

HEC MONTRÉAL

**La prise en compte de la capacité à payer
dans l'évaluation du risque de crédit des particuliers**

par

Hind Diboune

Sciences de gestion

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

Septembre 2008

© Hind Diboune, 2008

Table des matières

1	Introduction	9
2	Revue de littérature	12
2.1	Le risque de crédit	12
2.1.1	Introduction	12
2.1.2	Evaluation du risque de crédit	14
2.1.3	Cote de risque externe	15
2.1.4	Modélisation du risque de défaut	17
2.2	Capacité à payer	19
2.2.1	Capacité à payer : définition	20
2.2.2	Capacité à payer : problématique	24
2.3	Capacité à payer et risque de défaut	27
2.3.1	Les variables de la capacité à payer et la probabilité de défaut	27
2.3.2	Cote de capacité à payer et cote de risque de défaut	28
2.3.3	Cote de capacité à payer et probabilité de défaut	29
3	Données	32
3.1	Sources de données	32
3.2	Description des bases de données	32
3.2.1	Base de données Statistiques Canada	32
3.2.2	Base de données principale	35
4	Sélection des variables et hypothèses	37
4.1	Variables dépendantes	37
4.1.1	Les dépenses	37
4.1.2	Le défaut	38
4.2	Variables explicatives	39
4.2.1	Variables expliquant les dépenses	39
4.2.2	Variables explicatives du défaut	43
5	Statistiques descriptives	49
5.1	Moyenne des dépenses selon les personnes à charge et le type de logement	49
5.2	Comparaison des deux bases de données	50
5.3	Comparaison des taux de défaut selon la capacité à payer	51
5.4	Comparaison du groupe des défauts au groupe sans défaut	53
6	Description des modèles	56
6.1	Modèle d'estimation des dépenses	56
6.2	Construction de l'indice de capacité à payer	57
6.3	Estimation de la probabilité de défaut	61
7	Analyse des résultats	63
7.1	Estimation des dépenses	63
7.2	Estimation de la probabilité de défaut	66

TABLE DES MATIÈRES	4
8 Extensions de la recherche	74
9 Conclusion	79

Liste des tableaux

1	Informations incluses dans le pointage	16
2	Classement relatif moyen	18
3	Catégories d'âge	41
4	L'impact de la présence des personnes à charge sur les dépenses	49
5	L'impact du type de logement sur les dépenses	50
6	Comparaison des deux bases de données (avant pondération)'	51
7	Comparaison des deux bases de données (après pondération)	51
8	Taux de défaut pour les individus avec une capacité à payer négative	52
9	Taux de défaut pour les individus avec un remboursement mensuel supérieur à leur capacité à payer	52
10	Taux de défaut pour les cinq catégories d'écart	53
11	Comparaison des moyennes : défaut vs non défaut	54
12	Catégories d'écart	61
13	Régression MCO sur l'échantillon féminin	63
14	Régression MCO sur l'échantillon masculin	65
15	Régression Logit - Variable écart	67
16	Régression Logit - Toutes les variables	68
17	Régression Logit - Sans la variable écart	72
18	Régression Logit - Variables significatives	73
19	Régression MCO sur l'échantillon féminin (ajout de variables)	75
20	Régression MCO sur l'échantillon masculin (ajout de variables)	77

Table des figures

1	Dettes des ménages canadiens 1976-2006	13
2	Endettement par cote de risque pour les individus ayant des engagements financiers	29
3	Endettement par cote de risque pour les individus sans engagements financiers	30
4	Survie des prêts à terme	36
5	L'impact de l'âge sur le risque de défaut	44

Sommaire

Dans un contexte réglementaire de plus en plus strict, les banques sont contraintes à améliorer leur modèle de gestion du risque de crédit. L'intégration des informations relatives à la capacité à payer est devenue indispensable. Afin de disposer d'un bon indicateur de la capacité à payer de leurs clients, les banques doivent détenir toutes les informations sur les revenus et les dépenses de ces derniers. Ces institutions ont souvent accès aux données relatives aux engagements financiers et aux revenus de l'individu. Malheureusement, les informations concernant les dépenses demeurent toujours vagues ou inexistantes. Par conséquent, les banques ne peuvent avoir une image fiable de la capacité du client à payer le prêt qu'il demande. Dans ce mémoire, nous développons un modèle d'estimation des dépenses permettant aux institutions financières d'évaluer les dépenses des clients selon leurs caractéristiques. Nous développons également un indice de capacité à payer, et nous démontrons empiriquement son impact et son importance dans la gestion du risque de crédit.

mots clés : Gestion de risque, Capacité à payer, risque de défaut, cote de risque, credit scoring, estimation des dépenses, affordability, ability to pay.

Remerciements

En préambule à ce mémoire, je souhaite exprimer ici ma profonde gratitude à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce travail.

Tout d'abord, je tiens à remercier mon professeur et directeur de mémoire, Mr. Georges Dionne, pour sa disponibilité, ses conseils et ses suggestions très pertinentes qui m'ont aidé tout au long de ce travail.

Les remerciements les plus sincères vont également à Mme Denise Desjardins, dont l'aide et les recommandations me furent très précieuses. Son implication et sa rigueur m'ont permis de mener à bien ce mémoire.

Mes remerciements s'adressent aussi à toutes les personnes qui ont accepté de lire et d'examiner mon mémoire.

Enfin, je tiens à remercier ma famille et mes amis pour leurs encouragements et leur support inconditionnels. Je remercie particulièrement mes parents qui ont su, comme toujours, me donner le meilleur environnement pour l'accomplissement de mon travail. Merci !

1 Introduction

L'évolution des méthodes de financement ainsi que la grande expansion des crédits parmi la population nord américaine a conféré à la gestion de risque de crédit une grande importance. Les institutions financières, de par leur rôle d'intermédiaires, sont les plus concernées par ce risque. Selon K. Roszbach (2004), la décision des banques d'accorder ou non un prêt ainsi que la détermination des conditions du prêt est susceptible d'affecter non seulement le bien être et le niveau de consommation des individus mais aussi l'équilibre des marchés financiers (K. Roszbach, 2004). Une mauvaise évaluation de ce risque pourrait entraîner une crise financière qui s'étendrait à tous les secteurs économiques à travers le monde entier.

Lors de l'évaluation du risque de crédit, deux principaux éléments sont pris en considération : la probabilité de défaut et la capacité à payer. Bien que ces deux concepts semblent totalement indissociables, il serait intéressant de les définir séparément afin de distinguer le rôle de chacun d'eux dans cette évaluation.

La probabilité de défaut permet à la banque - selon ses objectifs et son aversion au risque - de déterminer le niveau de risque de ses clients, et ainsi décider d'accorder le prêt ou non. Les banques utilisent souvent des modèles de pointage pour évaluer le risque de défaut de leurs clients. Un modèle de pointage est défini comme étant un « ensemble de modèles décisionnels et de techniques sous jacentes qui ont pour but d'aider les banques dans l'octroi des crédits à la consommation » (Thomas et al, 2002). Cette méthode permet de calculer une cote de risque de défaut qui, une fois comparée à un seuil préétabli, permet à la banque de décider si elle accorde le prêt au client ou non. Les banques utilisent par la suite cette même cote de risque pour définir le taux d'intérêt du prêt.

La capacité à payer quant à elle représente le montant maximum qu'une personne peut encore payer sachant ses revenus, ses actifs liquides, ses dépenses et ses engagements financiers. Une bonne mesure de la capacité à payer du client permet à la banque

de définir des conditions de prêt adaptées à sa situation financière. De plus, cette information, utilisée à bon escient, est susceptible d'améliorer l'estimation de la probabilité de défaut permettant ainsi une évaluation du taux d'intérêt plus adéquate. Néanmoins la capacité à payer semble souvent délaissée au détriment d'autres aspects jugés plus intéressants. Actuellement, les banques utilisent des mesures très basiques pour calculer la capacité à payer. Souvent, ces mesures reflètent peu la réalité car elles omettent un élément des plus essentiels dans ce calcul : les dépenses des clients. Une bonne mesure de la capacité à payer doit inclure des données fiables sur les revenus, les engagements financiers, les actifs liquides et particulièrement les dépenses des individus. Cependant, ne disposant pas de sources fiables pouvant leur fournir les informations relatives aux dépenses, les banques se contentent des autres informations délaissant ainsi une composante essentielle de la capacité à payer. Ceci constitue la première problématique de ce mémoire : Comment construire une mesure fiable de la capacité à payer en tenant compte des dépenses ?

Par ailleurs, une fois cet indice calculé, il serait aussi intéressant de déterminer son rôle dans le processus décisionnel des banques. Ce qui nous amène à la deuxième problématique : durant quelle étape, et de quelle manière cet indice doit être utilisé pour améliorer le modèle interne de gestion du risque de crédit au sein d'une banque ? Quelle est la nature de la relation qui existe entre cet indice et la probabilité de défaut estimée ?

Afin de répondre à ces deux problématiques, nous avons développé une méthodologie qui se divise en trois étapes. La première consiste à construire un modèle d'estimation des dépenses. Pour ce faire, nous nous sommes basés sur les instructions exposées dans un des articles de SM. Finlay (2006). Le modèle obtenu a ensuite été appliqué à notre base de données principale afin de corriger les données relatives aux dépenses. La deuxième étape de notre méthodologie a été de développer un indice de capacité à payer à partir des informations dont nous disposons. Enfin nous avons vérifié au moyen de régressions logistiques, l'influence de ce nouvel indice sur la probabilité de défaut.

Notre étude représente une première étude empirique visant à construire un indice de capacité à payer fiable, capable d'aider la banque - de façon significative - durant la totalité de son processus de décision, l'hypothèse principale testée étant l'existence d'une valeur ajoutée à inclure l'indice de capacité à payer dans la gestion du risque de crédit.

Ainsi, le premier chapitre de ce mémoire contient une revue de littérature détaillée sur le risque de crédit : la probabilité de défaut, la capacité à payer et la relation entre ces deux concepts. Des définitions, ainsi que des théories économiques se rattachant au risque de crédit y sont introduites. Les deux bases de données utilisées lors de cette étude sont présentées dans le deuxième chapitre, les variables expliquées dans le troisième chapitre et enfin des statistiques descriptives sont exposées au quatrième chapitre. Nous expliquons par la suite notre méthodologie d'une manière détaillée au cinquième chapitre, et présentons les résultats au sixième chapitre. Une extension de la recherche est proposée dans le dernier chapitre, et une conclusion résumant l'ensemble des résultats de l'étude vient finalement clore cette étude.

2 Revue de littérature

2.1 Le risque de crédit

2.1.1 Introduction

De nos jours, le crédit à la consommation joue un rôle des plus importants dans la planification financière des ménages (K. Roszbach, 2004). L'utilisation de ce type de crédit est devenue avec le temps une pratique très répandue dans notre société. Nombreux sont ceux qui trouvent en cette pratique la solution idéale à leurs problèmes financiers. Les étudiants par exemple, voulant disposer d'un niveau de consommation plus élevé que leurs revenus (achat d'ordinateur, frais scolaires, . . .) choisissent souvent cette option. Les ménages ayant eu une baisse temporaire de leurs revenus, préfèrent recourir à des prêts afin de maintenir leur niveau de consommation habituel.

Il est donc évident que le crédit à la consommation est un élément qui contribue réellement au bien être des individus en leur permettant d'atteindre leur niveau de consommation désiré.

Étant dans une société de consommation par excellence, cette pratique ne cesse de croître jusqu'à menacer l'équilibre même des marchés financiers. En effet, le niveau d'endettement au Canada a nettement augmenté durant les dernières décennies et a atteint le chiffre record de 31081 dollars par personne en 2006, au terme d'un taux de croissance annuel moyen de 4.7%. Un taux qui devance de loin plusieurs indicateurs économiques canadiens tel que le PIB¹. La figure 1 illustre clairement cette tendance à la hausse.

¹"Où va l'argent ? La croissance de l'endettement des ménages au Canada", un rapport de l'association des Comptables Généraux Accrédités du Canada (CGA) - www.cga.org/canada-fr

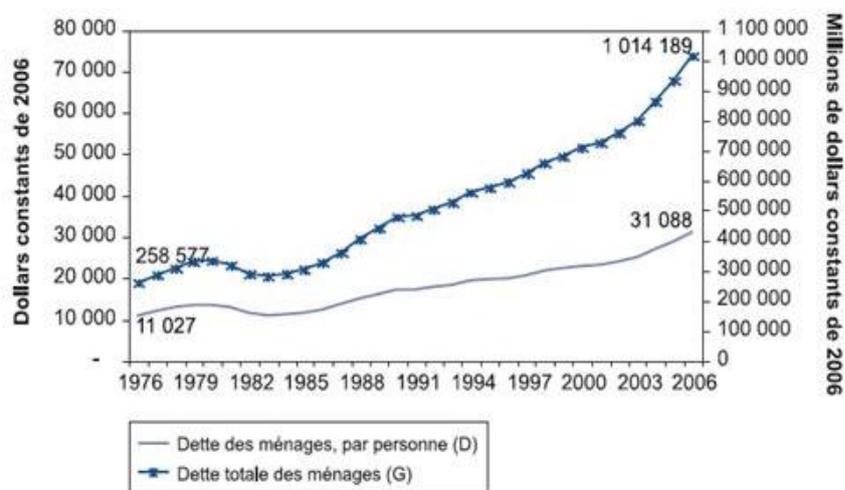


FIG. 1 – Dettes des ménages canadiens 1976-2006

Du point de vue des institutions financières, le crédit à la consommation représente une grande partie de leurs activités, étant donné qu'il constitue environ 34% du montant total des prêts accordés, toutes sortes confondues (K. Roszbach, 2004). Parallèlement, le risque lié à ce type de crédit n'est pas des moindres, la BIS (Bank of International Settlements) exige un capital réglementaire de 8% pour couvrir le risque des crédits à la consommation, contre seulement 4% pour les hypothèques. Ceci nous confirme la méfiance des organismes financiers internationaux quant à ce type de prêts (K. Roszbach, 2004).

Toutes ces informations confirment que le crédit à la consommation occupe une place importante dans le domaine financier, non seulement à cause de son volume de transactions mais aussi pour le risque qu'il engendre.

Ces facteurs ont incité les banques à améliorer leur processus de décision afin d'éviter les pertes liées au risque de crédit. Plusieurs méthodes et analyses ont été mises au point pour mieux évaluer ce risque tant appréhendé.

2.1.2 Evaluation du risque de crédit

Le processus de décision de la banque est composé de deux décisions distinctes et consécutives, la première représente la décision d'accorder ou non le prêt et la deuxième concerne les conditions du prêt : son montant, son échéance, et le taux d'intérêt appliqué.

En ce qui concerne la première décision, les banques évaluent le risque de défaut du client avant de décider si elle lui accorde le prêt ou non. Le risque de défaut représente le risque que l'emprunteur n'honore pas ses engagements vis-à-vis du prêteur. De manière plus claire, c'est la probabilité que l'emprunteur ne rembourse pas son prêt à la date fixée. Ce risque est perçu différemment selon qu'il s'agisse d'un prêt bancaire, d'une obligation, ou d'un produit dérivé, etc. Dans le présent mémoire, j'étudierai le risque de défaut lié aux prêts bancaires et plus précisément le risque de défaut lié au crédit à la consommation.

Comme nous l'avons constaté auparavant, accorder un prêt est devenu, pour une banque, une décision des plus difficiles qui requiert une étude et une évaluation très précises. En octroyant un prêt, une banque n'a pas de réelle certitude quant à son remboursement. Une mauvaise évaluation de ce risque pourrait donc entraîner des pertes assez conséquentes.

Plusieurs méthodes ont été conçues pour aider les banques à évaluer le risque de défaut des clients, la plus ancienne étant le jugement d'agents spécialisés. Cette méthode, très subjective, ne s'est pas révélée d'une grande efficacité car elle accordait beaucoup de valeur au jugement personnel et pouvait manquer de précision et même de rationalité (Thomas et al, 2002).

D'autres méthodes, plus quantitatives, ont été mises au point par la suite pour pallier à cette subjectivité du comité. Il s'agit de méthodes dites de « credit scoring »

ou de pointage. Ces méthodes consistent à attribuer à chaque client un score suivant ses caractéristiques personnelles, professionnelles et autres. Il s'agit donc d'établir une classification qui permet à la banque de trier ses clients selon le risque qu'ils représentent. Le degré de complexité de la méthode varie d'une banque à une autre. Il en existe qui utilisent des méthodes très élémentaires telles que l'attribution de scores sous forme de points pour chaque caractéristique du client, d'autres préfèrent quant à elles des modèles statistiques plus compliqués (K. Roszbach, 2004).

Afin d'évaluer le risque de défaut, les banques utilisent généralement une cote de risque. Il existe deux types de cotes de risque : une cote de risque externe fournie par une agence de crédit et une cote de risque interne construite à partir d'un modèle développé au sein même de l'institution et incorporant généralement la cote externe. Bien évidemment, les informations incluses dans ces deux scores sont différentes. Malheureusement nous ne pouvons avoir accès aux données et au modèle utilisés par l'institution financière pour la construction de la cote interne. Ce sont des informations confidentielles propres à chaque institution. Néanmoins, les données permettant de construire la deuxième cote de risque (la cote externe) sont disponibles auprès des bureaux de crédit.

