

211.3380.6

HEC MONTRÉAL

Analyse de survie de la relation entre les clients et le fournisseur dans le domaine des télécommunications

par

Sarmiza Davidescu

Sciences de la gestion

*Mémoire présenté en vue de l'obtention
du grade de maître ès sciences
(M.Sc.)*

Janvier 2006

© Sarmiza Davidescu, 2006

No. 28
2006

**AVIS DE CONFORMITÉ À LA POLITIQUE EN MATIÈRE D'ÉTHIQUE
DE LA RECHERCHE AVEC DES ÊTRES HUMAINS DE L'ÉCOLE HEC MONTRÉAL**

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait les exigences de notre politique en cette matière :

Titre du projet de recherche :

Analyse de survie de la clientèle dans l'industrie de téléphonie.

Chercheur principal :

Sarmiza-Maria Davidescu, étudiante, M.Sc., option marketing, codirigée par Denis Larocque et Pierre Balloffet.

Date de la déclaration du projet au Comité d'éthique de la recherche :

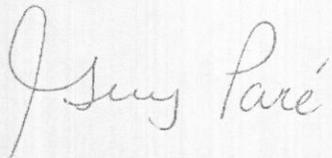
Le 21 avril 2004

Date d'approbation du projet :

Le 26 avril 2004

Date de l'émission de l'avis :

Le 26 avril 2004



Guy Paré, président,
Comité d'éthique de la recherche,
HEC MONTRÉAL

SOMMAIRE

Il est incontestable que la durée de la relation entre les entreprises et leurs clients joue un rôle majeur sur les stratégies de marketing adoptées par les dirigeants de ces entreprises. En effet, les entreprises essaient d'établir, de développer et de maintenir des relations durables avec leurs clients.

L'objectif de notre étude est de développer et d'évaluer un modèle pour la durée de la relation client – fournisseur afin d'identifier les facteurs influençant la période pendant laquelle un client reste abonné au service. Pour ce faire, nous avons utilisé une méthode statistique particulière : l'analyse de survie à temps discret.

Les données utilisées dans notre recherche proviennent d'une compagnie de télécommunications appelée Teleka (nom fictif donné à l'entreprise pour préserver sa confidentialité). La collecte de données s'est échelonnée sur une période de trois années. Les clients (des PME abonnées aux services de télécommunications fournis par Teleka) ont été suivis pendant une période de deux ans. Nous avons analysé l'évolution de la durée de vie du client pour deux services fournis par Teleka, le service local et le service Internet.

Les principaux facteurs influençant la décision du client de quitter le service local sont : le nombre de services auxquels le client est abonné, le nombre de numéros de téléphone par facture ainsi que l'utilisation du service interurbain d'un autre fournisseur. Compte tenu de ces facteurs, les résultats obtenus sont similaires pour les deux services analysés. Ces résultats peuvent constituer un outil efficace pour les gestionnaires désirant adapter les stratégies de marketing existantes afin de mieux cibler les clients et de répondre à leurs besoins.

TABLE DES MATIÈRES

SOMMAIRE.....	i
LISTE DE TABLEAUX.....	vi
LISTE DE FIGURES.....	vii
REMERCIEMENTS.....	viii
CHAPITRE 1 INTRODUCTION.....	1
CHAPITRE 2 REVUE DE LA LITTÉRATURE.....	4
2.1 La fidélisation de la clientèle.....	4
2.1.1 Le marketing transactionnel vers le marketing relationnel.....	5
2.1.2 Le concept de fidélisation.....	8
2.1.3 La gestion de la relation client.....	10
2.1.4 La contribution de la fidélisation sur la rentabilité de l'entreprise.....	13
2.1.5 L'attrition de la clientèle.....	14
2.2 L'exploration et l'entreposage de données.....	15
2.3 Les techniques et méthodes d'exploration de données.....	18
2.3.1 Les techniques d'exploration de données appliquées au repérage des clients susceptibles de partir ou de répondre à une offre.....	18
2.3.2 Les méthodes d'analyse de survie.....	19
2.4 L'applicabilité d'analyse de survie.....	21
2.4.1 La fidélisation de la clientèle dans l'industrie des télécommunications.....	21

2.4.2 Les applications d'analyse de survie dans l'industrie de l'assurance	26
CHAPITRE 3 MÉTHODOLOGIE	29
3.1 Objectifs de la recherche	29
3.2 Données	30
3.2.1 Composition des données	30
3.2.2 Variables dépendantes.....	30
3.2.3 Variables indépendantes	32
3.2.4 Collecte de données.....	34
3.3 La révision des données	35
3.4 Méthodes utilisées et implémentation	39
3.4.1 L'analyse de survie à temps continu.....	40
3.4.2 L'analyse de survie à temps discret.....	42
3.4.2.1 La fonction de risque à temps discret	43
3.4.2.2 Les prédicteurs et le risque à temps discret.....	44
3.4.2.3 Le modèle de Cox pour le risque à temps discret.....	44
3.4.2.4 L'adaptation de la base de données pour effectuer l'analyse.....	45
3.4.3 La sélection des variables et des modèles	49
CHAPITRE 4 RÉSULTATS	50
4.1 Le service local	50
4.1.1 Description de la variable cible : service local	51
4.1.2 La fonction de survie et la fonction de risque du service local	51

4.1.3 Les modèles de survie développés	54
4.1.3.1 Le meilleur modèle contenant le premier bloc de covariables.....	55
4.1.3.2 Le meilleur modèle contenant le deuxième bloc de covariables.....	56
4.1.3.3 Le meilleur modèle contenant le troisième bloc de covariables.....	59
4.2 Le service Internet	61
4.2.1 Description de la variable cible : service Internet	61
4.2.2 La fonction de survie et la fonction de risque du service Internet	62
4.2.3 Les modèles de survie développés	65
4.2.3.1 Le meilleur modèle contenant le premier bloc de covariables.....	65
4.2.3.2 Le meilleur modèle contenant le deuxième bloc de covariables.....	66
4.2.3.3 Le meilleur modèle contenant le troisième bloc de covariables.....	68
CHAPITRE 5 CONCLUSION GÉNÉRALE	72
5.1 Conclusion et discussion	72
5.2 Les principales limites.....	75
5.3 Les avenues de recherche.....	76
BIBLIOGRAPHIE.....	77

LISTE DES TABLEAUX

Tableau 3.1 Exemple de variable dépendante	32
Tableau 3.2 Présentation des intervalles de temps	34
Tableau 3.3 Cheminement des clients avant le nettoyage	37
Tableau 3.4 Cheminement des clients après le nettoyage	37
Tableau 3.5 Exemples de données des clients	46
Tableau 3.6 Données préparées pour l'analyse de survie en temps discret.....	48
Tableau 4.1 La distribution des clients en fonction du temps	51
Tableau 4.2 La fonction de survie et de risque pour tous les clients	52
Tableau 4.3 Le modèle incluant le premier groupe de covariables	55
Tableau 4.4 Le modèle incluant le deuxième groupe de covariables ...	58
Tableau 4.5 Le modèle incluant le troisième groupe de covariables ...	60
Tableau 4.6 La distribution des clients en fonction du temps	62
Tableau 4.7 La fonction de survie et de risque pour tous les clients	63
Tableau 4.8 Le modèle incluant le premier groupe de covariables	65
Tableau 4.9 Le modèle incluant le deuxième groupe de covariables ...	67
Tableau 4.10 Le modèle incluant le troisième groupe de covariables ..	69

LISTE DES FIGURES

Figure 2.1 Marketing One-to-One (Peppers et Rogers, 1993).....	7
Figure 2.2 Les étapes d'évolution des clients (Berry et Linoff, 2000, p. 74).....	12
Figure 3.1 Données censurées à droite.....	41
Figure 4.1 La proportion de clients abonnés au service local à la fin de chaque période	52
Figure 4.2 La proportion de clients ayant quitté le service local à chaque période parmi ceux ayant débuté la période	53
Figure 4.3 La proportion de clients abonnés au service Internet à la fin de chaque période.....	63
Figure 4.4 La proportion de clients ayant quitté le service Internet chaque période parmi ceux ayant débuté la période.....	64

REMERCIEMENTS

Tout d'abord, je tiens à remercier mon directeur de recherche, M. Pierre Balloffet et mon co-directeur, M. Denis Larocque, pour leur support constant et leurs commentaires pertinents au cours de la réalisation de cette étude.

Je tiens aussi à remercier Mme Sihem Taboubi et M. Olivier Bahn d'avoir lu ce travail.

J'aimerais aussi remercier Mme Sonia Benghiat et Mme Anne-Marie Bergeron pour m'avoir accordé la permission d'utiliser leur base de données, ainsi que M. Yves Lapierre pour toutes ses connaissances pratiques et les conseils qu'il m'a donnés pendant cette recherche.

Je voudrais, également, remercier Mme Karine Pouliot et Mme Ana Maer pour leurs corrections et leurs conseils linguistiques. Des remerciements spéciaux sont adressés à Kamel El-Hedhli pour ses conseils.

Finalement, un merci tout spécial à Victor, mon mari, pour son support et ses encouragements constants. Merci aussi à mes parents pour leur soutien moral tout au long de mes études.

CHAPITRE 1 INTRODUCTION

Selon le concept récent de marketing, la fidélisation du client et la gestion de la relation client sont considérées comme des éléments prioritaires étant donné la concurrence féroce dans chaque industrie et chaque secteur. Dans ce contexte, il est important d'analyser plus en détail le processus de gestion de la clientèle afin de mieux comprendre comment les activités d'une entreprise peuvent influencer sa relation avec les consommateurs. Garbarino et Johnson (1999) montrent qu'il y a une grande diversité entre les consommateurs en ce qui concerne leur décision de s'engager dans une relation avec une entreprise. D'après eux, il existe deux groupes de clients : un groupe incluant les clients moins orientés vers une relation, mais plus vers la satisfaction d'une transaction individuelle; et un autre regroupant les clients intéressés par une relation basée sur la confiance.

Dans les années 50, le concept de marketing correspondait à une technologie de commercialisation qui se traduisait par la communication de masse et la distribution de masse (Meyer-Waarden, 2002). Ce concept a créé une distance entre l'entreprise et ses clients. En même temps, à cause d'une offre excessive, les entreprises ont été obligées de diminuer les coûts de marketing et de conserver leur part de marché

et leurs clients. Pour ce faire, elles ont essayé de mettre en œuvre un ensemble de stratégies et de tactiques visant à recruter de nouveaux clients tout en veillant également à garder les anciens.

Plusieurs chercheurs ont conclu que retenir un client coûte moins cher qu'en acquérir de nouveaux et que les meilleurs clients sont les plus rentables (Berry, 1980; Bolton et Drew, 1994; Joo, Jun et Kim, 2002). Reichheld et Sasser (1990) ainsi que Reichheld (1996) indiquent qu'une augmentation du taux de fidélisation de 5 % permettrait d'augmenter le profit de 25 % à 85 % selon les secteurs d'activité, car les entreprises enregistrent, chaque année, une perte de 15 % à 20 % des clients. Pour cette raison, dans les années 80, une relation entre le client et le fournisseur s'est développée. La conservation des clients est devenue un objectif stratégique très important pour les entreprises. Le développement des technologies de l'information a contribué à l'atteinte de cet objectif en employant les termes *Customer Relationship Management (CRM)* ou *One-to-One Marketing*, qui sont impliqués directement dans le marketing orienté vers le client (Meyer-Waarden, 2002).

Dans le contexte économique actuel, « la vision du marché à court terme est remplacée par une approche à long terme » (Meyer-Waarden, 2002, p. 16) où la fidélisation de la clientèle joue un rôle essentiel. Ce mémoire présente le concept de fidélisation de la clientèle dans le

domaine des télécommunications et les facteurs qui influencent la décision du client de quitter le fournisseur. Tout particulièrement, nous cherchons à développer un modèle dynamique permettant d'étudier la durée de la relation entre le client et le fournisseur.

Ce travail débute par une revue de la littérature où sont présentés quelques aspects et remarques sur le concept de la fidélisation en marketing. L'apport de la technologie en marketing ainsi que des applications sur la conservation de la clientèle dans les domaines des télécommunications et de l'assurance sont également étudiés dans ce chapitre.

Par la suite, nous présentons la méthodologie adoptée dans ce mémoire. Ainsi, nous décrivons de manière succincte les méthodes statistiques auxquelles nous faisons appel afin de développer notre modèle.

Quant à la partie suivante, on y trouve une présentation des résultats obtenus ainsi qu'une discussion de ceux-ci. Finalement, nous présentons les conclusions et les limites de cette étude en traçant certaines avenues de recherche.

CHAPITRE 2 REVUE DE LA LITTÉRATURE

Dans la première partie de ce chapitre, nous allons passer en revue les études sur la fidélisation de la clientèle et la gestion de la relation entre le client et le fournisseur afin de mieux comprendre ces approches. Ensuite, le développement de la technologie et sa contribution au marketing seront présentés, suivis des techniques et méthodes statistiques appliquées pour retenir et fidéliser les clients. Enfin, nous exposerons des applications d'analyse de survie dans différents domaines.

2.1 La fidélisation de la clientèle

Les sections à venir présentent l'évolution du concept de marketing et de celui de fidélisation. En effet, dans un environnement économique moins stable, la fidélisation est un sujet de plus en plus discuté par des auteurs comme Reichheld et Sasser (1990), Jones et Sasser (1995), Rust et Zahoric (1993), et ce, depuis plusieurs années. On y étudie également un concept lié à la gestion de la relation client ou

Customer Relationship Management (CRM) et à la rentabilité de l'entreprise.

2.1.1 Le marketing transactionnel vers le marketing relationnel

Le concept de marketing a eu une évolution sans précédent. Il s'est toujours adapté aux fluctuations économiques et sociales. Cela a conduit à l'émergence du marketing relationnel (Kotler, 1992; Webster, 1992). Dans la littérature, certains auteurs (Peppers et Rogers, 1995; Garbarino et Johnson, 1999; Meyer-Waarden, 2002) considèrent qu'il y a un déplacement du marketing transactionnel vers le marketing relationnel.

Selon Webster (1992), ce déplacement est divisé en plusieurs étapes : les transactions, les transactions répétées, les relations à long terme, le partenariat entre le vendeur et l'acheteur, les alliances stratégiques, les organisations réseaux et l'intégration verticale. Il considère que les deux premières étapes représentent l'approche transactionnelle et que les autres étapes font partie de l'approche relationnelle.

