



Mémoire présenté
devant l'Institut de Science Financière et d'Assurances
pour l'obtention du diplôme d'Actuaire de l'Université de Lyon

le 11 avril 2013

Par : Adrien MANAS

Titre: Spécificités de la ligne d'affaire de la distribution directe du point de vue du réassureur

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Membres du jury de l'Institut des Actuaires

Entreprise :

Transamerica Reinsurance

Membres du jury I.S.F.A.

Directeur de mémoire en entreprise :

Mme Flavia BARSOTTI

Mme. Elsa RENOUF

M. Alexis BIENVENÛE

M. Areski COUSIN

Invité :

Mme Diana DOROBANTU

Mme Anne EYRAUD-LOISEL

M. Nicolas LEBOISNE

M. Stéphane LOISEL

Mlle Esterina MASIELLO

Mme Véronique MAUME-DESCHAMPS

M. Frédéric PLANCHET

Mme Béatrice REY-FOURNIER

M. Pierre RIBEREAU

M. Christian-Yann ROBERT

M. Didier RULLIERE

M. Pierre THEROND

Autorisation de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Secrétariat

Mme Marie-Claude MOUCHON

Signature du candidat

Bibliothèque :

Mme Patricia BARTOLO

RESUME

Mots-clés : *Distribution directe, réassurance, marketing direct/télémarketing, coûts d'acquisition, émergence du résultat, analyse du risque de chute, algorithme CART, régression logistique, risque de contrepartie, Solvabilité II*

Ce mémoire a pour objet l'étude des spécificités de la ligne d'affaire de la distribution directe pour le réassureur. En effet, ce mode de distribution sans intermédiaire entre l'assureur et l'assuré soulève de nombreuses questions relatives à la conception, au financement et au suivi du risque des produits ainsi distribués. En particulier, la prise en charge du financement de coûts d'acquisition importants à la souscription du contrat représente pour le réassureur la principale problématique lors de la mise en place d'une telle affaire.

La première partie de ce mémoire propose une revue du marché français de la distribution directe d'assurance. Les stratégies de distribution directe de produits d'assurance (marketing direct, télémarketing, multi-canal, etc.) sont notamment évoquées. Par la suite, nous étudions les acteurs et les produits rencontrés en France ainsi que le comportement des consommateurs à l'égard de ce mode de distribution. Un focus est ensuite fait sur le rôle du réassureur. Dans cette même partie contextuelle, nous nous intéressons aux problématiques légales associées, en particulier celles liées à la dématérialisation du contrat et au devoir d'information de l'assureur. Enfin, nous évoquons les opportunités de développement sur la sélection des risques, service de prédilection des réassureurs auprès de leurs cédantes, via des outils de sélection à distance.

La seconde partie traite d'une des problématiques majeures du réassureur lors de la mise en place d'un traité impliquant un financement important. Il s'agit de l'émergence du résultat dans différents référentiels comptables selon qu'ils permettent ou pas la mise en place de coûts d'acquisition différés. Un exposé des principes de prise en compte de ces coûts sous les référentiels comptables américain, français puis international (IFRS) est proposé suivi d'une illustration numérique.

Par la suite, une troisième partie est développée sur le risque de chute qui constitue le risque de type souscription le plus important pour ces affaires. En particulier, nous traitons de l'identification des facteurs de risque les plus discriminants et des prédictions grâce à l'utilisation des arbres de décision par l'algorithme CART et de la régression logistique. Les concepts théoriques sont présentés puis nous nous proposons une étude de cas présentant l'expérience de deux portefeuilles issus de campagnes de télémarketing dans deux pays différents.

Enfin, la dernière partie soulève la problématique du risque de contrepartie pour le réassureur en fonction de la structure de réassurance impliquant ou non des liens financiers avec un courtier distributeur. Pour l'application numérique, nous nous appuyons sur le cadre donné par Solvabilité 2 et les spécifications techniques du QIS5.

ABSTRACT

Keywords: *Distance selling, reinsurance, direct marketing / telemarketing, acquisition costs, accounting standards, surrender analysis, CART algorithm, logistic regression, counterparty risk, Solvency II*

This report aims to study the particularities of the direct distribution business line from a reinsurer's point of view. This distribution channel without any physical contact between the insurer and the policyholder raises many questions relating to the design of products, funding and risk management. The main particularity for the reinsurer of this type of contract is often the management of the financing of acquisition costs at the time the contract is written.

The first part of this report provides a review of the French market of direct distribution insurance products. Strategies of direct distribution of insurance products (direct marketing, telemarketing, multi-channel, etc.) are presented. Thereafter, players and products encountered in France together with consumer behavior in relation to this kind of distribution are studied. A focus is made on the role of the reinsurer. In the same section, we focus on the legal issues involved, particularly those relating to the absence of wet signature and the information due by the insurer. Finally, we discuss the development opportunities for risk selection, a key service offered by reinsurers to their ceding companies, via remote selection tools.

The second part of the report addresses one of the areas for interest of the reinsurer in the establishment of a treaty involving significant funding, namely the ability to implement deferred acquisition costs and the impact on profitability according to accounting standards and methods used for the profit emergence. A statement of the principles of recognition of these costs under U.S. GAAP, French GAAAP and IFRS GAAP is proposed followed by a numerical application.

Thereafter, a third part is developed on the lapse risk which is often the most important underwriting risk for this type of business well before any claim deviation risk. In particular, we discuss the identification of the most discriminating risk factors and we make predictions by using decision trees with CART algorithms and logistic regressions. The theoretical concepts are presented followed by a case study presenting the experience of two portfolios from telemarketing campaigns in two different countries.

The final part raises the issue of counterparty risk for the reinsurer according to the structure of reinsurance which may or may not involve a financial relationship with a broker. For the numerical implementation, we rely on the framework provided by Solvency 2 and the technical specifications of QIS5.

REMERCIEMENTS

*Je tiens tout d'abord à remercier **Elsa Renouf**, Directeur Technique et Développement Produits Europe, qui m'a permis d'effectuer ce mémoire autour d'une étude intéressante. Merci, Elsa, pour ton suivi tout au long de la réalisation de ce mémoire ainsi que pour m'avoir fait bénéficier de tes conseils et de ton expérience qui me suivront tout au long de ma vie professionnelle.*

Je remercie également l'ensemble du bureau parisien de Transamerica Reinsurance Europe qui m'a accueilli et guidé pendant plus d'un an.

*Je tiens à remercier plus particulièrement **Colin Kearney** et **Patrick Sabau** qui en partageant avec moi leur expérience ont permis d'enrichir sensiblement ce mémoire.*

*Merci à l'ensemble de l'équipe actuariat, **Anne-Claude Adam**, **Jennifer Hartley** et **Stéphanie Jos** pour leurs conseils techniques et leurs commentaires.*

*Je remercie ensuite l'I.S.F.A., notamment **Stéphane Loisel**, mon tuteur pédagogique, et **Pierre Théron**, pour leurs conseils et l'attention qu'ils ont pu apporter à mon travail.*

Je souhaite enfin remercier ma famille et Audrey qui ont su m'apporter leur confiance et leur soutien.

SOMMAIRE

RESUME	1
ABSTRACT	2
REMERCIEMENTS	3
SOMMAIRE	4
AVANT-PROPOS	7
INTRODUCTION	8
PARTIE I : LA DISTRIBUTION DIRECTE D'ASSURANCE EN FRANCE	9
Chapitre 1. Les techniques de distribution directe	10
Section 1.1. Caractéristiques des produits distribués par marketing direct	10
Section 1.2. Utilisation du marketing direct	11
Section 1.3. Les outils de communication utilisés en marketing direct.....	12
Section 1.4. Evolution récentes des techniques de marketing direct	14
Section 1.5. Le business model du marketing direct	14
Chapitre 2. Stratégies actuelles et perspectives d'évolution de la distribution directe de produits d'assurance en France.....	16
Section 2.1. Télémarketing et marketing direct : les acteurs historiques de la distribution directe.....	16
Section 2.2. Evolution des stratégies de distribution multi-canal des assureurs français	17
Section 2.3. Internet, un domaine d'innovation pour la distribution directe d'assurance	18
Section 2.4. Comportements des assurés.....	20
Chapitre 3. Le rôle du réassureur	22
Section 3.1. Rôle du réassureur dans la mise en place d'une structure de distribution directe	23
Section 3.2. Les atouts du réassureur dans l'utilisation de techniques de distribution alternatives pour la distribution de produits d'assurance	24
Chapitre 4. Les aspects légaux liés à la distribution directe de produits d'assurance	25
Section 4.1. Cadre légal.....	25
Section 4.2. L'obligation d'information de l'assureur et les sanctions en cas de non respect.....	26
Section 4.3. Le droit à renonciation de l'assuré.....	27
Section 4.4. La conclusion du contrat à distance.....	28
Chapitre 5. Tarification et souscription à distance : la télé-souscription.....	29
Section 5.1. Définition de la télé-souscription.....	29
Section 5.2. Avantages et limites de la télé-souscription	32
Section 5.3. Expérience dans d'autres pays et développement en France.....	33

PARTIE II : BESOIN DE FINANCEMENT ET PROFITABILITE SOUS DIFFERENTS REFERENTIELS COMPTABLES	35
Introduction : relations entre l'assureur et le distributeur pour un contrat d'assurance conclu à distance	35
Chapitre 1. Prise en compte des coûts d'acquisition sous différents référentiels comptables.....	36
Section 1.1. Définition et intérêts d'un référentiel comptable.....	36
Section 1.2. Coûts d'acquisition différés selon les normes comptables américaines	37
Section 1.3. Coûts d'acquisition différés selon les normes comptables françaises	38
Section 1.4. Frais d'acquisition reportés selon la norme IFRS contrats d'assurance phase 2	40
Chapitre 2. Application : émergence du résultat pour un produit avec commission d'acquisition	41
Section 2.1. Caractéristiques du contrat d'assurance modélisé	41
Section 2.2. Méthodes de calcul des provisions techniques	43
Section 2.3. Méthodologie retenue pour le modèle	44
Section 2.4. Résultats de l'application numérique.....	46
PARTIE III : ANALYSE DU RISQUE DE CHUTE.....	51
Chapitre 1. Produits étudiés et obtention d'une table d'étude.....	51
Section 1.1. Description des produits d'assurance	51
Section 1.2. Extraction et traitement des données entrant dans le périmètre de l'étude.....	52
Section 1.3. Etude des portefeuilles à travers les tables « contrats ».....	53
Section 1.4. Caractéristiques de la table d'étude après retraitement	53
Chapitre 2. Etude descriptive des résiliations sur les deux tables d'étude	55
Section 2.1. Intuitions et analyse préalable	55
Section 2.2. Commentaires et méthodologie choisie.....	56
Section 2.3. Etude des résiliations sur les portefeuilles XW et YZ.....	57
Chapitre 3. Les arbres de décision et l'algorithme CART	71
Section 3.1. Généralités	71
Section 3.2. L'algorithme CART	73
Section 3.3. Application de l'algorithme CART aux deux portefeuilles.....	83
Chapitre 4. Le modèle de régression logistique	93
Section 4.1. Présentation de la fonction logistique.....	93
Section 4.2. La régression logistique.....	94
Section 4.3. Aspects théoriques de la régression logistique	96
Section 4.4. Interprétation du modèle obtenu.....	99
Section 4.5. Application de la régression logistique aux deux portefeuilles	101

PARTIE IV : RISQUE DE CREDIT LIE A DES STRUCTURES DE FINANCEMENT.....	114
Chapitre 1. Risque de contrepartie chez le réassureur.....	114
Chapitre 2. Calcul du SCR pour le module risque de contrepartie	115
Section 2.1. La directive Solvabilité II : objectifs et organisation en piliers.....	115
Section 2.2. Le processus consultatif dans Solvabilité II : les études d'impacts quantitatives.....	116
Section 2.3. Calcul du SCR dans les spécifications techniques du QIS5 (Modèle Standard)	116
Chapitre 3. Spécifications du QIS5 pour le calcul du SCR du module du risque de contrepartie	117
Section 3.1. Calcul du capital requis pour une exposition de type 1	118
Section 3.2. Calcul du capital requis une exposition de type 2	119
Section 3.3. Calcul du capital requis pour l'ensemble des contreparties.....	119
Chapitre 4. Comparaison de deux structures de réassurance : avec/sans risque de crédit du courtier pour le réassureur dans le cas d'une clause de reprise sur commission initiale en cas de résiliation	119
Section 4.1. Hypothèses retenues	120
Section 4.2. Résultats de l'application.....	122
Section 4.3. Quelques mots pour conclure	124
CONCLUSION DU MEMOIRE	126
BIBLIOGRAPHIE	128
SITES INTERNET DE REFERENCES.....	129
ANNEXES	130
Annexe 1 : Articles de loi concernant la commercialisation à distance de contrats d'assurance	130
Annexe 2 : Exemple de grille de sélection emprunteur	133
Annexe 3 : Articles de loi concernant les commissions et frais d'acquisition à amortir	134
Annexe 4 : Table des correspondances pour l'algorithme CART sur le portefeuille XW	134
Annexe 5 : Table des correspondances pour l'algorithme CART sur le portefeuille YZ	134
Annexe 6 : Sorties obtenues avec le logiciel R pour l'algorithme CART dans l'étude du portefeuille XW sans la modalité « 0 » prime payée de la variable « NMP2 ».....	135
TABLE DES FIGURES	137
TABLE DES TABLEAUX.....	138

AVANT-PROPOS

La division de réassurance du groupe AEGON opérant sous la marque Transamerica Reinsurance (TARe) a été rachetée le 10 août 2011 par le réassureur européen Scor Global Life. Le contexte de ce mémoire s'appuie essentiellement sur la période précédant la vente.

INTRODUCTION

Transamerica Reinsurance (TARe), marque sous laquelle opèrent les entités de réassurance du groupe AEGON, se classe parmi les dix plus grands réassureurs de personnes au niveau mondial. TARe offre des solutions de réassurance vie à plus de 500 sociétés en Europe, Amérique du Nord, Amérique Latine et dans la zone Asie-Pacifique. J'ai effectué mon alternance au sein de l'entité TARe Europe basée à Paris.

Parmi les lignes d'affaires souscrites par TARe, figure la réassurance de portefeuilles issus de campagnes de marketing direct, de télémarketing ou plus largement de distribution à distance (incluant notamment Internet).

Largement développés dans les pays anglo-saxons mais apparus plus récemment en France, ces modes de distribution sont aujourd'hui des relais de croissance importants pour les organismes assureurs (même si les contrats d'assurance souscrits par ces canaux restent encore marginaux, de l'ordre de 3 % du total des souscriptions, en France, en 2010, d'après la Fédération Française des Sociétés d'Assurance). Dans ce contexte, il est important pour un réassureur d'apporter aux cédantes son expertise sur la conception, le financement et le suivi du risque de tels produits dont la principale particularité, pour celui-ci, est souvent la prise en charge du financement de coûts d'acquisition importants à la souscription du contrat (bien souvent à travers le versement d'une commission initiale à un intermédiaire).

Dans la première partie du mémoire, nous débutons par la présentation de cette ligne d'affaire en introduisant les concepts, les acteurs et les offres proposées sur le marché français. Nous nous intéressons ensuite aux problématiques légales associées, en particulier liées à la dématérialisation du contrat et au devoir d'information de l'assureur. Enfin, nous évoquons les opportunités de développement de la sélection des risques, service de prédilection des réassureurs auprès de leur cédante via des outils de sélection à distance.

Dans le cadre de la réassurance de cette ligne d'affaire, il est assez fréquent que le réassureur prenne en charge, totalement ou partiellement, le financement des coûts d'acquisition, ce qui pose des questions spécifiques que nous avons souhaité reprendre dans ce mémoire. Plus particulièrement, nous étudions les trois problématiques suivantes de manière théorique, puis en s'appuyant sur des applications numériques :

- Impacts des coûts d'acquisition différés sur la profitabilité du produit en fonction des référentiels comptables utilisés et des méthodes d'émergence du résultat (Partie 2).
- Risque de chute, identification des facteurs de risque les plus discriminants et prédictions (arbres de décision par l'algorithme CART et régression logistique). Afin d'illustrer cette partie, nous nous appuyons sur un exemple pratique présentant les études des résiliations de deux portefeuilles issus de campagnes de télémarketing dans deux pays différents (Partie 3).
- Revue du risque de contrepartie pour le réassureur en fonction de la structure de réassurance impliquant ou non le courtier distributeur. Pour l'application numérique, nous nous appuyons sur le cadre donné par Solvabilité 2 et les spécifications techniques du QIS5 (Partie 4).

PARTIE I : LA DISTRIBUTION DIRECTE D'ASSURANCE EN FRANCE

Depuis son apparition en France au début des années 1990, la distribution directe a connu un développement important et constitue aujourd'hui une réalité. Selon le rapport annuel 2010 de la Fédération Française des Sociétés d'Assurance (FFSA)¹, les contrats d'assurance issus de la distribution directe (marketing direct et télémarketing) et d'Internet représentent 3 % des cotisations totales pour les sociétés vie, de capitalisation et mixtes et 2 % pour les sociétés dommages. Ainsi, bien que cette part soit marginale et que les chiffres soient stables depuis 2004, l'existence de ces modes de distribution est bien réel.

	2004	2005	2006	2007	2008	2009	2010
SOCIÉTÉS VIE, DE CAPITALISATION ET MIXTES¹⁾							
- Réseaux de bancassurance	62	62	64	62	60	60	61
- Agents généraux	7	7	7	7	8	7	7
- Courtiers	13	13	12	13	14	14	13
- Saliariés	16	16	15	16	16	16	16
- Sociétés sans intermédiaires	-	-	-	-	-	-	-
- Autres modes (vente directe, Internet...)	2	2	2	2	2	3	3
SOCIÉTÉS DOMMAGES							
- Réseaux de bancassurance	9	9	9	9	10	10	11
- Agents généraux	35	35	35	35	35	35	34
- Courtiers	18	18	18	18	18	18	18
- Saliariés	2	2	2	2	2	2	2
- Sociétés sans intermédiaires	33	33	33	33	33	33	33
- Autres modes (vente directe, Internet...)	3	3	3	3	2	2	2

1. Nouvelles séries à partir de 2004; affinement de la rubrique « autres modes ».

Tableau 1 : Les cotisations selon la forme de distribution (en %)

Source : Rapport annuel 2010 de la FFSA

Dans ce contexte, le choix de la politique de distribution est un aspect très important dans la définition de la stratégie d'une compagnie d'assurance. A l'heure actuelle, plusieurs moyens pour atteindre les assurés potentiels existent: les agents généraux, les courtiers, le téléphone, Internet, etc. Dans de nombreux pays, en particulier chez les anglo-saxons où ils représentent de 25 % à 30 % du marché, les techniques de distribution directe se sont développées au cours des vingt dernières années. En France, malgré un démarrage plus tardif et des réticences plus importantes, la distribution directe commence à devenir un enjeu stratégique pour la plupart des assureurs. Dans cette situation, le réassureur peut jouer un rôle intéressant car il possède une expertise qu'il peut mettre à la disposition de ses partenaires, qu'ils soient assureurs ou distributeurs, grâce à son expérience acquise sur les marchés internationaux.

Par ailleurs, la distribution directe de produits d'assurance pose des problématiques légales liées à la dématérialisation du contrat d'assurance et à l'exigence de conseil et d'information à l'égard de l'assuré lorsque le contrat est conclu à distance. Enfin, l'évolution des stratégies de distribution et le recours de plus en plus fréquent aux techniques de communication à distance ont également permis une réflexion sur l'amélioration des techniques de sélection des risques et de tarification.

L'étude de la distribution directe de produits d'assurance est liée à différentes thématiques développées dans la première partie de ce mémoire. Celle-ci constitue la synthèse des recherches bibliographiques menées dans la découverte et l'étude de ce mode de distribution ainsi que d'une série d'entretiens avec les membres de l'équipe de Transamerica Reinsurance Europe. Au cours de ces échanges, la connaissance du marché français de la distribution directe de Patrick Sabau, directeur France de Transamerica Reinsurance, et

¹ Ce rapport est consultable sur http://www.ffsa.fr/sites/upload/docs/application/pdf/2011-06/raffsa2010fessentielffsa2010-3_2011-06-30_17-16-49_557.pdf

l'expertise de Colin Kearney, responsable souscription Europe, sur les techniques de souscription à distance ont été particulièrement précieuses.

Dans cette première partie, nous commençons par présenter quelques-uns des modes d'utilisation des systèmes de distribution directe (chapitre 1). Nous définissons en particulier le marketing direct et identifions les principales caractéristiques des produits distribués par cette technique ainsi que les outils de communication utilisés. Dans ce premier chapitre, nous attirons l'attention sur les évolutions les plus récentes des techniques de distribution directe ainsi que sur le *business model* utilisé par les distributeurs pour cibler les assurés potentiels. Dans le chapitre 2, nous analysons plus spécifiquement les stratégies actuelles de distribution directe de produits d'assurance ainsi que les perspectives d'évolution offertes, notamment, par Internet. Nous nous attachons aussi à étudier le comportement des assurés vis-à-vis de ces produits. Ensuite, nous expliquons les apports de la réassurance pour ce type d'affaire (chapitre 3). Enfin, dans les chapitres 4 et 5, nous étudions tour à tour les aspects légaux liés à la distribution de produits d'assurance à distance puis les spécificités de la sélection des risques et de la tarification à distance.

Chapitre 1. Les techniques de distribution directe

Les techniques de distribution conçues pour permettre à l'assureur de traiter directement avec l'assuré potentiel utilisent principalement trois méthodes qui sont :

- le publipostage,
- la publicité dans les journaux, les magazines, la radio ou la télévision,
- et le télémarketing.

Ces méthodes s'appuient sur une technique commerciale appelée marketing direct qui, en utilisant un ou plusieurs outils de communication, provoque une réaction immédiate et mesurable (par exemple une demande de renseignements ou un achat). Cette réaction est ensuite enregistrée pour être analysée puis mémorisée dans une base de données pour un usage ultérieur.

Un des intérêts de la distribution directe est de permettre à l'assureur de développer son chiffre d'affaires d'une façon différente de la distribution dite « classique » et de se démarquer de ses concurrents en touchant de nouvelles cibles d'assurés. Apparue au début des années 1990 dans le secteur des services financiers (banques et assurances), le marketing direct a connu depuis un développement important et est aujourd'hui utilisé par la grande majorité des assureurs.

Ce premier chapitre a pour objectif d'illustrer la mise en place et le suivi d'une campagne de marketing direct chez un assureur. Dans cette optique, nous étudions d'abord les caractéristiques des produits distribués ainsi que les techniques du marketing direct. Nous présentons ensuite les outils de communication utilisés ainsi que les évolutions récentes et les outils mathématiques du contrôle de la rentabilité d'une campagne de marketing direct.

Section 1.1. Caractéristiques des produits distribués par marketing direct

L'utilisation du marketing direct par les assureurs est fortement influencée par les caractéristiques des produits. Différents produits d'assurance peuvent être distribués au moyen du marketing direct mais certains d'entre eux sont plus adaptés que d'autres. Dès lors, les produits d'assurance doivent être conçus expressément pour le marketing direct si l'assureur souhaite retirer les avantages de cette technique de distribution. Ainsi, les garanties proposées en distribution directe concernent des risques bien spécifiques et sont très similaires chez les différents acteurs du marché.