La deuxième décision qui concerne les conditions du prêt sera discutée plus en détail dans la section 2.2 de la revue de littérature.

2.1.3 Cote de risque externe

Il existe plusieurs bureaux de crédit en Amérique du nord, les plus connus étant : TransUnion, Experian et Equifax. Ces agences calculent des scores compris entre 300 et 900, en se basant sur les informations du client en question (A. Carboni, 2007). L'historique des prêts représente la plus grande partie de ces informations. Un client ayant des antécédents prouvant une délinquance (retard d'un paiement, faillite, ...) aurait un score plus faible et donc moins de chances de se voir accorder un prêt.

Le tableau 1, issu du mémoire d’Alexandre Carboni, regroupe de manière claire les informations incluses dans le calcul de ces cotes selon les trois bureaux précédemment cités :

TAB. 1 – Informations incluses dans le pointage

	Transunion	Experian	Equifax
Informations personnelles			
Nom et prénom	Oui	Oui	Oui
Date de naissance et assurance maladie	Oui	Oui	Oui
Adresse complète	Oui	Oui	Oui
Durée à cette adresse	Non	Oui	Oui
Durée à l’adresse précédente	Oui	Non	Oui
Emploi actuel (durée)	Oui	Oui	Oui
Emploi précédent (durée)	Oui	Oui	Oui
Déclarations			
Déclaration du consommateur (cause de défaut)	Oui	Non	Oui
Informations de crédit			
Autres prêts et historique de paiement	Oui	Oui	Oui
Carte de crédit et historique de paiement	Tous	Tous	Tous
Comptes actuels et solde (historique)	Oui	Oui	Non
Registre public et autres informations			
Prêt garantie	Un	Tous	Tous
Faillite	Oui	Oui	Oui
Comptes en recouvrement	Oui	Oui	Oui
Demande de crédit			
Ceux qui ont regardés votre crédit	Oui	Oui	Oui
Caractéristiques			
Prix	30,95 \$	14,50 \$	23,45 \$
Fourchette de cote de risque	300 - 900	330 - 830	300 - 900

Il existe quelques différences quant aux informations incluses dans le calcul du pointage pour les trois bureaux de crédit étudiés. Une banque selon le bureau de crédit qu’elle choisit, aura a priori une cote de défaut différente des autres banques. De plus, comme nous pouvons le remarquer, la plus grande partie des informations utilisées pour établir la cote de risque externe concerne l’historique des prêts de l’individu. Aucune information relative à sa capacité à payer n’est contenue dans la cote externe.

2.1.4 Modélisation du risque de défaut

Selon L.C. Thomas, les modèles servant à estimer le risque de défaut diffèrent suivant l'objectif de la banque, selon son niveau d'aversion au risque, et selon si elle souhaite maximiser le profit ou minimiser le défaut des clients. En outre, ce même auteur affirme que ces modèles ont connu une évolution parallèle à l'évolution des objectifs des banques (LC. Thomas, 2002) .

Durant les dernières décennies, les institutions financières se sont lentement dirigées vers l'objectif de maximisation du profit au dépend de la minimisation de la probabilité de défaut. Les modèles de pointage ont évidemment suivi la même évolution et ont été à l'origine de plusieurs nouvelles méthodes utilisées actuellement, dont les techniques d'exploration des données (Datamining). Certains ont aussi développé des modèles intégrant une dimension comportementale du client, et ce en incluant les informations relatives aux comportements du client vis-à-vis des paiements périodiques de ses prêts(LC. Thomas, 2003). D'autres auteurs ont essayé de rassembler les deux objectifs dans un même modèle. C'est le cas de l'article "*Count Data model for a credit scoring system*" où Dionne et al. (1996) ont étudié conjointement la probabilité de défaut et les pertes associées aux non paiements des prêts².

Ceci dit, d'un point de vue économétrique, les deux approches les plus communes pour estimer le défaut sont la régression linéaire et la régression logistique, les deux stipulant que la probabilité de défaut (ou le log de $(p/1-p)$) dépend de certaines caractéristiques du client et les deux fournissent une probabilité de défaut que la banque compare avec son seuil préétabli pour décider si le client est un bon ou mauvais risque(LC. Thomas, 2003).

Des études plus récentes montrent l'utilisation d'autres méthodes telles que la méthode du plus proche voisin (Henley et Hand, 1996), les arbres de classification et les réseaux de neurones (Armingier, Enache et Bonne, 1997).

²Dionne et al, 1996

Certains auteurs ont essayé de classier ces différents modèles suivant leur efficacité et leur précision. Baesens (Baesens 2003, cité par LC. Thomas, 2003) a comparé la précision de dix sept méthodes en utilisant huit consommateurs différents avec huit bases de données et a regroupé les résultats obtenus dans le tableau 2 :

TAB. 2 – Classement relatif moyen

Méthode	Rég. lin.	Rég. log.	Prog. lin.	MVS	Rés. à neu- rones	Naïve Bayes	Arbres de class.	Plus proche voisin
Classement relatif en PCC	6.9	6.1	6.5	3.6	5.2	15.1	6.7	7.9

Les méthodes testées sont la regression linéaire (et sa variante quadratique), la regression logistique, la programmation linéaire, quatre variantes de la machine à vecteurs de support (MVS), quatre variantes de l'arbre de classification, deux variantes de la méthode du plus proche voisin, la méthode des reseaux de neurones, le modèle du Naïve Bayes et enfin la méthode des TAN (Tree-augmented Naïve Bayes). Ces méthodes ont été classifiées selon leur classement relatif moyen parmi les dix sept méthodes analysées. Ce classement a été établi selon la précision de leur classification, PCC étant le "percentage correctly classified" c'est-à-dire le pourcentage des observations classées correctement. Comme nous pouvons le constater, seule la méthode Naïve Bayes, mettant en oeuvre le théorème de Bayes avec de fortes (naïves) hypothèses, est significativement mauvaise.

Les banques choisissent la méthodes qui leur convient parmi celles précédemment énumérées, l'intègrent dans leur processus de décision et calculent leur probabilité de défaut.

Une fois la probabilité de défaut estimée, et le score calculé, la banque peut décider

si elle accorde le prêt ou non. Pour ce faire, elle compare la probabilité de défaut estimée (ou le score obtenu) avec un seuil de défaut préalablement établi « cut off value » et accepte les clients ayant une probabilité de défaut inférieure à ce seuil. Nous arrivons donc à la deuxième décision : les conditions du prêt. Sur quoi une banque se base-t-elle pour déterminer ces conditions ?

2.2 Capacité à payer

Rappelons que les conditions du prêt sont le montant, l'échéance et le taux d'intérêt. En ce qui concerne le taux d'intérêt, il est déterminé selon le risque que représente le client. Un client risqué se verra appliquer un taux d'intérêt plus élevé. Certaines études ont d'ailleurs estimé le taux d'intérêt conjointement avec la probabilité de défaut. Dionne et al (1996) ont utilisé l'équation d'équilibre de la dette standard pour calculer ce taux en se basant sur la probabilité de défaut estimée. La détermination du taux d'intérêt ne demande donc pas d'informations supplémentaires.

Cependant, afin de déterminer le montant à prêter et le terme, l'institution doit disposer d'informations sur la capacité à payer du client (revenu, actifs, dépenses, ...). Il est probable que ces informations soient incluses dans le calcul de la cote interne de la banque. Cela insinuerait donc qu'un bon nombre d'institutions financières utilisent un seul modèle pour déterminer à la fois si elle accorde le prêt ou non, ainsi que les conditions de l'éventuel prêt. Bien que plusieurs auteurs aient insisté sur le caractère multi décisionnel du processus d'accord du prêt (Durand, 1941 et Wonderlic, 1952), la majorité de la littérature ne traite que de la première décision : celle d'accorder ou non le prêt (SM Finlay, 2006).

Nous avons d'ailleurs remarqué précédemment la multitude des méthodes mises au point pour aider la banque dans sa première décision. Néanmoins, rares sont celles qui évaluent justement la capacité à payer afin d'aider la banque à déterminer les conditions du prêt, et elles sont d'autant plus rares pour les particuliers. Pour illustrer cela, une simple recherche sur le moteur « SSRN finance » avec comme mot clé : capacité à

payer, « ability to pay » ou même « affordability », ne donne aucun article qui traite de ce sujet dans le domaine financier (SM Finlay, 2006).

2.2.1 Capacité à payer : définition

"Il n'existe pas de définition largement acceptée de la capacité du client à payer le prêt demandé ni de ce qui pourrait le qualifier un client de surendetté ou non " (Consumer Affairs Directorate, 2003) (SM Finlay, 2006). Cette phrase nous informe clairement qu'il n'existe pas de définition généralement admise de la capacité à payer dans le monde financier. Cependant, quelques auteurs, ou plus précisément deux auteurs, ont tenté de définir ce concept de manière concrète. LC. Thomas et SM. Finlay ont essayé de fournir une définition mathématique de la capacité à payer.

LC. Thomas estime que la capacité à payer d'un individu au moment $t + 1$ est égale à son revenu et ses actifs liquides moins ses dépenses et les remboursements de ses prêts au moment t . Plus concrètement, l'auteur nous propose la formule suivante (LC. Thomas, 2005) :

$$\begin{aligned} (\text{Liquidités})_{t+1} &= (\text{Revenu})_t + (\text{liquidités})_t - (\text{Dépenses})_t \\ &\quad - (\text{Remboursement des prêts})_t \end{aligned}$$

Bien évidemment, les liquidités ici représentent la capacité à payer de la personne. Un client est en défaut si la somme de ces éléments est négative, car cela voudrait dire qu'il ne dispose pas de fonds lui permettant de payer son prêt. Il faudrait donc définir chacun de ces éléments pour pouvoir calculer de manière juste la capacité à payer.

En ce qui concerne le revenu, l'auteur ne précise pas quel genre de revenus nous pouvons inclure dans cette variable, nous ne savons donc pas s'il s'agit du revenu d'emploi seulement ou s'il faudrait ajouter les revenus relatifs aux placements, aux gains

occasionnels et autres. Il n'indique pas non plus s'il s'agit de revenus bruts ou de revenus nets.

Par ailleurs, l'auteur stipule que les liquidités n'incorporent pas les biens immobiliers du client mais seulement ses actifs immédiatement liquides tels que les placements financiers et l'épargne.

Concernant les dépenses, l'auteur suggère d'utiliser des enquêtes sur les revenus et les dépenses des personnes pour développer un modèle capable d'estimer les dépenses en fonction de certaines caractéristiques du client et de l'utiliser par la suite afin de calculer la capacité à payer. Malheureusement, rares sont les banques qui peuvent avoir accès à ce genre de données et l'auteur en est bien conscient. Il propose d'ailleurs une méthode alternative qui exige l'accès de la banque aux comptes courants des clients. La banque pourrait donc utiliser le montant mensuel total des crédits figurant dans le compte du client comme étant les revenus et le montant mensuel total des débits comme étant les dépenses et calculer ainsi la capacité à payer de l'individu pour le mois à venir comme étant la différence entre ces deux montants.

Dans cet article, l'auteur ne fournit pas plus d'informations sur la méthode d'évaluation de la capacité à payer et comme nous l'avons constaté plusieurs précisions manquent au calcul de cet indice. Nous ne pouvons donc nous baser uniquement sur cette définition pour bâtir notre indice de capacité à payer.

Finlay (2006), quant à lui, procure plus d'indications dans son article « *Predictive models of expenditure and over indebtedness for assessing the affordability of new consumer credit application* » le seul article qui traite de manière relativement détaillée la capacité à payer des particuliers. Il faut admettre que sa définition ne diffère pas beaucoup de la définition de LC. Thomas, néanmoins, elle demeure plus précise, mieux expliquée et surtout étudiée empiriquement.

Finlay définit donc la capacité à payer comme suit (SM Finlay, 2006) :

$$A = N - C - E$$

pour tout

$$N > 0; E \geq 0; C \geq 0$$

où A représente ici la capacité à payer tandis que N , C et E représentent respectivement le revenu, les engagements financiers et les dépenses.

L'auteur précise que le revenu dans cette formule représente le revenu net et inclut toutes les sources de revenus tels que le salaire, les bonus, les avantages sociaux, etc. Les engagements financiers désignent quant à eux tous les engagements financiers du client incluant l'hypothèque, les marges de crédit et bien évidemment les autres prêts. Pour ce qui est des dépenses, elles incluent toutes les dépenses hors crédits tels que le loyer, la nourriture, les voyages, les loisirs . . .

Ces deux types de dépenses (les engagements financiers et les dépenses hors crédits) sont traités différemment en raison de la différence de l'origine et de la méthode de collecte des informations nécessaires pour chacune d'entre elles. En effet, l'auteur estime que les banques ont généralement accès aux informations relatives aux engagements financiers de leurs clients via le rapport du bureau de crédit mais qu'ils n'ont pas de données fiables sur les dépenses réelles de leurs clients. C'est d'ailleurs le problème le plus important dans une étude comme celle-ci où le calcul de la capacité à payer revêt une grande importance. Nous discuterons cette problématique de façon plus approfondie dans la section 2.2.2 du mémoire.

SM. Finlay fournit aussi une définition de la capacité à payer sous forme de ratio. Il considère la capacité à payer relative comme étant une meilleure mesure puisqu'elle tient compte de l'inflation et des variations qui peuvent survenir dans le temps. La formule est donc comme suit :

$$A_r = 1 - \frac{C}{N} - \frac{E}{N}$$

A_r représente la capacité à payer relative de l'individu, c'est-à-dire $\frac{A}{N}$. C'est la capacité à payer absolue par rapport au revenu actuel (N) de l'individu, elle est donc en fonction des engagements financiers relatifs $\frac{C}{N}$ et des dépenses relatives $\frac{E}{N}$.

Cette définition est par la suite utilisée afin de calculer l'endettement des clients. L'endettement relatif est alors considéré comme étant la somme des engagements financiers et des dépenses relativement au revenu net. Ce qui nous donne une définition sous la forme suivante :

$$I = \frac{C}{N} + \frac{E}{N} = 1 - A_r$$

D'après cette définition, un client est surendetté si cet indice est supérieur à une valeur θ définie selon l'objectif et le niveau d'aversion au risque des banques. De manière générale, cette valeur est égale à un, mais il est à noter que certaines banques dont l'objectif est d'accorder le plus de prêts possibles peuvent affecter à θ une valeur supérieure à un. Cela voudrait dire que ces banques ne rejetteraient un client que s'il est sérieusement surendetté. On constate d'ores et déjà l'importance de la prise en compte de la capacité à payer dans le processus décisionnel de la banque.

Les deux définitions que nous venons de citer constituent les seules définitions de la capacité à payer présentes dans la littérature actuelle. Utiliser ces définitions requiert un volume d'informations considérable et surtout fiable pour pouvoir calculer la capacité à payer de manière juste et précise.

2.2.2 Capacité à payer : problématique

- **Collecte d'informations**

Il existe plusieurs approches qui nous permettent de disposer des informations nécessaires afin de pouvoir calculer les indices de capacité à payer précédemment discutés. L'approche la plus simple qu'une banque puisse utiliser est de demander directement aux clients des questions relatives à leurs revenus et leurs dépenses. Cependant cette approche a plusieurs limites (SM Finlay, 2006).

Tout d'abord, les banques opèrent dans un marché de plus en plus concurrentiel qui les oblige à raccourcir le plus possible leur formulaire de demande de financement. Plus le formulaire est long, plus la probabilité qu'un client le remplisse adéquatement est faible (Thomas et al, 2002). Nous comprenons donc que les banques préfèrent éviter d'allonger leur formulaire en rajoutant des questions détaillées sur les revenus et les dépenses du client.

Ce genre de problème se pose surtout pour les informations relatives aux dépenses. Souvent les banques exigent des informations sur les revenus de la personne et sur ses engagements financiers mais rares sont celles qui demandent aussi des informations sur les dépenses. Pour illustrer cette situation, SM Finlay a réalisé une enquête où trente formulaires de demande de financement à différentes banques ont été collectés dont 15 pour des prêts non sécurisés et 15 pour des cartes de crédit. Parmi les 30 formulaires, 12 ne demandaient aucune information concernant les dépenses, 16 posaient des questions vagues et assez basiques telles que « Quelles sont vos dépenses mensuelles » ou encore « À combien s'élève votre hypothèque ou loyer ? » et seulement 2 contenaient des questions plus détaillées sur les dépenses (Barclays et HSBC)(SM. Finlay, 2006). Nous constatons donc que le nombre des banques qui s'intéressent réellement aux informations relatives aux dépenses est très limité.

Par ailleurs, cette approche connaît une autre limite : les informations fournies par les clients concernant leurs dépenses ne sont pas nécessairement fiables. En effet, même si elles incluaient dans leurs formulaires des questions relatives aux dépenses, les banques ne peuvent être sûres de la fiabilité des informations données par les clients. Il est donc indispensable d'établir une nouvelle approche permettant la collecte d'informations sur les dépenses des clients afin de calculer un indice de capacité à payer réaliste.