Le marketing transactionnel est caractérisé par une tendance à court terme (Vence, 2002), par des relations temporaires, statiques et discontinues avec des partenaires (Coviello et Brodie, 2001). En plus, il met l'accent sur la part de marché (Peppers et Rogers, 1995).

L'approche transactionnelle est abordée comme la plus simple forme d'échange. Les transactions discrètes ont un début et une fin bien définis. Elles sont indépendantes les unes des autres et n'ont aucune interaction entre elles. En plus, cette approche transactionnelle se manifeste par un manque de communication entre les entreprises et les clients (Dwyer, Schurr et Oh, 1987).

Cette approche est définie par des relations impersonnelles, c'est-à-dire des relations anonymes, où l'accent est mis sur l'échange rapide d'un produit standardisé à un prix compétitif (Dwyer, Schurr et Oh, 1987; Day, 2000). Webster (1992) observe que, dans une telle approche, le prix est devenu l'élément le plus important pour conclure une transaction. Aussi, il n'y a aucune différenciation du produit et aucune référence à la marque (Webster, 1992).

Par contre, le marketing relationnel met l'accent sur les relations à long terme (Day, 2000; Coviello *et al*, 2002), sur le dynamisme et la personnalisation de la relation avec les clients et les partenaires, ainsi que sur le client, ses préférences et sa satisfaction (Deshpande, Farley et Webster, 1993).

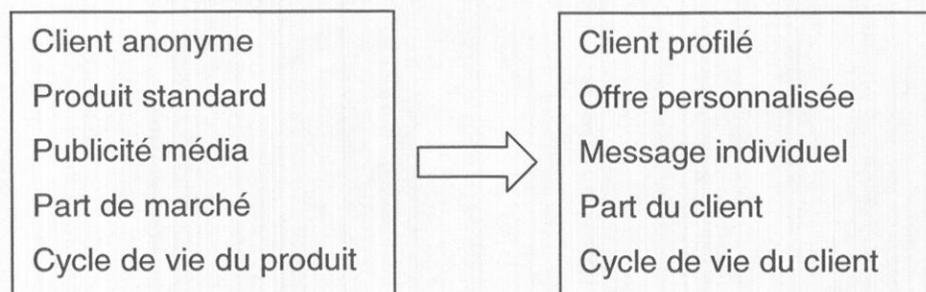
Plusieurs auteurs (Berry, 1995; Jackson, 1985; Morgan et Hunt, 1994) ont défini l'approche relationnelle comme l'ensemble des actions prises par l'entreprise pour attirer, conserver et consolider les relations avec ses clients. De plus, Morgan et Hunt (1994) ajoutent à cette relation

d'autres intervenants : fournisseurs, partenaires latéraux (concurrents, gouvernements) et partenaires internes (employés, départements). Selon Ricard et Perrier (1999), cette approche représente un engagement personnalisé à long terme entre le client et l'entreprise générant des bénéfices pour les deux parties.

Coviello et al (2002) divisent le marketing relationnel en : marketing par base de données (utiliser la nouvelle technologie), marketing interactif (développer des relations interpersonnelles entre l'entreprise et ses clients) et marketing de réseau (établir des relations inter-organisationnelles).

En conclusion, les deux approches ne sont pas nécessairement mutuellement exclusives et ce déplacement touche plusieurs éléments (figure 2.1). La nouvelle forme est caractérisée comme Marketing *One-to-One* par Peppers et Rogers (1993) :

Figure 2.1 Marketing One-to-One (Peppers et Rogers, 1993)



2.1.2 Le concept de fidélisation

Nous allons présenter des définitions de la fidélisation selon plusieurs auteurs, car aucun auteur ne tient compte à la fois de tous les éléments et aspects sur ce sujet.

D'un côté, la fidélisation, selon Gremler et Brown (1996), représente le comportement d'achat répété chez un certain fournisseur, s'expliquant par le fait que le client a une attitude positive envers ce fournisseur et qu'il considère ses services comme étant utiles.

Selon quelques auteurs (Rusbult, 1980; Barksdale *et al.*, 1997; Suh, 1994), l'intention d'achat répétée semble être influencée par trois facteurs : la valeur du client (section 2.1.3), la satisfaction du client (Rust et Zahoric, 1993) et la perception du coût de changement de fournisseur. D'autres auteurs considèrent que le comportement d'achat répété dépend aussi des autres variables comme les attitudes (Jacoby et Chestnut, 1978), la confiance et l'engagement (Morgan et Hunt, 1994).

La satisfaction envers le produit ou le service renforce l'attitude, puis la fidélité (Meyer-Waarden, 2002). Donc, la fidélité dépend de la qualité du service que le client reçoit et de la perception du client par rapport à ce service. Aussi, une relation avec le client basée sur la confiance et

l'engagement constitue un facteur essentiel pour développer la fidélité (Meyer-Waarden, 2002).

Nous avons vu que le coût de changement de fournisseur est également une variable importante dans la fidélisation étant donné les coûts liés à la recherche d'un autre fournisseur et à la commodité du client (Meyer-Waarden, 2002). En effet, les efforts pour changer de produit ou de fournisseur demandent plus d'énergie de la part des clients, et les résultats ne sont pas toujours satisfaisants. De plus, si le client détient plusieurs services, le coût de changement augmente et l'intérêt est diminué. En d'autres mots, la vente croisée est une stratégie qui diminue l'attrition ou la perte de la clientèle, en augmentant le nombre des services utilisés (Kamkura, Ramaswami et Srivastava, 1991).

D'un autre côté, Barlow (1992, cité par Meyer-Waarden, 2002, p. 57) considère que « la fidélisation est une stratégie qui identifie les meilleurs clients, les maintient grâce à une relation interactive à valeur ajoutée et axée sur le long terme, pour accroître leur rendement ». Cette définition est proche du concept de CRM que nous allons décrire plus loin.

L'idée commune présentée par des auteurs comme Reichheld (1996) et Meyer-Waarden (2002) est que la fidélisation coïncide avec les efforts défensifs de l'entreprise pour éviter le départ des clients et qu'elle est indispensable dans le contexte actuel.

La fidélisation est un élément important, peu importe le domaine d'activité, mais pour les entreprises qui offrent des services elle l'est encore plus pour trois raisons. Premièrement, la fidélité est plus grande chez les consommateurs de services que chez les consommateurs de biens (Zeithaml, 1981). Deuxièmement, les services aident la compagnie à développer des relations interpersonnelles avec les clients (fournir plus d'occasions pour l'interaction entre les gens) et, par la suite, à accroître les occasions de développer la fidélisation (Parasuraman *et al.*, 1985; Surprenant et Solomon, 1987). Troisièmement, le risque perçu est plus grand à l'achat de services qu'à l'achat de biens (Murray, 1991).

2.1.3 La gestion de la relation client

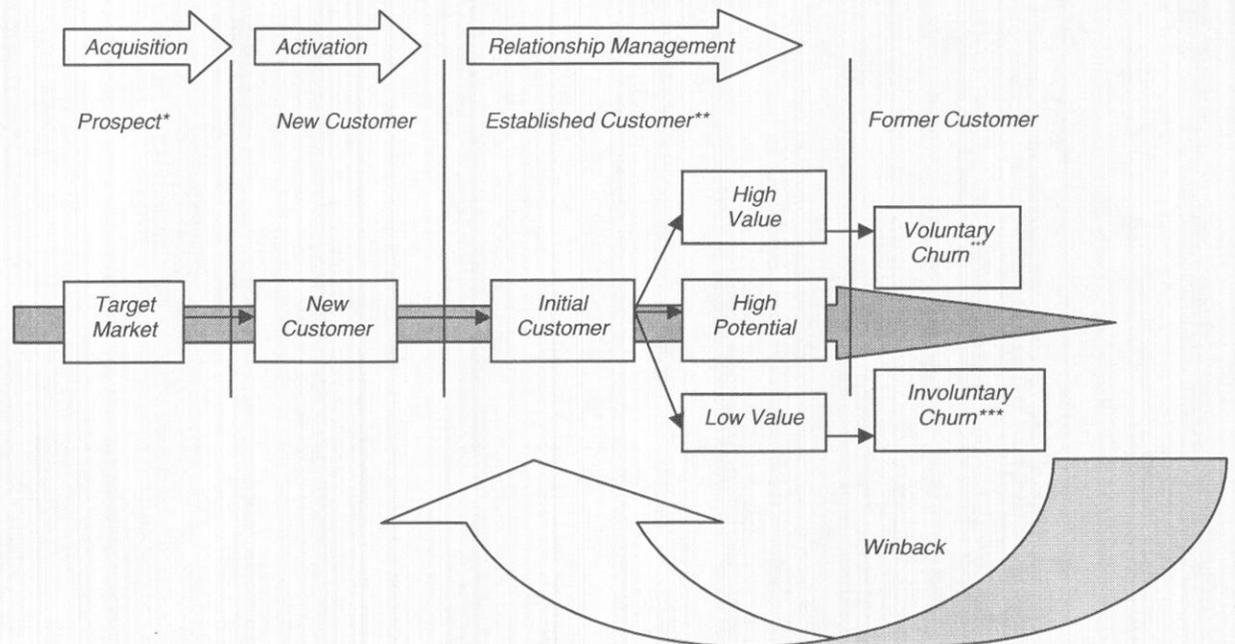
Selon Dwyer, Schurr et Oh (1987), l'entreprise cherche une relation stable et de longue durée avec le client. Plus tard, Crosby et Johnson (2000) prennent cette idée en l'incluant dans la définition de la gestion de la relation client ou *Customer Relationship Management* (CRM). Dans leur vision, la CRM est une stratégie d'affaires utilisée pour connaître davantage les besoins et les comportements du client afin de bâtir des relations durables, fortes et profitables avec lui. De bonnes relations avec le client constituent le cœur du succès dans les affaires. Berry et Linoff (2000, p. 80) résument la CRM en une seule phrase : « offrir le bon produit, pour le bon client, au bon moment et au bon

prix ». D'après eux, deux aspects sont à considérer dans l'approche de CRM. D'un côté, l'entreprise diminue ses coûts et augmente ses profits ; de l'autre, elle doit miser sur l'optimisation de la valeur de chaque client.

Une définition de la valeur du client est nécessaire, car elle est un élément utile dans le concept de CRM. Selon Schmittlein et Peterson (1994), la valeur actualisée du client ou *Life Time Value (LTV)* est le produit de l'espérance mathématique de profits réalisés par un client durant le temps t et la probabilité de survie de ce client au temps t . La LTV est donc une variable qui dépend du temps et qui estime la valeur d'un client dans le temps.

Un autre élément essentiel dans le marketing orienté vers le client est la durée de la relation client qui est corrélée à la valeur du client (Reichheld, 1996). La vie d'un client dans une entreprise peut être observée comme un cycle de vie caractérisé par trois étapes : l'introduction ou l'acquisition; le développement ou la maturité; et la rupture ou le déclin (Dwyer, Schurr et Oh, 1987). Autrement dit, on doit acquérir le client, le garder et développer une relation avec lui et sa rentabilité. Berry et Linoff (2000) ont un point de vue similaire, qui est illustré dans la figure 2.2.

Figure 2.2 Les étapes d'évolution des clients (Berry et Linoff, 2000, p. 74)



Source : Berry, M. J.A. et Linoff, G. S., *Mastering « Data Mining » – The Art and Science of Customer Relationship Management*, John Wiley & Sons, Inc., 2000, p. 74.

* *Prospects* sont les gens dans le marché cible qui ne sont pas encore clients.

** *Established Customer* utilise le service ou le produit.

*** *Voluntary churn* et *involuntary churn* sont expliqués dans la section 2.1.5.

La valeur du client et la durée de sa relation avec l'entreprise sont deux indicateurs fondamentaux dans la CRM. Grâce à eux, nous pouvons déterminer les segments plus et moins rentables afin d'appliquer la stratégie appropriée à chacun. En outre, une contribution importante dans la CRM est apportée par le développement des grandes bases de données, car elles réunissent d'information historique sur chaque client, en nous permettant d'évaluer les clients d'une manière individuelle et en leur présentant des offres personnalisées. Nous allons discuter plus loin du développement de la technologie de l'information.

2.1.4 La contribution de la fidélisation sur la rentabilité de l'entreprise

Selon Rosenberg et Crepiel (1984), Dawkins et Reichheld (1990), Reichheld et Sasser (1990) ainsi que Reichheld (1996), retenir un client coûterait jusqu'à six fois moins cher que d'en conquérir un nouveau. En effet, les coûts d'acquisition d'un client impliquent des dépenses en publicité et en promotion, des coûts de vente et d'autres coûts additionnels. Puisque ces coûts sont très élevés, ils doivent être amortis sur la durée de vie du client pour que celui-ci devienne rentable. Par conséquent, la fidélisation est la solution qui permet de réduire et d'amortir le budget d'acquisition de nouveaux clients.

Alors que les nouveaux clients sont non profitables pour une certaine période de temps suivant leur abonnement à un service, les clients existants sont plus ouverts à acheter d'autres services offerts par l'entreprise, et la relation avec eux est plus stable. En effet, selon Winer (2001), le revenu provenant des clients existants est deux fois plus grand que celui provenant des nouveaux clients. De plus, le client qui a une longue relation avec l'entreprise est prêt à payer un prix plus élevé que le nouveau client, car il apprécie déjà la relation avec la compagnie (Zeithaml *et al.*, 1996) et il est moins sensible au prix (Appiah-Adu, 1999). En fait, les clients fidèles sont profitables, car ils sont des utilisateurs fréquents.

En plus, les coûts pour servir un client qui est déjà satisfait du service et du fournisseur sont plus faibles que ceux liés à un nouveau client (Appiah-adu, 1999; Zeithaml *et al.*, 1996). Comme le client existant est souvent servi plus efficacement par l'entreprise, il est plus satisfait et recommande donc fréquemment et fortement la marque. Bref, Reichheld (1996, cité par Meyer-Waarden, 2002, p. 39) soutient que « les bénéfices des clients fidèles seraient croissants dans le temps et que l'on peut recruter de nouveaux consommateurs à un coût réduit en utilisant la fonction d'avocat du client fidèle, car celui-ci est supposé être un bon pour-parleur de l'entreprise ».

À l'unanimité, les chercheurs considèrent que la fidélisation est une source de revenu qui ne doit pas être négligée.

