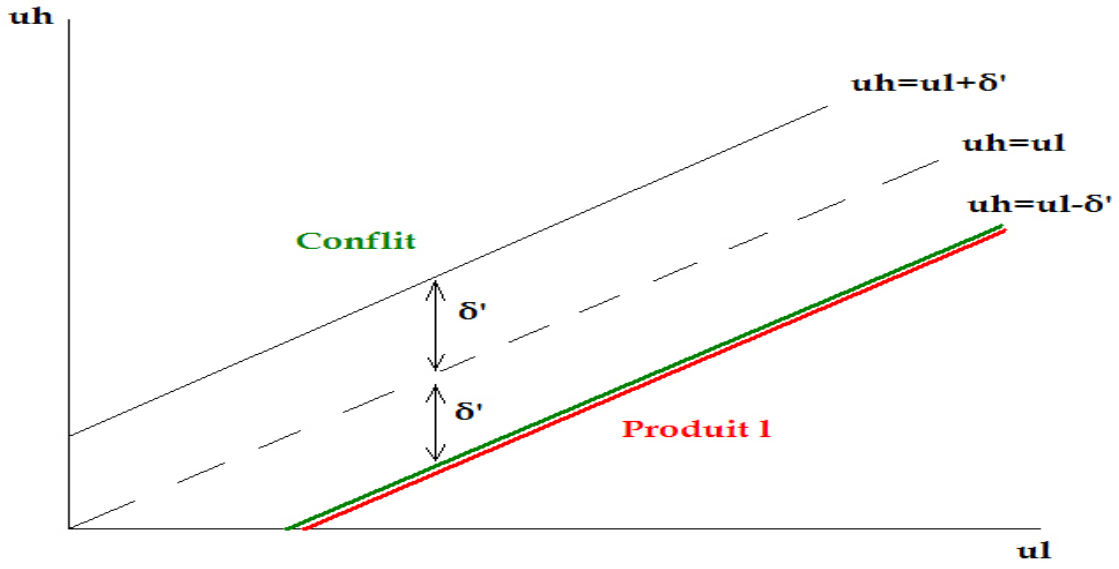
$$\hat{\beta} = \hat{\eta}_2$$

3.5.6 Modèle où y=1,4, respectivement y=2,4

Dans ces modèles, l'individu ne choisit que le produit h, respectivement le produit l, et le conflit. Ces choix signifient que l'individu n'apprécie pas du tout le produit l, respectivement le produit h, et qu'il ne se trouve pas en situation de refus.

Graphiquement, ces situations s'illustrent comme suit :





Du graphique du modèle $y = 1, 4$, les probabilités de choisir le produit h et de se trouver en situation de conflit s'écrivent :

$$P(y = h) = P(y = 1) = P(u_h > u_l + \delta') = P(\varepsilon_h - \varepsilon_l > \beta'(x_l - x_h) + \delta')$$

$$P(y = \text{Conflit}) = P(y = 4) = P(u_h < u_l + \delta') = P(\varepsilon_h - \varepsilon_l < \beta'(x_l - x_h) + \delta')$$

En exprimant ces probabilités en fonction des quantiles de la loi normale univariée, nous obtenons :

$$P(y = h) = P(y = 1) = 1 - \Phi\left(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}\right)$$

$$P(y = \text{Conflit}) = P(y = 4) = \Phi\left(\frac{\beta'(x_l - x_h) + \delta'}{\sqrt{2}\sigma}\right)$$

La fonction de vraisemblance d'un individu du modèle $y = 1, 4$ s'écrit :

$$L = \prod_{s=1}^{S} P(y_s = 1)^{y_{1s}} P(y_s = 4)^{y_{4s}}$$

avec S correspond au nombre total de choix effectués par un individu,

$$y_{1s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } h \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

et le logarithme de cette fonction L à maximiser est égal à :

$$\max \log L \quad \Longleftrightarrow \quad \max\left(\sum_{s=1}^{S} y_{1s} \log P_{1s} + y_{4s} \log P_{4s}\right)$$

En vérifiant l'identifiabilité des paramètres pour savoir si l'estimation des paramètres réduits donne directement celle des paramètres structurels :

$$\hat{\eta}_1 = \frac{\delta'}{\sigma} \implies \text{inconnues} \quad (\delta', \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

nous constatons qu'un des paramètres doit être normalisé. Soit $\sigma = 1$. Les paramètres

structurels sont directement déduits de l'estimation des paramètres réduits :

$$\hat{\delta}' = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

Les probabilités du modèle $y = 2, 4$ sont obtenues de la même manière que pour le modèle $y = 1, 4$:

$$P(y = l) = P(y = 2) = 1 - \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma})$$

$$P(y = \text{Conflit}) = P(y = 4) = \Phi(\frac{\beta'(x_l - x_h) - \delta'}{\sqrt{2}\sigma})$$

La fonction de vraisemblance L d'un individu de ce modèle s'écrit :

$$L = \prod_{s=1}^{S} P(y_s = 2)^{y_{2s}} P(y_s = 4)^{y_{4s}}$$

où S correspond au nombre total de choix effectués par un individu et

$$y_{2s} = \begin{cases} 1 & \text{lorsque l'individu choisit le produit } l \text{ dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

Le logarithme de cette fonction L à maximiser :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{2s} \log P_{2s} + y_{4s} \log P_{4s} \right)$$

permet d'obtenir les estimations des coefficients réduits, dont il faut vérifier l'identifiabilité des paramètres structurels.

Les équations associant les paramètres réduits aux paramètres structurels sont les suivantes :

$$\hat{\eta}_1 = \frac{\delta'}{\sigma} \implies \text{inconnues} \quad (\delta', \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

Etant donné que les paramètres sont sous-identifiés, nous normalisons le coefficient $\sigma = 1$.

Les paramètres structurels sont déduits de la manière suivante :

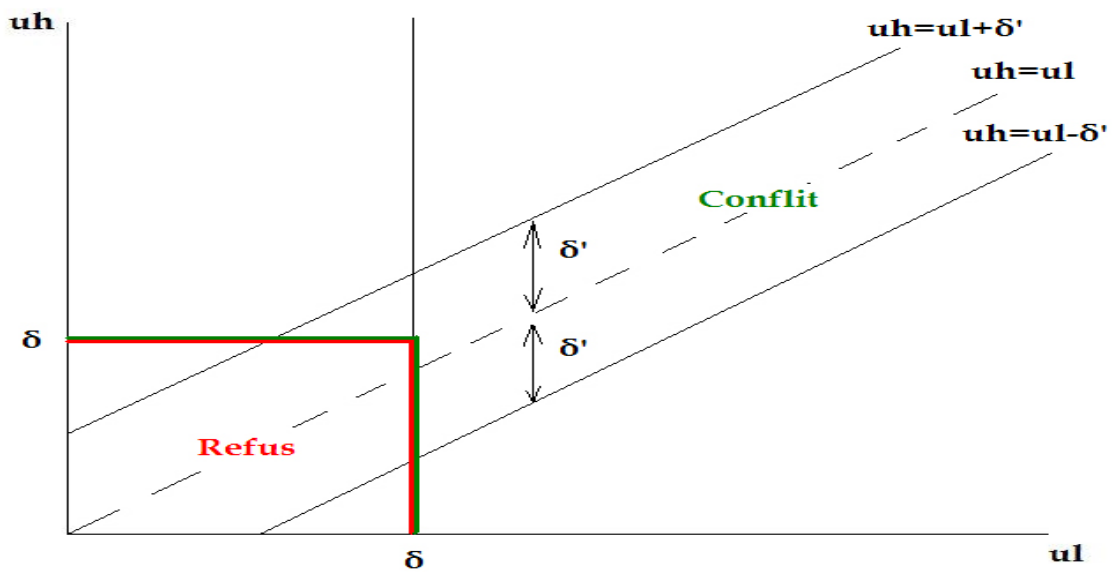
$$\hat{\delta}' = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

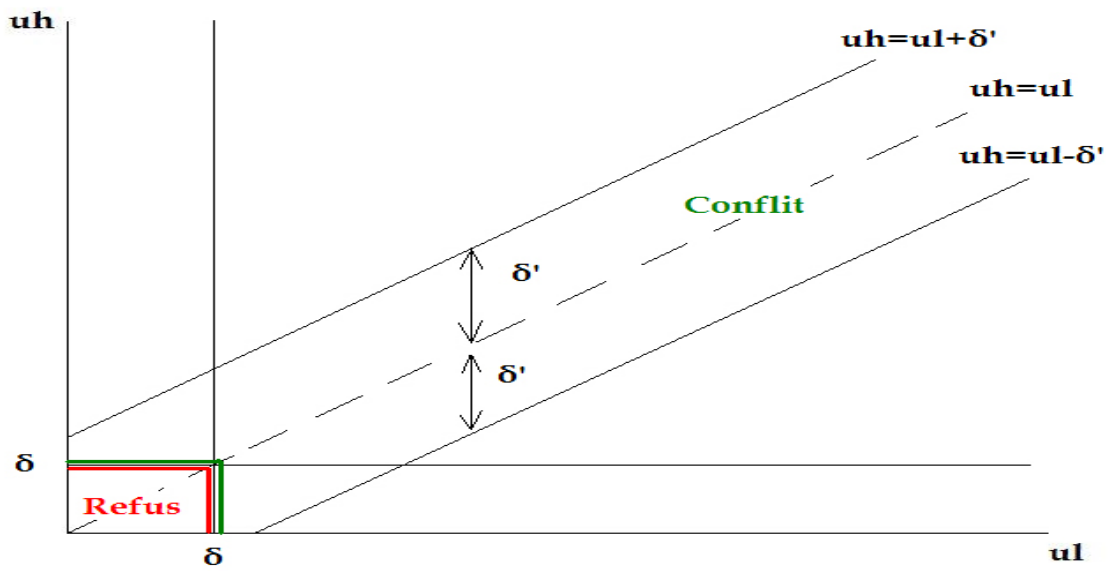
c'est-à-dire directement de l'estimation des paramètres réduits.

3.5.7 Modèle où y=3,4

Dans cette sous-section, nous supposons que l'individu est soit en conflit, soit qu'il n'apprécie aucun des produit h et l. De plus, selon les valeurs de δ et δ', l'illustration graphique de ce modèle diffère. En effet, lorsque $\delta > \delta'$, le graphique du modèle $y = 3,4$ se dessine de la manière suivante :



alors que lorsque $\delta < \delta'$, il est représenté comme suit :



L'expression des probabilités de ce modèle est, cependant, invariant aux différentes valeurs de δ et δ', étant donné que u_h, respectivement u_l, varie de $-\infty$ à ∞.

Ces probabilités sont définies comme suit :

$$P(\text{Refus}) = P(y=3) = P(u_h < \delta \cap u_l < \delta)P(\varepsilon_h < \delta - \beta' x_h, \varepsilon_l < \delta - \beta' x_l; \rho = 0)$$

$$P(\text{Conflit}) = P(y=4) = P(u_h > \delta \cup u_l > \delta) = P(\varepsilon_h > \delta - \beta' x_h \cup \varepsilon_l > \delta - \beta' x_l)$$

En procédant de la même manière que pour le modèle où $y = 1, 2, 3, 4$ et en appliquant les propriétés des probabilités $P(A \cup B) = P(A) + P(B) + P(A \cap B)$, les probabilités du modèle $y = 3, 4$ s'écrivent :

$$P(\text{Refus}) = P(y=3) = \Phi(\frac{\delta - \beta' x_h}{\sigma})\Phi(\frac{\delta - \beta' x_l}{\sigma})$$

$$P(\text{Conflit}) = P(y=4) = P(\beta' x_h + \varepsilon_h > \delta \cup \beta' x_l + \varepsilon_l > \delta)$$

$$= P(\varepsilon_h > \delta - \beta' x_h) + P(\varepsilon_l > \delta - \beta' x_l) - P(\varepsilon_h > \delta - \beta' x_h, \varepsilon_l > \delta - \beta' x_l; \rho = 0)$$

$$= 1 - \Phi(\frac{\delta - \beta' x_h}{\sigma})\Phi(\frac{\delta - \beta' x_l}{\sigma})$$

La fonction de vraisemblance d'un individu du modèle $y = 3, 4$ se formule de la manière suivante :

$$L = \prod_{s=1}^{S} P(y_s = 3)^{y_{3s}} P(y_s = 4)^{y_{4s}}$$

où S correspond au nombre total de choix effectués par un individu,

$$y_{3s} = \begin{cases} 1 & \text{lorsqu'aucun produit ne plaît au consommateur dans le sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

$$y_{4s} = \begin{cases} 1 & \text{lorsque l'individu est en situation de conflit face aux produits du sous-ensemble } s \\ 0 & \text{sinon} \end{cases}$$

et le logarithme de cette fonction L à maximiser :

$$\max \log L \quad \Longleftrightarrow \quad \max \left(\sum_{s=1}^{S} y_{3s} \log P_{3s} + y_{4s} \log P_{4s} \right)$$

Il suffit de vérifier l'identifiabilité des paramètres à l'aide des équations suivantes :

$$\hat{\eta}_1 = \frac{\delta}{\sigma} \implies \text{inconnues} \quad (\delta, \sigma)$$

$$\hat{\eta}_2 = \frac{\beta}{\sigma} \implies \text{inconnues} \quad (\beta, \sigma)$$

$$\implies \text{2 équations et 3 inconnues : sous-identification}$$

pour connaître l'expression des paramètres structurels. Etant donné que ces derniers sont sous-identifiés, nous en normalisons un, c'est-à-dire $\sigma = 1$, pour obtenir l'estimation directe des paramètres structurels :

$$\hat{\delta} = \hat{\eta}_1$$

$$\hat{\beta} = \hat{\eta}_2$$

Conclusion

Dans la littérature, le non choix est spécifié par une utilité ne comportant que des 0 selon Elrod, Louviere et Krishnakumar (1992) ou par un modèle hiérarchique selon Haaijer (1999). Le logiciel Sawthooth intègre le non choix sans spécifier, dans sa documentation relative à ce programme, la forme du modèle. La littérature n'étant pas riche sur ce domaine, il était donc intéressant de se pencher sur cette problématique. L'option de poser à 0 l'utilité et les attributs du modèle ne nous a pas semblé être une voie concluante. En effet, Haaijer (1999) a démontré le biais lié à cette méthodologie dans le cas d'attributs représentant des variables continues, telles qu'un prix. De plus, nous pensons qu'il y a également un biais, lorsque la variable indépendante est qualitative, étant donné qu'une valeur de 0 définit généralement la modalité de référence. L'idée d'un modèle hiérarchique semble donc plus satisfaisante du point de vue du biais. Toutefois, cette spécification implique une modélisation en plusieurs étapes, alors que nous préférons spécifier l'option "zéro" en une étape.

La définition du non choix a également été un autre aspect intéressant auquel il nous a semblé important de l'approfondir. L'option "zéro", telle qu'elle est appliqué dans la littérature de l'analyse conjointe, se réfère à une notion que nous nommerons de "rationnelle". En effet, l'individu connaît avec certitude ses préférences et il choisit donc l'option "zéro" que lorsque les produits proposés ne sont pas à son goût. Cependant, le comportement humain n'est pas aussi simple et aussi rationnel. En effet, les individus préfèrent également ne pas choisir lorsque les produits mis en vente sont attrayants et difficiles à départager. Cette prise de décision (ne pas choisir) découle du conflit engendré par la difficulté à émettre une préférence. Ce conflit est largement défini dans la littérature psychologique. Il est abordé dans l'un des articles de Tversky et Shafir (1992) où il est mis

en relation avec la notion de maximisation d'utilité.

La maximisation d'utilité est le concept primordial expliquant le choix. En effet, chaque produit a une utilité qui détermine, selon sa valeur, le niveau d'appréciation du produit. Toutefois, les modèles de l'option "zéro" décrit dans la littérature associent une utilité à ce non choix et appliquent la maximisation d'utilité afin de le classer parmi les autres produits. Or, Tversky et Shafir (1992) ont montré que cette méthodologie ne reflète pas la réalité et conduit donc à des résultats biaisés. Ces conclusions proviennent du fait que la maximisation d'utilité ne considère pas les non choix dus à un conflit. De plus, il n'existe pas de modèle dans la littérature incorporant le conflit.

La maximisation d'utilité n'échouerait-elle pas car elle compare des utilités de produits avec celle d'un non choix ? Il a donc fallu ne pas associer d'utilité à l'option "zéro" et déterminer le choix uniquement en fonction des utilités des produits. Simultanément, nous avons étendu ce point de vue en intégrant le non choix lié au conflit. L'idée de base reste la même, à savoir qu'aucune utilité n'est associée à l'option "zéro", qu'elle soit due à un choix rationnel ou à un conflit. Les différentes catégories ne sont donc sélectionnées qu'en fonction des utilités des produits, qui sont interprétables contrairement à celles associées à tous types de non choix.

D'un point de vue économétrique, notre spécification est innovante, car elle mélange les modèles à variable dépendante qualitative ordonnées à ceux à variable dépendante qualitative non ordonnée. En effet, notre spécification est un mélange des modèles de comparaison d'utilité et des seuils de réactions. La comparaison entre les utilités des produits persiste, mais elle est conditionnée par un seuil lié au non choix rationnel et au conflit.

L'autre aspect contraignant est celui de la définition de l'analyse conjointe. En effet, cette dernière a pour but d'obliger les individus à faire des choix lorsqu'ils sont confrontés à des produits "similairement" attrayants. En effet, il ne serait pas intéressant de demander de faire un choix entre des produits dominants et inférieurs étant donné que la réponse serait évidente. La conséquence de cette manière de procéder permet la connaissance des réelles préférences ou peut également amener l'individu à faire un choix au hasard dans le cas où le modèle est sans l'option "zéro" ou de choisir l'option "zéro" dans le cas où le modèle est avec l'option "zéro". L'introduction du non choix n'est-elle pas une voie de sortie dans les choix difficiles ? Est-elle finalement équivalente aux nombres de réponses au hasard dans le modèle sans l'option "zéro" ? C'est pourquoi, nous introduisons dans notre spécification le non choix dû au conflit, qui est une source d'information complémentaire au non choix rationnel. En effet, elle permet de connaître les véritables produits et donc les

attributs qui plaisent et qui ne plaisent pas à l'individu. Toutefois, le choix unique est une source d'information précise, mais dans le modèle classique, il peut aboutir à des réponses biaisées, ce qui n'est pas le cas avec notre nouvelle spécification. La raison est que, dans le questionnaire, il est proposé aux individus les options : un des produits, tous les produits ou aucun produit. Le choix de tous les produits représente le choix dû au conflit. Par conséquent, l'individu n'utilise pas le non choix comme une voie de sortie puisqu'il peut choisir plusieurs produits à la fois. Cependant, l'individu n'étant pas obligé d'émettre un seul choix, nous pourrions conclure que, selon la définition de l'analyse conjointe associée à un modèle compensatoire, ce dernier ne fait donc pas de compromis entre les attributs et ne révèle donc pas ses véritables valeurs. Cependant, les choix par ces compromis sont-ils issus d'une réflexion de l'individu sur ses propres valeurs ou sont-ils le fruit du hasard ? Pour répondre à cette question, il faudrait comparer le modèle sans et avec le refus. Cette comparaison entre le modèle de choix et de non choix ayant déjà été effectuée dans l'ouvrage de Haaijer (1999), dont la conclusion est une adéquation meilleure pour le non choix, nous permet de répondre à l'interrogation suivante : les individus utiliseraient-t-ils le non choix comme une voie de sortie ? La réponse est bien évidemment non, puisque l'adéquation de ce modèle est meilleure que celle du choix. Cependant, Haaijer (1999) remarque un pourcentage de non choix (proche de 50%) élevé. La question qu'il faut donc se poser est : y a-t-il un grand nombre de réponses au hasard dans le modèle de choix ? Si la réponse est oui, ceci pourrait expliquer la meilleure adéquation du modèle de non choix puisque le résultat de l'interrogation refléterait finalement les vrais goûts des consommateurs, contrairement au modèle de choix. Cependant, il n'est pas possible de vérifier cela.

En conclusion, notre nouvelle spécification intègre les quelques commentaires de la littérature. En effet, elle tient compte du biais lié à l'utilité du non choix par des 0, ainsi que celui lié à la prise de considération uniquement du non choix rationnel. Elle propose également une spécification en une étape. Par la suite, une application de notre nouveau modèle est effectuée. Elle utilise l'approche du maximum de vraisemblance pour estimer les paramètres de notre spécification de l'option "zéro" avec le refus et le conflit. En pratique, cette approche présente, toutefois, des problèmes de convergence. Ces derniers se manifestent essentiellement dans les petits échantillons. Etant donné que notre modèle est estimé de manière individuelle, il est donc fort probable que le maximum de vraisemblance ne soit pas convergent. Par conséquent, le chapitre suivant de notre thèse traite des solutions liées à cette non convergence du maximum de vraisemblance.

Chapitre 4

Notions de convergence et de séparation dans la méthode du maximum de vraisemblance

Introduction

Les modèles non linéaires à variable dépendante qualitative, tels que les modèles probit et logit, utilisent la méthode du maximum de vraisemblance (MV) pour estimer les paramètres inconnus. Cette méthode n'aboutissant pas toujours à des résultats selon la nature des données, l'existence d'une solution n'est donc pas assurée. L'absence de solutions est fréquemment constatée dans les échantillons de petite taille, c'est-à-dire en dessous de 50 observations.

Dans l'approche de l'analyse conjointe, il est fréquent d'effectuer des estimations sur des petits échantillons. Lorsque cette analyse provient d'un recueil de données du type classement ou notation, le problème lié aux échantillons petits n'a pas lieu d'être, étant donné que l'analyse est souvent effectuée à l'aide des moindres carrés ordinaires. Quant à la récolte de données par des choix, l'analyse par les modèles probit ou logit peut, soit être fait de manière agrégée, soit de manière individuelle. Dans le cas de l'agrégé, les données sont suffisamment nombreuses et ne révèlent que rarement des problèmes d'existence de solutions. Par contre, l'estimation individuelle s'effectue avec des petites tailles d'échantillon, étant donné que les observations représentent le nombre de choix de produits par individu. Par conséquent, afin de satisfaire une proportion de réponses fiables, il est indispensable de limiter le nombre d'observations. De tels échantillons dans l'analyse individuelle des préférences par des modèles probit ou logit aboutissent fréquemment à une estimation divergente d'un ou plusieurs des paramètres. Cet aspect divergent des

paramètres s'appelle la séparation. Elle est décrite dans la section suivante.

De nombreuses solutions à cette problématique ont été proposées dans la littérature. Des études comparatives ont également été menées. Ces résultats sont abordés dans ce chapitre. Dans la littérature, les solutions étant uniquement décrites pour les modèles logit à variable dépendante binaire, les aspects avancés de ces méthodologies ne sont donc pas abordés de manière aussi détaillée dans la suite de l'étude. La raison est que dans notre modèle de l'option "zéro", nous n'utilisons que la spécification probit pour estimer les paramètres inconnus. De plus, notre spécification de l'option "zéro" étant un mélange de modèles de comparaison et de seuils, les conclusions liées aux solutions présentées dans la littérature peuvent ne pas être valides et ne pas être facilement applicables et implémentables.

4.1 Notions de convergence et de séparation

Les estimations par le maximum de vraisemblance peuvent fournir plusieurs directions, telles que l'existence d'une solution unique, de plusieurs solutions et l'inexistence d'une solution. La solution unique est appelée dans la littérature "overlap", en français "recouvrement". Elle représente la situation idéale de la méthode du maximum de vraisemblance. Il suffit simplement de récolter les résultats de l'estimation sans pour autant faire appel à d'autres manipulations. A l'opposé, plusieurs maximums dans la log-vraisemblance obligent le chercheur à modifier certains points de la procédure. En effet, lors de la programmation informatique du maximum de vraisemblance, la plupart des logiciels spécifient des valeurs de départ pour l'algorithme d'estimation de la fonction MV. Ces valeurs permettent de lancer la procédure. Toutefois, des valeurs spécifiées proches d'un maximum local amène à des résultats proches de ce dernier, plutôt que de ceux du maximum global. Afin d'obtenir un optimum global, on teste plusieurs valeurs de départ. Nous tenons cependant à faire remarquer qu'il n'est pas toujours évident de repérer une situation de maximums multiples. En effet, les logiciels se contentent généralement de trouver un maximum quelconque et ne donnent pas le moindre avertissement quant à la nature du point trouvé. La raison est que si ce dernier était capable de différencier un maximum local à un global, il donnerait la solution associée au maximum global. Par conséquent, il est utile de tester plusieurs valeurs de départ dans les situations de maximums uniques et multiples. Dans le cadre de la fonction de vraisemblance du modèle logit, le problème de solutions multiples ne se pose pas, étant donné que la fonction est globalement concave et assure, par cette propriété, l'unicité d'un maximum (Amemiya (1985)).

Le maximum de vraisemblance n'est pas toujours capable de fournir une solution. Cette

situation, appelée "séparation", est fréquente dans les petits échantillons, avec des variables explicatives dont leurs valeurs prédictives dans le modèle sont très élevées vis-à-vis de la variable dépendante et/ou lorsque le rapport entre le nombre d'observations et le nombre de paramètres est faible ou inférieur à 10, à savoir si par exemple, $\frac{\text{nombre d'observations}}{\text{nombre de paramètres}} = \frac{50}{10} = 5 < 10$. La dernière raison, qui est susceptible de provoquer une séparation dans les données, a en réalité tendance à augmenter le biais d'estimation des paramètres du maximum de vraisemblance selon Bull, Mak et Greenwood (2002). En effet, l'estimateur du maximum de vraisemblance a généralement un biais d'ordre $O(n^{-1})$ selon Firth (1993). La méthode proposée dans l'article de Firth (1993) consiste à enlever ce biais. Par la suite, Heinze et Schemper (2002) ont découvert que cette méthodologie est également valable dans la résolution d'estimation des modèles logit comportant de la séparation dans ces données. Cette solution est décrite en détail dans la suite de cette section.

La séparation se distingue selon qu'elle est complète ou quasi-complète. Cette appellation est due à Albert et Anderson (1984). Il en est de même pour le terme "overlap". La séparation complète ou quasi-complète se définit de plusieurs manières. Techniquement, la séparation complète résulte en une division nette des modalités de la variable expliquée y_i en fonction d'une combinaison linéaire de ses variables explicatives $a'x_i$.

Soit l'exemple des jeans discuté dans la section sur les plans d'expérience pour illustrer la séparation complète. Dans cet exemple, nous ne nous intéressons qu'aux attributs suivants : la forme, la marque et la couleur. Les produits issus de la combinaison de ces variables sont :

Obs	Forme	Marque	Couleur
1	Slim	Gap	Claire
2	Slim	Gap	Foncée
3	Slim	H&M	Claire
4	Slim	H&M	Foncée
5	Bootcut	Gap	Claire
6	Bootcut	Gap	Foncée
7	Bootcut	H&M	Claire
8	Bootcut	H&M	Foncée

Afin d'illustrer la séparation complète dans les données, nous enlevons la première observation. Nous ne présentons donc que 7 produits à évaluer. La matrice des variables explicatives est composée des vecteurs x_i, $i = 1, \ldots, 7$, représentant les caractéristiques des produits :

$$X = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Soit la variable dépendante qualitative y représentant l'achat ou le non achat des 7 produits :

$$y_i = \begin{cases} 1 & \text{achat} \\ 0 & \text{non achat} \end{cases}$$

Si les réponses de la variable dépendante y sont égales à :

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

nous constatons qu'il y a une combinaison linéaire qui départage clairement l'achat du non achat. En effet, si, par exemple, $a' = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix}$, alors

$$\begin{cases} a'x_i > 0 & y_i = 1, \text{ pour } i = 4, 5 \\ a'x_i < 0 & y_i = 0 \text{ pour } i = 1, 2, 3, 6, 7 \end{cases}$$

En effet, pour $i = 4, 5$

$$a'x_4 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 0.5 > 0 \qquad a'x_5 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0.4 > 0$$

et pour $i = 1, 2, 3, 6, 7$

$$a'x_1 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -0.1 < 0 \qquad a'x_2 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -1 < 0$$

$$a'x_3 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = -1.1 < 0 \qquad a'x_6 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = -0.5 < 0$$

$$a'x_7 = \begin{pmatrix} 0.5 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -0.6 < 0$$

Dans la séparation quasi-complète, cette division des modalités de la variable dépendante y n'est pas aussi nette qu'avec la séparation complète. Reprenons l'exemple ci-dessus en y incluant cette fois-ci la première observation. Nous avons donc 8 produits dont la matrice des variables explicatives X est égale à :

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Supposons que le vecteur de la variable dépendante y soit le suivant :

$$y = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Ces données contiennent une séparation quasi-complète. La raison est qu'il existe un vecteur a, tel que :

$$\begin{cases} a'x_i \geq 0 & y_i = 1, \text{ pour } i = 1, 5 \\ a'x_i \leq 0 & y_i = 0 \text{ pour } i = 2, 3, 4, 6, 7, 8 \end{cases}$$

En effet, si par exemple, $a' = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix}$, alors pour $i = 1, 5$

$$a'x_1 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \geq 0 \qquad a'x_5 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 0.5 \geq 0$$

et pour $i = 2, 3, 4, 6, 7, 8$

$$a'x_2 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -0.5 \leq 0 \qquad a'x_3 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -1 \leq 0$$

$$a'x_4 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = -1.5 \leq 0 \qquad a'x_6 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0 \leq 0$$

$$a'x_7 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = -0.5 \leq 0 \qquad a'x_8 = \begin{pmatrix} 0.5 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1 \leq 0$$

Il suffit donc qu'au moins une observation ait une combinaison linéaire des variables explicatives nulle dans chaque modalité de la variable expliquée pour qu'il y ait séparation quasi-complète.

Cette définition de la séparation complète et quasi-complète est très technique et divise la variable dépendante en fonction de ces variables explicatives. En d'autres termes, les attributs prédisent parfaitement les modalités de la variable y. Cette définition apparaît dans certains logiciels d'estimation MV et n'est qu'associée à la séparation complète. En effet, lors de séparation quasi-complète, aucun avertissement du genre précédent n'est donné par la plupart des logiciels d'estimation. Il existe d'autres manières plus intuitives pour constater la séparation des données. Dans le cas de variables explicatives qualitatives, la séparation complète peut être observée dans un tableau représentant les modalités de la variable dépendante en fonction de celles de chaque attribut. Un tableau comportant toutes ses cases vides, à l'exception des diagonales, décrit la séparation complète. Quant à la séparation quasi-complète, elle est observée si dans les cases vides se trouvant en dehors des diagonales, au moins une ne l'est pas. La séparation se constate également dans les estimations par maximum de vraisemblance dont les valeurs des paramètres et des écart-types divergent vers $\pm\infty$. Cette constatation est valable pour tous les paramètres issus de données de séparation complète et pour certains dans le cas de la séparation quasi-complète, à savoir ceux dont l'apport prédictif est très élevé vis-à-vis de la variable

réponse. En pratique, les estimations des paramètres et des écart-types ne sont pas infinies, mais plutôt très élevées. De plus, dans ces situations de séparation, il est fréquent d'observer une valeur de l'écart-type bien plus grande que celle du paramètre.

L'explication de valeurs infinies dans les estimations est donnée par la monotonicité du logarithme de la fonction de vraisemblance dans la cas de la non existence d'une solution ou de la séparation. Cette fonction a en effet un aspect "plat", voir une asymptote horizontale au voisinage de 0 pour la séparation complète et au voisinage d'une valeur inférieure à 0 pour la séparation quasi-complète. Une variation très faible de cette fonction dans le voisinage mentionné ci-dessus provoque donc des changements élevés des paramètres estimés. C'est pourquoi, les estimations des paramètres issus d'observations comportant la séparation sont élevées dans les résultats des logiciels. Certains programmes concluent même à la convergence malgré les valeurs infinies des paramètres et des écart-types. La raison dépend du critère de convergence du logiciel. Si ce dernier est en fonction de la log-vraisemblance, il semble compréhensible qu'au voisinage des valeurs citées précédemment le changement de la fonction soit suffisamment petit pour conclure à la convergence. Par conséquent, il est préférable, comme le suggère Allison (2004) dans "*Numerical Issues in statistical computing for the social scientist*", de surveiller les variations des paramètres durant chacune des itérations pour confirmer la convergence de l'estimation du modèle. L'explication de la grande valeur des écart-types est, quant à elle, identique à celle des paramètres. En effet, la monotonicité du logarithme de la fonction de vraisemblance au voisinage de 0 ou d'une valeur inférieure à 0 selon le type de séparation, amène à des écart-types estimés très grands compte tenu que ces derniers sont issues de l'inverse de la matrice d'information de Fisher, elle-même dépendante de la dérivée seconde de la log-vraisemblance. En résumé, avec des données contenant la séparation, les variations des paramètres lors des itérations sont très grandes, les écart-types sont également élevés et la valeur de la fonction de la log-vraisemblance la plus haute s'approche de 0 pour la séparation complète et d'une valeur inférieure à 0 pour la séparation quasi-complète. Ces diverses constatations apportent des pistes pour détecter la séparation dans les données.

Le logiciel SAS dans sa procédure LOGISTIC se base sur certains points abordés ci-dessus pour détecter la séparation. Il recherche la séparation en calculant, au bout de la $8^{\text{ème}}$ itération seulement, la probabilité de la réponse observée pour chaque observation. Une valeur de 1 suppose une séparation complète, alors qu'une valeur inférieure, mais proche de 1, définit soit une séparation quasi-complète soit une valeur atypique. Dans le premier cas, le programme complète la recherche de séparation quasi-complète en analysant les variances des paramètres dans le cas d'observations standardisées, à savoir qu'elles ont été centrées puis réduites, afin de détecter une valeur supérieure à 5000. Si tel est le cas, le

programme arrête le processus et émet un avertissement signalant la non existence d'estimateurs MV. Il en est de même pour la séparation complète. L'avertissement conclut en disant que les valeurs se trouvant dans la sortie sont issues de la dernière itération et ce malgré la non existence de l'estimation MV. D'autres logiciels ne préviennent pas aussi clairement des situations de séparation, plus précisément celles de séparation quasi-complète. Par exemple, le programme TSP-GiveWin avertit l'utilisateur de la séparation complète en signalant que certaines variables prédisent parfaitement la variable réponse, alors qu'en ce qui concerne la séparation quasi-complète, un tel message n'apparaît pas. Il y a plus fréquemment un avertissement du type "échec à améliorer la fonction objective après x itérations". Ce signalement peut aussi bien correspondre à une situation de séparation quasi-complète qu'à toute autre situation. D'autres programmes n'informent ni de la séparation complète, ni quasi-complète. Compte tenu de la difficulté de détection de la séparation dans certains logiciels d'estimation, il est intéressant de compléter l'estimation MV en vérifiant les points suggérés ci-dessus ou en effectuant une analyse factorielle discriminante des données.

4.2 L'influence de la séparation sur les tests statistiques et les intervalles de confiance

L'estimation infinie des paramètres responsables de la séparation dans l'estimation MV ne permettant pas de calculer leur significativité par les tests t-statistiques ou de Wald, il faut donc trouver un autre moyen pour déterminer l'influence de ces derniers. Les tests statistiques du rapport de vraisemblance ne faisant appel qu'à la valeur de la log-vraisemblance et non aux valeurs d'estimation des paramètres et de la matrice de variance et de covariance asymptotique, sont donc applicables. Pour déterminer la significativité des paramètres, il suffit donc de comparer les log-vraisemblances du modèle complet et de celui auquel nous avons enlevé la variable testée.

La séparation des données empêche également le calcul des intervalles de confiance. En effet, ces dernières dépendent de la valeur estimée des paramètres et des écart-types. Il n'existe toutefois aucune autre alternative. Certaines solutions, décrites dans la suite de l'étude, aboutissent à des estimations des variables responsables de la séparation et ainsi permettent le calcul des intervalles de confiance. Nous tenons cependant à souligner que dans l'approche de la vraisemblance pénalisée, la manière de déterminer l'intervalle de confiance diffère de celle utilisée dans la méthode MV.

4.3 Solutions

L'approche du maximum de vraisemblance dans l'estimation des données n'est pas toujours appropriée pour déterminer les valeurs des paramètres inconnus du modèle. En effet, tel que nous venons de l'expliquer, en présence de séparation, la log-vraisemblance ne trouvant pas de maximum, fournit des valeurs infinies ou très élevées des paramètres avec pour conséquence une estimation peu fiable du modèle. Dans ces situations, les logiciels peuvent, soit ne pas fournir de valeurs aux paramètres provoquant la séparation, soit une valeur qui correspond, en réalité, à celle de la dernière itération du processus d'estimation. Le fait d'afficher une valeur n'est pas plus correcte que de n'en fournir aucune, étant donné que l'estimation du paramètre avec séparation n'est pas fiable. Par conséquent, il faut trouver d'autres moyens pour estimer les coefficients et les écart-types du modèle.

Quelques solutions relatées dans la littérature ont été proposées. Elles sont soit en rapport avec les paramètres ou les données :
– l'omission de la variable en cause
– le regroupement des modalités pour les variables explicatives qualitatives
– la modification des données
soit elles concernent d'autres méthodes d'estimation que celle de l'approche MV :
– l'inférence exacte
– l'approche bayésienne
– la vraisemblance pénalisée ou l'*a priori* de Jeffreys

Ces solutions sont décrites dans la suite. Dans le cadre des modèles d'estimation non linéaires, la littérature les décrit et les compare essentiellement pour la spécification logit. Par conséquent, les résultats fournis par la littérature en ce qui concerne les performances des solutions ne sont peut-être pas vérifiés dans notre spécification de l'option "zéro", qui combine des modèles probit de comparaison d'utilité et de seuil de réaction.
La méthode de vraisemblance pénalisée est essentiellement décrite pour les modèles logit à variable dépendante dichotomique. Bull, Mak et Greenwood (2002) étendent cette approche au logit multinomial. Néanmoins, les quelques logiciels dans lesquels la procédure est implémentée ne concernent que les modèles logit à variable dépendante dichotomique. Aucun programme à notre connaissance ne fournit l'estimation des paramètres des modèles multinomiaux logit et probit de cette méthodologie.