Les dépenses : Collecte d'informations

Nous avons vu précédemment que LC Thomas proposait aux banques d'utiliser les données d'une enquête gouvernementale sur les dépenses afin de bâtir un modèle capable d'estimer les dépenses. SM. Finlay, étudie cette proposition de manière empirique dans son article et nous fournit des précisions des plus intéressantes.

Pour bâtir son modèle de prévision des dépenses, SM. Finlay a utilisé des données d'une enquête gouvernementale britannique nommée « the UK Government's annual Expenditure and Food Survey, 2001 » pour estimer les dépenses, plutôt que des données issues des institutions financières. La base de données contenait près de 11975 observations réparties à travers l'Angleterre et pondérées de sorte à tenir compte des différences géographiques et du taux de non réponse.

Cette base de données avait l'avantage de fournir des informations détaillées sur les dépenses des ménages puisqu'elle incluait 400 catégories de dépenses différentes. En outre elle constituait un échantillon représentatif de la population susceptible de déposer une demande de financement auprès d'une banque. C'est principalement pour ces deux raisons que l'auteur a décidé de choisir ces données. L'étude n'incluait cependant pas de données sur les engagements financiers des ménages ni d'informations relatives au défaut. Afin de pallier à ce problème, l'au-

teur a utilisé une autre base de données issue d'un bureau de crédit pour bâtir un modèle capable d'estimer la probabilité de défaut, qu'il a appliqué à sa base de données principale par la suite.

Sa méthodologie était simple et claire : appliquer une régression linéaire aux données en utilisant le ratio $\frac{E}{N}$ - quotient des dépenses sur le revenu net - comme variable dépendante. Il a choisi une série de variables factices binaires comme variables explicatives. Il a réussi ainsi à construire un modèle solide qui lui permet d'estimer les dépenses de n'importe quel individu en se basant sur ses caractéristiques.

- **Utilisation des informations sur la capacité à payer**

Les banques utilisent actuellement deux approches pour prendre en considération la capacité à payer des clients. La première consiste à jumeler le modèle de pointage utilisé pour estimer le défaut avec les informations relatives à la capacité à payer. La logique derrière cette méthode est que le surendettement d'un client le mène irrémédiablement au défaut. Wilkinson et Tingay ont testé cette approche en effectuant une analyse empirique sur l'impact de l'intégration des informations relatives à la capacité à payer dans la construction d'un modèle de pointage pour estimer le défaut (Wilkinson et Tingay, 2004). Ils ont conclu alors que le développement d'un modèle de pointage avec des données relatives à la capacité à payer confère au modèle une légère augmentation marginale de la performance.

Par ailleurs, les modèles de pointage sont généralement basés sur des informations dont l'horizon varie entre 12 et 24 mois (Lewis, 1992 ; Rosenberg et Glait, 1994). Cet horizon peut ne pas être suffisant pour couvrir toute la période durant laquelle un client est censé maintenir ses paiements. Dans un tel cas, il serait donc peu prudent de se baser sur cet horizon.

La deuxième approche, différente de la première, traite quant à elle ces deux volets d'informations - capacité à payer et défaut - distinctement, et utilise deux indices pour chacun d'entre eux. Cette méthode se base sur le jugement d'un expert qui évalue la cote d'endettement de la personne et décide des conditions de prêts le cas échéant. Généralement, la cote utilisée est l'amortissement total de la dette noté ATD. Cet indice fournit une mesure relative des engagements financiers du client par rapport à ses revenus bruts. Les engagements financiers incluent le paiement du prêt ainsi que tous les autres engagements financiers mensuels du client. En général un prêt est accordé si cet indicateur est inférieur à 35%, ou dans certains cas à 40%. La somme des engagements financiers mensuels du client ne doit donc pas dépasser 40% de son revenu brut (A. Carboni, 2007).

Malheureusement, cet indicateur ne tient compte que des dépenses relatives aux engagements financiers. Il n'inclut pas les autres dépenses telles que le loyer, le transport, la nourriture ... Par conséquent, cet indice ne prend pas en considération les conditions et le niveau de vie de chaque client. Deux clients distincts peuvent avoir le même ATD avec des conditions de vie totalement différentes. L'un d'entre eux peut par exemple vivre dans une grande maison avec une famille constituée d'une femme et de deux enfants, tandis que l'autre vit seul dans un appartement. Il est évident que les deux individus auront des dépenses mensuelles différentes. Cet exemple, proposé par SM. Finlay, nous démontre clairement l'importance des données relatives aux dépenses (hors crédit) dans le calcul de la capacité à payer des clients.

2.3 Capacité à payer et risque de défaut

2.3.1 Les variables de la capacité à payer et la probabilité de défaut

Le lien entre la probabilité de défaut et les variables de capacité à payer - revenu, dépenses, engagements financiers, actifs tangibles - est plutôt confus. L'influence de

ces variables dans l'estimation de la probabilité de défaut change selon que la cote de risque de défaut est incluse dans le modèle ou non. Carboni a vérifié ce constat au moyen d'une régression logistique, et a conclu qu'en intégrant la cote de défaut, l'effet des variables de capacité à payer est moins significatif, et concerne seulement trois d'entre elles, à savoir : les dépenses, les engagements et surtout les actifs tangibles. L'impact de ces données demeure néanmoins logique, les dépenses et les engagements financiers augmentent la probabilité de défaut contrairement aux actifs tangibles (A. Carboni, 2007).

2.3.2 Cote de capacité à payer et cote de risque de défaut

Comme nous l'avons constaté, les articles qui proposent un indicateur de capacité à payer sont très rares. Seul l'article de SM. Finlay propose un indicateur de capacité à payer et vérifie le lien entre cet indicateur et la cote de risque de défaut (SM. Finlay, 2006).

A. Carboni a, quant à lui, construit deux indicateurs de capacité à payer. Le premier est celui suggéré par LC. Thomas (LC. Thomas, 2005) :

$$CAPA = \text{actifs tangibles} + \text{revenu} - \text{dépenses} - \text{paiement}$$

Et le deuxième est obtenu en utilisant uniquement les variables de capacité à payer significatives dans sa régression initiale sachant que la cote de défaut est prise en compte :

$$COTE2 = 2 \times \text{actifs tangibles} - \text{dépenses} - \text{paiement}$$

L'auteur vérifie ensuite, au moyen d'une régression linéaire, l'influence de la cote de risque de défaut utilisée par la banque sur ces deux indicateurs. La cote de risque de défaut s'avère très significative pour expliquer les deux indicateurs de capacité à payer. Nous pouvons donc affirmer qu'il y a une certaine relation entre la capacité à payer et la cote de risque de défaut aux particuliers (A. Carboni, 2007).

2.3.3 Cote de capacité à payer et probabilité de défaut

Finlay affirme dans son article que l'indice de capacité à payer et la probabilité de défaut sont fortement liés. En effet, il teste cette relation en deux étapes. La première consiste à vérifier la corrélation entre l'indicateur de capacité à payer et le risque de défaut. Cette corrélation s'est avérée significative à 99%. Il a ensuite utilisé des graphiques mettant en relation le défaut et l'endettement des personnes.

Nous rappelons que son indice d'endettement est égal à un moins la capacité à payer relative ($I = 1 - Ar$). La figure 2 nous montre clairement la corrélation négative qui existe entre l'endettement et le non défaut.

Il a ensuite comparé la figure 2 à la figure 3 qui inclut uniquement les individus sans engagements financiers : ceux dont l'indice d'endettement est constitué exclusivement des dépenses (hors crédits).

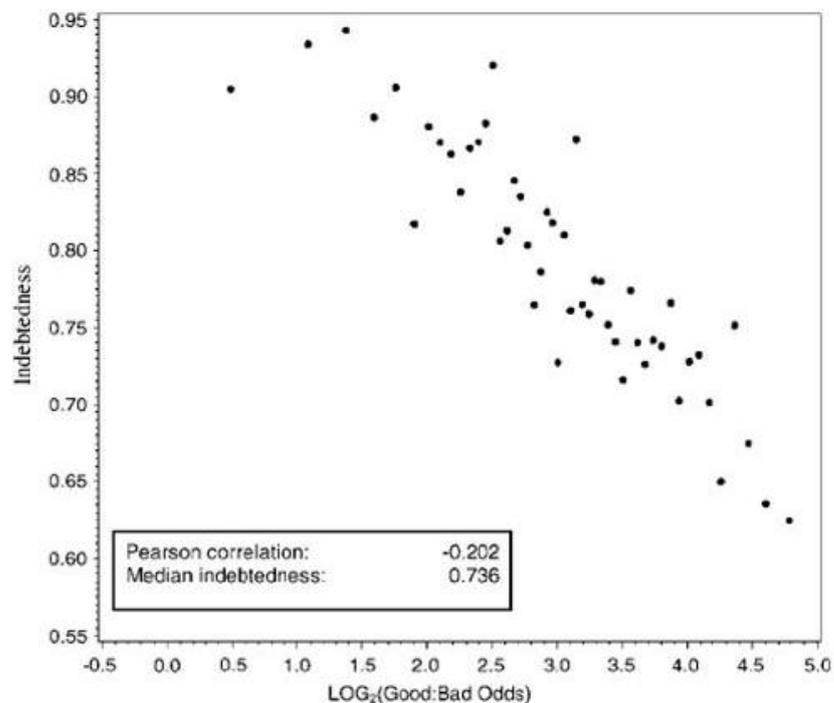


FIG. 2 – Endettement par cote de risque pour les individus ayant des engagements financiers

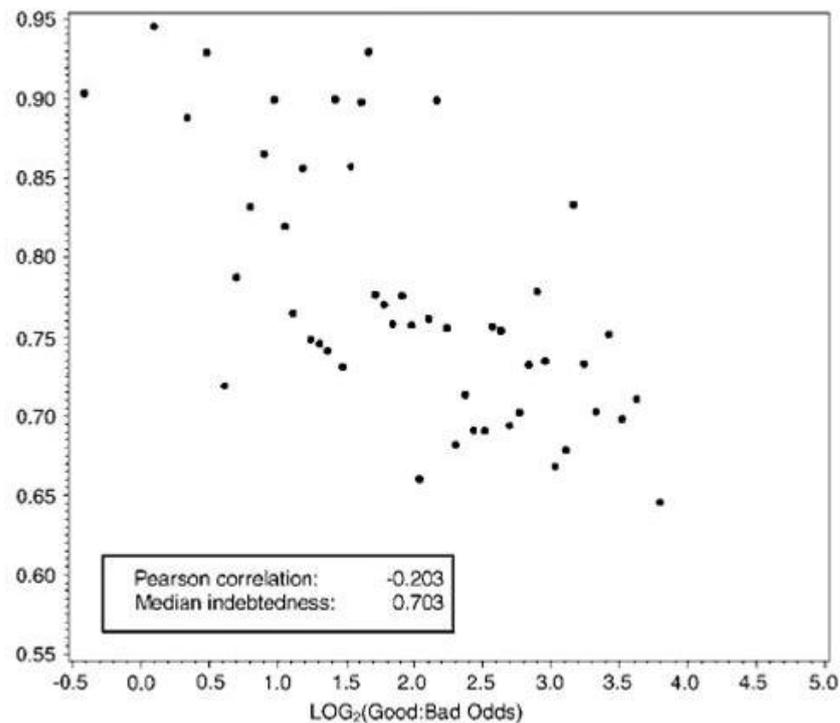


FIG. 3 – Endettement par cote de risque pour les individus sans engagements financiers

Même si la relation entre l'indice d'endettement et le risque de défaut demeure significative pour ce dernier groupe d'individus, nous remarquons qu'elle est moins accrue lorsque les individus n'ont pas d'engagements financiers. L'auteur en tire plusieurs conclusions.

Premièrement, ce résultat montre que pour ceux qui n'ont pas encore contracté un crédit, la cote de risque de défaut est un mauvais indicateur de leur capacité à payer un emprunt.

Ensuite, des individus ayant des cotes de risque acceptables (supérieures à 2.322 sur le graphique) au vu des critères de nombreuses banques, présentent tout de même un niveau d'endettement supérieur à la moyenne (0.703). L'auteur en déduit qu'il existe un nombre significatif de clients qui sont surendettés mais qui réussissent quand même à obtenir de nouveaux prêts car la banque se base uniquement sur leur cote de risque de défaut.

Enfin l'auteur conclut qu'il serait intéressant d'utiliser en plus du modèle d'estimation du défaut, un modèle qui prédit la capacité à payer avant d'accorder le prêt aux clients,

spécialement pour ceux n'ayant jamais contracté de prêts (SM. Finlay, 2006).

Carboni, tente aussi de vérifier si ses deux indicateurs de capacité à payer sont en mesure d'estimer significativement le défaut. Il procède de trois façons. D'abord, les deux cotes créées sont testées par régressions logistiques, d'abord seules et ensuite avec la cote de risque de défaut. Testés seuls, ces indicateurs parviennent à prédire le défaut. Par contre, les résultats sont moins convaincants lorsqu'on évalue les deux indicateurs en incluant la cote de défaut. L'auteur conclut donc que la capacité à payer n'apporte aucune valeur ajoutée lors de l'estimation du défaut.

Par la suite, pour s'assurer de la robustesse des résultats, il répète les mêmes régressions en incluant les variables explicatives susceptibles d'influencer le défaut. Les résultats obtenus amènent aux mêmes conclusions.

La troisième méthode est de construire des valeurs prédites de ses deux indicateurs et de vérifier ensuite si ces nouvelles valeurs sont significatives pour estimer le défaut. Les résultats obtenus confirment que les indicateurs de capacité à payer ne sont pas significatifs pour expliquer le défaut lorsque la cote de défaut est intégrée dans la régression. L'auteur a également vérifié si les résultats changent lorsque nous observons la cote de risque interne de l'institution. Les résultats obtenus sont quasiment identiques à ceux obtenus précédemment.

En conclusion, Carboni a démontré dans son mémoire que la capacité à payer n'est efficace pour prédire le défaut que lorsque la cote de défaut n'est pas prise en compte (A. Carboni, 2007).

Nous pouvons donc conclure que l'indice de capacité à payer tel qu'il a été développé par Carboni n'augmente pas la performance du modèle d'estimation du défaut et ne fournit pas d'informations supplémentaires à la banque sur le risque de défaut de ses clients.

Ce résultat pourrait être expliqué par le fait que la variable dépense utilisée lors de la construction de cet indice n'était pas bien documentée.

3 Données

L'objectif de ce mémoire étant d'abord de construire un indice de capacité à payer fiable, et donc de corriger la variable dépense. Nous avons eu besoin de deux bases de données. La première nous a permis de développer un modèle pour estimer les dépenses, et la deuxième, qui est notre base de données principale, a été utilisée pour toutes les autres analyses : construction d'indice de capacité à payer, estimation du défaut . . .

3.1 Sources de données

La première base de données est issue de Statistiques Canada. Elle a été collectée et analysée lors d'une enquête sur les dépenses des ménages au Canada. Elle inclut toutes les informations concernant les dépenses des ménages, mais ne fournit aucune information sur l'historique des prêts, les défauts et la cote de risque de ces ménages.

La deuxième base de données est issue d'une banque canadienne. Elle contient l'information relative aux défauts et aux cotes de risques des clients mais les données concernant les dépenses sont mal documentées. Il est à noter que Carboni a utilisé ces données dans son étude, ce qui peut avoir influencé ses résultats.

3.2 Description des bases de données

3.2.1 Base de données Statistiques Canada

La base de données utilisée pour l'enquête sur les dépenses des ménages contient des informations concernant 21315 ménages sélectionnés à travers toutes les provinces du Canada. Cette enquête s'est déroulée de janvier à mars 2006 au moyen d'un questionnaire rempli lors d'entrevues personnelles avec chacun des ménages sélectionnés. Cette base de données est disponible dans le centre de données de recherche de Statistiques Canada. Elle a été corrigée et pondérée par Statistiques Canada afin qu'elle soit

cohérente et représentative de la population canadienne.

Nous avons eu recours à cette base de données afin de remplacer la variable dépense dans la base de données principale. L'information concernant les dépenses était mal documentée et pouvait ainsi induire des erreurs lors de la construction de l'indice de capacité à payer et de l'estimation du défaut. L'analyse de la base de données principale a révélé que les dépenses déclarées par les clients sont nettement inférieures à leur niveau réel. La moyenne des dépenses mensuelles dans cette base de données ne dépassait pas les 265 dollars par mois. Sachant que ces dépenses devraient inclure le loyer ou l'hypothèque et toutes les autres dépenses, nous pouvons affirmer que l'information donnée par les clients relativement à cette question est complètement erronée.

Afin d'éviter ce problème nous avons développé un modèle permettant l'estimation des dépenses en nous basant sur le modèle de Finlay. Ce modèle a été par la suite appliqué à la base de données principale dans le but d'obtenir les dépenses estimées de cette population.