De manière générale, les deux principales caractéristiques de ces produits sont leur simplicité et leur prix abordable. Par conséquent, la façon dont les tarifs sont présentés est importante. En pratique, ceux-ci sont souvent donnés sous la forme d'une grille de tarifs différenciés selon la tranche d'âge et le sexe de l'assuré. Nous parlons alors de barèmes gradués puisque la prime augmente en fonction de la tranche d'âge de l'assuré à la souscription. Par ailleurs, une fois le contrat conclu, la prime n'augmente pas en fonction de l'âge atteint. En effet, une telle pratique pourrait entraîner un taux de résiliation plus important à l'âge où la prime augmenterait. De plus, pour attirer davantage d'assurés, une période de gratuité (en général le premier mois) est parfois présente et des options de montant de garanties peuvent être proposées. Du point de vue de la communication de l'assureur, l'accent est mis sur la protection offerte à l'assuré et sur la qualité de la prise en charge en cas de sinistre.

Ensuite, les primes sont présentées sous forme de paiements mensuels modiques plus attractifs pour le budget de l'assuré qu'une prime annuelle unique. Les compagnies considèrent en effet qu'il est plus facile de distribuer un produit à 9,95 € par mois plutôt qu'un produit à prime annuelle de 100 €. On peut noter qu'une fréquence mensuelle de primes entraîne des frais de gestion plus élevés pour l'assureur. Quant au mode de paiement utilisé, il s'agit en général de prélèvements bancaires automatiques.

Enfin, les produits distribués par marketing direct sont simples à émettre dans la mesure où la sélection médicale à la souscription est faible voire inexistante. Ces produits présentent donc un risque d'anti-sélection que l'assureur peut réduire en ajoutant des exclusions à la police, des délais de carence ou des franchises.

L'assurance emprunteur est un des produits les plus fréquemment rencontrés dans les pays anglo-saxons. Il est aussi possible de trouver différentes garanties en cas de décès accidentel, d'hospitalisation (en cas d'accident, de maladies graves, etc.), de prévoyance (décès toutes causes, rente éducation) ou pour des frais d'obsèques. Parmi ces produits, l'assurance hospitalisation est largement distribuée. Ce type de produit fera d'ailleurs l'objet d'une application sur le risque de chute pour les produits distribués en marketing direct dans la troisième partie de ce mémoire.

Section 1.2. Utilisation du marketing direct

Pour distribuer ses produits d'assurance par marketing direct, l'assureur doit procéder à des études de marché afin d'identifier les meilleures cibles à contacter et d'établir des tests destinés à évaluer la réaction des assurés potentiels ciblés par la compagnie.

Les personnes déjà référencées dans un fichier de consommateurs (titulaires d'une carte de crédit, clients d'un organisme de vente à distance (France Loisirs, La Redoute), etc.) représentent une cible privilégiée pour les assureurs utilisant le marketing direct. En effet, l'utilisation de bases de données déjà existantes, louées auprès de magasins, banques et établissements de crédit, permet à l'assureur de disposer d'un large échantillon d'assurés potentiels parfois déjà sensibilisés aux techniques de marketing direct. Les personnes âgées sont également visées par certaines campagnes de marketing direct avec des produits adaptés. En effet, cette catégorie d'assurés est sensible aux soins pris en charge par des produits de type assurance maladie complémentaire ou assurance hospitalisation et aux problématiques de dépendance et de frais d'obsèques.

Quant aux tests, ils jouent un rôle important dans la réussite d'une campagne de marketing. Ceux-ci sont effectués en amont et en aval du processus. En effet, les campagnes de marketing sont habituellement mises à l'essai pour un petit échantillon avant d'être étendues. Pour enregistrer, mesurer et suivre les taux de réponse résultant des tests, le département marketing du distributeur utilise des codes d'origine. Il s'agit de numéros spéciaux servant à identifier le fichier d'adresses utilisé, la campagne ou le test et l'outil de communication

utilisé pour distribuer le produit. Dès lors, en un sens, toute campagne de marketing direct est un test, ne serait-ce que pour comparer les résultats avec les prévisions faites en amont ou avec ceux des campagnes précédentes. Toutefois, il convient de s'assurer que les tests réalisés sont pertinents. En effet, l'introduction d'un trop grand nombre de variables dans un test peut rendre celui-ci difficilement interprétable si les responsables du marketing ne sont pas en mesure d'identifier la variable ou la combinaison de variables qui ont causé une variation du taux de réponse.

Section 1.3. Les outils de communication utilisés en marketing direct²

La sélection de l'outil de communication le plus apte à atteindre les assurés potentiels est un élément essentiel du marketing direct. D'abord limitées au publipostage, les techniques de marketing direct se sont étendues à la communication dans la presse, la radio et la télévision. Nous décrivons dans cette section les caractéristiques des principaux modes de communication du marketing direct.

1.3.1. Le publipostage

Le publipostage est un outil publicitaire, envoyé par courrier, qui décrit les offres de l'assureur. La publicité imprimée comprend ainsi généralement une lettre de présentation, un dépliant présentant le produit et un formulaire de demande de renseignements et/ou de proposition d'assurance.

Malgré son utilisation intensive, le publipostage reste une valeur sûre du marketing direct. Il permet de vendre et de générer des demandes de renseignements ou de rendez-vous. Cette technique est particulièrement efficace en amont d'une opération d'appels téléphoniques. Elle permet de s'appuyer sur un document concret déjà dans les mains de l'assuré potentiel. Pour être efficace, le publipostage doit utiliser une base de données avec des informations fiables et exhaustives pour permettre un envoi personnalisé.

Comparativement à d'autres outils de communication, le publipostage présente trois principaux avantages. En premier lieu, il se prête mieux à la personnalisation car il peut s'adresser à des personnes appartenant à des segments dont les caractéristiques sont connues. Ensuite, le publipostage est plus souple que la plupart des autres techniques de marketing direct. En effet, en utilisant le publipostage, le département marketing du distributeur détermine lui-même les caractéristiques du public visé, ce qui n'est pas toujours aisé lorsque la communication s'effectue par la radio ou la télévision. Le publipostage offre également plus de souplesse dans la conception et le mode de présentation du message publicitaire. Enfin, le publipostage est l'outil de communication qui se prête le mieux aux tests.

Toutefois, même si avec un bon ciblage le taux de réponse est acceptable (entre 0,5 et 3 %), le publipostage est un outil de communication dont le coût par assuré éventuel est élevé. En termes de coût d'une campagne, il faut compter entre 1 000 et 3 000 € pour réaliser un publipostage. Il faut ensuite ajouter l'impression, l'affranchissement et le coût de la location d'un fichier d'adresses de particuliers ou d'entreprises (entre 150 et 180 € pour 1 000 adresses).

²« *Marketing direct, le plébiscite des assureurs* », **L'Argus de l'assurance**, N°6 826, 21/03/2003 et sites Internet <http://www.le-marketing-direct.net/> et <http://www.companeo.com/pour-vendre-plus/prospecter-avec-les-bons-medias-:-mailing-emailing-telemarketing/checklist34>

1.3.2. La communication dans la presse, à la radio et à la télévision

La communication dans la presse désigne la publicité dans les magazines et les journaux. Le département marketing de l'assureur sélectionne les publications selon les caractéristiques démographiques, géographiques ou psychologiques de leurs lecteurs. De par sa nature, ce mode de communication touche un public large et se prête moins à la personnalisation que le publipostage. Cependant, en plaçant leur publicité dans des publications spécialisées, les assureurs peuvent cibler des groupes précis de lecteurs.

Un des avantages de la communication dans la presse est son coût relativement faible en comparaison du coût du publipostage. Cet avantage est cependant atténué par le fait que, généralement, un grand pourcentage des lecteurs ne fait pas partie de la cible visée par l'assureur. Ainsi, une grande part de la communication n'est d'aucune utilité et, par conséquent, les taux de réponse obtenus sont plus faibles que ceux produits par le publipostage. Cependant, à cause de son faible coût, la communication dans la presse peut, malgré un faible taux de réponse, donner des résultats acceptables.

La radio et la télévision sont utilisées pour la diffusion auprès du grand public de messages publicitaires dans de vastes zones géographiques. La segmentation dans la publicité diffusée au moyen de la radio ou de la télévision se fait principalement durant des émissions de type particulier, comme les événements sportifs, les journaux télévisés ou les téléfilms populaires où il est le plus probable que le public visé soit à l'écoute.

En marketing direct, la radio et la télévision offrent l'avantage de fournir un vaste choix de coûts. Les assureurs peuvent produire à faible coût des messages publicitaires radio ou dépenser plusieurs centaines de milliers d'euros dans la publicité télévisuelle. Un autre avantage de la communication radio et télévisuelle est la rapidité avec laquelle le public peut répondre au message du publicitaire. Toutefois, les principaux inconvénients sont la durée restreinte de transmission du message publicitaire et l'absence d'un service permanent servant de contact avec l'assuré potentiel. Ces deux points sont les raisons principales pour lesquelles la radio et la télévision sont surtout utilisées pour susciter des demandes de renseignement et supporter la publicité dans la presse plutôt que pour faire des offres de distribution directes.

1.3.3. Le télémarketing

Le télémarketing désigne essentiellement l'usage du téléphone pour la vente. Cet outil de communication est utilisé à la fois par l'assureur pour entrer en contact avec les assurés potentiels mais aussi par ces derniers pour obtenir des renseignements complémentaires ou souscrire un contrat.

Par rapport aux autres outils de communication, le télémarketing présente l'avantage de l'interactivité. En effet, la possibilité d'argumenter et de traiter les objections permet d'obtenir de bons taux de réponse (entre 3 et 5 %). Toutefois, les personnes contactées ont souvent l'impression d'une intrusion dans leur vie privée lorsque cet outil de communication est utilisé. La nécessité d'appeler en fin de journée et d'émettre des messages brefs sont des difficultés supplémentaires à la mise en place d'une campagne de télémarketing efficace.

Cependant, la possibilité offerte à l'assureur de proposer de nouveaux produits et d'encourager les assurés à améliorer leurs contrats en y augmentant les prestations ou en ajoutant des garanties rend attractif cet outil de communication.

Selon l'objectif de la campagne de télémarketing (prise de rendez-vous, identification d'assurés potentiels, enquête de satisfaction, etc.), les tarifs varient. Ainsi, pour contacter une centaine de personnes, il faut

compter entre 1 000 et 2 500 €. Le télémarketing peut donc être une technique très efficace mais nécessite des budgets conséquents pour obtenir des résultats significatifs.

1.3.4. Le e-marketing

Le e-marketing connaît depuis quelques années une croissance soutenue. Il s'agit d'une technique de communication qui permet de générer, suite à l'envoi d'un courrier électronique personnalisé, des demandes de devis, des prises de rendez-vous mais aussi des souscriptions pour les assureurs ayant une renommée suffisante pour créer la confiance. De plus, il bénéficie de taux de réponse plus élevés que les autres outils de communication (entre 3 et 10 %).

Section 1.4. Evolution récentes des techniques de marketing direct³

Aujourd'hui, la plupart des assureurs et des courtiers utilisent déjà les techniques de distribution évoquées précédemment et s'emploient désormais à développer les stratégies marketing de demain. Parmi ces tendances, deux phénomènes se dégagent.

Premièrement, les assureurs recourent de plus en plus fréquemment aux innovations technologiques liées aux *smartphones*. Aujourd'hui, ce marché représente en France 12,6 millions de clients potentiels et, même si cet outil de communication reste pour le moment réservé à une minorité, les assureurs parient sur son développement dans un futur proche. En effet, l'utilisation des *smartphones* pourrait permettre de contacter de futurs assurés en servant de support publicitaire, de proposer des services de déclaration de sinistres ou d'assistance, voire même devenir un outil de souscription pour des contrats comme l'assurance automobile ou la multirisque habitation.

La seconde tendance actuellement observée est le développement du marketing affinitaire. S'appuyant sur les techniques déjà existantes de distribution directe, le marketing affinitaire repose en plus sur une segmentation très précise de la population dans le but d'offrir des produits ultra-ciblés. Les principaux critères de segmentation retenus sont les étapes de la vie (naissance, études, vie de famille), les comportements (types d'achats, marques favorites) et le mode de vie (passion, métier ou habitat). Les produits pour les motards, les boulangers ou les expatriés sont des exemples de ce type de produits. Le principal intérêt du marketing affinitaire pour les assureurs est qu'il permet de profiter du bouche à oreilles au sein d'une communauté.

Section 1.5. Le business model du marketing direct

Bien que le marketing direct puisse être un bon outil pour la distribution, il n'en possède pas moins un « taux d'échec » élevé. En moyenne, seulement 1 % des personnes contactées répond à l'offre proposée. Cependant, une sélection attentive des outils de communication et un contrôle des coûts efficace peuvent assurer la rentabilité d'un faible taux de réponse. Ainsi, avec le publipostage, un taux de réponse de 0,5 % permet habituellement à un assureur d'atteindre ses objectifs de rentabilité.

Le principal objectif des distributeurs est de minimiser les coûts en maximisant le taux de réponse. Ainsi, lorsqu'ils conçoivent des produits et fixent le prix, les assureurs recourant au marketing direct établissent un

³ D'après les articles suivants « *La M-assurance fait rêver, mais des freins demeurent* », **L'Argus de l'assurance**, N°7 182, 10/09/2010 ; « *L'affinitaire a le vent en poupe* », **L'Argus de l'assurance**, N°7 206, 04/03/2011 et « *Generali investit le marché français de la m-assurance* » sur <http://www.mes-articles.net/finances/assurances>

coût admissible pour l'acquisition de chaque nouveau contrat ou euro supplémentaire de prime. La détermination de ce coût admissible fait intervenir plusieurs facteurs économiques et actuariels dont :

- Les techniques et frais de marketing direct prévus,
- Les taux de réponse prévus,
- Le taux de conversion prévu (c'est-à-dire le nombre de personnes qui concrétisent leur intention d'achat par le règlement de la prime),
- La moyenne prévue du montant de couverture,
- Le taux de sinistre prévu,
- Les charges de gestion projetées,
- La réserve mathématique du produit,
- Les profits et revenus de la compagnie produits par les objectifs d'investissements.

La détermination du coût admissible d'un produit permet à un assureur de déduire un taux de réponse minimal pour la campagne de marketing. Si les dépenses et les taux de réponse d'une campagne de marketing permettent d'acquérir de nouveaux assurés à un coût inférieur au coût admissible, la campagne est lancée. Si un profit trop faible ou des pertes sont projetés, il est fort probable que le coût de la campagne excédera son coût admissible et celle-ci est alors remise en question.

Les départements marketing du distributeur jugent des chances de réussite d'une campagne en calculant son seuil de rentabilité. Celui-ci correspond au niveau de ventes (en unités ou en euros) pour lequel les revenus d'un produit, vendu à un prix donné, égalent le total de ses coûts. Les pertes surviennent en deçà du seuil de rentabilité et les profits sont obtenus au-delà. En marketing direct, le seuil de rentabilité est le seuil théorique où les résultats du marketing produisent des revenus suffisants pour couvrir le coût admissible de l'acquisition de nouveaux assurés.

Illustrons ces notions par un exemple de calcul du seuil de rentabilité d'un produit.

<i>Hypothèses</i>	
Coût des envois postaux	200 € pour 1000 envois
Total des envois postaux	1 000 000
Coûts totaux de marketing	$(1\ 000\ 000 \div 1\ 000) \times 200\ € = 200\ 000\ €$
Coût d'acquisition admissible par euro de nouvelle prime annualisée	0,50 €
Prime annualisée moyenne prévue pour chaque nouveau contrat	100 €
Coût admissible projeté pour chaque nouveau contrat	$0,50 \times 100\ € = 50\ €$
Taux de conversion (taux des contrats émis convertis et produisant des primes)	75 %

Calculs établis d'après les hypothèses ci-dessus	
Nombre de conversions requises pour atteindre le seuil de rentabilité	$200\,000 \text{ €} \div 50 \text{ €} = 4\,000 \text{ contrats}$
Nombre de réponses requises pour atteindre le seuil de rentabilité en tenant compte du taux de conversion	$4\,000 \div 0,75 = 5\,333 \text{ réponses}$
Taux de réponse requis pour atteindre le seuil de rentabilité	$5\,333 \div 1\,000\,000 = 0,0053$
Pourcentage requis de réponses pour atteindre le seuil de rentabilité	$0,005333 \times 100 = 0,53 \%$

Source : *Adapté de Introduction to Direct Response Marketing in the Life and Health Insurance Industry*⁴,

Selon les calculs ci-dessus, l'assureur atteindrait le seuil de rentabilité s'il parvenait à obtenir de la campagne un taux de réponse de 0,53 %. Toutes choses égales par ailleurs, plus le coût admissible est élevé, plus faible est le taux de réponse nécessaire pour atteindre le seuil de rentabilité. Cependant, à cause des nombreuses hypothèses généralement introduites dans l'établissement initial du coût admissible, la prudence est de mise lors de la prévision du seuil de rentabilité. En effet, si les coûts réels de marketing, le taux de conversion ou la prime annualisée moyenne par contrat s'écartent des hypothèses, le seuil de rentabilité est différent.

Les assureurs établissent leurs hypothèses de base concernant le seuil de rentabilité non seulement d'après les coûts de la campagne de marketing mais aussi à partir des tests et des résultats antérieurs. En effet, sur certains marchés ou techniques de distribution, des facteurs tels que les coûts et les taux de conversion suivent des tendances identifiables. Bien que les projections de seuil de rentabilité ne soient pas infaillibles, elles constituent un premier moyen pour évaluer la rentabilité des projets de campagnes de marketing. A la suite de chaque campagne, les résultats effectivement obtenus sont évalués et utilisés pour prévoir et analyser les résultats probables de futures campagnes.

Chapitre 2. Stratégies actuelles et perspectives d'évolution de la distribution directe de produits d'assurance en France

Dans ce chapitre, nous commençons par dresser un panorama de la distribution directe de produits d'assurance grâce aux techniques du marketing direct à l'exception de l'utilisation d'Internet. Les acteurs de ce marché sont également présentés. Nous analysons ensuite les principaux résultats d'une étude menée par Accenture et l'Argus de l'Assurance concernant les évolutions actuelles des stratégies de distribution directe des assureurs français. Par la suite, le marché Internet, principal moteur actuel de la croissance de la distribution directe de produits d'assurance, est discuté. Enfin, nous nous intéressons aux comportements des assurés français face aux produits distribués par marketing direct.

Section 2.1. Télémarketing et marketing direct : les acteurs historiques de la distribution directe

Il est intéressant de constater que, de prime abord, les principaux assureurs du marché français sont peu nombreux à posséder une structure de distribution dédiée au marketing direct. Ainsi, Axa et sa filiale Direct

⁴ **American Council of Life Insurance/Health Insurance Association of America (1986)**, « *Introduction to Direct Response Marketing in the Life and Health Insurance Industry* », Washington (DC), p. 33 et 34

Assurance, présente sur le marché de la distribution directe depuis la fin des années 1990, fait figure d'exception. En effet, aujourd'hui les grands assureurs préfèrent bien souvent proposer des garanties directement via leur site Web et, s'ils souhaitent distribuer des produits par télémarketing ou marketing direct, ils utilisent plutôt un intermédiaire spécialisé, par exemple un courtier disposant de centres d'appels.

Ainsi, les acteurs de la distribution directe sont davantage des prestataires de services qui proposent à la fois des solutions en marketing, en produits, et en technologie que des assureurs à proprement parler. ReMark, filiale de Scor Global Life, est un exemple de ce type de sociétés.

Il existe toutefois des sociétés d'assurance qui se dédient, en partie ou totalement, à la distribution directe. Il convient de constater à ce sujet que, parmi ces assureurs, les plus connus sont anglo-saxons : MetLife, un des leaders mondiaux de l'assurance-vie et de la prévoyance salariée qui distribue ses produits via trois canaux, le courtage, les partenariats et la distribution directe, ou bien ADMS, filiale de l'assureur AEGON, spécialisée dans la conception de programmes d'assurance distribués en direct. Ce constat n'est pas dû au hasard. En effet, les pays anglo-saxons ont été les précurseurs de la distribution directe d'assurance et restent encore aujourd'hui les leaders du marché grâce à leur expertise y compris sur le marché français. Les assureurs français n'ont, jusqu'à présent, pas souhaité mettre en place de telles structures et préfèrent le plus souvent utiliser un courtier pour distribuer leurs produits en direct. Certains courtiers tels que SPB ou Filiassur sont ainsi spécialisés dans la distribution directe et disposent de leurs propres centres d'appels pour distribuer les produits de leurs partenaires assureurs. Par ailleurs, les assureurs français préfèrent aujourd'hui développer leur offre Internet via des filiales dédiées. Ce dernier point est plus largement évoqué dans une prochaine section.

Section 2.2. Evolution des stratégies de distribution multi-canal des assureurs français

Aujourd'hui, les assureurs utilisant les techniques du marketing direct ne se contentent plus de recourir à un unique outil de communication mais s'inscrivent au contraire dans une stratégie de diversification. Ainsi, les techniques du marketing direct, utilisées en appui du réseau physique traditionnel, doivent permettre à l'assuré de choisir son mode de contact en fonction de ses besoins et de ses disponibilités.

L'étude « stratégies de distribution multi-canal des assureurs français »⁵ menée par Accenture, cabinet de conseil aux entreprises, et par l'Argus de l'assurance, magazine d'actualité de l'assurance, entre le 11 décembre 2009 et le 5 mars 2010, offre l'opportunité de mieux saisir les enjeux du marché du marketing direct dans le secteur de l'assurance en France.

Un des apprentissages de l'enquête est que 93 % des participants sont en cours de définition ou de mise en œuvre d'une stratégie de distribution diversifiée. Tous prévoient que la part de leur volume d'affaires apportées par Internet augmentera et 65 % veulent également faire progresser la part du télémarketing. La part moyenne du chiffre d'affaires estimée grâce à l'association de ces deux techniques de communication serait de 10 % d'ici 2013. Ainsi, la distribution directe d'assurances individuelles en France pourrait représenter 12,5 milliards d'euros en 2013 contre 2,5 milliards d'euros en 2009.

Ces chiffres montrent une volonté d'investissement chez les assureurs. Toutefois, les stratégies mises en place diffèrent sensiblement. Ainsi, des assureurs comme AG2R-La Mondiale ou MMA ont opté pour une

⁵Ce rapport est consultable sur <http://www.accenture.com/fr-fr/Documents/PDF/EnqueteMulticanal.pdf>

Voir aussi « L'assurance à l'heure du multi-canal et d'Internet », L'Argus de l'assurance, N°7 166, 30/04/2010

uniformisation de leur offre quel que soit l'outil de communication utilisé à la souscription. En effet, partant du constat que les futurs assurés se renseignent sur Internet ou par téléphone, y font un devis, puis se rendent en agence, ces assureurs considèrent que proposer des produits ou des tarifs différents pourrait poser problème.

