

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DE LA DEMANDE ET TARIFICATION DE SERVICES DE
TÉLÉPHONIE FIXE

GAUTHIER ABRIAL

DÉPARTEMENT DE DÉPARTEMENT DE MATHÉMATIQUES ET DE
GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLOME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)

AVRIL 2008

© Gauthier Abrial, 2008.



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-41545-0
Our file Notre référence
ISBN: 978-0-494-41545-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

MODÉLISATION DE LA DEMANDE ET TARIFICATION DE SERVICES DE
TÉLÉPHONIE FIXE

présenté par: ABRIAL Gauthier

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

M. ROUSSEAU Louis-Martin, Ph.D., président

M. SAVARD Gilles, Ph.D., membre et directeur de recherche

M. MARCOTTE Patrice, Ph.D., membre et codirecteur de recherche

M. BASTIN Fabian, Doct., membre

RÉSUMÉ

Le développement des technologies et services téléphoniques et l'augmentation du nombre d'utilisateurs durant les dernières années ont contribué à une forte complexification de l'offre. Les utilisateurs, confrontés à cette offre, adoptent des comportements de plus en plus complexes. Pour faire face à ce phénomène, l'industrie doit adopter des outils lui fournissant une compréhension plus fine de ces comportements. Ce document décrit les travaux effectués dans ce domaine par l'auteur dans le cadre d'un projet de maîtrise recherche (M.Sc.A.) en Mathématiques appliquées à l'École Polytechnique de Montréal. Nous y étudions un modèle de demande proposé par Iyengar. Ce modèle se penche sur les conséquences de la tarification non linéaire sur la modélisation de la demande. Nous développons ici l'exploitation des données mises à disposition par un grand fournisseur de services téléphoniques, la mise en œuvre d'une méthodologie d'estimation des paramètres et les perspectives d'étude.

ABSTRACT

The development of phone-based telecommunication technologies and services and the growth of the subscribers in recent years contribute to a significant increase in the complexity of the supply. The customer, facing this supply, tends to have a more complex behavior. To confront this phenomenon, industry has to adopt some tools enabling it to have a finer understanding of these behaviors. This paper depicts the study led by the author in this field as part of a Master of Applied Science in applied mathematics. We study a demand model suggested by Raghuram Iyengar. This model focuses on the impact of the non-linear pricing schemes on the demand modeling. We fit this model to the data from a service provider, describe the way we estimate the parameters of the model and discuss the opportunities for future research.

TABLE DES MATIÈRES

RÉSUMÉ	iv
ABSTRACT	v
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	ix
LISTE DES ANNEXES	x
INTRODUCTION	1
CHAPITRE 1 REVUE DE LITTÉRATURE	5
1.1 Revenu management	5
1.2 Tarification et gestion des volumes de vente	8
1.3 Modèle de demande	11
1.4 Programmation bi-niveau	13
CHAPITRE 2 CADRE THÉORIQUE	14
2.1 Tarification non linéaire et contraintes budgétaires	14
2.2 Présentation du modèle	17
CHAPITRE 3 EXPÉRIMENTATIONS NUMÉRIQUES	21
3.1 Les données	21
3.1.1 Description	21
3.1.2 Prétraitement des données	23
3.1.3 Description de l'offre	24
3.1.4 Analyse statistique	25

3.1.5	Choix des forfaits	27
3.1.6	Analyse statistique des forfaits choisis	27
3.1.7	Choix de l'échantillon	28
3.2	Adaptation du modèle	29
3.2.1	Choix de l'optimisateur	29
3.2.2	Utilisation de l'optimisateur	30
3.2.3	Calcul du log-vraisemblance et de ses dérivées	33
3.3	Considérations pratiques	43
3.3.1	Considérations sur les dérivées secondes	43
3.3.2	Problèmes relatifs aux données	44
3.3.3	Étude de la fonction objectif	45
3.3.4	Problèmes numériques	46
3.4	Résultats numériques	48
CHAPITRE 4	PERSPECTIVES D'ÉTUDE	52
4.1	Évolution mixed-logit	52
4.2	Généralisation du modèle	54
4.3	Facteurs de corrélation temporelle et autres facteurs décisionnels	56
CONCLUSION		59
RÉFÉRENCES		61
ANNEXES		64

LISTE DES TABLEAUX

Tableau 3.1	Nombre d'usagers par forfait et par mois	26
Tableau 3.2	Nombre d'usagers changeant de forfait chaque mois	26
Tableau 3.3	Nombre d'usagers quittant la compagnie chaque mois	27

LISTE DES FIGURES

Figure 2.1	Schéma de tarification linéaire	15
Figure 2.2	Schéma de tarification non linéaire	16
Figure 2.3	Ensemble de contraintes budgétaires	17
Figure A.1	$U_{ijt} = f(\alpha_{11})$	65
Figure A.2	$U_{ijt} = f(\alpha_{12})$	65
Figure A.3	$U_{ijt} = f(\alpha_2)$	66
Figure A.4	$U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.3005$	66
Figure A.5	$U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.2045$	67
Figure A.6	$U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.1085$	67
Figure A.7	$U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.0125$	68

LISTE DES ANNEXES

ANNEXE A	FONCTION D'UTILITÉ	64
----------	------------------------------	----

INTRODUCTION

Le *revenue management* (ou *gestion du revenu*) constitue la manière dont une entreprise gère les rentrées de fonds associées à ses produits ou services. Le bénéfice généré par une entreprise faisant intervenir en grande partie les revenus et les coûts, le *revenue management* constitue, avec le *supply-chain management*, un des principaux outils dont il est possible de se servir pour augmenter les bénéfices. Le *supply-chain management* s'attachera à diminuer les coûts en optimisant par exemple l'approvisionnement, la production, la manutention, l'entreposage ou le délivrement d'un produit. Le *revenue management* aura quant à lui pour mission d'adapter l'offre de manière à optimiser les revenus qu'elle génère.

Les décisions qu'il est amené à prendre sont multiples. Elles concernent de nombreux facteurs, et interviennent dans des processus variés. Elles découlent de questions telles que : "Quand vendre?", "A qui vendre?", " Que vendre?", "A quel prix l'offrir?", "A quel prix accepter de le vendre?" Ce sont des décisions aussi bien stratégiques, par exemple concernant la gamme de produit développée, la segmentation du marché, le positionnement sur le marché, que tactiques, par exemple quand doit-on procéder à des promotions, doit-on faire des *quantity discount*.

Le revenu généré dépend des prix de vente et des quantités vendues, qui elles mêmes dépendent de la demande. Le prix est une variable à laquelle le manager a directement accès, mais la demande dépend d'une multitude de facteurs, et en particulier des différentes décisions prises par le manager. Une connaissance suffisante du comportement de cette demande permet de prévoir comment celle-ci réagirait en fonction des décisions que le manager pourrait prendre. Bien plus que de la prévoir, il s'agit d'être en mesure de provoquer une réaction désirée de la demande, de manière à ce que celle-ci devienne une variable du problème à laquelle le ma-

nager a accès. Le problème du revenue management est donc en grande partie un problème de demand management.

Dans la pratique, un des outils qui, ces dernières années, a été étudié et mis en œuvre avec succès dans le cadre du revenue management est la programmation bi-niveau. Cet outil permet de modéliser très efficacement les jeux de Stackelberg. Il consiste, dans le contexte du revenue management, en un programme mathématique dans lequel, au premier niveau, une entreprise optimise ses revenus ou ses bénéfices par rapport à un certain nombre de variables auxquelles elle a directement accès, et au second niveau, un ensemble de consommateurs ou un segment de marché optimise une utilité par rapport à d'autres variables (par exemple la consommation des produits de l'entreprise), en fonction de l'état des variables choisi par l'entreprise au premier niveau (par exemple le prix des produits). Ici, la modélisation de la demande est donc un élément central du problème, puisque c'est à travers les prévisions du modèle de demande situé au second niveau que le premier niveau va déterminer les décisions optimales que l'entreprise doit prendre.

Prévoir les réactions d'un marché n'est pas une pratique nouvelle. La contribution de la science à cette pratique réside surtout dans le fait de réduire l'incertitude de ces prévisions en augmentant le nombre de facteurs pris en compte, de représenter des situations plus complexes, avec un plus grand nombre de consommateurs possédant des comportements différents, tout en ayant la possibilité de différencier la réaction de chaque groupe de consommateurs. Elle réside aussi dans la capacité à systématiser l'analyse des données sur de grands échantillons de marché.

Un des marchés qui possède ces caractéristiques de complexité, de diversité et de taille tout en alliant un besoin d'exactitude des prévisions de la demande est le marché de la téléphonie. Les fournisseurs de services téléphoniques proposent des offres comprenant un prix forfaitaire incluant un certain nombre de minutes d'ap-

pel vers une certaine région, des tarifs pour les appels vers d'autres régions et pour les appels excédants les minutes incluses, ainsi que divers options payantes comme le répondeur ou les services de messagerie, ces options pouvant être regroupées sous une même offre. Ceci conduit à des comportements de la demande complexes. Les consommateurs sont aussi bien des particuliers, étudiants, familles ou retraités, que des entreprises. La demande est donc très diversifiée. Quasiment tout le monde possède au moins une ligne téléphonique, ce qui fait de ce marché un très gros marché. Enfin, la concurrence est rude. Beaucoup d'utilisateurs changent de fournisseurs d'accès régulièrement au fil de l'évolution des offres, ce qui rend l'exactitude et la finesse des prédictions des modèles de comportement extrêmement importants pour dégager un avantage concurrentiel.

Notre étude se place dans la continuité de l'étude de Iyengar (Iyengar, 2004). Alors que l'auteur y propose un modèle et identifie ses principales propriétés, la mise en œuvre n'est que partiellement abordée. Nous verrons dans ce mémoire dans quelle mesure l'application aux données dont dispose une entreprise peut poser des problèmes, nous définirons un ensemble de pré-traitements, et nous parlerons de leurs impacts sur la modélisation. Nous mettrons en place une série de procédures numériques et nous les décrirons très précisément. Nous soulèverons aussi un certain nombre d'enjeux scientifiques sous forme de perspectives d'études dans lesquelles nous proposerons en particulier un modèle issu de la généralisation de celui étudié. Celui-ci serait destiné à représenter avec encore plus de finesse le comportement du marché de la téléphonie mobile.

Dans notre étude, nous développons donc un modèle de demande pour la téléphonie fixe à partir de celui fourni par Iyengar (Iyengar, 2004) puis, après une étude des données mises à disposition par Bell, nous adapterons ce modèle à celles-ci. Le présent rapport commencera par une partie exposant rapidement l'état de la science en modélisation de la demande. Nous poserons certaines bases théoriques

et nous parlerons du modèle étudié dans une seconde partie, puis nous aborderons l'expérimentation numérique dans une troisième. Nous commencerons par y présenter les données sur lesquelles nous travaillerons, puis nous parlerons de la manière dont nous avons modifié le modèle ainsi que de certaines considérations pratiques et nous examinerons enfin les résultats numériques. Après avoir évoqué certaines perspectives d'études, nous conclurons dans une dernière partie.

CHAPITRE 1

REVUE DE LITTÉRATURE

Ce chapitre présente l'état de la recherche dans différents domaines en rapport avec notre étude. Nous commencerons par parler de revenue management, puis de tarification, de modélisation de la demande et enfin de programmation à deux niveaux.

1.1 Revenue management

Comme nous l'avons mentionné précédemment, le revenue management, dans son principe, est une pratique quasiment aussi vieille que le commerce lui même. Le début de sa formalisation en temps que pratique à part entière et qu'outil stratégique et tactique peut cependant être associé au système de réservation informatique SABRE qu'American Airline implémenta en 1966 et qui avait la capacité de contrôler l'inventaire des réservations. Comme nous l'explique Chiang (Chiang et al., 2007), ce n'est cependant qu'après l'*Airline Deregulation Act* de 1978 qui mit un terme au contrôle de l'état sur les prix pratiqués dans l'industrie du transport aérien que l'explosion de la concurrence mena à de très rapides innovations dans ce secteur.

On trouve de nombreuses définitions du revenue management dans la littérature, adaptées aux différents contextes dans lesquels il est employé. Il y est souvent fait référence sous le nom de *yield management* en référence à ses premières utilisations dans l'industrie de l'aviation civile qui concernait davantage la gestion des capacités des classes tarifaires. En voici quelques-unes :

- "Revenue Management is the art and science of predicting real-time customer demand at the micromarket level and optimizing the price and availability of products." - Robert G. Cross, Buckhiester.com
- "Systems & procedures to maximize results from the sale of a product or service in more or less fixed supply whose revenue producing ability diminishes with time." - Dr. Warren Lieberman, Buckhiester.com
- "Technique marketing de tarification flexible utilisée dans les services caractérisés par une forte présence de coûts fixes et une certaine inertie des capacités proposées (transport en commun, hôtellerie, ...). Le yield management consiste à maximiser le chiffre d'affaires généré en jouant sur les variables prix et le coefficient d'occupation à l'aide d'une politique de tarification différenciée. Dans ce cadre, les prix les plus bas sont proposés pour des réservations effectuées très en avance ou au contraire à la dernière minute. Les applications de yield management ont connu récemment de nouvelles applications avec le développement d'Internet qui favorise les procédures de gestion tarifaire en temps réel (exemple de la SNCF)." - definitions-marketing.com
- "Yield management is a sophisticated form of supply and demand management that balances both pricing strategies and inventory management. Its primary purpose is managing customer demand through the use of variable pricing and capacity management to maximize profitability. It is essentially the process of allocating the right type of capacity to the right customer at the right time at the right price. It focuses the service organization on maximizing profitability by applying disciplined tactics to forecast consumer behavior at the micro-market level and control inventory availability at each price level at any one time." - Robert G. Cross, Revenue Management - Hard-Core Tactics for Market Domination. New York : Broadway Books, 1997.

D'autre définition peuvent être trouvées sur Internet, par exemple sur le site : http://www.revdev-consultants.com/definition_yield.htm.

On observe depuis quelques années un intérêt accru de la part du milieu académique pour ce domaine de recherche. Ainsi, un certain nombre d'articles et de livres ont récemment permis de brosser un tour d'horizon du revenu management moderne et de fournir une introduction à ses différents concepts. Le livre de Cross (Cross, 1997), sans aborder de questions techniques, traite des questions fondamentales du revenue management. Il décrit l'histoire de cette discipline à travers ses premières applications dans l'industrie du transport aérien, et fournit une approche très qualitative de ses différentes problématiques. Le livre de Talluri et van Ryzin (Talluri et van Ryzin, 2004a), quant à lui, passe en revue sans rentrer dans les détails techniques, les différents outils et concepts rencontrés dans ce domaine et décrit de quelle manière ces différents outils viennent s'insérer dans les diverses pratiques du revenue management en fonction des industries et secteurs d'activités. Une autre excellente étude de cas a été faite par Yeoman et McMahon-Beattie (Yeoman et McMahon-Beattie, 2004).