2.1.5 L'attrition de la clientèle

La manière efficace de retenir les clients existants, d'augmenter les taux de fidélisation des bons clients et de diminuer le nombre des clients moins profitables est une question de plus en plus étudiée. Les études se concentrent donc souvent sur la diminution de la perte de la clientèle. Schellinck (2002) ainsi que Berry et Linoff (2000) affirment que l'attrition ou perte des clients (*churn*) peut être de deux types : involontaire et volontaire. L'attrition involontaire est la perte des mauvais clients qui sont rejetés par la compagnie à cause du non-paiement de leur facture

depuis plusieurs mois. Quant à l'attrition volontaire, il s'agit de la perte des clients qui quittent la compagnie pour différentes raisons : insatisfaction du service, décès, changement d'adresse, prix élevé en comparaison avec le revenu, changement des offres de la concurrence. Il est important de distinguer l'attrition volontaire de l'attrition involontaire, car il s'avère très coûteux d'effectuer une analyse de rétention sur les clients qui ne paient pas leurs factures. Selon Berry et Linoff (2000), il y a aussi l'attrition silencieuse qui représente la perte des clients abonnés au service, mais qui ne l'utilisent pas.

D'après une étude sur les raisons pour lesquelles les clients quittent un fournisseur, 4 % de la perte de la clientèle provient de consommateurs qui sont décédés ou qui sont déménagés, 5 % sont influencés par leurs amis, 9 % changent pour la compétition, 15 % sont insatisfaits de la qualité des services. Enfin, 67 % des défections peuvent s'expliquer, en partie, par l'absence de stratégie afin de garder le contact avec le client (Cohn, 2005 [en ligne]).

2.2 L'exploration et l'entreposage de données

Pour les entreprises, le développement de la technologie au cours des dernières années a eu un effet bénéfique. De plus en plus, les renseignements sur le client et le produit sont emmagasinés et traités

dans de grandes bases de données à moindre coût. Grâce au développement de bases de données avec l'archivage de renseignements individuels sur le client, l'entreprise peut satisfaire de plus en plus les besoins individuels du client en adaptant les offres et les produits. Lorsque nous parlons de l'emmagasinement et du traitement des données, nous nous référons aux notions d'entreposage de données (*Data Warehouse*) et d'exploration de données (*Data Mining*).

Les données sont à la base des décisions d'affaires. Berry et Linoff (1997) considèrent que l'archivage de données constitue la mémoire de l'entreprise et que l'exploitation de ces données représente l'intelligence de l'entreprise.

Un *Data Warehouse* est un entrepôt de données d'entreprise. Il contient des données provenant de différentes sources, qui sont « nettoyées » et régulièrement intégrées dans un format commun.

La *Data Mining* est définie comme le processus d'exploration et d'analyse des grandes bases de données dans le but de découvrir des modèles de prévision significatifs en utilisant les outils de la statistique ou de l'intelligence artificielle (Berry et Linoff, 1997).

Dans la vision de Berry et Linoff (2000), il y a deux approches : l'exploration de données directe et indirecte. La prospection de données directe regroupe trois tâches : classification, estimation et prévision. Son

but est d'utiliser les données disponibles pour construire un modèle qui décrit la variable cible en fonction des autres variables disponibles. Quant aux règles d'association et à l'analyse de regroupement, ces éléments font partie de l'exploration de données indirecte. Le but de ce type de prospection est de comprendre les données historiques ainsi que d'établir des relations entre toutes les variables sans avoir une variable cible. Les deux approches ne sont pas mutuellement exclusives. En fait, les deux sont en général utilisées de façon concomitante dans l'analyse de données.

La prospection et l'entreposage de données sont de plus en plus intégrés dans le processus de prise de décision des grandes entreprises et plus spécifiquement dans les départements de marketing. Plusieurs des applications d'exploration de données qui ont connu du succès ont d'ailleurs été expérimentées dans le domaine du marketing. Dans ce domaine, la prospection de données est utilisée pour les deux variables qui appartiennent à l'équation du profit : terme du coût et terme du revenu. Par exemple, l'exploration de données peut être utilisée pour minimiser les coûts du marketing en éliminant les appels et les lettres aux clients qui sont moins susceptibles de répondre à des telles offres.

En d'autres termes, la prospection et l'entreposage de données peuvent être considérés comme des outils stratégiques pour mieux gérer les différentes catégories de clients.

2.3 Les techniques et méthodes d'exploration de données

Dans cette section, nous présentons quelques techniques de la prospection de données ainsi que l'analyse de survie appliquées dans le domaine du marketing.

2.3.1 Les techniques d'exploration de données appliquées au repérage des clients susceptibles de partir ou de répondre à une offre

Les méthodes statistiques traditionnelles peuvent être adaptées aux besoins. Elles sont utilisées pour trouver des tendances et des corrélations parmi les données.

Les techniques de prospection de données, telles que la régression logistique, les arbres de décision et les réseaux de neurones ont permis de développer des modèles prédictifs avec lesquels on reconnaît les désabonnements afin de repérer les clients qui sont plus susceptibles d'abandonner les services d'une compagnie. À l'aide de ces techniques, on peut mieux comprendre le comportement passé et présent des clients afin d'anticiper leur comportement futur. Cela permet, par exemple, une intervention avant même que le client change de fournisseur. Au moment où un segment de clients qui vont partir est repéré, on peut se concentrer seulement sur ces clients et appliquer une stratégie pour les retenir.

De plus, ces techniques aident à découvrir quels clients sont susceptibles de répondre à une offre. Ainsi, le profit augmente, car les dépenses pour envoyer l'offre aux clients qui ne sont pas intéressés sont éliminées.

Jusqu'à maintenant, nous avons parlé seulement des méthodes qui permettent de reconnaître les clients disposés à abandonner le service ou susceptibles de répondre d'une manière affirmative à une offre. Un modèle de survie peut être développé afin de prévoir quand le client quittera le service ou quand le client achètera le prochain service. Dans la section suivante, nous allons présenter une brève description de ces analyses de survie.

2.3.2 Les méthodes d'analyse de survie

Nous avons vu que la durée de la relation client est un facteur important à considérer sans lequel nous ne pouvons pas anticiper ni le comportement du client, ni l'intention de quitter son fournisseur. Pour ce faire, les techniques d'analyse de survie sont mises en application.

Morita, Lee et Mowday (1993) exposent brièvement l'histoire du développement des analyses de survie. Celle-ci débute en 1693, quand l'astronome E. Halley propose un tableau de survie qui est conçu pour le prix de la rente viagère, et qui est fondé sur la probabilité conditionnelle de survie des individus. Jusqu'en 1950, les chercheurs se

concentrent dans leurs applications sur des interprétations descriptives de tableaux de survie. Après 1950, les statisticiens font de grands progrès quant aux caractéristiques des estimateurs de survie (Cox et Oakes, 1984; Tuma et Hannan, 1984).

Depuis la publication de l'article de Cox en 1972 sur les modèles statistiques pour des données de survie, les méthodes de survie (principalement les méthodes pour les données en temps continu) sont devenues très populaires dans différentes disciplines. Au début, elles ont été appliquées en médecine. Elles ont depuis touché d'autres domaines tels que les affaires, la sociologie et l'économie.

En général, un modèle de survie peut être considéré comme un modèle de transition à deux états : 1) vivant/actif et 2) mort. La seule transition possible est de l'état 1 à l'état 2. Quand l'individu arrive dans l'état 2, le processus de vie est considéré comme terminé. Dans le cadre d'un processus d'affaires, le client est « né » quand il commence à utiliser le service (état 1), et il est « mort » quand il quitte la compagnie (état 2).

Dans le domaine des affaires, l'analyse de survie est utilisée pour :

- reconnaître les clients hautement susceptibles de quitter la compagnie,
- estimer la durée de vie du client dans la compagnie,

- segmenter les clients en fonction de la durée de l'abonnement et en fonction du niveau d'utilisation du service,
- prévoir quand le client fera son prochain achat ou s'abonnera à un autre service (vente incitative et vente croisée).

Nous présentons maintenant quelques cas, selon différents domaines (télécommunications et assurance), où l'analyse de survie a été appliquée dans le but d'accroître la fidélisation de la clientèle.

2.4 L'applicabilité d'analyse de survie

Dans les sections suivantes, nous décrivons des applications de méthodes d'analyse de survie sur la conservation de clients dans le domaine des télécommunications et sur l'intention d'achats répétés dans le domaine de l'assurance.

2.4.1 La fidélisation de la clientèle dans l'industrie des télécommunications

L'industrie des télécommunications est propice à la constitution de larges bases de données. Ces bases de données contiennent trois types de variables : des données détaillées sur les appels, qui décrivent les échanges enregistrés à travers des réseaux de télécommunication; des données de réseau, qui décrivent l'état des composantes logicielles

(*software*) et des composantes matérielles (*hardware*) dans le réseau; et des données sur les clients. Pour bien gérer ce grand volume de données, les techniques d'exploration de données ont été employées.

Dans cette industrie, les clients peuvent choisir entre plusieurs fournisseurs de services et peuvent souvent changer de fournisseur avec une grande facilité. Il s'agit d'un marché très compétitif : les clients y demandent des services de bonne qualité à bas prix, alors que les fournisseurs se concentrent sur l'acquisition de nouveaux clients comme objectif d'affaires.

Or, selon Meltzer, pour les compagnies de télécommunications fixes et mobiles, le taux d'attrition annuel de la clientèle est de 30 % à 35 % et l'acquisition de nouveaux clients devient de plus en plus difficile et coûteuse. C'est pour cette raison que la fidélisation des clients devient de plus en plus importante. Plusieurs compagnies de télécommunications semblent également avoir observé que seulement 20 % à 30 % de leurs clients génèrent l'essentiel de leurs profits.

Encore une fois, il est très important de savoir quel client doit être retenu, quelle est la valeur de ce client et quel profit il apporte, l'objectif premier étant de garder les clients les plus profitables. Plusieurs compagnies de télécommunications développent des stratégies de fidélisation afin de maintenir les clients plus longtemps en leur fournissant de meilleurs services et produits.

Dans son étude, Li (1995) détermine non seulement pourquoi les clients renoncent aux services, mais aussi pourquoi ils partent prématurément.

L'article poursuit trois objectifs :

- 1) comprendre la dynamique de la compétition (Si la compagnie perd des clients au profit de la concurrence, quel est le taux de perte et quelle est la probabilité de changer pour un compétiteur?);
- 2) déterminer les facteurs qui influencent la durée de vie d'un client dans l'entreprise et construire des profils des clients fidèles et non fidèles;
- 3) développer un modèle dans le but de découvrir les clients hautement disposés à changer de fournisseur afin de les retenir plus efficacement.

Trois groupes de variables ont été utilisés dans l'analyse : des données historiques, des variables de services et des variables démographiques et de style de vie. Le taux de survie a été calculé pour découvrir la proportion des clients qui gardent le service pour une durée spécifiée. L'auteur a remarqué que, après six mois, plusieurs abonnés ont changé de fournisseur. Après 51 mois, environ 85 % des abonnés utilisent encore le service, mais, après 60 mois, le taux de changement de service augmente d'une manière assez importante. Le taux de risque indique que la propension de ce changement est plus élevée au début de l'abonnement ainsi qu'après une très longue période de service, les

clients qui ne sont pas satisfaits ayant la tendance de partir au début de l'engagement.

Le modèle à risque proportionnel est utilisé pour découvrir les clients ayant une très haute probabilité de renoncement au service et les facteurs qui les caractérisent. Li (1995) considère que l'effet des covariables sur le taux de risque peut être interprété en fonction du risque relatif de changer de fournisseur. Par exemple, le taux de changement des clients mariés est de 15 % inférieur à celui des clients célibataires. En segmentant les clients en fonction du risque, le gestionnaire peut créer des profils pour des clients hautement susceptibles de partir (clients à haut risque) et pour ceux qui sont faiblement disposés à quitter l'entreprise (clients à bas risque). Le modèle de survie permet d'attribuer un score à chaque client en fonction du risque estimé de changement de fournisseur. En conséquence, le gestionnaire pourra prendre des mesures pour conserver ces clients.

Pour garder leurs clients plus longtemps, les entreprises doivent recommander le bon service à chaque client. Elles doivent aussi parfois l'encourager à changer le service choisi initialement pour un service plus économique, même au détriment de leurs profits à court terme. D'après Joo, Jun et Kim (2002), cette stratégie leur permet de retenir un client fidèle et réaliser un profit à long terme. Ces auteurs ont ainsi démontré que les clients utilisant un plan d'appels approprié ont un taux

de fidélisation plus élevé. Dans leur recherche, Joo *et al.* (2002) développent un modèle de survie (fonction de risque) qui est adéquat pour analyser les données de durée, afin d'observer la différence de taux de fidélisation entre les clients ayant un plan d'appels correctement choisi et ceux ayant un plan d'appels non adapté. Un des trois plans d'appels examinés est considéré comme cible. Neuf variables, démographique et de comportement d'achat ont été incluses dans le modèle de fonction de risque. Ce modèle leur a permis d'introduire des observations censurées à droite (qui surviennent lorsqu'un client n'a pas connu l'événement observé pendant la période étudiée) et, aussi, d'inclure les covariables dépendantes du temps (les variables indépendantes changeant des valeurs au cours de la période observée). Les deux termes utilisés ci-dessus seront développés dans le troisième chapitre.

Les auteurs ont trouvé que le taux d'attrition est plus faible pour les hommes que pour les femmes et plus élevé pour les clients avec un mauvais plan qui paient 1 \$ supplémentaire par mois que pour ceux qui ont sélectionné le plan optimal. De même, les clients qui signent un contrat obligatoire, qui changent de téléphone cellulaire fréquemment, qui dépensent plus par mois et qui possèdent un plan d'appels ciblé connaissent le taux d'attrition le plus faible et restent plus longtemps avec le même fournisseur. Donc, en sélectionnant les bons clients,

l'entreprise augmente son profit à long terme et les clients se montrent plus satisfaits des services.

2.4.2 Les applications d'analyse de survie dans l'industrie de l'assurance

À cause de l'offre très variée de services financiers, les compagnies d'assurance se concentrent de plus en plus sur la fidélisation de leurs clients. Une façon de garder les clients est de leur proposer plusieurs services. À l'aide des techniques d'analyse de survie, Harrison et Ansell (2002) ont analysé les opportunités de vente croisée pour une compagnie d'assurance. Le but de cette étude est d'établir les clients les plus susceptibles d'acheter des produits additionnels de la même compagnie, la nature de ces produits et le moment d'achat.