4.3.1 Modification des variables ou des données

Les solutions proposées dans ce paragraphe modifient la structure des variables ou des données. En effet, la plupart des ouvrages proposent, entre autres, de résoudre la non convergence du modèle en supprimant la ou les variables explicatives responsables de la séparation quasi-complète. Dans le cas de la séparation complète, cette procédure n'est pas applicable, étant donné que ce sont toutes les variables qui provoquent la non convergence du modèle. Il est donc dénué de sens de retirer toutes les variables explicatives du modèle.

Une autre option applicable aux variables explicatives qualitatives est le regroupement des modalités responsables de cette situation avec celles qui sont convergentes. Cette solution est envisageable à condition que la variable en question ait plus de deux modalités et qu'après la transformation, il reste au moins deux niveaux au régresseur.

Une modification appliquée directement sur les données peut également constituer un moyen pour supprimer les problèmes de convergence du modèle. Ce changement, proposé par Clogg et al. (1991), s'effectue en ajoutant des observations supplémentaires aux différentes catégories de données, afin de ne pas avoir des cellules vides. L'estimation des ces nouvelles données s'effectue avec la méthode du maximum de vraisemblance.

L'omission de régresseurs ou le regroupement des modalités ne sont pas des solutions adéquates, étant donné que les variables responsables de la séparation ont en réalité une grande influence sur la variable dépendante. Par conséquent, en supprimant leur effet du modèle, nous n'obtenons pas les résultats associés aux variables importantes du modèle. L'approche proposée par Clogg et al. (1991), quant à elle, semble être moins efficace, selon Heinze et Schemper (2002) et Galindo-Garre et al. (2004), que celles présentées dans la suite de ce chapitre.

4.3.2 L'inférence exacte

L'inférence conditionnelle exacte est une alternative au maximum de vraisemblance dans les échantillons contenant de la séparation. Cette méthode est essentiellement décrite dans la littérature pour des variables dépendantes binaires logit. Nous ne savons donc pas si les propriétés mentionnées dans la littérature sont vérifiées dans notre spécification de l'option "zéro". Toutefois, étant donné que cette approche fournit des estimations dans les situations de séparation, nous allons donc brièvement la décrire.

La régression logistique exacte fut d'abord proposée par Cox (1970). Elle n'était toutefois

pas réalisable informatiquement compte tenu de l'avancée des algorithmes à cette période. Dès que les algorithmes eurent progressé, la réalisation de cette dernière fut envisageable. Désormais, l'inférence exacte offre une alternative intéressante au maximum de vraisemblance avec de très bons résultats et peut être exécutée informatiquement.

L'inférence exacte des paramètres inconnus β $(r \times 1)$ découle de la distribution de permutation des statistiques exhaustives des variables testées conditionnellement aux valeurs observées des statistiques exhaustives des autres coefficients.
Ces statistiques exhaustives des paramètres β du modèle logit binaire sont définies de la manière suivante :

$$t_r = \sum_i y_i x_{ri} \qquad \text{avec} \quad i = 1, \ldots, n$$

Leur valeur observée est, dans le cas des variables explicatives qualitatives, égale au nombre d'observations tel que $y_i = 1$ et $x_i = 1$. Dans un tableau de contingence entre la variable indépendante et la variable dépendante, cette valeur observée équivaut à n_{11} :

	y		
	1	0	Total
x	n_{11}	n_{12}	$n_{1.}$
	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

Si le modèle contient une constante, la valeur observée de cette dernière est :

$$m = \sum_{i=1}^{n} y_i$$

ou simplement toutes les valeurs telles que $y_i = 1$, c'est-à-dire $n_{.1}$.

L'estimation des paramètres s'effectue en maximisant la distribution conditionnelle de la statistique exhaustive du paramètre testé sachant les statistiques exhaustives des autres coefficients. Par exemple, si notre modèle logit contient une constante α et une variable explicative qualitative dont le coefficient est égal à β_1, la distribution conditionnelle de la statistique exhaustive t_1 de β_1 sachant t_0 de α est notée en faisant référence à la probabilité conditionnelle $f(t_1/\beta_1) = P(T_1 = t_1 | T_0 = m)$. Cette distribution conditionnelle peut être obtenue à l'aide de logiciels, tel que LogXact et SAS (PROC LOGISTIC, option EXACT). Pour les variables créant la séparation, la valeur de la statistique exhaustive de ces dernières est égale à sa valeur minimum ou maximum de sa distribution conditionnelle. Dans cette situation, la vraisemblance conditionnelle n'est pas définie. L'estimation des paramètres de la séparation s'effectue, par conséquent, en utilisant la méthode de la

médiane. L'estimation par la médiane a l'avantage de ne pas être biaisée.

L'inférence exacte ne s'applique pas si une des variables explicatives du modèle est continue et si la taille de l'échantillon est grand.

4.3.3 L'approche bayésienne

L'approche bayésienne se base sur le principe des lois conditionnelles énoncé par Bayes. En effet, dans un modèle avec Y la variable dépendante, X la matrice des variables explicatives et β les paramètres inconnus, le principe de Bayes définit la loi conditionnelle de β de la manière suivante :

$$\text{loi de } (\beta|Y) = \frac{\text{loi de } (Y|\beta) \cdot \text{loi de } (\beta)}{\text{loi de } Y}$$

Formellement, nous pouvons écrire cette expression de la manière suivante :

$$f(\beta|Y) = \frac{f(Y|\beta)f(\beta)}{f(Y)}$$

L'utilisation de l'approche de Bayes dans un modèle de régression a pour but de déterminer la loi conditionnelle des paramètres β afin de pouvoir en déduire une estimation de ces dernières. Par conséquent, étant donné que la densité de Y ne dépend pas des paramètres inconnus, $f(\beta|Y)$ devient proportionnelle à l'expression suivante :

$$f(\beta|Y) \propto f(Y|\beta)f(\beta)$$

La méthode bayésienne déduit en réalité ces coefficients inconnus en les supposant aléatoires, alors que le modèle de régression les suppose fixe et détermine leurs valeurs en utilisant des estimateurs appropriés aux propriétés de la spécification. De plus, si la variable dépendante du modèle est qualitative et contient H modalités, la loi de Y sachant β est une probabilité étant donné que Y ne prend que des valeurs discrètes. Cette probabilité conditionnelle dans les modèles multinomiaux s'écrit de la manière suivante :

$$P(Y|\beta) = \prod_{i=1}^{n} \prod_{h=1}^{H} P_{hi}^{y_{hi}}$$

où $P_{hi} = P(y_i = h)$ et $y_i = \begin{cases} 1 & \text{si l'individu } i \text{ choisit l'alternative } h \\ 0 & \text{sinon} \end{cases}$

La loi de β sachant Y devient donc :

$$f(\beta|Y) \propto P(Y|\beta)f(\beta) \equiv \prod_{i=1}^{n}\prod_{h=1}^{H} P_{hi}^{y_{hi}} f(\beta)$$

Pour déterminer l'expression $f(\beta|Y)$, il faut postuler une loi pour β, dite loi *a priori*. En effet, en supposant une loi *a priori* pour les coefficients aléatoires β, nous pouvons ainsi définir la loi *a posteriori* de β sachant Y. Les lois *a priori* les plus fréquemment mentionnées et utilisées dans la littérature sont la loi normale et l'*a priori* de Jeffreys.

Supposons que le nombre de paramètres inconnus du modèle à n observations s'élève à r. Une loi *a priori* normale de ces r paramètres inconnus est une loi multivariée normale de la forme :

$$f(\beta) = (2\pi)^{-\frac{n}{2}}|\Sigma_\beta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta-\mu_\beta)'\Sigma_\beta^{-1}(\beta-\mu_\beta)\right\}$$

Cette expression suppose la connaissance du vecteur de la moyenne μ_β et de la matrice de variance-covariance Σ_β de β.

La loi *a posteriori* $f(\beta|Y)$ des modèles multinomiaux, compte tenu de l'*a priori* normal, devient proportionnelle à l'expression suivante :

$$f(\beta|Y) \propto \prod_{i=1}^{n}\prod_{h=1}^{H} P_{hi}^{y_{hi}} \times (2\pi)^{-\frac{n}{2}}|\Sigma_\beta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta-\mu_\beta)'\Sigma_\beta^{-1}(\beta-\mu_\beta)\right\}$$

L'expression de l'espérance μ_β et de la matrice de variance-covariance Σ_β dépend de la loi *a priori* des β postulée. Une suggestion indiquée par Congdon (2001) dans un des articles de Galindo-Garre et al. (2004) est "*qu'en absence d'attente a priori sur la direction ou la taille des effets des variables, des priori plats peuvent être approximés dans BUGS*[1] *en prenant les distributions normales univariées avec une moyenne de 0 et une grande variance*". Cependant, Galindo-Garre et al. (2004) ajoutent dans leur conclusion que cette procédure qui spécifie des variances élevées n'est pas correcte avec les échantillons petits. Dans notre application de l'option "zéro" du chapitre 5, nous avons toutefois certains renseignements sur la moyenne et la variance des paramètres inconnus. En effet, si nous considérons le modèle agrégé de notre spécification, ce dernier peut être une source d'information pour les paramètres de la loi normale *a priori*.

[1]Le projet BUGS (Bayesian inference Using Gibbs Sampling) fait appel aux méthodes des chaînes de Markov de Monte Carlo (MCMC) dans l'analyse bayésienne de modèles statistiques complexes et a conduit au logiciel Winbugs. http ://www.mrc-bsu.cam.ac.uk/bugs/

L'autre *a priori* appliqué dans l'estimation des paramètres inconnus est celui de Jeffreys, qui a l'avantage d'être invariant à une transformation des paramètres. Ce dernier suppose que la loi des coefficients β est proportionnelle à la matrice d'information de Fisher $I(\beta)$ de la manière suivante :

$$f(\beta) \propto |I(\beta)|^{\frac{1}{2}}$$

où $|I(\beta)|$ correspond au déterminant de la matrice de Fisher.

Cette matrice se définit comme suit :

$$I(\beta) = -E\left(\frac{\partial^2 \ln P(Y|\beta)}{(\partial\beta)^2}\right)$$

Dans les modèles à variable qualitative dépendante multinomiale, la structure de la fonction $\frac{\partial^2 \ln P(Y|\beta)}{(\partial\beta)^2}$ ne dépendant pas de Y, l'espérance associée à cette dernière peut donc être enlevée :

$$I(\beta) = -\frac{\partial^2 \ln P(Y|\beta)}{(\partial\beta)^2}$$

Avec l'*a priori* de Jeffreys appliqué aux modèles multinomiaux, la loi conditionnelle $f(\beta|Y)$ de β devient proportionnelle à :

$$f(\beta|Y) \propto \prod_{i=1}^{n}\prod_{h=1}^{H} P_{hi}^{y_{hi}} \times \left|-\frac{\partial^2 \ln P(Y|\beta)}{(\partial\beta)^2}\right|^{\frac{1}{2}}$$

L'estimation des coefficients inconnus du modèle multinomial à partir de l'approche bayésienne s'effectue soit par le calcul du mode, soit par celui de la moyenne, soit par celui de la médiane. Ces valeurs découlent évidemment de la loi *a posteriori*.

L'estimation du mode *a posteriori* s'effectue en utilisant le maximum de vraisemblance sur la loi *a posteriori*. En effet, étant donné que le mode correspond à la valeur la plus élevée, maximiser la loi *a posteriori* revient à chercher le mode de la fonction. Cependant, Galindo-Garre et al. (2004) font remarquer dans leur article que l'utilisation de l'algorithme de Newton-Raphson avec des lois *a priori* normales ou des *a priori* de Jeffreys n'est pas adapté. Ils proposent un algorithme Newton-Raphson modifié pour estimer les coefficients du modèle. Ce dernier est décrit dans leur article. Nous ne l'abordons pas dans notre étude, étant donné que ce dernier n'est pas implémenté dans un logiciel et il n'est par conséquent pas évident de l'utiliser dans une application. Pour estimer les paramètres inconnus, il faut également implémenter la matrice d'information de Fisher du modèle multinomial. La forme de cette dernière étant complexe, il n'est donc pas facile

de l'introduire dans un programme.

Le calcul de la moyenne *a posteriori* s'effectue, quant à lui, de la manière suivante :

$$E(\beta|Y) = \int_{-\infty}^{+\infty} \beta \times f(\beta|Y) d\beta$$

La solution de cette intégrale n'étant pas résolvable analytiquement, il faut donc utiliser d'autres moyens pour la résoudre, tels que les simulations de Monte Carlo. En effet, il suffit de simuler S valeurs β_s issues de la loi *a posteriori* de $(\beta|Y)$ et de calculer :

$$E(\beta|Y) = \frac{1}{S} \sum_{s=1}^{S} \beta_s$$

Cependant, lorsque la loi *a posteriori* n'est pas définie comme une loi connue malgré le fait que nous avons sa forme structurelle, la simulation des paramètres β_s n'est donc pas faisable, entraînant l'impossibilité de calculer la moyenne.
Une autre option permettant l'estimation de la moyenne est de résoudre de manière numérique l'intégrale la définissant. D'autres types de simulations de Monte Carlo permettent le calcul de ces intégrales. Ces méthodes sont décrites plus en détail dans la suite du travail.

L'estimation par la médiane *a posteriori* s'effectue, quant à elle, en résolvant l'équation suivante :

$$\frac{1}{2} = \int_{-\infty}^{\beta_M} f(\beta|Y) d\beta$$

Cette intégrale n'est à nouveau pas résovable analytiquement. L'utilisation des simulations de Monte Carlo peut apporter une solution à condition de définir la loi *a posteriori* par une loi connue. Si tel est le cas, il suffit de simuler s β, notés β_s, et de calculer le rang des s valeurs et de sélectionner la valeur de ce rang dans l'échantillon des s β simulés. Cette procédure de simulations de Monte Carlo ne peut pas fonctionner si la loi *a posteriori* est inconnue, malgré la connaissance de la forme structurelle.
Un autre moyen pour déterminer la médiane est de résoudre numériquement l'intégrale la spécifiant. D'autres méthodes Monte Carlo de résolution d'intégrales peuvent également être utilisées pour déterminer cette intégrale.

En ce qui concerne l'efficacité des estimateurs, une étude comparative entre le mode et la moyenne menée sur deux modèles spécifiques logit par Galindo-Garre et al. (2004) ont démontré que l'estimateur du mode a un biais et une erreur quadratique médiane

résiduelle plus petits que celui de la moyenne. Par conséquent, le mode semble être plus efficace que la moyenne. De plus, les auteurs constatent que l'approche bayésienne obtient de meilleurs résultats que la méthode du maximum de vraisemblance et ce, quelque soit la distribution *a priori* choisie. Ils ajoutent que l'implémentation dans des logiciels de la méthode de Jeffreys n'existant pas, elle est donc plus difficile à concevoir.

4.3.4 La vraisemblance pénalisée ou l'*a priori* de Jeffreys

La méthode illustrée dans cette sous-section a été d'abord proposée dans la littérature par Firth (1993), puis elle a été reprise par Heinze et Schemper (2002) dans le but de résoudre la séparation dans les modèles à variable dépendante qualitative, plus précisément les modèles logit. Cette approche est essentiellement décrite dans le cadre de modèles binaires logit. Par conséquent, l'application et les résultats décrits dans la littérature peuvent ne pas être vérifiés dans le cadre de modèles multinomiaux probit spécifiés dans notre approche de l'option "zéro".

La méthodologie, proposée par Firth (1993), permet en réalité de diminuer les biais liés aux estimations par le maximum de vraisemblance (MV). D'autres méthodes, telles que le jacknife et la réduction de l'estimateur MV par la soustraction du biais d'ordre n à l'estimateur MV (biais corrigé (BC)), ont la même propriété que l'approche de Firth (1993), à savoir qu'elles réduisent le biais MV. Leur différence se situe sur leur manière de procéder. En effet, les approches jacknife et BC agissent directement sur les paramètres estimés par MV, alors que la méthode de Firth (1993) modifie la fonction de vraisemblance pour diminuer le biais d'estimation MV. Nous allons décrire brièvement les méthodes jacknife et BC, puis nous définirons la procédure d'application de la méthode de Firth (1993).

La méthode BC réduit donc le biais d'ordre n du maximum de vraisemblance en remplaçant dans l'équation de ce dernier, les paramètres β par leurs valeurs estimées par le maximum de vraisemblance $\hat{\theta}_{MV}$ et en soustrayant cette équation à l'estimateur MV, β_{MV}. Pour appliquer cette approche, il faut connaître la forme du biais du maximum de vraisemblance. Elle est résumée dans l'article de Firth (1993) de la manière suivante :

$$b(\beta_{MV}) = \frac{b_1(\beta_{MV})}{n} + \frac{b_2(\beta_{MV})}{n^2} + \dots$$

Par conséquent, l'estimateur BC se formalise comme suit :

$$\hat{\beta}_{BC} = \hat{\beta}_{MV} - \underbrace{\frac{b_1(\hat{\beta}_{MV})}{n}}_{b_1(\hat{\beta}_{MV})^*}$$

La méthode du jacknife consiste à trouver la moyenne et la variance d'un nouvel échantillon d'observations basées sur l'estimateur du maximum de vraisemblance, $\hat{\beta}_{MV}$. Ce nouvel échantillon est construit à l'aide de l'équation suivante :

$$\hat{\beta}_{*i} = n\hat{\beta}_{MV} - (n-1)\hat{\beta}_{MV(-i)}$$

où $\hat{\beta}_{MV(-i)}$ correspond à l'estimateur du maximum de vraisemblance avec toutes les données sauf la $i^{\text{ème}}$. Pour obtenir le nouvel échantillon, il suffit d'enlever la première observation et d'estimer le maximum de vraisemblance pour obtenir l'observation $\hat{\beta}_{MV(-1)}$, de refaire ce procédé en enlevant la deuxième observation pour obtenir $\hat{\beta}_{MV(-2)}$, et ainsi de suite jusqu'à la $n^{\text{ème}}$ observation $\hat{\beta}_{MV(-n)}$.
Ces n observations représentent donc le nouvel échantillon de données auquel nous cherchons la moyenne et la variance comme suit :

$$\hat{\beta}_{jacknife} = \hat{\beta}_* = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{*i}$$

$$\hat{V}(\hat{\beta}_*) = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\beta}_{*i} - \hat{\beta}_*)^2$$

L'estimateur jacknife, dont le biais d'ordre n est enlevé, correspond à la moyenne de ce nouvel échantillon d'observations $\hat{\beta}_{*i}$.

Ces deux méthodes ont, toutefois, le désavantage d'exiger l'estimation par le maximum de vraisemblance des paramètres avant d'être appliquées. Or, dans le cas de la séparation, l'estimation ne convergeant pas, il n'est donc pas possible d'utiliser ces méthodes pour obtenir des estimateurs.

La méthode de Firth (1993), quant à elle, n'a pas cet inconvénient, étant donné qu'elle modifie la fonction du maximum de vraisemblance et non pas les estimateurs. Cette modification, développé par Firth (1993) et appelée méthode de la vraisemblance pénalisée, permet de réduire le biais d'ordre n de l'estimateur MV. Ce fut Heinze et Schemper (2002) qui découvrirent que la méthode de Firth (1993) pouvait également être utile pour obtenir des estimations finies dans les échantillons présentant la séparation. La vraisemblance pénalisée correspond en réalité à l'approche bayésienne avec, comme distribution *a priori* des paramètres, l'*a priori* de Jeffreys.

Si la fonction de vraisemblance d'un modèle quelconque s'écrit $L(\beta)$, la fonction de vrai-

semblance pénalisée, quant à elle, s'exprime de la manière suivante :

$$L^*(\beta|Y) = L(\beta|Y)|I(\beta)|^{\frac{1}{2}}$$

Pour obtenir l'estimation des paramètres inconnus, il suffit de maximiser le logarithme de la fonction de vraisemblance pénalisée, qui s'écrit comme suit :

$$\ln\left(L^*(\beta|Y)\right) = \ln\left(L(\beta|Y)\right) + \frac{1}{2}\ln\left(|I(\beta)|\right)$$

La maximisation donne l'équation suivante :

$$\frac{\partial \ln\left(L^*(\beta|Y)\right)}{\partial\beta} = \frac{\partial \ln\left(L(\beta|Y)\right)}{\partial\beta} + \frac{1}{2}\left\{|I(\beta)|^{-1}\left[\frac{\partial|I(\beta)|}{\partial\beta}\right]\right\} = 0$$

Une des propriétés de cette procédure d'estimation est le fait qu'elle assure l'unicité et l'existence d'une solution. La raison est due au fait que les fonctions $\log|I(\beta)|$ et $L(\beta|Y)$ sont strictement concaves et que $L(\beta|Y)$ a une borne supérieure, alors que $\log|I(\beta)|$ n'a pas de borne inférieure. Cette propriété est valable à condition que la matrice X soit de rang complet. Elle est importante dans le cas de petits échantillons présentant généralement la séparation. En effet, elle permet d'obtenir des estimateurs finis.

L'expression de la matrice d'information de Fisher, ainsi que le modèle de vraisemblance pénalisée dans le cas de spécification de variable dépendante qualitative, est essentiellement formulée dans la littérature pour les modèles logit à variable dépendante dichotomique. Par conséquent, cette formulation ne nous est pas utile pour notre spécification de l'option "zéro". Néanmoins, Bull, Mak et Greenwood (2002) traitent du cas multinomial logit dans leur article. En effet, ils étendent l'approche de Firth (1993) au cas multinomial logit en spécifiant la forme du modèle de vraisemblance pénalisée et la matrice d'information de Fisher. Cette dernière est formulée comme suit :

$$I(\beta) = (X'_M M X_M)$$

où $X_M = X' \otimes I_H$ et M une matrice bloc diagonale avec comme éléments $\{m_{ihl}\}$, $h = 1, \ldots, H$ alternatives, $l = 1, \ldots, L$ alternatives et $i = 1, \ldots, n$:

$$m_{ihl} = \begin{cases} \mathrm{P}(y_i = h)(1 - \mathrm{P}(y_i = h)) & h = l \\ -\mathrm{P}(y_i = h)\mathrm{P}(y_i = l) & \text{sinon} \end{cases}$$

La vraisemblance pénalisée appliquée aux modèles multinomiaux se définit de la manière

suivante :

$$\frac{\partial \ln (L^*(\beta|Y))}{\partial \beta} = \frac{\partial \ln (L(\beta|Y))}{\partial \beta} - I(\beta) b_1(\hat{\beta}_{MV})^* = 0$$

où $b_1(\hat{\beta}_{MV})^*$ correspond au biais asymptotique d'ordre n du maximum de vraisemblance.

Ce biais s'écrit dans le cas du modèle multinomial comme suit :

$$b_1(\hat{\beta}_{MV})^* = -\frac{1}{2} I(\beta)^{-1} \{X_M' Q (X_M \otimes X_M) vec(I(\beta)^{-1})\}$$

où
$Q = \sum_i E_i Q_i (E_i \otimes E_i)'$,
$E_i = e_i \otimes I_H$, e_i est un vecteur $n \times 1$ formé que de 0 sauf à la $i^{\text{ème}}$ ligne,
$Q_i = \sum_{hlr} q_{ihlr} \iota_h (\iota_l \otimes \iota_r)'$,

$$q_{ihlr} = \begin{cases} \mathrm{P}(y_i = h)(1 - \mathrm{P}(y_i = h))(1 - 2\mathrm{P}(y_i = h)) & h = l = r \\ 2\mathrm{P}(y_i = h)\mathrm{P}(y_i = l)\mathrm{P}(y_i = r) & h \neq l \neq r \\ -\mathrm{P}(y_i = h)(1 - 2\mathrm{P}(y_i = h))\mathrm{P}(y_i = r) & h = l \neq r \\ -\mathrm{P}(y_i = h)\mathrm{P}(y_i = l)(1 - 2\mathrm{P}(y_i = r)) & h \neq r \neq h \text{ ou } r = l \end{cases}$$

et ι_h est un vecteur $H \times 1$ avec que des 0 sauf sur la ligne h.

Cette description du maximum de vraisemblance pénalisée n'est cependant pas implémentée dans les logiciels statistiques. En effet, les programmes de calcul ne proposent que cette approche pour les modèles logit dichotomiques. Son application et son implémentation dans le cadre de modèles multinomiaux logit est difficile et est encore plus complexe dans le cadre de modèles multinomiaux probit, compte tenu de l'expression plus compliquée des probabilités.
Les estimations des applications logit multinomiales, décrites dans l'article de Bull, Mak et Greenwood (2002), ont été implémentées en exécutant le langage de programmation GAUSS. Pour cela, les auteurs utilisent des équations itératives modifiées définies dans leur article.

Les tests statistiques appliqués dans les estimations par maximum de vraisemblance ne sont pas adéquats dans le cas du maximum de vraisemblance pénalisée. En effet, les tests de Wald, les t-statistiques et les tests du score supposent la symétrie de la fonction de vraisemblance. Cependant, la vraisemblance pénalisée n'étant pas symétrique, il faut donc utiliser d'autres tests statistiques utiles pour vérifier la significativité des paramètres et calculer des intervalles de confiance. Le test du ratio de vraisemblance peut

être utilisé en remplaçant les éléments de l'équation par le logarithme de la vraisemblance pénalisée. Quant aux intervalles de confiance, dans le cas de fonction symétrique, l'utilisation des intervalles de confiance de Wald est adéquate, alors que dans le cas de fonction asymétrique, il faut utiliser le profil de la fonction du logarithme de vraisemblance pénalisée et déterminer l'intervalle dans le voisinage de la solution.

Conclusion

La notion de séparation dans le maximum de vraisemblance des petits échantillons est très commune et oblige, lors de l'estimation des paramètres du modèle, à contourner la méthodologie envisagée. Les méthodes contournant le maximum de vraisemblance ne sont pas très nombreuses et ne sont qu'implémentées dans le cas de modèles logit dichotomiques. L'approche de Firth (1993) semble être la méthodologie la plus adaptée dans les échantillons avec la séparation. Cependant, la littérature la compare généralement dans des situations où le modèle a une spécification logit et une variable dépendante binaire. Notre modèle de l'option "zéro" n'a aucune de ces caractéristiques, étant donné qu'il utilise une spécification probit et qu'il est multinomial.
L'inférence exacte est également spécifiée dans la littérature avec des modèles logit dichotomiques. L'implémentation dans les logiciels ne concerne que cette catégorie de modèle à variable dépendante qualitative.
L'approche bayésienne semble également être une bonne alternative au maximum de vraisemblance. Sa description ne se limite pas essentiellement aux spécifications logit binaires. Son application demande souvent la connaissance des méthodes de simulations de Monte Carlo.

La littérature montre la supériorité des approches d'inférence exacte, bayésienne et du maximum de vraisemblance pénalisée dans les petits échantillons vis-à-vis du maximum de vraisemblance. Dans les modèles dichotomiques logit, la méthode apportant de meilleurs résultats est la vraisemblance pénalisée. En ce qui concerne les modèles probit multinomiaux, nous ne pouvons pas affirmer la supériorité de cette approche, étant donné que la littérature ne l'aborde pas.

Chapitre 5

Application de l'option "zéro" sur le marché de la VoD

Introduction

Ce chapitre a pour but d'illustrer à l'aide d'un exemple concret le modèle de l'option "zéro" avec le refus et le conflit. La littérature n'ayant qu'introduit l'option "zéro" avec le refus, cette illustration permet donc de statuer de l'influence du conflit dans l'option "zéro". La réalisation d'une application sous-entend la connaissance du produit à analyser, des attributs importants et du mode de recueil des préférences, l'utilisation de plans d'expérience adaptés au mode de recueil, la mise en place d'un questionnaire, le choix d'une population cible et l'estimation du modèle. Ces étapes font en réalité partie intégrante de la méthode de l'analyse conjointe. C'est pourquoi, certains ouvrages la qualifient de méthode d'étude intégrée.

Le produit étudié dans notre application est la Vidéo à la Demande (Video on Demand (VoD)). La compréhension du fonctionnement de ce dernier est complexe et diffère d'un pays à l'autre. Par conséquent, une section traitant de la Vidéo à la demande est introduite dans la suite de notre étude. Notre mode de recueil des données est la comparaison par paire. La raison est que nous aimerions que le nombre d'alternatives devant être choisies par chaque individu ne soit pas trop élevé. Or, en introduisant le refus et le conflit dans notre spécification, nous supposons qu'indépendamment des produits, il y a déjà deux alternatives (le conflit et le refus) inclues dans le questionnaire. Par conséquent, ajouter deux produits au conflit et au refus revient à proposer 4 alternatives aux individus. Ces 4 alternatives sont le nombre minimum pouvant être évalué par les individus dans le cas de notre spécification de l'option "zéro". Les plans d'expérience sont utilisés dans notre application afin de sélectionner des paires de produits. Il faut donc appliquer un plan

adapté à une comparaison par paire. Le détail de ces plans se trouve dans la section 5.3.

Les étapes de la création du questionnaire sont également traitées dans ce chapitre. L'application du modèle de l'option "zéro" à notre exemple tient compte des modifications découlant de la non convergence ou de la séparation complète. Les résultats sont analysés plus loin.

5.1 La Vidéo à la Demande (VoD)

La sélection des attributs pertinents dans la création d'un site de Vidéo à la demande (Video on Demand (VoD)) a exigé des connaissances informatiques ciblées, telles que la notion de tripleplay, de streaming, de DRM, etc. Ces informations ont été récoltées à travers les quelques sites de Vidéo à la Demande existant en France (principalement Canalplay, TF1Vision, Arte VOD et Ina) et en Angleterre (principalement Homechoice) au début de notre étude [1]. Nous nous sommes également intéressés à la Vidéo à la Demande existant aux Etats-Unis, étant donné que ces derniers sont déjà bien en avance dans ce domaine. Il faut toutefois noter que leur système est différent de celui des français et des anglais (Homechoice) (prise téléphonique versus le cable) [2]. Cette différence n'influe cependant pas sur la plate-forme de VoD. De plus, étant donné que la Suisse ne bénéficiait pas encore de site de VoD au début de notre application, nous avons donc ciblé les téléspectateurs suisses. La TSR (Télévision Suisse Romande) propose sur son site la retransmission de ses émissions. Elle l'associe à un site de VoD. Par VoD, nous entendons dans ce travail les sites qui proposent également des programmes n'étant pas liés à la chaîne de diffusion.

D'autres appuis dans la compréhension de la VoD ont été utilisés, tels que les journaux spécifiques à l'informatique, à savoir l'Ordinateur Individuel (L'OI), P2PMAG, Micro Actuel etc. et l'encyclopédie collaborative Wikipédia qui donne des définitions sur les différents aspects techniques liés à ce type de produit. Wikipédia définit la Vidéo à la demande comme étant "*un système permettant aux utilisateurs de choisir et de regarder le contenu de la vidéo à travers un réseau comme une partie d'un système de télévision interactive*". L'avantage de la VoD est que le consommateur choisit le moment où il ai-

[1] Aujourd'hui, en France, ce type de site internet est bien évidemment en forte augmentation en raison, entre autres, de la venue sur la télévision de cette VoD à travers Free, Club Internet, etc. Cependant, ces derniers utilisent souvent les plate-formes des sites internet de VoD, tel que celle de Canalplay, par exemple.

[2] Les Américains ont également le système TV par ADSL. En effet, le journal P2PMag met en avant l'avancée de l'Amérique dans la VoD en citant le site Cinemanow qui comporte un catalogue de 7500 vidéos. Quant aux anglais, ils ont également la VoD par le câble. En France, au début de notre application, Noos et Numéricâble ne semblaient pas encore offrir ce type de service. Par conséquent, nous axons notre étude principalement sur les systèmes les plus populaires.

merait regarder un programme.

La VoD peut être visualisée de différentes manières. Il existe, en effet, le mode streaming, le téléchargement temporaire et le téléchargement définitif. Wikipédia définit le mode streaming de la manière suivante : "*Le média streaming est un média qui est consommé (écouté ou regardé) pendant qu'il est livré. Streaming est plus une propriété de système de livraison qu'un média lui-même*". En réalité, le programme n'est pas chargé sur le PC. Il reste sur le serveur [3] qui le fournit. A titre d'exemple, les sites les plus connus qui proposent uniquement la vidéo sous forme de streaming sont Youtube et DailyMotion (français). Sur ces sites, il est possible de visionner un média sans pour autant le télécharger. Ces médias sont stockés sur plusieurs serveurs. A l'opposé, le média issu du téléchargement temporaire se stocke sur le PC et devient illisible après un certain laps de temps limite (généralement 1 mois). L'inconvénient est qu'il faut prévoir au moins 4 Go sur son PC pour le stocker durant la période de location (30 jours et 24 heures, 48 heures ou 72 heures après la première lecture). L'avantage est que dès que le média est chargé, la connexion à internet pour le visionnage n'est plus utile. Par conséquent, vous pouvez garder la vidéo (au maximum 30 jours) et la regarder, par exemple, à l'extérieur avec un portable. Le média issu du téléchargement temporaire peut également être visionné de suite, sans pour autant attendre d'être entièrement stocké sur le PC. Quant au téléchargement définitif, il correspond à l'achat du programme. Contrairement au téléchargement temporaire, il restera charger sur le PC et il n'y aura donc pas de date d'échéance.

En France et en Angleterre (Homechoice), ces services sont fournis à travers une connection ADSL (une ligne téléphonique), alors qu'aux Etats-Unis, ils proviennent principalement du câble. En Suisse, ces deux types de fournisseurs sont disponibles (Cablecom pour le câble et Swisscom pour l'ADSL). En Angleterre, le système de VoD par le câble existe également (TV Telewest et NTL). Toutefois, le système par la ligne téléphonique BT (British Telecom) (2003) est plus ancien que celui par le câble (2005). Le but de la VoD par le câble était de concurrencer les fournisseurs par satellite. En effet, étant donné que la VoD par satellite est difficilement applicable, le câble présente donc un avantage évident.

Pour un bon fonctionnement du mode streaming, l'internaute doit avoir une bande passante de bonne qualité (au moins 2 Mbits/s). Aux Etats-Unis, le câble offre une grande

[3] La définition d'un serveur selon Wikipédia est la suivante : "*Un serveur informatique, appelé serveur lorsque le contexte s'y prête, est un ordinateur ou un programme informatique qui rend service aux ordinateurs et logiciels qui s'y connectent à travers un réseau informatique, les clients. Ce service peut consister à stocker des fichiers, transférer le courrier électronique, héberger un site Web, etc. Il est possible pour un ordinateur d'être client et serveur en même temps. Les termes client et serveur viennent du fait qu'un client est demandeur d'un service et qu'un serveur rend ce service*".

bande passante. En France, le débit Internet au début de notre étude était de plus de 2 Mbits/s. Il était donc suffisant pour recevoir la VoD en streaming. Maintenant, il a bien évidemment augmenté. Un des fournisseurs d'accès internet, Free, propose même en 2007 de la fibre optique à ses abonnés. Ce mode de connexion apporte de très haut débit allant jusqu'à 100 Mbit/s. Ce mode de connexion passe par le système FTTH, c'est-à-dire Fiber to the Home se traduisant par la fibre à la maison. L'avantage de cette technologie est présenté dans "L'Ordinateur Individuel" d'octobre 2006 : "*en développant son infrastructure FTTH, le FAI va constituer son propre réseau de communication, jusqu'à l'abonné, sans avoir à emprunter les câbles téléphoniques de France Télécom sur les derniers mètres. De ce fait, c'est donc un rôle de véritable opérateur que Free tiendra dans l'avenir*". De plus, avec le FTTH, le débit annoncé est constant. Un tel débit permet la télévision en haute définition HD, qui est supérieure à un DVD en terme de qualité. Cette dernière ne fonctionne qu'avec un téléviseur adapté à cette technologie, c'est-à-dire HD Ready.

Wikipédia définit la fibre optique de la manière suivante : "*une fibre optique est un fil transparent très fin qui a la propriété de conduire la lumière et sert dans les transmissions terrestres et océaniques de données. Elle a un débit d'informations nettement supérieur à celui des câbles coaxiaux et supporte un réseau "large bande" par lequel peuvent transiter aussi bien la télévision, le téléphone, la visioconférence ou les données informatiques*".

Quant à la FTTH, Wikipédia la décrit comme étant "*une technologie visant à remplacer les technologies xDSL en installant de la fibre optique jusque chez l'abonné. Cette technologie définit un type d'infrastructure de communication permettant l'accès à Internet et aux services associés à des débits de 10 Mbit/s à 1 Gbit/s dans chaque sens, soit des débits très supérieurs à ceux accessibles via la paire de cuivre téléphonique.*

Cette technologie est déjà utilisée en milieux urbains en Asie du Sud-Est et aux Etats-Unis, ainsi que dans quelques agglomérations européennes, dont Pau où le déploiement d'un réseau tout fibre optique est donc d'actualité.

L'infrastructure fibre donne une garantie de satisfaction en termes d'usages, permettant à divers membres d'un foyer de simultanément surfer, téléphoner, visionner un film en haute définition, et caetera."