Certains documents traitent plus particulièrement d'un domaine d'application précis. Leur nombre est assez conséquent, et nous ferons ici uniquement référence à ceux sélectionnés par Talluri et van Ryzin (Talluri et van Ryzin, 2004a) dans leur introduction. Il s'agit de livres et d'articles qui, bien que traitant d'un secteur particulier, restent assez généraux dans les concepts et les outils utilisés. On citera notamment, dans le domaine du transport aérien, le livre de Daudel et Vialle (Daudel et Vialle, 1994) qui traite d'aspects très pratiques du revenue management sans en aborder la théorie, et les articles de Jenkins (Jenkins, 1995) et de Smith et al. (Smith et al., 1992) qui traitent d'un ensemble de règles pratiques dans ce domaine pour le premier, et d'une description des pratiques utilisées par American Airlines pour le second. On citera également le livre de Ingold et al. (Ingold et al., 2000) et l'article de Kimes (Kimes, 1989) qui fournissent des introductions au revenue management à travers des exemples du domaine de l'hôtellerie.

Enfin, un certain nombre d'articles présentent les différentes recherches qui ont été menées dans la littérature scientifique sur certains domaines du revenue management. Weatherford et Bodily (Weatherford et Bodily, 1992) passent en revue une quarantaine d'articles et propose une taxonomie de 14 éléments pour classer les hypothèses faites dans ces articles, de manière à classer ces articles. Kimes (Kimes, 2003) passe en revue plusieurs recherche portant sur le revenue management à la lumière desquels est menée une discussion sur les futures perspectives d'études. McGill et van Ryzin (McGill et van Ryzin, 1999) dépeignent l'histoire de la recherche en revenue management menée jusqu'en 1998. Il comporte en particulier une bibliographie de plus de 190 référence annotées. Il y est particulièrement fait référence à l'industrie du transport aérien et à ses problématiques, telles que l'*overbooking*, le *forecasting*, le contrôle de l'inventaire des sièges et la tarification. Enfin, Chiang et al. (Chiang et al., 2007) brossent un large tour d'horizon des recherches menées depuis 1999 en revenue management à travers une revue de littérature comportant plus de 220 références. L'article traite des différents domaines d'application du revenue management, des différents problèmes rencontrés dans ce domaine et discute des perspectives d'études.

Le revenue management emploie un certain nombre d'outils sous-jacents. La tarification et la modélisation de la demande en font partie. La section qui suit traite de tarification.

1.2 Tarification et gestion des volumes de vente

Comme mentionné plus haut, le revenue management tente de répondre à deux principales questions : "quelle quantité vendre" et "à quel prix vendre". Ces deux décisions sont au centre de l'élaboration de la tactique visant à maximiser les bénéfices. Dans notre étude, nous supposons que nous possédons une quantité illi-

mitée de produits (minutes de communication téléphonique), et que l'attribution d'une certaine quantité de minutes de communication à un utilisateur ne se fait pas au détriment d'un autre. Dans les faits, du point de vue technique, les informations transitent sur un réseau qui est sujet à de la congestion, mais nous n'en tenons pas compte de notre point de vue qui est essentiellement commercial. La question de la quantité ne se pose donc pas pour nous.

Reste la question de la tarification des services. Bon nombre de documents ont été écrits sur ce sujet. Certains livres proposent des introductions aux problèmes de tarification généraux, comme par exemple celui de Nagel et Holden (Nagel, 1994) qui propose un tour d'horizon des questions managériales en rapport avec la tarification, ou celui de Philips (Philips, 2005) qui aborde les techniques basiques de tarification dans le contexte du revenue management. Des articles font aussi une revue des différentes recherches menées dans le domaine académique, comme par exemple celui de Bitran et Caldentey (Bitran et Caldentey, 2003) qui passe en revue 88 articles portants sur la tarification dynamique, c'est à dire la manière d'établir une politique tarifaire en fonction de variables comme la demande, les stocks ou le temps, et ses relations avec le revenue management. Celui de Elmaghraby et Keskinocak (Elmaghraby et Keskinocak, 2003) propose lui aussi une revue de plus de 80 articles portant sur la tarification dynamique, et se concentre particulièrement sur les politiques tarifaires tenants compte de considérations sur l'inventaire.

Dans le domaine de la télécommunication en particulier, la littérature scientifique témoigne d'un intérêt récent très prononcé du milieu industriel pour les recherches de nouvelles techniques de tarification efficaces. En effet, comme expliqué par Luna (Luna, 2002), les opérateurs se sont lancés depuis quelques temps dans une guerre à la tarification. Cette guerre est d'autant plus d'actualité que, comme nous l'expliquent MacKie-Mason et al. (MacKie-Mason et al., 2000), l'importante augmentation récente des capacités informatiques et la forte diminution des coûts des ser-

vices associés ont profondément modifié les stratégies classiques de tarification et de *bundling* (pratique bien connue des opérateurs téléphoniques consistant à rassembler sous une même offre et un même tarif un ensemble de produits ou/et services). Les offres potentielles étant inédites, la réaction des consommateurs n'est pas encore bien définie et nécessite d'être étudiée. Certaines de ces stratégies de bundling ont d'ailleurs été étudiées par (Yikuan et Norsworthy, 1998). L'objet de l'étude est ici le regroupement sous un prix unique d'un certain nombre de minutes de communications locales, inter-étatiques et inter-continentales. L'élasticité croisée des demandes de ces trois services montre que celles-ci sont fortement corrélées et que l'utilisateur pourrait bénéficier de se voir proposer les trois services sous la forme d'une seule offre. D'autres études portant à la base sur des produits différents de ceux offerts par les opérateurs téléphoniques se révèlent très utiles, comme par exemple celui de Allenby, et al. (Allenby et al., 2004) qui explique comment gérer la tarification des offres consistant à diminuer le prix d'un package d'un type de produit en fonction de la taille du package. Bien que l'article s'adresse à la base à des produits tels que des packs de bouteilles, la problématique s'avère être tout à fait la même lorsqu'un opérateur téléphonique fixe le prix d'une gamme de forfaits incluant un nombre plus ou moins grand de minutes de communication "gratuites".

En ce qui concerne la tarification pure, la littérature scientifique foisonne de modèles de tarification de services de télécommunication. Cependant, tous ces modèles abordent la consommation du point de vue technique. Ils considèrent une consommation en terme de bande passante, ou plus généralement de consommation de ressources sur un réseau constitué de nœuds et d'arcs. Ces modèles, faisant pour la plupart appel à la théorie des graphes, traitent donc du coût technique de la communication, du transport de l'information. Il n'y a cependant que très peu d'études traitant de la consommation et de la tarification d'un point de vue plus global, plus marketing. Peu d'études considèrent la communication comme un ser-

vice intégré et traitent de l'offre faite au consommateur final. Yilmaz et Chen (Yilmaz et Chen, 2006), par exemple, se proposent d'étudier la tarification d'un certain nombre d'offre de télécommunication en prenant comme premier facteur la qualité de service sur le réseau. Altmana et al. (Altmana et al., 2006) travail sur la tarification de services différenciés en se basant sur la théorie des jeux pour modéliser les interactions entre utilisateurs et compagnie de télécommunication. Une étude de la tarification du transport de données sur un réseau se servant de la théorie des graphes et de la théorie des jeux pour représenter les jeux de Stackelberg est exposée dans l'article de Bouhtou et al. (Bouhtou et al., 2007). Les rares études qui abordent le sujet de la tarification de l'offre finale sont des études dont le sujet principal est la modélisation de la demande et qui étudient entre autre le comportement de cette demande lors de modifications tarifaires.

Nous allons maintenant aborder un second outil lié au revenue management : la modélisation de la demande.

1.3 Modèle de demande

La modélisation de la demande est une question centrale en revenue management. C'est en effet cette connaissance du comportement des consommateurs qui nous permet de prédire quelle serait la réaction du marché face à un changement de l'offre, et ainsi de prévoir du mieux possible les conséquences de tels changements sur les bénéfices ou l'image de marque de l'entreprise. Notre étude se porte sur un modèle de choix discret. Il s'agit d'outils servant à modéliser le choix fait par un utilisateur parmi plusieurs alternatives. L'ouvrage de Ben-Akiva et Lerman (Ben-Akiva et Lerman, 1985) a fait référence durant de nombreuses années. C'est un livre qui brosse très largement le sujet. Le livre de Train (Train, 2002) est lui aussi très complet et comporte une très grande partie sur la simulation nécessaire à

l'implémentation de modèle. Ces techniques sont maintenant rendues possibles par les avancées de l'informatique.

Des articles peuvent aussi être trouvés, traitant de sujet un peu plus précis. Talluri et van Ryzin (Talluri et van Ryzin, 2004b) tentent de modéliser le comportement des consommateurs lorsqu'ils sont face à une offre comportant une gamme de produits que l'on peut classer par qualité. Ils analysent la tendance que les utilisateurs vont avoir à acheter un produit ou service de qualité supérieure à celui qu'ils auraient initialement acheté, ou au contraire un produit de qualité inférieure, en fonction de l'évolution des prix. Dans le contexte de cette étude, les biens sont périssables, et disponibles en quantité limitée. L'article d'Allenby et al. (Allenby et al., 2004) traite pour sa part de la modélisation de la demande lorsque des articles sont vendus sous forme de packs, et que le prix à l'unité diminue en fonction de la taille du pack. C'est un cas qui peut s'appliquer au domaine de la téléphonie, dans lequel plus le nombre de minutes incluses dans un forfait est important, plus le prix de la minute est faible. Gilbert et al. (Gilbert et al., 2007) proposent de modéliser la demande d'un fournisseur de service en téléphonie fixe. Sans rentrer dans les détails, les utilisateurs sont regroupés selon un certain nombre de critères dans des *lifestyle* de manière à agréger les données et le modèle utilisé est du type mix-logit. Un des problèmes de ce modèle est le fait que la tarification est considérée comme étant proportionnelle au temps consommé, ce qui n'est pas le cas dans les faits. Le modèle présenté dans l'article d'Iyengar (Iyengar, 2004) tente lui aussi de modéliser la demande d'un opérateur téléphonique, mais tient compte de la non linéarité de la tarification. Il distingue de plus le processus de choix d'un forfait, avec une consommation espérée, et le processus de consommation réelle sous ce forfait.

Avant de rentrer dans le vif du sujet, une dernière section parlera d'un type de modèle qui se retrouve en revenue management : la programmation à deux niveaux.

1.4 Programmation bi-niveau

La programmation mathématique à deux niveaux est un outil particulièrement bien adapté à certain problèmes liés au revenue management, qu'il permet bien souvent de formuler de manière très naturelle. Nous n'entrerons pas ici dans les aspects théoriques car ce n'est pas le sujet de cette étude. Qualitativement, un programme bi-niveau consiste en deux programmes mathématiques dont la valeur des fonctions objectif dépend des variables des deux programmes. Un programme peut par exemple maximiser les bénéfices engendrés par une entreprise, et un autre minimiser des coûts de consommation auxquels des consommateurs vont faire face en achetant des produits dans cette entreprise ou chez le concurrent. Ainsi, le bénéfice de l'entreprise dépend des tarifs qu'elle applique (variables auxquelles elle a accès) et de la consommation des usagers qu'elle ne contrôle pas directement. Les coûts de consommation des usagers dépendent de leurs consommations (variables auxquelles ces usagers ont accès) et aux tarifs pratiqués par l'entreprise, auxquels les usagers n'ont pas accès. Les deux programmes (celui de l'entreprise et celui des usagers) sont donc liés. Si l'entreprise désire maximiser ses bénéfices, elle devra tenir compte du comportement des usagers. Ce type de formulation est très adaptée à la représentation de jeux de Stackelberg.

L'article écrit par Colson et al. (Colson et al., 2007) est une introduction à la programmation bi niveau qui, à partir d'un exemple simple, en explique les enjeux et les applications, discute de la théorie de la programmation bi niveau et de différentes méthodes de résolution. Pour plus de détails, on pourra par exemple se référer à l'article de Dempe (Dempe et al., 2002). On peut aussi trouver des exemples concrets de solutions de problèmes de revenue management dans des articles tels que celui de Côté et al. (Côté, 2003).

CHAPITRE 2

CADRE THÉORIQUE

Ce chapitre définit les principaux concepts théoriques et la plupart des notations utilisées dans la suite de cet article. Nous commencerons par expliquer à quoi correspond une tarification non linéaire et ce qu'elle implique du point de vue du budget des usagers, puis nous présenterons le modèle étudié.

2.1 Tarification non linéaire et contraintes budgétaires

Le modèle étudié a été développé par Iyengar (Iyengar, 2004). Il se penche sur les problèmes de tarification non-linéaire rencontrés dans la plupart des offres de services de télécommunication sans fils que nous observons actuellement sur le marché. Il divise de plus le comportement de l'utilisateur en un processus de choix qui consiste à choisir un forfait en début de mois en fonction d'une consommation qu'il prévoit, et un processus de consommation durant lequel il utilise des minutes de communication. Le but est de représenter le fait que la consommation effective de l'utilisateur peut différer des prévisions de consommation qui l'ont amené à choisir un forfait spécifique.

Dans un schéma de tarification non linéaire, le prix n'est pas constitué d'une composante fixe et d'une composante proportionnelle à la consommation. Le prix P ne dépend donc pas d'un coût fixe F , ou coût d'accès dans le domaine de la télécommunication, et d'un coût marginal p tel que $P = F + px$ où x représente la consommation, comme on peut le voir sur le graphique 2.1. Bien que ce soit le terme de tarification non linéaire qui soit utilisé dans la littérature, il serait plus cor-

rect de parler de tarification non affine, car l'équation précédente n'est pas linéaire mais affine puisqu'elle n'est pas complètement proportionnelle à la consommation ($P(x = 0) \neq 0$).