Cependant, une remise en cause de cette stratégie commence à apparaître chez des assureurs tels que Generali qui propose des contrats spécifiques à Internet, ou Maaf. Dès lors, nous assistons à la mise en place d'offres différenciées en fonction de l'outil de souscription choisi par l'assuré. L'effet sur les prix, avec des remises accordées pour des contrats conclus sur Internet ou par téléphone, semble d'ailleurs déjà avoir été accepté par le grand public.

Au final, la plupart des assureurs semblent privilégier un développement d'offres spécifiques puisque l'enquête révèle que 79 % des participants ont pour objectif d'adopter une stratégie différenciée par outil de souscription d'ici 2013.

En pratique, la part des souscriptions en ligne ne cesse de croître. Par exemple, chez AG2R-La Mondiale, en santé individuelle, sur 100 assurés ayant pris contact sur Internet, 60 % souscrivent totalement en ligne. Des différences de comportement générationnelles sont néanmoins encore observées. En effet, si pour les 20-35 ans l'utilisation d'Internet pour souscrire un contrat d'assurance s'est démocratisée, les générations intermédiaires mélangent Internet et le téléphone tandis que les plus de 50 ans s'informent plutôt sur Internet ou au téléphone avant de se rendre en agence. Toutefois, la part de la distribution directe n'est pas aussi importante chez tous les assureurs. A la MMA, la plupart des souscriptions se font encore en agence malgré une demande d'information importante par téléphone et sur Internet.

Ainsi, bien que réelle, l'importance actuelle de la distribution directe des produits d'assurance en France doit être modérée car elle ne représente encore que 3 % des souscriptions. Il convient de comparer ce chiffre à ceux des autres pays européens où 27 % des britanniques, 12 % des allemands et 10 % des italiens utilisent aujourd'hui les techniques de distribution directe pour souscrire leurs contrats d'assurance.

Section 2.3. Internet, un domaine d'innovation pour la distribution directe d'assurance⁶

Le constat précédent risque toutefois de changer, spécialement pour la distribution sur Internet. Dans une enquête menée en août 2010 par Assurland auprès de 5 400 internautes urbains, âgés de 25 à 55 ans et appartenant aux catégories socioprofessionnelles supérieures, 54 % des sondés considèrent qu'en matière d'assurance, Internet est avant tout un outil de souscription. En 2006, ils n'étaient que 15 %. Attention néanmoins aux résultats de cette étude qui cible principalement un public déjà sensibilisé à Internet et pas forcément représentatif de l'ensemble de la population. Par ailleurs, l'assurance sur Internet représente en 2009 entre 6 et 12 millions de visites de pages Web, 2 millions de demandes de devis et 100 000 contrats souscrits. Enfin, avec l'arrivée sur le marché de l'assurance des jeunes, familiarisés dès leur enfance à Internet et l'investissement massif des grands assureurs, la souscription sur Internet est amenée à connaître une forte croissance dans les prochaines années.

⁶D'après les articles suivants : « *La souscription en ligne monte en puissance* », **L'Argus de l'assurance**, N°7 183, 17/09/2010,

« *Assurance en ligne : le Web n'est encore qu'une vitrine* », **L'Argus de l'assurance**, N°7 173, 11/06/2010 et

« *Vente directe, l'Europe du Nord montre la voie au Sud* », **L'Argus de l'assurance**, N°7 076, 23/05/2008

De ce fait, il ressort que le marché de la distribution directe de produits d'assurance sera dans quelques années principalement le marché Internet. En effet, même si les techniques de marketing direct et télémarketing continuent à se développer, c'est Internet qui connaît la plus forte croissance, l'utilisation du Web étant aujourd'hui entrée dans les mœurs. Nous assistons aussi, grâce à Internet, à une démarche de comparaison effectuée par les assurés qui se servent du Web comme source d'information. Ainsi, en 2010, 61 % des internautes comparent les prix de leur assurance sur Internet selon le baromètre des cyberconsommateurs 2010 réalisé par l'institut de sondage Ifop. Dès lors, la souscription sur Internet, malgré une mise en place et une acceptation lente en France, est devenue un autre terrain de concurrence pour les assureurs.

Nous avons évoqué dans la section précédente la possibilité pour l'assureur d'adopter une stratégie tarifaire différenciée par outil de souscription. Les offres rencontrées sur Internet sont ainsi parfois bien différentes de celles proposées en agence. A ce titre, un des phénomènes actuellement rencontré sur le Web est celui des assurances à bas prix ou *low-cost*. Les principaux acteurs de ce marché parient sur la capacité de ces offres *low-cost* pour amorcer le décollage d'Internet. Il s'agit aujourd'hui d'Axa grâce à sa filiale « www.direct-assurance.fr », de la Macif par le biais du site internet « www.idmacif.fr », d'Allianz via « www.allsecur.fr » et de Groupama avec le site « www.amaguiz.com ».

Ainsi, les principaux acteurs français de l'assurance cherchent à s'implanter sur Internet et réalisent pour cela de lourds investissements à l'image de Groupama qui lors du lancement le 1 juillet 2008 de sa filiale 100 % Internet « Amaguiz.com » envisageait d'y investir entre 100 et 150 millions d'euros, principalement en communication⁷. De telles initiatives ont permis de faire la différence. En effet, alors que jusqu'en 2008, les acteurs de l'assurance sur Internet étaient de petites sociétés qui n'avaient pas les moyens de changer les habitudes françaises, l'arrivée des grandes sociétés d'assurances a bouleversé les pratiques existantes. Ainsi, qu'il s'agisse d'Axa, de Groupama ou d'Allianz, les grands assureurs ont investi dans le développement de la distribution directe sur Internet. Les sites des mutualistes tels que « www.idmacif.fr » ne sont pas en reste puisqu'il s'agit même des plus populaires (ils drainent 21 % du trafic global) en raison notamment de leur popularité auprès des assurés.

Par ailleurs, si officiellement, aujourd'hui, la distribution directe représente 3 à 4 % du marché, ce pourcentage reste entouré d'un certain flou. Quelle est, par exemple, la répartition entre les différents types (Internet, téléphone, marketing direct) de distribution directe ? De plus, même si les chiffres établis par la FFSA sont stables depuis 2004, il semble tout de même que ces chiffres tendent en pratique à augmenter. Ainsi, les contrats d'assurance issus du télémarketing étant souvent distribués par des courtiers, une fraction de la part de ces derniers vient, dans les faits, augmenter le pourcentage de la distribution directe dans les chiffres fournis par la FFSA. Il est donc fort probable que la part de la distribution directe représente plus que 2 ou 3 % des cotisations du marché (vraisemblablement aux alentours de 10 %). Est-il alors possible que les assureurs veuillent cacher cette tendance ? En effet, alors que leurs plans stratégiques prévoient de changer dans un futur proche leurs modes de distribution, cette volonté est pour l'instant masquée afin d'éviter d'entrer en conflit avec les réseaux de distribution classiques.

De ce fait, une autre problématique liée à l'arrivée des nouvelles offres d'assurance est de savoir comment Internet, et plus généralement les techniques de distribution directe, peuvent devenir des instruments au service des agents généraux⁸, intermédiaires traditionnels, lors de la souscription d'un contrat. Au-delà de

⁷ « Vente directe, l'Europe du Nord montre la voie au Sud », **L'Argus de l'assurance**, N°7 076, 23/05/2008

⁸ « L'avenir des agents passe par Internet », **L'Argus de l'assurance**, N°7 134, 11/09/2009

l'apparente menace que représente la distribution directe et Internet, étudions à travers l'exemple de Generali une des solutions qui peut être apportée pour créer une véritable synergie entre les deux réseaux de distribution. Nous avons expliqué précédemment que la distribution en ligne est considérée comme un moyen de bénéficier de tarifs plus avantageux. Dès lors, les assureurs conçoivent des produits spécifiques pour Internet moins chers mais aussi plus simples, afin de faciliter la souscription. Une différenciation entre l'offre disponible en agence et celle proposée en ligne s'est donc parfois opérée. Ainsi Generali, qui jusqu'à présent proposait sur le Web des produits auto et multirisque habitation identiques à ceux commercialisés en agences, vient de lancer à la demande des agents généraux une offre en ligne sans frais d'entrée et avec un choix restreint de supports. La personne qui aurait besoin de conseils peut s'adresser à un agent général, mais à des conditions tarifaires différentes.

Dès lors, comment vendre en agence des contrats avec frais d'entrée alors que des produits, certes différents, sont accessibles sans frais sur le Web ? Ceci passe par la différenciation des offres et des prix qui permet aux réseaux de se positionner sur des produits plus complexes et de mettre davantage en avant leur valeur ajoutée : analyse des besoins et de la situation des assurés, conseils, accompagnement lors d'un sinistre et prévention.

Pour conclure, la distribution directe ne remplacera probablement pas le modèle traditionnel dans les prochaines années car les assurés sont encore attachés à la proximité d'un conseiller. Néanmoins, malgré une part de marché encore faible, ces modes de distribution alternatifs, en particulier Internet, sont amenés à devenir des outils de communication incontournables pour les assureurs leur permettant de prendre contact avec de nouveaux futurs assurés avant de les faire venir en agence.

Section 2.4. Comportements des assurés⁹

Après avoir définis dans les sections précédentes l'organisation actuelle et les perspectives d'évolution de la distribution directe de produits d'assurance, cette dernière section cherche à étudier la position actuelle de l'assuré français vis-à-vis des produits d'assurance issus de la distribution directe. Après avoir rapporté un certain nombre de chiffres objectifs provenant de différents sondages, nous essayons d'appréhender les comportements de l'assuré afin de comprendre quels sont les freins encore existants au développement de la distribution directe en France.

2.4.1. Les données objectives : un attrait de plus en plus important pour les services financiers à distance

Aujourd'hui, dans les secteurs de la banque et de l'assurance, Internet constitue à la fois une importante source d'informations et un outil de gestion courant. En effet, selon les derniers chiffres publiés par la Fédération du e-commerce et de la vente à distance (FEVAD) (enquête conduite en mars 2010 auprès de 2 633 internautes par Médiamétrie et la FEVAD), Internet est considéré comme le premier outil d'information pour 36 % des personnes interrogées devant l'agence (23 %) et le téléphone (8 %). 60 % utilisent des

⁹ D'après les articles suivants : « *Les services financiers sur Internet séduisent* », **L'Argus de l'assurance**, Article Web, 11/05/2010, « *Un nouvel entrant sur le marché des comparateurs d'assurance en ligne* », **L'Argus de l'assurance**, Article Web, 07/05/2010 et « *Le boom de l'e-assurance : une réalité complexe. Bilan 2009 en chiffres avec Assurland.com* », consultable sur <http://assurances.annuairecommuniques.com/2010/02/le-boom-de-le-assurance-une-realite.html> , 19/02/2010

comparateurs pour s'informer sur l'assurance et 80 % des internautes y suivent leurs comptes, leurs placements financiers, leurs crédits ou leurs contrats d'assurance.

Cette enquête révèle également que la proximité d'un conseiller reste importante : 14 % des personnes interrogées envisagent de souscrire une offre entièrement à distance pour bénéficier de tarifs plus avantageux mais 79 % préfèrent se tourner vers un acteur disposant d'un réseau physique.

2.4.2. Les comportements de l'assuré : des freins encore existants mais des leviers d'amélioration possibles

La distribution directe de produits d'assurance a parfois une mauvaise image auprès du public. Ce constat est d'autant plus vrai pour le marketing direct et le télémarketing dont la réputation est de cibler des personnes peu ou mal informées. Pour autant, l'existence d'associations de défense des consommateurs, de forums d'informations sur le Web et de dispositifs légaux qui encadrent les pratiques des distributeurs et protègent les assurés offre de nombreux recours et incite les opérateurs à recourir à des pratiques honnêtes.

En effet, bien que le but de toute opération de marketing direct soit la distribution d'un produit ou d'un service, et que l'assurance ne fasse pas exception à cette règle, l'assureur va, à l'inverse d'un autre service, également chercher une cible d'assurés fidèles. En effet, alors que l'acquisition de la prime (généralement assez faible) a lieu de façon récurrente (mensuellement le plus souvent) tout au long de la vie du contrat, le coût d'acquisition d'un nouvel assuré représente au départ un investissement important pour l'assureur (achat de listes de numéros de téléphone ou de coordonnées personnelles, rémunération des centres d'appels et du courtier, taux de contact faible). Ainsi, il est dans son intérêt que l'assuré reste le plus longtemps possible dans son portefeuille afin d'assurer la rentabilité de l'affaire. La démarche de l'assureur ne peut alors pas être purement mercantile puisqu'il doit veiller à fournir des services de qualité pour fidéliser ses assurés.

Les services d'assurance distribués par marketing direct et télémarketing ont également l'image de produits chers et comportant de nombreux chargements. Ces chargements sont dus en partie aux coûts de financement et expliquent la rentabilité de tels produits dont le ratio sinistres sur primes se situe parfois à moins de 30 %. Or, contrairement à d'autres secteurs, l'assurance a l'obligation de communiquer sur ses frais. Le public peut alors se rendre compte des marges réalisées. Néanmoins, le niveau de marge atteint n'est pas choquant par rapport à d'autres secteurs (téléphonie mobile, logiciels informatiques, grande distribution) où il est également important (de l'ordre de 2,2 fois le coût de revient par exemple dans la grande distribution) mais pas nécessairement connu par le public.

De plus, contrairement aux autres secteurs, les marges dégagées dans le cas de l'assurance sont destinées à faire face au risque. Ainsi, seuls les produits d'assurance « risqués » (par opposition aux produits d'assurance épargne) sont vendus en marketing direct. En effet, les produits d'épargne ne sont pas adaptés car il n'est pas envisageable de réaliser des marges importantes sur ce type de contrats, l'assuré ne souhaitant pas ici se prémunir contre un risque mais placer son argent. Au contraire, pour un produit d'assurance « risqué », l'assuré, averse au risque, accepte de payer une prime plus élevée.

Par ailleurs, un point important pour l'assuré est le fait que la distribution directe s'inscrit aujourd'hui dans un cadre réglementaire, qui, bien que complexe, a le mérite d'exister. Ainsi, l'assureur doit s'acquitter des obligations d'information et de conseil propres à la distribution de l'assurance dont le traitement à distance, sans contact direct avec l'assuré, est loin d'être évident. Ceci explique l'émission, encore aujourd'hui, de papier pour justifier la délivrance de l'information et du conseil. Nous reviendrons plus

longuement sur les aspects légaux de la distribution de produits d'assurance par télémarketing et marketing direct dans un prochain chapitre.

Concernant Internet, il s'agit aujourd'hui du principal point d'entrée d'un assuré potentiel qui cherche à s'assurer sans passer par le réseau physique traditionnel. Que cela soit pour s'informer ou comparer, faire des devis ou conclure un contrat, l'assuré commence le plus souvent par faire une recherche sur le Web. Pour autant le contrat est rarement conclu sur Internet, l'assuré préférant encore souvent se diriger vers le réseau physique ou rappeler un numéro de téléphone dédié pour avoir un interlocuteur pouvant le conseiller et répondre à ses questions.

Ainsi, bien que la plupart des freins à la souscription sur le Web (sécurité du paiement, confiance de l'assuré, etc.) soient aujourd'hui levés laissant présager que l'utilisation d'Internet comme outil de contractualisation devienne de plus en plus courante, un certain nombre d'entraves à l'essor de la souscription sur le Web subsistent toujours. Un des obstacles encore existant est la problématique de la qualité de l'information fournie à l'assuré. En effet, bien qu'Internet constitue une source de renseignement, en pratique, le processus d'information n'est pas encore optimal pour le futur assuré. Ainsi, l'utilisation d'Internet ne se traduit pas encore concrètement par une simplification de la procédure de souscription. En effet, l'accès aux conditions contractuelles et aux tarifs ne peut se faire qu'après avoir rempli un questionnaire long et souvent inquisitoire. Par ailleurs, alors que le futur assuré attend en premier lieu de la clarté et de la précision sur les exclusions et les franchises, l'information précontractuelle est souvent lourde et surabondante. Dès lors, nous comprenons pourquoi les contrats d'assurance (et particulièrement les plus importants par leurs enjeux financiers, leur complexité ou leur durée) ne sont pas encore totalement souscrits via Internet. Malgré tout, il semble raisonnable de penser que la souscription en ligne présente suffisamment d'avantages pour l'assuré, plage horaire illimitée, possibilité d'effectuer des comparaisons tarifaires, coûts de distribution réduits, lecture du contrat à son rythme, etc., pour que cet outil de souscription se développe rapidement dans un futur proche.

Concernant la gestion des sinistres, les assurés cherchent d'abord à être rapidement et correctement indemnisés. A ce titre, un moyen de pression déterminant pour les assurés s'est construit à travers les échanges sur Internet via les forums de discussion et d'échange sur la qualité des produits ou services. L'image de l'assureur est un élément particulièrement important car un assuré satisfait en parle à son entourage ce qui constitue pour l'assureur un levier d'amélioration important de la perception de sa marque et un élément multiplicateur d'affaires nouvelles. Ainsi, selon une étude réalisée en mai 2011 auprès de 1 471 particuliers équipés de produits d'assurance (auto, MRH, santé) par le cabinet Molitor Consult¹⁰, spécialisé en marketing et communication en assurance, 52,6 % des internautes affirment qu'ils seront dans l'avenir plus attentif à la réputation sur Internet des assureurs. Il est d'ailleurs intéressant de constater à ce sujet que les internautes ont davantage tendance à se manifester en cas d'insatisfaction que de satisfaction : 49,3 % seraient prêts à ne pas recommander un assureur sur les réseaux sociaux en cas d'insatisfaction contre 30,4 % prêts à recommander en cas de satisfaction.

Chapitre 3. Le rôle du réassureur

Le rôle du réassureur dans le développement de la distribution directe est particulièrement important. En effet, il constitue souvent un support financier et stratégique pour les acteurs, assureurs ou distributeurs, souhaitant mettre en place une campagne de marketing direct.

¹⁰ Résultats de l'enquête disponibles sur le site de Molitor Consult : <http://www.molitorconsult-assurance.com>

Dans ce chapitre, nous commençons par analyser les relations entre les différents acteurs d'une campagne de marketing direct afin de mieux comprendre le rôle du réassureur. Nous présentons ensuite les atouts du réassureur dans l'utilisation des techniques du marketing direct pour la distribution de contrats d'assurance.

Section 3.1. Rôle du réassureur dans la mise en place d'une structure de distribution directe

La mise en place d'une campagne de marketing direct fait intervenir de nombreux acteurs. Le schéma ci-dessous a pour but de donner un aperçu des relations, parfois complexes, qui peuvent exister entre les différents protagonistes ainsi que de la place du réassureur dans une telle structure.

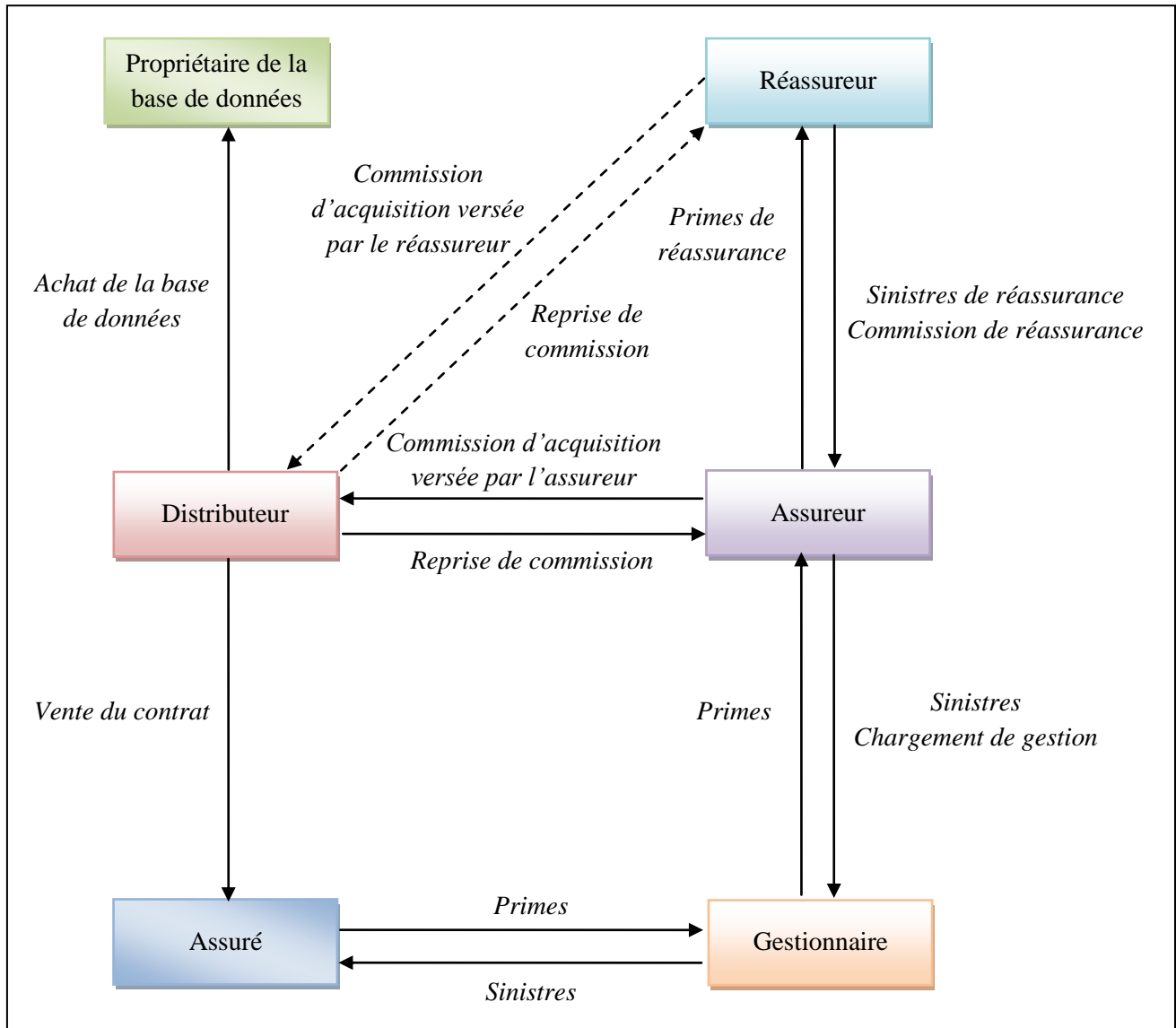


Figure 1 : Structure de distribution en marketing direct

La distribution directe de produits d'assurance implique un nombre d'interactions entre les différents protagonistes bien plus élevé que dans une structure de réassurance classique ne faisant intervenir que l'assureur, l'assuré et le réassureur. Lors d'une campagne de marketing direct, trois nouveaux acteurs entrent en jeu : le distributeur, le plus souvent un courtier, le gestionnaire et le propriétaire de la base de données.

Dans certains cas, l'assureur peut être le distributeur, le gestionnaire voire même le propriétaire de la base de données.

Cependant, la base de données est bien souvent fournie par d'autres acteurs de la distribution à distance qui opèrent dans des secteurs d'activité variés (par exemple entreprise de vente de livres à distance dans le domaine de la culture). Cette base est ensuite vendue au distributeur qui se sert du fichier transmis pour contacter les assurés potentiels pour le compte d'un assureur en échange du versement d'une commission d'acquisition pour chaque police placée. Le réassureur, en plus de ses missions habituelles de conception de produits et d'assistance sur la souscription des risques, finance aussi souvent partiellement ou totalement le paiement des commissions au distributeur. Le réassureur se pose donc comme un preneur de risque pour le compte de la cédante en l'aidant à financer sa structure de distribution.