En ce qui concerne le débit internet en Suisse, au début de notre application, il était trop faible, alors que désormais il est suffisant ($\sim$ 2 Mbits/s). Les Services Industriels de Genève (SIG) avait testé la fibre optique dans un des quartiers de Genève. Les habitants

bénéficiaient ainsi d'une connexion gratuite à haut débit. Cependant, ce projet ne s'est pas davantage développé.

Le mode de téléchargement temporaire impose également certaines contraintes. La taille du disque dur doit être suffisante afin de garder durant la location le média sur son ordinateur. Le P2PMag d'octobre 2006 donne les chiffres suivants : il faut compter 950 Mo la taille d'une vidéo d'1h30 et 1.30 Go celle de 2h. Ces critères peuvent différer selon le format de compression. De plus, le temps de téléchargement dépend de la taille du média, mais également de la vitesse de connexion. Là encore, P2PMag donne l'exemple d'une vidéo de 2h téléchargé en 1h50 avec un débit de 2 Mbit/s et en 20 minutes avec un débit de 10 Mbit/s.

5.1.1 Gestion numérique des droits (Digital Rights Management DRM)

Les DRM sont des "licences" intégrées aux contenus numériques comme les films, les chansons, etc. Ces DRM sont contenus dans les médias et gérés par les lecteurs vidéo du PC dans le but de vérifier ces "licences" et permettre le visionnage. Chaque lecteur a son propre format vidéo (wmv, DivX, avi, etc.). Par conséquent, selon ce format, le contenu d'un média ne peut pas être regardé sur n'importe quelle plate-forme vidéo. Le lecteur DivX permet le lancement d'un plus grand nombre de format que Windows Media Player. La gestion des DRM est liée au format. Par exemple, un programme ayant un format wmv ne pourra qu'être lu sur le lecteur de Windows Media Player. D'autres contraintes soulignées par le journal "L'Ordinateur individuel" de juillet-août 2006 montrent les limitations de ces DRM. En effet, ce dernier raconte que "*la plupart des sites de VoD protègent leurs programmes avec la DRM de Microsoft. Avec ce système, les vidéos doivent être lues avec les dernières versions du Lecteur Windows Media (9 ou 10), incompatibles avec Mac OS et Linux. Pour accéder à la VoD, il faut disposer d'un PC sous Windows (ou Windows XP, chez TF1Vision par exemple) et utiliser Internet Explorer 6*". Le magazine P2PMag d'octobre 2006 fait les mêmes constations, mais cite, quant à lui, l'obligation d'avoir la version 10 du lecteur Windows Media Player. Cette version "*assure la protection DRM des médias*". La version 9, quant à elle, "*ne prend pas en compte les wmv et DRM*". Cependant, contrairement à la version 9 qui est disponible pour Mac OS, la version 10 ne l'est pas et Microsoft n'a pas l'intention de la rendre disponible.
Les DRM liés au format Divx offrent plus de possibilités à l'internaute que ceux de Windows Media Player. En effet, ce dernier peut visionner le média sur un nombre fixé de machine ou le regarder sur un DVD autorisant la lecture Divx étant donné qu'il est autorisé à le graver.

Finalement, ces DRM orientent le fonctionnement du média sur un PC. Ces protections peuvent avoir une multitude de fonctionnalités. Par exemple, les sites américains empêchent par leur DRM le visionnage d'un programme en dehors de leur pays. Il en est de même pour la France et ses sites de VoD. Les DRM servent également à ne pas reproduire le média, à conserver le contenu durant une certaine période, à limiter le nombre de copies personnelles utiles pour son baladeur ou pour un autre ordinateur, etc.

Le fonctionnement de ces DRM est expliqué par Wikipédia de la manière suivante : "*Une architecture de gestion des droits numériques est basée sur le chiffrement des oeuvres. Seul un matériel ayant la clé de déchiffrement est capable de consulter l'oeuvre*". Cependant, il existe de nombreux logiciels permettant de contourner ces DRM en déchiffrant ces clés. C'est pourquoi, les fabriquants renouvellent ou mettent à jour leur protection. Ces protections sont également soutenus par des lois, telles que DMCA aux Etats-Unis et DAVDSI en France. Ces lois sont actualisées regulièrement. Toutefois, ces DRM créent plusieurs problèmes. En effet, certains contenus légaux peuvent ne pas fonctionner de manière optimale sur un PC. Par exemple, certains CD audio protégés ne peuvent qu'être lus sur un lecteur CD et pas sur un PC qui pourrait les copier. Windows Vista, qui a renforcé ces protections afin de satisfaire les artistes, impose certaines conditions pour obtenir un média HD en lecture optimale (P2PMAG). Wikipédia fait également remarqué que les fichiers protégés par des DRM ne peuvent pas être revendus car ils sont généralement lisibles que sur un PC. Or, lorsque nous achetons un média dans le commerce, nous pouvons le revendre. Wikipédia conclut en disant qu' "*en pratique, le DRM correspond davantage à une location ou à un droit d'usage provisoire techniquement surveillé, qu'à une vente*".

5.1.2 Intérêt de l'analyse de la Vidéo à la Demande

Au début de notre étude (début 2006), les entreprises, telles que Canalsat, TPS, Free, Aol, Alice, Neuf etc., commençaient à proposer partout en France les programmes TV par ADSL. En effet, le débit internet des français était suffisamment élevé pour proposer le tripleplay, à savoir la téléphonie sur IP qui correspond à la voix sur IP (VoIP), internet et la télévision par l'ADSL à un prix unique, au lieu du dualplay, téléphone et internet à un prix unique. La téléphonie sur IP permet de diminuer considérablement les prix, étant donné que les appels passent par internet et qu'ils n'ont donc plus besoin de France Télécom. Certains opérateurs proposaient des appels illimités vers les postes fixes français. Selon le journal suisse "Le Temps" de septembre 2006, cette politique s'émancipe vers les appels internationaux "*avec la gratuité des appels vers 23 pays pour AOL, 28 pour Free, 32 pour Neuf Télécom et 40 pour Club-Internet*". La téléphonie sur IP, la télévision

par l'ADSL et internet (tripleplay) présentent tout de même certains inconvénients. Ces inconvénients sont dans le cas d'internet, la quantité de pertes de paquets, la latence et la disponibilité du réseau, dans celui de la télévision, le nombre de gel d'images et sa qualité et dans celui de la téléphonie, la disponibilité du service et la qualité vocale. Etant donné que la disponibilité d'internet n'est pas de 100%, elle peut donc amener à des pannes du service qui sont généralement chiffrées de manière mensuelle. La télévision par ADSL requiert une connexion minimum de 4 Mbit/s.

En Suisse, la téléphonie sur IP ne semble pas faire partie des offres des opérateurs. En effet, tel qu'un des articles du Temps de septembre 2006 le dit, *Skype a apparemment de beaux jours devant lui. Swisscom, Sunrise et d'autres opérateurs proposent bien des offres de téléphonie via internet, mais elles sont plutôt compliquées et chères*". Skype est un logiciel de téléphonie sur IP. Il en est de même pour Wengo. Ces derniers peuvent être utilisés aussi bien pour des appels allant d'un ordinateur à l'autre, que pour des appels ordinaires sur des fixes ou des mobiles. Ils sont également fait pour recevoir des appels provenant de types de réceptions décrites ci-dessus. Pour cela, Skype ou Wengo fournit un numéro au PC qui coûtait en novembre 2006, selon Micro Actuel, 30 euros par an pour Skype et 15 euros pour Wengo. En plus de l'attribution payante d'un numéro, les appels avaient également un coût, à savoir pour un mobile ils étaient de 0.189 euros chez Skype, respectivement 0.15 euros chez Wengo, et sur un fixe ils étaient gratuits jusqu'à janvier 2007 chez Skype et de 0.01 euros chez Wengo. L'offre gratuite de Skype avait pour but d'augmenter son réseau de clients. Par contre, les appels d'un PC à un autre étaient, quant à eux, gratuits.
Quant à Cablecom, contrairement à ses concurrents, il semble s'intéresser à la téléphonie par le câble. Ce dernier a une offre de service téléphone et internet bon marché par le câble, s'affranchissant ainsi de Swisscom.

Une autre technologie surpassant le tripleplay est le quadriplay. Le principe est simple. Il fonctionne avec des mobiles comportant le système de communication GSM et Wi-Fi, GSM étant le réseau actuel. A proximité d'une borne Wi-Fi, la facture revient au FAI. Dans le cas contraire, le téléphone est en mode GSM. Cette offre dépendait de la conception de ce type de portable. Selon "L'OI" de septembre 2006, Neuf a créé un mobile avec la technologie décrite ci-dessus, appelé le Twin, autorisant ainsi le quadriplay. Ce type de service fonctionne de la même manière que les fixes en offrant les communications illimitées gratuites depuis un portable. Le premier qui a proposé cette offre quadriplay a été Neuf Télécom, suivi d'Orange, avec son système Unik. Ainsi, lorsque le mobile est en mode Wi-Fi, la facture revient au FAI. Ce système permet donc de diminuer les coûts de communication et d'avoir une meilleure vitesse d'envoi de médias ou d'images et d'accès

à internet. Ce système est encore plus intéressant dans les villes où il y a désormais des bornes Wi-Fi gratuites pour les utilisateurs. En effet, MicroActuel de septembre 2007 titre "*à Paris, le net c'est gratuit*". La capitale française équipe ses parcs, ses jardins, ses musées, ses bibliothèques, etc. de 400 bornes Wi-Fi gratuites. Dans le magazine "L'Ordinateur Individuel" d'octobre 2006, il est question "*des accès Wi-Fi gratuits*" existant aux Etats-Unis et en Angleterre. En effet, à Mountain View aux Etats-Unis, les bornes de Wi-Fi gratuites sont le fruit de Google (380). En Angleterre, des bornes (200) ont été installées par la société Telabria.
Les bornes ne se limitent pas qu'à celles qui sont gratuites. En effet, il y a également les payantes. En novembre 2006, selon Micro Actuel, il y avait 20 000 hotspots en France. Les payantes sont, par exemple, les hôtels, les cafés, les restaurants, etc. Quoique certains les offrent gratuitement.

Le tripleplay, proposé début 2006 en France, offrait, pour les développements futurs, la possibilité de mettre en place un système de Vidéo à la Demande [4]. Cette dernière amène les entreprises à réfléchir sur une plate-forme pertinente. L'intérêt de l'étude de la VoD était donc de permettre à ces entreprises de connaître les préférences des consommateurs sur les composants importants d'une plate-forme de VoD. Un autre type de support, le téléphone portable, peut également à l'avenir être utilisé pour ce type de service. En effet, les mobiles étant de plus en plus performants, les opérateurs offrent déjà la télévision en direct sur le portable, une sélection de vidéos à télécharger (VoD), l'achat de musiques, etc. Par exemple, en terme de VoD, Orange offre un catalogue de 2500 vidéos.
Un autre intérêt est, qu'étant donné qu'en octobre 2006 chaque site n'offrait pas les mêmes films, tel que le souligne le magazine Micro Actuel, "*chacun ayant des accords différents avec les majors*" et qu'il fallait donc naviguer entre ces sites de VoD pour obtenir toutes les vidéos récentes, une telle étude pouvait montrer l'importance d'un catalogue ayant toutes les nouveautés.

A ce moment, il n'y avait en Suisse pas toutes ces technologies. Les raisons de ces différences s'expliquent par le fait que "*le "dernier kilomètre" n'est effectivement pas ouvert à la concurrence*" (Le Temps de septembre 2006). Par conséquent, l'opérateur Swisscom ne souffrant pas de la concurrence des autres opérateurs ne s'est pas pressé pour faire évoluer le système. Cependant, à travers le câble, Cablecom a tout de même influencer Swisscom. En effet, cela se confirme par les propos d'un des article du Temps de septembre 2006 qui disait que "*mi-janvier 2006, Swisscom annonçait l'augmentation des débits de 600 kbit/s à 2000 kbit/s, toujours pour 49 francs par mois. Cablecom s'alignait*

[4]Cette mise en place n'était pas possible lorsque les programmes venaient par le satellite. En effet, nous avons vu que proposer de la VoD par le satellite était difficilement applicable.

quelques semaines plus tard avec une offre de 3000 kbit/s à 45 francs". Dès lors, il n'y a plus eu de nouvelles offres.

Par conséquent, au début de notre étude, la Suisse était bien plus en retard technologiquement parlant que la France. Compte tenu de la durée d'une telle étude, il semblait évident qu'à la fin des analyses, les Français proposeraient la VoD à leurs clients. En réalité, la VoD était déjà bien implémentée en France. Il existait déjà un site de VoD, netciné.com, à ce moment en France. Ce site a été créé en 2000 et devient en 2006 Canalplay du groupe Canal+. La Fnac proposait le téléchargement complet de films, ce qui ne correspondait pas à de la VoD, mais à de l'achat en ligne. En ce qui concernait les Suisses, la VoD ne semblait pas être d'actualité. C'est pourquoi, l'analyse s'est portée sur les consommateurs suisses. Néanmoins, les résultats de notre étude peuvent être également intéressants pour les consommateurs français.

Nous tenons, cependant, à souligner qu'en Suisse, la VoD semble être plus avancée à travers le téléviseur qu'à travers internet. En effet, en 2007, Swisscom offre la télévision via internet avec un service de VoD. Par contre, Cablecom ne montre aucune concurrence dans ce domaine. Sur internet, il n'y a que la TSR qui diffuse ses émissions. Toutefois, tel que nous l'avons mentionné précédemment, nous ne considérons pas ce type de service comme étant de la VoD, étant donné que le site ne propose que ses propres émissions.

Quant aux Etats-Unis, ces derniers étaient bien avancés sur la VoD. En effet, la VoD passait déjà par le câble. De plus, il existait également la VoD par internet. Un des plus connus est le site MovieLink. Wikipédia le définit comme suit : "*Movielink est une joint-venture (une entreprise commune ou coentreprise) entre 5 des 7 grands studios de cinéma hollywoodiens : Metro-Goldwyn-Mayer, Paramount Pictures Corporation, Sony Pictures Entertainment Inc., Universal Pictures et Warner Bros. Entertainment.*
Le service Movielink a été lancé en novembre 2002 aux Etats-Unis d'Amérique. Il permet le téléchargement par Internet de films produits par les studios partenaires. Chaque studio décide de sa politique tarifaire et de la date de disponibilité de ses films."
Au début de notre étude, on lisait dans LExpansion.com que Walt Disney venait également de permettre la diffusion de ses films par le site MovieLink. Il avait essayé huit mois auparavant de créer une coentreprise avec la 20th Century Fox, ce qui avait échoué. Le seul grand studio hollywoodien à ne pas diffuser ses films sur MovieLink était donc la 20th Century Fox.

5.1.3 Présentation de certains sites de VoD

En juillet-août 2006, le magazine "L'Ordinateur Individuel" consacrait un article sur les sites de VoD. Il en était de même pour le magazine Micro Actuel d'octobre 2006. Ces derniers comparaient plusieurs sites en y énumérant leurs avantages et leurs désavantages. De plus, il nous faisait part des limites liées à ces sites. Premièrement, il faut avoir un ordinateur pour visionner les programmes de ces sites. Pour obtenir les médias sur le téléviseur, il suffit de raccorder un câble supplémentaire adapté. Le Micro Actuel rajoutait que cet ordinateur doit être suffisamment puissant (2Ghz) avec un espace libre sur le disque dur de 2 à 5 Go, permettant ainsi de stocker les programmes téléchargés, et qu'il doit tourner sur Windows XP Media Center ou libellé Intel Viiv, avoir au moins 256 Mo de mémoire vive, une carte graphique de 64 Mo, une carte son et des hauts-parleurs 5.1 pour un rendu parfait, un téléviseur ou un écran d'ordinateur supérieur à 17". Deuxièmement, certains sites proposent leurs programmes en streaming, ce qui oblige l'internaute à les visionner de suite. A l'opposé, le téléchargement temporaire autorise le consommateur à garder durant 30 jours le film et 24 ou 48 heures après le premier lancement. L'achat du film correspond, quant à lui, au téléchargement définitif. Toutefois, le journal ne fait pas remarquer que l'internaute ne peut pas le revendre comme c'est le cas lorsque nous achetons un média dans le commerce. De plus, la copie sur un baladeur n'est pas permise par tous les sites. Troisièmement, il y a les limites liées au débit internet. En effet, un programme en streaming oblige le consommateur à avoir un débit d'au moins 2 Mbit/s afin que la qualité de l'image soit correcte. Pour le téléchargement, plus le débit est élévé, plus le film est rapidement téléchargé. Par exemple, avec un débit de 512 kbit/s, le programme sera chargé sur l'ordinateur en 3 ou 4 heures. Quatrièmement, la VoD impose au minimum la version 6 d'Internet Explorer et la version 9 de Windows Media Player, avec une préférence pour la version 10. Finalement, le nombre de médias proposés par les sites de VoD n'est pas très nombreux. De plus, ces programmes ne sont pas aussi récents que ceux proposés dans les vidéo clubs. Par conséquent, un consommateur désirant voir un film qui vient de sortir en DVD, devra attendre plus longtemps que s'il va au vidéo club, ce qui n'est pas le cas aux Etats-Unis. De plus, les tarifs sont très élévés. Une autre limite, qui n'est pas abordée par ces magazines, est la pauvreté vis-à-vis de la variété de programmes, contrairement aux vidéos clubs qui offrent tous les genres, tels que les films, les séries, les documentaires, les spectacles, etc.

Les sites analysé par "L'Ordinateur Individuel" en juillet et août 2006 et par Micro Actuel d'octobre 2006 sont Canalplay, Ina, ArteVoD, TF1Vision, Vodeo, VirginMega, France2 et France3, VoD.Wanadoo, Imineo, Cinezime, France 5 Vidéos. "L'Ordinateur Individuel" présentait un tableau abordant les caractéristiques des 6 sites de VoD les plus importants.

Ces caractéristiques ont bien évidemment évoluées depuis 2006. La raison qui nous amène à les présenter est que notre étude se base sur ces données. En effet, nous avons élaboré notre questionnaire au début 2007 et nous avons choisi nos attributs en fonction des informations présentées en 2006. Par exemple, en 2006, Canalplay offrait un catalogue de 1000 programmes et TF1Vision de 500. Maintenant, le catalogue de Canalplay s'élève à 3000 vidéos et celui de TF1Vision à 2734. De plus, il contient des films plus récents.
Le tableau des caractéristiques des sites de VoD issu de "L'Ordinateur Individuel" de juillet-août 2006 est le suivant :

	ARTE VOD	CANALPLAY	INA	TF1 VISION	VIRGINMEGA	VODEO
	www.artevod.com	www.canalplay.com	www.ina.fr	www.tfivision.fr	www.virginmega.fr	www.vodeo.tv
Contenu	Documentaires, émissions de la chaîne, films	Films, émissions pour enfants, mangas, films X	Emissions de télévision et radio, séries TV	Films, séries et émissions TV, documentaires, jeunesse, films X	Films, jeunesse, documentaires, séries TV, films X	Documentaires, reportages, fictions
Nombre de programmes	400	1000	100000	500	800	2000
Durée / prix de la location sans téléchargement (streaming)	Non disponible	Non disponible	Illimitée / gratuite	24 heures / de 2 à 5 euros (films X: 10 euros)	Non disponible	48 heures / 1 à 4 euros
Durée / prix de la location avec téléchargement	48 heures / de 1 à 4 euros	24 heures / de 1,50 à 5 euros (films X: 7 euros)	48 heures / de 1 à 3 euros	24 heures / de 2 à 5 euros (films X: 10 euros)	48 heures / de 0,50 à 5 euros (films X: 8 euros)	Non disponible
Prix d'achat	Non disponible	Non disponible	De 1 à 12 euros	De 10 à 15 euros	De 1 à 16 euros (films X: 25 euros)	De 2 à 8 euros
Débit vidéo	1500 Kbit/s	360, 600, 1500 ou 2500 (HR) Kbit/s	380 ou 1130 Kbit/s	1500 Kbit/s	1500 Kbit/s	350, 750 ou 1500 Kbit/s
Mode de paiement	CB	CB	CB, Internet plus	CB	CB, facture FT, cartes prépayées, SMS, Audiotel	CB, Paypal

D'autres magazines sorties autour de la même période présentaient le même type de résumé. Cependant, les informations pouvaient différer. C'est pourquoi, nous en montrerons plusieurs. Le P2PMag d'octobre 2006, par exemple, donne un catalogue de films du site Canalplay de 700, alors que dans "L'Ordinateur Individuel" il est de 1000.

Nom du Service	Titres	Catalogue	Principaux opérateurs VOD en France: Tarifs	Mode	Commentaires	Site Web
Arte VOD	400	Documentaires, fictions, magazines, films	0,99 à 4,99 €	Location 48h, 30 jours pour visualiser	Programmes en qualité DVD	www.artevod.com
Canal Play	700	Films	3,99 à 14,99 €	Achat ou Location 24h, 30 jours pour visualiser	Composant WMP DRM requis	www.canalplay.com
Cinézime	100	Films	0,50 à 16 €	Achat ou Location 48h, 30 jours pour visualiser	Composant WMP DRM requis	www.cinezime.fr
easyVOD	50	Films, séries, documentaires, jeunesse, théâtre	3,99 à 4,99 €	Location 48h, 30 jours pour visualiser	Lecteur multimédia sécurisé	www.easyvod.fr
Editions Montparnasse	534	Films, documentaires	6 € /location 10 € /vente	Achat ou Location 72 h, 30 jours pour visualiser	Une gravure possible via WMP	www.editionsmontparnasse.fr
Imineo	1500	Éducatifs, documentaires, animations, sports, films	2,99 à 16,99 €	Achat en téléchargement	Prix selon qualité VHS ou DVD	www.imineo.com
M6 HD	182	Films, fictions, séries, émissions	0,56 à 3,99 €	Location 48h, 30 jours pour visualiser	Logiciel M6 Vidéo nécessaire	www.m6video.m6.fr
Paris Première VOD	50	Émissions, théâtre, séries	Gratuit jusqu'en décembre 2006 puis 0,29 à 3,99 €	Streaming, Location	Logiciel ParisPremiereVOD	www.parispremierevod.fr
Télésavoirs	400	Documentaires scientifiques, éducatifs	1,80 €	Streaming	Flashplayer et Realplayer requis	www.telesavoirs.com
TF1 Vision	300	Séries, émissions, films, documentaires, fictions, spectacles, sports	1,99 à 4,99 €	Streaming, Location 24h, 30 jours pour visualiser	Streaming et location au même prix	www.tf1vision.com
Toondra	50	Courts-métrages d'animation	0,99 à 1,50 €	Achat en téléchargement	Lisible sans limitation	www.toondra.com
TPSVOD	70	Films, spectacles, émissions	3,99 à 4,99 €	Streaming, 24h, 30 jours pour visualiser	Offre découverte 0,49 € premier achat	www.tpsvod.fr
Virginmega	750	Films, clips, séries, sports, documentaires, animations, spectacles	0,99 à 4,99 €	Location 48h	Composant WMP DRM requis	www.virginmega.fr
Vodeo.tv	2000	Documentaires	2,99 à 5,99 €	Streaming, Location, Achat DVD, 30 jours pour visualiser	Prix en fonction du mode	www.vodeo.tv

L'évolution de ces sites peut être faite à l'aide du tableau comparatif proposé par "L'OI" de juin 2007 :

	Artevod	Canalplay	Ina	France tvod	Universciné	VirginMega	Vodeo
Adresse	www.artevod .com	www.canal play.com	www.ina.fr/ archivespourtous	www.france tvod.fr	www.univers cine.com	www.virgin mega.fr	www.vodeo.tv
Genres proposés	Films d'auteurs et émissions diffusés sur Arte	Films, documentaires, dessins animés et séries	Emissions de tele et de radio réalisées entre 1920 et 2006	Emissions de France Télévisions, films et spectacles grand public	Films d'auteurs contemporains	Films, documentaires, dessins animés et séries	Reportages et documentaires d'une sélection de chaînes télé
Nombre de programmes proposés	600	2 000	100 000	1 000	200	2 000	1 400
Présentation (bande-annonce ou extrait) / fiche descriptive / avis des internautes	Oui / oui / non	Oui / oui / oui	Oui / oui / non	Non / oui / non	Oui / oui / non	Oui / oui / non	Oui / oui / oui
Facilité d'emploi (recherche, paiement lecture)	Bien	Très bien	Bien	Médiocre	Moyenne	Très bien	Très bien
Films en VO sous-titrée	Oui	Rares	Non	Non	Oui	Rares	Non
Prix à la location	0,99 à 4,99€ les 48 heures	1,49 à 4,99€ les 48 heures	1 à 2€ les 48 heures	0,99 à 3€ les 24 heures	4,99€ les 48 heures	0,99 à 4,99€ les 48 heures	1 à 3,99€ les 48 heures
Achat des programmes	Oui	Oui	Oui	Oui	Non	Oui	Oui
Possibilité de regarder les programmes sur un baladeur vidéo	Seulement à l'achat	A l'achat et à la location, sur les baladeurs Archos	Seulement à l'achat sur les baladeurs certifiés DivX	Non	Non	Seulement à l'achat sur les baladeurs compatibles WMV DRM	A l'achat et à la location, sur les baladeurs Archos
Formules prépayées	Non	Forfait Cinéma à 15,99€, Enfants à 14,99€	Forfaits à 10, 20 ou 50€ donnant droit à des réductions	Non	Forfaits à 18,99, 27,99 et 49,99€	Non	Forfaits à 19, 29 et 49€, donnant droit à des réductions

Nous allons maintenant analyser ces différents sites. Un des défaut du site de VoD de Canalplay est qu'il a son propre logiciel de visionnage des programmes. Par conséquent, il faut télécharger ce logiciel (gratuit), ce qui rajoute à la liste du nombre de lecteurs de notre ordinateur. L'avantage de cette installation est que le lecteur est compatible avec tous les systèmes d'exploitation. Basé sur "L'Ordinateur Individuel" de juillet et août 2006, nous savons également que Canalplay offre certaines options non incluses dans les autres sites de VoD, telles que le contrôle parental ou le chapitrage comme sur un DVD. Un autre avantage cité par "L'Ordinateur Individuel" de juillet-août 2006 est que "*selon le débit de sa connexion, l'internaute a le choix entre divers niveaux de qualité d'image. Ainsi, l'accès au service reste possible avec une connexion à haut débit limitée tandis que les plus favorisés bénéficient d'une image en "haute résolution" (encodée à 2,5 Mbit/s) avec son multicanal (5.1)*". En 2006, le journal déplorait la durée de location de 24 heures. Désormais, elle est de 48 heures. A l'image des sites de téléchargement de musique, le transfert vers un baladeur vidéo Archos est autorisé par Canalplay. Cette option permet à l'utilisateur de regarder une vidéo lorsqu'il est en déplacement ou simplement de la visualiser sur l'écran de sa télévision à l'aide de la station DVR de l'Archos. Les prix de location n'ont pas changé entre 2006 et 2007. La location est entre 3.99 euros et 4.99 euros selon que ce soit une nouveauté ou selon son succès en salle. Dans le site de Canalplay, il n'y avait pas d'achat de programmes par téléchargement définitif. Désormais, il est possible d'acheter certains programmes à un prix compris entre 9.99 et 17.99 euros selon la

date de sortie et le succès dans les salles, ainsi que de les graver sur un support afin de les visionner sur le téléviseur. De plus, en 2006, le site était plus axé sur la VoD de films. Désormais, il y a aussi des séries TV. Un autre avantage récent est que malgré le fait que Canalplay n'a pas certaines nouveautés, nous pouvons tout de même voir les bandes annonces de ces dernières, à condition d'installer leur lecteur. De plus, la présentation du site est claire et agréable. Afin d'accéder à la location, il faut être inscrit en donnant son adresse e-mail et un mot de passe. Ce principe semble être commun à la plupart des sites de VoD.

Le site de l'Ina (Institut national de l'audiovisuel) offre, quant à lui, des programmes payants et gratuits en petit format. En effet, certaines anciennes émissions sont gratuites et peuvent être visionnées en petit format sur le site par la technologie streaming. Selon "L'Ordinateur Individuel" de juillet-août 2006, il s'agit de "*80% des 100'000 émissions disponibles*". La lecture des émissions en streaming s'effectue à l'aide du lecteur QuickTime de Apple. Ce logiciel a l'avantage de fonctionner avec tous les systèmes d'exploitation. Les émissions visibles en streaming deviennent payantes en grand écran. Certains programmes en streaming ne sont que des extraits, la diffusion dans sa totalité et en grand écran étant payante. Quant à l'amélioration au niveau du catalogue, elle n'est bien évidemment pas aussi flagrante que pour d'autres sites compte tenu de sa grande taille en 2006.
Le genre de programmes diffusés par l'Ina diffère de ceux de Canalplay, TF1Vision, etc. En effet, les vidéos de l'Ina sont des feuilletons, des séries, des grands entretiens, des discours fondateurs, des journaux télévisés, des émissions culturelles ou d'information, des événements sportifs ou des scènes de vie quotidienne, etc., alors que les autres sont essentiellement des films et des séries TV. En 2006, l'Ina avait déjà une durée de location de 48 heures. Elle offrait également la possibilité d'acheter ses programmes. Le format de fichier acheté est en DivX. D'ailleurs, l'Ina est le seul site à offrir ce type de format. De plus, il est compatible avec Mac OS et Linux. Comme le soulignait "L'Ordinateur Individuel" de juillet-août 2006, "*le petit plus est les programmes achetés peuvent être lus sur tout baladeur ou platine DVD certifiée DivX, après transfert ou gravure*". Un des inconvénients de ce site est l'installation de contrôle Active X à chaque lancement de vidéos en streaming.
Quant aux tarifs, ils évoluent en fonction de la durée des programmes. Pour les émissions d'une durée de moins de 10 minutes, ils sont de 1 euro lorsqu'il s'agit d'acualité et d'information et de 1,50 euros pour les divertissements, les variétés, la fiction, le sport et le cinéma. Les programmes compris entre 10 et 30 minutes sont au prix de 3 euros pour l'actualité, l'information, les documentaires et les magazines et 4 euros pour les autres. Et les vidéos d'une durée de plus de 30 minutes, telles que l'actualité, l'information, les documentaires et les magazines, coûtent 4 euros alors que les divertissements, les variétés,

la fiction, le sport et le cinéma sont à partir de 6 euros. Ces prix n'ont pas changé en 2007. Selon "L'Ordinateur Individuel", il existait également "*des formules prépayées à tarifs avantageux*" en 2006.

TF1Vision offrait, au début de notre étude, un nombre diversifié de catégories de programmes. En effet, en plus des films, il proposait des séries TV, des feuilletons et des documentaires de sa chaîne TF1, des spectacles et des concerts. En 2006, son catalogue représentait la moitié (500) de celui de Canalplay. Désormais, il a bien augmenté et se rapproche de Canalplay (2734). Le site est assez clair pour les recherches. Comme CanalPlay et grâce aux DRM, le site ne peut qu'être visionné depuis la France. Toutefois, en 2006, cela était également le cas pour les démos ou bandes annonces. Au début de notre étude, TF1Vision mettait déjà à disposition tous les types de téléchargement, c'est-à-dire le streaming, le teléchargement temporaire et définitif. En septembre 2006, les prix en téléchargement définitif correspondaient à ceux des DVD disponibles dans les magasins. De plus, à l'achat du média en ligne, le DVD était envoyé chez l'internaute. Le tarif de location, quant à lui, variait entre 2 et 5 euros. Le système de tarification gratuite avec des extraits de publicité a été testé par TF1Vision durant l'été 2006. En effet, le site proposait gratuitement la série TV "Le maître du zodiaque", quelques minutes après sa diffusion sur la chaîne. L'internaute pouvait visionner la vidéo durant une semaine. Il y avait trois passages publicitaire d'une minute trente en moyenne durant la diffusion sur le site. Le même système s'observe aux USA. En effet, le site d'ABC, la filiale de Walt Disney, offre gratuitement ses séries TV avec de la publicité (3 spots d'une minute par vidéos). AOL et Warner offrent également gratuitement des programmes.
Le site SpiralFrog fonctionne sur le même principe. Le magazine P2PMag de décembre 2006 et janvier 2007 y consacrait un article en y écrivant que "*Spiral Frog, qui ouvrira début décembre, veut offrir un téléchargement gratuit et légal d'audio et vidéos, financé par la publicité. Les revenus ainsi obtenus seront partagés entre SpiralFrog et les ayants droits. Warner Music - qui avait déjà signé avec YouTube lorsque du contenu de chez eux est utilisé - pourrait signer avec SpiralFrog, lequel collabore déjà avec Universal et EMI ainsi qu'avec le label indépendant Koch Records*". Toutefois, ce site n'est disponible qu'aux Etats-Unis et au Canada. Il s'étendra sûrement à d'autres pays.
Ce concept de gratuité ne semble pas si avant-gardiste. En effet, dans "L'OI" de juillet-août 2007, un article prône les avantages de la gratuité. Il semblerait que les sites de VoD payantes soient menacés par ceux tels que Joost, décrit ultérieurement, qui offrent la gratuité des vidéos avec des passages publicitaires.

"L'Ordinateur Individuelle" de juillet-août 2006 fait remarquer une caractéristique intéressante des médias diffusés sur le site de TF1Vision. Cette caractéristique est que *chaque*

vidéo bénéficie d'une note d'appréciation émanant des internautes, et d'une fiche descriptive et technique". De plus, comme avec Archos pour Canalplay, TF1Vision permet aux consommateurs de transférer ses fichiers vidéos sur un baladeur de la marque Creative Zen Vision. Le lecteur utilisé pour visionner les vidéos est le Media Player de Windows. Les DRM de ses médias ne fonctionnent qu'avec des versions récentes, elles-mêmes indisponibles sur un Mac OS ou Linux. En 2007, ces contraintes sont identiques et le système d'exploitation doit au moins tourner sous Windows XP et posséder au moins la version 6 du navigateur Internet Explorer. En 2006, la durée de location était de 24 heures. Dans "L'Ordinateur Individuel" de juillet-août 2006, le journal fait remarquer que TF1Vision n'offre pas l'image en "haute résolution", telle que le fait Canalplay.

Le site de VoD d'Arte (ArteVoD) est à l'image de la chaîne, axé sur le culturel. En 2006, le site contenait 400 programmes, alors que maintenant il en offre 927. Chaque vidéo contient une fiche très détaillée. Elle peut être conservée durant 30 jours et 48 heures après son premier lancement. Les tarifs varient entre 1.5 euros et 4 euros. En 2006, le site ne proposait que le téléchargement temporaire, alors qu'en 2007, ce mode a été échangé par le streaming et le site propose quelques vidéos en téléchargement définitif. Les types de programmes sont selon "L'Ordinateur Individuel" de juillet-août 2006, "*les émissions phares d'Arte - tous les documentaires, spectacles ou magazines, parfois disponibles dès le lendemain de leur diffusion - ainsi que quelques fictions et grands films classiques*". Chaque média s'accompagne d'un extrait en grand écran. L'achat temporaire ou définitif d'un média oblige le consommateur à être inscrit sur le site.

Un site de Vidéo à la Demande qui n'est pas le prolongement d'une chaîne TV est Vodeo.tv. Il est plus orienté documentaires et reportages. Selon "L'OI (Ordinateur Individuel)" de juillet et août 2006, "*le site proposait à l'achat et à la location, près de 2000 programmes d'une durée oscillant de 15 à 90 minutes. Surtout que certains n'ont jamais fait l'objet d'une diffusion télévisuelle*". De plus, tel que le magazine Micro Actuel d'octobre 2006 le soulignait, les programmes *proviennent majoritairement des chaînes de télévisions françaises*". En 2007, Vodeo.tv a augmenté son catalogue et met en ligne 3404 vidéos. Chaque programme est accompagné d'un résumé, de l'avis de l'internaute et de la rédaction. Les systèmes d'exploitation, tels que Mac OS et Linux, bénéficient de ces documentaires grâce à la possibilité de faire graver les contenus (jusqu'à 104 minutes) qui les intéressent et de se les faire envoyer (expédié sous 72 heures). Cependant, la gravure et l'envoi ont un coût d'à peu près 10 euros. Ce système correspond finalement à l'achat de l'émission et coûte donc plus cher à l'utilisateur. L'avantage de louer est le moindre coût.
De plus, il existe déjà des sites de location à domicile qui seraient plus avantageux en

terme de prix pour le consommateur utilisant Mac OS ou Linux. Ces vidéos clubs en ligne, tels que LocaFilm, Glowria, Dvdfly, Filmodebit, Cinesnape, Cinehome, Dvdpost, Dvdzage, etc., restent également plus avantageux en terme de prix pour les utilisateurs de PC. De plus, l'envoi et le renvoi des DVDs sont gratuits. Avec leur solution d'abonnement et de forfait pour certains, le consommateur peut visionner un DVD à un tarif meilleur marché que celui des vidéos clubs. De plus, le catalogue est bien plus fourni que ceux des VoD. En réalité, il correspond à celui des vidéos clubs classiques. Seul désavantage, il faut attendre de recevoir le ou les films pour les visionner. En effet, cette constatation est démontrée par Micro Actuel de novembre 2006 qui donne certains chiffres de leurs tests. Par exemple, chez Dvdfly, il leur a fallu attendre à peu près 5 jours. De plus, les DVD loués en ligne n'étant pas toujours disponibles, il faut donc créer une liste, ordonnée selon les priorités de chacun, de 5 à 20 choix selon le site et dès que l'un de ces films est libre, il est envoyé. Par conséquent, le consommateur ne connaît pas le film qu'il va recevoir. Les sites de location essaient, cependant, de privilégier les films prioritaires.