La première étape consiste à segmenter les clients en groupes à l'aide d'une analyse de regroupement, suivie d'une analyse discriminante permettant de prédire l'appartenance à un segment. Comme deuxième étape, une analyse de survie a été réalisée afin de prédire l'achat d'autres produits ou l'abandon prématuré de service. Deux cas ont été examinés : 1) le cas où il y a un deuxième achat et 2) le cas où il n'y a pas de deuxième achat. Dans le deuxième cas, le moment du deuxième achat n'a pas été observé; nous allons donc le considérer comme une observation censurée. Les covariables significatives incluses dans le

modèle sont l'âge, le sexe, le statut marital et un groupe de variables financières.

Les auteurs ont remarqué que les clients mariés, les gens plus âgés, les clients avec une expérience financière plus développée et les femmes ont plus de chances de se procurer le deuxième produit que les autres. Par contre, les hommes sont plus impliqués financièrement que les femmes. Ils prennent la décision d'achat en fonction du prix et des caractéristiques du produit, tandis que pour les femmes, la relation avec l'institution est plus importante. En conséquence, elles sont plus fidèles aux institutions financières que les hommes. Autant pour les femmes que pour les hommes, la probabilité de survie descend très rapidement dans les 1000 premiers jours pour ensuite ralentir et baisser plus lentement. En d'autres mots, les clients fidèles réaliseront un deuxième achat.

Sans aucun doute, l'analyse de survie est une technique très importante et de plus en plus utilisée afin de définir des stratégies de fidélisation de la clientèle dans différents domaines.

Puisque la fidélisation de la clientèle est essentielle dans le contexte économique actuel tel que vue tout au long de notre revue de la littérature, une étude sur ce sujet sera développée dans ce mémoire. Plus particulièrement, l'analyse de survie à temps discret sera appliquée sur des données appartenant au domaine des télécommunications afin

de bâtir un modèle dynamique sur la durée de la relation client – fournisseur.

Dans ce chapitre, nous avons examiné la littérature existante en relation avec notre sujet de recherche. C'est dans le chapitre suivant que nous présentons la méthodologie.

CHAPITRE 3 MÉTHODOLOGIE

Dans ce chapitre, nous allons décrire la méthodologie adoptée pour effectuer notre étude et atteindre les objectifs de notre recherche. Nous présenterons d'abord les objectifs de recherche. Ensuite, nous décrivons les données et discuterons de la manière dont nous les avons manipulées. Enfin, nous procéderons à la description des méthodes d'analyse de survie en général, puis à la mise en application des méthodes utilisées dans notre étude et à la présentation des critères de sélection de variables et de modèles.

3.1 Objectifs de la recherche

Cette recherche est une étude de cas pour Teleka. Teleka est une importante compagnie de télécommunications. Le but de notre recherche est de développer et d'évaluer un modèle pour la durée de la relation client – fournisseur afin de trouver les facteurs influençant le temps pendant lequel un client reste abonné au service. Pour ce faire, nous utiliserons une méthode statistique particulière : l'analyse de survie à temps discret.

3.2 Données

À ce stade, il est important de souligner qu'il s'agit d'une recherche effectuée en collaboration avec une compagnie privée et que les données et les méthodes sont soumises à un contrat de confidentialité. Par conséquent, il ne sera pas possible de présenter en détail toutes les étapes de cette recherche.

3.2.1 Composition des données

Dans notre étude, nous allons utiliser un échantillon qui provient du Canada. L'échantillon contient 4560 abonnés aux services fournis par Teleka. Les clients sont de petites et de moyennes entreprises (PME). Un client compte de deux types de variables : des variables cibles, indiquant si le client a renoncé aux services fournis par Teleka et des variables explicatives, décrivant l'historique du client avec le fournisseur de services. On y retrouve, entre autres, les transactions déjà effectuées par le client, les minutes utilisées pour chaque service, les revenus par chaque service, le nombre de services abonnés, etc.

3.2.2 Variables dépendantes

Il importe de prendre connaissance de la structure particulière des données, plus particulièrement celle de la variable cible (un service). Nous allons analyser l'évolution de la durée de vie du client avec les services fournis par Teleka. Cette compagnie offre aux clients 16

services différents, notamment, l'Internet de base ou haute vitesse, des services téléphoniques résidentiel et interurbain, des services pour les appels 1-800 ainsi que l'hébergement de pages Web.

La première variable dépendante que nous modélisons est le service local. Nous avons groupé les trois services locaux fournis par Teleka dans une seule variable qui est considérée comme variable cible dans notre recherche. Cette variable dépendante (service local) est une variable binaire qui prend la valeur 1 si le client est abonné à au moins un des trois services locaux. Cette variable prend la valeur 0 si le client n'est pas à aucun service local.

La deuxième variable cible qui est étudiée dans notre analyse est le service Internet. Cette variable prend la valeur 1 si le client est abonné au service et 0 s'il ne l'est pas. Les autres services sont utilisés dans notre recherche comme variables explicatives.

Chaque service est caractérisé par huit variables qui représentent les 8 périodes analysées (tableau 3.2). Par exemple, « servlocal_A » représente la variable « service local dans la période A ». Chaque variable est une variable binaire (prend la valeur 1 si le client est abonné au service en question et 0 s'il ne l'est pas). Nous présentons quelques exemples de cheminement client pour ce service (une des deux variables cibles) dans le tableau suivant (tableau 3.1).

Tableau 3.1 Exemple de variable dépendante

No d'identification du client (ID)	Servlocal_ _A	Servlocal _B	Servlocal _C	Servlocal _D	Servlocal _E	Servlocal _F	Servlocal _G	Servlocal _H
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	0
3	1	1	1	1	0	0	manquant	manquant
4	1	1	1	0	0	1	1	0
5	0	1	1	1	0	0	0	
6	manquant	0	0	1	1	1	1	0
7	1	0	0	0	1	1	manquant	manquant
8	manquant	1	1	1	0	1	1	1
9	0	0	0	1	1	1	0	0
10	1	1	0	1	manquant	manquant	manquant	manquant

La valeur « manquant » représente les données non-disponibles.

3.2.3 Variables indépendantes

Pour ce qui est des variables explicatives, ce sont soit des variables catégorielles, soit des variables continues dont les valeurs sont réelles. En même temps, les variables indépendantes peuvent être constantes dans les huit périodes analysées, ou elles peuvent varier dans le temps et prendre des valeurs différentes dans chaque période. Dans la prochaine section, nous allons présenter, plus en détail, les intervalles de temps pendant lesquels nous suivons les clients.

Certaines variables, telles que les minutes utilisées et les revenus sur les services, sont calculées d'une manière différente. Ces variables explicatives, qui varient en fonction du temps, représentent la moyenne des trois derniers mois de chaque période pour les cinq derniers intervalles de temps, et la moyenne d'un à deux mois pour les autres.

Il y a plusieurs types de variables explicatives utilisées dans notre recherche comme par exemple les revenus, les minutes utilisées par type de service, le nombre de services, les types de services, le nombre de numéros de téléphone par facture, le nombre de factures par client, etc. Le groupe de variables « revenus » comprend des revenus récurrents, des revenus non récurrents, d'autres qui sont réduits pour différentes raisons, des revenus par carte d'appels ou provenant du service sans frais (numéros 1-800) ou d'appels faits aux numéros 1-900, des taxes sur la facture, des revenus totaux sur la facture, etc. Les revenus récurrents représentent des revenus mensuels qui sont fixes comme l'abonnement au service local ou l'abonnement à Internet. Par contre, les revenus non récurrents sont différents d'un mois à l'autre. Par exemple, ils représentent des revenus d'appels interurbains ou le montant pour le modem acheté pour Internet haute vitesse. Les variables de type « minutes » comprennent les minutes utilisées pour toutes les catégories de services fournis par Teleka.

3.2.4 Collecte de données

Dans notre cas, les données utilisées proviennent de la base des données du Teleka. La collecte de données s'est échelonnée sur trois années distribuées en quatre temps appelés ici année 434, année 435, année 436 et année 437. Les clients en question sont des PME qui se sont abonnées aux services de télécommunications fournis par Teleka entre le 1^{er} septembre de l'année 434 et le 31 août de l'année 435. Ils ont ensuite été suivis pendant une période de deux ans. Donc, un client qui s'est abonné le 1^{er} septembre de l'année 434 a été suivi jusqu'au 31 août de l'année 436, et un autre qui a signé son contrat le 1^{er} septembre de l'année 435 a été suivi jusqu'au 31 août de l'année 437.

Nous sommes intéressés à analyser l'évolution de la relation entre Teleka et ses clients pendant une période de deux ans, divisée en huit intervalles inégaux présentés dans le tableau 3.2.

Tableau 3.2 Présentation des intervalles de temps

Période	Description
A	Le moment de l'acquisition de service (début du 1 ^{er} mois)
B	Après 1 mois de service (fin du 1 ^{er} mois)
C	Après 2 mois de service (fin du 2 ^e mois)
D	Après 3 mois de service (fin du 3 ^e mois)
E	Après 6 mois de service (fin du 6 ^e mois)
F	Après 12 mois de service (fin du 12 ^e mois)
G	Après 18 mois de service (fin du 18 ^e mois)
H	Après 24 mois de service (fin du 24 ^e mois)

La période A n'est pas vraiment une période : c'est le moment où le client s'abonne au premier service fourni par Teleka. L'abonnement peut se faire à n'importe quel moment pendant une année, entre le 1^{er} septembre de l'année 434 et le 31 août de l'année 435.

Par exemple, en observant la période B, nous pouvons affirmer que B est la fin du premier mois ou le début du deuxième mois. Nous savons à quelle période le client a quitté la compagnie et quand il a abandonné un service ou un numéro de téléphone. Nous pouvons aussi observer si un client quittant le service revient plus tard, mais nous ne tiendrons pas compte de cette situation. Nous prendrons en considération la période depuis l'abonnement du client à Teleka jusqu'au moment où il quitte la première fois la compagnie.

3.3 La révision des données

Le nettoyage et la manipulation des données sont des étapes principales dans notre recherche. Un effort important y a été consacré et un travail minutieux de vérification a été entrepris.

Dans notre analyse nous avons gardé les clients (les observations) abonnés au début de la période observée et nous allons les suivre jusqu'à la dernière période. Donc, nous analysons les observations pour lesquelles la variable cible prend la valeur 1 dans la période initiale (A)

et dans la deuxième période (B) (car les valeurs des variables explicatives en période C sont basées sur les valeurs de ces variables en A et B, tel qu'il est expliqué plus loin). En d'autres mots, les clients qui s'abonnent au service et l'utilisent au moins un mois sont les sujets de notre recherche. Dans l'exemple de cheminement présenté au tableau 3.1 nous avons utilisé seulement les clients avec les numéros d'identification (ID) : 1, 2, 3, 4 et 10; les autres ont été éliminés. Pour le service local offert par Teleka, nous avons trouvé que 18,3 % des clients ne sont pas abonnés à ce service pendant le premier mois d'utilisation d'autres services. Par conséquent, il reste 3 726 clients pour l'analyse. En ce qui a trait au service Internet, 84,4 % des clients sont éliminés de la base de données parce qu'ils ne sont pas présents ni au cours du premier mois d'abonnement, ni durant les autres périodes. Pour modéliser cette variable, nous avons conservé 712 clients.

Afin de procéder au nettoyage des données, nous avons analysé l'évolution des variables dépendantes durant huit périodes. En suivant le cheminement, nous avons observé que certains clients s'abonnent au service, restent quelques périodes, partent et reviennent. Quelques exemples de cheminement client sont présentés dans le tableau 3.1. Dans cette recherche, nous avons suivi les clients au cours de leur abonnement. Pour un client quittant le service et revenant plus tard, nous avons pris en considération seulement la première fois où il s'est abonné. Dès qu'il quitte le service une fois, nous considérons qu'il est

parti pour toujours. Ce qui nous intéresse est le temps avant son premier départ. Nous présentons, dans les tableaux suivants, les cheminements des clients avant (tableau 3.3) et après le nettoyage (tableau 3.4).

Tableau 3.3 Cheminement des clients avant le nettoyage

No d'identification du client (ID)	Servlocal_A	Servlocal_B	Servlocal_C	Servlocal_D	Servlocal_E	Servlocal_F	Servlocal_G	Servlocal_H
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	0
3	1	1	1	1	0	0	0	0
4	1	1	1	0	0	1	1	0
10	1	1	0	1	0	0	0	manquant

Tableau 3.4 Cheminement des clients après le nettoyage

No d'identification du client (ID)	Servlocal_A	Servlocal_B	Servlocal_C	Servlocal_D	Servlocal_E	Servlocal_F	Servlocal_G	Servlocal_H
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	0
3	1	1	1	1	0	0	0	0
4	1	1	1	0	0	0	0	0
10	1	1	0	0	0	0	0	0

Pour savoir si le nombre de services auxquels un client est abonné influence la durée d'abonnement, la variable explicative « sumvar » a été créée. Elle représente la somme de services auxquels le client est abonné moins le service en question. En effet, afin de prédire la variable

« service local », par exemple, « sumvar » est la somme de tous les services moins les trois services locaux. De même, pour prédire la seconde variable dépendante, soit « Internet », « sumvar » est la somme de tous les services moins « Internet ».

Les événements futurs sont prédits à l'aide des événements présents. Pour ce faire, nous utilisons les variables explicatives de la période précédente afin de modéliser la variable dépendante (variable cible) de la période actuelle. Par exemple, la variable explicative « revr10_B » (les revenus par rapport aux appels faits au numéro 1-900 en période B) est utilisée afin de modéliser le service local en période C.

Une autre façon de faire consiste à créer d'autres variables. Nous avons créé des variables qui sont calculées comme la différence entre une variable en période précédente et une variable en période présente afin de prédire la variable cible en période suivante. D'une manière générale, des nouvelles variables du type $(X_{t-2} - X_{t-1}) = X_t^*$ ont été créées comme nouvelles variables explicatives. De cette façon, la variable explicative « drevr10_C » a été créée comme différence entre « revr10_A » et « revr10_B » ayant pour but de prédire la variable cible « servlocal_C ».

Une troisième manière de prédire la variable dépendante est la

suivante : $X_t^{**} = \begin{cases} 1, & \text{si } (X_{t-2} - X_{t-1}) > 0 \\ 0, & \text{si } (X_{t-2} - X_{t-1}) \leq 0 \end{cases}$. Par exemple, « ddsumvar_t » est

la variable indicatrice montrant s'il y a eu une baisse du nombre de services entre deux périodes consécutives. Si le client est abonné à moins de service pendant la période $t-1$ que pendant la période $t-2$, il y a une baisse du nombre des services et « $ddsumvar_t$ » prend la valeur 1.