Enfin, le principe de la reprise de commission est le suivant : pour chacune des polices résiliées avant une période donnée (généralement un an), le distributeur rembourse la commission d'acquisition perçue à la souscription *pro rata temporis* de la durée pendant laquelle la police n'est plus dans le portefeuille. Par exemple, si l'assuré résilie au bout de trois mois, le distributeur doit rembourser à l'assureur et/ou au réassureur l'équivalent de $\frac{3}{4}$ de la commission d'acquisition. Ainsi, celle-ci n'appartient réellement au distributeur qu'après un an.

Section 3.2. Les atouts du réassureur dans l'utilisation de techniques de distribution alternatives pour les produits d'assurance

De par la nature de son métier et l'environnement international dans lequel il évolue, le réassureur a acquis une expérience précieuse pour ses partenaires. En effet, ces derniers peuvent s'appuyer sur l'expertise du réassureur sur les problématiques en rapport avec la gestion du risque, des capitaux et des données.

Concernant la gestion du risque, le réassureur possède les moyens techniques et humains nécessaires à la mise en place de procédures de sélection des risques efficaces. En effet, les réassureurs disposent à la fois d'une expérience approfondie des marchés et sont à la pointe des innovations dans ce domaine. Un exemple d'innovation dans les processus de sélection des risques et de tarification est d'ailleurs développé au chapitre 5.

Le réassureur profite également de son savoir-faire en matière de conception de produits d'assurance pour créer des produits à la fois simples, faciles à comprendre par l'assuré sans pour autant augmenter le risque de sinistre, et utilisant des processus de distribution adaptés aux besoins du type d'assurés ciblé.

Par ailleurs, le réassureur possède une expérience en gestion du capital et peut proposer des structures de financement des coûts d'acquisition à ses partenaires.

Enfin, le réassureur dispose d'une expertise en termes de gestion des bases de données qu'il peut mettre au service de ses partenaires, courtiers ou assureurs, afin de leur proposer des résultats d'expérience améliorant ainsi les processus de distribution et de sélection des risques. L'utilisation de modèles analytiques et prédictifs doit ainsi permettre au réassureur d'identifier les meilleures cibles d'assurés potentiels pour ensuite permettre aux centres d'appels de se concentrer sur des groupes spécifiques à fort potentiel.

Chapitre 4. Les aspects légaux liés à la distribution directe de produits d'assurance

Du fait des obligations liées aux conditions d'émission du contrat d'assurance, la distribution directe de tels produits a imposé une synthèse de différentes réglementations.

Section 4.1. Cadre légal

4.1.1. Origine et champ d'application

Toutes les formes de distribution directe d'assurance ont été visées par une ordonnance publiée en juin 2005¹¹. Jusqu'alors, la souscription par téléphone de contrats d'assurance, pourtant de pratique ancienne, n'avait fait l'objet d'aucune disposition dans les codes¹².

En effet, en droit français, le contrat d'assurance est formé dès la rencontre des consentements de l'assureur et de l'assuré. La conclusion du contrat pouvait ainsi avoir lieu sans qu'il y ait le moindre échange de papier et sans formalisme particulier.

Reste que, notamment en assurance vie, la souscription par téléphone était incompatible avec le devoir d'information préalable que doit respecter l'assureur avant de faire souscrire un contrat. Il paraît en effet compliqué de lire une notice d'information ou un projet de contrat à son interlocuteur au téléphone.

Le développement de la vente de produits financiers par téléphone puis par Internet a nécessité qu'un dispositif soit instauré pour protéger les droits des assurés. C'est l'objet de l'ordonnance n° 2005-648 du 6 juin 2005 qui a transposé la directive n° 2002/65/CE du Parlement européen du 23 septembre 2002 concernant la distribution directe de services financiers auprès des consommateurs dans le code des assurances (article L. 112-2-1 du code des assurances). Des dispositions similaires ont été introduites pour les mutuelles dans le code de la mutualité (article L. 221-18 I du code de la mutualité) et pour les institutions de prévoyance dans le code de la sécurité sociale (article L. 932-15-1 du code de la sécurité sociale). Le principe utilisé est celui de la technique des codes « pilotes » et « suiveurs » puisque des dispositions contenues dans le code de la consommation ont été introduites dans les différents codes mentionnés ci-dessus (articles L. 121-20-8, L. 121-20-9, L. 121-20-11, L. 121-20-13, L. 121-20-14, L. 121-20-15, L. 121-20-16 du code de la consommation)¹³.

4.1.2. Définition légale de distribution directe

Depuis le 1^{er} décembre 2005, date d'entrée en vigueur de cette ordonnance, toutes les formes de distribution à distance de l'assurance se trouvent réglementées de manière identique (article L. 121-20-8 du code de la consommation).

La technique de communication à distance est définie comme un moyen qui, sans présence physique simultanée du fournisseur et du consommateur, peut être utilisé pour la commercialisation d'un service entre les parties. Sont donc visés les systèmes de souscription par courrier, par téléphone et par Internet.

¹¹ Ordonnance n° 2005-648 du 6 juin 2005 relative à la commercialisation à distance de services financiers auprès des consommateurs. Consultable sur <http://admi.net/jo/20050607/ECOX0500090R.html>

¹² « La vente à distance soumise à ordonnances », *L'Argus de l'assurance*, N°6 962, 27/01/2006

¹³ Voir Annexe 1 : Articles de loi concernant la commercialisation à distance de contrats d'assurance

Pour que le texte s'applique, deux conditions doivent néanmoins être remplies :

- Le contrat d'assurance doit être entièrement conclu à distance. Le distributeur a utilisé, pour la formation du contrat, « exclusivement une ou plusieurs techniques de communication à distance jusqu'à, et y compris, la conclusion du contrat » (article L. 121-20-8 al. 1er du code de la consommation). L'essentiel réside donc dans le fait que les contractants n'aient pas été en présence l'un de l'autre à un moment quelconque du processus.
- Le distributeur doit s'adresser à un consommateur. La notion de consommateur est définie par l'article L. 121-20-16 2°a) du code de la consommation et précise que le consommateur d'assurance est « le souscripteur, personne physique, qui agit à des fins qui n'entrent pas dans le cadre de son activité commerciale ou professionnelle ». Est un consommateur la personne qui souscrit un contrat d'assurance couvrant un risque qu'il encourt durant sa vie privée. Si le même individu souscrit à distance un contrat pour garantir un risque relatif à son activité professionnelle, il ne bénéficie pas de la protection prévue par le dispositif.

Tous les contrats d'assurance vie ou non-vie et tous les services financiers (banque, investissement, épargne, retraite) sont soumis à cette ordonnance. Toutefois, dans le cas de contrats renouvelables par tacite reconduction, elle ne s'applique que lors de la conclusion du contrat. En effet, les reconductions ultérieures sont considérées comme étant des actes d'exécution et ne relèvent donc plus du processus de distribution.

Section 4.2. L'obligation d'information de l'assureur et les sanctions en cas de non respect

La loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique¹⁴ oblige une identification du distributeur qui doit décliner son identité, son adresse et ses immatriculations (en particulier son immatriculation à l'ORIAS, registre des intermédiaires en assurance, comme l'impose déjà l'article L. 512-1 du code des assurances).

L'article L. 112-2-1 III du code des assurances ajoute que, pour tout contrat conclu à distance, l'assureur doit, par écrit et avant la conclusion du contrat, communiquer à l'assuré les termes et conditions générales contractuels ainsi que les informations suivantes :

- la dénomination sociale de l'entreprise d'assurance contractante ;
- l'adresse de son siège social ;
- son numéro d'immatriculation au registre du commerce et des sociétés (ou au registre national des mutuelles pour ces dernières) ;
- les coordonnées de l'autorité chargée de son contrôle ainsi que, le cas échéant, l'adresse de la succursale qui propose la couverture ;
- le montant total de la prime (ou la base de calcul de celle-ci) ;
- la durée minimale du contrat ainsi que les garanties et exclusions prévues par celui-ci ;
- la durée pendant laquelle les informations fournies sont valables ;
- l'existence éventuelle et les modalités d'exercice du droit de renonciation ;
- les modalités de conclusion du contrat et de paiement de la prime ;
- l'éventuel coût supplémentaire spécifique à l'utilisation d'une technique de commercialisation à distance ;

¹⁴ **Loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique.** Consultable sur : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000801164&dateTexte>

- la loi applicable au contrat et la langue que l'assureur s'engage à utiliser pendant la durée du contrat ;
- les modalités d'examen des réclamations.

En outre, l'assureur devra garder trace que toutes les informations nécessaires ont bien été fournies de façon claire et complète à l'assuré. En pratique, pour s'assurer de la délivrance de cette information, il est possible pour l'assureur d'enregistrer les conversations téléphoniques (l'assuré devant être préalablement informé de l'enregistrement de la communication) ou de faire valider l'assuré sur Internet par l'intermédiaire d'une case prévue à cet effet.

En cas de non respect de ces obligations, l'assureur encourt un report du délai de renonciation de l'assuré jusqu'à la communication des informations ou des conditions contractuelles (article L. 112-2-1 II 1° du code des assurances).

Enfin, dans le cadre de leur devoir de vigilance à l'égard des assurés, il appartient à l'assureur de vérifier l'identité de l'assuré et de la personne payant les primes, si celle-ci est différente, avant la conclusion du contrat d'assurance (article L. 561-2 du code monétaire et financier). En pratique, l'assureur doit obtenir une copie d'un document officiel d'identification avec une photographie, par exemple, un passeport ou une carte nationale d'identité. Une telle identification n'est pas exigée quand la prime est payée par virement du compte de l'assuré, géré par une banque française réglementée par l'Autorité des Marchés Financiers.

Section 4.3. Le droit à renonciation de l'assuré

L'article L. 112-2-1 II 1° du code des assurances affirme que toute personne physique ayant conclu un contrat à distance, quelle que soit la technique, bénéficie d'un délai de renonciation de quatorze jours en assurance non-vie et de trente jours en assurance vie sans avoir à justifier de motif ni à supporter de pénalités. Dans les deux cas le délai court du jour où l'assuré est informé que le contrat est conclu. Néanmoins, le point de départ du délai est reporté jusqu'au jour où l'assuré reçoit l'information précontractuelle lorsque l'assureur n'a pas respecté cette obligation.

Pour exercer ce droit, l'assuré doit envoyer à son assureur une lettre de renonciation en recommandé avec demande d' accusé de réception. Toutefois, les polices d'assurance voyages ou bagages (ou des polices d'assurances similaires à court terme d'une durée inférieure à un mois), l'assurance automobile obligatoire et les contrats « exécutés intégralement par les deux parties à la demande expresse de l'assuré, avant que ce dernier n'exerce son droit de rétractation » (article L. 112-2-1 II 3° du code des assurances) échappent à cette faculté de renonciation.

Par ailleurs, l'assuré doit être informé de l'existence du droit de renonciation, de sa durée et des modalités pratiques de son exercice, notamment l'adresse à laquelle la notification de la renonciation doit être envoyée (article L. 121-20-16, 5° du code de la consommation) ainsi que du montant de la prime que l'assureur peut lui réclamer en contrepartie de la prise d'effet de la garantie, à sa demande expresse, avant l'expiration du délai de rétractation. Un modèle de lettre destiné à faciliter l'exercice du droit de renonciation doit également être communiqué.

Section 4.4. La conclusion du contrat à distance

Depuis la publication de l'ordonnance n° 2005-648 du 6 juin 2005 concernant la distribution à distance de contrats d'assurance auprès des assurés et sa transcription dans le code des assurances, l'assureur est tenu de communiquer à l'assuré, par écrit et en temps utile, les termes et conditions générales contractuels avant la conclusion du contrat. L'assureur a alors différents moyens de fournir cette information et de conclure le contrat.

4.4.1. Par signature vocale

En cas de conclusion du contrat par téléphone, l'acceptation peut se faire par signature vocale. Dès lors, l'assuré doit recevoir toutes les informations nécessaires lors de la discussion téléphonique qui doit être claire et sans équivoque. Autrement dit, l'acceptation doit être pure et simple. Par exemple, à la question « acceptez-vous l'offre d'assurance que nous vous soumettons ? », la réponse de l'assuré doit être « oui ». De plus, l'entretien devant être enregistré à fin de preuve, l'assuré doit, au préalable, être informé de l'enregistrement de la communication¹⁵.

Des contrats d'assurance peuvent donc être conclus oralement. Cependant, l'article L. 121-20-11 du code de la consommation, retranscrit dans l'article L. 112-2-1 du code des assurances, exige qu'une version écrite du contrat soit envoyée à l'assuré juste après l'acceptation de l'offre. Ainsi, lorsque le contrat d'assurance a été conclu oralement, une version écrite doit être envoyée à l'assuré juste après l'acceptation de l'offre et le délai de renonciation courra à compter du jour où l'assuré aura reçu cette information.

4.4.2. Par signature d'une lettre recommandée électronique

Le décret n° 2011-144 du 2 février 2011¹⁶ fixe les modalités d'application de l'article 1369-8 du code civil qui autorise l'envoi d'une lettre recommandée relative à la conclusion ou à l'exécution d'un contrat par courrier électronique. Le texte précise les obligations de l'opérateur, le « tiers chargé de l'acheminement » de la lettre recommandée par voie électronique. Préalablement à la mise en œuvre de la procédure d'envoi de la lettre recommandée électronique, l'utilisateur doit être informé de ses caractéristiques (avec ou sans avis de réception, contenu imprimé sur papier ou non, niveau de garantie contre les risques de perte, vol ou détérioration) et connaître l'identité du tiers chargé de l'acheminement.

Le décret fixe également les modalités relatives à l'identification de l'expéditeur et du destinataire ainsi que du prestataire qui assure, le cas échéant, la distribution de la lettre recommandée sous format papier. Il indique les mentions obligatoires que doit comporter la preuve de dépôt et de distribution (numéro d'identification de l'envoi, date et heure du dépôt électronique du message et, le cas échéant, l'identification du prestataire de services postaux chargé de la remise de la lettre recommandée imprimée sur papier : sa raison sociale, ses adresses postale et de courrier électronique).

Dans le cas d'une distribution de la lettre recommandée électronique dont le contenu a été imprimé sur papier, le décret prévoit une procédure de mise en instance de la lettre recommandée en cas d'absence du

¹⁵ Communication interne de Transamerica Re suite à une étude réalisée auprès d'un cabinet d'avocat spécialisé relative à la capacité de pratiquer la Télé-Souscription en France.

¹⁶ Décret n° 2011-144 du 2 février 2011 relatif à l'envoi d'une lettre recommandée par courrier électronique pour la conclusion ou l'exécution d'un contrat. Consultable sur <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023513151&dateTexte=&categorieLien=id>

destinataire. S'il s'agit d'une distribution électronique, le décret fixe la procédure permettant au destinataire d'accepter ou de refuser l'envoi pendant un délai de quinze jours.

Enfin, le tiers chargé de l'acheminement doit mettre à la disposition de l'utilisateur une adresse électronique et un dispositif lui permettant de déposer une réclamation.

4.4.3. Par validation en ligne

Dans le contexte de la distribution d'assurance sur Internet, il est important de rappeler qu'en droit français, le principe du consensualisme trouve application pour la formation du contrat d'assurance, de sorte que celui-ci est formé dès la rencontre des consentements de l'assureur et de l'assuré. Toutefois, nous avons vu que l'application du régime de la distribution à distance imposait une information spécifique du futur assuré. Par conséquent, afin de fournir à l'assuré l'information devant légalement être dispensée, le site Internet doit prévoir un écran spécifique dédié aux informations à remettre et s'assurer de la parfaite délivrance de cette information par la validation en ligne d'une case spécifique. Dès lors, sous réserve que l'assureur ait satisfait à son obligation d'information et de conseil par écrit auprès du client (art L 520-1 II du code des Assurances), la rencontre des volontés peut parfaitement se concevoir sur le site internet de la société distributrice.

Chapitre 5. Tarification et souscription à distance : la télé-souscription

Afin d'optimiser le processus de sélection des risques et de tarification conjointement à l'utilisation des méthodes de distribution présentées précédemment, Transamerica Reinsurance utilise et cherche à promouvoir sur le marché français la technique de la télé-souscription, largement développée dans les pays anglo-saxons depuis la fin des années 1990. La télé-souscription est une suite logique au télémarketing puisqu'elle permet d'évaluer le risque assuré au moment même où la vente du produit d'assurance est réalisée. Cependant, et même si les deux processus peuvent avoir lieu simultanément, il faut bien faire la différence entre le télémarketing, qui a pour objectif de distribuer un produit d'assurance, et la télé-souscription dont l'objectif est d'évaluer le risque amené par le futur assuré.

Section 5.1. Définition de la télé-souscription

D'après le *Dictionary of Insurance Terms*¹⁷, la télé-souscription est la procédure au cours de laquelle un sondeur possédant ou non une expérience en souscription pose des questions par téléphone à un assuré potentiel. En fonction des réponses fournies, le sondeur demande alors davantage de détails ou bien passe à une autre question. Cette méthode permet donc de recueillir des informations relatives au risque directement auprès du futur assuré, par le biais d'un entretien téléphonique enregistré.

5.1.1. Utilisation et développement

La télé-souscription peut être utilisée pour compléter un questionnaire de santé ou pour tarifier réellement si les capitaux sont sous le niveau de la grille de sélection (c'est-à-dire la grille qui, en fonction de l'âge de

¹⁷ **Rubin H. (2000)**, "Dictionary of Insurance Terms, 4th edition", Barron's Educational Series, p257

l'assuré et du niveau de garantie qu'il désire, détermine les examens médicaux à réaliser afin que l'assureur puisse décider d'accepter ou non le risque¹⁸).

Les États-Unis ont été le premier marché à la fin des années 1990 à utiliser le téléphone comme outil de collecte de données. Depuis, ce modèle a été suivi en Europe (principalement au Royaume-Uni et en Irlande), en Afrique du Sud et en Australie.

5.1.2. Procédure

Il existe deux modèles de télé-souscription, plus ou moins adaptés au type d'assuré et de réseau de distribution¹⁹ :

- Petit modèle : Dans ce modèle, le téléphone est utilisé pour obtenir de l'information additionnelle ou manquante sur le questionnaire de santé. Ce modèle n'aboutit pas à une décision quant à l'acceptation ou non du risque. Le coût moyen d'un entretien dans le petit modèle s'élève en moyenne à 20 €.
- Grand modèle : Dans ce second modèle, il est fait usage du téléphone pour obtenir de l'information additionnelle et tarifier réellement grâce à un manuel de tarification électronique. Le but de ce modèle est de pouvoir prendre une décision à l'issue de l'entretien. En moyenne, le traitement d'un futur assuré revient à environ 30 €.

Plusieurs options sont offertes aux acteurs de l'assurance pour mettre en place leur processus de télé-souscription. Certains réassureurs fonctionnent avec des logiciels de tarification spécialisés (c'est le cas de RGA avec la plateforme AURA® au Royaume-Uni et en Allemagne) utilisés par des sous-traitants en télé-entretien (Morgan Ash, par exemple) pour gérer le processus de télé-souscription auprès de l'assuré. On peut également citer le cas d'Allfinanz, société du groupe d'assurance Allianz, récemment rachetée à 100 % par Munich Re, qui fournit un moteur de règles pour la télé-souscription. Les entreprises clientes n'ont alors donc plus qu'à calibrer cet outil en fonction de leur politique de sélection des risques.

L'entretien de télé-souscription peut-être réalisé par un souscripteur ou par un conseiller en centre d'appel. Lors de l'entretien, le télé-souscripteur couvre une liste standard de questions médicales adaptées à la classe d'âge et au profil de risque de l'assuré potentiel. Les réponses sont directement enregistrées sous forme de données électroniques pour accélérer le traitement du dossier.

¹⁸ Voir Annexe 2 : Exemple de grille de sélection emprunteur

¹⁹ Documentation interne Transamerica Re et « Les défis de la télé-souscription », **Newsletter Scor**, Juin 2009

5.1.3. Aspects légaux

5.1.3.1. La réglementation relative à la protection des données personnelles²⁰

La collecte et la conservation des données personnelles par téléphone est sujette à des restrictions en vertu de la législation française.

En effet, celle-ci exige que le traitement des données personnelles et les moyens mis en œuvre pour ce traitement par une entité française, sur le territoire français, soient déclarés et autorisés par la Commission Nationale de l'Informatique et des Libertés (CNIL). Cette formalité préalable est désormais réalisée à travers la demande d'agrément auprès du ministère de la santé s'agissant de données personnelles mais également à caractère médical (article R. 1111-10 du code de la santé publique). Une fois obtenue cette autorisation, l'assureur est tenu à une obligation d'information à l'égard de l'assuré sur la finalité de l'information recueillie et sur ses droits d'opposition, d'accès et de rectification de ses données²¹. Par ailleurs, la limitation de la durée de conservation des données au temps strictement nécessaire aux finalités pour lesquelles elles sont collectées et traitées impose à l'assureur d'informer le futur assuré sur la durée de conservation de ses données.

L'assuré doit donc être informé de l'enregistrement de la communication et l'assureur doit recueillir le consentement exprès de l'assuré au traitement de ses données personnelles (article L. 1111-8 du code de la santé publique). En pratique, lorsque le questionnaire téléphonique est complété, il est rappelé à l'assuré que l'appel a été enregistré et que les réponses obtenues doivent être rigoureusement exactes. L'assuré est ensuite invité à donner son accord verbal, qui est dûment enregistré, tenant lieu de signature électronique.

5.1.3.2. La réglementation relative au secret médical

Le principe du secret médical, selon lequel toute personne prise en charge par un professionnel, un établissement, un réseau de santé ou tout autre organisme participant à la prévention et aux soins a droit au respect de sa vie privée et au secret des informations la concernant (article L. 1110-4 du code de la santé publique), s'applique aux sociétés d'assurance et de réassurance, au personnel salarié de ces sociétés et à l'ensemble du réseau d'intermédiaires.

Posé en des termes larges par le législateur dans le but de contraindre réellement tout détenteur potentiel d'une information médicale personnelle, il s'étend à la conservation des données sur support informatique et impose :

- un agrément de l'hébergeur Internet des données de santé, (article R. 1111-10 du code de la santé publique)
- un accord exprès de l'assuré, (article L. 1111-8 du code de la santé publique)
- des conditions particulières d'hébergement (articles L. 1111-8 et suivants du code de la santé publique).

²⁰ Documentation interne Transamerica Re

²¹ **Loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés**, version consolidée au 19 mai 2011. Consultable sur <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068624&dateTexte=20110817>

Ayant rempli les conditions préalables et obtenu l'agrément de la part du Ministère de la santé (articles R. 1111-9 et R. 1111-12 du code de la santé publique), l'assureur est libre d'interroger l'assuré sur son état de santé afin de déterminer le risque à couvrir et de pouvoir apprécier l'exactitude et la réalité de la déclaration des risques faite par l'assuré en cas de sinistre. Les questions ne doivent cependant pas porter sur la vie privée de l'assuré.