Un comparatif entre la VoD et ces sites de location en ligne a été donné par le magazine Micro Actuel de novembre 2006 :

Le site Vodeo.tv diffuse ses programmes en mode streaming et téléchargement définitif. Le temps de conservation après le premier lancement de la vidéo est de 48 heures. Les DRM intégrés dans les vidéos achetées autorisent à copier le contenu sur trois ordinateurs. Les tarifs de location sont compris entre 2 et 4 euros. Toutefois, Micro Actuel d'octobre 2006 rajoute que "*Vodeo.tv propose des forfaits, à ses gros consommateurs de vidéo, de 19 à 99 euros donnant droit à une économie pouvant aller jusqu'à 36 euros*".

VirginMega proposait déjà aux internautes de la musique en ligne. Ce dernier propose également des vidéos, telles que des films, des séries, des clips, des concerts, des programmes pour enfants et des documentaires. Le site mélange donc le téléchargement de

musique à la VoD. En 2006, le site proposait 800 programmes. En juin 2007, selon "L'OI", le site comporte 1000 programmes, ce qui correspond à une faible amélioration. Le site n'est pas très clair. La location a une durée de 48 heures. Pour chaque vidéo, il y a un petit descriptif, ainsi qu'un extrait et des photos, en plus de la bande annonce. Le mode de paiement, en plus de celui par carte bancaire, est très pratique, étant donné que le média est payé lors du reçu de la facture du fournisseur internet. Le visionnage des extraits et des bandes annonces fonctionnent très mal. En effet, l'image n'apparaît pas à l'écran. Est-ce que cela est dû au fait qu'il y a des DRM qui empêchent le visionnage depuis la Suisse ou la mauvaise qualité d'image ?

Les chaînes France2 et France3 proposent ensemble un site de VoD de leurs programmes, tels que des séries TV et des fictions ou des documentaires. Le site propose ces programmes à la location (entre 0,50 à 3 euros selon "L'OI" de juillet-août 2006) ou à l'achat (entre 2 à 7 euros selon "L'OI" de juillet-août 2006).
Quant à France 5, elle permet de revisionner gratuitement, pendant la semaine suivant sa diffusion sur la chaîne, ses émissions. Il y a également des programmes payants pouvant aller jusqu'à 3 euros. La chaîne ne propose pas le mode de téléchargement temporaire, uniquement le streaming et le téléchargement définitif. Le temps de conservation est de 24 heures après le premier lancement. En 2006, selon le magazine Micro Actuel, le site de vidéos de France 5 n'offrait que 16 émissions par semaine.
Les chaînes du groupe France Télévision, France 2, France 3, France 4, France 5 et RFO, proposent également un site de télévision à la demande, francetvod. Ce dernier offre quelques programmes gratuitement et d'autres payante en streaming ou en téléchargement définitif. La location est de 24 heures.
Un autre site de télévision à la demande est Joost. Ce site met à disposition gratuitement des vidéos et des chaînes TV avec un système peer-to-peer grâce aux divers accords signés par Joost. De ce fait et tel que P2PMag de juillet-août 2007 le souligne, "*les émissions diffusées par la plate-forme proviendront uniquement de partenaires et non pas des internautes (comme c'est le cas pour YouTube ou DailyMotion)*". La gratuité de ce site est due à la publicité de 3 minutes par heure.
Dans notre étude, nous nous intéressons essentiellement à la vidéo à la demande. Francetvod et Joost sont, quant à eux, des sites de télévision à la demande (ToD).

Les autres sites analysés sont celui de Wanadoo qui est réservé à ses abonnés, d'Imineo consacré aux loisirs et de Cinezime, la VoD du cinéma indépendant. 24h/24h Vidéo de Wanadoo propose, selon "L'OI" de juillet-août 2006, "*des films, des spectacles, des dessins animés et des documentaires. Les programmes ne sont disponibles qu'à la location à des tarifs variant de 1 à 5 euros. Un centaine de films courts humoristiques sont visualisables*

gratuitement". Le catalogue d'Imineo était en 2006 de 1400 programmes, alors qu'en 2007 il est de 3277. Il offre divers cours de musique et de formations, ainsi que des films et des documentaires. Les qualités VHS et DVD sont à la disposition du consommateur. Les prix de juillet-août 2006 cités dans "L'OI" étaient "*de 7 à 10 euros pour une qualité VHS et de 9 à 13 euros pour une qualité DVD*". En juin 2007, "L'OI" échelonne les prix des vidéos accessibles en téléchargement définitif uniquement à partir de 3.99 euros.

Quant au cinéma indépendant proposé en location et à l'achat par Cinezime, "L'OI" de juillet-août 2006 cite que "*les prix varient en fonction de la durée des vidéos choisies, et des formules de 10 à 50 euros permettant de profiter de réductions*". Toutefois, ce site a l'avantage de ne pas intégrer de DRM dans quelques unes de ses vidéos. Par conséquent, les utilisateurs de Mac OS et Linux peuvent également en bénéficier. De plus, il semble que selon le P2PMag de juillet-août 2007, "*un titre se vend deux fois plus sans DRM*". Cette constatation a été faite par le PDG de la FNAC. "*La FNAC a pris position en faveur des consommateurs en imposant la suppression des DRM pour un résultat plus probant en matière de vente*", selon "L'OI" de juin 2007.
En octobre 2006, le format de lecteur vidéo DivX avait l'intention de lancer son propre site de VoD (P2PMag). Ce dernier s'associerait avec une société suisse MC&C. En effet, selon la Tribune de Genève du 16 août 2007, la société suisse "*ClicMovies vient juste d'ouvrir*". Compte tenu de cette association, le format vidéo des fichiers téléchargés est en DivX et donc lisible sur un lecteur DVD ayant l'option de lecture DivX. Malgré l'association avec la Suisse, les prix restent en euros. De plus, le catalogue n'offre pas de nouveautés.

5.1.4 Les autres acteurs de la VoD

Nous venons de décrire la VoD à travers internet. Cependant, il existe un autre moyen de visualiser chez soi un programme à la demande. Ce service est offert par les fournisseurs du tripleplay, à savoir Neuf, Alice, Free, Club Internet, etc. En 2006, la VoD à travers ces derniers n'était pas répandue. En effet, seul Free offrait ce service. Toutefois, elle allait sûrement se développer et faire appel à l'interface des sites internets. Par conséquent, l'intérêt d'analyser la VoD était double, puisque le résultat serait aussi utile à internet qu'à la télévision. La télévision par le câble, offerte en France par Noos et Numéricâble, peut également être sensible à ce type d'étude. En effet, selon "L'OI" de juin 2007, la VoD était prévue à partir du 4 juin 2007.

En 2007, ce service commence à plaire aux consommateurs de ces fournisseurs. En effet, le 2 janvier 2007, le journal "Le Monde" titre "*les fournisseurs d'accès à Internet se mettent à la vidéo à la demande*". Il donne également les dates de lancement de la VoD par les

différents fournisseurs d'accès internet. "*C'est Orange qui fut le premier à ouvrir ce type de service, fin 2003. Si ce système a mis un certain temps à être adopté par le public, il commence à rencontrer le succès. En novembre, 150 000 programmes ont été loués. Free, de son côté, a ouvert son service vidéo à la demande en décembre 2005. Au premier semestre 2006, il revendiquait quelque 40% du marché, soit environ 750 000 programmes loués. Club Internet à l'automne, Alice pour ses clients de Paris depuis la mi-décembre et Neuf Telecom le 20 décembre ont suivi*".

Comme pour internet, des contraintes de débit s'imposent. En effet, pour assurer une qualité d'image irréprochable, le débit de connexion est très important. Une distance de connexion éloignée du domicile du consommateur à la central téléphonique rend le signal de réception et le débit bas. Chaque FAI (fournisseur d'accès internet) exige un niveau minimum de réception assurant une qualité d'image acceptable. Par exemple, Free propose 2,5 Mbit/s, alors que Neuf et Telecom en demandent le double et que Club Internet exige 6,5 Mbit/s et finalement Alice, le fournisseur le plus exigeant, propose un minimum de 8 Mbit/s. Les différence de ces débits sont au niveau de la qualité de réception. Plus le débit est élevé, plus la VoD est de bonne qualité.

Les programmes arrivant par ADSL sur le téléviseur sont, en réalité, en format Mpeg-2. En effet, tel que le journal Microactuel de septembre 2007 le dit, le : "*format Mpeg-2 est une opération incontournable pour transformer un format de type Secam (720 points × 576 lignes) en un flux vidéo correspondant au débit moyen de l'ADSL*". Cette compression oblige un débit élevé afin que le résultat soit de bonne qualité. Un nouveau format permet une compression avec un débit de connexion moins exigeant, le Mpeg-4. Club Internet l'a intégré dans son offre. L'avantage de ce format est décrit par le magazine Microactuel de septembre 2007 de la manière suivante : "*le format Mpeg-4, plus récent, compresse mieux les vidéos tout en délivrant une qualité d'image irréprochable*". De plus, ce dernier donne une définition des formats Mpeg : "*Norme de compression pour la vidéo numérique. Le Mpeg-2 correspond à une qualité DVD ; le Mpeg-4 permet de réduire au maximum le flux vidéo*".

Tout comme la VoD par internet, les FAI limitent le temps de location (24 heures ou 48 heures). Cette limite dépend du média et du fournisseur. Les vidéos sont disponibles en mode streaming, en téléchargement temporaire ou en téléchargement définitif. Le prix de la location varie de 0,50 euros le film ancien à 4 euros la nouveauté. En janvier 2007, selon Le Monde de janvier 2007, les catalogues de chaque FAI étaient de 1400 programmes (400 films) pour Orange, 2500 programmes (1000 films) pour Free et 1000 programmes (350 films) pour Club Internet. Toutefois, comme le souligne "L'OI" de juin 2007, "*le choix est*

plus vaste sur Internet : on y trouve non seulement des films mais aussi des émissions de télé, des documentaires et des reportages".

5.1.5 Présentation de la VoD proposée par les FAI

En octobre 2006, le magazine Micro Actuel comparait les différents services de VoD offerts par les FAI. Il les décrivait pour les fournisseurs de Free, d'Orange, de TPS-VoD et de Club-Internet. Dans Micro Actuel de septembre 2007, d'autres FAI sont testés, à savoir Neuf et Alice.

Orange propose un catalogue de 200 films, issus de Canalplay et d'un grand nombre de documentaires. De plus, le site d'Orange 24h/24h Vidéo est également disponible via Internet. La plate-forme offerte par le FAI est claire. Selon Micro Actuel d'octobre 2006, un point négatif a été avancé, à savoir que "*lors du test, une coupure de connexion les a contraints à racheter le film visionné, et ce, malgré la garantie d'accès de 24h*".

Quant à Free, ce dernier reprend les vidéos de Canalplay, alors qu'Alice reprend celles de Virginmega. De plus, selon Micro Actuel de septembre 2007, Alice utilise le format Mpeg-4 pour ses vidéos. Neuf propose un catalogue d'à peu près 1000 films en septembre 2007.

Tel que Micro Actuel d'octobre 2006 le dit, "*avec le lancement de son service de télévision en partenariat avec Microsoft (dont l'interface est excellente), Club-Internet ouvre la voie à la VoD*". Il propose 1000 programme en VoD à de bons prix. Chaque semaine, Club-Internet ajoute une cinquantaine de programmes. De plus, pour reprendre les mots de Micro Actuel, "*le FAI négocie au cas par cas les droits des émissions avec Studio Canal, TF1 Vidéo ou France Télévision Distribution*".

TPS-VoD offre des films (150), des spectacles et des concerts.

5.1.6 Les motivations et les raisons du choix des attributs de la VoD

Suite à l'étude des sites de Vidéo à la Demande existant au début de notre analyse, ainsi que des aspects liés à ce type de produit, nous avons envisagé plusieurs caractéristiques définissant la VoD. Ces caractéristiques étaient les suivantes :

Attributs	**Modalités**	**Attributs**	**Modalités**
Les genres du catalogue	Films Emissions de variétés Séries Documentaires Reportages Spectacles Concerts	Le niveau d'ancienneté du programme (selon les genres)	Nouveautés (8-9 mois après sa sortie en salle) Moyennement récents (9 mois - 1 an après sa sortie en salle) Anciens (1-5 ans après sa sortie en salle) Archives plus de 5 ans (après sa sortie en salle)
Temps de conservation	30 jours	Le nom du site (Canal, TF1, Ina, etc.)	
Service VoD disponible via	Internet Télévision (TriplePlay - FAI) Vidéoclub Location de DVD en ligne	Type de diffusion	Streaming Téléchargement temporaire Téléchargement définitif
Temps de conservation après le premier lancement	24 heures 48 heures 72 heures	Les autres supports de stockage	Baladeur numérique Aucun
Tarifs de location et d'achat	Abonnement Gratuit avec de la publicité Prix par programme	Protection DRM	Vidéo en contient Vidéo n'en contient pas
Visionnage à travers	Windows Media Player (9 ou 10) (incompatible avec Mac OS et Linux) DivX Lecteur propre au site	Disponibilité de	La bande annonce L'extrait Aucun
Inscription au site	Obligatoire Pas obligatoire	Qualité d'image (et gel d'image)	Faible (beaucoup) Moyenne (moyen) Haute (aucun)
Mode de paiement	Cash (Vidéo Club) Carte bancaire CB Visa Master Card	La quantité de programme dans le catalogue	plus de 1000 1000 700 500 moins de 500

Le choix de ces attributs s'explique de la manière suivante. Certains comblent des défauts des sites français de VoD fréquemment relevés par les magazines informatiques, tels que les catalogues ne sont pas assez fournis, les films ne sont pas assez récents et les tarifs sont

trop élevés. Compte tenu de ces informations, nous avons donc choisi de tester l'influence de ces constatations en retenant les attributs suivants : la quantité de programmes dans chaque catalogue, le niveau d'ancienneté des programmes, le genre de programmes par site et le prix. Pour le tarif, nous avons proposé des formules apparentées à la location en ligne de DVD, c'est-à-dire un abonnement dans lequel l'internaute est autorisé à visionner en streaming ou en téléchargement temporaire un nombre limité de films pour un tarif par mois unique. Un autre concept qui semble s'être développé en 2007 est la gratuité par la publicité. Nous avons donc également associé dans le tarif cet aspect-là. Quant au niveau d'ancienneté des programmes, nous avons supposé la nouveauté en fonction de sa sortie en salle. La valeur de 8 à 9 mois est issue des contraintes de diffusion légale en matière de films français. En effet, en France, les films français ne peuvent être au vidéo club que 8 mois après leur sortie en salle. Quant aux films américains, nous ne connaissons pas cette contrainte. Il en est de même pour la Suisse. C'est pourquoi, nous nous sommes basés sur les critères de la France en matière de diffusion.

D'autres attributs nous sont venus à l'esprit. Le nom du site, par exemple, a pour but de tester et d'identifier un éventuel effet de marque. De plus, le rapatriement d'une vidéo via Internet (site de VoD et location en ligne (exemple : Locafilm)) avec l'obligation d'avoir un ordinateur, via un téléviseur (FAI (TriplePlay)) ou via le vidéo club a également été retenu. Dans le rapatriement d'un média par Internet, il y a également le type de diffusion, c'est-à-dire le streaming, le téléchargement temporaire ou l'achat. Dans le même contexte, il y a les autres supports autorisant le transfert du média, tels qu'un baladeur (exemple IPod, Archos, etc.), qui peuvent être pris en compte. Une autre variable qui différencie un site de l'autre est la durée de location. Nous avons choisi les modalités 24 heures, 48 heures et 72 heures.

Un attribut, qui reprend une information du P2PMag, est l'intégration des DRM dans les médias. En effet, la FNAC faisait remarquer que l'absence des DRM augmentait la vente de vidéos. Toutefois, cet article datant de juillet-août 2007, nous n'avons donc pas considéré cet attribut dans notre étude. Il peut toutefois être intégré dans le cadre d'une autre analyse. Une variable pouvant être déterminante pour les utilisateurs de Mac OS ou de Linux est le type de lecteur compatible avec la vidéo. Cependant, tel que nous le mentionnons dans la suite, nous ne considérons que les individus avec un équipement PC tournant sous Windows 2000 ou XP. Les autres attributs, tels que la bande annonce, le mode de paiement, l'obligation de s'inscrire au site et la qualité de l'image, peuvent également avoir leur importance.

De cette sélection, nous avons dû faire une réduction pertinente des attributs à retenir.

En effet, une telle quantité d'attributs et de modalités associée à chacun représente une analyse moins fiable et trop complexe. Le nombre de paramètres à estimer aurait été d'au moins 34, étant donné que nous avons 16 attributs dont certains ont deux, trois, quatre, cinq ou sept modalités et d'autres pour lesquels les modalités ne sont pas encore déterminées. Les modalités de référence n'ont pas été comptées dans ce résultat. De plus, il est difficile avec une telle quantité d'obtenir un plan raisonnable compte tenu des contraintes liées à ce dernier. Une explication plus détaillée est donnée dans une autre section de ce chapitre.

Nous avons donc dû limiter les attributs à ceux les plus pertinents selon nos recherches, ainsi que diminuer les modalités des variables prises en compte dans notre étude. Les attributs n'ayant pas un intérêt primordial ont été d'abord le type de lecteur permettant le visionnage et la protection DRM. Les raisons ont été abordées précédemment. Pour le lecteur, étant donné que nous avons supposé que l'interviewé possédait un ordinateur tournant sous Windows 2000 ou XP, dans ce cas tous les players sont compatibles et donc indifférents au consommateur. Toutefois, le fait de devoir installer un nouveau lecteur est peut-être une variable influente, mais nous ne pensons pas qu'elle soit décisive dans le choix d'un site de VoD. En ce qui concerne les DRM, l'influence dans les ventes n'ayant été postulée qu'après le début de notre étude, nous ne l'avons donc pas prise en compte dans l'analyse. Toutefois, afin de ne pas alourdir les produits de VoD, nous avons également pensé que le terme de DRM et la notion de lecteur compatible étaient trop techniques et demandaient, par conséquent, une définition ou une explication supplémentaire durant l'interview. Au début de notre étude (janvier 2006 à janvier 2007), les interviewés choisis étant suisses, ils n'avaient donc certainement pas la même connaissance dans la VoD que les Français. Par conséquent, il fallait que les variables représentant les produits soient suffisamment claires. Certains attributs semblent plus compréhensibles désormais, étant donné que la technologie est plus avancée et que le nombre d'articles s'est depuis multiplié. En effet, un grand nombre de sujets traitant de la VoD sont parus entre janvier et septembre 2007.

Nous avons également enlever le type de diffusion et la qualité de l'image en appliquant le même argument, c'est-à-dire que ces variables sont très techniques et demandent, comme précédemment, des explications supplémentaires. De plus, l'équipement technique supposé de l'interviewé assure une bonne qualité d'image.
Le nom, l'inscription au site et le mode de paiement ne nous ont pas paru être des variables primordiales. Nous les avons donc enlevés du modèle. Le temps de conservation du média est de 30 jours sur tous les sites. Par conséquent, il n'est pas intéressant d'en faire une variable explicative. Quant à la disponibilité de la VoD via Internet, la télévison, etc. et

les autres supports de stockage, ils rajoutaient de la complexité à la compréhension de la Vidéo à la Demande. Par conséquent, afin que les produits présentés aux interviewés soient clairs, nous avons donc également enlevé ces variables de l'analyse.
L'utilité de cette dernière peut, toutefois, être constatée en interrogeant l'individu. En effet, il suffit d'ajouter une question du type, "Parmi les propositions suivantes, quel mode de disponibilité de la VoD préférez-vous ?"

☐ Internet : obligation d'avoir un ordinateur
☐ Télévision - TriplePlay - par les FAI
☐ Vidéoclub
☐ Location de DVD en ligne (exemple, Locafilm)

Une des modalités de cette question convient mieux aux consommateurs français qui ont le service de TriplePlay.
Le même type de question peut être posé pour l'attribut "type de diffusion".

Il ne reste donc plus que les attributs suivants :

Attributs	**Modalités**	**Attributs**	**Modalités**
Les genres du catalogue	Films Emissions de variétés Séries Documentaires Reportages Spectacles Concerts	Le niveau d'ancienneté du programme (selon les genres)	Nouveautés (8-9 mois après sa sortie en salle) Moyennement récents (9 mois - 1 an après sa sortie en salle) Anciens (1-5 ans après sa sortie en salle) Archives plus de 5 ans (après sa sortie en salle)
La quantité de programme dans le catalogue	plus de 1000 1000 700 500 moins de 500	Disponibilité de	La bande annonce L'extrait Aucun
Tarifs de location et d'achat	Abonnement Gratuit avec de la publicité Prix par programme	Temps de conservation	24 heures 48 heures 72 heures

Les modalités du catalogue étaient trop nombreuses. Nous avons donc opté pour un minimum, c'est-à-dire deux modalités. Ces dernières sont 600 et 1000 titres proposés. Elles sont basées sur les tableaux comparatifs de "L'OI" de 2006 se trouvant à la page 163. Dans ces tableaux, nous constatons que le catalogue dépasse rarement les 1000 pour les sites proposant des films. De plus, afin de ne pas prendre une valeur proche de 1000, nous avons donc choisi la moyenne entre 500 et 700, c'est-à-dire 600.

Le nombre de paramètres à estimer, c'est-à-dire 15, reste encore trop élevé. L'explication est dépendante des plans d'expérience et du type de recueil des données. Elle sera abordée par la suite. Afin de réduire ce nombre, il faut enlever des modalités superflues. Nous pouvons supprimer la modalité 72 heures de l'attribut durée. En effet, la durée de 72 heures n'est pas très répandue dans les sites de VoD et n'est donc pas très importante. La bande annonce ou l'extrait sont en quelque sorte très semblables. Etant donné que la bande annonce est plus fréquente, nous enlevons la modalité extrait.

Afin d'alléger davantage le nombre de paramètres, nous avons également remanié les modalités du genre de catalogue et du niveau d'ancienneté du programme. D'abord, la notation proposant les modalités de la variable genre amenait à l'impossibilité d'un site mélangeant les films, les séries, les concerts, etc. Nous avons décidé d'enlever les documentaires dans le but de ne pas alourdir les composants de cette modalités. Nous avons pensé rajouter une variable proposant un lien ou non à d'autres sites documentaires, mais nous y avons renoncé afin de ne pas avoir trop d'attributs dans le modèle. Les modalités de la variable genre ont donc été modifiées de la manière suivante :

Modalités
100% Films
75% Films 25% Séries TV
50% Films 25% Séries TV 10% Concerts 15% Spectacles
25% Films 75% Séries TV
100% Séries TV
25% Films 25% Séries TV 25% Concerts 25% Spectacles

Le nombre de modalités était encore trop élevé. En réalité, nous voulions satisfaire tous les types de consommateurs en proposant tous les styles de médias possibles. Finalement, nous nous en sommes tenus à un site de VoD sur les films et les séries. De plus, nous avons remarqué que le pourcentage de séries était souvent plus faible que celui des films dans les sites de location de DVD en ligne. Cette remarque peut être vérifiée dans le tableau ci-dessous sur le nombre de films et de séries. Par conséquent, nous avons opté pour des sites offrant soit que des films, soit que des séries, soit plus de films que de séries ou soit un partage égal de films et de séries. Les modalités de la variable genre sont donc devenues comme suit :

Modalités
100% Films
75% Films 25% Séries
50% Films 50% Séries
100% Séries

Nous avons dû faire le même type de remaniement pour la modalité du niveau d'ancienneté. En effet, nous avons constaté que notre notation amenait à des sites ne pouvant pas être un mélange de nouveautés, de moyennement récents, etc. Par conséquent, nous avons remanié cet attribut en enlevant d'abord une des modalités, c'est-à-dire "archives", qui donnait une description trop détaillée du niveau d'ancienneté.

Nous avons d'abord modifié cette variable de la manière suivante :

Modalités
100% Nouveauté
50% Nouveauté 50% Récent
50% Nouveauté 25% Récent 25% Ancien
50% Récent 50% Ancien
100% Ancien

Puis, nous avons également décidé d'enlever la modalité "récent" détaillant également de manière excessive la variable. De plus, étant donné que le nombre de nouveautés associé au nombre de catalogue atteignait des valeurs inexistantes dans les sites de location de DVD en ligne, nous avons donc dû modifier la notation des modalités de cette variable. Un résumé du nombre de catalogue de quelques sites de VoD, de location à domicile et d'annuaire de VoD est donné dans le tableau suivant :

Origine	Nom du site	Type	Films	Séries	Nouveautés Film	Nouveautés Série
Suisse	DVDFly	Location à domicile	5299	1008	-	-
Suisse	voirunfilm	Annuaire de VoD	1515	-	-	-
Suisse	DVDnet	Location à domicile	4340	630	140	96
France	Canalplay	VoD	1159	-	52	-
France	Glowria	Location à domicile	-	173	-	66

Les nouveautés séries correspondent à des programmes sortis en 2005 et 2006, compte tenu que notre étude a débuté en 2006. Par exemple, dans le site DVDnet, sur les 96 nouveautés séries, 15 sont sorties en 2006 et 81 en 2005. Pour le site Glowria, sur les 66, il y a en 15 en 2006 et 51 en 2005.

Pour les films et les séries, nous avons opté pour les valeurs de DVDnet. La raison est que nous avions l'impression que les films et les séries proposés par DVDFly apparaissaient plusieurs fois dans le site. De plus, le site DVDnet offre plus de renseignements sur le catalogue que les autres.

Les valeurs de ces dernières sont donc les suivantes :

4340	**Films dont**
140	Nouveautés Films
4200	Anciens Films
630	**Séries dont**
96	Nouveautés Séries
534	Anciennes Séries

Nous remarquons donc que l'association de la modalité 100% Nouveauté, 100% Films ou 100% Séries et 600 ou 1000 programmes signifie que le site propose 600 ou 1000 nouveautés films ou 600 ou 1000 nouveautés séries. Or, selon les sites DVDnet, cette quantité de nouveautés n'existe pas.

C'est pourquoi, nous avons préféré exprimer les modalités de la variable, niveau d'ancienneté, comme suit :

Modalités
- Un maximum de nouveauté films-séries
- Moitié des nouveautés et le reste ancien
- Que des anciens

Le variable tarif correspond en réalité au prix à la location. Il a été difficile de proposer une valeur au prix et à l'abonnement. En effet, nous ne connaissions pas le fonctionnement des droits d'auteur dans le cas de la diffusion de vidéos en streaming ou en téléchargement temporaire. De plus, étant donné que notre analyse se base sur les consommateurs suisses, la valeur locative des sites de VoD français n'est pas un bon exemple dans la tarification de vidéos suisses. En effet, les prix de location diffèrent d'un pays à l'autre. Par exemple, dans les vidéos clubs en France, le prix d'un DVD est d'à peu près 2 euros, alors qu'en Suisse, il est de 7 francs. Par conséquent, lorsque les journaux déclarent que la VoD est encore chère, elle compare des films d'une valeur de 4 à 5 euros à des films de 2 euros. De plus, étant donné qu'au début de notre étude il n'existait pas de site de VoD suisse, nous n'avons donc pas eu d'exemple pour proposer une tarification.

Une autre contrainte dans la détermination de la tarification a été les droits de diffusion. En effet, le fait que les sites de VoD ne présentent pas les mêmes nouveautés dans leur catalogue, tel que le souligne le Micro Actuel d'octobre 2006, est dû aux "*accords différents avec les majors*". Par conséquent, afin de proposer un prix de location en VoD, nous nous sommes renseignés auprès de la TSR et de Disques Office, distributeurs de films en Suisse, pour connaître le fonctionnement de ces droits.

Les renseignements donnés par la juriste de la TSR ont été sur les droits de diffusion. En effet, la TSR a deux types de contrats, c'est-à-dire l'exclusivité, valable pour une certaine période, et la priorité, utile vis-à-vis des programmes français. Le TSR comble les coûts de diffusion grâce aux redevances que paient les suisses, la publicité et le sponsoring, qui correspond au parrainage d'une émission. La négociation d'une diffusion ou d'une rediffusion s'effectue avec le distributeur. Cette négociation dépend du nombre de rediffusion, le territoire, le succès en salle du film, etc. En France, la sortie en DVD d'un film est réglementée (8 mois), alors qu'aux Etats-Unis, la réglementation est moins stricte. Un film américain qui a bien marché dans les salles se fera plus attendre en DVD. Ces informations nous ont permis de mieux comprendre le principe de négociation des prix. Toutefois, cela n'a pas suffit pour fixer une valeur aux programmes.
Quant aux distributeurs, les renseignements recueillis n'ont pas été suffisants pour déterminer un prix.

Nous avions également la répartition des recettes payantes des archives de l'Ina parue dans Micro Actuel d'octobre 2006, à savoir "*46% sont reversés aux ayants droit, 32% sont réinvestis dans la numérisation, la sauvegarde et la valorisation du patrimoine audiovisuel, 22% couvrent les frais financiers (TVA, taxes et transactions en ligne)*". Ces renseignements ne suffisaient également pas à choisir le prix de location d'une vidéo. Par conséquent, nous avons laissé la variable prix sans pour autant signaler une valeur.

Il en était de même pour la publicité. En effet, nous avons découvert qu'il était compliqué de fixer une durée pour la publicité. Nous avons d'abord pensé reprendre le temps des spots publicitaires proposé par TF1Vision lors de la diffusion de la série, "Le maître du zodiaque", la fréquence et la durée de ces spots étant de 3 publicités d'une minute 30. Or, nous avons appris par la juriste de la TSR que la publicité était réglementée en Suisse. En terme de législation de la publicité, la loi sur la radio et la TV (art. 19 de l'ORTV et art. 11 al. 2 de LRTV) dit que la publicité ne doit en principe pas excéder 15% du temps d'émission quotidien ni 20% d'une heure d'émission. Finalement, nous avons décidé de proposer la modalité d'un programme gratuit avec de la publicité sans pour autant mentionner la durée du ou des spots. Quant à l'abonnement, nous avons également préféré ne pas proposer de forfait ou de package moyennant un prix.

Compte tenu de toutes ces informations, nous avons décidé de retenir, dans notre étude, les attributs et les modalités suivants :

Modalités
Attribut A : Le nombre de programmes
1000 (1)
600 (2)
Attribut B : La composition du site en terme de films-séries
100% Films (1)
75% Films 25% Séries (2)
50% Films 50% Séries (3)
100% Séries (4)
Attribut C : La composition du site en terme nouveauté-ancien
Un maximum de nouveauté films-séries (1)
Moitié des nouveautés et le reste ancien (2)
Que des anciens (3)
Attribut D : Tarif
Payant (1)
Gratuit pub (2)
Abonnement (3)
Attribut E : Temps de conservation du film après le premier visionnage
24 heures (1)
48 heures (2)
Attribut F : Disponibilité de la bande annonce ou de l'extrait
Bande annonce disponible (1)
Aucun disponible (2)

5.2 Provenance de l'échantillon

La VoD requiert un équipement de base pour un fonctionnement optimal. C'est pourquoi, nous avons recherché des consommateurs (ceux qui participeront à l'interview) ayant accès à un tel équipement. En voici le descriptif qui est basé sur les différents journaux informatiques. Il reprend en réalité un grand nombre de contraintes citées dans la section présentant certains sites de VoD. Nous supposons tout d'abord que l'internaute, qui sera interviewé dans notre application, détient un ordinateur ou en a l'usage à travers une entreprise. Cette ordinateur doit être suffisamment puissant pour accueillir la VoD, c'est-à-dire qu'il a une puissance de 2 Ghz au minimum, avec un espace disque dur libre de 2 à 5 Go, qu'il tourne au minimum avec Windows 2000 ou XP, une mémoire vive d'au moins 256 Mo, une carte graphique de 64 Mo au minimum et qu'il a au moins la version d'Internet Explorer 6 et la version 9 de Windows Media Player (avec une légère préférence pour la 10). La connexion Internet doit être au minimum de 2 Mbit/s pour le visionnage en streaming.

Les étudiants universitaires ont généralement accès aux ordinateurs offerts par l'établissement. Ces derniers ont une technologie suffisamment avancée pour permettre la Vidéo à la Demande. De plus, nous avons supposé que la tranche d'âge des consommateurs pouvant être intéressés par ce type de produit serait autour des 25 à 35 ans. Par conséquent, interviewer des étudiants de l'Université semblait correspondre aux critères que nous avons avancés. Au début de ce chapitre, nous avons montré que la VoD en Suisse n'était pas aussi avancée qu'en France. Nous avions conclu que notre étude serait concentrée sur les internautes suisses. Par conséquent, nous avons opté pour les étudiants de l'Université de Genève. Cette dernière a effectivement un équipement de haut niveau, puisqu'elle est dotée de la fibre optique. Chaque étudiant peut accéder à l'aide d'un mot de passe personnel à ces ordinateurs dans les salles informatiques ou peuvent se connecter au réseau de l'Université avec leur propre ordinateur. Par conséquent, il semble évident que ces étudiants ont des connaissances suffisantes en informatique pour participer à cet interview.

Afin d'obtenir un nombre suffisant d'interviewés, nous avons préféré nous concentrer sur les étudiants de première année en sciences sociales. En effet, le nombre d'étudiants de cette section et de ce niveau d'étude sont bien plus nombreux que les autres. En février 2007, 120 étudiants ont ainsi participé à l'interview. Ces étudiants ont principalement entre 19 et 25 ans. Ils viennent du même milieu social. Nous pouvons donc dire que la population est homogène. Nous n'avons pas besoin de segmenter notre échantillon. Afin de s'assurer que la Vidéo à la Demande ait un sens pour les interviewés, nous avons pris une

population ciblée, dont la plupart ont certainement des connaissances technologiques et informatiques avancées. Avec cette manière de procéder, nous avons donc tenu compte du marché cible qui serait susceptible de s'intéresser à ce type de service. Par conséquent, la tranche d'âge sélectionnée devrait correspondre à celle ayant une part de marché grande.

5.3 Le plan d'expérience adapté à l'application

Dans la section 5.1, nous avons défini les attributs et les modalités de l'étude. Nous avons, en effet, 6 attributs, dont 3 ont 2 modalités, 2 ont 3 modalités et 1 a 4 modalités. Nous avons donc $2^3 \times 3^2 \times 4^1 = 288$ produits.
Toutefois, la combinaison de certaines modalités ne sont pas réalisables, compte tenu des informations sur la quantité de programmes. Par exemple, un produit composé de la modalité 1000 programmes (1) de l'attribut A et de la modalité 100% Séries (4) de l'attribut B correspond à un site proposant 1000 séries. En réalité, selon DVDnet, nous n'avons que 630 séries. Il faut donc enlever toutes les combinaisons où A=(1), B=(4), C=(1)(2)(3), D=(1)(2)(3), E=(1)(2) et F=(1)(2), c'est-à-dire 36 produits. En effet, en calculant le nombre de combinaisons où A et B sont fixes et C, D, E et F sont variables, nous avons $3^2 \times 2^2 = 36$ produits.
Les combinaisons où l'attribut A a la modalité 600 programmes (2), l'attribut B a la modalités 100% Séries (4), et l'attribut C a la modalité moitié nouveautés et le reste anciens (2), doivent également être enlevées. En effet, lorsque nous proposons ces produits, nous supposons que dans les 600 programmes, nous avons 48 nouvelles séries (la moitié des nouveautés) et que le reste correspond aux anciennes séries, c'est-à-dire $600 - 48 = 552$. Or, selon DVDnet, nous n'avons que 534 séries anciennes. Il manque donc 18 séries. Par conséquent, nous devons enlever les combinaisons dont les attributs A=(2), B=(4) et C=(2) sont fixes et les attributs D=(1)(2)(3), E=(1)(2) et F=(1)(2) sont variables, c'est-à-dire $3^1 \times 2^2 = 12$ produits.
Le même principe s'applique lorsque l'attribut A a la modalité 600 programmes (2), l'attribut B a la modalité 100% Séries (4), et l'attribut C a la modalité 100% anciens (3). En effet, étant donné que nous n'avons que 534 séries anciennes, la proposition de 600 anciennes séries n'est donc pas réalisables. Le nombre de produits à enlever s'élève à $3^1 \times 2^2 = 12$ produits, étant donné que les attributs A=(2), B=(4) et C=(3) sont fixes et les attributs D=(1)(2)(3), E=(1)(2) et F=(1)(2) sont variables.
Finalement, nous obtenons $288 - 36 - 12 - 12 = 228$ produits. Nous avons dû enlever 60 combinaisons.

Afin d'analyser notre spécification de l'option "zéro", nous choisissons de présenter les produits en comparaison par paires. La raison est que nous ne voulons pas imposer un trop

grand nombre d'alternatives à l'interviewé. En effet, dans le choix des alternatives, il faut tenir compte du refus et du conflit. Par conséquent, en proposant deux produits par choix en plus des options "zéros", nous obtenons le minimum d'alternatives, c'est-à-dire quatre. Compte tenu que nous avons 228 produits, nous pouvons donc créer $\frac{228\times227}{2} = 25'878$ paires.