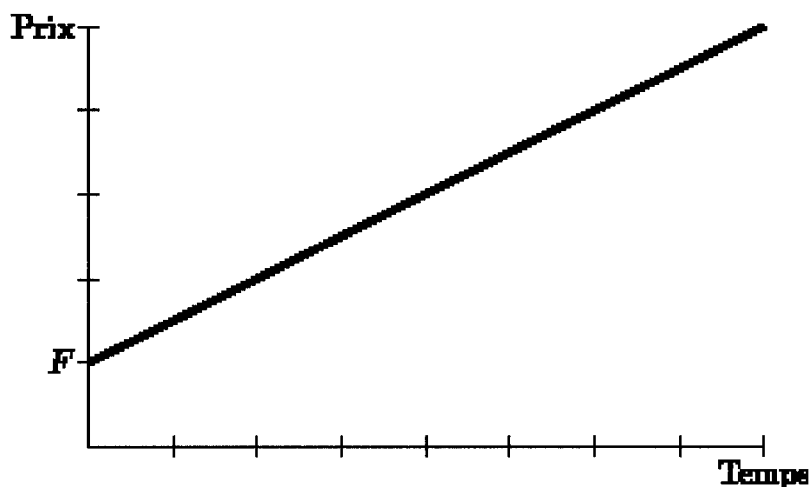


Figure 2.1 Schéma de tarification linéaire

Dans le cas étudié, qui est le cas généralement observé sur le marché de la télécommunication, la tarification est composée d'un coût d'accès F , d'un coût marginal p_1 pour toute consommation entre 0 et un nombre déterminé de minutes A , puis d'un second coût marginal $p_2 > p_1$ pour toute consommation excédant ce nombre de minutes. L'utilisateur paie alors :

$$P = F + p_1x \quad \text{si} \quad 0 < x \leq A$$

et

$$P = F + p_1A + p_2(A - x) \quad \text{si} \quad A < x < B$$

où B représente la quantité maximale matériellement consommable en un mois ($31 \times 24 \times 60$ minutes). Une représentation graphique de ce type de tarification peut être observée sur la figure 2.2

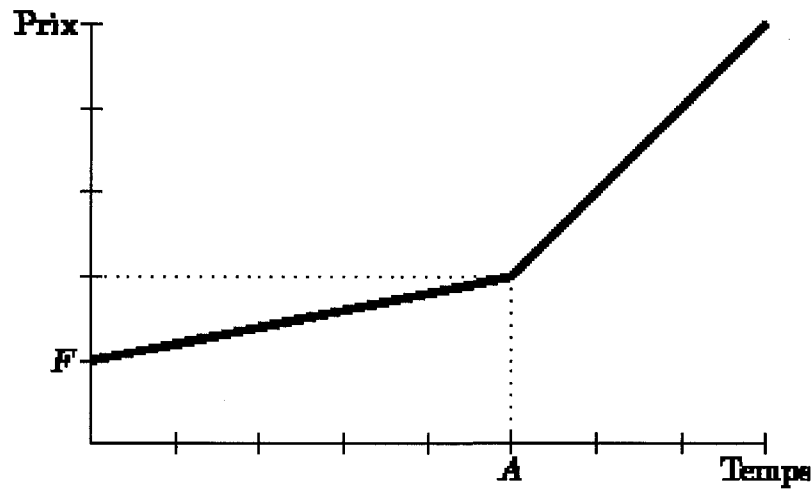


Figure 2.2 Schéma de tarification non linéaire

Ce type de tarification appartient à la catégorie des *increasing block pricing structures* car le coût marginal augmente avec la consommation. Il est appelé *two-tier increasing block pricing structure* car il est composé d'un coût fixe et de deux coûts marginaux. Il n'est pas rare que p_1 soit nul, on dit alors que A minutes sont incluses dans le forfait. Si on considère un usager bénéficiant d'un budget I chaque mois, cet usager attribue ce budget à la consommation de plusieurs choses. Il est donc amené à faire des arbitrages entre ces diverses consommations. Tout argent attribué à la consommation de minutes de télécommunication ne sera pas attribué à la consommation d'autre chose. Ce phénomène est pris en compte dans le modèle étudié en intégrant une commodité alternative x_2 , dont on normalise le coût unitaire à 1. L'usager consomme donc une certaine quantité x_2 et un certain nombre de minutes de télécommunication x_1 , ce qui mène aux contraintes suivantes sur la consommation :

$$p_1 x_1 + x_2 \leq I - F \quad \text{si} \quad 0 < x \leq A$$

et

$$p_2(x_1 - A) + x_2 \leq I - F - p_1 A \quad \text{si} \quad A < x < B$$

Le graphique 2.3 représente ce type de contraintes budgétaires.

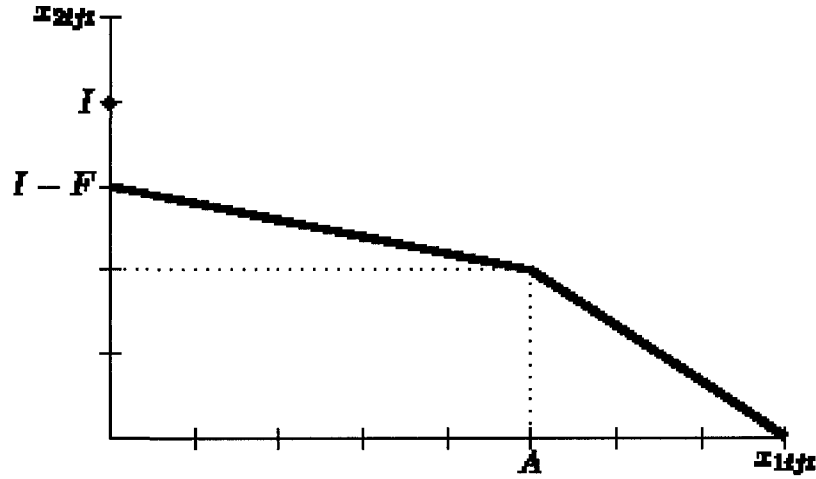


Figure 2.3 Ensemble de contraintes budgétaires

Une des spécificités du modèle décrit dans la section suivant est qu'il tient compte de ce type de tarification et de contraintes budgétaires.

2.2 Présentation du modèle

Disposant d'une gamme de forfaits, le modèle étudié s'attache à définir le forfait choisi par chaque usager au fil des mois et la consommation de ces usagers en terme de minutes. On définit l'utilité associée à un forfait j par un usager i au cours d'un mois t comme suit :

$$U_{ijt}(x_{1ijt}, x_{2ijt}) = \beta_{ij} + \alpha_{i11}x_{1ijt} + \alpha_{i2}x_{2ijt} + \alpha_{i12}x_{1ijt}^2 + (z'_{ijt}\gamma_i)x_{1ijt} + w'_{ijt}\varsigma_i + \epsilon_{ijt} \quad (2.1)$$

où x_{1ijt} représente le nombre de minutes consommées par l'utilisateur i sous le forfait j lors du mois t et x_{2ijt} représente la consommation d'unités de la commodité alternative précédemment évoquée. ϵ_{ijt} représente l'erreur d'estimation de l'utilité due au fait que le modèle ne prend en compte qu'un nombre limité des facteurs pouvant influencer cette utilité aux yeux de l'utilisateur. Cette erreur est un terme aléatoire i.i.d. double exponentiel. z_{ijt} et w_{ijt} sont des vecteurs contenant un certain nombre d'informations qui semblent raisonnablement influencer l'utilisateur dans son choix et dont nous aimerions connaître l'influence. Il peut par exemple s'agir du nombre de minutes consommées en dessous de A_j le mois précédent, ou de la présence ou non de publicité concernant un certain forfait. z_{ijt} contient les informations dont l'influence sur l'utilité du forfait serait proportionnel au nombre de minutes que l'utilisateur aurait l'intention de consommer, et w_{ijt} contient les informations dont l'influence serait indépendante de cette consommation.

Les autres termes sont les paramètres du modèle qui seront calibrés de manière à rendre le modèle le plus proche possible des valeurs effectivement observées. Ces termes sont aléatoires et suivent une distribution normale. Cette définition de l'utilité correspond à un modèle de choix discret de type mixed logit.

Pour chaque forfait, il existe une manière optimale de consommer, c'est-à-dire une valeur de x_{1ijt} et de x_{2ijt} qui maximise l'utilité. C'est l'utilité maximale que l'utilisateur peut retirer de ce forfait. Cette utilité se définit pour chaque utilisateur, chaque forfait et chaque mois en respectant les contraintes budgétaires comme suit :

$$\begin{aligned}
 U_{ijt}^{max} &= \max_{x_{1ijt}} U_{ijt}(x_{1ijt}, x_{2ijt}) \\
 \text{s.c. } & p_{1j}x_{1ijt} + x_{2ijt} = I_i - Fj \quad \text{si } 0 \leq x_{1ijt} \leq A_j \\
 & p_{2j}(x_{1ijt} - A_j) + x_{2ijt} = I_i - Fj - p_{1j}A_j \quad \text{si } A_j < x_{1ijt} \leq B
 \end{aligned}$$

Les contraintes sont convexes, et pour assurer la quasi-concavité de la fonction objectif (l'utilité) on impose la condition suffisante $\alpha_{i12} < 0$, ainsi que la condition $\alpha_{i2} > 0$ qui garantit un effet positif du revenu I sur l'utilité. On garantit ainsi l'existence et l'unicité de l'optimum.

En utilisant la condition nécessaire d'optimalité, on obtient une possibilité d'optimum lorsque la fonction d'utilité est maximisée en tenant compte de la première contrainte, et une autre lorsqu'elle est maximisée en tenant compte de la seconde contrainte :

$$x_{1ijt}^{candopt,1} = \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}$$

$$x_{1ijt}^{candopt,2} = \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}$$

L'existence de l'optimum garantit qu'il se trouve soit entre 0 et A , soit entre A et B , soit avant 0, soit en A , soit après B . Si l'optimum se trouve entre 0 et A , la valeur qui maximise l'utilité avec la première contrainte, $x_{1ijt}^{candopt,1}$, doit se trouver entre 0 et A . Si l'optimum se trouve entre A et B , la valeur qui maximise l'utilité avec la seconde contrainte, $x_{1ijt}^{candopt,2}$, doit se trouver entre A et B . Si $x_{1ijt}^{candopt,1} \leq 0$, l'optimum se trouve en 0. Si $x_{1ijt}^{candopt,2} \geq B$, l'optimum se trouve en B . Si $x_{1ijt}^{candopt,1} \geq A$ et $x_{1ijt}^{candopt,2} \leq A$, l'optimum se trouve en A . L'unicité de la solution nous apporte de plus la garantie que ces cinq cas sont mutuellement exclusifs.

$$U_{itj}^{max} = \begin{cases} U_{ijt}(0, x_{2ijt}) & \text{si } x_{1ijt}^{candopt,1} \leq 0 \\ U_{ijt}(x_{1ijt}^{candopt,1}, x_{2ijt}) & \text{si } 0 < x_{1ijt}^{candopt,1} < A_j \\ U_{ijt}(A_j, x_{2ijt}) & \text{si } x_{1ijt}^{candopt,1} \geq A_j \text{ and } x_{1ijt}^{candopt,2} \leq A_j \\ U_{ijt}(x_{1ijt}^{candopt,2}, x_{2ijt}) & \text{si } A_j < x_{1ijt}^{candopt,2} < B \\ U_{ijt}(B, x_{2ijt}) & \text{si } x_{1ijt}^{candopt,2} \geq B \end{cases}$$

On peut ainsi définir l'utilité maximale d'un forfait en fonction des paramètres du

modèle pour chaque utilisateur et pour chaque mois. Le forfait choisi par chaque utilisateur lors d'un certain mois est le forfait dont l'utilité maximale est la plus grande. Cette utilité intégrant un terme aléatoire ϵ_{ijt} , on accède donc à une probabilité de choix pour chaque forfait, qui est la probabilité que l'utilité maximale de ce forfait soit la plus grande. De part la distribution de ϵ_{ijt} , cette probabilité est pour un plan j est $P(j) = \frac{e^{U_j}}{\sum_i e^{U_i}}$. On connaît aussi le nombre de minutes optimal sous ce forfait.

Une des spécificités de ce modèle est, comme mentionné plus haut, le fait qu'une fois un forfait choisi, la consommation de l'utilisateur n'est pas le nombre de minutes optimal mais dévie de cette quantité par une erreur aléatoire distribuée selon une loi normale :

$$q_{ijt} = q_{ijt}^{opt} + \eta_{ijt}$$

Maintenant que le cadre théorique est posé, nous allons voir de quelle manière exploiter les éléments précédemment cités dans le cas pratique d'un grand opérateur en téléphonie fixe.

CHAPITRE 3

EXPÉRIMENTATIONS NUMÉRIQUES

Cette partie décrit la manière dont le modèle a été mis en œuvre dans le cadre d'une entreprise de téléphonie fixe. Nous analyserons dans une première partie les données qui ont été mises à notre disposition et nous expliquerons les pré-traitements qui ont été nécessaires pour les rendre exploitables. Nous expliquerons dans une seconde partie le travail qui doit être réalisé pour mettre en œuvre le modèle, nous décrirons les outils informatiques dont nous nous servirons et nous effectuerons le calcul de certaine fonction dont nous avons besoin. Nous parlerons dans la partie suivante des différents problèmes pratiques auxquels nous avons eu à faire face ainsi que la manière dont nous avons réglé ces problèmes, puis nous exposerons les résultats numériques dans une dernière partie.

3.1 Les données

Nous avons à notre disposition une importante quantité de données provenant d'un grand fournisseur de service en téléphonie fixe. Nous ferons ici l'analyse de ces données et exposerons un certain nombre de pré-traitements nécessaires à leur exploitation.

3.1.1 Description

Les données utilisées sont celles concernant la clientèle du principal opérateur en téléphonie fixe. Elles regroupent la consommation de ses utilisateurs sur 20 mois.

Elles concernent un total de plus de 100 millions d'entrées, chaque entrée correspondant aux informations relatives à un utilisateur pour un mois donné. Chaque utilisateur peut être suivi à travers les mois. Il est en effet identifié de manière unique par son numéro de téléphone à 10 chiffres auquel on adjoint un identifiant à trois chiffres. L'utilisation de cet identifiant est rendu nécessaire par le fait que le numéro de téléphone peut être réattribué à un autre utilisateur d'un mois à l'autre, ne faisant donc pas référence à lui seul à un utilisateur unique.

Les informations fournies sont :

- La composition de l'offre souscrite (forfaits, options, maintenance)
- La consommation (nombre d'appels, de minutes utilisées, d'accès aux services)
- La tarification (prix payé pour les options de base, pour l'accès au réseau, pour certain type de consommation)

Lors de l'examen de ces données, nous avons été particulièrement surpris du nombre de chiffres significatifs stockés. Les prix et les temps de communication sont en effet fournis avec 9 chiffres après la virgule, là où 2 sembleraient bien suffisants pour ne pas perdre d'information. Cela nous a semblé d'autant plus singulier, que les données sont extrêmement lourdes, tant dans leur stockage (plus d'une centaine de giga octets), que dans leur manipulation. À moins que cette précision n'ait un but qui nous aurait échappé, il semblerait très intéressant, voir nécessaire pour l'exploitation de ces données, de réécrire les archives et de stocker les prochaines données en diminuant la précision.

De plus, les informations fournies par ces données présentent des lacunes relativement importantes qui vont nécessiter par la suite un certain nombre d'hypothèses. La première de ces lacunes est le peu de détail concernant la consommation. En effet, nous disposons du nombre de minutes appelées en heures pleines et en heures creuses, du nombre de minutes appelées en intra-provincial, en inter-provincial, vers les États-Unis et outre mer, mais nous ne disposons pas du nombre de mi-

nutes appelées sur une tranche horaire ET vers une destination. De plus, le détail des destinations n'est pas fourni. La seconde lacune rencontrée concerne la tarification. Nous ne connaissons pas les détails de celle-ci. Il n'est par exemple pas fourni le prix de la minute de communication vers une certaine région. Enfin, même si nous connaissons la composition d'un forfait dans ses grandes lignes, les détails ne sont pas fournis, et il semble que l'offre change légèrement en fonction des utilisateurs et des mois.

Nous allons donc être amené à calculer un certain nombre de données, comme le coût de base moyen des forfaits qui nous intéressent et le prix moyen de la minute de communication vers les grandes catégories de zones citées plus haut.

3.1.2 Prétraitement des données

D'un point de vue pratique, les données sont fournies sous la forme de fichiers textes. Le format utilisé est le format CSV (comma-separated values) dans lequel les différentes valeurs d'une entrée sont séparées par des virgules, et chaque entrée correspond à une ligne. Chaque fichier comporte approximativement 6 millions de lignes, chaque ligne comportant 146 valeurs. Le premier travail a donc consisté à écrire un certain nombre de scripts en langage Perl et Python pour pouvoir accéder aux données pertinentes. Le Python est particulièrement intéressant car il comprend une bibliothèque permettant un accès simple et puissant aux fichiers de type CSV, tant en lecture qu'en écriture.