De même que pour ses informations personnelles, l'assuré dispose d'un certain nombre de droits concernant les données médicales qui ont été recueillies à son sujet : droit d'opposition, d'accès et de rectification. Il doit également être averti de la finalité des renseignements collectés, des procédures mises en place afin de sécuriser l'information recueillie et de la durée de conservation des données communiquées. Enfin, le sondeur doit nécessairement obtenir le consentement exprès du futur assuré au recueil et à la conservation de ses données médicales protégées afin d'assurer la légalité de sa démarche et la validité du contrat.

Section 5.2. Avantages et limites de la télé-souscription²²

5.2.1. Avantages

Il existe trois raisons principales pour une entreprise de recourir à la télé-souscription.

Tout d'abord, il a été constaté une diminution du nombre de fausses déclarations lorsque les contrats sont souscrits à l'aide de cette méthode. En effet, le fait que l'appel soit enregistré encourage les proposants à donner des réponses plus précises et plus complètes. Ainsi, selon les statistiques sur le marché britannique de Medical Direct, principal fournisseur en sous-traitance de gestion des risques et services médicaux pour les assureurs, il y aurait seulement de 3 à 4 % de fausses déclarations avérées via la télé-souscription à comparer aux 11 % de la sélection traditionnelle. En raison de ces résultats, l'assureur peut donc « alléger » les tableaux de sélection. En un mot, la télé-souscription engendre peu d'omissions et de fausses déclarations lors des demandes d'adhésion. De ce fait, les assureurs améliorent la sélection des risques, ce qui tend à réduire la sinistralité constatée conduisant à terme à une baisse des tarifs.

Ensuite, l'utilisation de la télé-souscription permet de réduire les coûts de souscription et le délai de prise de décision : puisque le nombre d'erreurs ainsi que l'usage aux rapports médicaux et bilans sanguins est diminué grâce à l'utilisation de questionnaires ciblés, le délai de traitement se trouve également réduit et la décision peut être communiquée dans les quarante-huit heures suivant réception. D'après l'expérience de Medical Direct sur le marché britannique, une décision finale peut être donnée dans 85 % des cas d'entretien de télé-souscription pour les couvertures décès et 65 % des cas pour les couvertures incapacité/invalidité. Ainsi, la télé-souscription permet de diminuer les coûts d'acquisition des contrats, 60 à 100 € sans télé-souscription en fonction des assurés et du montant de garantie, grâce au gain de temps important qu'elle procure.

Enfin, avec la télé-souscription, le futur assuré est au centre du processus : chaque appel est préparé et cherche à tendre vers l'efficacité et l'établissement d'une relation de confiance via les questions ciblées du sondeur.

²² Documentation interne Transamerica Re

5.2.2. *Limites*

Les réserves que l'utilisateur peut avoir à l'égard de la télé-souscription sont principalement dues à des limites légales en particulier concernant les données médicales. Une autre question soulevée est celle de la quantité de données que l'assureur risque d'avoir à traiter en utilisant la télé-souscription. En effet, si l'assuré fournit plus d'information au cours d'un entretien téléphonique, il y aura de manière concomitante besoin de recourir à davantage d'examen médicaux complémentaires.

De plus, l'efficacité d'un tel processus nécessite de recourir à des audits fréquents afin de s'assurer de la qualité des entretiens réalisés en n'oubliant pas que la finalité de la démarche reste la sélection médicale et non pas la collecte d'informations.

Enfin, comme pour tout processus de sélection médicale, l'utilisation de la télé-souscription entraîne un effet d'anti-sélection. Ainsi, l'assureur peut assister à l'attraction des bons risques si le niveau de sélection médicale opérée est élevé ou à l'afflux de mauvais risques dans le cas contraire.

Section 5.3. **Expérience dans d'autres pays et développement en France²³**

La pratique de la télé-souscription s'est largement développée aux Etats-Unis depuis la fin des années 1990. Une étude des pratiques associées à ce procédé et de son impact a été réalisée auprès de dix-huit assureurs américains utilisant la télé-souscription. Cette étude révèle que, parmi ces sociétés :

- 83 % utilisent la télé-souscription quels que soient les capitaux assurés sur les produits distribués par télé-souscription,
- 94 % utilisent la télé-souscription quel que soit l'âge des assurés,
- 50 % externalisent tous les entretiens téléphoniques,
- 33 % font leurs entretiens en interne,
- 17 % font les deux (centralisent et externalisent),
- 2/3 développent le questionnaire du télé-entretien en interne,
- 1/3 utilisent des questionnaires remis par des prestataires de service,
- 80 % ont expérimenté une réduction des rapports médicaux complémentaires.

Les pratiques de télé-souscription sont donc très variées mais cette technique se révèle globalement efficace et permet un gain de temps dans l'établissement de la police d'assurance. Aujourd'hui, de nombreux modèles existent et la télé-souscription est utilisée par la plupart des assureurs vie pour la gestion et la sélection des risques en assurance décès temporaire, maladie redoutée et incapacité.

En France, l'utilisation de la télé-souscription est beaucoup plus tardive et se situe encore dans sa phase de lancement. Néanmoins, de plus en plus d'assureurs français envisagent d'utiliser la télé-souscription pour des produits d'assurance emprunteur ou dépendance. Cette pratique semble bien adaptée au marché français et en particulier à l'assurance emprunteur. En effet, bien qu'il suffise la plupart du temps de remplir un bref questionnaire pour appliquer des taux forfaitaires, de nombreux dossiers exigent un questionnaire de santé rempli par un médecin, augmentant ainsi de quatre à cinq semaines le processus d'acceptation par l'assureur et d'approbation du prêt. Or, il a été démontré sur les marchés utilisant les techniques de télé-souscription, que les informations collectées par les professionnels de la télé-souscription réduisaient de manière significative la nécessité d'un questionnaire de santé. Par ailleurs, la loi n° 2010-737 du 1er juillet 2010 portant réforme de

²³ Documentation interne Transamerica Re

l'assurance emprunteur dite « loi Lagarde »²⁴ stipule que « le prêteur ne peut pas refuser en garantie un autre contrat d'assurance dès lors que ce contrat présente un niveau de garantie équivalent au contrat d'assurance de groupe qu'il propose. Toute décision de refus doit être motivée. ». Dès lors, le marché de l'assurance emprunteur est ouvert à la concurrence puisque l'assuré n'est plus obligé de souscrire à l'assurance de l'établissement prêteur. Dans un tel contexte, l'utilisation de la télé-souscription est amenée à faire la différence pour les assureurs qui réussiront ainsi à proposer de meilleurs tarifs à leurs assurés.

Du côté des réassureurs vie, où la sélection des risques est cruciale pour le développement de portefeuilles équilibrés, des initiatives sont également apparues. Citons notamment le lancement en novembre 2009 de SCOR Telemed²⁵, société du groupe SCOR Global Life dédiée à la télé-souscription. Cette nouvelle entité a produit un logiciel pour réaliser les télé-entretiens et possède un système de sélection des risques automatisé permettant de traiter les cas des plus simples aux plus complexes.

²⁴ Loi n° 2010-737 du 1er juillet 2010 portant la réforme de l'assurance emprunteur consultable sur <http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000022419094&categorieLien=id>

²⁵ Communiqué de presse de Scor à l'occasion du lancement de sa filiale dédiée à la télé-souscription consultable sur http://www.scor.com/images/stories/pdf/CP-2009/CP_2009_36_FR.pdf

PARTIE II : BESOIN DE FINANCEMENT ET PROFITABILITE SOUS DIFFERENTS REFERENTIELS COMPTABLES

L'utilisation des techniques de marketing direct pour la distribution de produits d'assurance pose la question du financement de coûts d'acquisition élevés pour l'assureur.

Introduction : relations entre l'assureur et le distributeur pour un contrat d'assurance conclu à distance

Cette introduction présente certaines caractéristiques des contrats d'assurance issus de la distribution directe afin de mieux appréhender les points développés dans la suite de cette partie.

Dans le cas de la distribution directe, les trois protagonistes sont en général le distributeur, l'assureur et l'assuré. La figure ci-dessous représente les principaux flux entre ces trois parties :

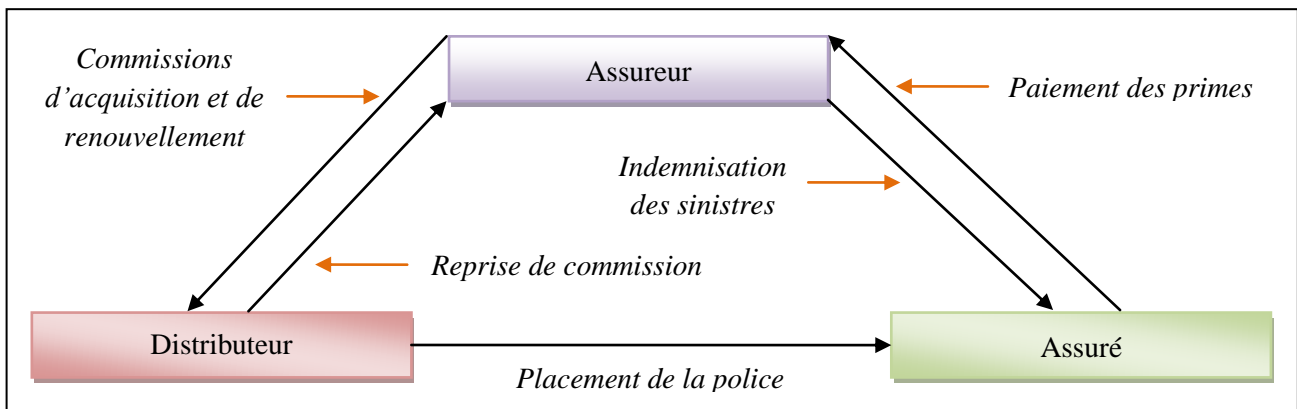


Figure 2 : Relations entre les différentes parties dans un contrat d'assurance conclu à distance

Analysons désormais chacun des flux afin de comprendre les motivations et intérêts de chaque partie.

L'assureur verse au distributeur (bien souvent un courtier) une commission d'acquisition pour chaque police placée. Cette commission d'acquisition, payée au commencement du contrat, est importante : elle peut être supérieure à 100 % de la prime annualisée. L'intérêt du distributeur est alors de placer le plus de contrats afin de toucher cette commission sans se soucier si le contrat est résilié ensuite. Cela va à l'encontre de l'intérêt de l'assureur qui veut garder les assurés le plus longtemps possible en portefeuille afin de rentabiliser la police émise. L'assureur cherche donc à se protéger en incitant le distributeur à avoir une politique de souscription de qualité. Les clauses de reprise de commission ou de participation aux bénéfices, les commissions de renouvellements dans le cadre de la gestion pour tiers sont des outils qui peuvent donner un alignement d'intérêt sur la fidélisation du portefeuille.

Rappelons ici le principe de la reprise de commission : pour chacune des polices résiliées avant une période donnée (généralement un an), le distributeur rembourse la commission d'acquisition perçue à la souscription *pro rata temporis* de la durée pendant laquelle la police n'est plus dans le portefeuille. Par exemple, si l'assuré résilie au bout de trois mois, le distributeur doit rembourser à l'assureur l'équivalent de $\frac{3}{4}$ de la commission d'acquisition. Ainsi, celle-ci n'appartient réellement au distributeur qu'après un an. Enfin, le distributeur touche également tous les ans, après la première année, une prime de renouvellement si l'assuré

n'a pas résilié. Dès lors, si l'assureur (ou le réassureur) engage des fonds à l'origine dont il ne tire un profit que plus tard, ces deux dispositifs, reprise de commission et commission de renouvellement, incitent le distributeur à rechercher des assurés qui ne résilieront pas immédiatement.

Il convient alors de se demander comment cet investissement initial est enregistré dans les comptes de l'assureur. Est-t-il considéré comme une charge à prendre en considération l'année de l'acquisition ou comme un investissement que l'assureur peut amortir sur la durée de vie du contrat ? La réponse dépend de deux facteurs : le type de contrat souscrit et le référentiel comptable dans lequel l'assureur se place.

Dans cette partie du mémoire, nous étudions la rentabilité d'un produit issu de la distribution directe sous différents référentiels comptables. Après avoir introduit les concepts et problématiques propres au financement originel des produits nécessitant un fort investissement initial, nous nous intéressons à l'impact de la méthode comptable utilisée sur l'émergence du profit pour un tel produit. Pour ce faire, nous avons développé dans le cadre du mémoire un modèle simplifié de calcul de performance afin de comparer les méthodes comptables et essayer d'obtenir une meilleure vision économique de la rentabilité.

Chapitre 1. Prise en compte des coûts d'acquisition sous différents référentiels comptables

La charge de première année liée à la distribution des produits souscrits à distance est importante et peut constituer une perte pour la compagnie d'assurance. Ce chapitre a pour objectif de présenter les leviers qui permettent une répartition des charges d'acquisition et le type d'émergence de résultats selon plusieurs référentiels comptables.

Les commissions et autres frais d'acquisition sont très supérieurs aux profits de première année pour les contrats souscrits à distance (pour rappel, ils peuvent être supérieurs à 100 % de la prime annualisée). L'assureur a alors intérêt à ne pas avoir à les comptabiliser en charge dès la première année mais à les considérer comme des frais d'investissement amortissables sur toute la durée de vie du contrat. L'assureur crée alors un actif pour représenter cet investissement. Il s'agit des coûts d'acquisition différés (*Deferred Acquisition Cost* ou DAC chez les anglo-saxons) qui permettent de lisser les dépenses de première année dans le temps et ainsi de réaliser un profit dès la première année. La rentabilité sous-jacente du contrat peut ainsi apparaître de manière adéquate. Toutefois, la définition et la comptabilisation des coûts d'acquisition différés ne sont pas les mêmes et ne s'appliquent pas aux mêmes contrats selon le référentiel utilisé. Donnons alors les éléments nécessaires pour comprendre ces différences.

Section 1.1. Définition et intérêts d'un référentiel comptable

L'ensemble des principes comptables généralement admis (*Generally Accepted Accounting Principles* ou GAAP) sont des normes, conventions et procédures qui définissent les pratiques comptables autorisées et devant être utilisées par une entreprise pour établir des documents comptables. Ces principes fournissent des critères à partir desquels les comptables peuvent établir ou juger des états financiers²⁶.

L'évolution de ces normes comptables subit aussi bien l'influence de l'environnement comptable international que de l'histoire, de l'économie ou de la politique. Ainsi, si aujourd'hui une volonté

²⁶ **Bernheim Y. (2000)**, « *L'essentiel des US GAAP, 2ème édition* », Maxima, p19-25

d'harmonisation des pratiques comptables a conduit à l'élaboration du référentiel international IFRS, le référentiel américain est toujours très présent sur la scène internationale du fait du poids économique des marchés financiers des Etats-Unis.

Chaque pays possède néanmoins ses propres GAAP. Nous parlons ainsi, par exemple, de GAAP français (ou French GAAP) ou de GAAP américains (ou US GAAP). Toutefois, les objectifs visés par ces GAAP peuvent différer d'un pays à l'autre. Aux Etats-Unis, et plus généralement dans les pays anglo-saxons, les normes GAAP servent à informer les investisseurs qui veulent que les entreprises mettent à leur disposition des informations leur permettant de considérer un prêt ou un investissement. En France, pour les comptes sociaux notamment, l'objectif est aussi d'ordre fiscal et statistique pour le besoin des différents régulateurs.

Section 1.2. Coûts d'acquisition différés selon les normes comptables américaines²⁷

Un principe essentiel des normes comptables américaines (US GAAP) est la nécessité de concordance entre la reconnaissance des dépenses et des revenus. Concrètement, un coût ne devient une dépense que lorsqu'il est reconnu en charge dans le compte de résultat. La dépense n'est donc pas le coût lui-même mais l'amortissement dans le temps de ce coût en contrepartie d'un revenu à reconnaître. Ainsi, certains coûts survenus lors de l'acquisition d'un contrat d'assurance ne doivent pas être comptabilisés comme une charge dans la période comptable au cours de laquelle ils sont engagés, mais comme un actif au bilan et progressivement amortis sur la durée de vie du contrat d'assurance. Les normes US GAAP définissent ces coûts « différables » comme les coûts dépendant directement de l'acquisition de nouveaux contrats. Cette définition assez générale est sujette à interprétation. Les frais de souscription au sens large (rémunérations des souscripteurs, examens médicaux, etc.) et autres commissions d'acquisition sont ainsi susceptibles d'entrer dans cette définition. Toutefois, tous les coûts d'acquisition n'entrent pas dans le périmètre des dépenses « différables ». A ce titre, la norme ASU 2010-26, en vigueur à partir de 2012, émise par le *Federal Accounting Standards Board*, concepteur des normes US GAAP, explicite les coûts qui peuvent être amortis ou non en DAC²⁸. Ainsi, les coûts associés aux nouveaux contrats mais qui ne varient pas directement avec l'acquisition de nouvelles polices ne peuvent pas être différés. Par exemple, la mise en place d'une campagne de marketing, le recrutement et la formation de nouveau personnel ou le développement de produits ne sont pas des coûts « différables » car ils ne dépendent pas directement de l'acquisition d'une nouvelle police. Ces coûts sont alors comptabilisés comme une charge pour la période.

Dans le contexte de ce mémoire, le montant des coûts d'acquisition différés (DAC) correspond au montant de la commission d'acquisition pour le réassureur. Ces coûts sont considérés comme un investissement financé par les profits futurs et peuvent donc être différés puis amortis sous forme de DAC. La technique d'amortissement des DAC dépend de la classification du produit. Pour les entreprises d'assurance (ou de réassurance) sur la vie, la norme FAS 60 doit s'appliquer. Selon cette norme, les coûts d'acquisition différés doivent être amortis annuellement en proportion des primes encaissées puis finissent par s'annuler à l'extinction du contrat. L'amortissement des DAC impacte les résultats annuels successifs. La compagnie enregistre par ce moyen un profit plus homogène au cours du temps.

²⁷ **Herget R. (2006)**, « *US GAAP for life insurers* », Society of Actuaries, p81-122

²⁸ « *New Accounting Rule – To DAC or not to DAC* », **The Messenger**, Mars 2012, Consultable sur : http://www.scor.com/images/stories/pdf/library/messengers/m1q12_web.pdf

Section 1.3. Coûts d'acquisition différés selon les normes comptables françaises

Pour les normes comptables françaises, il faut distinguer les comptes sociaux des comptes consolidés. D'une part, le code des assurances s'applique aux comptes sociaux des entreprises. D'autre part, dans une logique d'information financière harmonisée, les sociétés européennes cotées sont tenues de présenter leurs comptes consolidés selon les normes IFRS (Règlement n° 1606/2002 du Parlement européen et du Conseil²⁹).

1.3.1. Coûts d'acquisition différés reportés dans le code des assurances

Le code des assurances définit la notion de frais d'acquisition reportés en partant de l'idée que, pour conclure un contrat d'assurance et le faire entrer en portefeuille, une compagnie peut être amenée à engager, avant même la signature, des frais importants. Ces frais sont souvent enregistrés en charge dans l'année même où ils sont engagés. Pour autant, ces derniers ne sont pas toujours récupérés dans leur intégralité auprès de l'assuré selon le rythme auquel ils ont été acquittés : ainsi sont-ils, en général, recouverts tout au long de la durée du paiement des primes alors qu'ils ont été engagés pour l'essentiel à la souscription du contrat. L'objectif recherché par l'inscription à l'actif d'une partie de ces frais et leur étalement sur la durée de vie résiduelle du contrat (c'est-à-dire sur la période qui s'écoule jusqu'au terme du contrat) est un meilleur rattachement des charges aux produits.

Le code des assurances définit les frais d'acquisition reportés et leur mode d'évaluation et de comptabilisation dans les articles R. 332-33 et R. 332-35³⁰. Sont concernés par ce dispositif les contrats d'assurance à primes périodiques d'assurance sur la vie, d'assurance nuptialité-natalité, et de capitalisation. Les entreprises pratiquant ces opérations doivent inscrire à l'actif de leur bilan « la fraction non imputable à l'exercice des frais d'acquisition des contrats constatés en charge de l'exercice ». Ainsi selon l'article R. 332-35, « les entreprises (...) inscrivent au bilan les frais d'acquisition à reporter (...) ». Enfin, le code des assurances précise que « les frais d'acquisition à reporter sont inscrits à l'actif en fonction de la durée de vie résiduelle des contrats » (art R. 332-35).

La composition des frais d'acquisition devant être reportés n'est pas précisée. Les sociétés d'assurance restent donc libres d'incorporer différentes natures de frais dans la mesure toutefois où :

- ils correspondent à la définition de frais d'acquisition donnée par le code des assurances,
- ils respectent le plafonnement imposé par le code des assurances,
- ils ont bien été réellement engagés par l'entreprise.

Le montant des frais d'acquisition reportés est au plus égal à l'écart entre les montants de provisions mathématiques inscrites au bilan conformément à l'article L. 331-1 du code des assurances et le montant des provisions mathématiques qui seraient à inscrire si les chargements d'acquisition n'étaient pas pris en compte dans les engagements des assurés. Dès lors, bien que la notion de frais d'acquisition reportés s'apparente dans son approche au concept de coûts d'acquisition différés, les méthodes de calcul et de comptabilisation sont différentes.

²⁹ Règlement (CE) No 1606/2002 du Parlement Européen et du Conseil du 19 juillet 2002 sur l'application des normes comptables internationales. Consultable sur :

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:243:0001:0004:fr:PDF>

³⁰ Voir Annexe 3 : Articles de loi concernant les commissions et frais d'acquisition à amortir

Enfin, l'article R. 332-5 précise que « les frais d'acquisition reportés sont admis en représentation des provisions techniques ». Ainsi, selon l'article R. 331-5-1 du code des assurances, « la provision pour frais d'acquisition reportés doit être constituée pour un montant égal au montant des frais d'acquisition reportés en application des dispositions de l'article R. 332-35 » Ainsi, les frais d'acquisition inscrits à l'actif doivent être égaux, à tout instant, à la provision pour « frais d'acquisition reportés » inscrite au passif.