Afin de préparer notre base de données pour l'analyse de survie, nous avons procédé à une étape intermédiaire, soit la création de deux variables : la durée et la censure. La durée représente le temps, le nombre de périodes, d'avant lesquelles un client reste abonné au service. La censure prend la valeur 0 si le client a quitté le service avant les huit périodes ou prend la valeur 1 si le client est toujours abonné après les huit périodes. Dans ce cas, l'observation est donc censurée.

3.4 Méthodes utilisées et mises en application

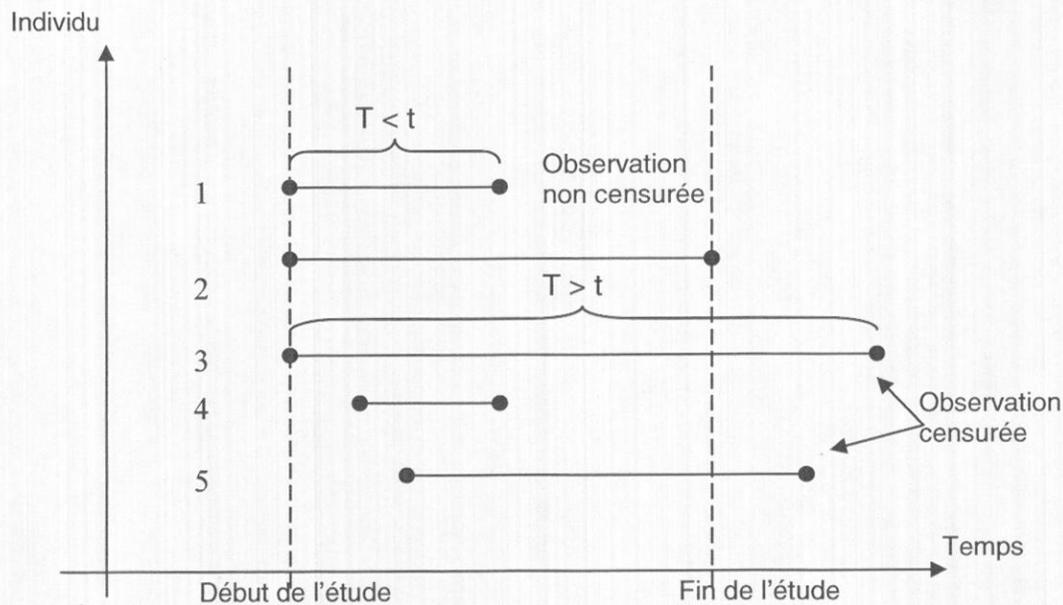
Dans ce chapitre nous allons décrire les méthodes d'analyse de survie, en débutant brièvement avec l'analyse de survie à temps continu. La prochaine partie, présentée en détails, est l'analyse de survie à temps discret, suivie de l'adaptation du fichier de données à la méthode appliquée dans cette étude. En dernier lieu, nous allons discuter des critères de sélection des variables et du modèle.

3.4.1 L'analyse de survie à temps continu

L'analyse de survie étudie l'occurrence d'un événement sur un intervalle de temps. Selon Allison (2003), un événement est une transition d'un état à un autre qui se produit à un moment donné. Dans notre étude, nous analysons le temps pendant lequel un client reste abonné à un service offert par l'entreprise et les effets des variables explicatives sur ce temps nommé « temps de survie ». La durée d'observation ou le temps de survie représente la différence entre le moment où l'événement (le client quitte le service) se produit et le temps initial (début de l'abonnement).

Si le temps observé est plus petit que le temps de survie, on parle d'une observation tronquée (censurée). Il y a deux type de cas censurés : la troncature à gauche et la troncature à droite (Allison, 2003). Dans notre étude, il existe des cas censurés à droite, représentés par les individus qui sont toujours abonnés au service à la fin de la période observée. Dans la figure suivante, il y a quelques exemples de cheminement qui expliquent les situations possibles.

Figure 3.1 Données censurées à droite



Un des éléments de base d'une analyse de survie est la fonction de survie qui constitue la probabilité de survivre au delà d'un certain temps, donc de ne pas vivre l'événement. Un autre élément important est la fonction de risque qui mesure le risque instantané de quitter l'état au temps t , étant donné que l'individu s'est rendu au temps t (il n'a pas quitté précédemment).

En analyse de survie, les variables explicatives sont nommées « prédictors » et leurs valeurs sont, soit fixes, soit variables dans le temps.

Lorsque le temps est mesuré en continu, le modèle de plus en plus utilisé pour analyser des données de survie est le celui de régression de Cox. Il modélise l'effet des prédictors sur la fonction de risque. Il peut

également tenir compte de prédicteurs dont la valeur change dans le temps.

Dans notre étude, le temps est mesuré d'une manière discrète et nous utilisons un modèle adapté à cette situation.

3.4.2 L'analyse de survie à temps discret

La méthode que nous avons appliquée à nos données est l'analyse de survie à temps discret décrite par Singer and Willett (1993). Pour les petites et moyennes entreprises, nous sommes intéressés à découvrir le risque qu'elles quittent leur fournisseur de service de télécommunication ainsi que les facteurs qui font varier ce risque.

Comme nous l'avons vu dans les sections précédentes, les variables dépendantes que nous modélisons sont le service local et le service Internet. L'événement ciblé, c'est-à-dire lorsque l'entreprise (client) abonnée a quitté le service, est non répété : une fois que l'événement s'est produit, il ne se produit pas une seconde fois. En d'autres termes, une fois que le client a quitté le service (Internet ou le service local), nous considérons qu'il l'a quitté pour toujours, même s'il s'y réabonne plus tard. Les modèles avec des événements répétés sont traités dans Allison (1984), Tuma et Hannan (1984) ainsi que dans Yamaguchi (1991).

Pour enregistrer l'occurrence d'événement en temps discret, il faut diviser le temps continu dans un nombre infini d'intervalles consécutifs $(0, t_1], (t_1, t_2], \dots, (t_{j-1}, t_j], \dots$. On considère j comme l'index de la période; la $j^{\text{ème}}$ période commence immédiatement après le temps t_{j-1} et continue jusqu'au temps t_j , en l'incluant. Dans notre recherche, pour un client qui s'est abonné le 1^{er} octobre de l'année 434, la période D est du 1^{er} décembre de l'année 434 au 31 décembre de l'année 434 et la période E s'étend du 1^{er} janvier de l'année 435 au 31 mars de l'année 435. Nous avons déjà vu que les intervalles de temps sont inégaux.

3.4.2.1 La fonction de risque à temps discret

Chaque client dispose d'une variable temps de survie dans l'état, nommée T , qui est soit censurée soit non-censurée. T est non-censurée, si l'individu a quitté l'état avant la fin de la période observée et elle est censurée s'il est toujours dans l'état à la fin de la période observée.

Le risque à temps discret, h_j , est défini comme la probabilité conditionnelle qu'un individu choisi aléatoirement expérimente l'événement cible dans la période j , étant donné qu'il ne l'a pas expérimenté dans une période antérieure :

$$h_j = P(T = j | T \geq j) \quad (2)$$

3.4.2.2 Les prédicteurs et le risque à temps discret

Nous sommes intéressés à identifier les variables explicatives qui ont un effet sur le risque de quitter.

Un ensemble de variables explicatives est associé à chaque individu. Les valeurs de ces variables peuvent être fixes ou variables dans le temps. Considérons X_1, X_2, \dots, X_p comme les variables explicatives. Pour celles qui peuvent varier dans le temps, leurs valeurs sont enregistrées à chaque période. Notons que leurs valeurs restent constantes à l'intérieur de chaque période. Soit h_{ij} la probabilité conditionnelle que le client i quitte en période j , étant donné qu'il n'a pas quitté précédemment :

$$h_{ij} = P(T_i = j | T_i \geq j, X_{1ij} = x_{1ij}, X_{2ij} = x_{2ij}, \dots, X_{pij} = x_{pij}) \quad (3)$$

Cette équation indique que le risque dépend des valeurs des prédicteurs.

3.4.2.3 Le modèle de Cox pour le risque à temps discret

Le modèle de risque à temps discret est :

$$h_{ij} = \frac{1}{1 + \exp[-(\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \dots + \alpha_J D_{Jij}) - (\beta_1 X_{1ij} + \beta_2 X_{2ij} + \dots + \beta_p X_{pij})]} \quad (4)$$

où $D_{1ij}, D_{2ij}, \dots, D_{Jij}$ sont des variables indicatrices (*binaires*) indiquant la période. Les paramètres $\alpha_1, \alpha_2, \dots, \alpha_J$ mesurent les effets des périodes tandis que les paramètres $\beta_1, \beta_2, \dots, \beta_p$ représentent les effets des prédicteurs sur le risque.

En appliquant la transformation logit au modèle (4), celui-ci devient :

$$\ln\left(\frac{h_{ij}}{1-h_{ij}}\right) = (\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \dots + \alpha_J D_{Jij}) + (\beta_1 X_{1ij} + \beta_2 X_{2ij} + \dots + \beta_p X_{p ij}) \quad (5)$$

Nous pouvons dire que le logit de la probabilité que l'événement cible se produise à la période j , étant donné qu'il ne s'est pas produit auparavant, est une combinaison linéaire de termes constants et des valeurs des prédicteurs multipliées par leurs paramètres. Le *odds ratio*

$\frac{h_{ij}}{1-h_{ij}}$ est important car il nous aidera à interpréter les modèles ajustés.

3.4.2.4 L'adaptation de la base de données pour effectuer l'analyse

Comme nous l'avons vu, deux situations possibles peuvent arriver à un client. Il peut quitter le service avant la fin de la collecte de données ou il est toujours abonné au service et il est censuré à la fin de la collecte de données. Pour identifier les deux situations, nous avons déjà créé la variable censure, C_i , qui prend la valeur 0 si l'individu i n'est pas censuré, donc s'il quitte le service avant la fin de la collecte de données,

et qui prend la valeur 1 si l'individu i n'a pas quitté le service et il reste abonné pendant toute la période observée.

Donc, chaque client représente une observation dans la base de données. Une observation, l'individu i , contient des variables comme : son numéro d'identification (id), la durée, la censure, les services pour chaque période et les prédicteurs pour chaque période. Afin de bien comprendre la base de données, en fait l'étape intermédiaire, nous présentons quelques exemples dans le tableau suivant :

Tableau 3.5 Exemples de données des clients

No d'identification du client (ID)	Durée servlocal	Censure servlocal	X ₁	X ₂₃	X ₂₄	X ₂₅	X ₂₆	X ₂₇	X ₂₈	X ₃₃	X ₃₄	X ₃₅	X ₃₆	X ₃₇	X ₃₈
1	8	1	1	0	0	0	1	1	1	1	3	3	3	4	4
2	6	0	0	1	1	1	1	0	0	3	5	5	5	0	0
3	3	0	1	1	0	0	0	0	0	5	0	0	0	0	0
5	8	0	1	1	1	1	1	1	1	2	2	2	3	3	3

Afin d'ajuster le modèle, il faut réorganiser le fichier de données. Dans le nouveau fichier de données, chaque individu aura $(j - 2)$ lignes, où j représente le nombre de périodes durant lesquelles le client est abonné au service. Nous avons déjà vu au début de ce chapitre que nous avons utilisé les événements présents pour prédire les événements futurs. En d'autres mots, nous avons employé les données des périodes A (période 1) et B (période 2) afin de modéliser la période C, la troisième période. Pour cette raison, nous avons commencé avec la troisième

période et nous avons suivi le client jusqu'à la huitième et dernière période. Donc un client quittant le service à la période 3 se verra attribuer une ligne.

Le tableau 3.6 indique comment les quatre individus du tableau 3.5 sont représentés dans le nouveau fichier d'analyse. Dans ce fichier, des variables indicatrices pour le temps, D_{ij} , sont introduites. D_{ij} prend la valeur 1 pour la ligne j et 0 pour les autres lignes.

Les variables explicatives qui sont constantes dans le temps (comme X_1) gardent la même valeur pour chaque observation. Par contre, les prédicteurs, comme X_3 , variant dans le temps prennent leurs valeurs à chaque période.

La dernière variable créée est Y qui indique si le client a quitté le service ($Y = 1$) à la période donnée.

Tableau 3.6 Données préparées pour l'analyse de survie en temps discret

No d'identification du client (ID)	D ₃	D ₄	D ₅	D ₆	D ₇	D ₈	X ₁	X ₂	X ₃	Y
1	1	0	0	0	0	0	1	0	1	0
1	0	1	0	0	0	0	1	0	3	0
1	0	0	1	0	0	0	1	0	3	0
1	0	0	0	1	0	0	1	1	3	0
1	0	0	0	0	1	0	1	1	4	0
1	0	0	0	0	0	1	1	1	4	0
2	1	0	0	0	0	0	0	1	3	0
2	0	1	0	0	0	0	0	1	5	0
2	0	0	1	0	0	0	0	1	5	0
2	0	0	0	1	0	0	0	1	5	1
3	1	0	0	0	0	0	1	1	5	1
5	1	0	0	0	0	0	1	1	2	0
5	0	1	0	0	0	0	1	1	2	0
5	0	0	1	0	0	0	1	1	2	0
5	0	0	0	1	0	0	1	1	3	0
5	0	0	0	0	1	0	1	1	3	0
5	0	0	0	0	0	1	1	1	3	1

Le fichier transformé qui est utilisé pour procéder à l'analyse est illustré ci-dessus.

3.4.3 La sélection des variables et des modèles

La sélection des variables est effectuée en fonction du but de notre étude. Dans cette recherche, nous voulons développer un modèle pour estimer les effets de certaines variables sur Y , en fait, nous désirons identifier les facteurs importants expliquant le risque de quitter. Nous allons comparer plusieurs modèles à l'aide du critère AIC (*Akaike Information Criterion*). Ce critère découle de la méthode d'estimation basée sur le maximum de vraisemblance.

Dans ce chapitre, nous avons présenté les besoins méthodologiques pour atteindre les objectifs de notre étude. En fait, le but est de développer un modèle afin d'identifier les facteurs qui influencent la décision des clients d'abandonner l'abonnement de Teleka. Nous avons exposé la préparation de données avec le nettoyage, la transformation de variables et la création de nouvelles variables, les méthodes utilisées dans l'analyse et les critères de la sélection pour un modèle. Dans le chapitre suivant l'analyse des résultats sera présentée.