Un premier script a été écrit permettant de sélectionner des catégories d'utilisateurs selon certains critères (forfaits choisis, consommation, prix payé) et renvoyant pour chaque utilisateur correspondant au critère, pour chaque mois, les informations qui nous intéressent, et uniquement elles (par exemple la consommation vers une

certaine région et le prix payé qui en résulte). Ce script permet en outre soit de sélectionner chaque mois tout utilisateur répondant aux critères spécifiés, soit de sélectionner les utilisateurs répondant à ces critères un certain mois, puis de les suivre au fil des mois. Les résultats sont stockés sous forme de fichier CSV pour être ensuite exploités dans Microsoft Excel.

Il a ensuite été écrit un certain nombre de scripts qui permettent, à partir des fichiers de base ou d'un échantillon fourni par le premier script, de calculer un certain nombre de données statistiques, comme le prix moyen de la minute consommée vers une destination dans le cadre d'un forfait ainsi que l'écart type de la distribution, ou encore le nombre moyen d'utilisateurs ayant un certain comportement.

Enfin, un dernier script permet de créer un fichier texte dont chaque ligne représentait un utilisateur, et comporte une série de chiffres séparés par des virgules, représentant le forfait choisi par cet utilisateur. Ce fichier est le fichier contenant les informations servants au calibrage du modèle.

3.1.3 Description de l'offre

Les forfaits recensés sont au nombre de 25. La plupart d'entre eux portent la mention "grandfathered" pour indiquer qu'ils ne sont actuellement plus offerts, mais peuvent encore être utilisés par certains utilisateurs. Les forfaits constituant l'offre actuelle sont au nombre de 9. Étant intéressés par les comportements de choix du marché, c'est sur ces forfaits que nous nous concentrerons, les autres ne pouvant plus être choisis, mais seulement abandonnés. La documentation fournie ne permet cependant pas de définir clairement toutes les caractéristiques de certains d'entre eux.

Les trois premiers forfaits, référencés BOTALIT, BOTALOT et BOTANYT, sont

des forfaits de tarification par bloc. Pour un prix de base, l'utilisateur peut appeler un certain temps sans coût supplémentaire, puis au-delà du nombre de minutes incluses dans le forfait, l'utilisateur paye un certain prix pour chaque minute appelée. Les deux premiers forfaits incluent 250 minutes et 1200 minutes consommables dans une certaine période de la journée, et le troisième 1200 minutes consommables n'importe quand pour un coût d'accès de base de respectivement \$6, \$12 et \$18. Un des problèmes qui se pose est le fait que l'utilisateur semble choisir une destination vers laquelle cette tarification s'applique, par exemple intra ou inter-provincial, mais aucune information n'est fournie sur ce choix.

Les trois forfaits suivants, identifiés FRATE, LDPXTRA et FRWW, sont assez mal identifiés. On sait seulement que le premier d'entre eux admet une structure de tarification par blocs, et que les deux autres concernent une consommation de minutes outre-mer.

Les trois derniers forfaits, LDP10C, LDP7C, LDP5C, enfin, offrent un prix de communication unique, respectivement de 10, 7 et 5 cents par minutes, pour un certain coût d'accès non fourni.

3.1.4 Analyse statistique

La clientèle de l'opérateur se répartit sur les forfaits précédemment décrits comme décrit dans le tableau 3.1.

Les différentes lignes correspondent au numéro du mois analysé, et la dernière à la moyenne sur les 5 mois. Les colonnes correspondent aux différents forfaits, la ligne "Mutl." correspondant au nombre d'utilisateurs cumulant plusieurs forfaits sur un même mois et la ligne "Tot." au total des utilisateurs. Ce total est bien plus élevé que la somme des utilisateurs des différents forfaits, car la plupart des utilisateurs

Tableau 3.1 Nombre d'usagers par forfait et par mois

Mois	BOTALIT	BOTALOT	BOTANY	FRATE	LDPXTRA
1	10614	38503	159488	1128917	13496
2	11788	40588	177342	1097281	21482
3	13093	42693	196615	1067020	30621
4	14261	43337	211677	1034947	40140
5	15565	44332	227454	1012363	48730
Moy	13064	41891	194515	1068105,6	30894

Mois	FRWW	LPD10C	LPD7C	LPD5C	Mult.	Tot.
1	295587	229880	7263	12536	27817	6735240
2	297747	246701	11395	19658	31201	6703711
3	297774	263589	15635	27590	40879	6652595
4	295415	278873	19715	35082	48591	6587522
5	294393	296210	23757	42296	56760	6544423
Moy	296183	263050,6	15553	27432	41049,6	6644698

ne possèdent pas de forfait, et un certain nombre en possèdent un qui n'est pas référencé ici.

On remarque que certains forfaits sont en croissance, la où d'autres régressent.

À la fin de chaque mois, certains utilisateurs changent de forfait, et d'autres quittent la compagnie, comme indiqué respectivement par les deux tableaux 3.2 et 3.3.

Tableau 3.2 Nombre d'usagers changeant de forfait chaque mois

Mois	BOTALIT	BOTALOT	BOTANY	FRATE	LDPXTRA
1	9542	34483	144391	1093343	1006
2	10680	36336	160357	1062652	1059
3	11711	37323	174561	1029492	1342
4	12774	38754	188607	1001716	1636
Moy.	11176,8	36724	166979	1046800,75	1261

Mois	FRWW	LPD10C	LPD7C	LPD5C	Mult.	Tot.
1	286655	211431	5044	7700	11968	1843668
2	288471	223887	8394	13103	18975	1858242
3	288045	234717	11716	18514	27822	1869962
4	285498	246070	14856	23811	36635	1884189
Moy.	287167,25	229026,25	10003	15782	23850	1864015,25

Tableau 3.3 Nombre d'utilisateurs quittant la compagnie chaque mois

Mois	BOTALIT	BOTALOT	BOTANY	FRATE	LDPXTRA
1	323	1354	4797	18995	632
2	371	1789	6254	22140	1386
3	583	2600	8826	24293	1503
4	607	2111	8103	21329	1903
Moy.	471	1963,5	6995	21689,25	1356

Mois	FRWW	LPD10C	LPD7C	LPD5C	Mult.	Tot.
1	6768	9337	252	470	1038	1865834
2	7975	11871	488	797	1353	1882972
3	8864	14186	682	1070	1846	1892421
4	9232	14448	835	1415	2261	1907414
Moy.	8209,75	12460,5	564,3	938	1624,5	1887160,25

3.1.5 Choix des forfaits

Nous choisissons, pour tester notre modèle, deux forfaits différents parmi ceux présentés ci-dessus. Les trois forfaits du milieu sont à exclure à cause du peu de connaissance que nous possédons sur eux. Les trois premiers forfaits possèdent une structure similaire. Nous choisissons donc le troisième (BOTANYT) qui présente l'avantage de posséder la même structure de tarification quelque soit la plage horaire durant laquelle les appels sont passés. Les trois derniers forfaits possèdent également une structure identique. Nous choisissons le premier (LDP10C) dans la mesure où c'est le plus populaire et que nous possédons donc un échantillon d'utilisateurs plus grand. Nous considérons aussi les utilisateurs n'ayant souscrit à aucun forfait.

3.1.6 Analyse statistique des forfaits choisis

La tarification linéaire de 10 cents/min du forfait LDP10C est respectée pour les communications intra urbaines, inter urbaines et vers les États-Unis. Plus le nombre de minutes consommées est élevé, plus le coût marginal de la minute est en effet proche de 10 cents/min, mais même pour de petites consommations, la tarification

ne s'en éloigne pas trop. Pour le forfait BOTANYT, nous observons en effet que le prix payé pour des consommations inférieures à 1200 minutes est nul, soit en intra urbain, soit en inter urbain, mais pas pour les deux zones. Nous constatons que certains utilisateurs ne respectent pas cette tarification, mais cela reste assez marginal. Un problème rencontré est que très peu d'utilisateurs excédant les 1200 minutes, d'où un certain manque d'informations sur leur comportement.

3.1.7 Choix de l'échantillon

Les utilisateurs choisis pour faire partie de l'échantillon devront répondre aux critères suivants :

- Avoir souscrit à l'un des forfaits étudiés ou à aucun forfait
- N'avoir souscrit qu'à un seul forfait par mois
- Avoir une consommation simple :
 - Consommation intra urbaine, extra urbaine et vers les États-Unis pour LDP10C
 - Consommation intra urbaine, extra urbaine et vers les États-Unis pour les utilisateurs sans forfaits
 - Consommation exclusivement intra urbaine OU exclusivement inter urbaine

La raison du premier critère est assez évidente. Nous n'avons pas choisi de rassembler les utilisateurs n'appartenant à aucune de ces trois catégories dans un groupe "Autre" à cause des différences dans les structures de tarification de ces forfaits. Nous n'aurions pas pu mettre en relation le temps de communication et le prix payé.

Le second critère provient du fait que le modèle choisi nécessite de pouvoir attribuer à chaque minute consommée un prix, ce qui devient compliqué si l'utilisateur possède plusieurs forfaits.

Le troisième critère est rendu nécessaire par la connaissance limitée que nous possédons de la tarification et de la consommation. Pour le forfait LDP10C, l'étude des données nous montre que le prix de 10 cents/min est respecté tant que la consommation s'effectue sur ces trois zones, mais pas en dehors. La structure par blocs du forfait BOTANYT semble s'appliquer à seulement une et une seule des deux zones citées. Les utilisateurs sans forfait devront consommer vers les mêmes zones que ceux des deux autres forfaits pour rendre leur consommation comparable à celle des deux autres catégories d'utilisateurs.

Maintenant que nous possédons un échantillon de données exploitables, nous allons nous pencher dans la prochaine section sur l'implémentation du modèle.

3.2 Adaptation du modèle

Nous allons décrire ici le travail à effectuer pour implémenter le modèle. Nous décrirons la forme du problème mathématique, les outils informatiques que nous avons choisi pour résoudre ce problème et nous effectuerons un certain nombre de développements mathématiques nécessaires à cette résolution.

3.2.1 Choix de l'optimisateur

Le modèle précédemment décrit comporte un certain nombre de paramètres. Ces paramètres doivent être choisis de manière à ce que les probabilités de choix du modèle soient le plus proche possible des probabilités de choix observées dans les données historiques. Pour un utilisateur i , soit $\omega_i = (\beta_{i1}, \dots, \beta_{in}, \alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \gamma_i, \varsigma_i)$, nous définissons pour cela la fonction de vraisemblance $LH_i(\omega_i)$ qui est la probabilité de la série des choix faits sur l'ensemble des périodes par un utilisateur dans

les données historiques, en fonction de la valeur des paramètres choisis. Si notre modèle reproduisait parfaitement les choix faits par cet utilisateur, cette probabilité serait égale à 1. Comme c'est une probabilité et qu'elle ne peut donc pas dépasser cette valeur, pour faire en sorte que notre modèle reproduise du mieux qu'il le peut les choix faits par l'utilisateur, nous devons maximiser la valeur de $LH_i(\omega_i)$ en fonction de ω_i .

Cette procédure d'étalonnage du modèle et d'estimation des paramètres s'écrit donc mathématiquement :

$$\begin{aligned} \max_{\omega_i} \quad & LH_i(\omega_i) \\ \text{s.c.} \quad & \omega_i \in S \end{aligned}$$

où S est l'ensemble admissible de ω_i . Dans les faits, nous maximiserons le logarithme de $LH_i(\omega_i)$, car celui-ci se manipule plus simplement au cours de certains calculs.

Pour réaliser cette maximisation, nous utilisons IpOpt, une bibliothèque d'optimisation selon une méthode de point intérieur développée en C++. IpOpt traite les problèmes du type :

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.c.} \quad & a \leq h(x) \leq b \\ & c \leq x \leq d \end{aligned}$$

Ici, la fonction objectif f est la vraisemblance, le vecteur x est le vecteur des paramètres du modèle (les α_{i11} , α_{i12} , α_{i2} et β_{ij}). Il n'y a à priori pas de contraintes.

3.2.2 Utilisation de l'optimisateur

Cette bibliothèque communique avec nous à travers une classe héritée dont nous spécifions les membres de manière à créer notre modèle. Le travail à effectuer pour utiliser cette classe est, dans les grandes lignes, le suivant.

IpOpt fait appel à la fonction *get_nlp_info* pour spécifier la taille du problème, c'est-à-dire le nombre de variables et le nombre de contraintes du problème, ainsi qu'éventuellement la densité des matrices hessienne et jacobienne. Pour chaque utilisateur, il y a une variable β par forfait et trois variables α . Si nous considérons deux forfaits, il y en a donc 5 fois le nombre d'utilisateurs. Il n'y a pas de contraintes, et nous ne fournissons aucune information sur la densité des matrices.

IpOpt fait appel à la fonction *get_bounds_info* pour spécifier les bornes sur les variables et les contraintes. En accord avec notre modèle, les α_{i12} sont négatifs et les α_{i2} sont positifs.

IpOpt fait appel à la fonction *get_starting_point* pour fournir un point de départ pour la résolution du problème. Nous donnons dans un premier temps comme valeur initiales à nos variables les valeurs suivantes de manière à les rapprocher de 0 tout en respectant le domaine admissible :

$$\begin{aligned}\beta_{ij} &= 0 \\ \alpha_{11} &= 0 \\ \alpha_{12} &= -0.0005 \\ \alpha_2 &= 0.0005\end{aligned}$$

IpOpt fait appel à la fonction *eval_f* pour fournir la valeur de la fonction objectif pour une certaine valeur des variables. IpOpt nous fournit une liste brute de variables. Il s'agit alors de stocker ces variables dans une structure permettant aux fonctions calculant le likelihood d'accéder aux variables en fonction de leur signification, par exemple à la variable α_{i11} du 3^{ème} utilisateur, ou à la variable β_{ij} du 7^{ème} utilisateur pour le plan 2. Une fois cela fait, nous appelons la fonction de calcul du log-vraisemblance.

IpOpt fait appel à la fonction *eval_grad_f* pour fournir la valeur du gradient de la fonction objectif pour une certaine valeur des variables. De même que précédemment, il faut dans un premier temps stocker les variables dans la structure adéquat, puis appeler la fonction calculant la dérivée du likelihood en fonction de chaque variable.

Le programme principal crée une instance de notre classe, définit certains paramètres relatifs à la résolution, puis appelle les fonctions *Initialize* et *OptimizeTNLP* pour respectivement initialiser et lancer la résolution du problème. Parmi ces paramètres, nous pouvons citer les suivants.