1.3.2. Les frais d'acquisition reportés dans les comptes consolidés

Depuis 2005, toutes les sociétés cotées en Europe sont tenues de présenter leurs comptes consolidés selon les normes IFRS (règlement n° 1606/2002 sur l'application des normes comptables internationales). Le Comité de Règlementation Comptable (CRC) a rendu différents recommandations et avis. Dans le cadre de ce mémoire, il convient de s'attacher plus particulièrement au règlement n° 2000-05 du CRC relatif aux règles de consolidation des entreprises régies par le code des assurances et des institutions de prévoyance régies par le code de la sécurité sociale ou par le code rural³¹. En effet, le paragraphe 30012 « Frais d'acquisition reportés » indique que « les charges différées relatives aux frais d'acquisition des contrats doivent être enregistrées de la façon suivante :

- Activité vie : les frais d'acquisition sont reportés dans la limite des marges nettes futures des contrats considérés, y compris la marge financière dûment justifiée, notamment lorsqu'il existe un écart entre le taux d'actualisation retenu et le taux de rendement prévisionnel des actifs prudemment évalué. Ils sont amortis sur la base du rythme de reconnaissance de ces marges futures, réévaluées à la clôture de chaque exercice. Le cas échéant, ils font l'objet d'un amortissement exceptionnel dans la mesure où les marges futures deviennent insuffisantes eu égard au plan d'amortissement ;
- Activité non vie : le calcul des frais d'acquisition reportés est effectué sur une base cohérente avec celle utilisée pour le report des primes non acquises. Ces frais sont amortis sur la durée résiduelle des contrats considérés ;
- Une information est fournie dans l'annexe comptable de la société sur les modalités et les durées d'amortissement pour ce poste de l'actif du bilan consolidé. »

Le règlement n° 2000-05 du CRC autorise donc le report des frais d'acquisition. En assurance vie, les frais directement affectables à l'acquisition des contrats d'assurance sont portés à l'actif des comptes consolidés dans la limite de la valeur actuelle des bénéfices futurs des contrats. Ces frais sont amortis sur la durée de vie des contrats en fonction de la cadence d'émergence des marges futures pour chaque génération de contrats, les marges futures étant déterminées à partir d'hypothèses économiques (taux d'actualisation, taux de rendement des actifs et taux de résiliation). Ainsi, chaque année, par famille homogène de produits, la valeur actuelle probable des marges futures est comparée au total des frais d'acquisition reportés nets des amortissements déjà comptabilisés dans le passé. Si cette valeur est inférieure, un amortissement exceptionnel est constaté en charge. En assurance non vie, les frais d'acquisition afférents aux primes non acquises sont reportés et inscrits à l'actif du bilan.

³¹ Consultable sur <http://admi.net/jo/20010119/ECOT0020238A.htm>

Section 1.4. Frais d'acquisition reportés selon la norme IFRS contrats d'assurance phase 2

Les principes comptables de prise en considération des frais d'acquisition reportés pour les comptes consolidés des sociétés cotées sont issus de la norme « IFRS 4, contrats d'assurance » qui traite de la comptabilisation des contrats d'assurance. Cette norme, en vigueur depuis 2005, est provisoire dans l'attente d'une norme définitive dont les contours ne sont pas encore totalement connus à ce jour. De ce fait, l'*International Accounting Standards Board* (IASB), concepteur des normes IFRS, travaille actuellement sur la phase 2 du projet pour remplacer la norme IFRS 4 actuelle. Cette section revient sur les conséquences de la mise en place de la norme assurance phase 2, espérée pour 2013 ou 2014, sur les frais d'acquisition reportés tels que nous les connaissons dans la phase 1 de la norme. Le document de référence utilisé est l'*exposure draft*³² (ou « prénorme ») publié le 30 juillet 2010 par l'IASB et qui donne les derniers avancements des travaux menés.

Dans la phase 2 de la norme, les coûts d'acquisition directement induits par un nouveau contrat comme, par exemple, les commissions versées aux intermédiaires d'assurance et tous les frais liés à la souscription, seraient inclus dans les flux de trésorerie servant à déterminer la provision technique au lieu d'être différés sous forme d'un actif. Tous les autres frais d'acquisition devraient être enregistrés en charge aussitôt qu'ils sont encourus. Les frais à reporter seraient donc les frais supportés par la souscription d'un contrat supplémentaire et qui n'auraient donc pas été encourus si le contrat n'avait pas été émis. Les frais qui ne passeront pas immédiatement en charge devraient de ce fait être plus restreints que ceux actuellement reportables selon le règlement CRC n° 2000-05. Pour rappel, ce règlement autorise le report des frais d'acquisition (directs voire indirects) dans la limite des marges futures positives des contrats. A l'image de l'évolution de la définition des DAC en US GAAP, les normes IFRS tendent donc aussi à donner une définition plus rigoureuse des frais d'acquisition reportés.

Concernant leur comptabilisation, selon l'*exposure draft*, les frais d'acquisition variables, inclus à l'origine du contrat dans les flux de trésorerie estimés viendraient en déduction de la marge résiduelle (définie dans l'*exposure draft* comme « la marge destinée à éliminer tout gain à la souscription et à reconnaître le profit au fur et à mesure que le service est rendu, et relâchée sur la période de couverture du contrat selon un rythme déterminé à la souscription puis bloqué »). Le profit différé que représente cette marge se trouverait donc diminué au passif puisque les frais d'acquisition différés enregistrés directement en déduction du passif d'assurance impacteraient le résultat par le biais de l'amortissement de la marge résiduelle nette.

³² Consultable et téléchargeable sur

<http://www.ifrs.org/Current+Projects/IASB+Projects/Insurance+Contracts/Exposure+draft+2010/Exposure+draft+and+Comment+letters.htm>

L'approche proposée par l'*exposure draft* est ainsi illustrée ci-dessous.

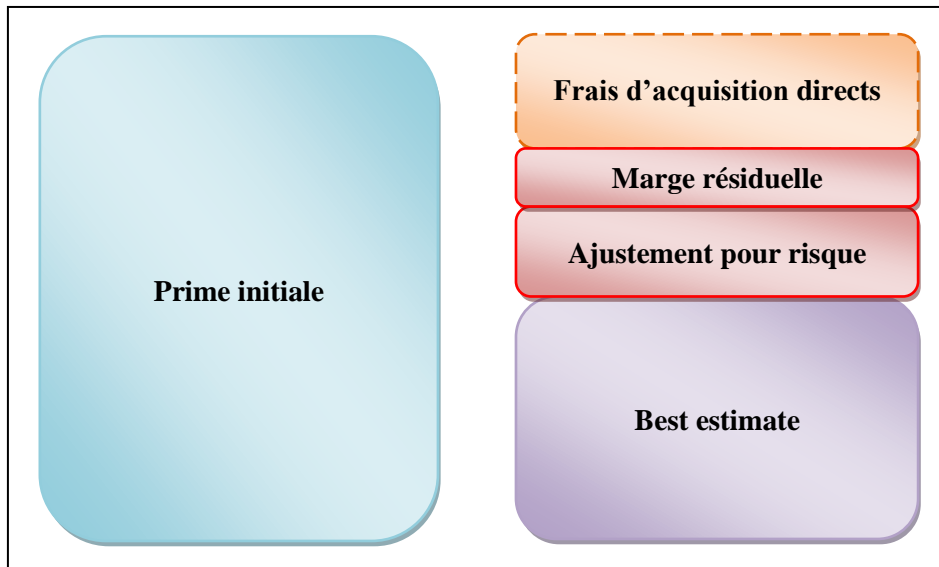


Figure 3 : Approche par les flux du calcul des provisions techniques proposée par l'exposure draft

Source : Norme IFRS assurance phase 2 : vers quel modèle de reporting financier³³

En conclusion, l'assiette et la nature des frais d'acquisition pouvant être constatés a une importance dans la reconnaissance des profits. Aujourd'hui, cette question n'est toujours pas complètement tranchée dans le cadre de la mise en place de la norme « IFRS 4, contrats d'assurance ». Malgré une volonté de convergence des normes américaines et des normes IFRS, des différences subsistent en particulier quant à la façon de reconnaître les frais d'acquisition différés. En réassurance, cette problématique reste toutefois plus marginale dans la mesure où ces frais, dans notre cas les commissions d'acquisition, sont facilement qualifiables.

Chapitre 2. Application : émergence du résultat pour un produit avec commission d'acquisition

Dans ce chapitre, nous avons voulu comparer l'influence de la méthode de calcul des provisions techniques et le type de prise en compte des coûts d'acquisition différés sur l'émergence du résultat pour un contrat d'assurance avec financement initial. Pour cette application, un modèle simplifié de calcul de performance a été réalisé. Nous nous sommes placés dans un cadre purement théorique afin de comparer les options d'émergence du résultat. Ainsi, le type de produit, vie ou non-vie, n'a pas été pris en compte. Il convient néanmoins de rappeler que la nature de produit peut avoir une importance comme évoqué dans le chapitre précédent avec les normes comptables françaises.

Section 2.1. Caractéristiques du contrat d'assurance modélisé

Le produit d'assurance modélisé dans cette application numérique est un contrat garantissant le versement d'indemnités journalières (IJ) pour une hospitalisation liée à une maladie ou un accident. Ce type de produit, issu du télémarketing prend en compte des commissions d'acquisition et de renouvellement. A ce titre, nous avons considéré la structure décrite en figure 2 de l'introduction de la partie II du mémoire. Nous nous plaçons du point de vue de l'assureur (ou du réassureur) qui reçoit des primes, indemnise les sinistres et paie

³³ Consultable sur <http://www.pierretherond.fr/?tag=ifrs>

une commission d'acquisition initiale et des commissions de renouvellement annuelles au distributeur. Nous ne considérons pas de clause de reprise de commission. Le versement des primes, la gestion des sinistres et le calcul des provisions techniques sont réalisés mensuellement. Nous avons opté pour une agrégation annuelle pour la présentation des résultats. Le tableau ci-dessous fait la synthèse des hypothèses retenues.

<i>Nombre de contrats en portefeuille</i>	100 000
<i>Age moyen des assurés</i>	30 ans
<i>Taux de chargement annuel (en % de prime annuelle)</i>	65
<i>Frais de gestion (en % de prime annuelle)</i>	5
<i>Commission d'acquisition (en % de prime annuelle)</i>	100
<i>Commission de renouvellement (en % de prime annuelle)</i>	20
<i>Taux d'imposition sur les résultats (en %)</i>	12,5
<i>Incidence moyenne (en % annuel)</i>	4,0
<i>Garantie journalière (en euros)</i>	100
<i>Nombre de jours moyen de versement de la garantie</i>	10
<i>Taux de rendement des actifs (en %)</i>	4,50
<i>Taux d'actualisation (en %)</i>	4,75

Tableau 2 : Principales hypothèses techniques retenues pour l'application numérique

Par ailleurs, la loi suivante a été retenue :

<i>Mois</i>	<i>Taux de résiliation mensuel (en %)</i>
0	15,0
1	7,0
2	5,0
3	2,5
4	2,5
5	2,5
6	2,5
7	2,5
8	2,5
9	2,5
10	2,5
11	2,5
12	2,5
> 12	1,5

L'application numérique couvre une période de 30 ans à l'issue de laquelle la persistance des assurés en portefeuille est nulle.

Section 2.2. Méthodes de calcul des provisions techniques

Au-delà des hypothèses retenues, l’émergence du résultat dépend directement de la méthode de calcul des provisions techniques. Dans cette application, deux méthodes ont été implémentées à savoir les approches par la *Net Premium* et par la *Gross Premium*. Cette section présente le calcul théorique des provisions techniques selon ces deux approches. Dans les formules présentées, la notion de valeur actuelle probable est abrégée en VAP.

2.2.1. Utilisation de la Net Premium pour le calcul des provisions techniques

Dans l’approche par la *Net Premium*, la provision technique à la date t est égale à :

$$NPReserve_t = VAP_t (\text{engagement de l'assureur}) - VAP_t (\text{engagement de l'assuré})$$

où

$$\begin{aligned} \text{engagement de l'assureur} &= \text{prestations futures} \\ \text{engagement de l'assuré} &= \text{paiements futurs de la Net Premium} \end{aligned}$$

et la Net Premium est une valeur théorique obtenue selon le principe d’équivalence actuarielle :

$$VAP_0 (\text{engagement de l'assureur}) = VAP_0 (\text{engagement de l'assuré})$$

La *NPReserve* est donc une provision technique prospective dont le calcul est basé sur le risque porté par l’assuré et ne fait intervenir ni le montant réel de la prime payée par l’assuré, ni les charges et commissions futures. Deux façons de calculer la *NPReserve* sont présentées dans l’application : la *NPReserveBestEstimate* et la *NPReserveAvecPAD*. En effet, initialement, dans l’équation du calcul de la Net Premium, les flux sont établis dans le cas d’un scénario qui est considéré comme le plus probable (*best estimate scenario*). Si ces hypothèses se réalisent, des profits apparaissent proportionnellement aux primes. Cependant, il faut prendre en considération la possibilité que ces éléments puissent être modifiés. D’où la mise en place d’une provision pour déviation (*Provision for adverse deviation - PAD*). Cette provision est utilisée pour étudier l’éventualité selon laquelle les hypothèses seraient moins favorables que prévues. Par conséquent, cette provision est calculée en prenant des hypothèses plus conservatives que celle du *best estimate scenario*. Dès lors, il est intéressant de quantifier le rôle des PAD sur le calcul de la *NPReserve* et l’émergence du résultat.

Nous comparons dans l’application numérique les trois scénarii suivants :

- Les résultats sans DAC avec provisions *Net Premium*,
- Les résultats sans DAC avec provisions *Net Premium* + PAD,
- Les résultats avec DAC avec provisions *Net Premium* + PAD (US GAAP).

Le tableau ci-dessous présente les hypothèses retenues pour le calcul des PAD.

<i>Incidence moyenne</i>	+ 10 %
<i>Taux de résiliation mensuel</i>	+ 10 %
<i>Taux d’actualisation</i>	- 0,25 %

2.2.2. Utilisation de la Gross Premium pour le calcul des provisions techniques

Dans l'approche par la *Gross Premium*, la provision technique à la date t est égale à :

$$GPReserve_t = VAP_t (\text{engagement de l'assureur}) - VAP_t (\text{engagement de l'assuré})$$

où

$$\text{engagement de l'assureur} = \text{prestations futures} + \text{charges et commissions futures}$$

$$\text{engagement de l'assuré} = \text{paiements futurs de la Gross Premium}$$

et la *Gross Premium* est une valeur théorique obtenue selon le principe d'équivalence actuarielle :

$$VAP_0 (\text{engagement de l'assureur}) = VAP_0 (\text{engagement de l'assuré})$$

Il s'agit à nouveau d'une provision technique prospective mais dans laquelle toutes les dépenses futures (charges et commissions) sont prises en compte. Dans cette approche, la commission d'acquisition est incluse à l'origine dans les provisions techniques. Ainsi, le montant de la provision diminue lorsque la commission d'acquisition est payée.

Les normes comptables irlandaises utilisent l'approche par la *Gross Premium* avec prise en compte de PAD, scénario que nous étudions dans l'application numérique. Les hypothèses retenues pour le calcul des PAD sont identiques à celles de l'approche par la *Net Premium*.

Section 2.3. Méthodologie retenue pour le modèle

Le but de l'application numérique est de produire des graphiques donnant les flux de trésorerie annuels, en nominal et en pourcentage des primes, dans le cadre des hypothèses présentées précédemment.

2.3.1. Cadre général du modèle

La gestion du produit modélisé étant mensuelle pour le versement des primes, des indemnités et le calcul des provisions, nous avons dans un premier temps réalisé les différents calculs selon un pas de temps mensuel. Dans un second temps, nous avons agrégé les quantités obtenues de façon à pouvoir donner des résultats après impôt et des flux de trésorerie annuels. La figure ci-dessous illustre la structure générale du modèle après agrégation des différentes quantités calculées sur une base mensuelle.

Duration (annuelle)	Persistance (début d'année)	Primes payées	Sinistres	Provisions (début d'année)	Variation de provisions
0					
1	100%	5 225 801	1 752 276	0	84 372
2	60%	3 907 417	1 310 206	84 372	66 407

Commission	Frais de Gestion	Résultat technique	Commission d'acquisition	DAC amortis	Résultat financier
		-7 157 503	7 157 503	0	
0	-261 290	3 127 863		0	1 898
-781 483	-195 371	1 553 950		0	5 291

Capital (début d'année)	Intérêt sur Capital	Résultat avant impôt	Résultat après impôt	Flux de trésorerie	DAC (fin d'année)
1 256 142		-7 157 503	-6 262 815	-7 518 957	-
917 128	48 899	3 178 659	2 781 327	3 120 341	-
688 418	36 125	1 595 366	1 395 945	1 624 656	-

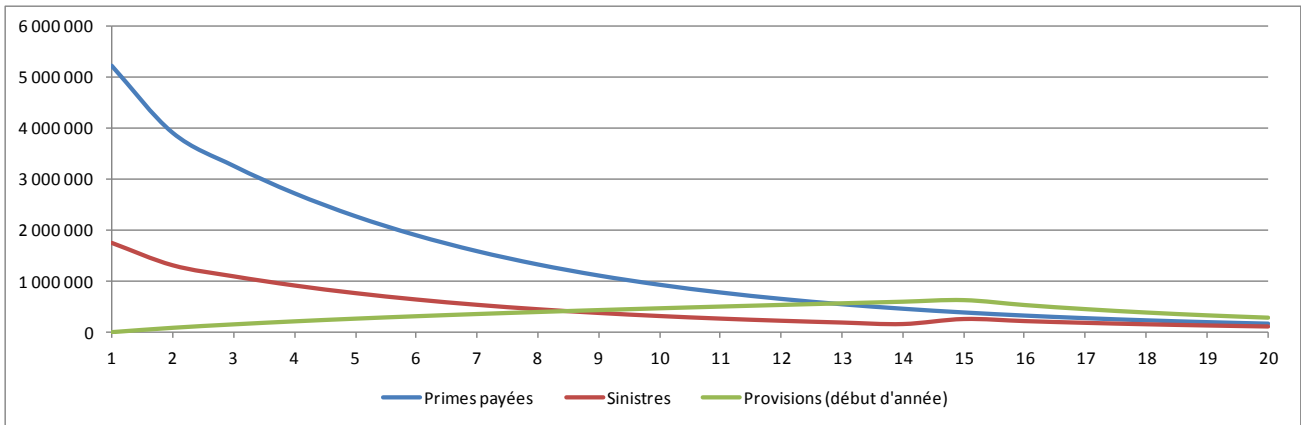


Tableau 3 : Structure générale du modèle comptable annuel

Il convient de préciser que la première ligne est destinée à calculer le montant de la commission d'acquisition et du capital requis. Ce dernier, dont la formule est donnée par le modèle interne de la compagnie ou le régulateur, représente le capital à immobiliser par les actionnaires pour mettre en place l'affaire. Ce capital entre notamment en jeu dans le calcul du *Return on Equity* présenté dans la section suivante.

2.3.2. Influence des coûts d'acquisition différés

L'application numérique développée cherche à illustrer l'importance pour le réassureur de pouvoir, dans une approche *Net Premium*, mettre en place des coûts d'acquisition différés pour lisser son résultat et éviter de constater une perte très importante au cours de la première année du contrat.

Pour rappel, le montant des coûts d'acquisition différés (DAC) est égal au montant de la commission d'acquisition. En effet, les DAC sont un actif mis en place à la souscription du contrat de façon à annuler au niveau du résultat les pertes initiales résultant des charges d'acquisition du contrat. Dès lors, deux techniques d'amortissement sont possibles et étudiées dans l'application. Dans le référentiel comptable américain, les DAC sont amortis annuellement selon le rythme auquel le réassureur encaisse les primes tandis que selon les normes IFRS, les DAC doivent être amortis en fonction de la cadence d'émergence des marges c'est-à-dire en fonction de l'émergence des profits. Dans les deux cas, l'amortissement des DAC impacte les résultats annuels successifs mais moins brutalement qu'en impactant dès l'origine le financement du contrat. Nous donnons dans les paragraphes suivants, les éléments de calcul de l'amortissement des DAC selon les deux référentiels.

Selon les normes US GAAP, les coûts « différables » sont amortis proportionnellement aux primes de l'année. Le montant de DAC à reconnaître au bilan en fin d'année s'obtient alors à partir de la formule suivante :

$$DAC_n = DAC_{n-1} + \text{Intérêts sur } DAC_{n-1} - DAC_n^{\text{amortis}}$$

où

- DAC_{n-1} est égal au montant des DAC à la fin de l'exercice précédent ;
- *Intérêts sur DAC_{n-1}* sont les intérêts générés par les DAC;
- DAC_n^{amortis} reflète le montant de DAC passé en charge au cours de l'exercice.

Mathématiquement, au cours d'un exercice n , le montant de DAC amorti est égal à :

$$KDAC \times \text{Primes de l'année}_n$$

où

$$KDAC = \frac{DAC_0}{VAP (\text{Primes})}$$

Le coefficient $KDAC$ est le pourcentage des primes payées par les assurés nécessaire pour faire face aux DAC générés par l'affaire.

Le principe d'amortissement des DAC selon l'émergence des marges est très similaire à ce qui précède. La principale différence provient de la formule de calcul du coefficient $KDAC$. En effet, dans l'approche IFRS, ce coefficient est calculé ainsi :

$$KDAC = \frac{DAC_0}{VAP (\text{Marges Brutes})}$$

où

$$\text{Marge Brute}_n = \text{Résultat Technique}_n - \text{Frais de Gestions}_n + \text{Résultat Financier}_n$$

Ici, le coefficient $KDAC$ est le pourcentage des marges brutes nécessaire pour faire face aux DAC générés par l'affaire.

Notons que dans une approche *Gross Premium* pour le calcul des provisions techniques, il n'y a pas lieu de comptabiliser de DAC car ils sont « inclus » dans cette approche à travers le montant de la *Gross Premium*. En effet, par rapport au calcul théorique de la *Net Premium* qui ne prend en compte que le risque, la *Gross Premium* inclut les différents chargements destinés à financer le contrat. De cette façon, le financement des coûts d'acquisition est impacté progressivement tout au long de la vie du contrat.

Section 2.4. Résultats de l'application numérique

La dernière section de cette partie présente les résultats de l'application numérique.

2.4.1. Présentation des indicateurs de mesure de l'émergence du résultat

Deux indicateurs de la mesure des profits ont été utilisés dans l'application numérique : le taux de rendement des capitaux propres (« *Return on Equity* » ou ROE en anglais) et la marge sur primes payées.

Le rendement des capitaux propres est égal au rapport entre le résultat net après impôt et le capital investi par les actionnaires (c'est-à-dire les capitaux propres) et permet de mesurer les profits dégagés par la société avec l'argent investi par les actionnaires. Ce ratio permet de comparer la rentabilité d'une société par rapport aux autres entreprises intervenant dans le même secteur d'activité. Dans le cadre de cette étude, l'étude de ce ratio est intéressant car les produits avec commission d'acquisition nécessitent d'importants investissements initiaux et font donc davantage appel aux actionnaires.

$$ROE = \frac{\text{Résultat Net}}{\text{Capitaux Propres}}$$

Dans l'application numérique développée dans ce mémoire, le ROE calculé est plus spécifiquement égal au taux de rendement interne (TRI) des flux de trésorerie successifs.

$$ROE = TRI (\text{Flux de Trésorerie})$$

La marge sur primes payées est égale au rapport entre la valeur actuelle probable des flux de trésorerie et la valeur actuelle probable des primes payées.

$$\text{Marge} = \frac{VAP(\text{Flux de Trésorerie})}{VAP(\text{Primes})}$$

Procédons désormais à l'analyse comparative des différents scénarii étudiés.