Le nombre de paires étant impraticable pour un interviewé, il faut donc les réduire. Pour cela, nous décidons d'utiliser en association avec les plans en bloc incomplets équilibrés, les plans D-optimaux. Les plans D-optimaux permettent de réduire le nombre de produits et les plans en bloc incomplets équilibrés de placer ces produits en paire dans un nombre de blocs fixés. Le nombre de blocs ou de choix dépend, quant à lui, du nombre de paramètres à estimer. Etant donné que nous avons 3 attributs à 2 modalités, 2 attributs à 3 modalités et 1 attribut à 4 modalités, le nombre de paramètres inconnus s'élève donc à 16. Les variables explicatives étant qualitatives et chaque modalité par attribut étant codée à l'aide d'une variable indicatrice, nous devons donc enlever une variable de référence à chaque attribut. Par conséquent, nous obtenons 10 paramètres à estimer. Le nombre de blocs ou de choix par individu doit être supérieur à 10+1=11.
Un récapitulatif de toutes les informations sur le nombre de produits, de paires, de paramètres à estimer et de blocs minimums se trouve dans le tableau suivant :

Nombre de Produits totaux	288
Nombre de Produits enlevés	36+12+12=60
Nombre de Produits finaux	**228**
Nombre de Paires totales	$\frac{228\times227}{2} = 25'878$
Nombre de Paramètres estimés	**10**
Nombre de Questions minimales	10+1=11

L'application des plans D-optimaux requière la connaissance du nombre réduit de produits que l'on désire obtenir. Cette connaissance découle des plans en blocs incomplets équilibrés. En utilisant la théorie de ces plans, illustrée dans le chapitre 2, nous pouvons trouver cette valeur. Nous savons que nous voulons deux produits par blocs, c'est-à-dire que $k = 2$. Le nombre de blocs se calcule de la manière suivante :

$$b = \frac{a!}{(a-k)!k!}$$

Par exemple, si nous décidons de réduire les 228 produits à
- 10 produits ($a = 10$), nous aurons 45 blocs ($b = 45$)
- 9 produits ($a = 9$), nous aurons 36 blocs ($b = 36$)
- 8 produits ($a = 8$), nous aurons 28 blocs ($b = 28$)

– 7 produits ($a = 7$), nous aurons 21 blocs ($b = 21$)
– 6 produits ($a = 6$), nous aurons 15 blocs ($b = 15$)

Selon ces informations, il semble évident que proposer 45, 36 ou 28 choix est irréaliste. En effet, l'individu ne peut répondre de manière concentrée et fiable à tant de questions. Par contre, les 21 et les 15 choix semblent être plus censés. Nous optons finalement pour 21 questions, dans lesquelles nous disposerons les paires de produits. Des 228 produits initiaux, le plan D-optimal doit donc les réduire à 7. Dès que l'application du plan D-optimal est effectuée, nous pouvons utiliser le plan en blocs incomplets équilibrés sur les 7 produits sélectionnés. Dans ce plan, les produits apparaissent 6 fois ($r = 6$) et les paires 1 fois ($\lambda = 1$). En effet, en utilisant les équations $N = ar = bk$ et $\lambda = \frac{r(k-1)}{a-1}$, nous obtenons $r = \frac{bk}{a} = \frac{21\times 2}{7} = 6$ et $\lambda = \frac{6(2-1)}{7-1} = 1$.

Le plan résultant par cette méthode de choix, c'est-à-dire l'association des plans D-optimaux et des plans en blocs incomplets équilibrés, ne nous convient pas. Les raisons, décrites dans la section 2.9, sont la faible variabilité des produits sélectionnés par cette approche (7 produits sur les 25'878). Compte tenu de cette restriction, nous optons, finalement, pour une autre méthode, qui consiste à n'appliquer que les plans D-optimaux sur toutes les paires de produits. En effet, nous appliquons un des algorithmes des plans D-optimaux sur la matrice d'expérience du modèle, afin qu'il sélectionne un nombre fixe de lignes de cette matrice. Le nombre de lignes choisies représente le quantité de questions ou de choix proposés à l'interviewé. Précédemment, nous avions opté, en accord avec les contraintes des plans en blocs incomplets équilibrés, pour 21 paires de produits. Nous pensons que 20 paires conviennent mieux. En effet, une vingtaine de questions suffisent et diminuent ainsi modestement la tâche de l'interviewé. Par conséquent, les plans D-optimaux vont sélectionner 20 lignes sur les 25'878 paires.

La matrice d'expérience du modèle dépend des différences des modalités des attributs des produits présentés en paires. Avant d'appliquer les plans D-optimaux à cette matrice d'expérience, il faut enlever les paires inutiles, décrites dans la section 2.10.
Cette réduction a été programmée sur Matlab et le code informatique concernant la suppression des produits infaisables sur le logiciel TSP-GiveWin.

Dans la suite, nous décrivons en détail les étapes de cette programmation. Elle s'effectue en 4 étapes. La première est réalisée dans le programme SAS. Il permet de créer les 288 produits selon les modalités des attributs retenus dans la sous-section 5.1.6. Nous copions

et collons ces 288 produits depuis l'output de SAS dans un fichier texte :

Produit	A	B	C	D	E	F
1	1	1	1	1	1	1
⋮						
288	2	4	3	3	2	2

Dans cette étape, les modalités des attributs ne sont pas représentées sous la forme de variables indicatrices. Cette transformation sera faite dans la deuxième étape. Nous continuons la mise en oeuvre de notre base de données en créant un autre fichier texte contenant que les 228 produits faisables sur les 288.

La deuxième étape est effectuée à l'aide du logiciel TSP-GiveWin. Cette étape permet la construction d'un fichier texte représentant les 25'878 paires des 228 produits. Dans cette base, nous avons dans les colonnes, la paire, le produit 1, la produit 2 et un vecteur de différence des attributs des produits 1 et 2 :

Paire	Produit 1	Produit 2	DA1	DB1	DB2	DB3	DC1	DC2	DD1	DD2	DE1	DF1
1	1	2	0	0	0	0	0	0	0	0	0	1
⋮												
25'878	263	264	0	0	0	0	0	0	0	0	0	1

Les troisième et quatrième étapes sont réalisées dans le logiciel Matlab. Elles éliminent d'abord les paires évidentes et celles qui ne comparent qu'un attribut. La réalisation de ces étapes passe par la création d'un fichier txt qui contient les paires à supprimer. Cette étape donne 24'936 paires au lieu des 25'878 :

Paire	Produit 1	Produit 2	DA1	DB1	DB2	DB3	DC1	DC2	DD1	DD2	DE1	DF1
1	1	4	0	0	0	0	0	0	0	0	1	1
⋮												
24'936	263	264	0	0	0	0	0	0	0	0	0	1

La dernière étape supprime les paires dont le vecteur des différences est le même en valeur absolue que celui d'autres paires. Cette dernière étape nous fournit notre base de données finale :

Paire	Produit 1	Produit 2	DA1	DB1	DB2	DB3	DC1	DC2	DD1	DD2	DE1	DF1
1	1	4	0	0	0	0	0	0	0	0	1	1
⋮												
791	245	264	0	0	0	1	-1	0	0	1	1	1

Finalement, nous obtenons 791 paires, soit une réduction $\frac{(25'878-791)}{25'878} \times 100$ =96.94%. En effet, au début de l'étude, nous avions 25'878 paires.

En résumé, nous avons donc les informations suivantes :

Nombre de Produits totaux	288
Nombre de Produits enlevés	36+12+12=60
Nombre de Produits finaux	**228**
Nombre de Paires totales	$\frac{228\times227}{2} = 25'878$
Nombre de Paires enlevées	25'087
Nombre de Paires finales	**791**
Pourcentage de réduction	96.94%
Nombre de Paramètres estimés	**10**
Nombre de Questions minimales	10+1=11
Nombre de Questions proposées	**20**

Le tableau suivant résume les étapes de création de notre base de données, qui ne contient que les produits et les paires désirées :

Etapes	**Logiciel**	**Parcours**
1	SAS	Création des 288 produits
2	GiveWin	Elimination des 60 produits infaisables
3	Matlab	Suppression des paires évidentes et celles qui comparent un attribut
4	Matlab	Suppression des paires dont la différence est identique en valeur absolue

Ces étapes de programmation ont donc permis de construire une base de données sur laquelle nous appliquons un plan D-optimal, afin de réduire le nombre de paires à 20 choix.

L'application de ce plan D-optimal a été réalisée plusieurs fois, afin de sélectionner le plan le plus efficace, c'est-à-dire celui dont le déterminant est le plus élevé. La raison est que l'algorithme utilisé dans le programme SAS peut fournir des plans différents. L'explication de cette différence est donnée dans la section 2.7.

Les résultats des applications du plan D-optimal sur notre matrice d'expérience ont donc été effectués à l'aide du logiciel SAS. Nous avons lancé 15 fois la procédure de détermination des plans D-optimaux sur les 791 paires. Les plans à 20 choix récoltés sont les suivants :

	Plan 1		**Plan 2**		**Plan 3**		**Plan 4**	
	D-Efficiency= 68.7983		D-Efficiency= 68.7983		D-Efficiency=68.7320		D-Efficiency= 68.7983	
Choix	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2
1	1	198	1	199	1	214	1	199
2	1	215	1	214	1	181	1	214
3	13	247	13	246	1	245	1	232
4	1	262	1	263	13	256	17	261
5	37	232	37	232	37	229	49	246
6	53	261	53	261	37	263	37	263
7	97	258	97	259	101	263	97	259
8	73	253	73	253	73	260	73	253
9	1	172	1	172	1	164	1	172
10	5	156	5	156	17	178	5	156
11	5	59	5	58	5	60	5	58
12	13	62	13	63	13	67	13	63
13	5	106	5	107	1	84	1	81
14	169	255	169	254	149	261	145	260
15	37	81	37	81	41	108	41	107
16	181	260	181	260	205	258	205	254
17	229	264	229	264	229	261	229	264
18	221	263	221	262	217	254	221	262
19	1	21	1	21	1	21	1	21
20	1	29	1	29	1	?	1	29

	Plan 5=Plan 2		**Plan 6**		**Plan 7**		**Plan 8**	
	D-Efficiency=68.7983		D-Efficiency= 68.7983		D-Efficiency=68.7320		D-Efficiency= 68.5888	
Choix	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2
1	1	199	1	198	1	203	1	193
2	1	214	1	207	5	214	13	214
3	13	246	5	240	1	248	5	238
4	1	263	13	261	1	263	25	262
5	37	232	49	243	37	226	37	251
6	53	261	37	258	49	257	37	256
7	97	259	97	254	101	263	85	260
8	73	253	77	261	73	253	77	262
9	1	172	1	175	1	160	1	173
10	5	156	1	156	17	178	5	155
11	5	58	1	59	1	54	1	55
12	13	63	17	70	13	63	5	72
13	5	107	1	77	5	93	13	99
14	169	254	145	256	169	254	145	259
15	37	81	37	106	37	99	37	78
16	181	260	209	263	185	264	197	261
17	229	264	229	260	229	264	241	255
18	221	262	217	263	217	258	217	261
19	1	21	145	157	1	33	1	24
20	1	29	5	34	1	7	1	26

	Plan 9		**Plan 10**		**Plan 11**		**Plan 12=Plan 2**	
	D-Efficiency=68.7983		D-Efficiency=68.7983		D-Efficiency=68.7983		D-Efficiency=68.7983	
Choix	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2
1	1	203	1	199	1	213	1	199
2	1	211	1	181	1	187	1	214
3	13	247	1	232	17	249	13	246
4	29	261	1	258	1	264	1	263
5	41	240	41	249	37	231	37	232
6	37	258	37	263	49	258	53	261
7	85	262	85	259	101	264	97	259
8	73	253	97	262	73	254	73	253
9	1	178	1	172	1	168	1	172
10	1	148	17	180	5	177	5	156
11	1	50	5	58	5	58	5	58
12	17	70	1	72	1	64	13	63
13	1	98	13	105	1	104	5	107
14	145	264	173	263	169	253	169	254
15	37	81	37	80	37	82	37	81
16	193	255	193	254	185	263	181	260
17	241	260	221	262	229	263	229	264
18	221	263	217	255	217	257	221	262
19	1	17	1	21	1	17	1	21
20	5	35	1	29	13	28	1	29

	Plan 13=Plan 4		**Plan 14=Plan 4**		**Plan 15=Plan 1**	
	D-Efficiency=68.7983		D-Efficiency=68.7983		D-Efficiency=68.7983	
Choix	Produit 1	Produit 2	Produit 1	Produit 2	Produit 1	Produit 2
1	1	199	1	199	1	198
2	1	214	1	214	1	215
3	1	232	1	232	13	247
4	17	261	17	261	1	262
5	49	246	49	246	37	232
6	37	263	37	263	53	261
7	97	259	97	259	97	258
8	73	253	73	253	73	253
9	1	172	1	172	1	172
10	5	156	5	156	5	156
11	5	58	5	58	5	59
12	13	63	13	63	13	62
13	1	81	1	81	5	106
14	145	260	145	260	169	255
15	41	107	41	107	37	81
16	205	254	205	254	181	260
17	229	264	229	264	229	264
18	221	262	221	262	221	263
19	1	21	1	21	1	21
20	1	29	1	29	1	29

La valeur maximale du D-efficiency est 68.7983. Plusieurs plans ont cette valeur maximale d'efficacité. D'autres, tels que les plans 3, 7 et 8, ont une efficacité au dessous de cette valeur maximale. Par conséquent, ces derniers ne sont pas pris en considération dans le choix d'un plan. Dans les plans restants (D-efficiency maximum), certains sont similaires. En effet, le plan 1 est sorti deux fois dans le lancement de la procédure du plan D-optimal

(plan 15 = plan 1), alors que les plans 2 et 4 sont apparus trois fois (plan 5, 12 = plan 2 ; plan 13, 14 = plan 4). Quant aux plans 6, 9, 10 et 11, ils ne se sont manifestés qu'une fois.
Dans chacun de ces plans, certains produits apparaissent plusieurs fois. Le tableau suivant résume ces valeurs :

Plan 1		**Plan 2**		**Plan 4**		**Plan 6**	
Produit	Fréquence	Produit	Fréquence	Produit	Fréquence	Produit	Fréquence
1	6	1	6	1	7	1	6
5	3	5	3	5	2	5	2
13	2	13	2			37	2
37	2	37	2			145	2
2 fois		**3 fois**		**3 fois**		**1 fois**	

Plan 9		**Plan 10**		**Plan 11**	
Produit	Fréquence	Produit	Fréquence	Produit	Fréquence
1	7	1	8	1	7
37	2	37	2	5	2
				37	2
1 fois		**1 fois**		**1 fois**	

Parmi ces 7 plans, nous désirons en sélectionner qu'un. Notre critère de choix dépend du nombre de fois que le plan est apparu lors de la réalisation de la procédure du plan D-optimal, ainsi que de la variété des produits. Selon le tableau ci-dessus, nous remarquons que les plans 1, 2 ,6 et 11 proposent un grand nombre de fois les mêmes combinaisons d'attributs dans les paires de choix. De plus, le produit 1 semble être représenté abondamment dans tous les plans. Sa fréquence d'apparition est légèrement plus faible dans les plans 1, 2 et 6. Les plans 2 et 4 sont, quant à eux, ceux qui sont sortis le plus souvent.

Compte tenu de ces informations, il semble qu'il n'y a pas une évidence dans la sélection de l'un de ces plans. Cependant, nous pensons que le plan 4 comble mieux nos critères de choix, et ce malgré une fréquence du produit 1 légèrement supérieure. Cependant, ce désavantage n'est pas très flagrant et est donc négligeable. Quant aux autres critères, le plan 4 les satisfait au maximum. En effet, ce dernier apparaît un nombre maximum de fois et il n'a que deux produits qui sont représentés plusieurs fois dans les paires de choix.

Pour toutes ces raisons, nous avons donc opté pour le plan 4. Nous le redonnons de manière plus complète, en y ajoutant le vecteur des modalités des attributs correspondant aux produits présentés dans les choix.

Plan 4

D-Efficiency= 68.7983			**Attributs Produit 1**						**Attributs Produit 2**					
Choix	Produit 1	Produit 2	A	B	C	D	E	F	A	B	C	D	E	F
1	1	199	1	1	1	1	1	1	2	2	2	2	2	1
2	1	214	1	1	1	1	1	1	2	2	3	3	1	2
3	1	232	1	1	1	1	1	1	2	3	2	1	2	2
4	17	261	1	1	2	2	1	1	2	4	1	3	1	1
5	49	246	1	2	2	1	1	1	2	3	3	2	1	2
6	37	263	1	2	1	1	1	1	2	4	1	3	2	1
7	97	259	1	3	3	1	1	1	2	4	1	2	2	1
8	73	253	1	3	1	1	1	1	2	4	1	1	1	1
9	1	172	1	1	1	1	1	1	2	1	3	1	2	2
10	5	156	1	1	1	2	1	1	2	1	1	3	2	2
11	5	58	1	1	1	2	1	1	1	2	2	3	1	2
12	13	63	1	1	2	1	1	1	1	2	3	1	2	1
13	1	81	1	1	1	1	1	1	1	3	1	3	1	1
14	145	260	2	1	1	1	1	1	2	4	1	2	2	2
15	41	107	1	2	1	2	1	1	1	3	3	3	2	1
16	205	254	2	2	3	1	1	1	2	4	1	1	1	2
17	229	264	2	3	2	1	1	1	2	4	1	3	2	2
18	221	262	2	3	1	2	1	1	2	4	1	3	1	2
19	1	21	1	1	1	1	1	1	1	1	2	3	1	1
20	1	29	1	1	1	1	1	1	1	1	3	2	1	1

Dans la suite de notre travail, nous utilisons ce plan afin de construire un questionnaire, dont le but est d'obtenir les valeurs de la variable dépendante de notre modèle des préférences individuelles.

5.4 Mise en oeuvre de l'interview

L'interview consiste à faire passer un questionnaire, durant le cours du Professeur G. Antille, aux étudiants de première année en Sciences Sociales de l'Université de Genève. Ce questionnaire est issu du résultat du plan D-optimal représenté dans le tableau de la section précédente. En associant les modalités des attributs à ce plan D-optimal, nous pouvons ainsi illustrer les produits de chaque paire et créer le questionnaire. L'ordre des choix doit cependant varier d'un individu à l'autre, afin d'assurer un résultat non biaisé. Dès que les interviewés ont passé le questionnaire, il s'en suit une étape de tri des observations. Le résultat fournit ainsi notre base de données permettant l'estimation de notre spécification de l'option "zéro".

La suite de cette section consiste donc à décrire en détail les étapes de la réalisation du questionnaire, de l'interview et de la création de la base de données.

5.4.1 Le questionnaire

Le questionnaire, dont sont soumis les étudiants des Sciences Sociales, est issu du plan D-optimal. En utilisant la description des modalités des attributs sélectionnées dans la section 5.1, et en nous basant sur le codage du plan retenu dans la section 5.3, nous pouvons définir les produits associés à chaque choix. Ces derniers sont décrits dans l'annexe 4.

Afin de nous assurer de la fiabilité des réponses des interviewés, nous avons ajouté un choix supplémentaire dans le questionnaire. Ce choix propose des produits dont l'un est très clairement supérieur à l'autre. Dans ces produits, les modalités non ordonnées sont identiques. Quant aux modalités ordonnées, les valeurs supérieures en terme de préférence sont regroupées dans un des produits et dans l'autre, ces valeurs sont inférieures. Ce choix, appelé question test, est le choix 21 de l'annexe 4.

Dans cette présentation du questionnaire (annexe 4), nous remarquons que, dans chaque choix, les produits issus du plan D-optimal sont très différents et concordent avec la définition de l'analyse conjointe. En effet, le but de l'analyse conjointe est de présenter des produits qui diffèrent en grand nombre de points, afin d'amener le consommateur à un choix par compromis. Le critère de sélection des plans D-optimaux permet d'obtenir ces différences dans les produits. En effet, étant donné que les plans D-optimaux proviennent de la maximisation du déterminant, ce résultat est donc atteint lorsque les modalités des attributs sont très éloignées les unes des autres.

Les notations associées à certaines variables des produits peuvent ne pas être assez claires

pour l'interviewé. Ces variables, telles que la composition et le niveau d'ancienneté, doivent être chiffrées de manière plus précise. En effet, un site de VoD qui contient 75% de films et 25% de séries n'apporte pas une idée précise et claire du contenu du site. Etant donné que nous connaissons le nombre de programmes par choix et le nombre de nouveautés films-séries (DVDnet), c'est-à-dire les valeurs des nouveautés sont de 140 pour les films et de 96 pour les séries, nous pouvons donc modifier cette notation en échangeant les pourcentages par des chiffres. Par exemple, dans le choix 1, nous savons que le produit 1 ou le site de VoD 1 propose 1000 programmes ne contenant que des films (100% Films) dont toutes les nouveautés existantes y font partie. Il existe 140 nouveautés films. Nous pouvons donc exprimer le contenu du site comme étant composé de 140 nouveautés films et de 860 films anciens (1000-140). De plus, afin que l'interviewé situe la nouveauté film ou série dans le temps, nous introduisons à côte des ces dernières une date de sortie du programme. Tous les produits du questionnaire ont donc ainsi été modifiés.

Afin de donner au questionnaire un aspect ressemblant à un site de VoD, nous avons réalisé une page d'avant-garde représentant une plate-forme de site fictif de Vidéo à la Demande. Quant à l'aspect des autres pages du questionnaire, nous avons gardé le contour de l'avant-garde et avons décrit, à l'intérieur de ces dernières, les paires de produits.
Nous avons également ajouté dans ce questionnaire une page sur les données personnelles des étudiants. Le but de cette manoeuvre a été de rendre l'interview plus sérieux et plus réaliste pour les étudiants et de repérer les individus qui n'ont jamais utilisé internet. La raison est que ces individus sont moins propices à connaître la VoD et les technologies associées à internet et à ce type de service. De plus, dans ces données personnelles, il y a également des renseignements sur le niveau de connaissances technologiques des étudiants, telles que la notion d'ADSL, de débit internet et de VoD.

Un aspect important a également été ajouté aux paires de produits. Il s'agit de l'option "zéro", plus précisément du refus et du conflit. Le refus étant caractérisé par une non appréciation des produits, il a donc été introduit dans le questionnaire par le biais de la case "Aucun". Cette appellation correspond à l'idée qu'aucun produit de la paire ne plaît à l'interviewé. Quant au conflit, il représente l'incertitude du consommateur dans le choix. L'individu ne peut donc pas se déterminer à choisir un produit. Afin de reconnaître cette situation de conflit, nous proposons à l'individu la case de sélection des deux produits, appelée "Les deux". Nous supposons que dans les situations de conflit, le consommateur préfère opter pour les deux produits.

L'aspect général du questionnaire soumis aux interviewés est finalement présenté en annexe 5.

La fiche représentant les données personnelles a été déplacée au début du questionnaire durant l'interview. A l'aide du programme SAS, nous avons lancé 171 fois un plan aléatoire permettant de changer l'ordre des choix pour les 171 questionnaires conçus. Chaque individu a donc eu un ordre différent de présentation des questions. Cette manoeuvre sert à éviter les biais liés à l'influence des réponses dues aux choix précédents. En variant l'ordre, l'influence diffère d'un individu à l'autre et enlève le biais lié à cette situation.

5.4.2 L'interview

Lors de l'interview, nous avons dû donner certaines précisions. En effet, avant de soumettre ce questionnaire aux étudiants, nous leur avons expliqué ce que nous entendions par la case "Aucun" et "Les deux". Les explications sont que le choix "Aucun" correspond au fait qu'aucun des 2 produits présentés ne convient et "Les deux" à la situation amenant l'étudiant à ne pas pouvoir départager ces 2 produits. Quant aux produits 1 et 2, le fait de cocher une de ces cases sous-entend qu'un de ces produits correspond à la demande de l'étudiant en terme de VoD.
Ces indications ont été conçues de sorte qu'elles n'influent pas sur le choix des étudiants. Nous avons également brièvement décrit la Vidéo à la Demande en disant qu'elle consiste à visionner un programme (films, séries, etc.) dès que l'individu le décide. Ce programme peut être visionné sur l'ordinateur à travers l'ADSL ou sur un téléviseur, comme en France. Quant aux directives liées à cet interview, nous avons simplement précisé que ce questionnaire a pour but de cerner les caractéristiques les plus importantes pour chacun d'entre eux dans le choix d'un produit et qu'il leur suffit de donner leurs préférences parmi des produits potentiels. Ces produits sont des sites de VoD et il faut, à chaque page, cocher une des 4 propositions.

5.4.3 Résultat de l'interview et données retenues

L'interview a duré à peu près 20 minutes. Il y a eu 120 étudiants qui ont répondu au questionnaire. Parmi ces 120 étudiants, nous avons trié les questionnaires des interviewés, de sorte à ne retenir que les réponses satisfaisantes. Nous avons dû éliminer certains questionnaires dont les réponses n'étaient pas cohérentes. Ces réponses incohérentes étaient soit :

- incomplètes, c'est-à-dire certaines ou même toutes les questions n'ont pas été cochées
- multiple, c'est-à-dire dans certains choix, plusieurs cases ont été sélectionnées
- incorrectes vis-à-vis de la question test, c'est-à-dire que :
  - l'individu a choisi le produit 2, qui est nettement inférieur au produit 1

  - l'individu a choisi la case, aucun des deux ou les deux, ce qui n'a pas de sens, compte tenu qu'un des produits est clairement supérieur à l'autre
- erronées, c'est-à-dire que l'individu a sélectionné la même case dans tous les choix ou dans la majorité des questions. Dans les questionnaire où la même réponse a été choisie, la case sélectionnée est aucun des produits me plaît. Ce type de réponse signifie que l'étudiant n'est pas intéressé par la VoD et n'est donc pas un client potentiel à ce type de service. Il apporte donc aucun renseignement dans l'estimation des préférences de la VoD.

En supprimant les questionnaires incohérents, nous obtenons une base de données de 74 individus. Cette réduction est assez conséquente, étant donné que nous avions 120 étudiants qui avaient participé au sondage. Les raisons peuvent être multiples. Elles peuvent provenir, par exemple, du fait que nous ne proposions pas en échange de ce questionnaire un cadeau par tirage au sort pour motiver les interviewés. Une autre raison serait simplement que certains étudiants n'ont pas trouvé intéressant de répondre à ce sondage, parce qu'ils n'ont peut-être pas de télévision, qu'ils vont très peu sur internet, etc.

Avec les 74 questionnaires, nous avons associé les réponses données lors de cette interview aux vecteurs des attributs. Ces vecteurs ont été modifiés afin d'être exprimés sous la forme de variables indicatrices. Il a fallu ensuite supprimer les modalités de référence, à savoir la dernière modalité de chaque attribut. Ces modalités de référence sont les suivantes :

Modalités
Attribut A : Le nombre de programmes
600 (2)
Attribut B : La composition du site en terme de films-séries
100% Séries (4)
Attribut C : La composition du site en terme de nouveautés-anciens
Que des anciens (3)
Attribut D : Tarif
Abonnement (3)
Attribut E : Temps de conservation du film après le premier visionnage
48 heures (2)
Attribut F : Disponibilité de la bande annonce ou de l'extrait
Aucun disponible (2)

Dans les questionnaires retenus, les étudiants n'ont pas tous coché la case "conflit" du questionnaire. En effet, nous avons constaté les combinaisons de réponses suivantes :
- Produit 1 et 2 ($y = 1, 2$) (3 individus)
- Produit 1, 2 et refus ($y = 1, 2, 3$) (45 individus)
- Produit 1, 2, refus et conflit ($y = 1, 2, 3, 4$) (22 individus)

– Produit 1, 2 et conflit ($y = 1, 2, 4$) (3 individus)
– Produit 1 et refus ($y = 1, 3$) (1 individu)

Selon ces combinaisons, le modèle d'estimation est différent. Cet aspect est détaillé dans le chapitre 3, section 3.5 Cas spéciaux.
La base de données, servant à cet estimation de notre spécification de l'option "zéro", contient les 20 choix des 74 interviewés (1480 données), ainsi que les vecteurs des variables indicatrices des modalités des attributs (sans les variables de référence).
La suite de l'étude consiste à estimer le modèle de l'option "zéro" avec ces données et de commenter ces résultats.

5.5 Application du modèle de l'option "zéro" avec le refus et le conflit

5.5.1 Le maximum de vraisemblance

Nous avons estimé les paramètres de notre spécification individuelle de l'option "zéro" avec la méthode du maximum de vraisemblance. Cependant, l'output donnait l'avertissement "Failure to improve objective function after x itérations".

Cette non convergence et la petite taille des échantillons individuels nous laisse supposer que les données individuelles présentent une séparation. Par conséquent, dans la sous-section suivante, nous effectuons une analyse discriminante factorielle et un tableau de classement dans SAS afin de vérifier cette séparation.

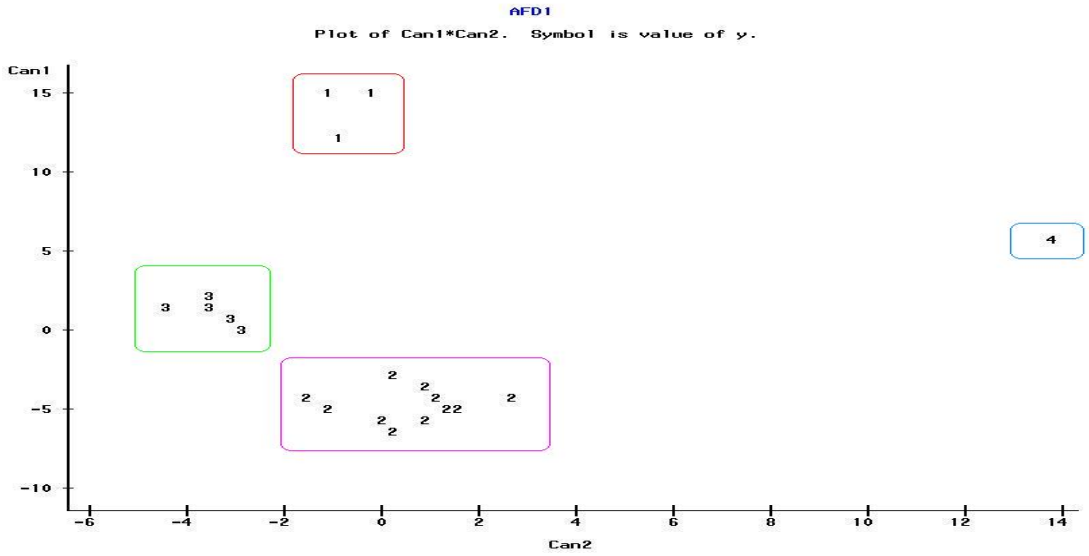
Quant aux données agrégées, l'estimation combinée par maximum de vraisemblance des différents modèles, représentant les combinaisons des alternatives cochées dans les questionnaires, ne présente pas de séparation comme l'illustre la sous-section suivante. Par conséquent, nous avons pu estimer les coefficients de notre modèle de l'option "zéro". Les résultats se trouvent dans la sous-section 5.6.6 Le modèle agrégé.

5.5.2 Vérification de la séparation

Nous vérifions l'existence de séparation complète ou quasi-complète à l'aide de l'analyse factorielle discriminante (AFD). Nous avons effectué cette analyse pour chaque individu, ainsi que pour le modèle agrégé. Les graphiques AFD des 74 individus présentent tous de la séparation. Ils sont issus du programme SAS. Afin d'illustrer cette séparation, nous ne présentons que le graphique AFD d'un seul des 74 individus. Les commentaires valent

également pour les autres individus.

Par exemple, le graphique de l'analyse factorielle discriminante de l'individu 1 se présente de la manière suivante :



Nous remarquons une nette division entre les différentes alternatives proposées dans le questionnaire. L'échantillon présente clairement une séparation des données. Cette séparation se vérifie dans le tableau de classement des alternatives, dans lequel les éléments en dehors de la diagonale sont nuls :

Number of Observations and Percent Classified into y

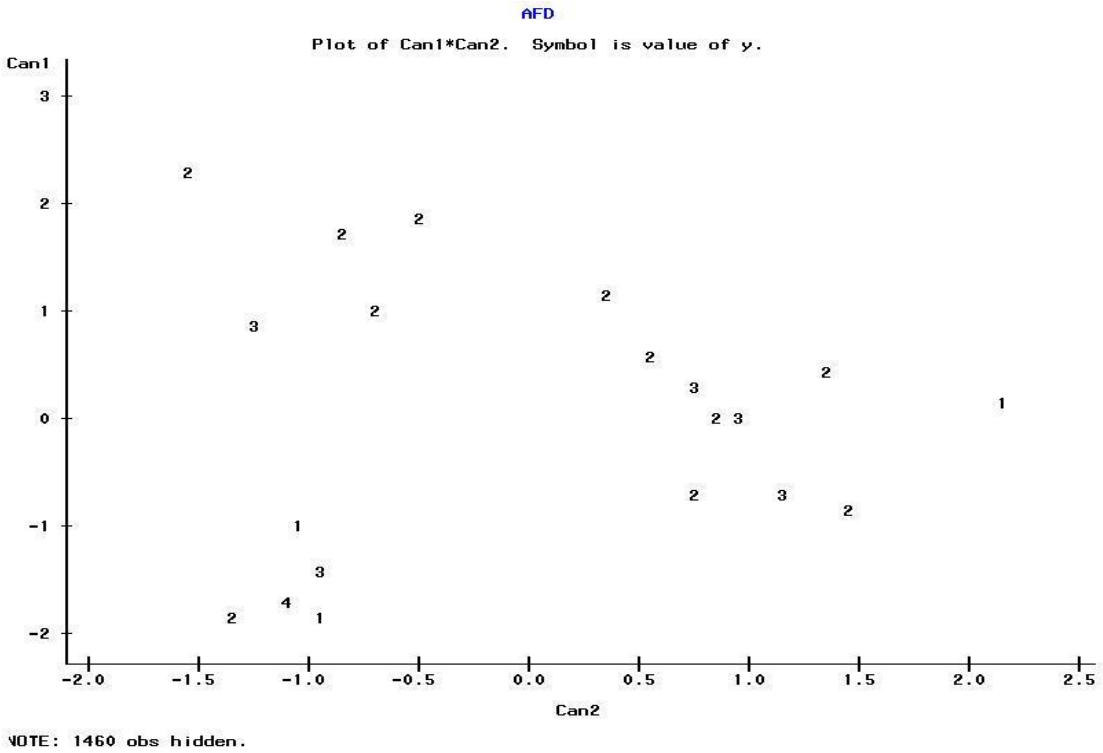
From y	1	2	3	4	Total
1	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
2	0 0.00	11 100.00	0 0.00	0 0.00	11 100.00
3	0 0.00	0 0.00	5 100.00	0 0.00	5 100.00
4	0 0.00	0 0.00	0 0.00	1 100.00	1 100.00
Total	3 15.00	11 55.00	5 25.00	1 5.00	20 100.00

Pour certains individus, toutes les observations ne se trouvent pas sur le graphique AFD. La raison est que certaines de ces observations sont cachées par d'autres. Dans de tels

cas, nous pouvons nous assurer de la séparation en créant la tableau de classement dans le logiciel SAS. Les tableaux de classement correspondant aux autres individus montrent la même séparation que celui de l'individu 1 et ne sont par conséquent pas présentés dans ce travail.

La raison la plus probable de cette séparation est la quantité faible de données par individu. Dans l'analyse conjointe, ces situations sont fréquentes, étant donné qu'il faut respecter la fiabilité des réponses des interviewés. En effet, proposer un nombre de choix élevé pourrait biaiser les réponses. En règle générale, un échantillon de 50 observations a une probabilité élevée d'obtenir des réponses présentant une séparation. De plus, la quantité de paramètres inconnus du modèle (12) en comparaison au 20 choix proposés influe également sur la séparation. Ce rapport étant faible, c'est pourquoi les données comportent la séparation.

Quant au modèle agrégé, il n'a pas de séparation dans ces données comme l'illustre le graphique issu de l'analyse factorielle discriminante AFD (SAS) :



De plus, étant donné que 1460 observations sont cachées dans le graphique, nous vérifions la non séparation à l'aide du tableau de classement. Les cases non nulles du tableau de classement confirment la non séparation dans les données du modèle agrégé :

Number of Observations and Percent Classified into y

From y	1	2	3	4	Total
1	283	36	197	20	536
	52.80	6.72	36.75	3.73	100.00
2	44	249	130	42	465
	9.46	53.55	27.96	9.03	100.00
3	29	71	317	6	423
	6.86	16.78	74.94	1.42	100.00
4	14	14	22	6	56
	25.00	25.00	39.29	10.71	100.00
Total	370	370	666	74	1480
	25.00	25.00	45.00	5.00	100.00

Le fait que les modèles individuels ne comportent pas beaucoup de données (20) influe sur la séparation dans les données. Par contre, le modèle agrégé qui compte 1480 données, n'a pas de séparation dans ces données. Par conséquent, l'estimation par maximum de vraisemblance de ce dernier est faisable. Ces résultats sont utiles dans l'estimation individuelle des données comme l'illustre la suite du travail.

5.5.3 L'approche bayésienne et les simulations de Monte Carlo dans l'option "zéro"

Tel que nous venons de le souligner dans la sous-section précédente, nos données individuelles contiennent de la séparation et rendent difficiles l'estimation par le maximum de vraisemblance. Par conséquent, nous devons contourner cette méthode avec une approche adaptée aux situations présentant la séparation. Dans le chapitre 4 sur la notion de séparation et de convergence, nous avons abordé ces différentes méthodes. L'approche conseillée dans la littérature concerne le maximum de vraisemblance pénalisée. Par conséquent, nous avons essayé de l'appliquer dans notre spécification de l'option "zéro".