- *mu_strategy* permet d'établir le type de stratégie d'adaptation du paramètre de barrière. Il est ici mis sur *adaptive* pour utiliser une stratégie adaptative, par opposition à la stratégie par défaut qui est la stratégie monotone (Fiacco-McCormick).
- *tol* permet de régler la tolérance d'arrêt de l'optimisation. Nous la mettons ici sur $1e - 5$.
- *hessian_approximation* permet d'établir quelles valeurs de la matrice Hessienne du Lagrangien utiliser lors de l'optimisation. Réglée par défaut sur *exact* pour spécifier d'utiliser des valeurs fournies par le programme, nous la mettons sur *limited-memory* pour demander au programme d'effectuer une approximation de type quasi-Newton à mémoire limitée.
- *check_derivative_for_naninf*, établie à *yes* demande de vérifier la valeur des dérivées transmises à l'optimisateur. Si celles-ci sont indéfinies (NaN) ou infinies (INF), l'optimisateur le signale et n'en tient pas compte. Cette option sera particulièrement utile pour régler certains problèmes de *scaling* dont nous parlerons en 4.2.7.
- *derivative_test*, établie à *first-order*, demande à l'optimisateur d'effectuer avant le démarrage de l'optimisation un test de la valeur des dérivées du premier ordre

fournies par le programme. Les dérivées sont testées au point de départ fourni par le programme, et celle ayant une valeur suspecte sont signalées. Cette option se révèle très utile en phase de débogage du programme.

3.2.3 Calcul du log-vraisemblance et de ses dérivées

Nous allons effectuer dans cette section les développements nécessaires à l'obtention des deux principales informations requises par IpOpt (et plus généralement par n'importe quel processus d'optimisation). Les étapes suivies dans cette section sont aussi celles suivies numériquement, chacune d'elles correspondant à une fonction faisant appel aux précédentes.

Nous allons commencer par calculer la forme explicite de la fonction d'utilité. Comme nous l'avons vu lorsque nous avons présenté le modèle, l'expression de l'utilité dépend des valeurs de x_{1ijt} et de x_{2ijt} , qui dépendent elles-mêmes des valeurs de $x_{1ijt}^{candopt,1}$ et $x_{1ijt}^{candopt,2}$. L'utilité s'exprime :

$$U_{ijt}(x_{1ijt}, x_{2ijt}) = \beta_{ij} + \alpha_{i11}x_{1ijt} + \alpha_{i2}x_{2ijt} + \alpha_{i12}x_{1ijt}^2 + (z'_{ijt}\gamma_i)x_{1ijt} + w'_{ijt}\varsigma_i + \epsilon_{ijt}$$

Les conditions d'optimalité du premier ordre nous fournissent deux optima potentiels :

$$x_{1ijt}^{candopt,1} = \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}$$

$$x_{1ijt}^{candopt,2} = \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}$$

En prenant en compte l'ensemble de contraintes budgétaires, cinq cas mutuellement exclusifs sont possibles, comme expliqué dans la description du modèle, en fonction des cinq régions du domaine d'étude : les points extrêmes 0, A_j et B et les points appartenant à $]0, A_j[$ et à $]A_j, B[$.

1. Si la valeur du premier optimum potentiel est comme suit :

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq 0,$$

alors, de part les contraintes, les valeurs optimales de x_{1ijt} et x_{2ijt} sont :

$$x_{1ijt} = 0 \quad \text{et} \quad x_{2ijt} = I_i - F_j.$$

L'utilité exprimée uniquement en fonction des paramètres est donc :

$$U_{ijt}(x_{1ijt}, x_{2ijt}) = \beta_{ij} + \alpha_{i2}(I_i - F_j) + w'_{ijt}\varsigma_i + \epsilon_{ijt}.$$

2. De même, si

$$0 < \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < A_j,$$

alors

$$x_{1ijt} = \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \quad \text{et} \quad x_{2ijt} = I_i - F_j - p_{1j}x_{1ijt}.$$

Donc

$$\begin{aligned} U_{ijt}(x_{1ijt}, x_{2ijt}) &= \beta_{ij} + x_{1ijt}(\alpha_{i11} - \alpha_{i2}p_{1j} + z'_{ijt}\gamma_i) + \alpha_{i2}(I_i - F_j) \\ &\quad + \alpha_{i12}x_{1ijt}^2 + w'_{ijt}\varsigma_i + \epsilon_{ijt}, \end{aligned}$$

$$U_{ijt}(x_{1ijt}, x_{2ijt}) = \beta_{ij} - \frac{(\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i)^2}{4\alpha_{i12}} + \alpha_{i2}(I_i - F_j) + w'_{ijt}\varsigma_i + \epsilon_{ijt}.$$

3. Si

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq A_j,$$

et

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq A_j,$$

alors

$$x_{1ijt} = A_j \quad \text{et} \quad x_{2ijt} = I_i - F_j - p_{1j}A_j.$$

Donc

$$\begin{aligned} U_{ijt}(x_{1ijt}, x_{2ijt}) &= \beta_{ij} + \alpha_{i11}A_j + \alpha_{i2}(I_i - F_j - p_{1j}A_j) + \alpha_{i12}A_j^2 \\ &\quad + (z'_{ijt}\gamma_i)A_j + w'_{ijt}\varsigma_i + \epsilon_{ijt}. \end{aligned}$$

4. Si

$$A_j < \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < B,$$

alors

$$x_{1ijt} = \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \quad \text{et} \quad x_{2ijt} = I_i - F_j - p_{1j}A_j - p_{2j}x_{1ijt} + p_{2j}A_j.$$

Donc

$$\begin{aligned} U_{ijt}(x_{1ijt}, x_{2ijt}) &= \beta_{ij} + x_{1ijt}(\alpha_{i11} - \alpha_{i2}p_{2j} + z'_{ijt}\gamma_i) \\ &\quad + \alpha_{i2}(I_i - F_j + A_j(p_{2j} - p_{1j})) + \alpha_{i12}x_{1ijt}^2 + w'_{ijt}\varsigma_i + \epsilon_{ijt}, \end{aligned}$$

$$\begin{aligned} U_{ijt}(x_{1ijt}, x_{2ijt}) &= \beta_{ij} - \frac{(\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i)^2}{4\alpha_{i12}} \\ &\quad + \alpha_{i2}(I_i - F_j + A_j(p_{2j} - p_{1j})) + w'_{ijt}\varsigma_i + \epsilon_{ijt}. \end{aligned}$$

5. Si

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq B,$$

alors

$$x_{1ijt} = B \quad \text{et} \quad x_{2ijt} = I_i - F_j - p_{2j}B + (p_{2j} - p_{1j})A_j.$$

Donc

$$\begin{aligned} U_{ijt}(x_{1ijt}, x_{2ijt}) &= \beta_{ij} + \alpha_{i11}B + \alpha_{i2}(I_i - F_j - p_{2j}B + (p_{2j} - p_{1j})A_j) \\ &\quad + \alpha_{i12}B^2 + (z'_{ijt}\gamma_i)B + w'_{ijt}\varsigma_i + \epsilon_{ijt}. \end{aligned}$$

De l'expression de l'utilité, nous dérivons alors la probabilité Logit de choix d'un forfait j :

$$P(j) = \frac{e^{U_j}}{\sum_{j'} e^{U_{j'}}}.$$

La vraisemblance par période t et utilisateur i est alors :

$$LH_{i,t} = P(j)$$

avec j le choix fait par l'utilisateur i à la période t dans l'historique, et le log-vraisemblance par période t et utilisateur i est :

$$LLH_{i,t} = \ln(LH_{i,t}).$$

Nous pouvons enfin exprimer le log-vraisemblance du modèle :

$$LLH_i = \sum_t LLH_{i,t}.$$

Nous allons maintenant passer au calcul de la dérivée de l'utilité par rapport aux

paramètres du modèle. De même que pour la valeur de l'utilité, les expressions des dérivées de l'utilité dépendent des valeurs de $x_{1ijt}^{candopt,1}$ et $x_{1ijt}^{candopt,2}$.

1. Si

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq 0,$$

alors

$$\begin{aligned} \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \beta_{ij'}} &= 0 \quad \forall j' \neq j \\ &= 1 \quad \text{si } j' = j, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i11}} &= 0, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i12}} &= 0, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i2}} &= I_i - F_j. \end{aligned}$$

2. si

$$0 < \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < A_j,$$

alors

$$\begin{aligned} \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \beta_{ij}} &= 1 \quad \forall j, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i11}} &= \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i12}} &= \frac{(\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i)^2}{4\alpha_{i12}^2}, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i2}} &= -p_{1j} \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} + I_i - F_j. \end{aligned}$$

3. Si

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq A_j,$$

et

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq A_j,$$

alors

$$\begin{aligned} \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \beta_{ij'}} &= 0 \quad \forall j' \neq j \\ &= 1 \quad \text{if } j' = j, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i11}} &= A_j, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i12}} &= A_j^2, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i2}} &= I_i - F_j - p_{1j}A_j. \end{aligned}$$

4. Si

$$A_j < \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < B,$$

alors

$$\begin{aligned} \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \beta_{ij}} &= 1 \quad \forall j, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i11}} &= \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}}, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i12}} &= \frac{(\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i)^2}{4\alpha_{i12}^2}, \\ \frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i2}} &= -p_{2j} \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \\ &\quad + I_i - F_j + A_j(p_{2j} - p_{1j}). \end{aligned}$$

5. Si

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq B,$$

alors

$$\begin{aligned}
\frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \beta_{ij'}} &= 0 \quad \forall j' \neq j \\
&= 1 \quad \text{if } j' = j, \\
\frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i11}} &= B, \\
\frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i12}} &= B^2, \\
\frac{\partial U_{ijt}(x_{1ijt}, x_{2ijt})}{\partial \alpha_{i2}} &= I_i - F_j - p_{2j}B + (p_{2j} - p_{1j})A_j.
\end{aligned}$$

Nous allons maintenant calculer la dérivée de la probabilité de choix d'un forfait j . Soit a_i une variable quelconque intervenant dans l'expression d'une des utilités de l'utilisateur i . Les différentes dérivées par rapport à z_{ijt} et à w_{ijt} ne seront pas exposées ici car, comme nous le verrons dans la prochaine partie, ces termes ne seront pas pris en compte dans les applications pratiques de notre modèle.

$$\begin{aligned}
\frac{\partial P(j)}{\partial a_i} &= \frac{\frac{\partial U_j}{\partial a_i} e^{U_j}}{\sum_{j'} e^{U_{j'}}} - \frac{e^{U_j} \sum_{j'} \frac{\partial U_{j'}}{\partial a_i} e^{U_{j'}}}{\left(\sum_{j'} e^{U_{j'}} \right)^2} \\
&= \frac{e^{U_j}}{\sum_{j'} e^{U_{j'}}} \left[\frac{\partial U_j}{\partial a_i} - \sum_{j'} \frac{\partial U_{j'}}{\partial a_i} \frac{e^{U_{j'}}}{\sum_{j''} e^{U_{j''}}} \right] \\
&= P(j) \left[\frac{\partial U_j}{\partial a_i} - \sum_{j'} \frac{\partial U_{j'}}{\partial a_i} P(j') \right]
\end{aligned}$$

La dérivée de la vraisemblance par période t et utilisateur i s'exprime :

$$\frac{\partial LH_{i,t}}{\partial a_i} = \frac{\partial P(j)}{\partial a_i},$$

avec j choix fait par l'utilisateur i à la période t dans l'historique, et la dérivée du

log-vraisemblance par période t et utilisateur i est :

$$\frac{\partial LLH_{i,t}}{\partial a_i} = \frac{\partial \ln(LH_{i,t})}{\partial a_i} = \frac{\partial LH_{i,t}}{\partial a_i} \frac{1}{LH_{i,t}}.$$

Enfin la dérivée du log-vraisemblance du modèle :

$$\frac{\partial LLH_i}{\partial a_i} = \sum_t \frac{\partial LLH_{i,t}}{\partial a_i}.$$

Nous allons maintenant calculer les dérivées secondes des différentes fonctions. Nous commençons par la matrice hessienne de l'utilité. Encore une fois, il existe 5 cas mutuellement exclusifs, en fonction des valeurs de $x_{ij t}^{candopt,1}$ et $x_{ij t}^{candopt,2}$.

1. Si

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq 0,$$

alors

$$\begin{matrix} & \dots & \beta_i & \dots & \alpha_{11} & \alpha_{12} & \alpha_2 \\ \dots & \left(\begin{array}{cccccc} \dots & \dots & \dots & \dots & \dots & \dots \\ \beta_i & \dots & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_{11} & \dots & 0 & \dots & 0 & 0 & 0 \\ \alpha_{12} & \dots & 0 & \dots & 0 & 0 & 0 \\ \alpha_2 & \dots & 0 & \dots & 0 & 0 & 0 \end{array} \right) \end{matrix}$$

2. Si

$$0 < \frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < A_j,$$

alors

$$\begin{array}{c}
\begin{array}{ccccc}
& \dots & \beta_{ij} & \dots & \\
& & & & \alpha_{i11} & & \alpha_{i12} & & \alpha_{i2} \\
\dots & \left(\begin{array}{ccc}
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots
\end{array} \right. & & & & & & & \left. \begin{array}{ccc}
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots
\end{array} \right)
\end{array}
\end{array}$$

3. Si

$$\frac{\alpha_{i2}p_{1j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq A_j,$$

et

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \leq A_j,$$

alors

$$\begin{array}{c}
\begin{array}{ccccc}
& \dots & \beta_i & \dots & \alpha_{11} & \alpha_{12} & \alpha_2 \\
\dots & \left(\begin{array}{ccc}
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots
\end{array} \right. & & & & & & & \left. \begin{array}{ccc}
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & \dots & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots \\
\dots & 0 & \dots
\end{array} \right)
\end{array}
\end{array}$$

4. Si

$$A_j < \frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} < B,$$

alors

$$\begin{array}{ccccccc}
& \dots & \beta_{ij} & \dots & \alpha_{i11} & \alpha_{i12} & \alpha_{i2} \\
\begin{array}{l} \dots \\ \beta_i \\ \dots \\ \beta_{ij} \\ \alpha_{i12} \\ \alpha_{i2} \end{array} & \left(\begin{array}{ccccccc}
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & 0 & \dots & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & 0 & \dots & \frac{-1}{2\alpha_{i12}} & \frac{\alpha_{i11}-\alpha_{i2}p_{2j}+z'_{ijt}\gamma_i}{2\alpha_{i12}^2} & \frac{p_{2j}}{2\alpha_{i12}} & \dots \\
\dots & 0 & \dots & \frac{\alpha_{i11}-\alpha_{i2}p_{2j}+z'_{ijt}\gamma_i}{2\alpha_{i12}^2} & -\frac{(\alpha_{i2}p_{2j}-\alpha_{i11}-z'_{ijt}\gamma_i)^2}{2\alpha_{i12}^2} & p_{2j}\frac{\alpha_{i2}p_{2j}-\alpha_{i11}-z'_{ijt}\gamma_i}{2\alpha_{i12}^2} & \dots \\
\dots & 0 & \dots & \frac{p_{2j}}{2\alpha_{i12}} & p_{2j}\frac{\alpha_{i2}p_{2j}-\alpha_{i11}-z'_{ijt}\gamma_i}{2\alpha_{i12}^2} & \frac{-p_{2j}^2}{2\alpha_{i12}^2} & \dots
\end{array} \right)
\end{array}$$