2.4.2. Comparaison des scénarii étudiés dans l'application

Le tableau ci-dessous présente le niveau de ROE et de marge sur primes payées obtenu pour les différents scénarii étudiés :

Modèle	ROE	Marge
(1) NP sans DAC ni PAD	12,54 %	8,97 %
(2) NP sans DAC avec PAD	12,06 %	9,01 %
(3) NP avec DAC amortis selon les primes et PAD	61,59 %	9,01 %
(4) NP avec DAC amortis selon les marges et PAD	56,00 %	9,01 %
(5) GP avec PAD	27,65 %	7,92 %

Tableau 4 : Comparaison des scénarii étudiés dans l'application

L'analyse de ce tableau fait ressortir plusieurs éléments :

- La prise en compte d'une provision *for adverse deviation* (PAD) impacte à la baisse le niveau de ROE mais n'a que très peu d'influence sur la marge (comparaison des scénarii (1) et (2)). Ce constat est cohérent car, lorsqu'une PAD est utilisée, le sur-provisionnement initial entraîne un décalage dans le temps de l'émergence du profit. De plus, le niveau de provision à mettre en place à l'origine étant plus important, le capital investi par les actionnaires doit être plus élevé.
- La possibilité de reconnaître des DAC augmente significativement le ROE (comparaison des scénarii (1) et (2) d'une part, et (3) et (4) d'autre part). En effet, la prise en compte de DAC permet de différer l'impact des coûts d'acquisition de la police la première année. Ainsi, les profits sont repartis de façon homogène au cours de la durée de vie du contrat.
- Les niveaux de ROE obtenus dans les scénarii (3) et (4) diffèrent peu. La méthode d'amortissement des DAC semble ne pas avoir d'influence sur les profits reconnus.
- Les niveaux de marge des scénarii (2) à (4) sont égaux. En effet, les primes reçues sont identiques dans ces trois scénarii, et, bien que les flux de trésorerie ne soient pas identiques, leur valeur actuelle probable est la même dans les trois scénarii.
Par ailleurs, la marge dans le scénario (1) est également très proche des autres scénarii car les primes reçues sont les mêmes. De plus, les flux de trésorerie ne sont que légèrement inférieurs, les intérêts sur le capital investi étant moins importants qu'avec prise en compte d'une PAD.

- Le scénario (5) fait apparaître une différence de résultats. L'approche par *Gross Premium* donne des résultats intermédiaires par rapport aux scénarii avec *Net Premium* avec ou sans DAC pour le ROE. En effet, si les primes payées sont les mêmes quelle que soit l'approche, les provisions, et par conséquent le niveau de capital investi par l'assuré, ainsi que les flux de trésorerie varient considérablement. La prise en compte d'une *Gross Premium* permet de prendre en compte le coût de la commission d'acquisition dans le calcul des provisions. Ainsi, même si ce mécanisme n'est pas aussi efficace que l'utilisation de DAC, il permet d'atténuer l'effet de la perte de première année et d'obtenir un ROE assez élevé. Toutefois, la marge est plus faible que dans les scénarii (1) à (4) car les intérêts financiers sur les provisions et sur le capital sont nettement moins importants avec cette approche.

Pour compléter cette analyse, deux graphiques sont présentés ci-dessous. Le premier illustre les résultats après impôt en nominal tandis que le second présente le résultat après impôt en fonction du montant de prime payée par les assurés.

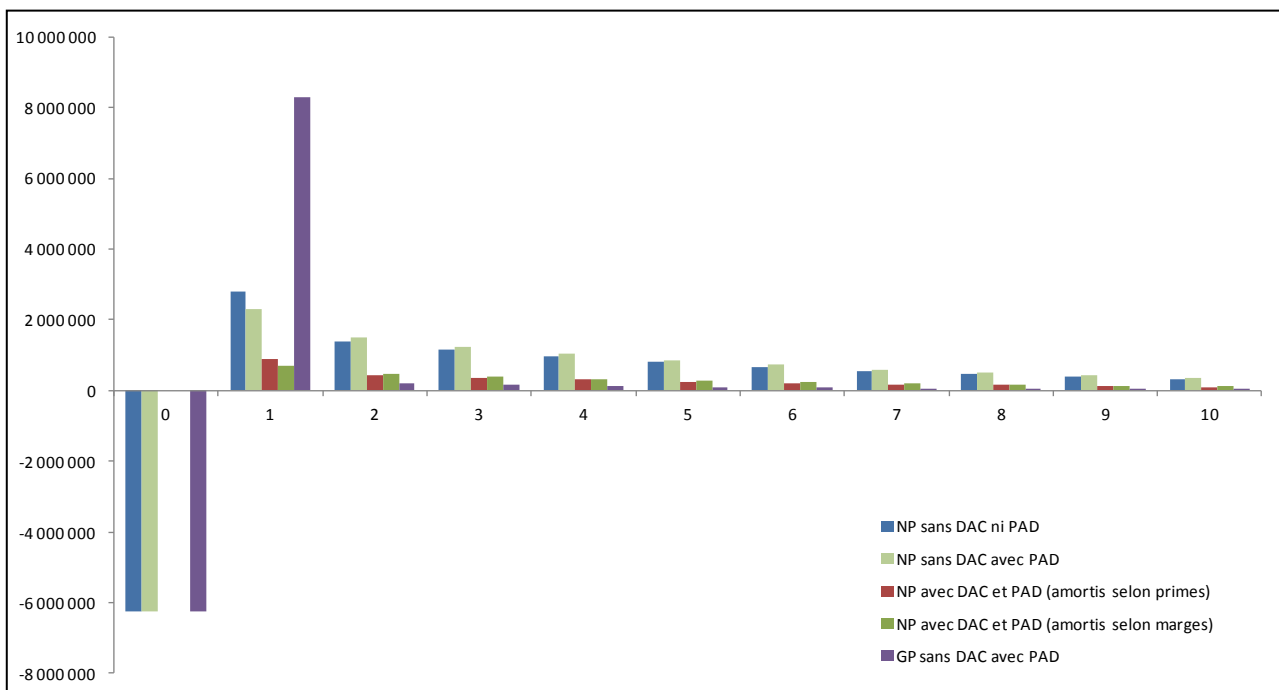


Figure 4: Résultats après impôt

Source : Propres calculs

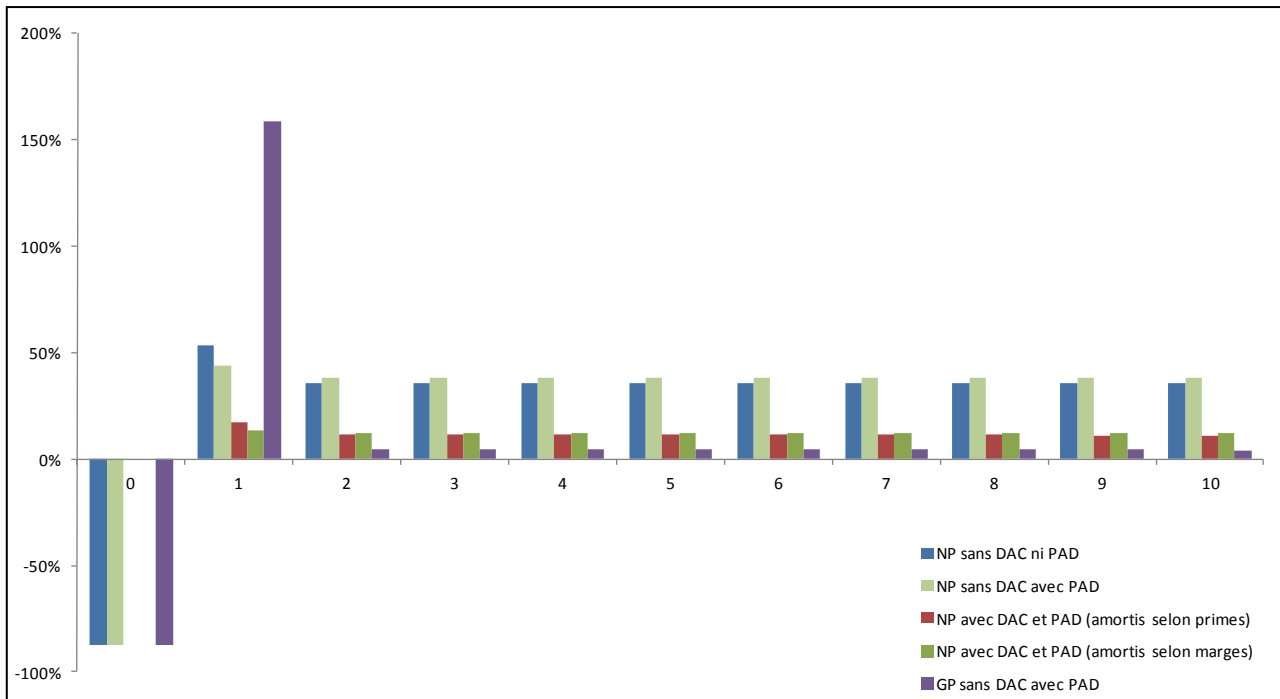


Figure 5: Résultats en pourcentage des primes

Source : Propres calculs

Ces deux graphiques permettent de mieux appréhender les résultats précédemment présentés et la spécificité de chacun des scénarii.

Ainsi, les scénarii avec DAC, laissent apparaître un résultat mieux réparti au cours de la durée de vie du contrat. Contrairement aux approches *Net Premium* sans DAC, la perte de première année est annulée, mais les profits aux cours des années suivantes sont moins importants en termes de montant et de pourcentage des primes. Le principe des DAC apparaît ainsi de façon plus « visuelle ».

Par ailleurs, l'utilisation de la *Gross Premium* pour le calcul de la provision fait apparaître une reconnaissance des profits différente des approches par la *Net Premium*. La perte de première année est équivalente à l'approche *Net Premium* sans DAC, mais est ensuite largement compensée la deuxième année par une reprise de provision très importante générant un profit considérable supérieur à 150 % des primes payées de l'année. Par la suite, le niveau de profit se stabilise à un niveau inférieur aux différentes approches par *Net Premium*.

En conclusion, le modèle développé pour ce mémoire permet de mieux appréhender l'avantage de l'utilisation de DAC pour obtenir un profit plus régulier dans le temps. Concrètement, l'intérêt quant à la rentabilité se mesure au niveau du ROE de l'affaire, nettement plus élevé lorsque des DAC peuvent être mis en place. Par contre, le gain de rentabilité en termes de marge est nul car la valeur actuelle probable des flux de trésorerie reste la même avec ou sans DAC. La méthode d'amortissement des DAC a également peu d'influence sur la rentabilité même s'il convient de signaler un ROE plus élevé lorsque les DAC sont amortis selon le rythme d'encaissement des primes. Enfin, l'utilisation de l'approche par la *Gross Premium* pour le calcul des provisions techniques entraîne une émergence du profit différente de l'approche par la *Net Premium*. Ainsi, le relâchement de la provision suite au paiement de la commission d'acquisition la première année entraîne un profit très important la deuxième année. Cette approche avec une volatilité importante du résultat les premières années du contrat donne un meilleur ROE au détriment d'un niveau de marge plus faible.

A l'issue de cette partie du mémoire, nous constatons que la méthode comptable imposée par la législation et le type de produit d'assurance distribué peut considérablement changer la mesure de la profitabilité d'une affaire et la reconnaissance des profits au cours du temps. Le réassureur, de par son envergure internationale, peut assumer le financement des commissions d'acquisition et, dans une certaine mesure, bénéficier d'arbitrage comptable en fonction de la législation des différents pays dans lesquels il opère. Au delà de l'arbitrage comptable, la possibilité d'amortir des coûts d'acquisition différés donne une meilleure vision économique du risque.

PARTIE III : ANALYSE DU RISQUE DE CHUTE

Nous avons évoqué dans la partie II du mémoire la nécessité pour l'assureur de rechercher, pour ses produits issus de la distribution directe, des clients « fidèles ». En effet, l'acquisition d'un nouvel assuré représente à la fois une opportunité mais également un coût important pour l'assureur. Ce dernier peut avoir plusieurs origines, notamment la rémunération du distributeur. Dès lors, l'assureur doit veiller à conserver ses assurés le plus longtemps possible afin d'amortir ces coûts d'acquisition et de garantir la rentabilité du portefeuille. La modélisation des polices résiliées (ou chutes) du portefeuille doit permettre, à travers la sélection et l'étude de variables explicatives, de choisir les meilleures cibles d'assurés afin de contrôler les coûts d'acquisition et de piloter la rentabilité du portefeuille existant.

En pratique, la question n'est pas évidente. En effet, la mise en place ou la révision d'une sélection des assurés potentiels, aussi méthodique et précise soit elle, est soumise d'une part au manque d'information, dû à l'impossibilité matérielle d'acquérir certaines données (coûts d'obtention trop élevés, informations non réclamées à l'assuré à l'origine, etc.), et, d'autre part, à des enjeux marketing difficilement modélisables mathématiquement parlant.

Dans cette partie, nous cherchons donc à appréhender le risque de chute pour des produits d'assurance issus de la distribution directe et, plus particulièrement, du télémarketing. Nous nous appuyons pour cela sur l'expérience de deux portefeuilles issus de deux pays différents que nous comparons à travers plusieurs études. Pour ce faire, nous présentons dans un premier chapitre les portefeuilles utilisés et la façon dont nous avons procédé pour obtenir une table d'étude pour chaque pays. Dans un second chapitre, une mise à jour et un développement approfondi d'une étude descriptive précédemment réalisée par l'équipe de Transamerica Reinsurance Europe sont menés par rapport à différentes variables d'analyse. Enfin, les deux derniers chapitres se basent sur l'article « *Surrender triggers in life insurance : classification and risk predictions* »³⁴ de X.Milhaud, S.Loisel et V.Maume-Deschamps dans lequel les arbres de classification avec l'algorithme CART et la régression logistique sont présentés. Les aspects théoriques de ces techniques sont évoqués avant de développer une application pratique de ces modèles afin de comparer les portefeuilles et challenger les méthodes utilisées par Transamerica Reinsurance Europe.

Par ailleurs, nous tenons à préciser que les chiffres et résultats présentés dans cette étude ont été modifiés afin de respecter la confidentialité des données utilisées tout en conservant l'esprit des conclusions.

Chapitre 1. Produits étudiés et obtention d'une table d'étude

Ce chapitre présente les portefeuilles étudiés ainsi que les données disponibles pour réaliser l'analyse du risque.

Section 1.1. Description des produits d'assurance

Les deux portefeuilles étudiés sont constitués de produits issus du télémarketing garantissant le versement d'indemnités journalières (IJ) pour une hospitalisation liée à une maladie ou un accident. Ce type de produit

³⁴ Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance : classification and risk predictions* », Hal

est un « classique » parmi les offres de contrats d'assurance de personnes issus du télémarketing. En général, ces produits proposent plusieurs options de versement pouvant aller de 15 à 120 euros par jour d'hospitalisation après une période de franchise de 24 heures. De plus, l'indemnisation est doublée en cas d'accident pour ces produits qui prennent aussi en charge les sinistres à l'étranger dans une certaine limite de jours (de 60 à 90 jours). Toutefois, la durée de versement des indemnités journalières est limitée bien que majorée en cas d'accident. Enfin, afin d'inciter les assurés potentiels à souscrire au produit, des remises sont accordées en cas d'adhésion de plusieurs membres d'une même famille et le premier mois de couverture est gratuit.

Section 1.2. Extraction et traitement des données entrant dans le périmètre de l'étude

Dans cette section, nous expliquons comment nous avons procédé à partir des données brutes afin d'obtenir les deux tables (XW et YZ) qui servent de support à l'étude réalisée par la suite.

Issues de portefeuilles récents, les données dont nous disposons sont initialement étudiées sur un historique de 25 mois pour le produit XW et de 15 mois pour le produit YZ. Nous nous intéressons donc aux résiliations des premiers mois suivant l'acquisition des contrats. Ceci est déjà significatif pour ce type de produit dont les taux de résiliation après la première année sont souvent faibles.

La première étape consiste à extraire l'ensemble des données correspondant à ce périmètre. Toutefois, afin de bien appréhender cette étape, il convient de commencer par décrire les bases de données elles-mêmes. Celles-ci sont chacune constituées de deux types de table :

- Les tables « contrats » regroupent les informations sur les caractéristiques des contrats et des assurés. Elles contiennent notamment les dates de début et de fin d'engagement (si la police n'est plus en vigueur) qui sont indispensables dans le cadre de cette étude. Par ailleurs, outre les variables explicatives qui serviront à l'analyse, il est essentiel d'avoir accès à un identifiant de contrat. Il est alors important de garder à l'esprit qu'un identifiant peut correspondre à plusieurs assurés (mais bien à un seul contrat) car l'étude porte sur les contrats et non sur les assurés. En effet, les coûts d'acquisition sont plutôt déterminés au niveau du contrat et ne peuvent pas être facilement définis au niveau d'un assuré.
- Les tables « flux de trésorerie » contiennent les informations sur les primes, les commissions et les sinistres.

Pour chacune des bases, il faut d'abord regrouper les assurés par police. Cette manipulation est effectuée sur les tables de type « contrats » et sur les tables de type « flux de trésorerie ». Pour chacun des portefeuilles, ces deux tables sont ensuite fusionnées pour obtenir la table d'étude. Les tables obtenues font alors l'objet d'un travail d'analyse afin d'obtenir la certitude que les données utilisées sont pertinentes et fiables. Ainsi, il faut comprendre pourquoi un certain nombre d'enregistrements sont présents uniquement sur la table « contrat » ou sur la table « flux de trésorerie » mais n'apparaissent pas dans la table d'étude (pour qu'un identifiant de contrat apparaisse sur cette table il doit être présent dans les deux tables d'origine). Il convient également de déterminer pour quelles raisons, sur un certain nombre de contrats, la somme des flux de primes payées est négative ou nulle.

Section 1.3. Etude des portefeuilles à travers les tables « contrats »

Les caractéristiques des deux tables « contrats » étudiées sont synthétisées dans le tableau ci-dessous :

	XW	YZ
Nombre d'enregistrements	80 000	50 000
N°contrat	✗	✓
N°adhésion (ID contrat)	✓	✓
N°adhérent	✗	✓
Nom et prénom de l'assuré	✓	✓
Nom et prénom de l'adhérent	✓	✗
Date de naissance de l'assuré	✓	✓
Date d'effet du contrat	✓	✓
Date de fin d'effet du contrat	✓	✓
Montant de l'IJ	✓	✓
Prime émise TTC	✓	✓
Statut de la police (en vigueur, résilié ou rétracté)	✓	✗
Motif du statut	✓	✗
Type d'assuré (assuré principal, conjoint ou enfant)	✓	✗
Modifications du contrat	✓	✗
Campagne	✗	✓

Tableau 5 : Principales caractéristiques des tables « contrats »

Section 1.4. Caractéristiques de la table d'étude après retraitement

Pour les deux portefeuilles, le périmètre d'étude des données doit être modifié. En effet, les tables « flux de trésorerie » ne disposent pas du même nombre de mois que les tables « contrats » dans leur historique. La décision a donc été prise de restreindre le périmètre des données utilisées aux périodes pour lesquelles nous disposons d'enregistrements dans la table de « flux de trésorerie ». Par ailleurs, n'ont pas été retenues les données des premiers mois pour le portefeuille YZ car le nombre d'enregistrements a été jugé insuffisant pour effectuer des études statistiques.

Après avoir traité et fusionné les différentes tables présentes dans chacune des bases de données, les tables contiennent respectivement 30 000 enregistrements pour XW et 15 000 pour YZ. Toutefois, nous devons encore faire face à un certain nombre d'erreurs dans la table obtenue du fait de la quantité importante de données. Il a déjà été évoqué en particulier l'existence d'un certain nombre d'enregistrements pour lesquels le nombre de primes payées est négatif, nul ou non entier. Il convient d'essayer de déterminer l'importance de ces erreurs et de tenter, dans la mesure du possible, de les expliquer.

Sur les 30 000 lignes de la table XW, nous en comptons seulement 10 avec un nombre de primes payées négatif. Cette quantité est négligeable et nous excluons donc ces polices de l'étude. Par ailleurs, nous avons

2 000 polices sans aucune prime versée. Après analyse, il s'agit de polices résiliées, de rétractations ou de polices en vigueur. Dans ce dernier cas, deux explications sont possibles : soit nous nous trouvons dans le premier mois de la police qui est gratuit, soit ces polices auraient dû être résiliées car en regardant la table des « flux de trésorerie » nous nous apercevons alors qu'il y a eu des impayés. Toutefois, ce nombre reste relativement faible et nous parons à ce problème en étudiant les résiliations après le premier mois de contrat (c'est-à-dire après le premier paiement).

Enfin, le nombre de polices où le nombre de primes payées n'est pas entier reste limité (1 000). Deux raisons expliquent ce phénomène : des modifications de la cotisation (il s'agit de la majorité des cas) et des primes théoriques qui ne correspondent pas aux primes réellement versées. Pour ces polices, nous avons choisi d'arrondir le nombre de primes payées à l'entier inférieur.

Quant à la table YZ, sur 15 000 polices, le nombre de primes payées est négatif pour 50 cas et nul pour 7 000 autres cas. Les traitements opérés sont les mêmes que pour le portefeuille XW, mais nous remarquons que le nombre de polices sans paiement est bien plus important que pour le pays XW (47 % des polices en YZ contre moins de 7 % en XW). Un tel écart est surprenant mais peut être expliqué par plusieurs facteurs : un système de gestion moins performant en YZ (de nombreuses primes ne sont pas enregistrées), des impayés plus importants, etc.

Par ailleurs, nous comptons 1 000 polices avec un nombre de primes payées non entiers. Nous constatons, comme en XW, que cela est dû à des changements dans la prime payée ou à des écarts entre la prime annoncée et celle réellement payée. Nous avons également décidé d'arrondir à l'entier inférieur.

In fine, la table d'étude XW compte 29 990 polices sur 23 mois et la table YZ 14 950 polices sur 11 mois. Les figures ci-dessous donnent l'évolution mensuelle du nombre de polices pour les portefeuilles XW et YZ.

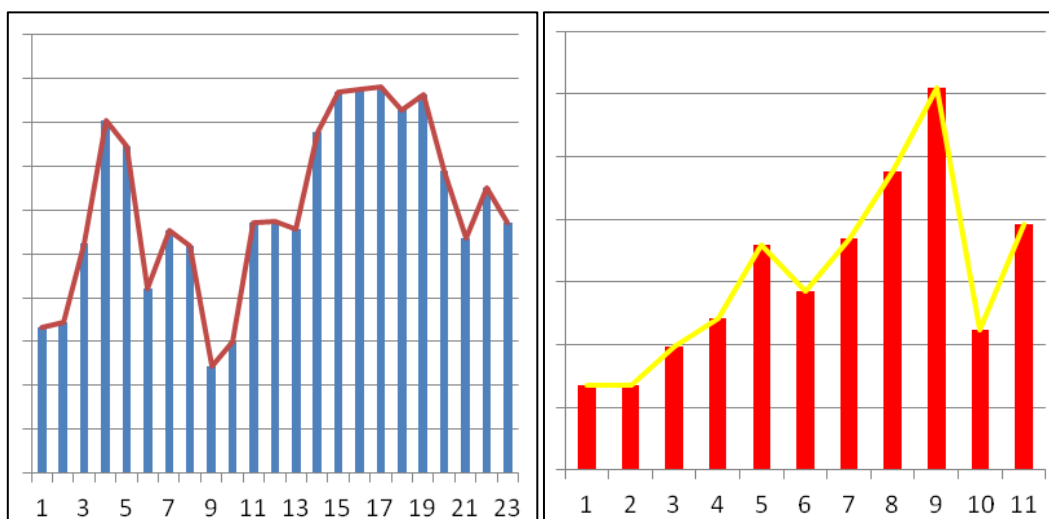


Figure 6: Evolution mensuelle du nombre de polices des portefeuilles XW (à gauche) et YZ (à droite)

Source : Propres calculs à partir de données réelles

Pour le portefeuille XW, nous constatons une croissance rapide au lancement du programme puis une baisse entre le cinquième et le huitième mois, suivie d'une nouvelle phase de croissance et enfin une phase de décroissance à compter du dix-neuvième mois. Concernant le portefeuille YZ, le nombre de contrats souscrits au début de la période d'observation sont négligeables par rapport aux niveaux atteints ultérieurement. Nous avons donc une phase de lancement caractérisée par une croissance rapide du nombre de contrats souscrits

jusqu'au cinquième mois. Il y a ensuite un décrochage suivi d'une reprise de la croissance et d'un nouveau décrochage à la fin de la période d'observation.