Les deux bases de données étant assez différentes, nous avons dû établir une liste de corrections à appliquer à la base de données de Statistiques Canada afin qu'elle ait des caractéristiques et des pondérations semblables à la base de données principale. Premièrement, les observations dans cette base de données ne représentaient pas des individus mais des ménages. Or, dans la base de données principale, les observations représentaient les prêts contractés par une seule personne (pas de prêts joints). Il fallait donc corriger l'échantillon de Statistiques Canada afin de détenir les observations concernant un individu et non un ménage. L'idée a été de ne sélectionner que les ménages comprenant une seule famille économique et qui étaient composés soit d'une seule personne soit d'un parent unique avec dépendants (enfants ou autres). Nous avons effectué cette sélection afin de disposer des observations qui ne prenaient en compte qu'une seule source de revenu. Deuxièmement, cette base de données concernait toutes les provinces canadiennes, or, nous savons pertinemment que le niveau de vie, les dépenses ainsi

que les revenus varient d'une province à l'autre. Par ailleurs, la banque canadienne pour laquelle nous devons établir cette étude opère principalement au Québec. Nous avons donc préféré exclure les autres provinces et ne garder que les informations collectées dans la province du Québec. Ensuite, nous avons rencontré le problème lié à l'âge des individus inclus dans la base de données. La base de données de l'enquête comprenait des personnes relativement plus âgées que celles de la base de données de la banque. Nous avons alors effectué une sélection afin de disposer de catégories d'âge comparables. Par ailleurs, nous avons remarqué la présence d'observations contenant des valeurs manquantes dans nos données que nous avons été contraints de supprimer. De plus, nous avons appliqué un traitement aux valeurs extrêmes afin de supprimer toute valeur pouvant être aberrante et capable de compromettre les résultats. Autre différence : le pourcentage des femmes dans chaque base de données. La base de données de l'enquête comprenait nettement plus de femmes que la base de données principale. Afin d'y remédier, nous avons scindé cet échantillon en deux parties : un échantillon ne concernant que les hommes et un autre ne contenant que les femmes. nous avons effectué cette scission pour plusieurs raisons. Premièrement, cela nous permettra d'avoir un pourcentage de femmes proche de celui existant dans la base de données principale. Deuxièmement cette opération nous facilitera la pondération de cette base de données par rapport aux autres variables (l'âge et les dépendants). Enfin, l'utilisation de deux échantillons distincts nous permettra de mieux cerner les différences en matière de dépenses entre les hommes et les femmes ce qui n'aurait pas été le cas si nous avions utilisé un seul échantillon les rassemblant tous les deux. Ces deux échantillons ont été pondérés par la suite de façon à obtenir des échantillons comparables à l'échantillon principal auquel nous devons appliquer les modèles construits.

Nous avons finalement obtenu deux échantillons distincts : le premier incluant 912 observations et ne concernant que les femmes et le deuxième comprenant 1000 observations et ne concernant que les hommes. Chacun de ces deux échantillons devait servir à construire un modèle : le premier estimant les dépenses pour les femmes et un autre qui estime les dépenses des hommes.

3.2.2 Base de données principale

Notre base de données principale est plus importante tant en nombre qu'en volume d'informations. Elle contient un total de 34495 observations qui représentent tous les prêts à la consommation ouverts existants dans la base de données de la banque au 31 décembre 2007. Ces prêts sont accompagnés de toutes les informations disponibles au moment de l'obtention du prêt (caractéristiques du client, historique de crédit, actifs et passifs ...) ainsi que des informations courantes du prêt telles que les remboursements, les retards de paiement, la cote de risque ...

Il est à noter que nous avons sélectionné des prêts notés (scorés) au moins une fois et au plus deux fois après qu'ils ont été accordés. Sachant que l'attribution de score à un prêt en cours se fait annuellement, l'échantillon contient des prêts âgés de deux à trois ans. Nous avons décidé d'adopter cette sélection après avoir effectué une analyse d'une autre base de données issue de la même institution, qui nous a informé que la plupart des prêts faisait défaut durant leur deuxième année et au plus tard durant leur troisième année. Les prêts âgés de plus de trois ans ont donc une probabilité de défaut plus faible. Aussi avons nous décidé de garder les prêts ayant au moins deux ans pour disposer de plus d'informations et au plus trois ans pour avoir un taux de prêts faisant défaut suffisant.

La figure 4 montre que, dans l'ancien échantillon, seul 5.6% des prêts en défaut, ont été en défaut après trois ans.

Cet échantillon a également subi une série de traitements afin d'obtenir des informations cohérentes et sans valeurs manquantes ni aberrantes.

Premièrement, nous avons supprimé toutes les observations ayant des valeurs manquantes spécifiquement lorsqu'il s'agissait de variables importantes dans notre étude. Deuxièmement, les valeurs extrêmes ont aussi été supprimées en utilisant la même mé-

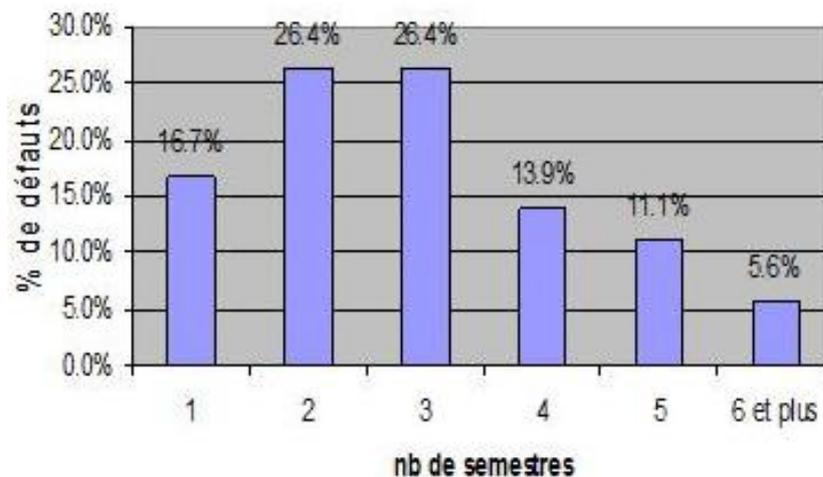


FIG. 4 – Survie des prêts à terme

thode que celle employée dans la base de données de Statistiques Canada.

Nous avons également remarqué que cette base de données contenait beaucoup de doublons, c'est-à-dire, des observations identiques qui se répétaient, et qui devaient, elles aussi, être éliminées.

Comme nous l'avons précédemment mentionné, cette base de données regroupe toutes les informations nécessaires à notre analyse, sauf la variable dépense qui était malheureusement mal documentée. Nous avons donc appliqué le modèle construit à partir de la première base de données afin d'estimer les dépenses dans la base de données principale et remplacer la variable erronée.

Finalement nous avons obtenu un échantillon de 28312 observations, avec un taux de défaut de 1.99%, soit 564 défauts contre 27748 non défauts. Taux que nous avons jugé suffisant pour notre étude.

La partie suivante discutera de la sélection des variables utilisées, et de leur rôle dans chacun des modèles employés.

4 Sélection des variables et hypothèses

La sélection des variables est une étape cruciale dans une étude statistique. Le choix des variables peut affecter l'efficacité et la performance du modèle. Il est donc très important de bien choisir les variables à intégrer dans les modèles employés. Dans ce mémoire, le choix des variables est d'autant plus important vu que notre étude se divise en deux étapes principales. Chaque étape nécessite donc un choix de variables et un modèle distincts. Par ailleurs, la variable dépendante du premier modèle sera une variable explicative dans le deuxième. Il est donc indispensable d'effectuer le choix des variables de chaque étape de façon prudente afin d'utiliser les informations susceptibles d'apporter une valeur ajoutée à chaque modèle.

4.1 Variables dépendantes

4.1.1 Les dépenses

Le premier modèle vise à estimer les dépenses. Il est donc évident que la variable dépendante de ce modèle n'est autre que la variable dépenses. D'après Finlay, cette variable devrait inclure toutes les dépenses qu'une personne pourrait effectuer dans sa vie courante telles que les dépenses relatives à l'alimentation, l'habillement, le transport... Elle exclut néanmoins les dépenses rares telles que l'achat de véhicules, l'achat d'électroménagers, les vacances, les rénovations et les frais de déménagement, et ce à cause du caractère exceptionnel de ces dépenses (SM. Finlay, 2006). L'intégration de ces dépenses exceptionnelles pourrait biaiser nos estimations par la suite en augmentant visiblement les dépenses estimées des individus. Par exemple, une personne n'achète pas une voiture tous les mois pour que cette dépense soit incluse dans l'estimation de ses dépenses mensuelles.

L'avantage de notre première base de données réside dans le fait qu'elle contient tous les types de dépenses qu'un ménage peut avoir, nous avons donc pu inclure toutes les dépenses que Finlay avait incluses dans son article en tenant compte des exclusions

proposées par l'auteur.

Par ailleurs, nous avons aussi inclus les impôts dans les dépenses car nous ne disposons pas de cette information dans la base de données principale. Dans son article, SM. Finlay n'incluait pas cette dépense, cependant il utilisait les revenus nets d'impôt. Puisque les revenus dont nous disposons dans l'échantillon de la banque sont des revenus bruts, nous avons décidé d'intégrer cette dépense dans la variable dépense.

Finalement, nous avons préféré utiliser les dépenses absolues et non les dépenses relatives au revenu net comme ce qui a été le cas dans l'article de Finlay. Ce sera donc une variable continue qui pourra prendre les valeurs de zéro jusqu'à la plus haute dépense. Nous tenons à préciser qu'il s'agit ici de dépenses mensuelles et non annuelles.

Une fois le modèle construit, et la variable estimée, elle remplacera la variable dépense qui existe déjà dans la base de données principale. Nous pourrions par la suite l'utiliser pour réaliser la deuxième étape de notre étude.

4.1.2 Le défaut

Le défaut représente la variable dépendante du modèle servant à estimer la probabilité de défaut. Cette variable est une variable binaire qui peut prendre la valeur 0 ou 1 selon que le prêt est en défaut ou non.

$$\text{La variable défaut} = \begin{cases} 1 & \text{si défaut} \\ 0 & \text{si non défaut} \end{cases}$$

La variable défaut sera donc utilisée en tant que variable dépendante dans notre deuxième modèle qui comptera également plusieurs variables explicatives. Ceci nous permettra d'identifier l'impact de certains éléments sur la probabilité de défaut des clients et spécialement ceux relatifs à la capacité à payer.

4.2 Variables explicatives

4.2.1 Variables expliquant les dépenses

La base de données de Statistiques Canada contient une multitude de variables pouvant expliquer le niveau des dépenses de chaque ménage et ce tels que le revenu, le nombre des dépendants, l'âge, le sexe ... Même les informations les plus détaillées sont présentes, par exemple nous pouvons connaître le nombre de voitures achetées, l'utilisation d'internet ainsi que l'utilisation d'un téléphone portable.

SM. Finlay inclut plusieurs variables explicatives dans son modèle. Il inclut entre autres toutes les variables que nous venons de citer. Néanmoins, nous ne pouvions répliquer totalement son modèle car la base de données principale à laquelle nous devons appliquer le modèle ne contenait pas toutes ces variables. Malheureusement, les seules variables en commun entre ces deux échantillons pouvant être employées pour l'estimation des dépenses étaient au nombre de cinq. Nous citons l'âge, le revenu, le sexe, le nombre de dépendants et enfin le type du logement (propriétaire ou locataire).

En réalité elles ne sont que quatre variables vu que la base de données de Statistiques Canada a été déjà divisée en deux parties : un échantillon pour les femmes et un deuxième pour les hommes. La variable relative au sexe ne pourra être incluse dans les régressions.

- **Le revenu**

Le revenu est un élément qui influence nécessairement le niveau des dépenses des individus. De manière générale, plus le revenu d'une personne est élevé, meilleur est son niveau de vie et plus élevées sont ses dépenses. Dans son article, Finlay a fortement insisté sur l'utilisation du revenu lors de l'estimation des dépenses. Il a également fourni plusieurs explications quant aux éléments devant être inclus dans le calcul de cette variable. Selon l'auteur, toutes les sources de revenus dont dispose la personne doivent être intégrées. Ceci comprend les revenus d'investis-

sement, de placements, les avantages sociaux et bien sur le revenu d'emploi. Ces revenus devraient néanmoins être nets d'impôt.

Comme nous l'avons expliqué précédemment, les revenus inclus dans la base de données principale (l'échantillon de la banque) sont des revenus bruts. Les revenus utilisés dans la base de données de Statistiques Canada sont aussi des revenus bruts. Nous avons donc utilisé les revenus bruts au lieu des revenus nets proposés par l'auteur. Ceci n'a pas affecté nos résultats puisque les impôts ont été intégrés dans les dépenses et seront donc soustraits au revenu lors de la construction de l'indice de capacité à payer.

Par ailleurs, Finlay a suggéré de convertir la variable revenu qui est à l'origine une variable continue, en variables dichotomiques. Il a utilisé vingt variables dichotomiques pour définir les vingt catégories de revenus, chaque catégorie comprenant ainsi 5% de la population totale. Ce changement a été motivé par la nature de la variable dépendante utilisée dans ce modèle par l'auteur. Étant une mesure relative prenant des valeurs souvent entre zéro et un, l'utilisation de variables dichotomiques (qui prennent aussi les valeurs zéro et un) au lieu de variables continues s'avère plus intéressante.

Dans notre modèle nous avons considéré la variable dépenses en tant que valeur absolue et qui peut donc prendre des valeurs éloignées de zéro et de un. De plus, intégrer des sous classes de revenus dans notre modèle ne nous fournirait pas d'informations supplémentaires. Nous avons donc préféré garder la variable revenu telle qu'elle était.

En conclusion, la variable revenu inclura toutes les sources de revenus disponibles dans la base de données, sur une base mensuelle, et sera une variable continue qui prendra des valeurs strictement positives. Cette variable est supposée influencer positivement les dépenses.

- **L'âge**

Dans la littérature, l'âge est un élément qui influence légèrement le niveau des dépenses. Finlay a intégré cette variable en la divisant en deux variables dichotomiques : la première prenant la valeur un si l'âge est inférieur à vingt et un ans et la deuxième concerne les personnes âgées de plus de vingt et un ans. Il a intégré la première variable dans sa régression (SM. Finlay, 2006).

Dans notre étude nous avons divisé cette variable en six catégories, chaque catégorie ayant une dichotomique pour la représenter. Les catégories sont définies comme le montre le tableau 3 :

TAB. 3 – Catégories d'âge

catégorie d'âge	description
Age1	les individus âgés de moins de 25 ans
Age2	ceux âgés entre 25 et 35 ans
Age3	ceux âgés entre 35 et 45 ans
Age4	ceux âgés entre 45 et 55 ans
Age5	ceux âgés entre 55 et 65 ans
Age6	ceux âgés de plus de 65 ans

Cinq de ces variables ont été intégrées dans le modèle pour éviter la colinéarité. L'influence de cette variable sur les dépenses demeure ambiguë, plusieurs hypothèses peuvent être formulées, la plus vraisemblable étant l'augmentation des dépenses avec l'âge.

- **Les dépendants**

Il est évident qu'avoir des personnes à charge est un élément qui influence réellement le niveau des dépenses d'un individu. Un parent ayant des enfants dépense nécessairement plus qu'une personne sans enfants. Les dépenses liées à l'éducation, l'habillement, l'alimentation... augmentent considérablement. C'est pour cette raison qu'il est indispensable d'intégrer cette information dans le modèle. Par souci de simplicité, cette donnée est représentée par une variable dichoto-

mique qui nous informe si la personne en question a des personnes à charge ou non.

$$\text{dépendant} = \begin{cases} 1 & \text{si présence de personnes à charge} \\ 0 & \text{sinon} \end{cases}$$

- **Le type de logement**

Le type de logement signifie ici le mode d'occupation du logement, si c'est une propriété de l'individu ou un logement loué. Généralement, un individu qui est propriétaire de sa maison doit payer des taxes additionnelles au gouvernement, les dépenses liées à l'entretien de sa maison et de son jardin ainsi que des primes d'assurances en plus. Un propriétaire a donc plus de dépenses qu'un simple locataire. Ceci nous confirme l'importance de cette information dans l'estimation des dépenses. Cette variable est représentée dans notre modèle par une variable dichotomique (propriétaire) qui indique si l'individu en question est propriétaire de son logement ou non.

$$\text{propriétaire} = \begin{cases} 1 & \text{si l'individu est propriétaire} \\ 0 & \text{sinon} \end{cases}$$

- **Le sexe**

Le sexe est aussi une information qui pourrait influencer l'estimation des dépenses. Néanmoins Finlay a préféré ne pas tenir compte de cette donnée dans son estimation. Après réflexion, nous avons décidé d'en tenir compte mais pas en l'intégrant en tant que variable explicative dans notre modèle mais plutôt en divisant l'échantillon en deux parties distinctes et en construisant deux modèles différents : un concernant les femmes et le deuxième concernant les hommes. Cela nous permet de tenir compte des différences existantes en matière de dépenses entre les deux sexes.