CHAPITRE 4 RÉSULTATS

Ce chapitre est consacré à la présentation des résultats de recherche. Nous commençons avec une brève description des données en fonction du temps et nous poursuivons par une analyse de la courbe de survie de chaque variable dépendante (service local et service Internet). Ensuite, les modèles de survie incluant différents blocs de prédicteurs sont développés.

4.1 Le service local

Dans cette section, la variable dépendante retenue correspond au service local. Nous présentons d'abord la distribution des clients dans le tableau 4.1. Nous avons déjà vu que les individus analysés sont les clients abonnés au service local au début de la période observée pendant au moins un mois, donc la première période étudiée est la période C ou 3. Nous avons utilisé les variables explicatives correspondant aux périodes A et B pour prédire la variable cible de la période C (utiliser le présent afin de prédire le futur) et ainsi de suite.

4.1.1 Description de la variable cible : service local

Le mouvement des clients en fonction du temps est présenté dans le tableau 4.1.

Tableau 4.1 La distribution des clients en fonction du temps

Période	Clients abonnés au début de la période	Clients désabonnés à la fin de la période	Clients censurés à la fin de la période
(0)	(1)	(2)	(3)
3 - C	3726	130	0
4 - D	3596	171	0
5 - E	3425	350	0
6 - F	3075	462	0
7 - G	2613	310	0
8 - H	2303	253	2050

La colonne (1) représente les clients abonnés au service local au début de chaque période. La colonne (2) contient les clients qui ont quitté à la fin de chaque période. La colonne (3) inclut les clients censurés. À la fin de la période observée, 1675 clients (45 % de l'échantillon) ont quitté le service local et 2050 (55 %) clients y sont encore abonnés.

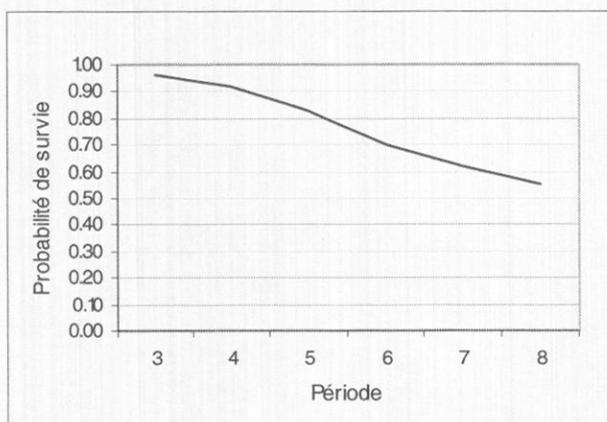
4.1.2 La fonction de survie et la fonction de risque du service local

La probabilité de survie et les risques sont fournis dans le tableau 4.2.

Tableau 4.2 La fonction de survie et de risque pour tous les clients

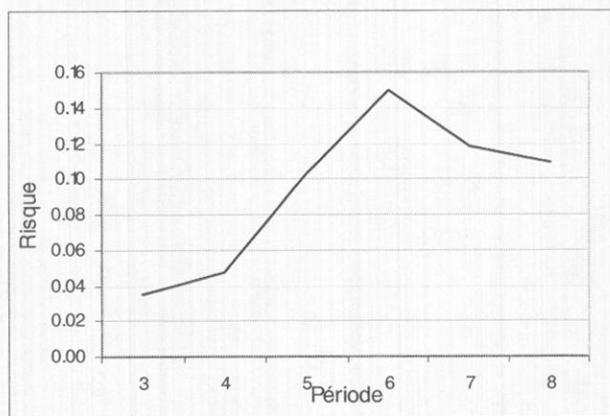
Période (0)	Probabilité de survie jusqu'à la fin de la période (1)	Risque de quitter pendant la période (2)
3 - C	0,9651	0,0349
4 - D	0,9192	0,0476
5 - E	0,8253	0,1022
6 - F	0,7013	0,1502
7 - G	0,6181	0,1186
8 - H	0,5502	0,1099

La probabilité de survie est la proportion d'individus étant toujours abonnés au service à la fin de la période. En examinant la fonction de survie (tableau 4.2, figure 4.1), nous observons que 96,51 % des clients sont abonnés au service local plus de deux mois, 91,92 % plus de trois mois, 82,53 % plus de six mois, 70,13 % plus de douze mois, 61,81 % plus de dix-huit mois et 55,02 % plus de deux ans.

Figure 4.1 La proportion de clients abonnés au service local à la fin de chaque période

La fonction de risque représente la probabilité que l'individu quitte le service au cours d'une période, étant donné qu'il était abonné au début de la période. À la figure 4.2, nous avons tracé la fonction de risque pour notre échantillon.

Figure 4.2 La proportion de clients ayant quitté le service local à chaque période parmi ceux ayant débuté la période



Nous observons que 3,49 % des 3726 clients ayant débuté un deuxième mois quittent le service local avant la fin des deux premiers mois, 4,76 % des 3596 clients abonnés au début du troisième mois quittent le service avant la fin du mois. Après trois mois d'abonnement et jusqu'au douzième mois, nous remarquons une croissance rapide du risque. En effet, à la fin de la première année (période 6), il atteint un maximum de 15,02 % pour ensuite atteindre 10,99 % à la fin de la deuxième année (période 8).

4.1.3 Les modèles de survie développés

Nous avons développé des modèles de régression selon plusieurs groupes de covariables présentés à la section 3.3.

Premièrement, nous utilisons les covariables de la période précédente pour prédire la variable cible de la période actuelle.

Deuxièmement, nous avons utilisé un autre groupe de covariables. Ce groupe contient des variables calculées comme la différence entre une variable en période précédente et une variable en période présente afin de prédire la variable cible en période suivante. D'une manière générale, les nouvelles variables créées sont du type $(X_{t-2} - X_{t-1}) = X_t^*$.

Troisièmement, le dernier groupe utilisé contient des covariables du

$$\text{type } X_t^{**} = \begin{cases} 1, & \text{si } (X_{t-2} - X_{t-1}) > 0 \\ 0, & \text{si } (X_{t-2} - X_{t-1}) \leq 0 \end{cases}$$

Pour un groupe de prédicteurs donnés, le meilleur modèle, c'est-à-dire le meilleur sous-ensemble de covariables, est choisi à l'aide du critère AIC parmi tous les sous-ensembles possibles. Un fois le modèle choisi, l'effet de chaque covariable est discuté.

4.1.3.1 Le meilleur modèle contenant le premier bloc de covariables

Le tableau 4.3 contient le meilleur modèle avec le premier groupe de covariables.

Tableau 4.3 Le modèle incluant le premier groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-3,2298	0,0955	<0,0001	0,040
d4	-2,9521	0,0835	<0,0001	0,052
d5	-2,1829	0,0592	<0,0001	0,113
d6	-1,7220	0,0535	<0,0001	0,179
d7	-1,9817	0,0634	<0,0001	0,138
d8	-2,0583	0,0696	<0,0001	0,128
nontlkn	0,4691	0,1498	0,0017	1,599
revn3n	-0,0073	0,0019	0,0002	0,993
revn4n	0,0003	0,0002	0,0820	1,000
revr1n	0,0008	0,0002	0,0003	1,001
revr9n	0,0340	0,0070	<0,0001	1,035
revr10n	0,0148	0,0036	<0,0001	1,015

AIC=10798; -2LL=10774; Chi-carré=12801 (signif., p-value<0,0001)

En plus des six variables indicatrices des périodes (d3 à d8), ce modèle contient six covariables qui sont :

- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,
- « revn3n » = les taxes sur la facture,

- « revn4n » = les revenus totaux sur la facture,
- « revr1n » = les revenus pendant les heures de bureau,
- « revr9n » = les revenus par rapport aux appels faits au numéro 411,
- « revr10n » = les revenus par rapport aux appels faits aux numéros 1-900.

Dans ce modèle (tableau 4.3), tous les prédicteurs sont fortement significatifs (p -value $< 0,001$), sauf la variable « revn4n » (p -value = 0,082). Le coefficient positif du prédicteur « nontlkn » (0,47) indique un effet positif. En fait, à chaque période, le risque de quitter le service local est 60 % plus élevé pour les clients qui utilisent le service d'appels interurbains d'autres compagnies de télécommunications par rapport aux autres clients. S'il y a une augmentation de 1 \$ des revenus par rapport aux appels faits au numéro 411, le risque de quitter le service local de Teleka augmente de 3,5 %.

4.1.3.2 Le meilleur modèle contenant le deuxième bloc de covariables

Le meilleur modèle, avec le deuxième groupe de covariables, toujours selon le critère AIC, est présenté au tableau 4.4. Le meilleur modèle inclut huit prédicteurs :

- « dsumvar » = la différence du nombre de services auxquels le client est abonné entre deux périodes consécutives,

- « dnfc » = la différence du nombre de factures par le client entre deux périodes consécutives,
- « dntf » = la différence du nombre de numéros de téléphone par facture entre deux périodes consécutives,
- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,
- « drevn3 » = la différence des taxes sur la facture entre deux périodes consécutives,
- « drevn4n » = la différence des revenus totaux sur la facture entre deux périodes consécutives,
- « drevr3n » = la différence des revenus par carte d'appels entre deux périodes consécutives,
- « drevr10n » = la différence des revenus par rapport aux appels faits aux numéros 1-900 entre deux périodes consécutives.

Tableau 4.4 Le modèle incluant le deuxième groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-3,3423	0,0911	<0,0001	0,035
d4	-3,0212	0,079	<0,0001	0,049
d5	-2,2015	0,0576	<0,0001	0,111
d6	-1,7434	0,0509	<0,0001	0,175
d7	-2,003	0,0612	<0,0001	0,135
d8	-2,0918	0,0674	<0,0001	0,123
dsumvar	0,2023	0,0597	0,0007	1,224
nontlkn	0,4615	0,149	0,0020	1,586
dnfc	-0,1979	0,1304	0,1291	0,820
dntf	0,0687	0,0323	0,0332	1,071
drevn3	0,0047	0,0018	0,0081	1,005
drevn4	-0,0004	0,0002	0,0385	1,000
drevr3	-0,0028	0,0017	0,0901	0,997
drevr10	-0,0126	0,0042	0,0024	0,987

AIC=10841; -2LL=10813; Chi-carré=12779 (signif., p-value<0,0001)

Dans ce modèle, toutes les variables explicatives ont un apport significatif (p-value < 0,05), sauf les variables « drevr3 » (p-value = 0,09) et « dnfc » (p-value = 0,129). On peut noter que si la différence des services utilisés par le client augmente d'une unité (le nombre de services par le client diminue), le risque de quitter le service local augmente de 22,4 % ou est multiplié par 1,224. Le risque de quitter le service local est plus élevé (59 %) pour les clients qui utilisent le service d'appels interurbains d'autres compagnies de télécommunications. Le

risque de quitter s'accroît de 7 % si la différence des numéros de téléphone par facture augmente d'une unité (le nombre de numéros diminue).

4.1.3.3 Le meilleur modèle contenant le troisième bloc de covariables

Le meilleur modèle avec le troisième groupe de covariables est présenté au tableau 4.5. Le meilleur modèle contient onze prédicteurs :

- « ddsuvar » = indique s'il y a une baisse du nombre de services auxquels le client est abonné entre deux périodes consécutives,
- « ddntf » = indique s'il y a une baisse du nombre de numéros de téléphone par facture entre deux périodes consécutives,
- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,
- « ddminm1 » = indique s'il y a une baisse des minutes utilisées pendant les heures de bureau entre deux périodes consécutives,
- « ddrevr2 » = indique s'il y a une baisse des revenus liés aux appels qui sont traités par une opératrice entre deux périodes consécutives,
- « ddrevr5 » = indique s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-800 entre deux périodes consécutives,
- « ddrevr9 » = indique s'il y a une baisse des revenus par rapport aux appels faits aux numéros 411 entre deux périodes consécutives,

- « ddrevr10 » = indique s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-900 entre deux périodes consécutives.

Tableau 4.5 Le modèle incluant le troisième groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-3.4449	0.0909	<.0001	0.032
d4	-3.1968	0.0831	<.0001	0.041
d5	-2.2936	0.0590	<.0001	0.101
d6	-1.9965	0.0608	<.0001	0.136
d7	-2.3638	0.0715	<.0001	0.094
d8	-2.4193	0.0757	<.0001	0.089
ddsumvar	0.5847	0.1630	0.0003	1.794
nontlkn	0.4281	0.1503	0.0044	1.534
ddntf	1.1823	0.1224	<.0001	3.262
ddminm1	0.2114	0.0557	0.0001	1.235
ddrevr2	0.1941	0.0982	0.0480	1.214
ddrevr5	0.4017	0.1103	0.0003	1.494
ddrevr9	0.1444	0.0581	0.0129	1.155
ddrevr10	0.5282	0.1443	0.0003	1.696

AIC=10716; -2LL=10688; Chi-carré=12842 (signif., p-value<0,0001)

Dans ce modèle (tableau 4.5), toutes les variables explicatives ont un apport significatif (p-value < 0,05). Notamment, le risque de quitter le service local augmente de 79 % s'il y a une baisse du nombre de service utilisé par le client et de 226 % s'il y a une baisse du nombre de

numéros de téléphone par facture. Le risque de quitter le service local pour les clients qui n'utilisent pas le service d'appels interurbains de Teleka est 53 % plus élevé que pour les autres. Aussi, le risque de quitter le service local augmente de 23 % s'il y a une baisse des minutes utilisées par le client pendant les heures de bureau et de 15 % et plus s'il y a une baisse des revenus par rapport aux différents appels faits par le client.

4.2 Le service Internet

La variable dépendante correspondant au service Internet est analysée dans cette section. Nous présentons d'abord la distribution des clients dans le tableau 4.6. Les individus analysés sont les clients abonnés au service Internet au début de la période observée pendant au moins un mois, donc la première période étudiée est la période C ou 3. Tout comme dans le cas du service local, nous avons utilisé les variables explicatives correspondant aux périodes A et B pour prédire la variable cible de la période C et ainsi de suite.

4.2.1 Description de la variable cible : service Internet

Une présentation de la distribution des clients pour chaque période est fournie dans le tableau 4.6.