Implémenter la vraisemblance pénalisée demande l'intégration du logarithme du déterminant de la matrice d'information de Fisher dans l'équation de la log-vraisemblance. La structure de la matrice d'information de Fisher pour les modèles multinomiaux est très complexe en comparaison avec celle du modèle dichotomique. De plus, contrairement à la spécification logit, les probabilités du modèle probit sont également plus difficiles

dans leur formulation. Néanmoins, les programmes des logiciels statistiques contiennent généralement des commandes facilitant la transformation d'une variable en fonction de la répartition normale. Suite à plusieurs essais, nous nous sommes rendus compte que l'implémentation du déterminant de la matrice d'information de Fisher du modèle multinomial n'était pas évidente. Par conséquent, nous avons opté pour une autre approche dans l'estimation des paramètres de la spécification de l'option "zéro", c'est-à-dire l'approche bayésienne.

Dans les cas où l'individu n'avait pas choisi le refus et le conflit, la variable dépendante du modèle étant dichotomique ($y = 1, 2$), nous avons quand même pu utiliser l'implémentation de la méthode de la vraisemblance pénalisée dans le logiciel S-PLUS (logistf) pour estimer les paramètres du modèle. Pour l'individu qui n'a choisi que le produit 1 et le refus, l'estimation de maximum de vraisemblance pénalisée n'a pas pu être effectuée, étant donné que la matrice $X'X$ n'est pas de rang complet. Nous avons également estimé les paramètres du modèle $y = 1, 2$ avec l'approche bayésienne décrite par la suite. Une comparaison des résultats des deux approches est commentée dans la suite du travail. Nous avons essayé de recréer les résultats de S-PLUS en implémentant nous-mêmes la vraisemblance pénalisée du modèle $y = 1, 2$ dans un autre programme TSP-GiveWin, afin de comprendre le processus et de l'appliquer ensuite dans le cas multinomial. Cependant, nos résultats ne correspondaient pas à ceux de S-PLUS. Par conséquent, nous avons choisi l'autre méthode d'estimation des paramètres du modèle de l'option "zéro", c'est-à-dire l'approche bayésienne.

L'approche bayésienne suppose une connaissance de la loi *a priori* des β. Etant donné que le modèle agrégé ne contient pas de séparation, telle que l'illustre la sous-section 5.5.2, nous avons donc utilisé cette information dans la distribution *a priori*. Les valeurs d'estimation du modèle agrégé, ainsi que sa matrice de variance covariance ont été utilisées comme les moyennes et les variances des paramètres β d'une loi normale. La loi *a priori* des β se formule donc comme suit :

$$\beta \sim N(\hat{\beta}_{\text{agrégé}}, \hat{\Sigma}_{\text{agrégé}})$$

où $\hat{\beta}_{\text{agrégé}}$ correspond aux moyennes du modèle agrégé et $\hat{\Sigma}_{\text{agrégé}}$ à la matrice de variance covariance du modèle agrégé.

L'approche bayésienne appliquée à notre spécification de l'option "zéro" consiste à définir la distribution *a posteriori* de $\beta|Y$ par individu à partir de la loi *a priori* de β et la loi connue de $Y|\beta$ de notre spécification de l'option "zéro" :

$$f(\beta|Y) = \underbrace{\prod_{i=1}^{20}\prod_{h=1}^{4} P(y_i = h)^{y_{hi}}}_{P(Y|\beta)} \varphi(\beta)$$

où $y_{hi} = \begin{cases} 1 & \text{si l'individu } i \text{ choisit l'option } h \\ 0 & \text{sinon} \end{cases}$

et $\varphi(\beta)$ l'expression de la fonction de densité d'une loi normale multivariée d'espérance $\hat{\beta}_{\text{agrégé}}$ et de matrice de variance covariance $\hat{\Sigma}_{\text{agrégé}}$. Etant donné que dans le modèle où $y = 1, 2, 3, 4$, nous avons 12 paramètres inconnus, cette expression de la fonction de densité s'écrit :

$$\varphi(\beta) = (2\pi)^{-\frac{12}{2}} \left|\hat{\Sigma}_{\text{a}}\right|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\beta-\hat{\beta}_{\text{a}})'\hat{\Sigma}_{\text{a}}^{-1}(\beta-\hat{\beta}_{\text{a}})}$$

où $\hat{\beta}_{\text{a}}$ correspond à $\hat{\beta}_{\text{agrégé}}$ et $\hat{\Sigma}_{\text{a}}$ à $\hat{\Sigma}_{\text{agrégé}}$.

La formulation de $P(Y|\beta)$ dans l'expression $f(\beta|Y)$ correspond à un individu qui a sélectionné les produits 1 et 2, le refus et le conflit dans les 20 choix du questionnaire. Pour obtenir l'expression de cette dernière dans les autres situations, c'est-à-dire $y = 1, 2$, $y = 1, 2, 3$, $y = 1, 2, 4$ et $y = 1, 3$, il suffit de remplacer l'expression $P(Y|\beta)$ par le produit des probabilités des produits 1 et 2 dans le modèle $y = 1, 2$, par le produit des probabilité des produits 1 et 2 et du refus dans le modèle $y = 1, 2, 3$, etc. L'expression $(2\pi)^{-\frac{12}{2}}$ dans $\varphi(\beta)$ change également d'un modèle à l'autre. En effet, par exemple, dans le modèle $y = 1, 2, 3$, le nombre de paramètres inconnus s'élève à 11, étant donné que le coefficient associé au conflit n'intervient pas dans cette spécification. Par conséquent, la formulation $(2\pi)^{-\frac{12}{2}}$ du $y = 1, 2, 3, 4$ devient $(2\pi)^{-\frac{11}{2}}$ dans le modèle $y = 1, 2, 3$. Les mêmes modifications sont effectuées pour les autres spécifications. Le vecteur d'espérance $\hat{\beta}_{\text{a}}$ et la matrice de variance covariance $\hat{\Sigma}_{\text{a}}$ ont également été transformés pour chaque situation afin d'obtenir uniquement les éléments de β qui influent sur le modèle. Par exemple, dans le modèle $y = 1, 2, 3$, nous avons enlevé la ligne du vecteur $\hat{\beta}_{\text{a}}$ correspondant au paramètre expliquant le conflit et les lignes et les colonnes de la matrice de variance covariance $\hat{\Sigma}_{\text{a}}$ de ce même coefficient.

La distribution *a posteriori* $f(\beta|Y)$ n'étant pas clairement définissable, nous avons donc choisi d'appliquer des simulations de Monte Carlo pour calculer les estimations des pa-

ramètres inconnus de notre spécification de l'option "zéro" par l'approche bayésienne. Plusieurs critères d'estimation sont envisageables, c'est-à-dire l'estimation par le mode, la moyenne et la médiane. Nous avons opté pour le mode, étant donné que nous recherchons le maximum de la fonction $f(\beta|Y)$. La moyenne est également un bon estimateur du maximum, à condition que la distribution soit symétrique. Or, nous ne savons pas si cette dernière est symétrique. Dans l'analyse du maximum pénalisé, nous savons que la fonction de vraisemblance pénalisée n'est pas symétrique et que par conséquent, les tests statistiques et les intervalles de confiance de Wald ne sont pas corrects. En conclusion, nous avons pensé qu'il est préférable de prendre le mode, qui correspond au maximum, quelque soit la forme de la distribution.

Les simulations Monte Carlo, dans le modèle $y = 1, 2, 3, 4$, ont consisté à générer 10'000 vecteurs β de paramètres inconnus issus de la loi normale multivariée d'espérance $\hat{\beta}_{\mathrm{a}}$ et de matrice de variance covariance $\hat{\Sigma}_{\mathrm{a}}$. Ces 10'000 échantillons ont permis de calculer les valeurs de la distribution *a posteriori* $f(\beta|Y)$. De toutes ces valeurs, nous avons sélectionné le nombre le plus élevé et nous avons recherché la valeur du vecteur β, qui a généré ce maximum. Cette valeur correspond aux estimations des paramètres inconnus β. Les résultats et les commentaires se trouvent dans la suite du travail.
A quelques modifications près, nous avons procédé de la même manière pour les autres modèles correspondant aux différentes réponses fournies par les interviewés.

5.6 Résultats

Dans cette section, nous présentons et commentons les résultats des estimations individuelles et agrégées de notre spécification de l'option "zéro". Les résultats sont regroupés selon les combinaisons des réponses fournies par chaque interviewé. Ces combinaisons sont, par exemple, le choix uniquement des produits 1 et 2 dans le questionnaire ($y = 1, 2$) ou le choix des produits 1 et 2 et du refus ($y = 1, 2, 3$), etc. Les résultats découlent de l'application de l'approche bayésienne sur notre modèle de l'option "zéro". L'implémentation de cette méthode a été exécutée sur le logiciel Matlab. Cette approche ne fournissant pas les écart-types des paramètres inconnus, nous avons donc utilisé une estimation qui consiste à associer à l'écart-type, la valeur de la moitié de l'étendue des quantiles, correspondant à 25% et à 75% des observations de la distribution *a posteriori*. Cette méthode permet, en réalité, de déterminer la valeur associée aux 25% des observations en dessus et en-dessous du mode. Dans une loi normale, l'écart-type correspond à 15.87% et 84,13% des observations ou 15.87% en dessous et en dessus de la moyenne ou du mode, car la loi normale est symétrique. Etant donné que nous ne savons pas si la distribution *a posteriori* est une loi normale, nous avons donc opté pour une estimation des écart-types par la méthode des 25% et des 75% des observations.

L'estimation des 25% et des 75% de la distribution s'effectue en simulant des valeurs à partir de la loi *a posteriori* et en les triant dans le but de connaître celles qui se situent à 25% et à 75% des observations. Cependant, ne connaissant pas la loi *a posteriori*, nous avons donc procédé différemment. Plutôt que de trier les valeurs des β, nous avons trié celles de la loi *a posteriori* $f(\beta|Y)$ pour les dessiner en fonction de chaque β (annexe 6). Ensuite, nous avons situé la valeur de $f(\beta|Y)$ correspondant aux 25% des observations en dessus et en dessous du mode[5]. Nous avons reporté la valeur des β associée à celle des $f(\beta|Y)$ des 25% des observations en dessus et en dessous du mode, afin d'obtenir les étendues de chaque paramètre. Sur les graphiques (annexe 6), nous avons représenté le mode par un trait vert. Les valeurs reportées des 25% en dessus et en dessous de ce mode (points-traits rouges) nous permettent de calculer l'étendue, c'est-à-dire la différence des quantiles ou des β associés aux 25% en dessus et en dessous du mode. L'écart-type correspond à la moitié de l'étendue. L'implémentation de cette procédure et les graphiques ont été exécutés sur Matlab.

Dans la suite de cette section, nous illustrons les résultats de l'estimation de l'option "zéro" à l'aide de tableaux présentant la valeur estimée des paramètres, leurs écart-types et l'importance de chaque attribut. Cette dernière valeur est calculée selon la formule

[5]Le mode correspond à l'estimateur de nos paramètres inconnus.

décrite à la page 32 du chapitre des plans d'expérience.
Nous présentons également les parts de marché de deux produits réels, à savoir les sites Canalplay et TF1Vision et celles d'un produit fictif appelé Idéal. Les modalités définissant nos produits réels reflètent les caractéristiques disponibles au début de notre étude (octobre 2006). Par conséquent, certaines de ces modalités peuvent ne pas être adaptées aux caractéristiques actuelles, étant donné que ces dernières ont évolué. Cependant, les résultats apportent, tout de même, certaines informations utiles dans la mise en place de ce type de produit.

5.6.1 Modèle y=1,2

Dans ce modèle, les interviewés ont choisi seulement les produits 1 et 2, c'est-à-dire deux options sur quatre. Etant donné que les données présentent de la séparation, nous avons donc dû opter pour une autre méthode que celle du maximum de vraisemblance. L'estimation des utilités partielles par l'approche bayésienne et le calcul des écart-types par la procédure des 25% et 75% des observations ont fourni les résultats pour un des individus du modèle $y = 1, 2$ suivants :

Individu 22

Variable	Utility	Standard Error	Importance (% Utility Range)
Catalogue :1000	0.173	0.069	4.32%
Catalogue :600	-0.173	0.069	
Composition :100% Films	0.315	0.095	33.82%
Composition :75% Films 25% Séries	0.858	0.102	
Composition :50% Films 50% Séries	0.676	0.076	
Composition :100% Séries	-1.850	0.274	
Nouveauté :max	0.702	0.0619	25.46%
Nouveauté :moitié	0.633	0.059	
Ancien :max	-1.336	0.121	
Tarif :payant	-0.224	0.054	31.54%
Tarif :gratuit pub	1.375	0.059	
Tarif :abonnement	-1.150	0.113	
Conservation :24h	-0.120	0.052	2.99%
Conservation : 48h	0.120	0.052	
Disponibilité : Bande annonce	-0.073	0.059	1.84%
Disponibilité : Aucun	0.073	0.059	

Les signes des modalités des attributs des individus du modèle $y = 1, 2$ semblent, en général, correspondre à la réalité. En effet, nous constatons une préférence pour la location de 48 heures, pour un catalogue de 1000 programmes et pour la gratuité des émissions avec de la publicité. De plus, un site composé d'un maximum de nouveautés et de 75% de films et 25% de séries semble prévaloir dans les goûts de ces individus. Quant à la bande annonce, elle ne semble pas primordiale dans le choix du site de VoD. Les attributs qui influencent essentiellement le choix des sites de Vidéo à la Demande sont la composition, le niveau de nouveauté et le tarif. En effet, selon le pourcentage d'importance, ces attributs sont déterminants dans la sélection d'un site de Vidéo à la Demande. Par contre, la quantité de programme, la durée de conservation ou la disponibilité de la bande annonce ne semblent pas avoir d'importance dans le choix du site pour ces individus.

Les écart-types sont raisonnables vis-à-vis des utilités partielles estimées. En effet, ils sont inférieurs à la valeur estimée des paramètres inconnus. Dans l'ensemble, les résultats d'estimation par l'approche bayésienne du modèle $y = 1, 2$ sont donc satisfaisants.

Les résultats d'estimation des autres individus du modèle $y = 1, 2$ sont très similaires. Par conséquent, nous ne les mettons pas dans ce travail. Quant aux commentaires, ils s'appliquent également pour tous les autres individus du modèle $y = 1, 2$. Pour les autres catégories de modèles, nous procédons de la même manière, c'est-à-dire que nous n'illustrons les résultats qu'à l'aide d'un individu dans chaque groupe, étant donné qu'en général, les estimations sont très similaires d'un individu à l'autre. De plus, les commentaires d'un individu de chaque catégorie de modèles s'appliquent également pour tous les autres individus du même groupe.

Etant donné que le modèle $y = 1, 2$ est dichotomique, il existe des logiciels tels que S-Plus dans lesquels la méthode du maximum de vraisemblance pénalisée est implémentée. Nous avons voulu la tester, afin de comparer les résultats avec ceux de l'approche bayésienne. La raison est que la littérature référence cette approche comme étant la plus adaptée dans le cadre d'un modèle à variable dépendante binaire logit. Notre spécification de l'option "zéro" étant probit, les propriétés de la méthode du maximum de vraisemblance pénalisée ne sont peut-être pas applicables à ce type de modèle.

Les résultats d'estimation de l'individu 22 avec le logiciel S-Plus du maximum de vraisemblance pénalisée sont dans le tableau suivant :

Individu 22

Variable	Utility	Standard Error	Importance (% Utility Range)
Catalogue :1000	-0.792	1.127	9.52%
Catalogue :600	0.792	1.127	
Composition :100% Films	-0.626	1.960	36.76%
Composition :75% Films 25% Séries	-1.976	1.890	
Composition :50% Films 50% Séries	-1.528	1.754	
Composition :100% Séries	4.131	5.605	
Nouveauté :max	-2.172	1.411	37.07%
Nouveauté :moitié	-1.816	1.141	
Ancien :max	3.988	2.553	
Tarif :payant	0.334	1.273	5.42%
Tarif :gratuit pub	-0.579	1.045	
Tarif :abonnement	0.245	2.319	
Conservation :24h	0.698	1.119	8.31%
Conservation :48h	-0.698	1.119	
Disponibilité : Bande annonce	0.242	1.092	2.89%
Disponibilité : aucun	-0.242	1.092	

Likelihood ratio test=9.925 on 10 df, p=0.447, n=20

Les résultats issus du maximum de vraisemblance pénalisée ne sont pas cohérents en terme de signe. Par exemple, l'utilité partielle de la modalité 1000 de l'attribut catalogue de l'individu 22 est négative, alors que le signe de la modalité 600 est positive. Il est plus probable que l'individu 22 préfère avoir 1000 programmes dans un site de VoD que 600. De plus, les individus semblent préférer des sites de VoD payants, plutôt que gratuits. Cette constatation est également peu probable. La durée de conservation ne semble également pas refléter la réalité, étant donné que le signe des modalités de cet attribut donne une préférence à la modalité 24h, plutôt que celle de 48h.

Les écart-types sont plus élevés que les utilités partielles. Ce fait révèle une mauvaise adéquation de l'approche du maximum de vraisemblance pénalisée dans notre spécification

de l'option "zéro". Par conséquent, la méthode bayésienne est plus adaptée à notre modèle que la vraisemblance pénalisée.

5.6.2 Modèle y=1,2,3,4

Les individus ayant choisi les 4 alternatives, c'est-à-dire le produit 1 et 2, le refus et le conflit, nous intéressent particulièrement. En effet, dans ce cas, le modèle intègre les deux types d'option "zéro". De plus, nous pouvons ainsi constater si le conflit apporte une explication supplémentaire dans le comportement des préférences des consommateurs.

L'estimation des paramètres inconnus a été effectuée par l'approche bayésienne et celle des écart-types par la procédure des 25% et des 75% des observations. Les résultats pour un des individus du modèle $y = 1, 2, 3, 4$ sont illustrés dans le tableau suivant :

Individu 2

Variable	Utility	Standard Error	Importance (% Utility Range)
Refus :δ	1.271	0.083	
Conflit :δ'	0.476	0.036	
Catalogue :1000	0.210	0.056	5.55%
Catalogue :600	-0.210	0.056	
Composition :100% Films	0.240	0.084	31.63%
Composition :75% Films 25% Séries	0.751	0.097	
Composition :50% Films 50% Séries	0.654	0.078	
Composition :100% Séries	-1.645	0.260	
Nouveauté :max	0.658	0.067	25.05%
Nouveauté :moitié	0.581	0.073	
Ancien :max	-1.240	0.140	
Tarif :payant	-0.224	0.054	33.39%
Tarif :gratuit pub	1.377	0.060	
Tarif :abonnement	-1.153	0.114	
Conservation :24h	-0.094	0.056	2.48%
Conservation : 48h	0.094	0.056	
Disponibilité : Bande annonce	-0.071	0.058	1.88%
Disponibilité : Aucun	0.071	0.058	

Les résultats sont très semblables pour les autres individus du modèle $y = 1, 2, 3, 4$. Par conséquent, les commentaires peuvent également être appliqués pour les autres individus de ce modèle.

Nous constatons, de manière générale, que les signes des coefficients de tous les individus du modèle $y = 1, 2, 3, 4$, ainsi que leur grandeur, reflètent la réalité et que les valeurs des écart-types sont convenables vis-à-vis des utilités partielles. En effet, dans le cas d'attributs ordonnés, tels que la quantité de programmes, la durée de conservation et le tarif entre un site payant et gratuit, il semble évident qu'un individu préférera avoir un nombre de programmes ou une durée de conservation les plus élevés ou qu'il choisira un site offrant des films gratuitement plutôt que payants. Avec les résultats de l'estimation par l'approche bayésienne, nous constatons que la modalité 24h de l'attribut conservation, respectivement la modalité 600 de l'attribut quantité de programmes, est systématiquement inférieure à la modalité 48h, respectivement à la modalité 1000, ou négative, alors que 48h, respectivement 1000, est positive. La même affirmation s'observe pour les modalités gratuit et payant, ainsi que pour les modalités gratuit et abonnement de l'attribut tarif. En effet, la valeur de la modalité gratuit est systématiquement supérieure à celle de la modalité payant, respectivement à celle de la modalité abonnement. De plus, l'option abonnement ne semble pas être préférée à celle du tarif payant par film ou par série.

Les coefficients représentant le conflit ou le refus sont significatifs pour tous les individus, étant donné que le rapport entre le paramètre estimé et l'écart-type est suffisamment grand (largement supérieur à 2). Par exemple, pour l'individu 2, ce rapport est le suivant :

$$\text{ratio refus} = \frac{1.271}{0.083} = 15.299$$
$$\text{ratio conflit} = \frac{0.476}{0.036} = 12.929$$

Les résultats significatifs de la variable du conflit confirment les comportements incertains des consommateurs face à des produits similaires. Par conséquent, l'introduction du conflit apporte une information supplémentaire dans les préférences des consommateurs et permet ainsi d'améliorer la précision des estimations.

Une autre constatation générale est que les attributs dont l'importance est la plus élevée et est significative, sont la composition, le niveau de nouveauté et le tarif. Par conséquent, les choix des individus sont motivés essentiellement par ces variables. Le fait qu'un site propose la bande annonce d'un film, une durée de conservation plus grande que celle d'autres

sites ou un catalogue plus fourni n'est pas réellement déterminant dans la sélection des sites de VoD. Par contre, une quantité de nouveautés s'approchant de celle des vidéos clubs, la gratuité des films et des séries, à l'exception des nouveautés, ou une composition de films et de séries correcte influencent bien plus le comportement des consommateurs dans le choix de la Vidéo à la Demande.

Ces diverses affirmations montrent que notre modèle de l'option "zéro" reflète les préférences réelles des consommateurs et permettent donc de conclure que l'approche bayésienne fournit des résultats cohérents et significatifs dans le cas de séparation dans les données. De plus, la modalité conflit dans le questionnaire offre aux interviewés un aspect de la prise de décision existant dans la réalité, étant donné que cette modalité est significative et qu'elle a été cochée par certains interviewés. L'ajout de la modalité conflit apporte donc une information supplémentaire à la notion de refus existant dans la littérature.

5.6.3 Modèle y=1,2,3

Dans cette situation, les interviewés n'ont pas sélectionné le conflit. Ils ont choisi les produits 1 et 2 et le refus. Aucun produit de VoD présenté dans le questionnaire n'a donc été suffisamment similaire pour que les individus soient incertains dans leurs choix.

Les résultats pour un des individus des utilités partielles issus de l'approche bayésienne et des écart-types par la méthode 25% et 75% des observations sont représentés dans le tableau suivant :

Individu 3

Variable	Utility	Standard Error	Importance (% Utility Range)
Refus :δ	1.341	0.082	
Catalogue :1000	0.197	0.058	4.95%
Catalogue :600	-0.197	0.058	
Composition :100% Films	0.304	0.089	34.38%
Composition :75% Films 25% Séries	0.874	0.122	
Composition :50% Films 50% Séries	0.683	0.076	
Composition :100% Séries	-1.862	0.288	
Nouveauté :max	0.721	0.067	25.58%
Nouveauté :moitié	0.593	0.066	
Ancien :max	-1.314	0.134	
Tarif :payant	-0.217	0.056	30.56%
Tarif :gratuit pub	1.324	0.060	
Tarif :abonnement	-1.107	0.116	
Conservation :24h	-0.119	0.052	3.00%
Conservation : 48h	0.119	0.052	
Disponibilité : Bande annonce	-0.059	0.056	1.50%
Disponibilité : Aucun	0.059	0.056	

Les résultats des autres individus sont très similaires et les commentaires sont donc valables pour tous les individus du modèles $y = 1, 2, 3$.

Les mêmes constatations que pour les modèles $y = 1, 2$ et $y = 1, 2, 3, 4$ peuvent être dites. En effet, les attributs composition, niveau de nouveauté et tarif ont les importances

les plus élevées et les plus significatives. Les autres variables, c'est-à-dire la quantité de programme, la durée de conservation du média et la disponibilité de la bande annonce, n'influencent pas le choix des sites de VoD. La grandeur et le signe des paramètres sont cohérents avec la réalité. Les écart-types sont inférieurs aux valeurs des coefficients estimés. La variable refus est significative, compte tenu de la grande valeur du rapport entre le coefficient estimé et l'écart-type. L'attribut bande annonce a, pour certains individus du modèle $y = 1, 2, 3$, un écart-type légèrement plus élevé que son utilité partielle estimée. Toutefois, étant donné que l'importance de cette variable n'est pas significative, cette affirmation n'est pas essentielle et n'influe pas sur la qualité des résultats.

En résumé, l'approche bayésienne appliquée au modèle $y = 1, 2, 3$ fournit encore une fois de bons résultats d'estimation.

5.6.4 Modèle y=1,2,4

Dans cette situation, les interviewés apprécient tous les produits présentés dans le questionnaire. En effet, ils n'ont pas coché la case refus, qui correspond à aucun des produits ne me plaît. Par contre, certains produits se sont avérés être très similaires pour ces individus et ont donc mené à une incertitude dans le choix. En résumé, les interviewés ont sélectionné les produits 1 et 2 et le conflit.

Les résultats pour un des individus des estimations par l'approche bayésienne des utilités partielles et des écart-types par la méthode 25% et 75% des observations sont résumés dans le tableau suivant :

Individu 18

Variable	Utility	Standard Error	Importance (% Utility Range)
Conflit :δ'	0.476	0.036	
Catalogue :1000	0.210	0.056	5.55%
Catalogue :600	-0.210	0.056	
Composition :100% Films	0.240	0.084	31.63%
Composition :75% Films 25% Séries	0.751	0.097	
Composition :50% Films 50% Séries	0.654	0.078	
Composition :100% Séries	-1.645	0.260	
Nouveauté :max	0.658	0.067	25.05%
Nouveauté :moitié	0.581	0.073	
Ancien :max	-1.240	0.140	
Tarif :payant	-0.224	0.054	33.39%
Tarif :gratuit pub	1.377	0.060	
Tarif :abonnement	-1.153	0.114	
Conservation :24h	-0.094	0.056	2.48%
Conservation : 48h	0.094	0.056	
Disponibilité : Bande annonce	-0.071	0.058	1.88%
Disponibilité : Aucun	0.071	0.058	

Les résultats des autres individus du modèle $y = 1, 2, 4$ sont très similaires et les commentaires peuvent donc également être appliqués pour ces derniers.

Encore une fois, nous constatons les mêmes propriétés que pour les modèles où $y = 1, 2$, $y = 1, 2, 3, 4$ et $y = 1, 2, 3$, c'est-à-dire une plus forte importance pour les attributs composition, niveau de nouveauté et tarif, une moindre importance pour la quantité de programmes proposés par le site, la durée de conservation et la disponibilité de la bande annonce, une forte influence de la gratuité des programmes qui ne sont pas des nouveautés, une composition 75% Films 25% Séries qui prime sur les autres modalités et un maximum de nouveautés légèrement plus appréciable par les consommateurs que la moitié des nouveautés.

Le conflit est significatif et apporte ainsi une information supplémentaire au mode des préférences des consommateurs dans leurs choix des sites de VoD. Les écart-types sont inférieurs aux utilités partielles estimées, à l'exception de la bande annonce pour certains individus.
L'estimation du modèle de l'option "zéro" par l'approche bayésienne est donc dans son ensemble satisfaisante.

5.6.5 Modèle y=1,3

L'individu 62 n'a choisi que le produit 1 ou le refus, ce qui signifie qu'il n'a pas apprécié les produits 2 du questionnaire et qu'il a trouvé certains des produits 1 largement supérieurs aux produits 2. Les résultats d'estimation du modèle de l'option "zéro" par l'approche bayésienne et des écart-types par la méthode 25% et 75% des observations de cet individu sont dans le tableau suivant :

Individu 62

Variable	Utility	Standard Error	Importance (% Utility Range)
Refus :δ	1.271	0.083	
Catalogue :1000	0.210	0.056	5.55%
Catalogue :600	-0.210	0.056	
Composition :100% Films	0.240	0.084	31.63%
Composition :75% Films 25% Séries	0.751	0.097	
Composition :50% Films 50% Séries	0.654	0.078	
Composition :100% Séries	-1.645	0.260	
Nouveauté :max	0.658	0.067	25.05%
Nouveauté :moitié	0.581	0.073	
Ancien :max	-1.240	0.140	
Tarif :payant	-0.224	0.054	33.39%
Tarif :gratuit pub	1.377	0.060	
Tarif :abonnement	-1.153	0.114	
Conservation :24h	-0.094	0.056	2.48%
Conservation : 48h	0.094	0.056	
Disponibilité : Bande annonce	-0.071	0.058	1.88%
Disponibilité : Aucun	0.071	0.058	

Les constations mentionnées précédemment peuvent être répétées dans ce cas-là. En effet, les attributs ayant la plus forte importance n'ont pas changé, la grandeur et les signes des paramètres sont cohérents avec la réalité, les écart-types sont inférieurs aux utilités partielles estimées et le refus est significatif.

Etant donné que la variable dépendante est dichotomique, nous pouvons également estimer ce modèle avec le maximum de vraisemblance pénalisée de S-Plus. Cependant, cette méthode n'a pas fonctionné, étant donné que la matrice $X'X$ n'est pas de rang complet.

5.6.6 Le modèle agrégé

Les résultats du modèle agrégé sont issus de l'application de la méthode du maximum de vraisemblance sur une combinaison des différents modèles associés aux types de réponses fournis par les interviewés. En d'autres termes, nous avons combiné les modèles $y = 1, 2$, $y = 1, 2, 3, 4$, etc. en un seul modèle nous donnant l'estimation des paramètres inconnus $\beta = \begin{pmatrix} d & d1 & \beta_1 & \dots & \beta_{10} \end{pmatrix}$. En effet, dans cette estimation agrégée, nous avons associé le modèle $y = 1, 2$ aux individus qui ont choisi les produits 1 et 2, le modèle $y = 1, 2, 3, 4$ aux individus qui ont sélectionné les produits 1 et 2, le refus et le conflit, etc. et nous avons effectué une estimation regroupée de ces différents modèles. Cette procédure a été implémentée sur le logiciel TSP-GiveWin. Les résultats de cette estimation du modèle agrégé de l'option "zéro" se trouvent dans le tableau suivant :

Modèle agrégé

Variable	Utility	Standard Error	Importance (% Utility Range)
Refus :δ	1.308	0.112	
Conflit :δ'	0.471	0.053	
Catalogue :1000	0.212	0.085	3.14%
Catalogue :600	-0.212	0.085	
Composition :100% Films	0.262	0.121	35.51%
Composition :75% Films 25% Séries	0.800	0.128	
Composition :50% Films 50% Séries	0.676	0.112	
Composition :100% Séries	-1.597	0.362	
Nouveauté :max	0.690	0.090	29.71%
Nouveauté :moitié	0.625	0.088	
Ancien :max	-1.315	0.178	
Tarif :payant	-0.235	0.079	23.49%
Tarif :gratuit pub	1.350	0.081	
Tarif :abonnement	-1.115	0.160	
Conservation :24h	-0.114	0.078	3.39%
Conservation : 48h	0.114	0.078	
Disponibilité : Bande annonce	-0.053	0.082	1.58%
Disponibilité : Aucun	0.053	0.082	

L'estimation du modèle agrégé a été effectué sur une population homogène, étant donné que ce sont des étudiants universitaires genevois de première année et qu'ils sont par conséquent issus du même milieu social et qu'ils sont dans la même tranche d'âge. Si la population avait été hétérogène, nous aurions dû former des groupes homogènes et estimer les utilités partielles pour chacun de ces groupes.

L'annexe 7 donne les valeurs d'estimation des paramètres sans les variables de référence, ainsi que leurs t-statistiques et leurs p-valeurs. Les résultats des variables de référence sont obtenus en supposant que la somme des modalités par attribut est égale à 0. Ces variables de référence sont en réalité les dernières modalités de chaque attribut, c'est-à-dire :

- catalogue : 600
- composition site : 100% Séries
- composition nouveauté : que des anciens
- tarif : abonnement
- temps de conservation : 48 heures
- bande annonce : aucun

Les résultats du modèle agrégé montrent que les catégories de l'option "zéro", le refus et le conflit, sont significatives. En effet, leurs t-statistiques équivalent à 11.6384 pour le refus et 8.82505 pour le conflit (annexe 7). Ces valeurs semblent suffisamment élevées pour conclure à la significativité de ces paramètres. Les p-valeurs de ces coefficients confirment cela, car ils sont égales à 0 (annexe 7). Par conséquent, le conflit et le refus apportent une explication supplémentaire dans le choix des préférences des consommateurs dans la Vidéo à la Demande. Ils ne sont donc pas superflus dans l'analyse des préférences des individus. D'après la valeur estimée du refus, les individus apprécient les produits dont l'utilité est supérieure à 1.30832. D'après la valeur estimée du conflit, des produits dont la différence est inférieure à 0.471899 provoquent une indécision ou une incertitude dans le choix des consommateurs.

Les attributs qui ont la plus grande importance sont la composition du site, le niveau de nouveauté et le tarif. Ces derniers influencent la décision du consommateur de manière significative. La variable dont l'importance est la plus élevée est la composition du site. Cet attribut est donc le plus déterminant dans le choix du site de VoD. Il s'en suit le niveau de nouveauté, puis le tarif. Etonnement, les individus ne portent pas la plus grande importance au tarif. Par conséquent, un site bien réalisé peut fonctionner malgré le fait qu'il soit payant. Néanmoins, la grandeur de la gratuité est plus élevée que celles des attributs concernant la composition et le niveau de nouveauté. Par conséquent, la

valeur associée à la gratuité a une influence élevée sur l'utilité du produit qui contient cette modalité.
Quant à la quantité de programmes dans le catalogue, la durée de conservation et la disponibilité de la bande annonce, elles ne semblent pas être décisives dans le choix d'un site de Vidéo à la Demande d'après leurs importances. De plus, d'après le signe négatif associé à la modalité bande annonce, le fait de proposer ce service est plus désavantageux que de ne pas le proposer. Il faut, cependant, sous-peser cette remarque, étant donné que le niveau d'importance de cet attribut n'est pas significatif.

Le signe et la grandeur des coefficients estimés sont cohérents avec la réalité. En effet, pour les attributs pouvant être rangés dans un ordre de préférence universel, tels que la quantité de programmes et la durée de conservation, ils sont négatifs pour les plus petites valeurs et positifs pour les plus grandes. Par conséquent, la valeur la plus faible de la durée de conservation (24h), respectivement de la quantité de programmes (600), est plus basse que celle qui est la plus élevée, c'est-à-dire 48h pour la durée de conservation, respectivement 1000 pour la quantité de programmes. La même constatation s'observe entre les modalités gratuit et payant ou abonnement. En effet, il semble correct que les paramètres associés au tarif payant ou à l'abonnement soient inférieurs à la gratuité d'un média. Les résultats d'estimation des coefficients de l'attribut tarif reflètent bien cette réalité.

L'estimation des coefficients associés à la variable composition du site montre une préférence des individus concernant une disposition de 75% Films et 25% Séries. Il s'en suit la modalité 50% Films et 50% Séries, puis la caractéristique 100% Films. Consacrer un site exclusivement aux séries TV ne semble pas réjouir le consommateur de VoD.

Un niveau de nouveauté maximum est un critère important dans le choix du site. Cependant, le fait d'avoir la moitié des nouveautés disponibles sur le marché des vidéos clubs s'approche, en terme de satisfaction, de la modalité maximum de nouveautés. Par contre, un site qui n'offre que des programmes anciens ne reflète pas réellement les préférences des consommateurs dans la Vidéo à la Demande.

En règle générale, les résultats d'estimation du modèle agrégé de l'option "zéro" sont cohérents et satisfaisants d'un point de vue statistique et économique. Ils reflètent bien la réalité des préférences dans la Vidéo à la Demande. Ils montrent la significativité et donc l'importance de l'indécision dans le choix des produits similaires (conflit), ainsi que l'indifférence sur certains autres sites de VoD (refus). Ces résultats permettent finalement de créer des sites de VoD plus adaptés aux préférences des individus. Dans la suite, nous

exposons les parts de marché de certains sites existant et d'un site correspondant à l'idéal des individus selon les résultats d'estimation.

5.7 Les parts de marché

Les résultats de l'estimation agrégée de notre spécification de l'option "zéro" nous ont permis de déterminer les attributs et les modalités influençant au mieux la sélection d'un site de Vidéo à la Demande. A partir de ces informations, nous avons créé le site de VoD qui pourrait satisfaire la majorité des individus. Nous l'avons appelé le site Idéal. Ses caractéristiques sont les suivantes :

- Catalogue : 1000 programmes
- Composition : 75% Films 25% Séries
- Nouveauté : maximum
- Tarif : gratuit pub
- Conservation : 48h
- Disponibilité : aucun

Le but de la détermination du produit Idéal est en réalité le calcul des parts de marché de ce dernier par rapport à d'autres sites de VoD existants déjà sur le marché. Pour cela, nous avons choisi les sites existants de Canalplay et de TF1Vision. Ces derniers sont ceux dont la popularité a été la plus prononcée dans la Vidéo à la Demande des films et des séries. Au début de notre étude (octobre 2006), les caractéristiques de ces sites n'étaient pas les mêmes que maintenant. Ils ont, en effet, bien évolué. Cependant, afin de concorder avec le début de notre étude, nous avons dû définir ces sites de VoD en fonction des caractéristiques apparaissant au début de notre étude (juillet 2006). Pour cela, nous nous sommes appuyés sur les divers journaux informatiques de juillet à octobre 2006, tels que L'Ordinateur Individuel (L'OI), P2PMag et MicroActuel, et sur nos visites directes sur ces sites au moment de notre étude. Nous avons dû adapter certaines caractéristiques de ces sites selon les modalités des attributs considérés dans notre étude.