- **Autres variables**

Comme nous l'avons cité antérieurement, il existe plusieurs autres variables pouvant influencer les dépenses. Cependant, nous allons dans un premier temps nous contenter des quatre variables énoncées et par la suite intégrer d'autres variables pour une éventuelle extension de l'étude.

4.2.2 Variables explicatives du défaut

Plusieurs articles dans la littérature traitent du défaut des particuliers exposant ainsi une série de variables pouvant expliquer ce phénomène. Vu la multitude des variables pouvant être incluses dans le modèle, nous avons d'abord testé la significativité de la majorité des variables proposées, et nous avons par la suite choisi les variables significatives par rapport à l'estimation du défaut et exclu celles non significatives. Les variables sélectionnées sont :

- **La cote de risque**

La cote de risque représente le score que la banque attribue au client en analysant sa demande de financement ainsi que sa situation financière. Ce score est associé à la probabilité de défaut estimée du client. Il est aussi considéré comme étant l'élément principal sur lequel se base la banque pour décider si oui ou non elle accorde le prêt.

Il existe deux types de cote de risque, la première est une cote externe élaborée par un bureau de crédit et la deuxième est une cote interne souvent établie à partir de la première cote en utilisant un modèle propre à l'institution en question.

Bien que la cote externe soit un bon indicateur de la probabilité de défaut, nous avons préféré utiliser la cote interne, puisqu'elle englobe le score externe ainsi que d'autres informations.

Cette cote prend les valeurs de 1 à 9 selon le risque que présente l'individu. Plus la valeur de la cote est élevée plus le risque que présente l'individu analysé est élevé aussi. Cette variable influencera donc positivement la probabilité de défaut.

- L'âge

Comme dans l'autre modèle, l'âge est une variable essentielle pour l'estimation du défaut. Elle est présente dans quasiment tous les modèles traitant de la probabilité de défaut.

Thomas (2000) affirme dans son article l'influence de cette variable sur la probabilité de défaut. Il démontre cette tendance dans la figure 5.

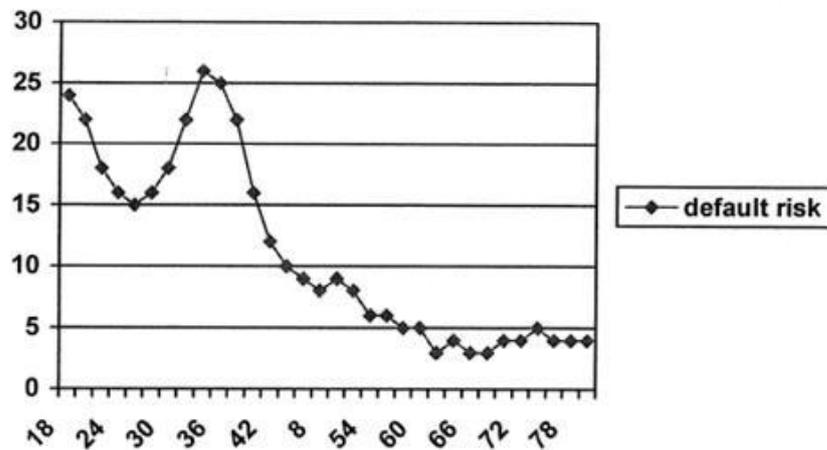


FIG. 5 – L'impact de l'âge sur le risque de défaut

Cette variable sera donc incluse dans notre modèle en tant que variable dichotomique avec la même répartition que celle présentée pour le modèle d'estimation des dépenses. Elle aura éventuellement un effet ambigu mais globalement négatif sur la probabilité de défaut.

- **Le nombre de lignes de crédits**

Le nombre de crédits ouverts reflète en quelque sorte l'état financier de l'individu. Ceci nous donne une image sur le niveau d'endettement du client et donc sur les engagements financiers que ce client doit honorer. Nous avons exposé auparavant l'éventuelle relation qui existe entre le niveau d'endettement et le risque de défaut de l'individu, il est donc évident que cette variable influence positivement la probabilité de défaut.

Cette information est représentée par une variable discrète positive appelée CREDITACTIF qui regroupe les prêts et les marges de crédit ouverts au nom du client.

$$CREDITACTIF = \begin{cases} 0 & \text{si aucun prêt ouvert} \\ \text{Nombre de prêts ouverts} & \text{sinon} \end{cases}$$

Notons que, dans la base de données principale, le nombre de prêts ouverts varie entre 0 et 7 prêts au maximum pour chaque client.

- **Montant du prêt**

Le montant du prêt est représenté ici par une variable continue (Montant), qui prend des valeurs strictement positives. Elle est supposée influencer positivement la probabilité de défaut. Plus le montant à rembourser est élevée plus la probabilité que le client ne rembourse pas la totalité de ce prêt augmente.

- **Solde**

Le solde signifie ici le total des engagements financiers de la personne au moment de l'obtention du prêt. Cette variable nous donne une image sur le niveau d'endettement de l'individu. Cette information est également représentée par une variable continue strictement positive qui aurait un impact positif sur la probabilité de défaut.

- **Type de travail**

Cette variable nous fournit l'information quant au statut de l'individu au sein de son travail, c'est-à-dire s'il est propriétaire de sa compagnie ou un salarié. Nous utiliserons ici une variable dichotomique qui prend les valeurs zéro ou un et sera définie comme suit :

$$\textit{salarie} = \begin{cases} 1 & \textit{si l'individu est salarié} \\ 0 & \textit{sinon} \end{cases}$$

L'effet de cette variable sur le risque de défaut de l'individu demeure mitigé. D'une part une personne ayant sa propre compagnie est supposée disposer de plus de fonds pour payer ses dettes qu'un salarié. D'autre part, cette même personne a plus de risque de rencontrer des problèmes financiers (crise du marché par exemple) et a plus de responsabilités financières et aurait donc une probabilité élevée de faire de défaut.

- **Ancienneté du client**

Cette variable mesure la période pendant laquelle l'individu a été un client de la banque et ce au moment de l'obtention du prêt. C'est une variable discrète positive calculée en fonction du nombre de mois d'ancienneté. L'effet de cette variable est supposé négatif, car un client qui a accumulé des années d'ancienneté au sein de sa banque devrait être plus fidèle dans ses engagements vis-à-vis de cette même banque.

- **Garantie**

La variable garantie nous informe si le client a fourni une garantie à son prêt ou non. Un prêt qui est octroyé avec garantie est supposé plus sûr qu'un prêt sans garantie. Cette variable influencerait négativement la probabilité de défaut liée au prêt. Elle sera intégrée donc dans notre modèle et sera représentée par une variable dichotomique qui est définie comme suit :

$$Garantie = \begin{cases} 1 & \text{si présence de garantie} \\ 0 & \text{sinon} \end{cases}$$

- **Ancienneté du prêt**

Nous allons reprendre ici la variable développée par Dionne et al qui mesure l'ancienneté du prêt. Cette variable nommée DUREEA mesure la période (en mois) s'étalant du moment de l'ouverture du prêt au moment de l'extraction des données pour ceux qui n'ont pas fait défaut, ou au moment du défaut pour le groupe ayant déjà fait défaut (Dionne et al, 1996).

- **Variables relatives à la capacité à payer**

Le sujet de notre mémoire se base sur l'intégration des variables de la capacité à payer dans le processus de décision de la banque. Cela implique, entre autres, la prise en compte de ces variables lors de l'estimation de la probabilité de défaut.

Nous avons longuement réfléchi à cette éventualité et à la meilleure méthode pour la réaliser. Nous avons exposé antérieurement les différents problèmes rencontrés dans la littérature et la difficulté d'établir un lien clair et direct entre la capacité à payer d'un individu et sa probabilité de défaut.

Après réflexion, nous avons décidé d'intégrer cette information en tant qu'indicateur qui prend en considération toutes les variables pouvant influencer cette donnée. Nous avons donc construit une variable sous forme d'indicateur en prenant comme référence l'indicateur de LC. Thomas (2005) et celui de SM Finlay (2006). Quelques rectifications ont été également ajoutées à ces définitions afin d'obtenir une cote capable de refléter la capacité du client à payer le prêt qu'il demande.

Cet indicateur est défini comme suit :

$$\textit{Capacité} = \textit{revenu} + \textit{tangibles}_r - \textit{dépenses estimées}$$

Il est à noter que la capacité à payer calculée est une capacité mensuelle. Nous allons expliquer cette définition plus en détail dans la partie méthodologie.

Par ailleurs, nous avons construit un autre indicateur à partir de ce dernier que nous avons nommé « écart », et qui compare la capacité à payer calculée avec le paiement mensuel du prêt accordé à l'individu étudié. C'est celui que nous allons utiliser dans notre modèle servant à estimer le défaut. Cet indicateur a été divisé en cinq catégories, chaque catégorie comprenant environ 20% de la population. Ces catégories seront par la suite représentées au moyen de cinq variables dichotomiques, la première catégorie représentant tous ceux dont le paiement mensuel excède leur capacité à payer mensuelle calculée. La dernière catégorie comprend quant à elle ceux dont la capacité à payer mensuelle est beaucoup plus élevée que leur paiement mensuel du prêt. Nous nous attendons à ce que cette variable influence négativement la probabilité de défaut. Les clients ayant une capacité à payer plus élevée que leur paiement mensuel auraient moins tendance à faire défaut et vice versa. Le calcul ainsi que l'explication de ces deux indicateurs seront étendus dans la section 6.2 du mémoire.

5 Statistiques descriptives

Dans cette section nous allons présenter les statistiques descriptives des variables utilisées dans les deux modèles ainsi que les comparaisons intéressantes entre les différents groupes d'individus.

5.1 Moyenne des dépenses selon les personnes à charge et le type de logement

L'évolution des dépenses par rapport à quelques variables est très importante dans notre étude. Cela nous aide à connaître le sens dans lequel chacune des caractéristiques de l'individu influence le niveau de ses dépenses. Nous avons donc tenté de visualiser la différence entre les dépenses de quelques groupes d'individus.

Premièrement, nous avons établi la relation entre la présence de personnes à charge et le niveau de dépense. Le tableau 4 nous fournit un aperçu de l'impact de cette variable sur les dépenses.

TAB. 4 – L'impact de la présence des personnes à charge sur les dépenses

Dépendant	Nb	Nb cumulatif	Moyenne des dépenses
0	1656	1656	2675.57
1	256	1912	3314.94

Rappelons que la variable dépendant est une variable dichotomique qui représente la présence ou non de personne à charge. Cette analyse confirme l'hypothèse selon laquelle la présence de personnes à charge augmenterait les dépenses de l'individu. Les individus ayant des personnes à charge dépensent en moyenne 3315 CAD, ce qui est bien plus élevé que les dépenses moyennes des personnes seules qui sont de 2675 CAD.

Si on s'attarde sur la relation entre le type de logement et le niveau des dépenses, on remarque qu'un propriétaire dépense nettement plus qu'un locataire. Ceci est démontré

dans le tableau 5

TAB. 5 – L’impact du type de logement sur les dépenses

Propriétaire	Nb	Nb cumulatif	Moyenne des dépenses
0	1068	1068	2200
1	844	1912	3471

En conclusion, nous pouvons dire que les deux variables citées ont un impact important sur le niveau des dépenses de l’individu et c’est la raison pour laquelle elles sont intégrées dans le modèle de prévision.

5.2 Comparaison des deux bases de données

L’objectif de cette partie est de comparer les deux bases de données afin de s’assurer que les deux populations sont distribuées de manière quasiment identique. Ceci nous permettra par la suite d’appliquer le modèle d’estimation des dépenses construit à partir de la première base de données à notre base de données principale.

Le tableau 6 présente les statistiques descriptives des deux échantillons par rapport aux variables importantes. Ces statistiques ont été calculées à partir de l’échantillon corrigé mais non encore pondéré.

En observant le tableau 6, nous remarquons que la population n’est pas distribuée de manière identique au sein de ces deux échantillons. Une pondération de l’échantillon de Statistiques Canada est donc nécessaire afin de pouvoir utiliser le modèle construit à partir de cet échantillon dans la base de données principale.

Le tableau 7 présente les statistiques descriptives des mêmes variables après pondérations. Nous remarquons que la population du premier échantillon ressemble davantage à celle de l’échantillon principal. Nous avons gardé cet échantillon pour la construction du modèle d’estimation des dépenses.

TAB. 6 – Comparaison des deux bases de données (avant pondération)

Variabiles	Description	BD principale	BD StatCan
RÉVENU	Revenu brut mensuel moyen	3400	3080
DÉPENSES	Dépenses mensuelles moyennes	280	2600
AGE	Age moyen de la population	40	38
PROPRIÉTAIRE	Pourcentage des propriétaires	58%	48%
DÉPENDANT	Pourcentage des personnes avec dépendants	8%	18%
HOMME	Pourcentage d'hommes	66%	40%
Total		28312	3476

TAB. 7 – Comparaison des deux bases de données (après pondération)

Variabiles	Description	BD principale	BD StatCan
RÉVENU	Revenu brut mensuel moyen	3400	3310
DÉPENSES	Dépenses mensuelles moyennes	280	2759
AGE	Age moyen de la population	40	39
PROPRIÉTAIRE	Pourcentage des propriétaires	58%	50%
DÉPENDANT	Pourcentage des personnes avec dépendants	8%	10%
HOMME	Pourcentage d'hommes	66%	63%
Total		28312	1912

Par ailleurs, nous avons présenté la moyenne des dépenses dans les deux populations afin de souligner la grande différence qui existe entre ce que déclare les clients comme dépenses dans leur demande de financement et la réalité. Ceci explique la nécessité de remplacer la variable dépense dans la base de données principale par une variable estimée selon les caractéristiques de chaque individu.

5.3 Comparaison des taux de défaut selon la capacité à payer

Une fois que le modèle d'estimation des dépenses a été développé, nous l'avons appliqué à la base de données principale afin d'estimer les dépenses. Nous avons par la

suite utilisé les dépenses estimées pour calculer notre indice de capacité à payer, qui à son tour a servi au calcul de la variable écart. Notre objectif étant de définir l'impact de la capacité à payer sur la probabilité de défaut, il serait donc intéressant de visualiser la distribution des défauts selon les catégories d'écart, ainsi que le taux de défaut pour les individus ayant une capacité à payer négative.

Le tableau 8 expose le taux de défaut moyen pour les personnes ayant une capacité à payer négative.

TAB. 8 – Taux de défaut pour les individus avec une capacité à payer négative

DÉFAUT	Fréquence	Pourcentage	Freq cumulée	Pourc cumulée
0	1863	96,58%	1863	96,58%
1	66	3,42%	1929	100%

Sachant que le taux de défaut pour le groupe ayant une capacité à payer positive est de 1.6%, nous pouvons dire que les individus ayant une capacité à payer négative se caractérisent par un taux de défaut moyen 2 fois plus élevé. La capacité à payer influence donc négativement la probabilité de défaut. Ce qui confirme notre hypothèse selon laquelle la capacité à payer de l'individu a nécessairement un impact sur son risque de défaut qui s'avère maintenant être négatif.

Le tableau 9 nous démontre également que les individus ayant un paiement de prêt mensuel plus élevé que leur capacité à payer ont également un taux de défaut plus élevé.

TAB. 9 – Taux de défaut pour les individus avec un remboursement mensuel supérieur à leur capacité à payer

DÉFAUT	Fréquence	Pourcentage	Freq cumulée	Pourc cumulée
0	3308	96,36%	3308	96,36%
1	125	3,64%	3433	100%

L'analyse du taux de défaut dans chacune des catégories d'écart nous confirme également ce constat. Comme nous pouvons le voir dans le tableau 10, le nombre des défauts baisse considérablement à travers les cinq catégories d'écart. La première catégorie d'écart, qui comprend les individus ayant un remboursement mensuel plus élevé que leur capacité à payer, se caractérise par un nombre de défaut relativement élevé et qui atteint 191 défauts. Parallèlement, la dernière catégorie d'écart qui concerne ceux ayant une capacité à payer nettement plus élevé que leur remboursement, ne contient que 53 défauts.

TAB. 10 – Taux de défaut pour les cinq catégories d'écart

Catégories d'écart	Nombre	Somme des défauts	Moyenne des défauts
1	5358	191	3,56%
2	5826	163	2,79%
3	5702	85	1,49%
4	5705	72	1,26%
5	5721	53	0,92%

Le taux défaut baisse considérablement passant de 3.56% pour le premier groupe à 0.92% pour le dernier groupe. Nous pouvons donc prévoir une influence positive de cet écart sur la probabilité de défaut (si la catégorie 5 est prise comme référence).

5.4 Comparaison du groupe des défauts au groupe sans défaut

Dans cette section nous comparons, au sein de la base de données de la banque, les caractéristiques des prêts ayant fait défaut avec ceux n'ayant pas fait défaut. Le tableau 11 nous expose clairement les différences entre ces deux groupes.