5. Si

$$\frac{\alpha_{i2}p_{2j} - \alpha_{i11} - z'_{ijt}\gamma_i}{2\alpha_{i12}} \geq B,$$

alors

$$\begin{array}{ccccccc}
& \dots & \beta_i & \dots & \alpha_{11} & \alpha_{12} & \alpha_2 \\
\begin{array}{l} \dots \\ \beta_i \\ \dots \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_2 \end{array} & \left(\begin{array}{ccccccc}
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & 0 & \dots & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & 0 & \dots & 0 & 0 & 0 & 0 \\
\dots & 0 & \dots & 0 & 0 & 0 & 0 \\
\dots & 0 & \dots & 0 & 0 & 0 & 0
\end{array} \right)
\end{array}$$

Les dérivées secondes de la probabilité de choix d'un forfait j sont alors :

$$\begin{aligned}
\frac{\partial^2 P(j)}{\partial a_1 \partial a_2} &= P(j) \left[\left[\frac{\partial U_j}{\partial a_1} - \sum_{j'} \frac{\partial U_{j'}}{\partial a_1} P(j') \right] \left[\frac{\partial U_j}{\partial a_2} - \sum_{j'} \frac{\partial U_{j'}}{\partial a_2} P(j') \right] \right. \\
&\quad \left. + \frac{\partial^2 U_j}{\partial a_1 \partial a_2} - \sum_{j'} P(j') \left(\frac{\partial^2 U_j}{\partial a_1 \partial a_2} + \left[\frac{\partial U_{j'}}{\partial a_2} - \sum_{j''} \frac{\partial U_{j''}}{\partial a_2} P(j'') \right] \right) \right].
\end{aligned}$$

Les dérivées secondes de la vraisemblance par période t et utilisateur i sont :

$$\frac{\partial^2 LH_{i,t}}{\partial a_1 \partial a_2} = \frac{\partial^2 P(j)}{\partial a_1 \partial a_2},$$

avec j choix fait par l'utilisateur i à la période t dans l'historique, et les dérivées secondes du log-vraisemblance par période t et utilisateur i :

$$\frac{\partial^2 LLH_{i,t}}{\partial a_1 \partial a_2} = \frac{\partial^2 LH_{i,t}}{\partial a_1 \partial a_2} \frac{1}{LH_{i,t}} - \frac{\partial LH_{i,t}}{\partial a_1} \frac{\partial LH_{i,t}}{\partial a_2} \frac{1}{LH_{i,t}^2}.$$

Enfin, les dérivées secondes du Log-Likelihood du modèle sont :

$$\frac{\partial^2 LLH_i}{\partial a_1 \partial a_2} = \sum_t \frac{\partial^2 LLH_{i,t}}{\partial a_1 \partial a_2}.$$

Nous avons ici décrit l'implémentation du modèle tel qu'elle devrait être faite en théorie. Cependant, la pratique nous oblige à tenir compte de certains phénomènes dont nous allons parler dans la prochaine section.

3.3 Considérations pratiques

Durant l'implémentation de la procédure d'étalonnage du modèle, nous avons rencontré un certain nombre de problèmes pratiques. Nous les exposons ici, ainsi que les solutions que nous y avons apportées.

3.3.1 Considérations sur les dérivées secondes

Les expressions des dérivées secondes deviennent assez complexes. Le calcul de chacune d'elle nécessite le calcul d'un certain nombre de fonctions sous-jacentes, ce qui rend le calcul assez coûteux en ressources. De plus, de part la relative complexité des expressions, nous courons le risque d'introduire en les implémentant un certain nombre d'erreurs qui seraient difficiles à détecter. Nous choisissons donc, comme

annoncé précédemment, d'avoir recours à une approximation de type quasi-Newton à mémoire limitée de la matrice hessienne de la fonction objectif.

Le calcul de la matrice hessienne de l'utilité qui reste, elle, très simple, nous permet d'avoir une certaine intuition du comportement de cette utilité, et nous permettra dans les sections suivantes de vérifier que l'aspect de cette fonction d'utilité est cohérente, et par la même occasion qu'elle a bien été implémentée.

3.3.2 Problèmes relatifs aux données

Les données que nous possédons posent un certain nombre de problèmes.

Tout d'abord, il peut sembler que l'absence d'information quand au revenu I_i des utilisateurs peut être problématique. En réalité, en examinant les expressions de l'utilité présentées dans la section précédente, nous nous rendons compte que ce revenu n'intervient qu'à travers le terme $\alpha_2 I_i$ qui s'ajoute à l'utilité, terme indépendant du forfait j . Soit $U_j^i = U_j - \alpha_2 I_i$, l'influence sur la probabilité Logit est la suivante :

$$\begin{aligned} \frac{e^{U_j^i}}{\sum_i e^{U_j^i}} &= \frac{e^{U_j - \alpha_2 I_i}}{\sum_i e^{U_i - \alpha_2 I_i}} \\ &= \frac{e^{U_j} e^{-\alpha_2 I_i}}{\sum_i e^{U_i} e^{-\alpha_2 I_i}} \\ &= \frac{e^{U_j}}{\sum_i e^{U_i}} \\ &= P(j) \end{aligned}$$

La suppression du terme de revenu dans tous les calculs n'a aucune influence sur la valeur de la probabilité Logit, et donc sur le log-likelihood. Ce terme peut donc

être remplacé par 0 ou par toute autre constante, le modèle sera étalonné de la même manière, et conduira aux mêmes probabilités.

Un autre problème, beaucoup plus embêtant, concerne le fait qu'un nombre infime d'utilisateurs, voir nul lors de certaines périodes, excèdent le nombre de minutes offert par les forfaits. Ainsi, nous ne sommes pas en mesure de calibrer γ_i et ς_i , les paramètres relatifs à z_{ijt} et à w_{ijt} . Faute d'informations pertinentes à intégrer à ces termes, nous décidons de les supprimer dans les applications pratiques du modèle.

3.3.3 Étude de la fonction objectif

Le comportement de la fonction objectif en fonction des différents paramètres n'est pas particulièrement intuitif. Quelques tentatives d'optimisation nous révèlent que la probabilité de choix d'un forfait peut facilement tendre vers 0, faisant tendre le log-vraisemblance vers $-\infty$, ce qui a pour conséquence d'avoir un système très mal conditionné. Nous avons donc besoin de posséder une connaissance plus fine de cette fonction objectif de manière à déterminer un domaine d'étude approprié pour les paramètres et, si nécessaire, un pré-traitement permettant d'obtenir un système mieux conditionné. Nous nous intéresserons uniquement à l'influence des paramètres α_{11} , α_{12} et α_2 , les paramètres β_j ne faisant que contribuer linéairement à l'utilité.

Nous procédons pour cela à l'ajout d'une fonction à notre code en C++. Cette fonction, après initialisation du problème, considère un ensemble de points uniformément répartis sur un sous-ensemble borné de l'ensemble admissible. Elle génère ensuite un script Python qui, à l'aide des bibliothèques NumPy, Math-PlotLib et PyLab, affiche l'évolution de l'utilité en fonction de chaque paramètre (A.1, A.2 et A.3) et les lignes de niveau de la fonction objectif en fonction de α_{11}

et α_2 pour des valeurs fixées de α_{12} (A.4, A.5, A.6 et A.7).

L'examen des expressions de la fonction d'utilité et de ses dérivées nous révèle que ces fonctions, définies à la base sur 5 domaines distincts, peuvent être prolongées par continuité aux extrémités de ces domaines, faisant de l'utilité une fonction de classe $C^1 \forall (\alpha_{11}, \alpha_{12}, \alpha_2) \in (]-\infty, +\infty[,]-\infty, 0[,]0, +\infty[)$.

Comme nous pouvons le constater sur les graphiques A.1, A.2 et A.3 en annexe, et conformément aux expressions de la fonction d'utilité et de ses différentes dérivées, la fonction d'utilité est concave par rapport à α_{11} , α_{12} et α_2 , composée de trois zones linéaires par rapport à ces deux variables, reliées entre elles par des arcs de paraboles.

La continuité et la dérivabilité en tout point du domaine d'étude garantit les mêmes propriétés pour le log-vraisemblance, ce qui est très appréciable du point de vu de l'optimisation. Cependant, rien ne garantit à priori la concavité du log-vraisemblance.

3.3.4 Problèmes numériques

À la lumière de ce que nous avons appris de la fonction d'utilité, nous pouvons identifier certains problèmes numériques qui doivent être résolus.

Tout d'abord, examinons le comportement du modèle lorsqu'un utilisateur ne souscrit qu'à un seul et même forfait durant toute la période d'étude. La probabilité de choix du forfait souscrit est 1, là où la probabilité de choix des autres forfaits est nulle. Pour tenter de reproduire ces probabilités de choix, la maximisation du log-vraisemblance va faire tendre l'exponentielle de l'utilité des forfaits qui ne sont pas choisis vers 0 en minimisant l'utilité de ces forfaits. Si l'exponentielle pou-

vait être nulle, ceci suffirait à obtenir les probabilités observées, mais l'exponentiel étant toujours strictement positive, l'optimisateur va également tenter de maximiser l'exponentielle du forfait choisi en maximisant son utilité. Cependant, comme nous l'avons vu précédemment, la fonction d'utilité n'est ni majorée, ni minorée. Le problème n'est donc pas borné. Nous allons résoudre ce problème de manière assez simple, en bornant les différentes variables. Les bornes seront établies de manière à ce que d'une part le domaine d'étude comporte chacun des différents domaines mentionnés précédemment, et d'autre part en faisant varier chaque variable indépendamment, il soit possible d'obtenir des valeurs de l'utilité suffisamment grandes ou suffisamment petites pour que l'exponentielle d'une utilité puisse être négligeable par rapport à celui d'une autre. Les bornes, choisies de manière empirique, sont les suivantes :

$$\begin{aligned} -10 &\leq \beta_{ij} \leq 10 \\ -100 &\leq \alpha_{11} \leq 400 \\ -0.3005 &\leq \alpha_{12} \leq -0.0005 \\ 0.0005 &\leq \alpha_2 \leq 30 \end{aligned}$$

Le problème se pose de manière similaire dès lors que certains forfaits ne sont jamais choisis par un utilisateur, et que la probabilité de choix de ce forfait est donc nulle.

Un autre problème rencontré, provient de l'ordre de grandeur des utilités. Comme nous pouvons le voir sur les graphiques (A.4, A.5, A.6, A.7, A.1, A.2 et A.3) en annexe, l'utilité est comprise sur une plage de valeurs assez étendue, et peut facilement prendre des valeurs de l'ordre de $1e10$. Or, IpOpt travaille sur des valeurs comprises entre $1e62$ et $1e-62$. L'exponentiel de la probabilité Logit atteint un ordre de grandeur de $1e62$ lorsque l'utilité atteint approximativement la valeur de

142. L'utilité ne peut donc prendre que des valeurs d'une grandeur très limitée. Pour remédier à ce problème, avant chaque calcul d'une probabilité Logit, nous prenons la plus grande utilité, et nous la retranchons à toutes les utilités. Ainsi, les utilités sont toutes négatives ou nulles. Une telle opération ne change pas la probabilité de choix. En effet, comme nous l'avons vu précédemment, l'ajout d'une constante quelconque à toutes les utilités ne change pas les probabilités Logit. On pourrait envisager d'ajouter une certaine valeur comprise entre 0 et 142 aux utilités avant d'y retrancher la valeur de la plus grande utilité, de manière à éviter que l'exponentiel des utilités les plus faible tende vers 0.

Enfin, le dernier problème rencontré provient du fait que, certains forfaits n'étant jamais choisis par certains utilisateurs, leur probabilité de choix est nulle. Lors du calcul du log-likelihood associé à cette alternative, le logarithme n'étant pas défini en 0, l'optimisation n'aboutit pas. Nous réglons ce problème en demandant à la fonction de calcul des probabilités Logit de ne jamais renvoyer de probabilité nulle. Si une telle probabilité est calculée, elle est remplacée par une probabilité très faible ($1e-15$). Le logarithme reste ainsi défini, et l'alternative sera tellement peu souvent choisie que, du point de vue probabiliste, la probabilité garde la même signification.

Une fois ces problèmes résolus, l'optimisation peut être effectuée. Nous allons discuter des résultats numériques dans la prochaine section.

3.4 Résultats numériques

Nous effectuons deux types d'études. Nous étudions dans un premier temps le comportement d'utilisateurs indépendants, puis celui d'un ensemble d'utilisateurs.

La première étude a été menée sur trois personnes. La première a choisi le forfait BOTANY sur toutes les périodes, la seconde a choisi le forfait LDP10C sur toutes les

périodes, et la troisième a alterné entre deux forfaits. Les deux tableaux suivants montrent respectivement les solutions trouvées par l'optimisateur pour ces trois personnes.

Log-Likelihood	1.641758e - 007
Nombre d'itérations	32
β_1	9.709099
β_2	-9.173421
α_{11}	2.774542
α_{12}	-0.1911135
α_2	7.015845e - 004

Log-Likelihood	0.000000
Nombre d'itérations	7
β_1	7.436224e - 005
β_2	-1.487248e - 004
α_{11}	37.78432
α_{12}	-0.2292519
α_2	6.478812

Log-Likelihood	1.909543
Nombre d'itérations	60
β_1	9.992509
β_2	6.589173
α_{11}	3.501393
α_{12}	-0.2396853
α_2	0.002371479

On voit que, si il y a une alternance dans le choix des forfaits, le modèle a du mal à reproduire le comportement de l'utilisateur. Cela s'explique par le fait que les termes expliquant l'évolution du choix de l'utilisateur à travers le temps étaient contenus dans les paramètres γ_i et ς_i , que nous avons dû supprimer car les données lacunaires nous empêchaient de les évaluer. Dans le cas du choix d'un seul forfait, le modèle reproduit par contre parfaitement le choix de l'utilisateur, le log-vraisemblance étant de 0 ou très proche de 0, et donc la vraisemblance de 1.

α_{11} prend une valeur positive, dans les deux cas. Cette valeur peut être interprétée

comme l'effet de la consommation optimale sur l'utilité d'un forfait. Il est cohérent que plus la consommation optimale est importante, plus l'utilisateur attribue une utilité importante au forfait. α_2 reflète l'effet du revenu sur l'utilité, ou encore l'effet de la part du budget qui peut être attribué à autre chose qu'à la consommation de minutes de communication. Il semble cohérent que la valeur de ce paramètre soit inférieur à celle de α_{11} , puisque, bien que l'utilisateur attribue une utilité positive au fait de pouvoir conserver son argent pour faire autre chose que téléphoner, l'utilité de pouvoir téléphoner doit être plus important que celle de conserver son argent pour que l'utilisateur décide de prendre un forfait téléphonique. α_{12} est le paramètre de concavité de la fonction d'utilité, et les β_j capturent les effets de facteurs qui ne sont pas pris en compte mais qui agissent de manière globale sur toutes les périodes et, lorsque les données sont agrégées, sur tous les utilisateurs.

Le modèle est en revanche assez sensible à la variation du point de départ. En effet, lorsque nous faisons varier celui-ci, l'optimum change. La valeur de la vraisemblance ne change pas, et les remarques faites précédemment sur la signification des paramètres sont toujours respectées, mais la valeur des paramètres peut varier de manière assez importante.