Par exemple, le catalogue de Canalplay en juillet-août 2006 était de 1000 programmes selon L'OI, alors que le P2PMag d'octobre 2006 disait qu'il était de 700, le MicroActuel d'octobre 2006 de 500 et nos recherches en septembre 2006 montrait un catalogue de 1159 films. Par conséquent, nous avons opté pour un catalogue de 1000 programmes. En 2006, le site de Canalplay était plus axé sur la VoD de films. La caractéristique de l'attribut composition du site est donc 100% Films. En ce qui concerne le niveau de nouveauté, les magazines informatiques ne présentant pas cet aspect, si ce n'est qu'ils disent que les

films de ces sites ne sont pas assez récents. Nous avons donc utilisé les informations recueillies sur l'accès direct au site en septembre 2006. Nous avions constaté à cette période que le nombre de nouveautés s'élevait à 52 pour les films. Ces nouveautés devaient plus correspondre à des films de l'année 2005 que 2006. Etant donné que nous avions considéré comme nouveautés les films des années 2005 et 2006, nous avons donc gardé cette valeur. Cette dernière est inférieure à la moitié des nouveautés ($\frac{140}{2}$). Nous avons tout de même décidé d'opter pour la caractéristique moitié nouveautés pour définir l'attribut niveau de nouveauté du site Canalplay. En effet, nous pensions qu'il aurait été erroné de catégoriser le site Canalplay dans la modalité que anciens. Les autres informations ont été plus précises. En effet, le site Canalplay ne proposait pas de bande annonce au début de notre étude et la durée de conservation était de 24 heures. De plus, ce dernier n'offre que des programmes payants. En résumé, les caractéristiques du site Canalplay selon nos critères d'attribut sont les suivantes :

- Catalogue : 1000 programmes
- Composition : 100% Films
- Nouveauté : moitié
- Tarif : payant
- Conservation : 24h
- Disponibilité : aucun

Nous avons procédé de la même manière pour définir le site de TF1Vision. Au début de notre étude, les magazines informatiques ne donnaient pas des quantités de programmes du site de TF1Vision identiques. L'OI de juillet-août 2006 postulait une valeur de 500 programmes, alors que MicroActuel d'octobre 2006 donnait une valeur de 300 et P2PMag de 230. La visite directe du site en septembre 2006 ne permettait pas de déterminer le nombre de programmes de manière fiable, étant donné que certains films apparaissaient plusieurs fois dans diverses catégories. Nous avons approximé le nombre de programmes à 600, compte tenu des 500 proposés par L'OI de juillet-août 2006. De plus, étant donné que TF1Vision proposait quelques séries sans pour autant dépasser le nombre de films à dispositions, nous avons postulé pour une composition de 75% Films et 25% Séries. En terme de nouveautés, nous avons constaté qu'en visitant le site en septembre 2006, il y en avait quelques unes, aussi bien pour les séries que pour les films. Cependant, pour les mêmes raison que précédemment, nous n'avons pas pu déterminer la quantité de nouveautés par film et par série. Nous avons donc choisi de définir cet attribut par la modalité moitié de nouveautés, afin de tenir compte du fait que le site en proposait quelques unes.

La durée de conservation au début de notre étude était de 24 heures. Le site proposait, gratuitement et en avant-première, la diffusion d'une de ses séries phare durant la semaine qui suivait son passage télévisé sur la chaîne. Nous avons, quand même, choisi la modalité payante de l'attribut tarif, car les autres programmes étaient, quant à eux, tous payants. Au début de notre étude, le site proposait des démos de ces programmes. Par conséquent, nous avons choisi la modalité bande annonce dans l'attribut disponibilité. En résumé, en accord avec les attributs de notre étude, les caractéristiques du site de VoD de TF1Vision sont les suivantes :

- Catalogue : 600 programmes
- Composition : 75% Films 25% Séries
- Nouveauté : moitié
- Tarif : payant
- Conservation : 24h
- Disponibilité : bande annonce

Dans la suite du travail, nous avons calculé d'abord les probabilités de choix ou d'achat individuelles, puis les parts de marché. Les parts de marché peuvent être déduites à partir du modèle individuel ou du modèle agrégé.

5.7.1 Les probabilités de choix

Pour chaque individu, nous pouvons, à partir des utilités partielles estimées, calculer les probabilités individuelles d'achat des sites existants sur le marché, à savoir Canalplay et TF1Vision, et du site Idéal. La procédure de calcul est détaillée pour l'individu 2 dans les tableaux suivants :

Individu 2			
Canalplay	Attributs	Modalités	Utilités partielles
	Catalogue	1000	0.210
	Composition	100% Films	0.240
	Nouveauté	moitié	0.581
	Tarif	payant	-0.224
	Conservation	24h	-0.094
	Disponibilité	aucun	0.071
Utilité u_i i=Canalplay	Somme des utilités partielles =utilité du site Canalplay pour l'individu 2		0.785
Probabilités de choix individuelles en %	$\frac{u_i}{\sum_i u_i}$ (Bradley-Terry-Luce)		16.77%
	$\frac{\exp(u_i)}{\sum_i \exp(u_i)}$ (logit)		7.85%

	Individu 2		
TF1Vision	Attributs	Modalités	Utilités partielles
	Catalogue	600	-0.210
	Composition	75% Films 25% Séries	0.751
	Nouveauté	moitié	0.581
	Tarif	payant	-0.224
	Conservation	24h	-0.094
	Disponibilité	bande annonce	-0.071
Utilité u_i i=TF1Vision	Somme des utilités partielles =utilité du site TF1Vision pour l'individu 2		0.733
Probabilités de choix individuelles en %	$\frac{u_i}{\sum_i u_i}$ (Bradley-Terry-Luce)		15.65%
	$\frac{\exp(u_i)}{\sum_i \exp(u_i)}$ (logit)		7.45%

	Individu 2		
Idéal	Attributs	Modalités	Utilités partielles
	Catalogue	600	0.210
	Composition	75% Films 25% Séries	0.751
	Nouveauté	moitié	0.658
	Tarif	payant	1.377
	Conservation	24h	0.094
	Disponibilité	bande annonce	0.071
Utilité u_i i=Idéal	Somme des utilités partielles =utilité du site Idéal pour l'individu 2		3.163
Probabilités de choix individuelles en %	$\frac{u_i}{\sum_i u_i}$ (Bradley-Terry-Luce)		67.56%
	$\frac{\exp(u_i)}{\sum_i \exp(u_i)}$ (logit)		84.69%

Pour le calcul des probabilités individuelles de choix, nous avons utilisé la technique consistant à diviser l'utilité d'un des sites en fonction de la somme des utilités de tous les sites, Canalplay, TF1Vision et Idéal, et celle qui divise l'exponentielle de l'utilité d'un des sites en fonction de la somme des exponentielles de tous les sites. La première méthode est associée aux auteurs Bradley-Terry-Luce, alors que la seconde au terme logit. En résumé, les probabilités individuelles d'achat de chaque site avec ces différentes méthodologie se formulent comme suit :

$$\text{Bradley-Terry-Luce :} \qquad \frac{u_i}{\sum_i u_i}$$

$$\text{logit :} \qquad \frac{\exp(u_i)}{\sum_i \exp(u_i)}$$

En procédant de la même manière que ci-dessus, nous avons calculé les probabilités individuelles de choix des sites Canalplay, TF1Vision et Idéal des autres individus.

Ces résultats se trouvent dans le tableau suivant :

		Probabilités de choix individuelles en % (Bradley-Terry-Luce)			**Probabilités de choix individuelles en % (logit)**		
Modèle	Individu	**Canalplay**	**TF1Vision**	**Idéal**	**Canalplay**	**TF1Vision**	**Idéal**
1,2,3,4	2	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	3	16.39	17.50	66.10	7.14	7.55	85.29
1,2,3,4	4	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	5	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	7	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	8	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	10	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	11	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	14	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	15	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	16	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	17	15.55	15.91	68.53	7.20	7.33	85.46
1,2,4	18	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	20	15.07	19.29	65.62	6.98	8.60	84.40
1,2,3,4	21	15.55	15.91	68.53	7.20	7.33	85.46
1,2	22	16.83	17.81	65.34	7.31	7.68	84.99
1,2,3	23	16.82	16.90	66.26	7.91	7.94	84.14
1,2,3	24	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	29	16.82	16.90	66.26	7.91	7.94	84.14
1,2,3	31	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	33	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	34	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	36	16.77	15.65	67.56	7.85	7.45	84.69
1,2,4	41	15.16	15.58	69.25	6.86	6.99	86.14
1,2,3	44	16.77	15.65	67.56	7.85	7.45	84.69
1,2,4	45	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	46	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	48	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	50	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	52	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	53	15.55	15.91	68.53	7.20	7.33	85.46
1,2,3	54	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	55	15.07	19.29	65.62	6.98	8.60	84.40
1,2,3	56	16.14	15.91	67.93	8.19	8.11	83.68
1,2,3	57	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	59	16.82	16.90	66.26	7.91	7.94	84.14
1,2,3,4	60	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	61	16.77	15.65	67.56	7.85	7.45	84.69
1,3	62	16.77	15.65	67.56	7.85	7.45	84.69

		Probabilités de choix individuelles en % (Bradley-Terry-Luce)			**Probabilités de choix individuelles en % (logit)**		
Modèle	Individu	**Canalplay**	**TF1Vision**	**Idéal**	**Canalplay**	**TF1Vision**	**Idéal**
1,2,3	63	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	64	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	66	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	67	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	68	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	72	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	73	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	74	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	75	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	76	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	79	15.07	19.29	65.62	6.98	8.60	84.40
1,2	81	15.55	15.91	68.53	7.20	7.33	85.46
1,2,3	82	15.07	19.29	65.62	6.98	8.60	84.40
1,2,3	83	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	84	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	86	15.55	15.91	68.53	7.20	7.33	85.46
1,2,3	91	16.82	16.90	66.26	7.91	7.94	84.14
1,2,3,4	92	15.07	19.29	65.62	6.98	8.60	84.40
1,2,3,4	94	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	95	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	97	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	98	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	100	15.07	19.29	65.62	6.98	8.60	84.40
1,2	103	15.24	15.97	68.77	7.09	7.33	85.56
1,2,3	107	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	108	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	110	16.39	17.50	66.10	7.14	7.55	85.29
1,2,3	111	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	112	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	115	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3,4	116	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	117	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	118	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	119	16.77	15.65	67.56	7.85	7.45	84.69
1,2,3	120	15.55	15.91	68.53	7.20	7.33	85.46

En règle générale, les probabilités individuelles d'achat, fournies par les méthodes (Bradley-Terry-Luce) ou (logit), du produit Idéal sont largement plus élevées que celles de Canalplay ou de TF1Vision. De plus, il semble que les sites Canalplay et TF1Vision se partagent de manière équivalente le marché de la VoD, étant donné que leurs probabilités individuelles de choix sont très similaires. Dans la suite, nous effectuons la même analyse pour le cas agrégé.

5.7.2 Les parts de marché

Le calcul des parts de marché a été effectué de deux manières, à savoir à partir des probabilités individuelles de choix ou à partir du modèle agrégé.

Les parts de marché issues des probabilités individuelles ont été effectuées en calculant la moyenne des probabilités individuelles de choix de chaque scénario :

$$\text{Parts de marché} = \left(\frac{\sum \text{probabilités de choix}}{\text{nombre d'individu (74)}} \right)$$

Nous obtenons ainsi les parts de marché des méthodes Bradley-Terry-Luce et logit. Les résultats de ces dernières sont les suivants :

	Parts de marché en %		
	Canalplay	**TF1Vision**	**Idéal**
Bradley-Terry-Luce	16.47	16.12	67.39
logit	7.68	7.56	84.74

Nous observons les mêmes constatations qu'avec les probabilités individuelles de choix, c'est-à-dire que le scénario Idéal offre une part de marché bien plus élevée que celles de Canalplay ou de TF1Vision. De plus, ces derniers ont des parts de marché très similaires.

Le calcul des parts de marché aurait également pu être effectué en calculant la moyenne des utilités individuelles et en utilisant sur ces moyennes la méthode de Bradley-Terry-Luce et logit. Les résultats sont très similaires à ceux effectués en utilisant la moyenne des probabilités. Cependant, cette méthode n'est pas correcte, car elle calcule des moyennes d'utilités dont leurs valeurs d'échelle peuvent différer d'un individu à l'autre. En effet, si une des utilités individuelles est plus grande que celles des autres individus, le poids de l'individu associé à cette utilité plus élevée, dans le calcul des parts de marché, ne sera pas le même que celui des autres individus. Utiliser les probabilités individuelles permet donc de normaliser ces valeurs d'échelle à des valeurs individuelles comparables.

Les parts de marché calculées avec les probabilités individuelles de choix ne tiennent, cependant, pas compte du refus et du conflit. Nous ne considérons ces situations que lors du calcul des parts de marché à partir du modèle agrégé. La raison est qu'avec le modèle agrégé, nous pouvons calculer les utilités agrégées des scénarios, afin de les comparer aux valeurs du refus et du conflit. Dans le cas individuel, nous ne pouvons pas obtenir des utilités agrégées correctes compte tenu des échelles de mesure non comparables.

Les parts de marché issus du modèle agrégé s'obtiennent, quant à elles, en calculant les utilités des scénarios à partir des résultats du maximum de vraisemblance agrégé. Ces utilités sont les suivantes :

Utilités			**Refus**
Canalplay	**TF1Vision**	**Idéal**	δ
0.749	1.021	3.053	1.308

Différence			**Conflit**
Canalplay-Tf1Vision	**Canalplay-Idéal**	**TF1Vision-Idéal**	δ'
0.271	2.303	2.031	0.471

Selon les valeurs de ces tableaux, nous remarquons que les sites de Canalplay et TF1Vision sont dans la zone de refus, étant donné que leurs utilités sont inférieures à δ. Par conséquent, dans un marché où les sites Canalplay, TF1Vision et Idéal coexistent, les produits Canalplay et TF1Vision ne sont pas choisis par les consommateurs. Nous en concluons que le seul produit intéressant est le site de VoD Idéal et que par conséquent, il détient des parts de marché de 100%, alors que Canalplay et TF1Vision ont 0%. De plus, selon les valeurs des différences, il semble que les produits Canalplay et TF1Vision sont difficilement départageables, compte tenu de leur similitude. En effet, la différence entre les sites Canalplay et TF1Vision est inférieure à la valeur δ' du conflit.

Des parts de marché de 100% ne nous renseignent pas sur l'influence des sites de Canalplay et de TF1Vision dans la Vidéo à la Demande. C'est pourquoi, nous avons décidé qu'il serait intéressant de calculer ces influences d'achat dans un marché comportant que les sites de Canalplay et de TF1Vision. Etant donné que nous ne comparons que deux produits, nous pouvons déduire les probabilités des alternatives produit 1 (Canalplay), produit 2 (TF1Vision), refus et conflit à l'aide des équations de probabilités de notre spécification de l'option "zéro". En effet, ces dernières s'expriment de la manière suivante :

- $P(\text{Choix h}) = P(y=1) = P_1 = 1 - \Phi(\frac{\delta-\beta' x_h}{\sigma}) - \Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}) + \Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}, \frac{\delta-\beta' x_h}{\sigma}; \frac{1}{\sqrt{2}})$
- $P(\text{Choix l}) = P(y=2) = P_2 = \Phi(\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\delta-\beta' x_l}{\sigma}, \frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}; -\frac{1}{\sqrt{2}})$
- $P(\text{Refus}) = P(y=3) = P_3 = \Phi(\frac{\delta-\beta' x_h}{\sigma})\Phi(\frac{\delta-\beta' x_l}{\sigma})$
- $P(\text{Conflit}) = P(y=4) = P_4 = \begin{cases} \Phi(\frac{\delta-\beta' x_h}{\sigma}) + \Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}) - \Phi(\frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}) \\ -\Phi(\frac{\beta'(x_l-x_h)+\delta'}{\sqrt{2}\sigma}, \frac{\delta-\beta' x_h}{\sigma}; \frac{1}{\sqrt{2}}) - \Phi(\frac{\delta-\beta' x_h}{\sigma})\Phi(\frac{\delta-\beta' x_l}{\sigma}) \\ +\Phi(\frac{\delta-\beta' x_l}{\sigma}, \frac{\beta'(x_l-x_h)-\delta'}{\sqrt{2}\sigma}; -\frac{1}{\sqrt{2}}) \end{cases}$

où σ a été posé à 1, afin de tenir compte de l'identification des paramètres.

En remplaçant les valeurs des paramètres par celles estimées par le maximum de vraisemblance du modèle agrégé :

$$\begin{array}{llll} \delta = 1.308 & \delta' = 0.471 & \beta_1 = 0.212 & \beta_2 = 0.262 \\ \beta_3 = 0.800 & \beta_4 = 0.676 & \beta_5 = 0.690 & \beta_6 = 0.625 \\ \beta_7 = -0.235 & \beta_8 = 1.350 & \beta_9 = -0.114 & \beta_{10} = -0.053 \end{array}$$

nous obtenons les résultats suivants :

Probabilités

Canalplay	TF1Vision	Refus	Conflit
0.186	0.291	0.436	0.085

Nous constatons que la probabilité de refus est élevée, mais qu'elle n'est pas de 100%, c'est-à-dire qu'elle ne suppose pas que les produits Canalplay et TF1Vision ne sont pas appréciés par tous les consommateurs. Leur part de marché est donc supérieure à 0%, lorsque ces sites coexistent dans le marché de la VoD.

En résumé, nous avons donc sur 100 individus, 19 qui choisissent Canalplay, 29 qui préfèrent TF1Vision, 44 qui n'apprécient aucun des produits et 8 qui sont indécis dans le choix de ces sites. Afin de calculer les parts de marché, nous transformons ces probabilités en tenant compte des situations de conflit et de refus. Les 44% du refus ne correspondent pas à un potentiel d'achat des produits et doivent donc être exclus du calcul des parts de marché. Les 8% se partagent entre le produit Canalplay et TF1Vision, compte tenu de l'incertitude du consommateur. Nous supposons que ce partage est équivalent d'un site à l'autre et nous décidons donc d'ajouter la moitié de cette valeur aux probabilités Canalplay et TF1Vision. Afin d'obtenir un pourcentage de 100%, étant donné que la probabilité de refus a été enlevé du calcul des parts de marché, il faut donc réajuster les probabilités Canalplay et TF1Vision de la manière suivante :

Probabilité avec conflit

Canalplay	**TF1Vision**
0.187+0.042	0.291+0.043
=0.229	=0.334

Somme des probabilités avec conflit

0.563

Parts de marché

Canalplay	**TF1Vision**
$\frac{0.229}{0.563} =$	$\frac{0.334}{0.563} =$
0.407	**0.593**

Cette méthode, qui intègre le refus et le conflit, nous montre que dans un marché où il n'y a que les individus qui sont intéressés par les produits Canalplay et TF1Vision, les parts de marché sont de 40.71% pour Canalplay et de 59.29% pour TF1Vision. De plus, les parts de marché du site TF1Vision sont plus élevées que celles de Canalplay. Ces résultats montrent que le refus et le conflit sont une source d'information supplémentaire dans la compréhension du comportement de prise de décision. En effet, notre spécification de l'option "zéro" permet dans le calcul des parts de marché de ne tenir compte que des consommateurs potentiels et d'intégrer ceux qui hésitent dans leur choix. Par conséquent, nos parts de marché représentent bien les consommateurs potentiellement prêt à acheter le type de produit proposé.

Conclusion

Le marché de la Vidéo à la Demande est un marché en pleine expansion et méconnu en Europe au début de notre étude (octobre-novembre 2006). Ce marché ne cesse d'évoluer au risque de remplacer les vidéo clubs. Nous nous sommes intéressés à ce produit en considérant des attributs s'approchant de ceux proposés par les vidéo clubs. Nous étions conscients que ces services, n'étant pas encore intégrés dans la VoD, pourraient à l'avenir bientôt faire partie des nouveaux sites de VoD. Nous avons donc testé certains attributs pertinents à travers notre spécification de l'option "zéro", qui considère le refus et le conflit dans le choix des produits. La sélection de ces attributs pertinents à été motivée par les divers magazines informatiques traitant du sujet. Certains de ces attributs sont le reflet des sites de VoD disponibles au début de notre étude. En une année, l'évolution de la VoD a été grandissante et améliorée dans les propositions des services disponibles. Cependant, certains de nos attributs, dont leur importance et leur influence sont grandes, n'ont pas encore été intégrés dans les sites de VoD actuels, alors que d'autres caractéristiques, dont notre étude montre leur moindre importance, ont été améliorées. Ces améliorations sont essentiellement l'augmentation des catalogues, de la durée de conservation d'un film et l'introduction des bandes annonces. Or, ces attributs montrent une importance négligeable pour le consommateur.

Notre spécification de l'option "zéro" a été effectuée de manière individuelle, afin de se conformer au but premier de l'analyse conjointe, qui est une étude du comportement individuel dans le choix des préférences. Cependant, lorsque le recueil de données est de type choix, c'est-à-dire la sélection d'un produit parmi plusieurs, l'estimation des paramètres oblige le chercheur à utiliser la méthode du maximum de vraisemblance. Cette méthode présente le défaut de ne pas pouvoir converger, lorsque l'échantillon contient de la séparation dans ces données.

Dans notre spécification de l'option "zéro", nous avons utilisé les comparaisons par paires pour recueillir nos données, dans le but d'intégrer le refus et le conflit dans notre modèle. L'estimation individuelle des paramètres inconnus par notre spécification de l'option "zéro" a donc fait intervenir le maximum de vraisemblance. Nous nous sommes retrouvés confronter au problème de séparation dans nos données, étant donné que les échantillons individuels ne comportaient que 20 observations. En réalité, la séparation est fréquente dans les échantillons de petites tailles, c'est-à-dire inférieurs à 50 observations, et lorsque la proportion du nombre d'observations par rapport au nombre d'attributs est faible, inférieure à 10. Notre modèle estime 12 paramètres pour 20 observations, ce qui donne une proportion de 1.67. Compte tenu de cette séparation, nous avons donc dû trouver une alternative au maximum de vraisemblance. La littérature suggère la vraisemblance pénalisée dans les cas dichotomique logit. Elle affirme que cette méthodologie apporte de meilleurs résultats que les autres approches applicables avec la séparation. Notre modèle de l'option "zéro" n'étant pas de spécification logit, mais probit, et étant multinomial, nous avons donc constaté que l'affirmation précédente n'était pas valable dans notre cas et nous avons donc opté pour l'approche bayésienne. Cette approche a demandé l'intégration de simulations de Monte Carlo dans l'estimation des paramètres inconnus. Les résultats fournis par cette approche décrivent comme nous l'attendions la Vidéo à la Demande.

Nous avons pu constater que la quantité de programmes, la durée de conservation d'un média et la disponibilité de la bande annonce n'avait pas une grande importance dans l'échantillon d'étudiants sélectionné dans l'interview. Par contre, la gratuité des programmes, la composition du site en terme de films et de séries et de nouveautés semblent être primordiaux dans le choix de sites de VoD. Le refus et le conflit apportent également des informations supplémentaires dans la prise de décision.

Cet apport supplémentaire a pu être constaté lors du calcul des parts de marché. En effet, en calculant les parts de marché à partir de la moyenne des probabilités individuelles des sites Canalplay, TF1Vision et Idéal [6], nous avons obtenu une faible part de marché non nulle pour les sites Canalplay et TF1Vision, et une part majoritaire pour le site Idéal. Lorsque nous avons introduit le refus et le conflit, nous avons observé à partir du modèle agrégé, que les utilités des sites Canalplay et TF1Vision sont plus faibles que la valeur du refus, impliquant la non satisfaction de ces sites. De plus, la différence d'utilités de ces sites étant inférieure à la valeur du conflit, montre une similitude de ces produits amenant à une incertitude dans le choix. Le site Idéal a donc une part de marché de 100% lorsque

[6] Le site Idéal a été créé de sorte qu'il contient les éléments préférés des individus selon les résultats de l'estimation.

le refus et le conflit sont intégrés.

Le refus des sites de Canalplay et TF1Vision s'explique par le fait que lors du questionnaire, les individus ont été confrontés à des caractéristiques inexistantes sur le marché, mais supérieures à celles offertes par Canalplay et TF1Vision. Ils ont donc intégré ces attributs comme une réalité, et ne se sont pas contentés du peu que leur proposait les sites de VoD existant sur le marché. Dans un marché où le site Idéal serait proposé, les sites Canalplay et TF1Vision n'auraient donc pas leur place. Ces affirmations montrent à quel point nous pouvons encore améliorer le marché de la Vidéo à la Demande.

Nous aurions pu n'intégrer que les caractéristiques existantes dans le marché de la VoD, afin d'obtenir les parts de marché actuelles des sites de VoD. Cependant, le but de ce type d'étude n'est pas de reproduire les résultats de parts de marché déjà connues par les fournisseurs de VoD, mais de rechercher l'effet d'autres attributs sur le marché de la VoD.

Conclusion et perspectives

Les implémentations de l'analyse conjointe sont multiples. Elles dépendent de la manière, dont sont récoltées les observations. Dans notre travail, nous distinguons l'analyse conjointe par la notation ou par le classement de celle par le choix. La notation, respectivement le classement, demande à l'individu d'associer une valeur, respectivement un ordre, à des scénarios ou des produits hypothétiques. Le choix consiste, quant à lui, à effectuer une sélection unique d'un de ces produits. Les scénarios, auxquels l'individu est confronté, sont une partie du nombre total de produits ou de combinaisons d'attributs. Ils sont déterminés à l'aide des plans d'expérience. Le plan utilisé pour réduire le nombre de produits est, soit un plan réduit orthogonal, soit un plan D-optimal. Les analyses exigeant, par exemple, des classements de produits différents selon des groupes d'individus, peuvent également faire appel à une autre catégorie de plans d'expérience, appelée les plans en bloc incomplets équilibrés. Une fois que les plans d'expérience sont choisis et appliqués sur les combinaisons d'attributs totales, il reste à soumettre les produits résultants à un échantillon d'individus, afin qu'ils les notent, les classent ou en sélectionnent un. Cette étape permet de récolter les données de la variable dépendante du modèle de l'analyse conjointe, dans le but d'estimer les paramètres inconnus ou les utilités partielles.

Ces estimations des utilités partielles des modèles d'analyse conjointe varient selon l'approche envisagée, c'est-à-dire la notation et le classement, ou le choix unique. Pour la première catégorie, les méthodes des moindres carrés ordinaires (OLS) ou MONANOVA sont celles les plus fréquemment utilisées dans ce genre d'étude. L'analyse des préférences par le choix unique estime, quant à elle, ces coefficients inconnus par le maximum de vraisemblance (MV) sur des spécifications économétriques, qui sont soit logit, soit probit. Des observations récoltées, en soumettant dans les interviews des comparaisons par paires, par triplets, etc. ou en demandant à l'individu de se prononcer par un oui ou par non sur l'appréciation des produits, exigent également l'estimation des utilités partielles par la méthode MV sur des modèles logit ou probit.

Dans les comparaisons par paires, par triplets, etc., nous pouvons ajouter un choix supplémentaire, à savoir une option permettant aux consommateurs de ne sélectionner aucun

scénario proposé dans la paire, le triplet, etc. Cette option permet de rendre plus réel un interview ou un questionnaire. Le logiciel Sawtooth propose cette alternative de non choix dans son programme CBC, sans pour autant décrire le procédé d'application. La littérature est, quant à elle, très pauvre à ce sujet. Dans notre travail, nous avons donc développé cet aspect du non choix. Nous l'avons également étendu, en y intégrant des notions psychologiques. En effet, le non choix, appelé l'option "zéro" dans notre travail, est défini comme étant dû à la non appréciation des scénarios (refus) ou à l'indécision dans le choix des produits (conflit). La définition du non choix dans la littérature diffère de la notre, étant donné qu'elle ne considère que le refus et pas le conflit. Cependant, tel que les auteurs Tversky et Shafir (1992) le démontrent dans un de leurs articles, le non choix ne signifie pas uniquement que les individus n'apprécient pas les produits proposés. Ils peuvent également effectuer des non choix lorsqu'ils ne sont pas certains dans leurs préférences.

Dans cette thèse, nous avons donc donné une nouvelle spécification du non choix dû au refus et au conflit. La raison est que la littérature ne fournit que très peu de modèles intégrant le refus, qui, en plus, ne sont pas très satisfaisants et que le non choix dû au conflit n'a pas encore été modélisé, malgré son apport évident dans sa compréhension. Etant donné que l'intégration du non choix se prête mieux dans les modèles de comparaisons par paires, par triplets, etc., nous avons donc décidé de spécifier notre modèle de l'option "zéro" dans ce cadre-là. L'introduction des non choix dû au refus et au conflit a été effectuée en ajoutant dans les comparaisons par paires, par triplets, etc. les options "aucun produit ne me plaît" et "tous les produits me plaisent". De plus, afin que le nombre d'alternatives ne soit pas trop élevé, nous avons préféré utiliser les comparaisons par paires, qui donnaient ainsi le minimum possible, à savoir deux produits et deux non choix par question. Les produits ont été représentés dans le modèle par des utilités, alors que les non choix dépendaient de seuils liés aux utilités des produits. La spécification, utilisée dans notre modèle de l'option "zéro", est de la forme probit et elle mélange les approches de comparaison d'utilités et de seuils de réaction, pour tenir compte du fait qu'une utilité associée au non choix ne peut s'interpréter comme dans le cas des produits.

Etant donné que notre spécification probit à quatre alternatives de l'option "zéro" utilise la comparaison par paires pour récolter ces observations, l'estimation des utilités partielles a donc dû être effectuée par la méthode du maximum de vraisemblance (MV). Cette procédure rencontre certains problèmes d'estimations dans les échantillons de petites tailles, tels que c'est généralement le cas dans les modèles individuels de l'analyse conjointe. En effet, des observations présentant de la séparation dans les données peuvent mener à la non convergence des estimations par MV. Quelques alternatives au maximum

de vraisemblance sont proposées dans la littérature, telles que l'inférence exacte, l'approche bayésienne et la méthode du maximum de vraisemblance pénalisée.

Afin de mettre en oeuvre toutes ces connaissances et ces nouveautés en terme de non choix, nous avons illustré notre spécification de l'option "zéro" à l'aide d'un exemple sur le marché de la Vidéo à la Demande (VoD). Nous avons analysé ce modèle de manière individuelle et agrégée, en estimant les paramètres inconnus par la méthode MV. L'estimation individuelle n'a pas pu être effectuée, car les données comportaient de la séparation. Nous avons donc contourné le problème en estimant les coefficients du modèle par l'approche bayésienne. Les résultats d'estimations par cette méthode, effectués sur une population d'étudiants universitaires de première année, ont été probants d'un point de vue économique et statistique. Ils révèlent une forte influence du contenu en terme de films et de séries et de niveau de nouveautés, ainsi que du tarif de location des médias. La gratuité avec de la publicité dans le tarif prime sur la proposition d'un abonnement ou d'un simple prix par programme. Les consommateurs donnent également leur préférence sur une composition de site de VoD de 75% Films et 25% Séries. Ils ne sont pas très attirés par des sites dont la proportion d'anciennes émissions est de 100%. Ils préfèrent un mélange entre des nouveaux et des anciens programmes. Cette préférence est légèrement supérieure lorsque les nombre de nouveautés est maximum, plutôt que de moitié. Les services, tels la disponibilité de la bande annonce, la quantité de programmes dans le site et la durée de location, ne sont pas d'une très grande importance.

En ce qui concerne les non choix dus au refus et au conflit, l'estimation de notre modèle sur le marché de la VoD montrent une significativité des paramètres associés à ces derniers. Par conséquent, l'introduction du conflit et du refus n'est pas sans importance et apporte bien des informations supplémentaires sur le comportement des individus dans la prise de décision.

L'estimation de notre modèle de l'option "zéro" par l'approche bayésienne dans le cas dichotomique a fourni de meilleurs résultats d'un point de vue économique et statistique que la vraisemblance pénalisée. Cette dernière, qui est recommandée dans la littérature des modèles binaires logit avec séparation dans les données, n'a pas eu le même effet dans notre spécification dichotomique probit. Notre choix, orienté sur l'approche bayésienne, a été plus concluant et montre que la vraisemblance pénalisée, une alternative à la non convergence du maximum de vraisemblance due à la séparation des données, n'est pas forcément adaptée à toutes les spécifications. Dans notre cas précis, nous avons pu constater que l'approche bayésienne est préférable.

En comparant certains sites réels de VoD, à savoir Canalplay et TF1Vision et un site créé suite à notre étude, appelé Idéal, nous avons remarqué que Canalplay et TF1Vision se situent dans la partie refus de notre modèle, à savoir qu'ils ne sont pas choisis par les consommateurs. De plus, s'ils étaient choisis, ils seraient en conflit, c'est-à-dire trop similaires pour que l'individu puisse faire un choix avec certitude. L'introduction du conflit a donc un avantage, comme l'illustre cet exemple, à savoir qu'il permet pour deux concurrents dont l'offre de services est très similaire, de proposer le petit plus qui fera pencher la balance en sa faveur. Par exemple, il suffit qu'un de ces sites offre un cadeau ou une prestation supplémentaire gratuite pour que le choix des consommateurs s'oriente sur lui. Dans un marché où il n'y a que les sites Canalplay et TF1Vision, leurs parts de marché montrent une préférence pour le site TF1VIsion.

En conclusion, nous avons spécifié un modèle de l'option "zéro" qui comporte le refus et le conflit. Par cette spécification, nous avons confirmé l'intérêt de l'introduction du conflit dans les modèles de prise de décision. Une analyse intéressante, qui n'a pas pu être entreprise dans ce travail, aurait été de comparer notre modèle à d'autres spécifications, telles qu'un modèle présentant le choix des produits à l'aide des comparaisons par paires sans pour autant intégrer le refus et le conflit, ou un modèle de choix qui se présente sous la forme de comparaisons par paires, mais qui propose uniquement le refus et pas le conflit ou vice-versa. D'autres perspectives pourraient être entreprises. Elles se situent du point de vue économétrique et informatique. En effet, notre modèle pourrait être spécifié sous la forme logit et estimé par l'approche bayésienne et celle du maximum de vraisemblance pénalisée pour constater la supériorité des méthodes. Pour ce faire, il faudrait que la procédure du maximum de vraisemblance pénalisée soit implémentée dans un logiciel et commercialisée. La programmation de cette procédure à un cas multinomial pourrait, en effet, être d'un grand intérêt.
L'extension de notre modèle de l'option "zéro" à une comparaison par triplets des produits avec introduction du refus et du conflit serait également une perspective intéressante. La réalisation graphique serait plus complexe, ainsi que le développement des probabilités associées au cinq alternatives proposées dans un questionnaire.

Annexes

Annexe 1

Tableau tiré de l'ouvrage de Montgomery (1976) sur le nombre de carrés latins standards et non standards selon le nombre de niveaux :

Taille	3×3	4×4	5×5	6×6	7×7	$p \times p$
Exemples de carrés latins	A B C B C A C A B	A B C D B C D A C D A B D A B C	A B C D E B A E C D C D A E B D E B A C E C D B A	A B C D E F B C F A D E C F B E A D D E A B F C E A D F C B F D E C B A	A B C D E F G B C D E F G A C D E F G A B D E F G A B C E F G A B C D F G A B C D E G A B C D E F	A B C . . . P B C D . . . A C D E . . . B ⋮ P A B . . . (P-1)
Nombre de carrés latins standards	1	4	56	9408	16 942 080	-
Nombre total de carrés latins	12	576	161 280	818 851 200	61 479 419 904 000	$p!(p-1)!\times$ (le nombre de carrés latins standards)

Annexe 2

Tableau tiré de l'ouvrage de Federer (1955) récapitulant les plans en bloc incomplets équilibrés concevables selon les différents paramètres[7] :

Résolu							
a	b	r	k	a	b	r	k
4	6	3	**2**	13	26	6	3
4	4	3	3	13	13	4	4
5	10	4	**2**	13	13	9	9
5	10	6	3	15	35	7	3
5	5	4	4	15	15	7	7
6	15	5	**2**	15	15	8	8
6	10	5	3	16	20	5	4
6	15	10	4	16	24	9	6
6	6	5	5	16	16	6	6
7	21	6	**2**	16	16	10	10
7	7	3	3	19	57	9	3
7	7	4	4	19	19	9	9
7	7	6	6	19	19	10	10
8	28	7	**2**	21	70	10	3
8	14	7	4	21	21	5	5
8	8	7	7	21	30	10	7
9	36	8	**2**	25	50	8	4
9	12	4	3	25	30	6	5
9	18	8	4	25	25	9	9
9	18	10	5	28	63	9	4
9	12	8	6	28	36	9	7
9	9	8	8	31	31	6	6
10	45	9	**2**	31	31	10	10
10	30	9	3	37	37	9	9
10	15	6	4	41	82	10	5
10	18	9	5	49	56	8	7
10	15	9	6	57	57	8	8
10	10	9	9	64	72	9	8
11	55	10	**2**	73	73	9	9
11	11	5	5	81	90	10	9
11	11	6	6	91	91	10	10
11	11	10	10				
	Non	résolu		Pas	de	solution	
36	45	10	8	15	21	7	5
46	46	10	10	21	28	8	6
46	69	9	6	22	22	7	7
51	85	10	6	29	29	8	8

[7] Ce tableau a peut-être été complété.