Il existe des différences très significatives entre le groupe des défauts et celui des non défauts. Premièrement, le groupe des défauts a une cote de risque interne moyenne de trois contrairement au deuxième groupe qui se caractérise par une cote moyenne de cinq. Les prêts ayant fait défaut ont donc été définis par le modèle interne de la

TAB. 11 – Comparaison des moyennes : défaut vs non défaut

Variables	Description	Groupe des non défauts	Groupe des défauts	Test t
SCORE	cote de risque externe (entre 0 et 1011)	635	375	<.0001
COTRISK	cote de risque interne (de 1 à 9)	3	5	<.0001
REVENU	Revenu brut mensuel	3485	3221	<.0001
DÉPENSEP	Dépenses mensuelles estimées	2836	2655	<.0001
RATIODEP	= dépensep / Revenu	0.81	0.85	.0044
CAPACITÉ	= revenu + tangibler - dépensep	1333	863	<.0001
SOLDE	Solde total des crédits et marges détenus	14463	15865	.05
TANGIBLE	= actif total - passif total	34172	11155	0.0019
TANGIBLER	= tangible / terme du prêt	684	297	0.0023
ATD	= (paiement du prêt + (engagements financiers mensuels * 100)) / (revenu mensuel brut)	30	32	0.0002
AGE	Âge du client	41	38	<.0001
ANCIEMEMB	Ancienneté du client (en mois)	186	132	<.0001
TAUX	Taux d'intérêt appliqué	9	10	<.0001
DUREEA	Âge du prêt	35	30	<.0001

banque comme étant plus risqués que les autres ce qui s'est matérialisé à travers un taux d'intérêt appliqué plus élevé. Cette variable aurait donc un effet positif significatif sur la probabilité de défaut.

Nous remarquons également que le revenu des individus au sein du groupe des défauts est significativement moins élevé que celui des individus dans l'autre groupe. Le ratio des dépenses relatives au revenu nous confirme ce constat puisque ce ratio est moins élevé chez le groupe des non défauts. Par ailleurs la capacité à payer est aussi significativement plus élevé chez le groupe des non défauts. Ce qui indique que cette variable influencerait l'estimation du défaut dans le sens négatif.

Les individus ayant fait défaut sont plus endettés que les autres vu que leur solde de crédits est significativement plus élevé. De plus, l'ATD (Amortissement Total de la Dette) du groupe des défauts est supérieur à celle qui caractérise le groupe des non défauts. Cela confirme notre hypothèse selon laquelle un individu surendetté se caractérise par un risque de défaut plus élevé. Ces deux variables (ATD et Solde) sont supposées influencer positivement la probabilité de défaut.

Les clients plus anciens et plus âgés sont moins sujets à des défauts, cela confirme l'hypothèse selon laquelle la variable représentant l'ancienneté du client influencera négativement la probabilité de défaut.

Enfin, le groupe des non défauts ont une valeur tangible moyenne nettement plus élevée que celle du groupe des défauts. Cela implique deux choses : soit que les membres du premier groupe ont plus d'actifs à leur compte ce qui les aide à toujours honorer leurs engagements financiers, soit que ceux du deuxième groupe ont beaucoup de dettes à leur passif ce qui rend leur remboursement plus difficile.

6 Description des modèles

Dans ce mémoire, nous avons fait appel à deux modèles distincts. Chacun de ces modèles sert à estimer une variable différente. Le premier modèle a été construit afin d'estimer les dépenses. Nous avons ensuite utilisé ce modèle afin de corriger la variable dépense dans notre base de données principale et pouvoir ainsi calculer notre indice de capacité à payer et construire une nouvelle variable nommée « écart ». Cette variable a été intégrée dans le deuxième modèle qui sert à estimer la probabilité de défaut.

Nous allons dans un premier temps présenter le modèle d'estimation des dépenses, pour ensuite expliquer le calcul de notre indice de capacité à payer et de la variable écart, et enfin exposer le modèle que nous avons utilisé pour estimer la probabilité de défaut.

6.1 Modèle d'estimation des dépenses

La problématique essentielle que nous avons rencontrée lors de la construction de l'indice de capacité à payer à partir de notre base de données principale a été le manque d'information concernant les dépenses. La variable dépense dans cet échantillon était mal documentée et ne reflétait nullement le niveau réel des dépenses des individus.

Pour pallier à ce manque d'informations nous avons adopté la méthode utilisée par Finlay et qui consistait en la construction d'un modèle d'estimation des dépenses à partir d'une autre base de données.

Comme l'auteur, nous avons utilisé la méthode des moindres carrés ordinaires (MCO) pour estimer ces dépenses. Notre objectif étant de définir l'impact de certaines variables explicatives (l'âge, le revenu, le type de logement et les personnes à charge) sur la variable dépendante Y (les dépenses).

L'équation générale utilisée lors de cette régression s'exprime comme suit :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

Où i représente les ménages avec $i = 1, 2, 3 \dots N$

les β_k sont les coefficients estimés lors de la régression.

Les variables explicatives utilisées représentent celles précédemment citées dans la section 4.2.1. La composition du modèle changera selon les variables explicatives incérées. Cette méthode suppose également l'absence d'hétéroscédasticité et d'auto-corrélation entre les résidus. Des tests ont été effectués afin de vérifier ces hypothèses et quelques corrections ont été apportées.

Nous avons également testé la corrélation entre les quatre variables utilisées dans la régression, corrélation qui s'est avérée assez faible et permettant ainsi l'utilisation de ces variables dans notre modèle.

6.2 Construction de l'indice de capacité à payer

Thomas (2005) ainsi que Finlay (2006) nous ont exposé deux définitions complémentaires de la capacité à payer. Nous avons utilisé ces deux informations afin de construire un nouvel indice qui ne tenait pas compte seulement des actifs et passifs de l'individu, mais aussi des conditions du prêt qu'il demande.

Notre but étant d'obtenir un indice de capacité à payer mensuel, qui reflète le montant maximum que le client peut payer par mois. Il fallait donc prendre en considération toutes les sources de revenus possibles, les actifs liquides ainsi que les engagements financiers et les dépenses du client.

Concernant les revenus, nous avons utilisé le total des revenus brut mensuels, comme nous l'avons expliqué dans la section décrivant les variables.

Les dépenses ont été estimées grâce au modèle construit auparavant et qui prenait

en compte quatre caractéristiques du client. Ces dépenses étaient aussi mensuelles et incluaient les impôts.

Les actifs liquides et les engagements financiers étaient tous les deux rassemblés dans une seule variable nommée actif tangible. Cette variable est considérée comme étant la différence entre tous les actifs liquides et tous les passifs du client et ce au moment de l'obtention du prêt. Elle comprend donc deux informations importantes relatives à l'état financiers de l'individu : les actifs immédiatement liquides qui peuvent être utilisés dans le remboursement du prêt en cas de difficulté financière et les engagements financiers qui peuvent constituer un obstacle au paiement du prêt. Ces deux données sont indispensables au moment du calcul de la capacité à payer. Cependant, le problème de cette variable résidait dans le fait qu'elle n'avait pas de dimension temporelle. C'est une variable absolue qui résulte d'additions et de soustractions d'éléments ayant chacun une dimension temporelle différente. Il fallait donc trouver un moyen qui nous permettrait de convertir cette donnée en donnée mensuelle.

Après réflexion, nous avons trouvé une solution qui non seulement donnait une dimension temporelle mensuelle à cette variable mais qui créait aussi un lien direct avec le prêt demandé. Nous avons décidé de diviser la variable 'actifs tangibles' par le terme du prêt demandé en mois. Il est clair que cette durée ne correspond pas à la durée des éléments qui constituent cette variable. Cependant, nous avons considéré cette variable comme une information sur l'actif net de l'individu, actif net qui peut être utilisé pour le remboursement du prêt en cas de difficulté financière.

Les actifs tangibles représentent en réalité les actifs liquides disponibles du client après soustraction de tous les autres engagements. Ils comprennent l'épargne à terme, les placements financiers, les liquidités . . . Thomas et Finlay incluent cette information dans le calcul de la capacité à payer car ils estiment qu'un individu faisant face à une situation financière difficile pourrait vendre ses placements boursiers ou réclamer une partie de son épargne pour payer ses engagements financiers. Diviser ces actifs par le terme du prêt nous permettra de connaître la portion que pourrait remplir ces actifs tangibles dans le remboursement du prêt tout au long de son terme. En additionnant

cette portion au revenu mensuel et en retranchant les dépenses nous aurions la capacité totale du client pendant toute la durée du prêt. Ceci bien sur ne prendra pas en compte les éventuels changements dans les revenus ou les dépenses, tels qu'une augmentation ou une baisse du salaire, ou l'ajout d'un autre prêt. Néanmoins, nous aurions l'information sur la capacité à payer du client au moment de la demande du prêt.

La formule de la capacité à payer sera donc comme suit :

$$\text{Capacité} = \text{revenu} + \text{tangible}_r - \text{dépenses estimées}$$

sachant que :

$$\text{tangible}_r = \frac{\text{actifs tangibles}}{\text{terme du prêt}}$$

Cette définition nous informe sur le montant mensuel dont disposera le client tout au long du terme du prêt, incluant ces actifs liquides et soustraction faite de ses engagements financiers, et ce pour rembourser son prêt selon le terme accordé.

De plus, cette définition aidera également la banque à définir les conditions du prêt, après que celui-ci soit accordé. Nous utiliserons un exemple pour expliquer cette dernière information.

Exemple Un client ayant un revenu mensuel brut de 3000 CAD, des dépenses estimées de 2000 CAD et des actifs tangibles totaux de 30000 CAD. Il demande un prêt de 50000 CAD qui lui est accordé après avoir vérifié que sa probabilité de défaut ne dépassait pas le seuil prédéterminé. Reste à définir le terme et le montant mensuel du prêt. Le client demande un terme de 5 ans (et donc 60 mois) par exemple et la banque calcule sa capacité à payer à partir de ce terme, elle obtient donc :

$$\text{Capacité} = 3000 + \frac{30000}{60} - 2000 = 1500$$

Ce client dispose donc d'une capacité à payer mensuelle maximale de 1500 CAD sur une période de 60 mois. Ce qui veut dire que, au moment de la demande du prêt,

ce client peut honorer un engagement financier allant jusqu'à 1500 CAD par mois sur une période de 5 ans.

La banque pourrait donc lui prêter au maximum un total de $1500 \times 60 = 90000$ CAD comprenant le capital et les intérêts et ce pour le terme que le client a demandé. Vu que, pour cet exemple, le montant maximum dépasse le montant demandé, la banque pourrait donc suggérer au client de prendre un terme plus court ou de lui prêter moins que sa capacité à payer maximale.

En général, la deuxième éventualité est la plus courante et la plus recommandée, et ce pour éviter qu'un changement imprévu (négatif) de la situation financière du client n'affecte le remboursement de son prêt. Il faudrait donc définir un pourcentage de la capacité à payer qui serait attribué au remboursement mensuel du prêt et qui éviterait le défaut du client.

Ceci nous amène à la deuxième étape, qui est la construction de la variable écart. Cette variable représente l'écart existant entre la capacité à payer mensuelle d'un individu et le paiement mensuel de son prêt qui lui a été accordé par la banque. Il est à noter que c'est un écart relatif (en pourcentage) qui est exprimé sous la forme :

$$Ecart = \frac{\text{paiement mensuel du prêt}}{\text{capacite à payer mensuelle}}$$

L'écart est donc le pourcentage de la capacité à payer destiné au paiement du prêt. En général cet écart ne devrait pas dépasser la valeur 1, qui veut dire que le paiement mensuel devrait être au plus égal à la capacité à payer maximale ce qui est déjà assez risqué.

Cet écart ne représente pas un simple ratio d'endettement car il comprend les informations relatives aux dépenses estimées de la personne ce qui n'est pas le cas de ce dernier. De plus, un ratio d'endettement inclut en général les actifs non liquides tels que les biens immobiliers, alors que la variable écart ne comprend que les actifs immédiatement liquides.

L'intérêt de cette variable réside dans son caractère comparatif, c'est à dire qu'elle nous permet de comparer le remboursement mensuel du prêt accordé par la banque et la

capacité à payer calculée de l'individu. Nous avons pu ainsi déterminer l'impact de cet écart sur le comportement du client.

Afin de découvrir l'effet des différents pourcentages (ou seuils) sur la probabilité de défaut et ainsi définir un seuil qui n'influencerait pas positivement le défaut, nous avons divisé cette variable en cinq catégories comme nous l'avons expliqué dans le chapitre précédent. Ces catégories comprennent cinq groupes d'individus répartis selon l'écart relatif entre le paiement de leur prêt et leur capacité à payer calculée. Ceci nous permettra de connaître par la suite à partir de quel seuil cet écart influence positivement le défaut. Cette information aurait été difficile à cerner si nous avions laissé cette variable en tant que variable continue. Le tableau suivant fournit les informations quant aux catégories et aux pourcentages utilisés.

Le tableau 12 présente les cinq catégories d'écart.

TAB. 12 – Catégories d'écart

Catégories d'écart	Description
Ecart1	si l'écart < 0 ou si l'écart > 1
Ecart2	si $0,424 < \text{écart} < 1$
Ecart3	si $0,241 < \text{écart} < 0,424$
Ecart4	si $0,134 < \text{écart} < 0,241$
Ecart5	si $0 < \text{écart} < 0,134$

6.3 Estimation de la probabilité de défaut

La modélisation de la probabilité de défaut demande un traitement différent de celui opéré précédemment lors de la modélisation des dépenses. La variable dépendante étant une variable dichotomique, nous ne pouvons plus utiliser la méthode des moindres carrés ordinaires mais une méthode qui peut tenir compte du caractère binaire de cette variable. Dans la littérature, les auteurs utilisent souvent l'un de ces deux modèles : le modèle Logit ou le modèle Probit. Les deux modèles sont presque identiques, la seule différence étant l'épaisseur des queues du modèle Logit par rapport au Probit ainsi que

la loi de distribution des résidus.

Dans le cas présent, nous avons choisis le modèle Logit afin de modéliser notre probabilité de défaut. Cette méthode suppose un modèle latent sous la forme suivante :

$$y_i^* = x_i' \beta + \varepsilon_i$$

y_i^* étant une variable latente non observable, β le vecteur des coefficients estimés lors de la régression et enfin x_i' le vecteur des variables explicatives du modèle.

Ainsi qu'un modèle observable tel que :

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

Sachant que y_i est notre variable dépendante qui représente le défaut.

Le modèle Logit suppose une distribution symétrique des résidus par rapport à zéro comme le stipule la loi logistique. Ainsi la probabilité qu'un client soit en défaut peut être exprimée comme suit :

$$P(Y = 1) = F(x_i' \beta) = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

où x_i' est le vecteur qui inclut toutes les variables explicatives que nous avons utilisées lors de la régression. Ces variables ont été énumérées et décrites dans les sections 4 et 5 du mémoire. Nous rappelons l'intégration de la variable écart autant que variables dichotomiques dans ce modèle. Il est à noter que nous avons divisé les variables continues telles que solde et montant par 1000 et ce pour obtenir des coefficients plus convenables.

7 Analyse des résultats

7.1 Estimation des dépenses

Le modèle utilisé afin d'estimer les dépenses se présente comme suit :

$$\begin{aligned} \text{Dépenses} = & \beta_0 + \beta_1 \times \text{revenu} + \beta_2 \times \text{dépendant} + \beta_3 \times \text{propriétaire} + \\ & + \beta_4 \times \text{Age1} + \dots + \beta_9 \times \text{Age6} + \varepsilon_i \end{aligned}$$

Dans cette régression, nous avons utilisé la troisième catégorie d'âge comme référence.

Nous tenons à rappeler que notre base de données a été divisée en deux parties : un échantillon féminin et un échantillon masculin. Cette répartition était une solution pour faciliter la pondération de l'échantillon, et qui nous a permis également de tenir compte des différences pouvant exister entre le niveau de dépenses chez les hommes et chez les femmes. Il est à noter que nous avons utilisé 70% de chaque échantillon pour la régression principale, et les 30% restants pour la validation du modèle.

Le tableau 13 nous présente les résultats de la régression pour l'échantillon féminin :

TAB. 13 – Régression MCO sur l'échantillon féminin

Variables	Coefficient	Erreur standard	test t	Pr > t
Constante	208.5353	61.059	3.42	0.0007
REVENU	0.7257	0.0121	59.47	<.0001
DEPENDANT	274.575	61.202	4.85	<.0001
PROPRIO	201.167	53.225	3.41	0.0002
AGE1	183.0773	104.1290	1.76	0.0717
AGE2	62.5720	63.398	0.98	0.3269
AGE4	143.763	62.6239	2.20	0.0213
AGE5	286.32	84.6121	3.32	0.0007
AGE6	-37.7735	148.477	-0.25	0.7890
R^2 ajusté	0.83			
Nombre d'observations	640			

Nous remarquons que toutes les variables sont très significatives pour expliquer les dépenses sauf l'âge qui s'avère peu significatif.

L'influence des variables sur le niveau des dépenses rejoint nos prévisions. Le revenu a un impact positif sur le niveau des dépenses, ce qui est logique, car plus le revenu d'un individu est élevé, plus il a tendance à consommer et à élever son niveau de vie : plus belle maison, plus belle voiture, plus de loisirs...