Nous considérons dans un second temps un ensemble d'utilisateurs. Nous supposons qu'ils partagent un comportement semblable. Comme les termes du modèle qui traduisent l'évolution du choix de l'utilisateur à travers le temps et comme chaque période est supposée identique aux autres, les choix deviennent indépendant et les observations ne sont donc plus ordonnées par période. Considérons un ensemble de N utilisateurs faisant un choix par mois durant M mois et ayant tous le même comportement, c'est à dire le même ensemble de paramètres ω , le problème d'estimation des paramètres est donc le même que si nous considérons 1 utilisateur faisant un choix par mois durant $N \times M$ mois.

Nous prenons un ensemble de 10000 utilisateurs, et nous sélectionnons tous les utilisateurs qui répondent aux critères précédemment décrits, que nous suivons à travers 10 mois. L'ensemble final comporte 599 utilisateurs, soit 5990 observations. Le résultat obtenu est le suivant :

Log-Likelihood	1.909549
Nombre d'itérations	155
β_1	1.445656
β_2	7.926147
α_{11}	$1.930082e + 002$
α_{12}	-0.1290808
α_2	$1.263853e - 003$

Comme nous pouvons l'observer, les résultats restent cohérents avec les remarques faites précédemment. Le nombre d'itérations a été multiplié par 2.6, ce qui est d'autant plus négligeable par rapport à l'augmentation du nombre d'utilisateurs (599 fois plus nombreux) que le temps de calcul n'est que de 35 secondes. La vraisemblance prend une valeur quasiment identique au cas où un utilisateur change de forfait d'un mois sur l'autre. On peut donc en déduire que cette vraisemblance correspond aux limites du modèle lorsque le choix de forfait change à travers les observations (mois ou utilisateurs) sans que cela soit explicable (les termes temporels ayant été supprimés et le choix des utilisateurs étant aléatoire).

En faisant varier le point de départ, nous pouvons observer une réduction du nombre d'itérations, mais pas une augmentation significative. La valeur de la vraisemblance reste la même, mais celle des paramètres peut varier, tout en restant cohérente avec les observations faites.

CHAPITRE 4

PERSPECTIVES D'ÉTUDE

Cette partie proposera un certain nombre de perspectives d'études qui pourraient être menées. Nous parlerons tout d'abord d'une évolution vers un modèle mixed-logit, puis nous proposerons une généralisation de notre modèle pour des offres plus complexes telles que celles rencontrées en téléphonie mobile et nous finirons par parler de la corrélation temporelle.

4.1 Évolution mixed-logit

Tel qu'il est ici décrit, le modèle est du type logit. Dans sa formulation originel, il est du type mixed-logit.

Les modèle mixed-logit considèrent que l'incertitude que nous avons sur le comportement des utilisateurs est issue de deux facteurs. D'une part, l'utilité que nous avons exprimée de manière explicite, dite utilité observée, dépend d'un certain nombre de variables observées, par exemple de la consommation (bien que notre cas soit particulier car la consommation n'est pas réellement une variable puisque qu'elle est définie comme maximisant l'utilité). Il y a pourtant un certain nombre de facteurs qui ne sont pas pris en compte, soit parce que nous ne les connaissons pas, soit parce que nous ne pouvons pas les observer. La spécification logit, comme la spécification mixed-logit, tient compte de cette erreur en ajoutant le terme d'erreur ϵ_{ijt} , aléatoire i.d.d. double exponentiel. C'est la part de l'utilité qui n'est donc pas observée. Le mixed-logit, lui, considère de plus que nous avons une connaissance statistique du comportement d'un groupe d'utilisateurs et, pour refléter cet

aspect, stipule que les paramètres ne sont pas fixés, mais répartis aléatoirement de manière à introduire une hétérogénéité dans le comportement d'un utilisateur à l'autre. Cette hétérogénéité, dans notre cas, est défini comme suit :

$$\text{soit } \omega_i = (\beta_{i1}, \dots, \beta_{in}, \alpha_{11}, \alpha_{12}, \alpha_2, \gamma_i, \varsigma_i), \quad \omega_i \sim N(S_i \mu, \Lambda)$$

avec S_i matrice contenant les informations démographiques relatives à l'utilisateur i , μ le vecteur de niveau de population et Λ la matrice de covariance de la population. L'étalonnage du modèle consiste donc à déterminer les propriétés de la distribution (μ et Λ) qui reproduisent le mieux les données historiques. Sont alors définis la vraisemblance conditionnelle, fonction de paramètres déterministes :

$$L_i(\omega_i, \sigma^2) = \prod_{t=1}^{T_i} L_{i,t}(\omega_i, \sigma^2)$$

et la vraisemblance inconditionnelle, qui est l'espérance des vraisemblances conditionnelles sur une certaine distribution :

$$L_i = \int \int \dots \int \prod_{t=1}^{T_i} L_{i,t}(\omega_i, \sigma^2) f(\omega_i) d\omega_i$$

Nous avons opté pour cette simplification car, comme nous pouvons le voir, la vraisemblance ne prend pas une forme explicite. La maximisation de cette expression devient alors délicate, et sort du cadre de notre étude. Un certain nombre de méthodes peuvent cependant être rencontrées dans la littérature.

4.2 Généralisation du modèle

En examinant les spécificités du modèle, nous réalisons que sa structure peut se généraliser de manière à particulièrement bien représenter des problèmes très actuels et aux enjeux économiques très importants du domaine de la télécommunication. Il s'agit des problèmes de modélisation de la demande en téléphonie mobile. Lorsque nous observons les offres de téléphonie mobile, elles se présentent quasiment toutes de la manière suivante : pour un prix de base, il est proposé à l'utilisateur un certain temps de communication non tarifée en journée (par exemple 100 minutes), un certain temps de communication soir et week-end non tarifée (par exemple 500 minutes), et un certain nombre de messages textes (SMS) inclus (par exemple 10 messages). Au delà de ces limites, chacun de ces trois types de consommation est facturé à un certain prix. Bien que le nombre de SMS émis soit une quantité discrète, il ne semble pas trop irréaliste de le traiter comme un nombre continu. Nous parlerons donc par la suite de "minutes" pour faire référence de manière indifférenciée à une unité de consommation.

Ainsi, chaque type k de communication peut être représenté par :

- $x_{ijt k}$ nombre de minutes du type de consommation k consommées par l'utilisateur i dans le cadre du forfait j durant la période t
- A_{jk} nombre de minutes du type de consommation k offertes par le forfait j
- p_{1k} prix de la minute avant A_{jk} pour le type de consommation k dans le cadre du forfait j (0 dans notre cas)
- p_{2k} prix de la minute après A_{jk} pour le type de consommation k dans le cadre du forfait j

Et nous pouvons directement procéder à la généralisation du modèle précédent.

L'utilité se définit alors :

$$\begin{aligned}
 U_{ijt}(\{x_{ijtk}\}_{k \in \{1, \dots, n\}}, x_{ijt(n+1)}) &= \beta_{ij} + \sum_{k=1}^n (\alpha_{ik1}x_{ijtk} + \alpha_{ik2}x_{1ijtk}^2 \\
 &\quad + (z'_{ijtk}\gamma_{ik})x_{1ijtk}) + \alpha_{i(n+1)}x_{ijt(n+1)} \\
 &\quad + w'_{ijt}\varsigma_i + \epsilon_{ijt}
 \end{aligned}$$

La consommation optimale est obtenue comme suit :

$$\begin{aligned}
 U_{itj}^{max} &= \max_{\{x_{ijtk}\}_{k \in \{1, \dots, n\}}} U_{ijt}(\{x_{ijtk}\}_{k \in \{1, \dots, n\}}, x_{ijt(n+1)}) \\
 \text{s.c.} \quad &\sum_{k=1}^n \delta_{x_{ijtk}-A_{jk}}^{|A_{jk}-x_{ijtk}|} (p_{2jk}(x_{ijtk} - A_{jk}) + p_{1jk}A_{jk}) \\
 &\quad + \sum_{k=1}^n \delta_{A_{jk}-x_{ijtk}}^{|A_{jk}-x_{ijtk}|} (p_{1jk}x_{1ijtk}) + x_{2ijt} = I_i - F_j
 \end{aligned}$$

où $\delta_i^j = 1$ si $i = j$ et $\delta_i^j = 0$ si $i \neq j$ (Symbole de Kronecker)

Les conditions nécessaires d'optimalité nous fournissent des optima potentiels similaires à ceux du modèle de base :

$$\begin{aligned}
 x_{1ijtk}^{candopt,1} &= \frac{\alpha_{i(n+1)}p_{1jk} - \alpha_{ik1} - z'_{ijtk}\gamma_{ik}}{2\alpha_{ik2}} \\
 x_{1ijtk}^{candopt,2} &= \frac{\alpha_{i(n+1)}p_{2jk} - \alpha_{ik1} - z'_{ijtk}\gamma_{ik}}{2\alpha_{ik2}}
 \end{aligned}$$

La valeur de l'utilité optimale qui en résulte s'exprime elle aussi de manière simi-

laire :

$$U_{itj}^{max} = \begin{cases} U_{itjk}(0, x_{2itjk}) & \text{if } x_{1itjk}^{candopt,1} \leq 0 \\ U_{itjk}(x_{1itjk}^{canopt,1}, x_{2itjk}) & \text{if } 0 < x_{1itjk}^{candopt,1} < A_{jk} \\ U_{itjk}(A_{jk}, x_{2itjk}) & \text{if } x_{1itjk}^{candopt,1} \geq A_{jk} \text{ and } x_{1itjk}^{candopt,2} \leq A_{jk} \\ U_{itjk}(x_{1itjk}^{canopt,2}, x_{2itjk}) & \text{if } A_{jk} < x_{1itjk}^{candopt,1} < B \\ U_{itjk}(B, x_{2itjk}) & \text{if } x_{1itjk}^{candopt,2} \geq B \end{cases}$$

$$\forall k \in \{1, \dots, n\}$$

4.3 Facteurs de corrélation temporelle et autres facteurs décisionnels

Dans l'implémentation du modèle que nous avons faite, nous ne pouvions pas calibrer γ_i et ς_i car l'échantillon ne contenait aucune information sur l'influence du dépassement du nombre de minutes incluses lors d'une certaine période sur le choix du forfait utilisé durant la période suivante. De plus, il n'existe que 3 forfaits incluant un bloc de minutes, dont 2 très similaires. Les alternatives sont donc assez limitées. En appliquant le modèle à la téléphonie mobile, nous disposerions sur-ement d'un échantillon significatifs d'utilisateurs dépassants le nombre de minutes incluses. De plus, ce type de téléphonie offre un certain nombre de forfaits de structure similaire qui, en fonction du prix de base payé, comprennent plus ou moins de minutes incluses. Dans ces conditions il deviendrait donc possible d'étudier les changements des utilisateurs vers des forfaits plus ou moins chers lors du dépassement du nombre de minutes incluses ou en fonction des changement de politiques tarifaires.

L'auteur du modèle d'origine proposait les variables suivantes :

State dependence : Les études montrent qu'un utilisateur a peu tendance à changer de forfait, même si il en existe un plus avantageux que celui actuellement souscrit. Cela peut s'expliquer par la démarche à faire pour changer de forfait, par une certaine appréhension du changement, et par le fait que les inconvénients de renoncer aux options actuelles semble plus évidents que les avantages que peuvent procurer de nouvelles options. Tout ceci est représenté par un ensemble contenant une variable par forfait disponible. La variable correspondant au forfait choisi en $t - 1$ vaut 1, et toutes les autres sont nulles.

Past quantity : Cette variable, qui contient la consommation en période $t - 1$, tente de capturer l'effet de la consommation antérieure sur la consommation actuelle (qui peut par exemple témoigner d'une certaine "accoutumance" au confort que procure le téléphone).

Past underage revenue : Cette variable considère le temps que l'utilisateur aurait pu consommer en période $t - 1$ sans dépasser le temps inclus dans le forfait, et multiplie ce temps par le prix de la minute de communication au-delà du bloc de temps inclus. Cette variable cherche ainsi à exprimer un "manque à gagner" en terme financier, tel que pourrait le percevoir l'utilisateur. Si celui-ci consomme plus que le temps inclus, alors cette variable est nulle.

Past overage revenue : Cette variable, symétriquement à la précédente, exprime l'argent dépensé en consommation au-delà du temps inclus dans le forfait.

Cumulative underage/overage revenue : Ces deux variables sont la somme des deux précédentes sur toutes les périodes antérieures. Elle tentent de capturer l'effet de l'accumulation de la "sur consommation" ou de la "sous consommation".

Promotion : Cet ensemble contient une variable par forfait. Chaque variable est établie à 1 si une promotion a été faite concernant ce forfait en période $t - 1$ et vaut 0 sinon. Nous sortons ici du contexte de la corrélation temporelle. Nous cherchons à capturer l'effet de la promotion sur le choix du consommateur.

Une étude pourra donc être menée dans une de ces trois directions. L'évolution vers le mixed-logit nécessite une bonne connaissance des méthodes d'estimation des paramètres. L'étude du modèle généralisé proposé et la prise en compte des facteurs de corrélation temporelle nécessitent de posséder un ensemble de données approprié.

CONCLUSION

Ce n'est que très récemment que la recherche académique a commencé à adapter certains outils scientifiques au marché de la télécommunication. Ces outils ont cependant été utilisés avec succès dans beaucoup d'autres domaines, et nous en avons une bonne maîtrise.

Le modèle que nous avons étudié, initialement proposé par Iyengar (Iyengar, 2004), va dans ce sens en tentant d'intégrer la structure de tarification non linéaire, structure classique dans l'industrie des services de télécommunication, aux modèles de choix discrets. Il tente de plus de tenir compte du fait que l'utilisateur choisi son forfait en se basant sur une espérance de consommation. Comme nous avons pu le voir, les données que possèdent les opérateurs téléphoniques ne se prêtent pas forcément à ce genre d'études. Leur traitement peut être aussi long que le travail de modélisation lui-même, et l'absence de certaines informations peut nous amener à abandonner certaines parties du modèle. L'implémentation du modèle peut être assez longue, et les outils nécessaires nécessitent de solides bases en informatique. Nous avons aussi pu constater qu'aux aspects théoriques de la modélisation s'ajoutent certains aspects pratiques qui nécessitent un minimum d'expérience, comme par exemple les problèmes de mise à l'échelle de la fonction d'utilité. En définitive, le modèle se comporte de manière cohérente, mais les simplifications que nous avons dû y apporter pour remédier au manque de données diminuent sensiblement ses performances.

En améliorant la compréhension que nous avons des réactions du marché en fonction de l'évolution des offres, le but à plus long terme serait de développer un nouveau modèle de tarification qui saurait bénéficier de la finesse de ces prévisions. Il serait ainsi possible de dégager un avantage compétitif en terme de différenciation

des catégories d'utilisateurs, de capacité à anticiper les réactions de celles-ci et finalement d'efficacité de la tarification.