Tableau 4.6 La distribution des clients en fonction du temps

Période	Clients abonnés au début de la période	Clients désabonnés à la fin de la période	Clients censurés à la fin de la période
(0)	(1)	(2)	(3)
3 - C	712	45	0
4 - D	667	36	0
5 - E	631	90	0
6 - F	541	92	0
7 - G	449	65	0
8 - H	384	51	333

La colonne (1) représente les clients abonnés au service Internet au début de chaque période. Nous avons déjà présenté les colonnes (2) et (3) du tableau 4.6 dans la section 4.1.1. À la fin de la période observée, 379 clients (53 % de l'échantillon) ont quitté le service Internet et 333 (47 %) clients y sont encore abonnés.

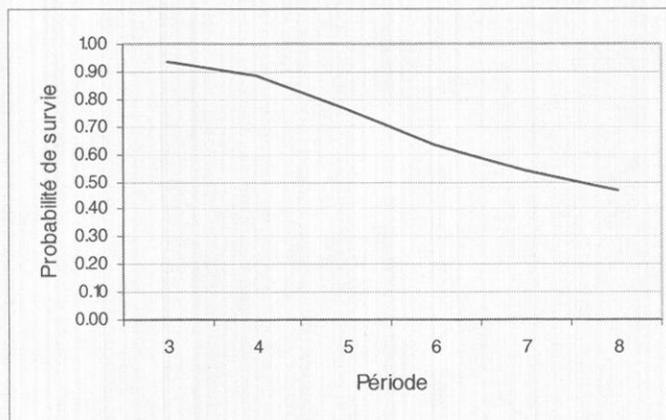
4.2.2 La fonction de survie et la fonction de risque du service Internet

La probabilité de survie et les risques sont présentés dans le tableau 4.7.

Tableau 4.7 La fonction de survie et de risque pour tous les clients

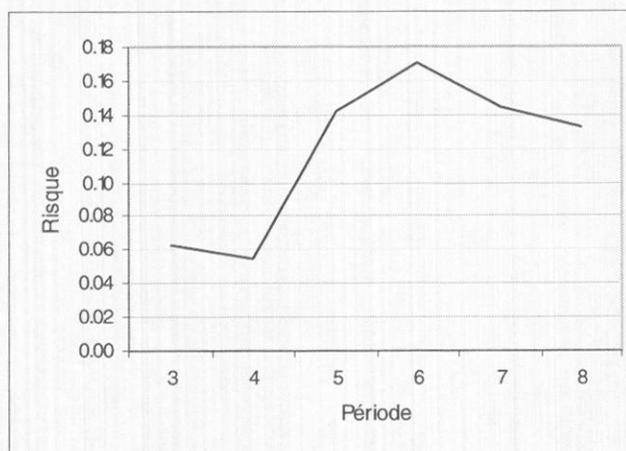
Période (0)	Probabilité de survie jusqu'à la fin de la période (1)	Risque de quitter pendant la période (2)
3 - C	0,9368	0,0632
4 - D	0,8862	0,0540
5 - E	0,7598	0,1426
6 - F	0,6306	0,1701
7 - G	0,5393	0,1448
8 - H	0,4677	0,1328

Comme nous l'avons déjà mentionné dans la section 4.1.2, la probabilité de survie est la proportion d'individus étant toujours abonnés au service à la fin de la période. En examinant la fonction de survie (tableau 4.7, figure 4.3), nous observons que 93,68 % des clients sont abonnés au service Internet plus de deux mois, 88,62 % plus de trois mois, 75,98 % plus de six mois, 63,06 % plus de douze mois, 53,93 % plus de dix-huit mois et 46,77 % plus de deux ans.

Figure 4.3 La proportion de clients abonnés au service Internet à la fin de chaque période

La fonction de risque représente la probabilité que l'individu quitte le service au cours d'une période, étant donné qu'il y était abonné au début de la période. À la figure 4.4, nous avons tracé la fonction de risque pour notre échantillon.

Figure 4.4 La proportion de clients ayant quitté le service Internet chaque période parmi ceux ayant débuté la période



Nous observons que 6,32 % des 712 clients ayant débuté un deuxième mois quittent le service Internet avant la fin des deux premiers mois. Pendant la deuxième période, il y a une diminution du risque et seulement 5,40 % des 667 clients abonnés au début du troisième mois quittent le service avant la fin du mois. Après trois mois d'abonnement et jusqu'au douzième mois, nous apercevons une croissance rapide du risque. En effet, à la fin de la première année (période 6), il atteint un maximum de 17,01 % et à la fin de la deuxième année (période 8), il baisse jusqu'à 13,28 %.

4.2.3 Les modèles de survie développés

Nous avons développé des modèles de régression selon plusieurs groupes de covariables présentées aux sections 3.3. et 4.1.3. Pour un groupe de prédicteurs donnés, le meilleur modèle, c'est-à-dire le meilleur sous-ensemble de covariables, est choisi selon le critère AIC, parmi tous les sous-ensembles possibles. Le meilleur modèle incluant chaque bloc de prédicteurs est présenté.

4.2.3.1 Le meilleur modèle contenant le premier bloc de covariables

Le meilleur modèle incluant le premier bloc de prédicteurs est fourni au tableau 4.8.

Tableau 4.8 Le modèle incluant le premier groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-2,0937	0,2005	<.0001	0,123
d4	-2,2484	0,2153	<.0001	0,106
d5	-1,1549	0,1766	<.0001	0,315
d6	-0,9052	0,1856	<.0001	0,404
d7	-1,0909	0,1987	<.0001	0,336
d8	-1,1855	0,2117	<.0001	0,306
sumvarn	-0,2072	0,0473	<.0001	0,813
nontlkn	1,1551	0,3464	0,0009	3,174
revn1n	0,0012	0,0007	0,0828	1,001

AIC=2288; -2LL=2270; Chi-carré=2078 (signif., p-value<0,0001)

Ce modèle contient trois covariables :

- « sumvarn » = le nombre de services utilisés par le client,
- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,
- « revn1n » = les revenus mensuels.

Dans ce modèle (tableau 4.8), l'effet de la variable « sumvarn » et de la variable « nontlkn » sont significatifs (p -value $< 0,001$). Les revenus mensuels « revn1n » n'ont pas d'effet significatif (p -value = 0,082). Les clients qui utilisent le service d'appels interurbains d'autres compagnies (« nontlkn ») présentent un risque plus élevé (217 %) de quitter par rapport aux autres clients. Le coefficient négatif de la variable « sumvarn » indique une diminution du risque de quitter le service Internet. En d'autres termes, pour chaque baisse de 1 (un) du nombre de services utilisés par le client, le risque de quitter le service Internet augmente de 23 %.

4.2.3.2 Le meilleur modèle contenant le deuxième bloc de covariables

Le tableau 4.9 contient le meilleur modèle, avec le deuxième groupe de covariables. Le meilleur modèle inclut les cinq prédicteurs suivants :

- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,

- « dntf » = la différence du nombre de numéros de téléphone par facture entre deux périodes consécutives,
- « drevn2 » = la différence des revenus non récurrents entre deux périodes consécutives,
- « drevn4 » = la différence des revenus totaux sur la facture entre deux périodes consécutives,
- « drevr3 » = la différence des revenus par carte d'appels entre deux périodes consécutives.

Tableau 4.9 Le modèle incluant le deuxième groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-2,5811	0,1582	<.0001	0,0760
d4	-2,8216	0,1729	<.0001	0,0600
d5	-1,7178	0,1188	<.0001	0,1790
d6	-1,6323	0,1171	<.0001	0,1950
d7	-1,8336	0,1367	<.0001	0,1600
d8	-1,8948	0,1526	<.0001	0,1500
nontlkn	0,9609	0,3454	0,0054	2,6140
dntf	0,2830	0,0879	0,0013	1,3270
drevn2	-0,0045	0,0010	<.0001	0,9950
drevn4	0,0034	0,0009	<.0001	1,0030
drevr3	-0,0064	0,0020	0,0015	0,9940

AIC=2273; -2LL=2251; Chi-carré=2082 (signif., p-value<0,0001)

Dans ce modèle (tableau 4.9), toutes les variables explicatives ont un apport significatif (p-value < 0,01). Notamment, le risque de quitter le

service Internet est plus élevé (161 %) pour les clients qui utilisent le service d'appels interurbains d'autres compagnies de télécommunications. Le risque de quitter s'accroît de 32,7 % si la différence du nombre de numéros de téléphone par facture augmente d'une unité (si le nombre de numéros baisse).

4.2.3.3 Le meilleur modèle contenant le troisième bloc de covariables

Le meilleur modèle choisi inclut dix prédicteurs (tableau 4.10) :

- « ddsuvar » = indique s'il y a une baisse du nombre de services entre deux périodes consécutives,
- « ddntf » = indique s'il y a une baisse du nombre de numéros de téléphone par facture entre deux périodes consécutives,
- « nontlkn » = indique si le client utilise le service d'appels interurbains d'autres compagnies de télécommunications,
- « ddminm1 » = indique s'il y a une baisse des minutes utilisées pendant les heures de bureau entre deux périodes consécutives,
- « ddrevn2 » = indique s'il y a une baisse des revenus non récurrents entre deux périodes consécutives,
- « ddrevn4 » = indique s'il y a une baisse des revenus totaux sur la facture entre deux périodes consécutives,
- « ddrevr2 » = indique s'il y a une baisse des revenus liés aux appels qui sont traités par une opératrice entre deux périodes consécutives,

- « ddrevr3 » = indique s'il y a une baisse des revenus par carte d'appels entre deux périodes consécutives,
- « ddrevr5 » = indique s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-800 entre deux périodes consécutives,
- « ddrevr10 » = indique s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-900 entre deux périodes consécutives.

Tableau 4.10 Le modèle incluant le troisième groupe de covariables

Covariables (x)	Coefficients (β)	Écart-type	p-value	Ratio de risque ($\exp(\beta)$)
d3	-2.3886	0.1900	<.0001	0.092
d4	-2.6036	0.2057	<.0001	0.074
d5	-1.4382	0.1628	<.0001	0.237
d6	-1.6567	0.1391	<.0001	0.191
d7	-2.0363	0.1619	<.0001	0.131
d8	-1.9824	0.1717	<.0001	0.138
ddsumvar	0.7771	0.3803	0.0410	2.175
nontlkn	0.9261	0.3559	0.0093	2.525
ddntf	1.6006	0.2689	<.0001	4.956
ddminm1	0.3046	0.1267	0.0162	1.356
ddrevn2	-0.2690	0.1650	0.1030	0.764
ddrevn4	-0.3252	0.1650	0.0488	0.722
ddrevr2	0.3446	0.2300	0.1341	1.411
ddrevr3	-0.6677	0.3629	0.0658	0.513
ddrevr5	0.4180	0.1999	0.0365	1.519
ddrevr10	0.9612	0.2882	0.0009	2.615

AIC=2246; -2LL=2214; Chi-carré=2109 (signif., p-value<0,0001)

Dans ce modèle (tableau 4.10), toutes les variables explicatives ont un apport significatif ($p\text{-value} < 0,05$) excepté les variables « ddrevn2 » ($p\text{-value} = 0,103$), « drevr2 » ($p\text{-value} = 0,134$) et « drevr3 » ($p\text{-value} = 0,065$). On observe que le risque de quitter le service Internet augmente de 117 % s'il y a une baisse du nombre de services utilisés par le client et de 395 % s'il y a une baisse du nombre de numéros de téléphone par facture. Le risque de quitter le service Internet pour les clients qui n'utilisent pas le service d'appels interurbains de Teleka est 152 % plus élevé par rapport aux autres clients. Aussi, le risque de quitter le service Internet augmente de 35 % s'il y a une baisse des minutes utilisées par le client pendant les heures de bureau, de 51 % s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-800 et de 161 % s'il y a une baisse des revenus par rapport aux appels faits aux numéros 1-900. De plus, le risque de quitter diminue de 39 % s'il y a une augmentation des revenus totaux sur la facture.

Dans ce chapitre, nous avons présenté les modèles prédictifs développés ayant pour but de connaître les facteurs qui ont un effet significatif sur les variables cibles : le service local et le service Internet. Les trois blocs de prédicteurs créés précédemment nous ont servi à développer et interpréter les modèles.

Nous pouvons conclure que le risque de quitter le fournisseur est le plus élevé à la fin de la première année avec 15,02 % pour le service local et

17,01 % pour le service Internet. Il est principalement influencé par le nombre de services utilisés par le client, le nombre de numéros de téléphone par facture ainsi que le facteur indiquant si le client utilise le service interurbain d'un autre fournisseur. De plus, des variables comme les minutes utilisées et les différents types de revenus jouent un rôle non négligeable sur le risque de quitter le fournisseur.

Le risque augmente si le nombre de services par facture baisse ainsi que si le nombre de numéros de téléphone par facture diminue. La relation client – fournisseur est donc moins stable. Dans ce cas, nous pouvons imaginer que le client n'est pas satisfait du service ou bien qu'il cherche à diminuer ses dépenses.

Également, le risque de quitter est plus élevé pour les clients utilisant le service interurbain d'un autre fournisseur que pour les autres clients. Par exemple, un client qui est abonné au service local fourni par Teleka et qui utilise le service interurbain d'un autre fournisseur est plus susceptible de quitter Teleka. De plus, le fait qu'un client utilise le service interurbain d'un autre fournisseur a un effet plus significatif sur la décision de quitter le service Internet fourni par Teleka.

Dans la section suivante, nous allons exposer la conclusion générale, les recommandations aux gestionnaires, les limites de notre recherche et les avenues de recherche.

CHAPITRE 5 CONCLUSION GÉNÉRALE

Ce chapitre porte sur les conclusions qui découlent des résultats de l'étude. Ensuite, nous présentons les principales limites de la recherche suivies d'avenues pour des études futures.

5.1 Conclusion et discussion

L'objectif principal visé par cette étude a été de développer un modèle prédictif de la durée de la relation client – fournisseur dans le contexte d'une compagnie de télécommunications et à identifier les facteurs susceptibles d'influencer cette relation. Un deuxième objectif a été de fournir aux gestionnaires des recommandations afin de les aider à mieux gérer la relation avec leur clientèle. Voici donc un bref rappel de ces deux objectifs respectivement d'ordre théorique et pratique.