Par ailleurs, ce résultat est conforme aux conclusions de Finlay décrivant la relation entre le revenu et les dépenses (SM. Finlay, 2006).

La présence des personnes à charge augmente considérablement les dépenses. Dans le cas présent, la présence de personnes à charge augmente les dépenses mensuelles d'une femme d'environ 274 CAD en moyenne, avec plus ou moins 61 CAD de différence. D'autre part, les propriétaires ont également un niveau de dépenses plus élevé d'environ 183 CAD en moyenne par mois, ce qui affirme encore nos prévisions.

Quand à l'âge, son impact demeure ambiguë et peu significatif. Les individus dont l'âge se situe entre 25 et 35 ans et entre 45 et 55 ans dépensent significativement plus que ceux âgés de 35 à 45 ans. L'effet des deux autres catégories d'âge est non significatif.

Il est à noter qu'en utilisant ce modèle, nous obtenons des coefficients globalement significatifs pour toutes les variables et un R^2 de 0.83. Nous avons donc retenu le modèle présent pour l'étape de la validation.

L'étape de validation nous a confirmé que ce modèle fournissait des estimations très proches de la réalité. Cette conclusion a été tirée après avoir comparé la moyenne des dépenses estimées et celle des dépenses réelles au moyen d'un test t. Nous avons par la suite effectué une deuxième régression ayant comme variable dépendante les dépenses réelles et comme seule variable explicative les dépenses estimées. Le coefficient obtenu était de 0.98 qui est une valeur proche de 1 et la constante s'avérait aussi proche de zéro. Ceci nous réaffirme la robustesse du modèle. Il est donc retenu pour l'estimation

des dépenses de la population féminine de la base de données principale.

L'application du modèle sur la population masculine nous produit des résultats globalement identiques à ceux obtenus précédemment dans l'autre population. Comme le montre le tableau 14, les signes des coefficients sont semblables à ceux de la première régression, seule l'influence de l'âge varie légèrement.

TAB. 14 – Régression MCO sur l'échantillon masculin

Variabes	Coefficient	Erreur standard	test t	Pr > t
Constante	601.44	59.2543	10.09	<.0001
REVENU	0.6126	0.0109	58.47	<.0001
DEPENDANT	469.9720	79.3716	5.85	<.0001
PROPRIO	147.7671	49.9801	3.00	0.0028
AGE1	33.0773	91.7290	0.36	0.7117
AGE2	64.5720	65.4090	0.99	0.3269
AGE4	-129.9088	60.7326	-2.15	0.0323
AGE5	-151.742	75.9761	-2.00	0.0421
AGE6	-174.682	157.4777	-1.10	0.2690
R^2 ajusté		0.81		
Nombre d'observations		700		

L'influence des variables demeure semblable à nos prévisions. Dans ce cas, la présence de personnes à charge augmente les dépenses d'un montant moyen supérieur à celui obtenu auparavant au sein de la population féminine. Les dépenses augmentent d'environ 470 CAD en moyenne au lieu de 274 CAD pour les femmes. Concernant l'âge, nous remarquons avec étonnement que les deux premières catégories d'âge dépensent plus que la troisième catégorie. Cependant, les coefficients ne sont pas significatifs. Contrairement à cela, les catégories d'âge plus avancées (supérieure à 45 ans) dépensent moins que la troisième catégorie, les coefficients étant globalement significatifs. Cette tendance pourrait être expliquée par le fait que des personnes de plus de 45 ans ont déjà complété le paiement de leur maison, ne prennent plus en charge leurs enfants, ou encore ne dépensent pas beaucoup en loisirs (sorties, voyages ...).

Ce modèle a été validé de la même manière que le modèle féminin, et les résultats ont également confirmé la robustesse de ses estimations. Nous avons donc retenus ce modèle sans retirer les variables relatives à l'âge vu que dans ce cas elles étaient globalement significatives.

Il est à noter que les signes obtenus lors des deux régressions sont cohérents avec les résultats obtenus par Finlay (2006) dans son article.

Nous avons par la suite appliqué les deux modèles à la base de données principale ce qui nous a permis de remplacer la variable dépenses dans cet échantillon par une variable estimée qui reflète le niveau réel des dépenses selon les caractéristiques de chaque individu. Nous avons utilisée cette variable afin de construire la variable écart que nous avons intégrée dans notre deuxième régression.

7.2 Estimation de la probabilité de défaut

Dans cette partie, notre objectif principal est de déterminer l'impact de la nouvelle variable écart sur la probabilité de défaut afin de vérifier l'existence d'un seuil (ou d'un pourcentage) qui augmente significativement le risque de défaut. Pour ce, nous avons inclut les quatre premières catégories d'écart dans notre régression et nous avons comparé quelques variantes du modèle afin de retenir le modèle le plus complet.

Tout d'abord, nous avons utilisé un modèle incluant la variable écart seulement, ceci nous permet de confirmer l'influence de cette variable sur la probabilité de défaut.

$$P(\text{défaut}) = \beta_0 + \beta_1 \times \text{ecart1} + \dots + \beta_4 \times \text{ecart4} + \varepsilon$$

Le tableau 15 présente les résultats obtenus.

L'impact de la variable écart est clairement démontré lors de cette régression. En effet, tous les coefficients sont significativement positifs et ce relativement à la dernière et meilleure catégorie d'écart. Réduire l'écart entre le remboursement mensuel du prêt

TAB. 15 – Régression Logit - Variable écart

Variabes	Coefficient	Erreur standard	Khi 2	Pr > Khi 2
Constante	-4.6762	0.1367	1146.23	<.0001
Ecart1	1.3744	0.1563	77.4315	<.0001
Ecart2	1.1236	0.1592	49.4866	<.0001
Ecart3	0.4817	0.1760	7.6999	0.0032
Ecart4	0.3123	0.1819	2.9469	0.0826
Log de vraisemblance		139.7		
Nombre d'observations		28312		

et la capacité à payer de l'individu augmente son risque de défaut. Ceci rejoint les résultats de nos statistiques descriptives et réaffirme ainsi nos prévisions.

Afin de s'assurer de la valeur ajoutée de cette variable au modèle interne de la banque nous avons inclut dans le modèle la cote de risque interne de cette institution ainsi que d'autres variables susceptibles d'influencer le défaut.

Le modèle principal utilisé s'exprime comme suit :

$P(\text{défaut}) =$

$$\begin{aligned} & \beta_0 + \beta_1 \times \text{ecat1} + \dots + \beta_4 \times \text{ecat4} + \beta_5 \times \text{score} + \beta_6 \times \text{solde} + \beta_7 \times \text{montant} + \\ & + \beta_8 \times \text{propritaire} + \beta_9 \times \text{AGE2} + \dots + \beta_{13} \times \text{AGE6} + \beta_{14} \times \text{salarie} + \beta_{15} \times \text{garantie} \\ & + \beta_{16} \times \text{ancienneté} + \beta_{17} \times \text{Creditactif} + \beta_{18} \times \text{DUREEA} + \varepsilon \end{aligned}$$

Le tableau 16 résume les résultats obtenus lors de cette régression :

TAB. 16 – Régression Logit - Toutes les variables

Variabes	Coefficient	Erreur standard	Khi 2	Pr > Khi 2
Constante	-7.8357	0.3093	641.15	<.0001
Ecart1	0.7703	0.1820	17.488	<.0001
Ecart2	0.7232	0.1763	16.696	<.0001
Ecart3	0.2312	0.1856	1.6004	0.2012
Ecart4	0.2140	0.1865	1.3169	0.2523
COTRISK	0.5112	0.0260	385.37	<.0001
SOLDE	0.0026	0.00351	0.5696	0.4432
HOMME	0.5467	0.1010	29.313	<.0001
MONTANT	-0.0119	0.00716	2.5316	0.1113
PROPRIÉTAIRE	-0.0781	0.1092	0.5282	0.4686
AGE2	0.3296	0.1287	6.6079	0.0102
AGE3	0.6045	0.1371	19.612	<.0001
AGE4	0.7720	0.1599	23.273	<.0001
AGE5	0.4021	0.2255	3.1828	0.0711
AGE6	-0.0119	0.5377	0.0005	0.9892
SALARIE	-0.1945	0.1137	2.1246	0.1476
GARANTIE	-0.2695	0.1129	5.5177	0.0335
ANCIENMEMB	-0.007	0.00445	3.0018	0.0839
CREDITACTIF	0.1529	0.0394	15.088	<.0001
DUREEA	0.0174	0.00261	43.493	<.0001
Log de vraisemblance	753			
Nombre d'observations	28312			

La variable "Ecart"

Les coefficients des deux premières catégories d'écart demeurent significatifs lors de l'estimation de la probabilité de défaut contrairement aux coefficients des deux catégories suivantes. Ceci indique qu'à partir d'un certain seuil, l'écart entre le remboursement mensuel du prêt et la capacité à payer augmente la probabilité de défaut de l'individu. Rappelons que ce résultat est obtenu en incluant la cote de risque interne de la banque dans la régression. Cela implique que la variable écart confère une certaine valeur ajoutée au modèle interne de la banque puisque l'impact de cette variable sur le risque de défaut demeure significatif même en présence de la cote interne.

Dans le modèle présent, la probabilité de défaut augmente quand cet écart dépasse 42% : Si le paiement mensuel du prêt constitue plus de 42% de la capacité à payer

du client, son risque de défaut devient plus élevé. Cette valeur représente donc le seuil que l'écart ne devrait pas dépasser si la banque ne désire pas que la capacité à payer du client n'affecte son risque de défaut et qu'il devienne alors plus risqué. Nous en concluons alors que l'intégration d'informations relatives à la capacité à payer dans le modèle servant à estimer le défaut augmente sa performance et lui attribue une plus grande précision. Par ailleurs, ces résultats rejoignent d'une certaine manière l'étude de Wilkinson et Tingay (2004) qui avaient déjà prouvé que le développement d'un modèle de pointage avec des données relatives à la capacité à payer confère au modèle une augmentation marginale de la performance.

Ainsi, après avoir estimé les dépenses du client, et calculé sa capacité à payer, la banque pourrait calculer l'écart entre le paiement mensuel basé sur le terme et le montant demandé par le client et sa capacité à payer. Parallèlement, la banque estimera la probabilité de défaut du client. Si cette probabilité est inférieure au seuil préétabli elle calcule par la suite le paiement mensuel maximal qu'elle pourra lui attribuer et qui sera égal à 42% de sa capacité à payer. Ce paiement sera comparé avec la mensualité demandée : s'il est supérieur la banque accorde le prêt et la mensualité demandée et pourrait même proposer au client un prêt plus élevé, sinon la banque pourrait alors proposer la mensualité maximale qu'elle a calculée suivant le terme demandé ou modifier le terme afin d'avoir une mensualité moins élevée.

Si nous reprenons l'exemple exposé dans la section 6.2, nous aurions une mensualité maximale égale à : $1500 \times 0.42 = 630$ CAD. La banque pourrait donc accorder à ce client un prêt équivalent à un paiement mensuel de 630 CAD par mois pendant 60 mois ce qui nous donne un montant total de 37800 CAD intérêts inclus.

Il est évident, que le seuil de 42% peut être remplacé par un autre seuil suivant les objectifs et le niveau d'aversion de risque de la banque. Si cette dernière adopte un objectif de rentabilité par excellence, elle pourrait attribuer au seuil une valeur plus élevée. Nous recommandons néanmoins de ne pas définir un seuil supérieur à un, car

dans ce cas l'individu sera en situation de surendettement et son risque de défaut sera élevé.

Autres variables

La cote de risque interne est très significative lors de l'estimation du risque de défaut. Le signe de son coefficient rejoint nos prévisions. Un client ayant une cote de risque plus élevée est nécessairement plus risqué.

L'âge est globalement significatif, mais le sens de son impact n'est cependant pas clair. Les résultats de la régression nous informent qu'un client âgé entre 25 et 55 ans est apparemment plus risqué qu'un client âgé de moins de 25 ans. Ces résultats peuvent être expliqués par les dépenses et les responsabilités qui deviennent de plus en plus importantes avec l'âge et qui mènent irrémédiablement à une situation financière difficile pouvant causer un défaut.

Dans le modèle présent, les hommes sont significativement plus risqués que les femmes. En effet, le coefficient de la variable "homme" est positif et significatif à 1%.

La présence d'une garantie au moment de l'octroi du prêt baisse significativement la probabilité de défaut du client. Un client qui fournit une garantie a plus d'incitations pour rembourser son prêt, la motivation principale étant la perte de son bien. Ceci implique que ce client est supposé être moins risqué qu'un client qui obtient un prêt sans fournir une garantie. Ce résultat rejoint nos prévisions relatives à l'influence de cette variable sur la probabilité de défaut.

Le nombre de lignes de prêts et de marges de crédit ouvertes ont quand à elles une influence positive sur le risque de défaut du client. Comme le montre le tableau, le coefficient lié à cette variable est significativement positif : plus un individu a de prêts (tous types confondus) plus il a de chance de commettre un défaut. Ceci confirme nos

prévisions et se conforme aux conclusions de SM. Finlay. Cet auteur, a démontré dans son article la relation positive qui existe entre la probabilité de défaut et le surendettement. Un client ayant un nombre élevé d'engagements financier peut être considéré comme surendetté, ce résultat rejoint donc la littérature puisqu'il réaffirme l'hypothèse selon la quelle un individu surendetté présente un risque de défaut plus élevé.

L'âge du prêt joue aussi un rôle important dans l'estimation du défaut. Le coefficient de cette variable (DUREEA) s'avère significativement positif lors de la régression : Un prêt plus ancien présente un risque plus élevé de faire défaut. Ceci implique que le temps a un effet positif sur le risque du client. Par ailleurs, ce résultat est conforme à la littérature puisque la nature de la relation entre cette variable et la probabilité de défaut a déjà été déterminée par Dionne et al (1996) et la conclusion fut semblable.

Enfin, les variables relatives au solde, au montant du prêt, à l'ancienneté du client, et celles définissant les propriétaires et les salariés ne sont pas significatives dans ce modèle.

Ensuite nous avons effectué une régression logistique qui inclut toutes les variables utilisées dans le dernier modèle mais sans la variable écart. Le but de cette régression est de cerner l'impact de ces variables sur le défaut sans l'intégration de la variable écart. Cela nous permettra par la suite de comparer les deux modèles (avec et sans la variable écart) afin de connaître l'impact de l'intégration de la variable écart sur les autres variables.

Le tableau 17 nous expose les résultats obtenus pour cette régression :

Nous remarquons que, globalement, les résultats présentés dans le tableau 17 ne diffèrent pas des résultats du tableau 16. Les quelques exceptions concernent la variable propriétaire et la variable salarié dont les coefficients gardent le même signe mais qui deviennent respectivement significatives à 10% et à 5%. Ces mêmes variables n'étaient pas significatives dans la régression incluant la variable écart. L'intégration de cette dernière influence donc la significativité de ces deux variables. Rappelons que

TAB. 17 – Régression Logit - Sans la variable écart

Variabes	Coefficient	Erreur standard	Khi 2	Pr > Khi 2
Constante	-7.2841	0.2752	700.76	<.0001
COTRISK	0.5383	0.0255	445.73	<.0001
SOLDE	0.00026	0.00344	0.0490	0.9442
HOMME	0.4445	0.0991	20.128	<.0001
MONTANT	-0.0051	0.0069	0.5416	0.4603
PROPRIÉTAIRE	-0.2683	0.1025	6.8545	0.0086
AGE2	0.3014	0.1284	5.5118	0.0189
AGE3	0.4918	0.1371	19.612	0.0003
AGE4	0.5846	0.1560	14.039	0.0002
AGE5	0.2074	0.2227	0.8628	0.3518
AGE6	-0.2289	0.5362	0.1820	0.6697
SALARIE	-0.2349	0.1327	3.1390	0.0145
GARANTIE	-0.2764	0.1131	5.9707	0.0135
ANCIENMEMB	-0.0063	0.00441	2.0211	0.1551
CREDITACTIF	0.1634	0.0388	17.778	<.0001
DUREEA	0.0181	0.00263	47.345	<.0001
Log de vraisemblance	7521			
Nombre d'observations	28312			

la variable écart inclut dans son calcul les dépenses qui sont estimées en fonction du caractère propriétaire de l'individu et d'autres caractéristiques. il est donc possible qu'il existe une certaine corrélation entre ces deux variables qui a eu comme conséquence l'annulation partielle de l'influence de la variable propriétaire sur la probabilité de défaut, et ce, une fois la variable écart intégrée dans la régression.

La dernière différence concerne la variable mesurant l'ancienneté du membre à la banque. Cette variable devient non significative, sachant qu'elle l'était à 10% dans le deuxième modèle.

Enfin, nous avons réestimer la probabilité de défaut en n'incluant que les variables significatives dans la régression précédente. Le tableau 18 nous expose les résultats obtenus.

Les résultats tels que reportés dans ce tableau sont globalement semblables à ceux

obtenus dans la régression précédente.