Comme nous l'avons vu au cours de cette étude, certains modèles de base récemment proposés dans la littérature académique peuvent être adaptés aux problématiques récentes du domaine de la téléphonie. Qui plus est, il existe un fort potentiel d'étude et d'amélioration dans ce domaine, potentiel d'autant plus intéressant que la compétition féroce qui règne dans le milieu industriel de la téléphonie peut rendre le moindre avantage compétitif décisif.

RÉFÉRENCES

- ALLENBY, G. M., SHIVELY, T. S., YANG, S., GARRATT, M. J. (2004), A Choice Model for Packaged Goods : Dealing with Discrete Quantities and Quantity Discounts, *Marketing Science*, Vol. 23, No. 1, Winter 2004, pp. 95-108
- ALTMANA, E. , BARMANB, D., EL AZOUZIC, R., ROSD, D. et TUFFIN, B. (2006), Pricing differentiated services : A game-theoretic approach, *Computer Networks* Volume 50, Issue 7 , 15 May 2006, pp. 982-1002
- BEN-AKIVA, M. et LERMAN, S. (1985), *Discret-Choice Anlysis : Theory and Application to Travel Demand*, MIT Press, Cambridge, MA
- BITRAN, G. et CALDENTHEY, R. (2003), An overview of pricing models for revenue management, *Manufacturing & Service Operations Management*, Vol. 5, pp.203-230
- BOUHTOU, M., GRIGORIEV, A., HOESEL, S. v., VAN DER KRAAIJ, A. F., Spieksma, F. C., Uetz, M. (2007), Pricing bridges to cross a river, John Wiley & Sons, Ltd, *Naval Research Logistics*, Vol. 54 ; N.4, pp. 411-420
- Boyd, E.A. et Bilegan, I.C. (2003), Revenue management and e-commerce, *Management Science*, Vol. 49, pp.1363-1386
- CHIANG, W-C., CHEN, J.C.H. et XU, X. (2007), An overview of research on revenue management : current issues and future research, *Int. J. Revenue Management*, Vol. 1, No. 1, pp.97-128.
- COLSON, B., MARCOTTE, P., SAVARD, G. (2007), An overview of bilevel optimization, *Annals of Operations Research*, Volume 153, Number 1, September 2007 , pp. 235-256(22)
- CÔTÉ, J.-P., MARCOTTE, P., SAVARD, G. (2003), A Bilevel Modeling Approach to Pricing and Fare Optimization in the Airline Industry, *Journal of Revenue and Pricing Management* 2, p. 23-36

- CROSS, R. (1997), Revenue management : hard-core tactics for market domination, New York : Broadway Books
- DAUDEL, S. et VIALLE, G. (1994), Yield Management : Application to Air Transport and Other Service Industries, Paris : Les Presses de L'Institut du Transport Aérien
- DEMPE, S. (2002), Foundations of Bilevel Programming, Kluwer Academic Publishers
- ELMAGHRABY, W. et KESKINOC AK, P. (2003), Dynamic pricing in the presence of inventory considerations : research overview, current practices, and future directions, Management Science, Vol. 49, pp.1287-1309
- GILBERT, F., MARCOTTE, P., SAVARD, G. et SCHOEB, A. (2007), A Behavioral Model of Bell Canada's Long-Distance Customers, Publication
- INGOLD, A., MCMAHON-BEATTIE, U. et YEOMAN, I. (2000), Yield Management, Strategies for the Service Industries. London : Thomson
- IYENGAR, R. (2004), A Structural Demand Analysis for Wireless Services under Nonlinear Pricing Schemes, Columbia University
- JENKINS, D. (1995), Handbook of Airline Economics, New York : McGraw-Hill
- KIMES, S. E. (1989), Yield Management : A tool for capacity-constrained service firms, Journal of Operations Management, No. 8, pp.348-363.
- KIMES, S.E. (2003), Revenue management : a retrospective, Cornell Hotel and Restaurant Administration Quarterly, Vol. 44, pp.131-138
- LUNA L. (2002), Wireless sector descends into battle over pricing, Telephony ; Sep 16, 2002 ; 243, 10 ; ABI/INFORM Trade & Industry pp. 18
- MCGILL, J.I. et VAN RYZIN, G. (1999), Revenue management : research overview and prospects, Transportation Science, Vol. 33, pp.233-256
- MACKIE-MASON, J. K., RIVEROS, J. F., GAZZALE, R. S. (2000), Pricing and Bundling Electronic Information Goods : Experimental Evidence, MIT Press

NAGEL, N. T., HOLDEN, R. K. (contributor) (1994), *The Strategy and Tactics of Pricing : A Guide to Profitable Decision Marketing*, Prentice-Hall, Englewood Cliffs, 2nd edition

PAK, K. et PIERSMA, N. (2002-2003), *Airline revenue management : an overview of OR techniques 1982-2001*, Econometric Institute Report EI

PHILIPS, R. (2005), *Pricing and Revenue Optimization*, Stanford University Press

SMITH, B. C., Leimkuhler, J. F., Darrow, R. M. (1992), *Yield Management at American Airlines*, *Interface*, No. 22, pp.8-31.

TALLURI, K.T. et VAN RYZIN, G.J. (2004), *The theory and practice of revenue management*, Boston : Kluwer Academic Publishers

TALLURI, K. van RYZIN, G. (2004), *Revenue management under a general discrete choice model of consumer behavior*, *Management Science archive* Volume 50, Issue 1, pp. 15-33

TRAIN, K. E. (2002), *Discrete Choice Methods with Simulation*, University of California, Berkeley and National Economic Research Associates, Inc.

WEATHERFORD, L.R. et BODILY, S.E. (1992), *A taxonomy and research overview of perishable-asset revenue management : yield management, overbooking, and pricing*, *Operations Research*, Vol. 40, pp.831-844

YEOMAN, I. et MCMAHON-BEATTIE, U. (2004), *Revenue Management and Pricing : Case Studies and Applications*. London : Thomson

YIKUAN, L. et NORSWORTHY, J.R. (1998), *The analysis of market efficiency : a case of bundling telecomservices*, *Engineering and Technology Management*, 11-13 Oct 1998 pp.413-418

TILMAZ, O. et CHEN, I.-R. (2006), *Utilizing call admission control to derive optimal pricing of multiple service classes in wireless cellular networks*, *Proceedings of the 12th International Conference on Parallel and Distributed Systems*, Vol. 1, pp. 605-612

ANNEXE A**FONCTION D'UTILITÉ**

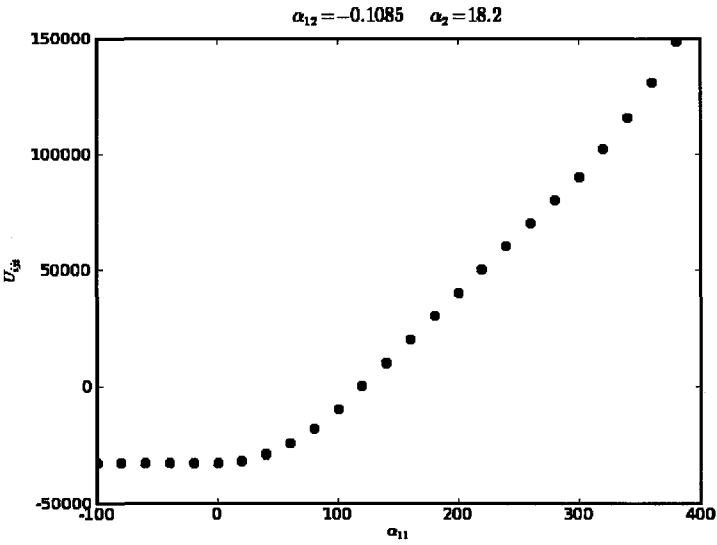


Figure A.1 $U_{ijt} = f(\alpha_{11})$

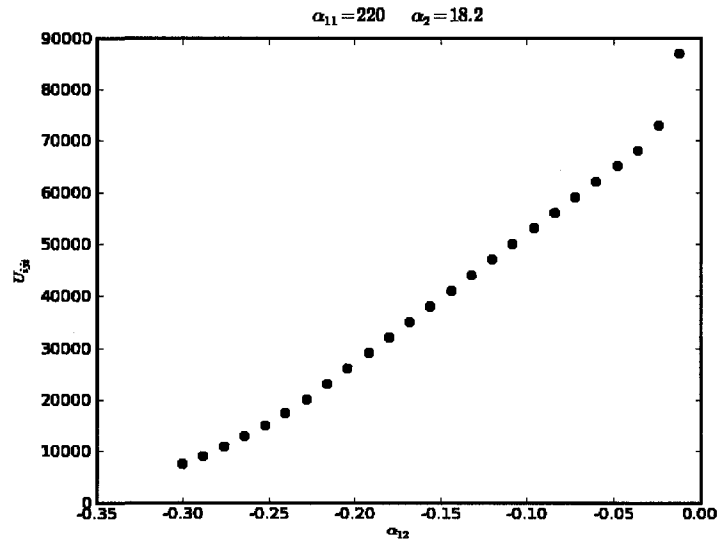
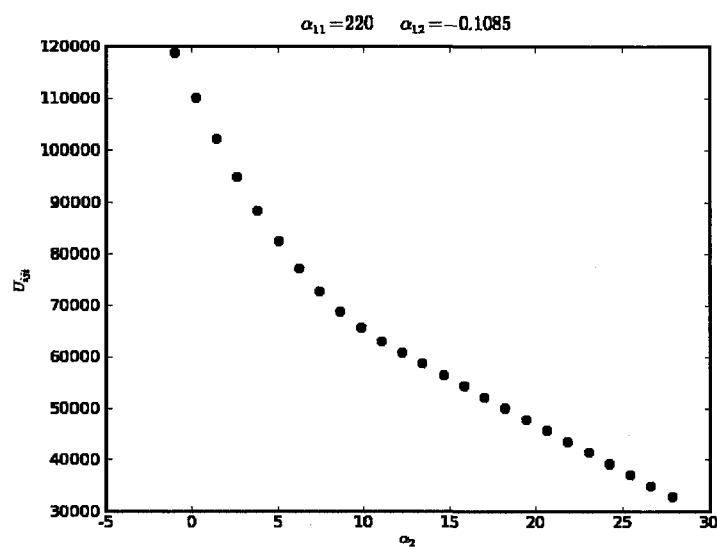
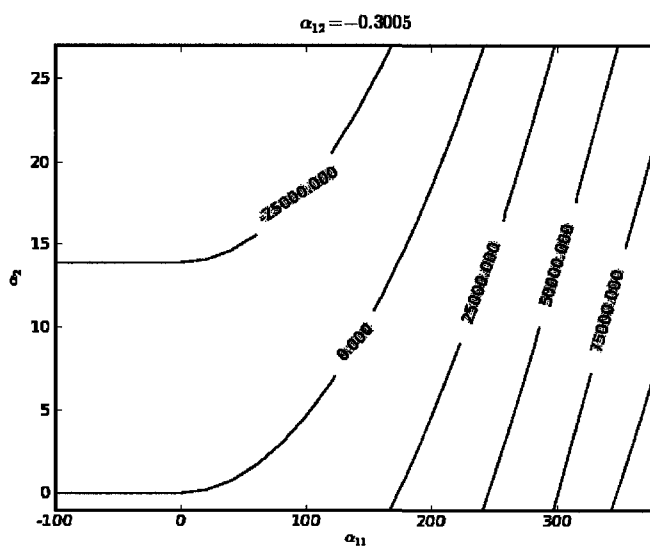


Figure A.2 $U_{ijt} = f(\alpha_{12})$

Figure A.3 $U_{ijt} = f(\alpha_2)$ Figure A.4 $U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.3005$

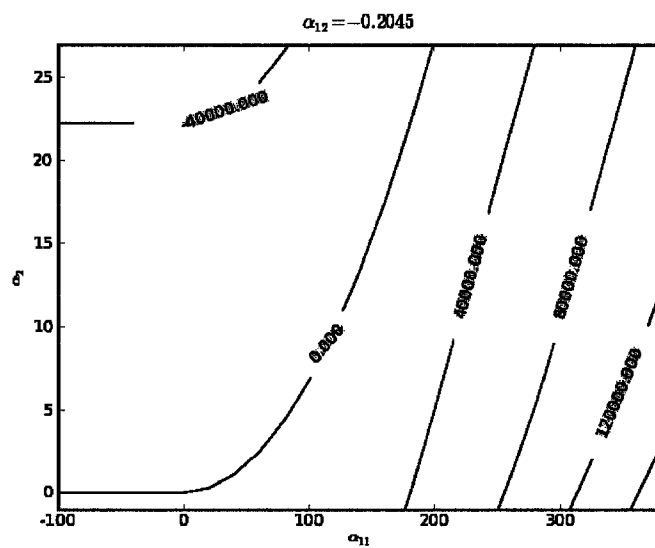


Figure A.5 $U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.2045$

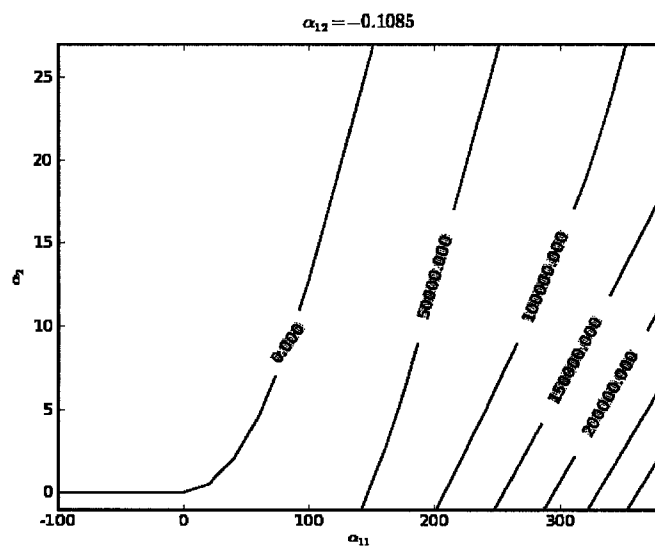


Figure A.6 $U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.1085$

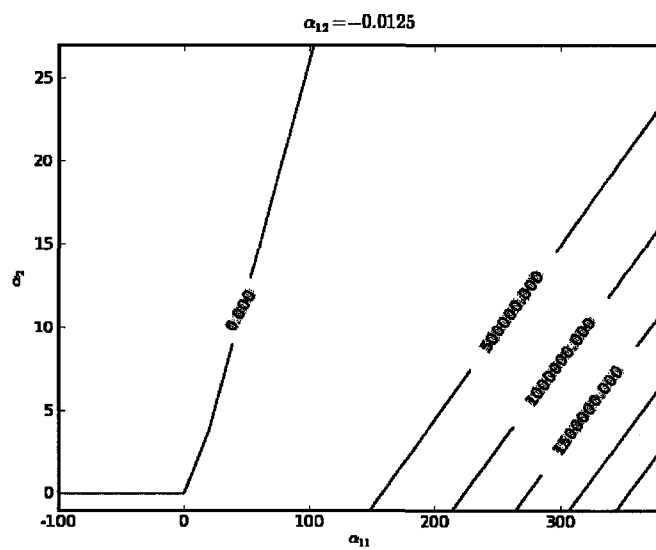


Figure A.7 $U_{ijt} = f(\alpha_{11}, \alpha_2)$ $\alpha_{12} = -0.0125$