Sur le plan théorique, nous croyons que le travail que nous avons réalisé a permis de faire une synthèse de la littérature sur la fidélisation de la clientèle, sur la gestion de la relation entre le client et le fournisseur ainsi que sur les méthodes statistiques appliquées dans le domaine d'affaires. Dans la littérature, Kamakura, Ramaswami et Srivastava (1991) ont affirmé que la vente croisée peut être considérée comme une stratégie qui diminue la perte de clients tout en augmentant le nombre de services utilisés par le client. Selon notre recherche, nous

pouvons ajouter que le nombre de services utilisés par un client est un facteur ayant un effet considérable sur la durée de la relation client – fournisseur. En effet, le risque de quitter le fournisseur est plus élevé s'il y a une baisse des services utilisés par le client. En d'autres termes, la vente croisée est une stratégie qui réduit l'attrition de la clientèle et qui, de plus, renforce et soutient la relation client – fournisseur.

Ainsi, nous trouvons que les méthodes statistiques traditionnelles et l'évolution de la technologie ont un apport substantiel dans la gestion de la relation avec le client. Les grandes bases de données gérées par l'entreprise nous ont aidés à mieux connaître le comportement du client dans le temps et d'analyser le cycle de vie du client dans une entreprise.

Aussi, nous avons développé un modèle prédictif pour expliquer la relation entre le client et le fournisseur ainsi que l'effet des facteurs sur la durée de cette relation. Notre modèle est appliqué sur une base de données qui provient d'une compagnie de télécommunications. Nous avons été intéressés par la longévité des clients qui utilisaient un ou plusieurs services offerts par l'entreprise.

Dans cette optique, si le nombre de services augmente, le risque de quitter l'entreprise diminue. La relation entre le client et le fournisseur se veut donc plus stable.

D'un autre côté, si le client diminue le nombre de numéros de téléphone par facture, le risque de quitter augmente. Dans ce cas, ce qui joue un rôle important sur le risque de quitter est, en fait, la taille de la PME, sa viabilité ou bien sa satisfaction à l'égard du service offert.

Les clients utilisant le service interurbain d'un autre fournisseur sont plus susceptibles de quitter le présent fournisseur pour la concurrence. De plus, le risque de quitter le service Internet est plus élevé que le risque de quitter le service local.

Sur le plan pratique, le modèle que nous avons développé dans cette étude constitue un outil stratégique pour le gestionnaire. Dans le contexte actuel, il est difficile d'identifier, de sélectionner et de segmenter les clients profitables. Li (1995) considère que l'analyse de survie est un outil très important pour les entreprises afin de mieux comprendre les raisons pour lesquelles les clients quittent en faveur de la concurrence. L'effet de prédicteurs sur le risque de quitter est facile à interpréter. Le gestionnaire peut donc élaborer des profils pour des clients hautement susceptibles de partir et pour ceux qui sont peu disposés à quitter le fournisseur. Dans ces conditions, il peut choisir la meilleure stratégie pour chaque groupe de clients afin de fidéliser ceux à fort potentiel et les garder longtemps.

Une stratégie de fidélisation serait de promouvoir le plus grand nombre de services pour un client en lui offrant des forfaits, des promotions et des gratuités pouvant le satisfaire selon l'historique de ses achats.

5.2 Les principales limites

Les principales limites de notre étude portent sur la base de données. Premièrement, les variables contenues dans la base de données décrivent le comportement du client (historique d'achat, nombre de services utilisés, etc.). Une analyse de survie demande aussi, des variables démographiques, des variables qui expliquent l'état financier du client (PME) ainsi que des variables qui décrivent des facteurs incontrôlables (l'activité de la compétition, la communication de masse, etc.).

Deuxièmement, une autre limite porte sur le nombre de périodes analysées. Dans notre étude, nous avons analysé seulement huit intervalles de temps. Il est souhaitable d'avoir plus de périodes afin de mieux modeler le processus et d'avoir des résultats plus pertinents.

Troisièmement, une autre limite concerne la longueur des intervalles de temps. Selon la littérature (Li, 1995), dans l'analyse de survie, il est préférable d'utiliser des intervalles égaux. Les données dont nous disposons sont observées à des intervalles de temps différents, ce qui résulte en des intervalles inégaux (tableau 3.2).

En conclusion, il est important de mentionner que les limites de cette étude peuvent influencer les résultats obtenus. En conséquence, de nouveaux modèles peuvent être développés en considérant les limites présentées précédemment.

5.3 Les avenues de recherche

Une autre voie de recherche serait d'inclure de nouveaux prédicteurs, par exemple, la satisfaction du client (Rust et Zahoric, 1993) et la longueur des intervalles de temps (Allison, 2003).

Nous pourrions également effectuer la validation du modèle qui permet d'estimer le risque pour chaque client et de trier le risque en ordre décroissant. Ce tri nous permettrait d'estimer si le modèle est viable.

Enfin, une segmentation des clients pourrait être faite afin de cibler chaque groupe de clients de façon spécifique pour appliquer des stratégies appropriées à chacun d'entre eux.

BIBLIOGRAPHIE

Allison, P.D. (1984). *Event history analysis : Regression for longitudinal event data*, Sage University Paper Series on Quantitative Applications in Social Sciences, Series No.07-046, Sage, Beverly Hills.

Allison, P.D. (2003). *Survival Analysis Using SAS : A practical guide*, SAS Institute Inc., Cary, NC.

Appiah-adu, K. (1999). *Marketing effectiveness and customer retention in the service sector*, The Service Industries Journal, Vol. 19, No. 3, Juillet, 26-41.

Barksdale, H.J.Jr., Johnson, J.T. et Suh, M. (1997). *A relationship maintenance model: A comparison between managed health care and traditional fee-for-service*, Journal of Business Research, No. 40, 237-247.

Barlow, R. (1992). *Relationship Marketing – The Ultimate in Customer Services*, Retail Control, Mars., 29-37.

Berry, L.L. (1980). *Services Marketing is Different*, Business, No. 30, Mai, 24-29.

Berry, L.L. (1995). *Relationship Marketing of Services-Growing Interest, Emerging Perspectives*, Journal of the Academy of Marketing Science, Vol. 23, No. 4, 236-245.

Berry, M.J.A. et Linoff, G.S. (1997). *Data Mining Techniques for Marketing, Sales and Customer Support*, John Wiley & Sons, Inc.

Berry, M. J.A. et Linoff, G. S. (2000). *Mastering « Data Mining » – The Art and Science of Customer Relationship Management*, John Wiley & Sons, Inc., chap.1, 4, 8.

Bolton, R. et Drew, J. (1994). *Linking Customer Satisfaction to Services Operations and Outcomes*, Service Quality: New Directions in Theory and Practice, Roland T. Rust and Richard L. Oliver eds.

Cohn, T (2005). *The Value of Customer Retention: A Business without a Customer Retention Plan Will Lose Revenue and Market Share*
<http://www.marketingprinciples.com/customerretention/default.asp?cat=123>

Cox, D.R. (1972). *Regression models and life tables*, Journal of Royal Statistical Society, Series B, No. 34, 187-202.

Cox, D.R. et Oakes, D. (1984). *Analysis of survival data*, Chapman & Hall, London.

Coviello, N.E. et Brodie, R.J. (2001). *Contemporary Marketing Practices of Consumer and Business-to-Business Firms: How Different are They?*, The Journal of Business & Industrial Marketing, Santa Barbara, Vol. 16, No. 5, 382.

Coviello, N.E., Brodie, R.J., Danher, P.J. et Johnston, W.J. (2002). *How Firms Relate to their Markets: An Empirical Examination of Contemporary Marketing Practices*, Journal of Marketing, Chicago, Vol. 66, No. 3, 33-46.

Crosby, L.A. et Johnson, S.L. (2000). *Customer relationship management: Is technology the new Holy Grail for customer loyalty?*, Marketing Management, fall, 4-5.

- Dawkins, J. et Reichheld, F.F. (1990). *Customer Retention as a Competitive Weapon*, Directors and Boards, 14, 4 Summer. 42-47
- Day, G.S. (2000). *Managing Market Relationships*, Journal of the Academy of Marketing Science, Vol. 28, No. 1, 24-30.
- Deshpande, R., Farley, J., Webster, F. (1993). *Corporate culture, customer orientation, and innovativeness*, Journal of Marketing; Jan, Vol. 57, No. 1, 23-38.
- Dwyer, R., Schurr, P. et Oh, S. (1987). *Developing Buyer Seller Relationships*, Journal of Marketing, Vol. 52, 21-34.
- Garbarino, E. et Johnson, M.S. (1999). *The Different Roles Of Satisfaction, Trust And Commitment In Customer Relationships*, Journal of Marketing, Vol. 63 (April), 70-87
- Gremler, D.D. et Brown, S.W. (1996). *Service loyalty: its nature, importance, and implications*, in Edwardson, B., Brown, S.W., Johnston, R et al., *Advancing Service Quality: A Global Perspective*, International Service Quality Association, 171-180.
- Halley, E. (1693). *An estimate of the degrees of the mortality of mankind, drawn from curious tables of the birth and funerals of the city of Breslav, with an attempt to ascertain the price of annuities upon lives*, Philosophical transactions of the Royal Society of London, 17, 596-610.
- Harrison, T. et Ansell, J. (2002). *Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities*, Journal of Financial services Marketing, Mar 2002, Vol. 6, No. 3, 229.
- Jacoby, J. et Chestnut, R.W. (1978). *Brand Loyalty : Measurement and Management*, J. Wiley and Sons, NY.

Jackson, B.B. (1985). *Build Relationship that Last*, Harvard Business Review, Nov-Dec, 120-128.

Joo, Y.H., Jun, J.K. et Kim, B.D. (2002). *Encouraging customers to pay less for mobile telecommunication services*. Journal of Database Management, Vol. 9, No. 4, 350-359.

Jones, T.O. et Sasser, W.E. (1995). *Why Satisfied Customer Defect?*, Harvard Business Review, Nov/Dec, 89-99.

Kamakura, W.A., Ramaswamy, S.N. et Srivastava, R.K. (1991). *Applying latent trait analysis in the evaluation of prospects for cross-selling in the financial services*, International Journal for Research in Marketing, Vol. 8, 329-349.

Kotler, P. (1992). *Marketing's New Paradigm: What's Really Happening Out There*, Planning Review, Vol. 20, No. 5, 50-52.

Li, S. (1995). *Survival Analysis: Just as in life, customers are 'born' and they 'die'. Here's how to prevent their premature demise*, Marketing Research, Fall 1995/Winter 1996, Vol. 7, No. 4, 16.

Meyer-Waarden, Lars (2002). *Les sources d'efficacité des programmes de fidélisation – Une étude empirique sur la base d'un panel single source*, thèse de doctorat, Université de Pau et des Pays de l'Adour

Meltzer, M. *Customer Retention and Acquisition for Telecommunications*, e-Solutions In The Age Of 'e', AMT Consulting.

Morita, J.G., Lee, T.W. et Mowday, R.T. (1993). *The regression - analog to survival analysis - A selected application to turnover research*, Academy of Management Journal, Vol. 36, No. 6, 1430-1464.

Morgan, R. et Hunt, S. (1994). *The commitment-trust theory of relationship marketing*, Journal of Marketing, Vol. 58, No. 3, 20-38

Murray, K.B. (1991). *A test of services marketing theory: consumer information acquisition activities*, Journal of Marketing, Vol. 55, January, 10-25.

Parasuraman, A., Zeithaml, V.A. et Berry, L.L. (1985). *A conceptual model of service quality and its implications for future research*, Journal of Marketing, Vol. 49, Fall, 41-50.

Peppers, D. et Rogers, M.(1993). *The One to One Future: Building Relationships On Customers at a Time*, New York: Currency/Doubleday

Peppers, D. et Rogers, M. (1995). *A New Marketing Paradigm: Share of Customer not Market Share*, Managing Service Quality, Bedford, Vol. 5, No. 3, 48-51.

Reichheld, F., et Sasser, W.E. (1990, September-October). *Zero defections: Quality comes to services*, Harvard Business Review, Vol. 68, 105-111.

Reichheld, F. (1996). *The loyalty effect, the hidden force behind growth, profits and lasting value*, Harvard Business School Press.

Ricard, L. et Perrier, J. (1999). *Explaining and Evaluating the Implementation of Organizational Relationship Marketing in the Banking Industry: Clients' Perception*, Journal of Business Research, Vol. 45, 199-209.

Rosenberg, L. et Czepiel, J. (1984). *A Marketing Approach to Consumer Retention*, Journal of Consumer Marketing, Vol. 1, Spring, 45-51.

Rusbult, C. E. (1980). *Commitment and Satisfaction in Romantic Associations: A test of the Investment Model*, Journal of Experimental social Psychology, Vol. 16, 172-186 et *Satisfaction and Commitment in Friendships*, Representative Research in Social psychology, Vol. 11, 96-105.

Rust, R. et Zahoric, A. (1993). *Customer satisfaction, customer retention and market share*, Journal of Retailing, Vol. 69, No. 2

Schellinck, T. (2002). *Voluntary Churn Analysis*, Term Project For BUSI 6420, Meloche Monnex

Schmittlein, D.C. et Peterson, R.A. (1994). *Customer Base Analysis : An Industrial Purchase Process Application*, Marketing Science, Vol. 13, No. 1, 41-67

Singer, J.D. et Willett, J.B. (1993). *It's about time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events*, Journal of Educational Statistics, Vol. 18, No. 2, 155-195.

Surprenant, C.F. et Solomon, M.R. (1987). *Predictability and personalization in the service encounter*, Journal of Marketing, Vol. 51, April, 86-96.

Suh, M. (1994). *An examination of the client-professional service provider relationship maintenance from the clients' perspective*, Unpublished Dissertation, Georgia State University.

Tuma, N. B. et Hannan, M. T. (1984). *Social dynamics: Models and methods*, Academic Press, New York.

Vence, D. V. (2002). *Marketers Always will Rely on Transactional Angle*, Marketing News, Chicago, Vol. 36, No. 13, 1-9.

Webster, F. E. (1992). *The Changing Role of Marketing in The Corporation*, Journal of Marketing, Vol. 56, 1-17.

Winer, R.S. (2001). *A Framework for customer relationship management*, California Management Review, summer, Vol. 43, No. 4, 89-105.

Yamaguchi, K. (1991). *Event history analysis*, Sage, Newbury Park, CA.

Zeithaml, V.A. (1981). *How consumer evaluation processes differ between goods and services*, Donnelly, J.H. and George, W.R. (Eds), Marketing of Services, American Marketing Association, Chicago, IL, 186-190.

Zeithalm, V.A., Berry, L.L. et Parasuraman, A. (1996). *The behavioural consequences of service quality*, Journal of Marketing, Vol. 60, April, 31-46.