TAB. 18 – Régression Logit - Variables significatives

Variabes	Coefficient	Erreur standard	Khi 2	Pr > Khi 2
Constante	-7.9457	0.2908	705.15	<.0001
Ecart1	0.7663	0.1719	19.488	<.0001
Ecart2	0.6932	0.1663	17.096	<.0001
Ecart3	0.2048	0.1829	1.2804	0.2512
Ecart4	0.1989	0.1833	1.0869	0.2923
COTRISK	0.5212	0.0260	425.37	<.0001
homme	0.5273	0.1003	27.666	<.0001
AGE2	0.3146	0.1287	6.1879	0.0131
AGE3	0.5775	0.1341	18.782	<.0001
AGE4	0.7320	0.1599	21.273	<.0001
AGE5	0.3705	0.2261	2.7428	0.0911
AGE6	-0.0379	0.5377	0.0041	0.9451
SALARIE	-0.1995	0.1337	2.2346	0.1376
GARANTIE	-0.314	0.1029	8.5177	0.004
CREDITACTIF	0.1509	0.0391	18.288	<.0001
DUREEA	0.0170	0.00261	41.493	<.0001
Log de vraisemblance	749.67			
Nombre d'observations	28312			

8 Extensions de la recherche

Nous avons expliqué auparavant la raison pour la quelle nous avons utilisé seulement 4 variables dans le modèle servant à l'estimation des dépenses. Nous ne disposons pas d'autres variables communes entre les deux bases de données, et c'était donc les seuls variables que nous pouvions intégrer dans notre modèle afin de pouvoir l'appliquer par la suite dans la base de données principale.

Nous avons néanmoins tenté de construire un modèle plus complet comprenant un nombre plus élevé de variables et ce en utilisant la base de données de statistiques Canada. Nous nous sommes basé sur le modèle développé par SM. Finlay pour choisir les nouvelles variables à un inclure dans notre modèle(SM. Finlay, 2006). Finalement, nous avons choisis les variables suivantes :

- La variable « ordinateur » qui précise si la personne dispose d'un ordinateur à la maison ou non.
- La variable « internet » qui nous informe si l'individu dispose d'une connexion internet à la maison ou non
- La variable « voiture » indiquant le nombre de véhicules que l'individu détient.
- La variable « nbnumtel » qui représente le nombre de téléphones inscrits au compte de l'individu. Les cellulaires sont exclus.
- La variable « nbcell » qui nous informe sur le nombre de cellulaires dont l'individu dispose.

Les deux premières variables sont des variables dichotomiques prenant les valeurs zéro ou un selon si le caractère qu'elles représentent existe ou non. Les autres variables sont quand à elles des variables discrètes prenant des valeurs allant de 0 jusqu'à 6.

Le tableau 19 présente les résultats obtenus pour l'échantillon féminin :

Les résultats tels que reportés dans le tableau nous informent que la présence d'un ordinateur ainsi que l'accès à internet à la maison n'est pas une source significative de

TAB. 19 – Régression MCO sur l'échantillon féminin (ajout de variables)

Variables	Coefficient	Erreur standard	Test t	Pr > t
Constante	-59.831	78.2543	-0.77	0.4401
REVENU	0.6898	0.0179	54.82	<.0001
DEPENDANT	201.409	59.3716	3.62	0.0008
PROPRIO	116.617	53.7801	2.16	0.0314
ordinateur	34.56	83.11	0.41	0.6880
internet	11.2844	79.3445	0.15	0.8801
voiture	92.7707	45.5968	2.07	0.0418
nbcell	263.5743	38.0206	6.35	<.0001
nbnumtel	177.6689	62.6553	2.38	0.0045
AGE1	161.0773	101.7290	1.59	0.1168
AGE2	39.5720	61.4090	0.64	0.5269
AGE4	167.9088	60.7326	2.75	0.0062
AGE5	332.742	82.9761	4.02	<.0001
AGE6	49.682	145.4777	0.34	0.7323
R^2 ajusté	0.85			
Nombre d'observations	640			

dépenses additionnelles. Les coefficients liés à ces deux variables sont positifs mais non significatifs. La présence d'un ordinateur dans la maison peut ne pas engendrer des dépenses supplémentaires, mais l'accès à internet demande un abonnement mensuel ainsi que des frais de services connexes. Il est donc étonnant que cette variable soit non significative. En outre, ce résultat ne se conforme pas aux preuves empiriques de Finlay qui mettent en évidence une relation positive et significative entre le niveau des dépenses et l'accès à internet (SM. Finlay, 2006).

La variable liée au nombre de voitures que détient l'individu a un coefficient positif et significatif : plus le nombre de voitures appartenant à l'individu est élevé plus le niveau de dépenses augmente. En effet, les femmes ayant des voitures, sont sujettes à des dépenses additionnelles relatives à ces voitures telles que les réparations, le carburant... D'après les résultats, chaque véhicule détenu augmente les dépenses mensuelles d'environ 93 CAD en moyenne. Par ailleurs, Finlay a tiré la même conclusion de ses résultats, les voitures appartenant à l'individu ont tendance à augmenter ses dépenses (SM. Finlay, 2006).

Le nombre de numéros de téléphones joue aussi un rôle important dans les dépenses mensuelles des personnes. Le coefficient lié à cette variable est positif et significatif à 1%, ce qui implique que plus l'individu a de numéros de téléphones plus il dépense. En réalité, disposer d'un numéro de téléphone implique un abonnement mensuel et donc des dépenses additionnelles, ce qui explique l'impact de cette variable sur le niveau des dépenses. Dans le cas présent, chaque numéro de téléphone ajoute en moyenne 177 CAD aux dépenses mensuelles de l'individu.

Enfin, le nombre de cellulaires détenus a un effet significatif et positif sur les dépenses des femmes. Le coefficient lié à cette variable est positif et significatif à 1%, ce qui permet de confirmer les résultats de Finlay qui a eu la même conclusion (SM. Finlay, 2006). Dans ce modèle, un cellulaire augmente les dépenses mensuelles d'environ 260 CAD en moyenne.

En ce qui concerne les autres variables, leur impact est semblable à celui enregistré dans le premier modèle, seule la valeur des coefficients varie.

Il est à noter que le R^2 de cette régression est de 0.85 comparé à un R^2 de 0.83 pour le modèle construit pour la banque. De plus, la constante est nettement inférieure à celle du premier modèle, ce qui prouve que ce modèle capte plus d'effets que le premier modèle. En conclusion, nous pouvons dire que ce modèle fournira des estimations plus précises des dépenses.

Le même modèle a été appliqué à l'échantillon masculin, et les résultats sont reportés dans le tableau 20

TAB. 20 – Régression MCO sur l'échantillon masculin (ajout de variables)

Variables	Coefficient	Erreur standard	Test t	Pr > t
Constante	469.533	72.8543	7.09	<.0001
REVENU	0.5986	0.0116	52.8	<.0001
DEPENDANT	424.409	80.1716	5.92	<.0001
PROPRIO	131.534	51.8943	2.93	0.010
ordinateur	4.896	79.4657	0.06	0.9515
internet	120.617	78.4770	1.54	0.1228
voiture	23.996	35.0968	0.67	0.5056
nbnmtel	-8.1576	51.0206	-0.15	0.8709
nbccl	80.8924	40.0853	1.96	0.05
AGE1	-1.8503	91.7290	-0.03	0.9868
AGE2	35.5720	66.4090	0.54	0.5969
AGE4	-112.9088	60.7325	-1.86	0.0635
AGE5	-122.042	76.9761	-1.62	0.1097
AGE6	-116.682	159.007	-0.73	0.4602
R^2 ajusté	0.82			
Nombre d'observations	700			

En ce qui concerne l'échantillon masculin, les résultats sont quelque peu différents. Le nombre de numéros de téléphones et de voitures détenus n'ont pas un impact significatif sur les dépenses des individus et ce contrairement à l'échantillon féminin.

Seul le nombre de cellulaires demeure significatif dans ce modèle. Le coefficient lié à cette variable est positif et significatif à 10%. Cependant, il est nettement inférieur au coefficient de la régression sur l'échantillon féminin. Ceci implique que le nombre de cellulaire détenus influence moins le niveau des dépenses des hommes. En d'autres termes, les femmes, en moyenne, dépensent plus au téléphone que les hommes. Ceci est d'autant plus confirmé par la valeur du coefficient estimé de cette variable. En effet, un cellulaires augmente les dépenses mensuelles des hommes d'environ 55 CAD en moyenne. Cette valeur est nettement inférieure aux dépenses additionnelles du cellulaire pour les femmes et qui sont de 160 CAD en moyenne par mois.

Nous précisons que le R^2 de ce modèle est de 0.82, il est donc légèrement supérieur au R^2 du modèle appliqué à la base de données principale. Par ailleurs, la constante

est aussi inférieure à celle du premier modèle. Ce modèle est donc plus performant que le modèle principal.

Nous constatons donc que ces deux derniers modèles sont susceptibles de fournir des estimations plus précises et plus proche de la réalité. Par ailleurs, certaines de ces variables peuvent être significatives compte tenu de la période d'analyse. Par exemple, l'utilisation d'internet va devenir une activité commune dans les années à venir. Le principal message à retenir pour les banques, est de bien suivre l'évolution temporelle des analyses de dépenses par Statistiques Canada afin d'améliorer continuellement son modèle.

9 Conclusion

L'objectif principal de ce mémoire était de construire un indice de capacité à payer fiable qui aurait un impact sur la performance et l'efficacité de la gestion de risque de crédit. Pour ce faire, nous avons besoin d'informations précises concernant toutes les composantes de cet indice. Malheureusement, le manque d'informations relatives aux dépenses nous a incités à utiliser une autre méthode afin de disposer de l'information demandée. Ceci représentait notre première problématique : la nécessité d'utiliser une méthode qui nous permettait de déterminer les dépenses des clients afin de l'intégrer dans l'indice de capacité à payer. Afin de résoudre ce problème, nous avons développé un modèle visant à estimer les dépenses et ce à partir d'une base de données indépendante issue de Statistique Canada. Après l'étape de validation, ce modèle a été appliqué à la base de données principale de la banque afin de corriger les données relatives aux dépenses et de calculer notre indice de capacité à payer. Le calcul de cet indice incorporait non seulement les actifs et les passifs du client mais aussi une dimension du prêt qu'il demande. Nous n'avons pas calculé une capacité à payer absolue, mais la capacité du client à payer le prêt qu'il demande. Cet indice représente le montant mensuel maximum que le client peut payer sachant ses revenus, ses dépenses, ses actifs liquides, ses engagements financiers ainsi que le terme du prêt demandé. Nous avons par la suite démontré comment cet indice pouvait être utilisé afin d'aider la banque à déterminer le montant et le terme du prêt.

Dans le but de matérialiser la relation entre l'indice de capacité à payer développé et la probabilité de défaut, nous avons construit une autre variable nommée « écart ». Cette variable représente le pourcentage du paiement mensuel du prêt accordé par la banque relativement à sa capacité à payer calculée. Elle permet de distinguer les personnes surendettées, qui sont représentées par celles ayant un remboursement mensuel supérieur à leur capacité à payer mensuelle. Nous avons par la suite démontré que ces personnes ont un taux de défaut supérieur au taux de défaut des personnes n'étant pas en situation de surendettement. Afin de visualiser l'impact de cette variable sur la probabilité de défaut, nous l'avons intégré dans le modèle estimant le risque de défaut.

Les résultats nous ont démontrés qu'à partir d'un certain seuil cette variable augmente significativement la probabilité de défaut. En effet, si le remboursement mensuel du prêt accordé dépasse 42% de la capacité à payer du client, sa probabilité de défaut se voit affecter et il devient alors plus risqué. Cette conclusion nous a permis d'affirmer notre hypothèse initiale et ainsi confirmer que la prise en compte de la capacité à payer dans l'évaluation du risque de crédit offre une valeur ajoutée non seulement pour la détermination des conditions du prêt mais aussi pour le modèle interne servant à estimer la probabilité de défaut. Tenir compte de cet indice permet à la banque d'améliorer son modèle interne en intégrant l'impact du surendettement (et donc de la capacité à payer) sur la probabilité de défaut ce qui améliore l'estimation.

Notre étude procure à la banque canadienne un premier modèle d'estimation des dépenses qui pourrait être appliqué à toutes ses bases données pourvu qu'elles aient les mêmes pondérations. Elle propose par la suite une formule de calcul de l'indice de capacité à payer qui prend en considération toutes les informations nécessaires dont les dépenses estimées. La banque pourrait par la suite calculer l'écart relatif entre la mensualité désirée et la capacité à payer du client et le comparer avec un seuil qu'elle a défini auparavant selon ses objectifs et son niveau d'aversion au risque. Par ailleurs, nous conseillons un seuil de 42% si la banque ne veut pas affecter la probabilité de défaut de ses clients. Cependant, le seuil de 42% n'est pas parfaitement endogène, une extension de cette recherche serait la modélisation de ce seuil de manière tout à fait endogène.

Finalement, nous avons développé un deuxième modèle d'estimation des dépenses en intégrant d'autres informations non disponibles dans la base de données de la banque. Nous avons constatés que ces modèles étaient plus complets et plus performants que les premiers. Nous recommandons donc aux banques d'inclure dans leurs formulaires plus d'informations relatives aux caractéristiques du client (nombre de voitures, accès à internet, nombre de cellulaires...) et ce en suivant l'évolution des analyses des dépenses. Ceci leur permettra d'utiliser des modèles plus performants et ainsi de cal-

culer un indice de capacité à payer encore plus précis, ce qui se reflètera sur tout le processus décisionnel de la banque. Une autre ouverture de recherche serait d'estimer les dépenses en utilisant un modèle non linéaire qui pourrait mieux refléter la relation entre les dépenses et les différentes variables qui l'expliquent.

Références

- [1] CARBONI, Alexandre (2007). «*Analyse de la prise en compte de la capacité à payer dans la cote de risque pour les défauts des particuliers*», mémoire de maîtrise, Montréal, HEC Montréal
- [2] DE ANDRADE, F.W.M , LC. Thomas, (2007) «*Structural models in consumer credit*», School of Management, University of Southampton, Southampton,
- [3] ARMINGER, Gerhard, Daniel ENACHE et Thorsten BONNE, (1997). «*Analyzing credit risk data : a comparison of logistic discrimination, classification tree analysis and feedforward networks*», Computational Statistics, n°12, 293-310,
- [4] Baesens, B., (2003). «*Developing Intelligent Systems for Credit scoring using Machine Learning Techniques*», thèse de doctorat, Katholieke Universiteit Leuven, cité par : THOMAS, LC., (2003). «*Consumer Credit Modeling : Context and Current Issues*», School of Management, University of Southampton, Southampton.
- [5] Consumer Affairs Directorate,(2001). «*Report by the Task Force on Tackling Over-indebtedness*», Department of Trade and Industry, London, UK.
- [6] Consumer Affairs Directorate, (2003). «*Second Report by the Task Force on Tackling Over-indebtedness*», Department of Trade and Industry, London, UK.
- [7] Association des Comptables Généraux Accrédités du Canada (CGA), (2007). «*Où va l'argent? La croissance de l'endettement des ménages au Canada*», Rapport de l'association des Comptables Généraux Accrédités du Canada, <www.cga.org/canada-fr>.
- [8] DIONNE, Georges, M. ARTIS et M. GUILLEN, (1996). «*Count data models for a credit scoring system*», Journal of Empirical Finance, 3, p. 303-325,
- [9] HAND, DJ. et WE. HENLEY, (1997). «*Statistical Classification Method in Consumer Credit Scoring : A Review*», Journal of the Royal Statistical Society, vol. 160, n° 3, p. 523-541,
- [10] HAND, DJ., RW. OLIVER et LC. THOMAS, (2005). «*A survey of the issues in consumer credit modelling research*», Journal of the Operational Research Society, p.1-10,

- [11] ROSZBACH, Kasper, (2004). «*Bank lending policy, credit scoring and the survival of loans*», Revue of Economics and Statistics.
- [12] THOMAS, LC., (2000). «*A survey of credit and behavioural scoring : forecasting financial risk of lending to consumers*», International Journal of Forecasting, n° 16, p. 149-172.
- [13] THOMAS, LC., DB. EDELMAN et JN. CROOK, (2002). «*Credit Scoring and its Applications*», SIAM : Philadelphia.
- [14] THOMAS, LC., (2003). «*Consumer Credit Modeling : Context and Current Issues*», School of Management, University of Southampton, Southampton.
- [15] THOMAS, LC., (2005) «*Modelling the Credit Risk for Portfolios of Consumer Loans : Analogies with the corporate loan models*», School of Management, University of Southampton, Southampton.
- [16] FINLAY, SM., (2006). «*Predictive models of expenditure and over-indebtedness for assessing the affordability of new consumer credit applications*», Journal of the Operational Research Society, n°57, p. 655-669.
- [17] WILKINSON G. et J. TINGAY, (2004). «*The use of affordability data - does it add real value ?*», cité par : Thomas LC, Edelman DB and Crook JN (eds) «*Readings in Credit Scoring. Foundations, Developments, and Aims*», Oxford University Press, Oxford, UK, pp 73-91.