Annexe 3

La matrice des variables explicatives X $((H+1)\times S\times N)$ du modèle de l'analyse conjointe, ainsi que les vecteurs des paramètres β et des erreurs ε s'écrivent :

$$X = \begin{pmatrix} x_{011} \\ x_{111} \\ x_{211} \\ x_{311} \\ \vdots \\ x_{H11} \\ x_{021} \\ x_{121} \\ x_{221} \\ x_{321} \\ \vdots \\ x_{H21} \\ \vdots \\ x_{0SN} \\ x_{1SN} \\ x_{2SN} \\ x_{3SN} \\ \vdots \\ x_{HSN} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \quad \text{et} \quad \varepsilon = \begin{pmatrix} \varepsilon_{011} \\ \vdots \\ \varepsilon_{HSN} \end{pmatrix}$$

où $x_{hsi} = \begin{pmatrix} x^1_{hsi} & \dots & x^K_{hsi} \end{pmatrix}$

L'utilité de l'alternative h du sous-ensemble s de l'individu i s'exprime en fonction de la matrice X :

$$u_{hsi} = X\beta + \varepsilon$$

Dans le modèle MNL de l'analyse conjointe, on estime les paramètres en modifiant la matrice x_0 de sorte qu'elle soit égale à 0. Par conséquent, la matrice X devient la matrice X^* $(H \times S \times N)$:

$$X^* = \begin{pmatrix} 0 \\ x_{111} - x_{011} \\ x_{211} - x_{011} \\ x_{311} - x_{011} \\ \vdots \\ x_{H11} - x_{011} \\ 0 \\ x_{121} - x_{021} \\ x_{221} - x_{021} \\ x_{321} - x_{021} \\ \vdots \\ x_{H21} - x_{021} \\ \vdots \\ x_{1SN} - x_{0SN} \\ 0 \\ x_{2SN} - x_{0SN} \\ x_{3SN} - x_{0SN} \\ \vdots \\ x_{HSN} - x_{0SN} \end{pmatrix}$$

Les matrices X_{hl}, $h, l = 1, \ldots, H$ et $l > h$ décrites dans les modèles MNL à effets croisés avec l'option "zéro" s'expriment comme suit :

$$X_{12} = \begin{pmatrix} x_{011} \\ \mathbf{x_{211}} \\ \mathbf{x_{111}} \\ x_{311} \\ \vdots \\ x_{H11} \\ x_{021} \\ \mathbf{x_{221}} \\ \mathbf{x_{121}} \\ x_{321} \\ \vdots \\ x_{H21} \\ \vdots \\ x_{0SN} \\ \mathbf{x_{2SN}} \\ \mathbf{x_{1SN}} \\ x_{3SN} \\ \vdots \\ x_{HSN} \end{pmatrix} \quad X_{13} = \begin{pmatrix} x_{011} \\ \mathbf{x_{311}} \\ x_{211} \\ \mathbf{x_{111}} \\ \vdots \\ x_{H11} \\ x_{021} \\ \mathbf{x_{321}} \\ x_{221} \\ \mathbf{x_{121}} \\ \vdots \\ x_{H21} \\ \vdots \\ x_{0SN} \\ \mathbf{x_{3SN}} \\ x_{2SN} \\ \mathbf{x_{1SN}} \\ \vdots \\ x_{HSN} \end{pmatrix} \quad \ldots \quad X_{1H} = \begin{pmatrix} x_{011} \\ \mathbf{x_{H11}} \\ x_{211} \\ x_{311} \\ \vdots \\ \mathbf{x_{111}} \\ x_{021} \\ \mathbf{x_{H21}} \\ x_{221} \\ x_{321} \\ \vdots \\ \mathbf{x_{121}} \\ \vdots \\ x_{0SN} \\ \mathbf{x_{HSN}} \\ x_{2SN} \\ x_{3SN} \\ \vdots \\ \mathbf{x_{1SN}} \end{pmatrix} \quad \text{etc.}$$

Annexe 4

CHOIX 1	**Produit 1**	**Produit 199**
	1000 programmes	600 programmes
	100% Films	75% Films 25% Séries
	Maximum nouveautés	Moitié du maximum nouveautés
	Payant	Gratuit pub
	24 heures	48 heures
	Bande annonce	Bande annonce
CHOIX 2	**Produit 1**	**Produit 214**
	1000 programmes	600 programmes
	100% Films	75% Films 25% Séries
	Maximum nouveautés	Que anciens
	Payant	Abonnement
	24 heures	24 heures
	Bande annonce	Aucun
CHOIX 3	**Produit 1**	**Produit 232**
	1000 programmes	600 programmes
	100% Films	50% Films 50% Séries
	Maximum nouveautés	Moitié du maximum nouveautés
	Payant	Payant
	24 heures	48 heures
	Bande annonce	Aucun
CHOIX 4	**Produit 17**	**Produit 261**
	1000 programmes	600 programmes
	100% Films	100% Séries
	Moitié du maximum nouveautés	Maximum nouveautés
	Gratuit pub	Abonnement
	24 heures	24 heures
	Bande annonce	Bande annonce
CHOIX 5	**Produit 49**	**Produit 246**
	1000 programmes	600 programmes
	75% Films 25% Séries	50% Films 50% Séries
	Moitié du maximum nouveautés	Que anciens
	Payant	Gratuit pub
	24 heures	24 heures
	Bande annonce	Aucun

CHOIX 6	**Produit 37**	**Produit 263**
	1000 programmes	600 programmes
	75% Films 25% Séries	100% Séries
	Maximum nouveautés	Maximum nouveautés
	Payant	Abonnement
	24 heures	48 heures
	Bande annonce	Bande annonce
CHOIX 7	**Produit 97**	**Produit 259**
	1000 programmes	600 programmes
	50% Films 50% Séries	100% Séries
	Que anciens	Maximum nouveautés
	Payant	Gratuit pub
	24 heures	48 heures
	Bande annonce	Bande annonce
CHOIX 8	**Produit 73**	**Produit 253**
	1000 programmes	600 programmes
	50% Films 50% Séries	100% Séries
	Maximum nouveautés	Maximum nouveautés
	Payant	Payant
	24 heures	24 heures
	Bande annonce	Bande annonce
CHOIX 9	**Produit 1**	**Produit 172**
	1000 programmes	600 programmes
	100% Films	100% Films
	Maximum nouveautés	Que anciens
	Payant	Payant
	24 heures	48 heures
	Bande annonce	Aucun
CHOIX 10	**Produit 5**	**Produit 156**
	1000 programmes	600 programmes
	100% Films	100% Films
	Maximum nouveautés	Maximum nouveautés
	Gratuit pub	Abonnement
	24 heures	48 heures
	Bande annonce	Aucun
CHOIX 11	**Produit 5**	**Produit 58**
	1000 programmes	1000 programmes
	100% Films	75% Films 25% Séries
	Maximum nouveautés	Moitié du maximum nouveautés
	Gratuit pub	Abonnement
	24 heures	24 heures
	Bande annonce	Aucun

CHOIX 12	**Produit 13**	**Produit 63**
	1000 programmes	1000 programmes
	100% Films	75% Films 25% Séries
	Moitié du maximum nouveautés	Que anciens
	Payant	Payant
	24 heures	48 heures
	Bande annonce	Bande annonce
CHOIX 13	**Produit 1**	**Produit 81**
	1000 programmes	1000 programmes
	100% Films	50% Films 50% Séries
	Maximum nouveautés	Maximum nouveautés
	Payant	Abonnement
	24 heures	24 heures
	Bande annonce	Bande annonce
CHOIX 14	**Produit 145**	**Produit 260**
	600 programmes	600 programmes
	100% Films	100% Séries
	Maximum nouveautés	Maximum nouveautés
	Payant	Gratuit pub
	24 heures	48 heures
	Bande annonce	Aucun
CHOIX 15	**Produit 41**	**Produit 107**
	1000 programmes	1000 programmes
	75% Films 25% Séries	50% Films 50% Séries
	Maximum nouveautés	Que anciens
	Gratuit pub	Abonnement
	24 heures	48 heures
	Bande annonce	Bande annonce
CHOIX 16	**Produit 205**	**Produit 254**
	600 programmes	600 programmes
	75% Films 25% Séries	100% Séries
	Que anciens	Maximum nouveautés
	Payant	Payant
	24 heures	24 heures
	Bande annonce	Aucun
CHOIX 17	**Produit 229**	**Produit 264**
	600 programmes	600 programmes
	50% Films 50% Séries	100% Séries
	Moitié du maximum nouveautés	Maximum nouveautés
	Payant	Abonnement
	24 heures	48 heures
	Bande annonce	Aucun

CHOIX 18	**Produit 221**	**Produit 262**
	600 programmes	600 programmes
	50% Films 50% Séries	100% Séries
	Maximum nouveautés	Maximum nouveautés
	Gratuit pub	Abonnement
	24 heures	24 heures
	Bande annonce	Aucun
CHOIX 19	**Produit 1**	**Produit 21**
	1000 programmes	1000 programmes
	100% Films	100% Films
	Maximum nouveautés	Moitié du maximum nouveautés
	Payant	Abonnement
	24 heures	24 heures
	Bande annonce	Bande annonce
CHOIX 20	**Produit 1**	**Produit 29**
	1000 programmes	1000 programmes
	100% Films	100% Films
	Maximum nouveautés	Que anciens
	Payant	Gratuit pub
	24 heures	24 heures
	Bande annonce	Bande annonce
CHOIX 21	**Produit**	**Produit**
	1000 programmes	600 programmes
	75% Films 25% Séries	75% Films 25% Séries
	Maximum nouveautés	Maximum nouveautés
	Gratuit pub	Payant
	48 heures	24 heures
	Bande annonce	Aucun

Annexe 5

VIDEO
on
DEMAND
Recherche rapide
OK
FILMS
Nouveautés
Genres
SERIES
Nouveautés
Genres
Inscription
Contact
THE DAVINCI CODE
THE DEVIL WEARS PRADA
MERYL STREEP
ANNE HATHAWAY
LOST
SEASON 2
EXTENDED EXPERIENCE
PRISON BREAK
FRIENDS

1

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	70 Nouveautés-Films (05-06) 48 Nouveautés-Séries (05-06) 380 Films (paru avant 05 au ciné) 102 Séries (paru avant 05 à la TV)
Payant, prix par film	Gratuit avec coupure publicité (excepté les nouveautés-films)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

2

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	450 Films (paru avant 05 au ciné) 150 Séries (paru avant 05 à la TV)
Payant, prix par film	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

3

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	70 Nouveautés-Films (05-06) 48 Nouveautés-Séries (05-06) 230 Films (paru avant 05 au ciné) 252 Séries (paru avant 05 à la TV)
Payant, prix par film	Payant, prix par film ou par épisode (série)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

4

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
70 Nouveautés-Films (05-06) 930 Films (paru avant 05 au ciné)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

5

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
70 Nouveautés-Films (05-06) 48 Nouveautés-Séries (05-06) 680 Films (paru avant 05 au ciné) 202 Séries (paru avant 05 à la TV)	300 Films (paru avant 05 au ciné) 300 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Gratuit avec coupure publicité
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

6

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 610 Films (paru avant 05 au ciné) 154 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

7

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
500 Films (paru avant 05 au ciné) 500 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Gratuit avec coupure publicité
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible

❑Aucun ❑Les deux

8

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 360 Films (paru avant 05 au ciné) 404 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Payant, prix par épisode (série)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible

❑Aucun ❑Les deux

9

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	600 Films (paru avant 05 au ciné)
Payant, prix par film	Payant, prix par film
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

10

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	140 Nouveautés-Films (05-06) 460 Films (paru avant 05 au ciné)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

11

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	70 Nouveautés-Films (05-06) 48 Nouveautés-Séries (05-06) 680 Films (paru avant 05 au ciné) 202 Séries (paru avant 05 à la TV)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

12

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
70 Nouveautés-Films (05-06) 930 Films (paru avant 05 au ciné)	750 Films (paru avant 05 au ciné) 250 Séries (paru avant 05 à la TV)
Payant, prix par film	Payant, prix par film ou par épisode (série)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

13

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 360 Films (paru avant 05 au ciné) 404 Séries (paru avant 05 à la TV)
Payant, prix par film	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

14

❑Produit 1	❑Produit 2
600 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 460 Films (paru avant 05 au ciné)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film	Gratuit avec coupure publicité
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

15

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 610 Films (paru avant 05 au ciné) 154 Séries (paru avant 05 à la TV)	500 Films (paru avant 05 au ciné) 500 Séries (paru avant 05 à la TV)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible
❑Aucun	❑Les deux

16

❑Produit 1	❑Produit 2
600 programmes dont	600 programmes dont
450 Films (paru avant 05 au ciné) 150 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Payant, prix par épisode (série)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

17

❑Produit 1	❑Produit 2
600 programmes dont	600 programmes dont
70 Nouveautés-Films (05-06) 48 Nouveautés-Séries (05-06) 230 Films (paru avant 05 au ciné) 252 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Payant, prix par film ou par épisode (série)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	48h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

18

❑Produit 1	❑Produit 2
600 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 160 Films (paru avant 05 au ciné) 204 Séries (paru avant 05 à la TV)	96 Nouveautés-Séries (05-06) 504 Séries (paru avant 05 à la TV)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

19

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	70 Nouveautés-Films (05-06) 930 Films (paru avant 05 au ciné)
Payant, prix par film	Abonnement, prix par mois (nombre fixe de programmes loués)
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible

❑Aucun ❑Les deux

20

❑Produit 1	❑Produit 2
1000 programmes dont	1000 programmes dont
140 Nouveautés-Films (05-06) 860 Films (paru avant 05 au ciné)	1000 Films (paru avant 05 au ciné)
Payant, prix par film	Gratuit avec coupure publicité
24h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	Bande annonce disponible

❑Aucun ❑Les deux

21

❑Produit 1	❑Produit 2
1000 programmes dont	600 programmes dont
140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 610 Films (paru avant 05 au ciné) 154 Séries (paru avant 05 à la TV)	140 Nouveautés-Films (05-06) 96 Nouveautés-Séries (05-06) 310 Films (paru avant 05 au ciné) 54 Séries (paru avant 05 à la TV)
Gratuit avec coupure publicité (excepté les nouveautés-films)	Payant, prix par film ou par épisode (série)
48h de conservation après le 1er visionnage	24h de conservation après le 1er visionnage
Bande annonce disponible	
❑Aucun	❑Les deux

Données personnelles

Age: ☐ Sexe: ☐ M ☐ F

Avez-vous internet à la maison? ☐ Oui ☐ Non

Si oui, utilisez-vous internet? ☐ Parfois ☐ Moyennement ☐ Souvent

Avez-vous déjà entendu parler de la vidéo à la demande (VoD)? ☐ Oui ☐ Non

Avez-vous déjà eu l'occasion d'utiliser un système de VoD? ☐ Oui ☐ Non

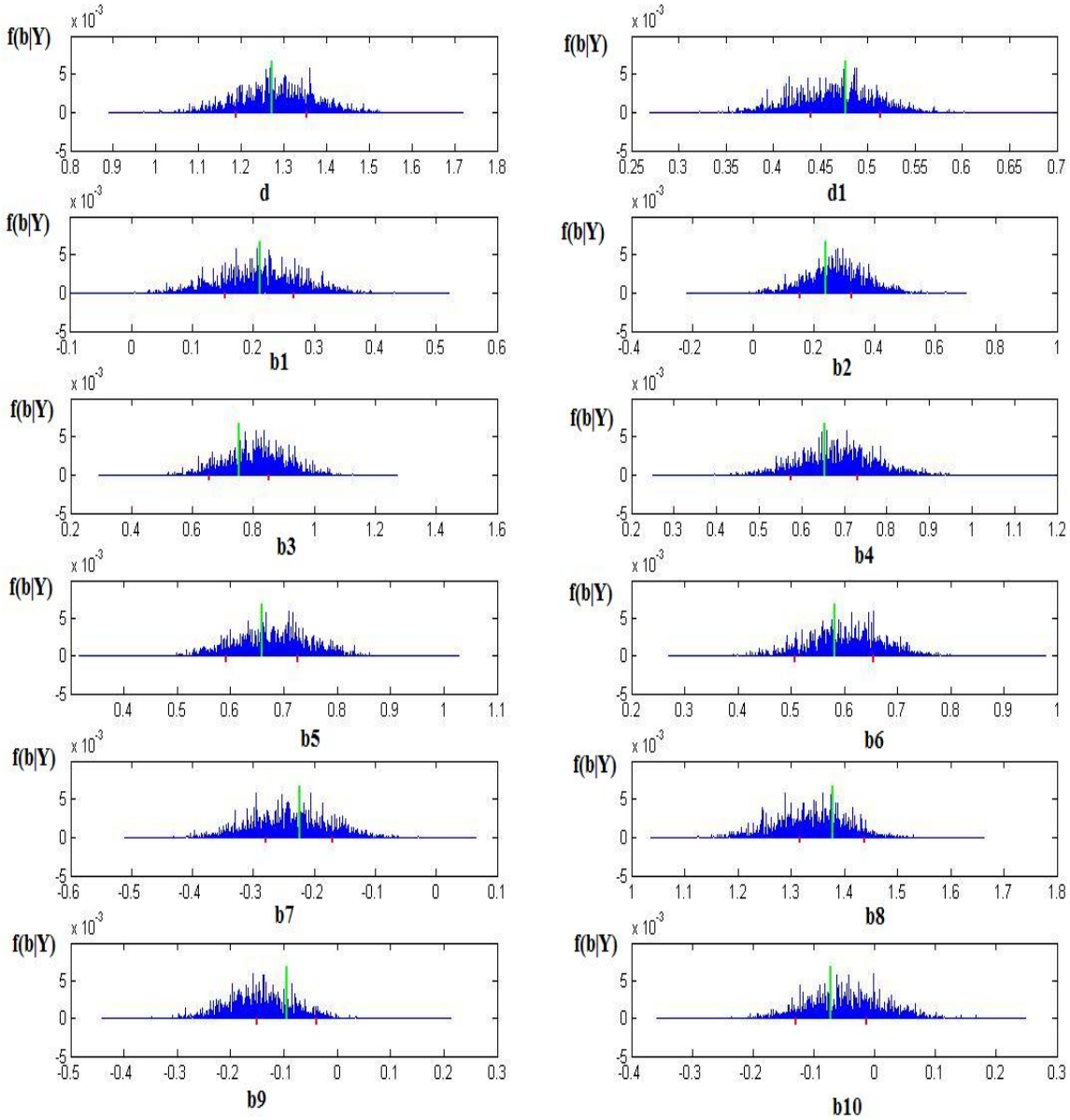
Avez-vous de l'ADSL à la maison? ☐ Oui ☐ Non

Connaissez-vous le débit internet que vous avez à la maison? ☐ Oui ☐ Non

Si, oui de combien est-il? ☐

Annexe 6

Ces graphiques présentent les valeurs de la distribution *a posteriori* $f(\beta|Y)$ en fonction de chaque paramètre $\beta' = \begin{pmatrix} d & d1 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 & \beta_{10} \end{pmatrix}$. Le trait vert représente le mode calculé par l'approche bayésienne et les points-traits rouges aux 25% des observations en dessous et en dessus du mode. Nous n'illustrons que les graphiques de l'individu 2 du modèle $y = 1, 2, 3, 4$. Pour les autres individus, le principe est le même.



Annexe 7

CONVERGENCE ACHIEVED AFTER 19 ITERATIONS

618 FUNCTION EVALUATIONS

Parameter	Estimate	Standard Error	t-statistic	P-value
D	1.308	0.112	11.638	[0.000]
D1	0.471	0.053	8.825	[0.000]
B1	0.212	0.085	2.493	[0.013]
B2	0.262	0.121	2.163	[0.030]
B3	0.800	0.128	6.239	[0.000]
B4	0.676	0.112	6.009	[0.000]
B5	0.690	0.090	7.659	[0.000]
B6	0.625	0.088	7.080	[0.000]
B7	-0.235	0.079	-2.965	[0.003]
B8	1.350	0.081	16.615	[0.000]
B9	-0.114	0.078	-1.450	[0.147]
B10	-0.053	0.082	-0.648	[0.516]

où

D	Refus
D1	Conflit
B1	Catalogue : 1000
B2	Composition : 100% Films
B3	Composition : 75% Films 25% Séries
B4	Composition : 50% Films 50% Séries
B5	Nouveauté : max
B6	Nouveauté : moitié
B7	Tarif : payant
B8	Tarif : gratuit pub
B9	Conservation : 24h
B10	Disponibilité : Bande annonce

Bibliographie

Albert A. et Anderson J. A. (1984), "On the existence of maximum likelihood estimates in logistic regression models", *Biometrika*, 71, 1-10

Altman M., Gill J., McDonald M. P. (2004), "*Numerical issues in statistical computing for the social scientist*", Chapter 10 : Convergence Problems in Logistic Regression, Allison P., Wiley

Amemiya (1985), "*Advanced Econometrics*", Cambridge, MA : Harvard University Press

Anderson (1962) N. H., "Application of an additive model to impression formation", *Science*, 138, 817-18

Anderson N. H. (1970), "Functional measurement and psychophysical judgment", *Psychological Review*, 77, 153-70

Anderson N. H. (1981), "*Foundations of Information Integration Theory*", New York : Academic Press

Anderson N. H. (1982), "*Methods of Information Integration Theory*", New York : Academic Press",

Anderson S. P., De Palma A. et Thisse J. F. (1992), "*Discrete Choice of Product Differentiation*", Cambridge MIT Press

Bagozzi R. P. (1994), "*Advanced Methods of Marketing Research*", Blackwell Business

Baron J. et Ritov I. (1994), "Reference Points and Omission Bias", *Organizational Behavior and Human Decision Processes*, 59 (September), 475-498

Beattie J. et Barlas S. (1992), "Predicting Perceived Differences in Tradeoff Difficulty", *working paper*, University of Sussex

Ben-Akiva M., Lerman S. R. (1985), *Discrete Choice Analysis : Theory and application to travel demand*, Mit Press, Cambridge, MA

Benammou S., Saporta G., Swissi B. (2007), "Une procédure de réduction du nombre de paires en analyse conjointe", *Journal de la Société Française de Statistique*, tome 148, n°4, 57-76

Bull S. B., Mak C., Greenwood C. M. T. (2002), "A modified score function estimator for multinomial logistic regression in small samples", *Computational Statistics and Data Analysis*, 39, 57-74

Carroll J. D. (1972), "*Individual Differences and Multidimensional Scaling*", In : R. N. Shepard, A. K. Romney, and S. B. Nerlove (eds.), Multidimensional Scaling : Theory and Applications in the Behavioral Sciences (Vol. 1), New York : Seminar Press

Cattin P. et Wittink D. R. (1982), "Commercial use of conjoint analysis : a survey", *Journal of Marketing*, 46, 44-53

Cattin P. et Wittink D. R. (1986), "Commercial use of conjoint analysis : an update", *Presented at the ORSA/TIMS Marketing Science Meetings*, Richardson, TX, March 12-15

Clogg C. C., Rubin D. B., Schenker N., Schultz B. et Weidman L. (1991), "Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression", *Journal of the American Statistical Association*, 86, 68-78

Congdon P. (2001), "*Bayesian Statistical Modelling*, Chichester, UK : John Wiley

Cox D. R. (1970), "*Analysis of Binary Data*", New York : Wiley

Dagnelie P. (2003), *Principe d'expérimentation : Planification des expériences et analyse de leurs résultats*, Les presses agronomiques de Gembloux

De Sarbo, Ramaswamy et Cohen (1995), "Market Segmentation with Choice-Based Conjoint Analysis", *Marketing Letters*, 6(2), 137-148

Dhar R. (1997), "Consumer Preference for a No-Choice Option", *Journal of Consumer Research*, Vol. 24, Septembre

Domencich T. et McFadden D. (1975), "*Urban Travel Demand : A Behavioral Analysis*", North-Holland

Dussaix A.-M., Saporta G., Carle P., Darmon R.-Y., Grimmer J.-F., Morineau A. (1998), "*L'Analyse Conjointe, la Statistique et le Produit Idéal : Méthodes et Applications*, Cisia·Ceresta

Elrod T., Louviere J. J. et Davey K. S. (1992), "An Empirical Comparison of Rating-Based and Choice-Based Conjoint Models",*Journal of Marketing Research*, 29 (3), 368-377

Elrod T., Louviere J. J. et Krishnakumar S. D. (1992), "An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models", *Journal of Marketing Research*, Vol. XXIX, Août

Evrard Y., Pras B. , Roux E. (2003), "*MARKET : Etudes et Recherches en Marketing*", 3e Edition, Dunod

Faivre J.-Ph., Pioche A. (1976), "L'Analyse des Décisions d'Achat par le Modèle Trade-off : Problèmes Théoriques et Méthodologiques", *Revue Française du Marketing*, Cahiers 64-65

Federer W. T. (1955), "*Experimental Design : Theory and Application*", The Macmillan Company

Festinger L. (1957), "*A Theory of Cognitive Dissonance*", ROW, Peterson and Company

Festinger L. (1964), "*Conflict, Decision and Dissonance*", Stanford University Press

Firth D. (1993), "Bias reduction of maximum likelihood estimates", *Biometrika*, 80, 1, pp. 27-38

Fishman G. S. (2006), "*A First Course in Monte Carlo*", Duxbury Advanced Series

Furlan R. et Corradetti R. (2004), "An Empirical Comparison of Conjoint Analysis Models on a Same Sample", *Giornata di Studio "La Conjoint Analysis" 27 septembre 2004*, Universita degli Studi Salerno

Galindo-Garre F., J. K. Vermunt et W. P. Bergsma (2004), "Bayesian Posterior Estimation of Logit Parameters with Small Samples", *Sociological Methods and Research*, Vol. 33, No 1, 88-117

Geweke J., Keane M. et Runkle D. (1994), "Alternative Computational Approaches to Inference in the Multinomial Probit Model", *Review of Economics and Statistics*, no 76, 609-632

Green P. E (1984), "Hybrid Models for Conjoint Analysis : An Expository Review", *Journal of Marketing Research*, 21(May), 155-169

Green P. E., Goldberg S. M. et Montemayos M. (1981), "A Hybrid Utility Estimation Model for Conjoint Analysis", *Journal of Marketing*, 45(Winter), 33-41

Green P. E. et Krieger A. M. (1993), "*Conjoint Analysis with Product-Positioning Applications*", In : J. Eliashberg and G. L. Lilien (eds), Handbooks in Operations Research and Management Science, Vol. 5, Marketing, North-Holland, Amsterdam

Green P. E. et Rao V. R. (1971), "Conjoint Measurement for Quantifying Judgemental Data", *Journal of Marketing Research*, 8, 355-363

Green P. E. et Srinivasan V. (1978), "Conjoint Analysis in Consumer Research : Issues and Outlook", *Journal of Consumer Research*, 5, 103-123

Green P. E. et Wind Y. (1973), "*Multiattribute Decisions in Marketing : A Measurement Approach*", Hinsdale, IL : Dryden Press

Greene W. H. (1993), "*Econometrics analysis*", Fifth Edition, Prentice Hall

Gustafsson A., Herrmann A. , Huber F. (2003), "*Conjoint Measurement : Methods and Applications*", Third Edition, Springer

Haaijer M. E. (1999), "*Modeling Conjoint Choice Experiments with the Probit Model*", Labyrint Publications

Hajivassiliou V., McFadden D. et Ruud P. (1992), "Simulation of Multivariate Normal Orthant Probabilities : Methods and Programs", *Document de travail 1021*, Yale University, Cowles Foundation

Hajivassiliou V. et Ruud P. A. (1994), "Classical Estimation Methods for LDV Models using Simulation", *Handbook of Econometrics IV*, Amsterdam, North Holland

Hammond K. R. (1955), "Probabilistic functioning and the clinical method", *Psychological Review*, 62, 255-62

Harbi S. (2004), *Sur Différentes Approches Multidimensionnelles de Liaisons entre Deux Ensembles et leurs Applications à l'Analyse Conjointe : Régression PLS / Approche du PCAR*, Thèse, Université de Tunis El Manar, Faculté de Sciences Economiques et de Gestion de Tunis, Sous la Direction des Professeurs Gilbert Saporta et Samir Ben Ammou

Heinze G. (2006), "A comparative investigation of methods for logistic regression with separated or nearly separated data", *Statistics in Medecine*, 25, 4216-4226

Heinze G., Ploner M. (2004), "Technical Report 2/2004 : A SAS macro, S-PLUS library and R package to perform logistic regression without convergence problems", *Section of Clinical Biometrics, Department of Medical Computer Sciences, Medical University of Vienna*, Vienna, http ://www.meduniwien.ac.at/msi/biometrie/programme/fl/tr2_2004.pdf

Heinze G., Ploner M. (2003), "Fixing the nonconvergence bug in logistic regression with SPLUS and SAS", *Computer Methods and Programs in Biomedicine*, 71, 181-187

Heinze G., Schemper M. (2002), "A solution to the problem of separation in logistic regression", *Statistics in Medicine*, 21, 2409-2419

Hoffman P. J. (1960), "The paramorphic representation of clinical judgment", *Psychological Bulletin*, 57, 116-31

Hoffman P.J., Slovic P. et Rorer L.G. (1968), "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment", *Psychological Bulletin*, 69, 338-49

Huber J. (1997), "What we have learned from 20 years of Conjoint Research : When to use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments", *Sawtooth Software Conference Proceedings*

Janis I. L. et Mann L. (1977), "*Decision Making*", New York : Free Press

Johnson R. M. (1974), "Trade-off Analysis of Consumer Values", *Journal of Marketing Research*, 11, 121-127

Johnson R. M. (1997), "Individual Utilities from Choice Data : a New Method", *Sawtooth Software Conference Proceedings*

Johnson R. M. (1997), "ICE : Individual Choice Estimation", *Technical Paper Series*, Sawtooth Software, Inc.

Kamakura W. A. et Srivastava R. K. (1984), "Predicting Choice Shares Under Conditions of Brand Interdependence", *Journal of Marketing Research*, 21, November, 420-434

Kamakura W. A., Wedel M. et Agrawal J. (1994), "Concomitant Variable Latent Class Models for Conjoint Analysis", *International Journal of Research in Marketing*, 11(5), 451-464

Karni E. et Schwarz A. (1977), "Search Theory : The Case of Search with Uncertain Recall", *Journal of Economic Theory*, 16 (October), 38-52

Kessels R., Goos P. and Vandebroek M. (2008), "Optimal designs for conjoint experiments", *Computational Statistics and Data Analysis*, 52, 2369-2387

Kruskal J. B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data", *Journal of the Royal Statistical Society*, Series B (March), 251-263

Kuhl J. (1986), "*Human Motivation : From Decision Making to Action Control*", in New Directions in Research on Decision Making, ed. B. Brehmer, H. Jungermann, P. Lourens, and G. Sevon, Amsterdam : North-Holland

Kuhfeld W. F. (2005), "*Marketing Research Methods in SAS : Experimental Design, Choice, Conjoint, and Graphical Techniques*", SAS 9.1 Edition, SAS Institute Inc., Cary, NC, USA, http ://support.sas.com/techsup/technote/ts722.pdf

Lancaster K. (1966), "A new approach to consumer theory", *Journal of Political Economy*, 74, 132-57

Liquet J.-C. (2001), "*Cas d'Analyse Conjointe*", Editions TEC&DOC

Lollivier S. (2006), "*Econométrie avancée des variables qualitatives*", Economica

Louviere J. J. (1988a), "*Analyzing Individual Decision Making : Metric Conjoint Analysis*", Sage University Series on Quantitative Applications in the Social Sciences, Series No. 67, Newbury Park, Ca : Sage Publications , Inc.

Louviere J. J. (1988b), "Conjoint analysis modeling of stated preference : a review of theory, methods, recent developments and external validity", *Journal of Transport Economics and Policy*, 10, 93-119

Louviere J. J., Hensher D. A., Swait J. D. (2000), "*Stated Choice Methods : Analysis and Application*", Cambridge University Press

Louviere J. J. et Timmermans H. J. P. (1990), "A Review of Recent Advances in Decompositional Preference and Choice Models", *Tijdschrift voor Economische en Sociale Geografie*, 81(3), 214-225

Louviere J. J. et Woodworth G. (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments : An Approach Based on Aggregate Data", *Journal of Marketing Research*, Vol. 20, No.4, 350-67

Luce R. D. (1959), "*Individual Choice Behavior*", New York, Wiley

Luce R. D. et Tukey J. W. (1964), "Simultaneous Conjoint Measurement-A New Type of Fundamental Measurement", *Journal of Mathematical Psychology*, 1, 1-27

Maddala G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press

Mahagan V., Green P. E. et Goldberg S. M. (1982), "A Conjoint Model for Measuring Self and Cross-Price/Demand Relationships", *Journal of Marketing Research*, 19(3), 334-342

McFadden D. (1974), "The Measurement of Urban Travel Demand", *Journal of Public Economics*, 3, 303-28

McFadden D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, no 57, 995-1026

Mehta C. R., Patel N. R. (1995), "Exact Logistic regression : theory and examples", *Statistics in Medicine*, vol. 14, 2143-2160,

Montgomery D. C. (1976), *Design and Analysis of Experiments*, Second Edition, Wiley

Montgomery H. (1989), "*From Cognition to Action : The Search for Dominance in Decision Making*", in Process and Structure in Human Decision Making, ed. Henry Montgomery and Ola Svenson, New York : Wiley

Orme B. (2000), "Hierarchical Bayes : Why All the Attention ?", *Research Paper Series*, Sawtooth Software, Inc.

Pekelman D. et Sen S. (1974), "Mathematical Programming Models for the Determination of Attribute Weights", *Management Science*, 20, 1217-1229

Ritov I. et Baron J. (1990), "Reluctance to Vaccinate : Omission Bias and Ambiguity", *Journal of Behavioral Decision Making*, 3 (December), 263-277

Saporta G. (2006), "*Probabilités, analyse des données et statistique*", 2ème édition, Technip

Scholnick E. K. et Wing C. S. (1988), "Knowing When You Don't Know : Developmental and Situational Considerations", *Developmental Psychology*, 24 (March), 190-196

Sjoberg L. (1980), "Volitional Problems in Carrying through through a Difficult Decision", *Acta Psychologica*, 12 (August), 123-132

Slovic P. et Lichtenstein S. (1971), "Comparison of Bayesian and regression approaches to the study of information processing in judgment", *Organizational Behavior and Human Performance*, 6, 649-744

Spranca M., Minsk E. et J. Baron (1991), "Omission and Commission in Judgment and Choice", *Journal of Experimental Social Psychology*, 27 (January), 76-105

Srinivasan V. (1980), "Comments on "On Conjoint Analysis and Quantal Choice Models"", *Journal of Business*, 53(3.2), S46-S50

Srinivasan V. et Shocker A. D. (1973a), "Linear Programming Techniques for Multidimensional Analysis of Preferences", *Psychometrika*, 38, 337-369

Srinivasan V. et Shocker A. D. (1973b), "Estimating the Weights for Multiple Attributes in a Composite Citerion Using Pairwise Judgments", *Psychometrika*, 38, 473-493

Srinivasan A. D. et Shocker V. (1977), "LINMAP (Version II) : An FORTRAN IV computer program for analyzing ordinal preference (dominance) judgments via linear programming techniques for conjoint measurement", *Journal of Marketing Research*, 14, 101-103

Stern S. (1992), "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models", *Econometrica*, no 60, 943-952

Thomas A.(2000), "*Econométrie des variables qualitatives*", Dunod

Thurstone L. (1927), "A Law of Comparative Judgment", *Psychological Review*, no 34, 273-286

Tversky A. et Shafir E. (1992), "Choice under Conflict : The Dynamics of Deferred Decision", *Psychological Science*, Vol. 3, No 6 (November), 358-361

Wittink D. R. et Cattin P. (1981), "Alternative Estimation Methods for Conjoint Analysis : A Monte Carlo Study", *Journal of Marketing Research*, 18 (February), 101-106

Young F. W. (1972), "*A model for polynomial Conjoint Analysis algorithms*", in : Shepard, R., Romney, A. K. and Nerlove, S. B., eds., Multidimensional Scaling - Theory and Applications in Behavioral Sciences, New York, 69-104

Zorn C. (2005), "*A solution to separation in binary response models*", Political Analysis, Volume 13, Number 2, 157-170

(2000), "*Preference Structure Measurement : Conjoint Analysis and Related Techniques : A Guide for Designing and Interpreting Conjoint Studies*", Second Edition, Intelligest

Documentation du logiciel SAS

Documentation du logiciel TSP

Les journaux informatiques :

- L'Ordinateur Individuel de
  - mars 2006
  - mai 2006
  - juillet-août 2006
  - septembre 2006

  - octobre 2006
  - juin 2007
  - juillet-août 2007

- Le Hors-Série de L'Ordinateur Individuel (octobre-novembre 2006)

- P2PMag de
  - octobre 2006
  - décembre 2006-janvier 2007
  - juillet-août 2007

- Micro Actuel de
  - octobre 2006
  - novembre 2006
  - septembre 2007

- Micro hebdo (septembre 2006)

- Le Monde (janvier 2007)

- Le Temps (septembre 2006)

Les sites informatiques de VoD :

- Canalplay, TF1Vision, Ina, Vodeo, ArteVoD, VirginMega, France 2, France 3, France 5 Vidéos, VoD.Wanadoo (24h/24h Vidéo (Orange)), Imineo, Cinezime, SpiralFrog, joost, Allocine, Locafilm, Glowria, DVDFly, francetvod, ClicMovies, Neuf, Alice, Free, Club Internet, TPS-VoD, Orange, Homechoice, ABC, Cinemanow, netciné, clubic.com, lexpansion.com, linternaute.com, www.ratiatum.com

Le site informatique de l'encyclopédie libre de Wikipédia (français et américain)

Les sites informatiques sur les simulations de Monte Carlo :

- Computational Science Education Project (1991-1995), "*Introduction aux méthodes Monte Carlo*"