La synthèse des principales caractéristiques des deux portefeuilles est donnée dans le tableau suivant :

	XW	YZ
Prime brute annuelle moyenne	330 €	220 €
Age moyen	50 ans	40 ans

Tableau 6 : Synthèse sur les portefeuilles XW et YZ

Le portefeuille XW, plus ancien que le portefeuille YZ, contient davantage de polices même si le nombre de polices souscrites par mois est légèrement supérieur en YZ. Par ailleurs, les compositions des portefeuilles sont assez différentes. Les assurés du portefeuille YZ sont notamment plus jeunes en moyenne que ceux du portefeuille XW. Enfin, le montant de la prime brute annuelle moyenne est sensiblement plus grand en XW qu'en YZ (330 € contre 220 €). Après ce premier travail de retraitement et d'analyse de cohérence des données, les tables sont estimées suffisamment fiables pour s'attacher à réaliser l'étude des résiliations.

Attention néanmoins à ne pas oublier que les données disponibles ne sont pas les mêmes selon qu'il s'agit du produit XW ou YZ. Par exemple, alors que nous avons uniquement à disposition pour étudier les résiliations la variable « date de résiliation de la police » en YZ, en XW nous pouvons également nous servir des variables « statut de la garantie », qui renseigne si le produit est en vigueur, résilié ou s'il y a eu rétractation de l'assuré, et « motif du statut ». Il est ainsi plus aisé de faire une analyse des résiliations sur le produit XW.

Chapitre 2. Etude descriptive des résiliations sur les deux tables d'étude

Dans ce chapitre, nous réalisons, à partir d'une étude descriptive des résiliations réalisée par Transamerica Reinsurance Europe, une analyse mise à jour et complétée des résiliations sur les deux tables d'étude à travers une approche uni-variée. En effet, une fois les données clairement déterminées, il semble intéressant d'effectuer ces travaux d'analyses préliminaires avant de commencer la modélisation.

Section 2.1. Intuitions et analyse préalable

Les principales variables d'analyse disponibles sur les deux portefeuilles « hospitalisation » sont l'âge de l'assuré, le sexe de l'assuré, sa situation familiale et le niveau de garantie souscrit.

Tout d'abord, l'âge semble être une variable déterminante. En effet, les comportements de l'assuré sont très liés à l'âge, encore plus lors d'une souscription par téléphone. L'intuition est qu'il existe un effet générationnel dans la façon dont les assurés résilient leur contrat. Or, il a été constaté lors de l'étude statistique des portefeuilles XW et YZ que le second était beaucoup plus « jeune » que le premier (pour rappel, l'âge moyen de l'assuré est de 40 ans en YZ contre 50 ans en XW). Il est donc intéressant d'étudier cette variable. A priori, le sexe de l'assuré devrait avoir moins d'importance dans l'explication des résiliations. Nous allons vérifier cette intuition par la suite.

La situation familiale de l'assuré peut également expliquer les résiliations. En effet, un assuré qui souscrit pour plusieurs membres de sa famille doit avoir un intérêt plus fort pour le produit. Par ailleurs, il bénéficie d'une remise sur la prime dans le cas où plusieurs personnes sont couvertes par la police. En contrepartie, le

montant total de prime à payer chaque mois est plus important, ce qui peut inciter le détenteur de la police à résilier ou modifier son contrat. Il est donc intéressant d'étudier l'impact réel de la situation familiale sur le comportement de l'assuré.

Enfin, le nombre de résiliations semble assez naturellement dépendre du niveau de garantie souscrit : plus la prime est élevée, plus la charge qu'elle représente pour l'assuré est importante et plus il aura tendance à résilier. Ce constat peut toutefois être nuancé car il est possible que les assurés prêts à payer une prime élevée soient ceux qui disposent de plus de moyens financiers et qu'ils résilient moins. L'analyse à suivre tente de déterminer quelle intuition est la plus juste.

Section 2.2. Commentaires et méthodologie choisie

Les taux de résiliation obtenus sont pondérés par les primes. En effet, il semble cohérent de ne pas analyser les résiliations en termes de nombre de polices car cette information n'est pas suffisante pour connaître leur impact financier. Dans l'absolu, les deux méthodes devraient être équivalentes si une segmentation initiale de la base de données est réalisée afin d'obtenir des groupes homogènes. A titre indicatif, les arbres de classification obtenus par l'algorithme CART étudié dans le prochain chapitre permettent de réaliser de tels groupes homogènes et peuvent donc être utilisés préalablement à l'étude des résiliations si l'on souhaite réaliser une étude en termes de nombre de polices. Ce n'est toutefois pas l'approche retenue pour cette section dans laquelle nous effectuons d'abord une étude sur l'ensemble du portefeuille où nous nous intéressons au taux de résiliation et au niveau de persistance des polices (c'est-à-dire à la proportion de polices encore en vigueur par rapport au nombre initial) en utilisant le nombre de primes payées par l'assuré. Nous analysons ensuite l'impact de plusieurs variables (âge, sexe, situation familiale de l'assuré principal, montant de la garantie souscrite, campagne de télémarketing (uniquement dans le cas du portefeuille YZ pour cette dernière variable)) sur le niveau de persistance des polices par rapport au niveau de l'ensemble du portefeuille. Nous serons alors en mesure de confirmer ou d'infirmer nos intuitions sur le profil de risque en fonction des caractéristiques de l'assuré et de la police.

L'approche retenue pour obtenir des taux de persistance est basée sur le principe de la méthode de Kaplan Meier. Cette technique est couramment utilisée par les actuaires afin de déterminer des taux de sortie (taux de mortalité, taux d'entrée en incapacité/invalidité, taux de rachat, etc.). Dans le cadre de cette étude, nous adoptons cette méthode de calcul des taux de résiliation et des niveaux de persistance mensuels en raisonnant sur les primes et non sur les polices. Introduisons donc les notations et la méthodologie utilisées³⁵.

Soit,

- S : loi discrète de persistance de la forme $(a_i, s_i)_{i \in \{1, \dots, m\}}$, les a_i étant les dates (ici, il s'agit plus exactement de nombres de mois), et les s_i les valeurs prises par S en a_i ,
- T : durée de vie du contrat en nombre de mois,
- q_i : probabilité de résilier entre les dates a_i et a_{i+1} ,
- n_i : total des primes en a_i ,
- d_i : total des primes des personnes chutant entre a_i et a_{i+1} .

L'estimateur de Kaplan Meier s'appuie sur la remarque suivante : la probabilité de ne pas résilier au-delà de la date $t > s$ peut s'écrire :

³⁵ Source : **Planchet F., Thérond P. (2006)**, « *Modèles de durée, applications actuarielles* », Economica

$$S(t) = \mathbb{P}(T > t | T > s) \times \mathbb{P}(T > s) = \mathbb{P}(T > t | T > s) \times S(s)$$

En renouvelant l'opération, nous faisons apparaître des produits de termes en $(T > t | T > s)$.

Si nous choisissons comme instants de conditionnement les dates a_i , nous nous ramenons à estimer des probabilités de la forme :

$$p_i = \mathbb{P}(T > a_i | T > a_{i-1})$$

Un estimateur naturel de $q_i = 1 - p_i, \forall i \in [1; m]$ est $q_i = \frac{d_i}{n_i}$.

Dès lors, nous obtenons une expression de s_i , estimateur de S en $a_i, \forall i \in [1; m]$:

$$s_i = \prod_{j/j \leq i} \left(1 - \frac{d_j}{n_j}\right)$$

Nous obtenons ainsi de façon empirique un estimateur non paramétrique de la fonction de persistance des polices dans le portefeuille. Dans la section suivante, nous analysons de façon détaillée les résultats obtenus à l'aide de cette méthode sur les deux portefeuilles.

Section 2.3. Etude des résiliations sur les portefeuilles XW et YZ

Dans cette section, nous étudions le taux de résiliation et la persistance des polices en portefeuille en termes de montant de prime reçue selon différents critères : l'âge de l'assuré principal, le sexe de l'assuré principal, sa situation familiale, le montant d'indemnité journalière souscrit et, pour le portefeuille YZ, la campagne de télémarketing correspondant à la vente.

2.3.1. Etude sur l'ensemble du portefeuille

Nous commençons par étudier les résiliations pondérées par les primes sur les deux portefeuilles afin d'avoir une première idée du comportement des assurés dans les deux pays et une référence par rapport à laquelle nous étudierons ensuite les résiliations selon d'autres critères. Les résiliations sont étudiées en considérant seulement les personnes ayant payées au moins une prime pour ne pas prendre en compte les polices sans paiement. Par conséquent, nous n'avons pas renseigné le taux de résiliation au mois « 0 ».

Enfin, nous rappelons à nouveau que les chiffres et les résultats fournis dans ce chapitre ont été modifiés afin de respecter la confidentialité des données mais de conserver l'esprit des conclusions.

2.3.1.1. Portefeuille XW

Nous disposons pour le portefeuille XW de données fiables sur 23 mois. Toutefois, nous ne donnons des résultats que pour une analyse sur 18 mois car nous estimons qu'au delà le recul n'est pas suffisant pour avoir des chiffres cohérents. Les résultats obtenus sont présentés ci-dessous.

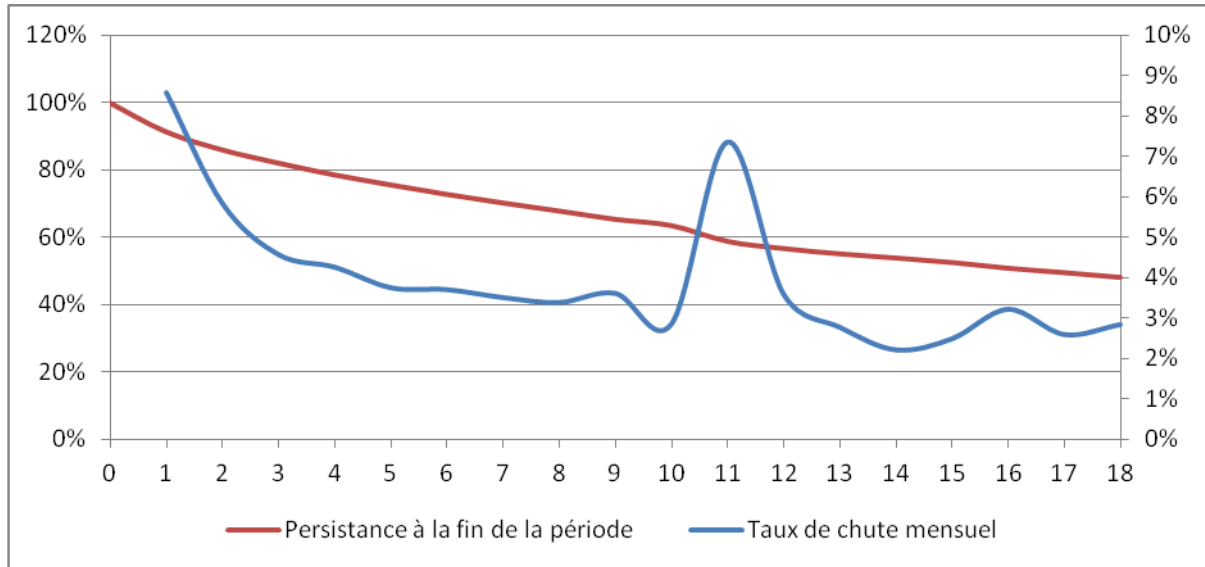


Figure 7 : Résiliations sur le portefeuille XW
Source : Propres calculs à partir de données réelles

Après un taux de résiliation de 9 % le premier mois, nous constatons une décroissance du taux mensuel qui se stabilise entre 3 % et 4 % à partir du quatrième mois avec toutefois un pic à 7 % à la date anniversaire du contrat. Après 18 mois, 48 % des polices sont toujours en portefeuille.

2.3.1.2. Portefeuille YZ

Nous disposons pour le portefeuille YZ de données fiables sur 11 mois. Toutefois, nous ne donnons des résultats que pour une analyse sur 9 mois car nous estimons qu'au-delà le recul n'est pas suffisant pour avoir des chiffres cohérents. Les résultats obtenus sont présentés ci-dessous.

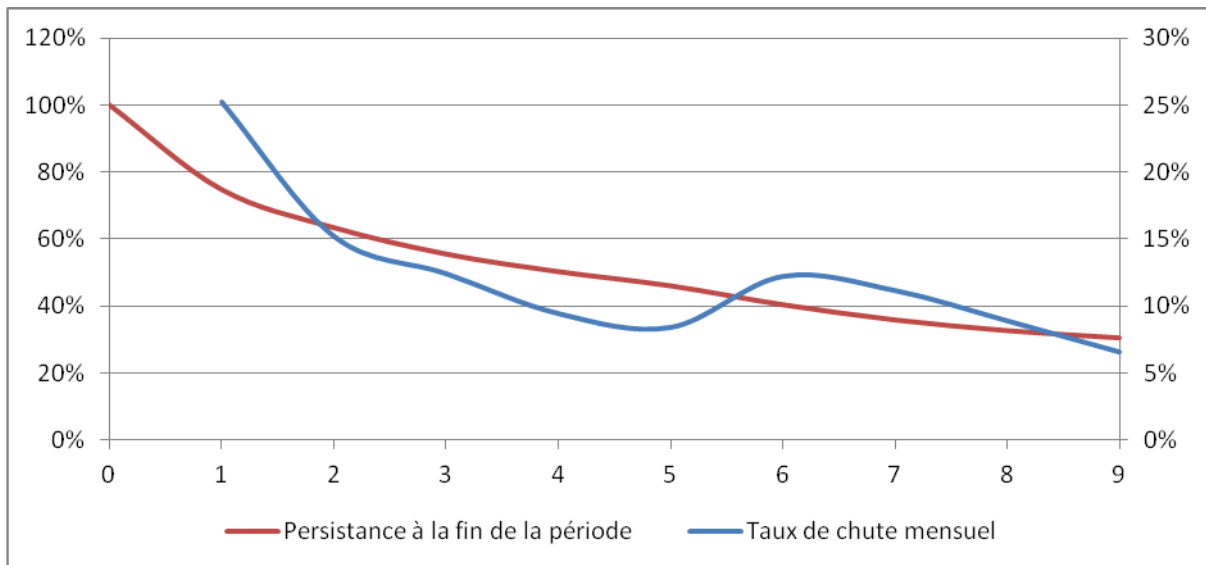


Figure 8 : Résiliations sur le portefeuille YZ
Source : Propres calculs à partir de données réelles

Nous observons à nouveau une diminution du taux de résiliation au cours des trois premiers mois de 25 % à 12 %. Cependant, passée cette première période, le taux de résiliation ne semble pas se stabiliser et reste très volatil. Après 9 mois, seules 31 % des polices sont toujours en portefeuille.

2.3.1.3. Commentaires et comparaisons

Les polices YZ sont résiliées de manière nettement plus sévère que les polices XW. Ainsi, alors qu'après 6 mois 73 % du portefeuille XW initial est encore présent, ce chiffre n'est que de 41 % en YZ. Il est difficile de trouver des explications « toutes faites » pour comprendre ce phénomène. Toutefois, et malgré des similarités entre les deux produits, la façon dont les assurés résilient et les différences de mode de gestion des impayés en XW et en YZ ont pu avoir une influence sur les résiliations.

La différence est significative et il convient également de prêter attention aux comportements des assurés qui sont différents en XW et en YZ. Ainsi, en XW, un assuré a tendance à prévenir son assureur de sa volonté de résilier son contrat ou de se rétracter ce qui se traduit par un taux de résiliation plus élevé le premier mois et par un pic à la date anniversaire du contrat. En YZ, les assurés ne paient tout simplement plus leur prime quand ils ne veulent plus du contrat, d'où la forte volatilité du taux de résiliation qui peut varier significativement d'un mois sur l'autre. Ces comportements sont difficilement modélisables mathématiquement mais elles ont une importance lorsque nous nous intéressons à la dynamique des résiliations sur les portefeuilles. Nous assistons, de ce fait, à des résiliations plus importantes et plus rapides sur le portefeuille YZ après le premier mois.

2.3.2. Etude en fonction de la tranche d'âge

Quatre tranches d'âge sont distinguées: 18-34 ans, 35-44 ans, 45-54 ans et 55-65 ans.

2.3.2.1. Portefeuille XW

Le comportement des différentes tranches d'âge par rapport au portefeuille global est étudié en comparant les niveaux de persistance des polices aux valeurs trouvées précédemment. Les résultats sont donnés dans la figure ci-dessous.

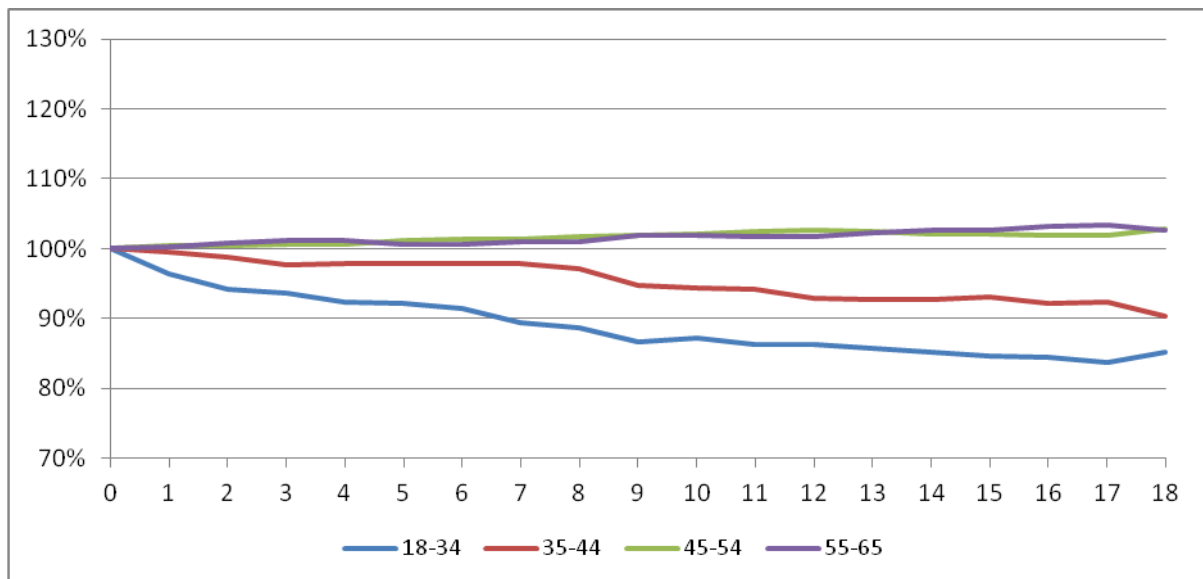


Figure 9 : Comparaison de la persistance après le premier paiement par tranche d'âge par rapport au portefeuille global en XW

Source : Propres calculs à partir de données réelles

Nous notons de visu que les tranches d'âges « jeunes » sont celles qui résilient le plus : plus les assurés sont « jeunes », plus la persistance des polices est faible par rapport au portefeuille global. Cette tendance est

constante dans le temps et nous remarquons que plus le nombre de mois écoulés est grand, plus la différence en termes de persistance des polices est importante entre les groupes. Ainsi, après 18 mois, la persistance de la tranche 18-34 ans représente 85 % de la persistance du portefeuille global, contre 90 % pour la tranche 35-44 ans et 103 % pour les tranches 45-54 ans et 55-65 ans.

2.3.2.2. Portefeuille YZ

La même étude est réalisée pour le portefeuille YZ. Les résultats sont présentés dans la figure ci-dessous.

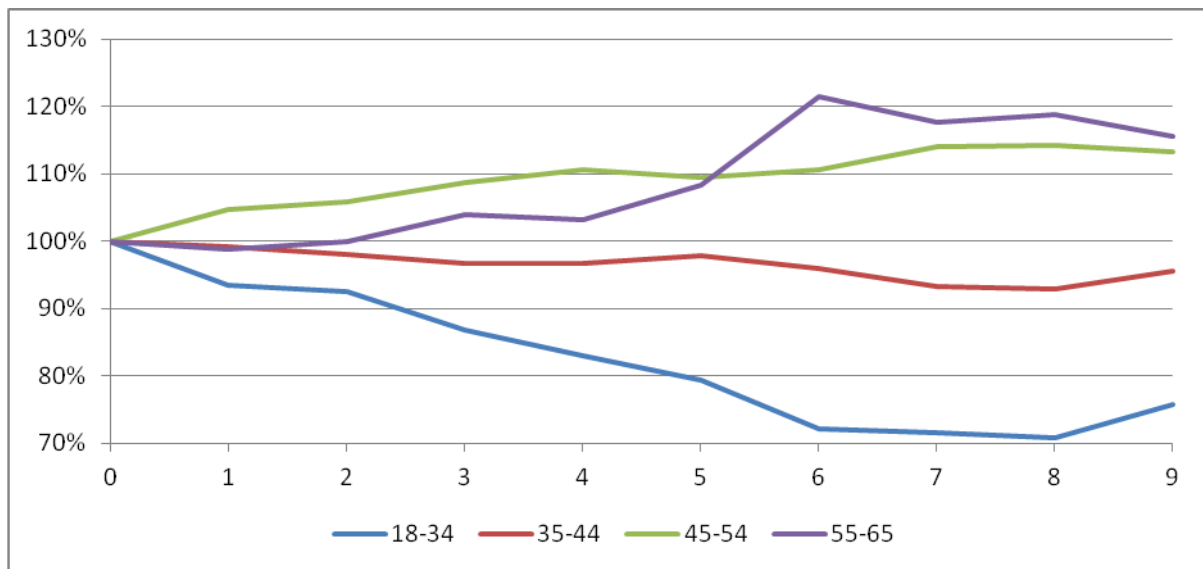


Figure 10 : Comparaison de la persistance après le premier paiement par tranche d'âge par rapport au portefeuille global en YZ

Source : Propres calculs à partir de données réelles

En YZ, les tranches d'âges « jeunes » résilient aussi davantage. Ainsi, après 9 mois, la persistance de la tranche 18-34 ans représente 76 % de la persistance du portefeuille global, contre 96 % pour la tranche 35-44 ans, 113 % pour la tranche 45-54 ans et 115 % pour la tranche 55-65 ans.

2.3.2.3. Commentaires et comparaisons

Même si les tendances sont identiques pour les deux portefeuilles (les « jeunes » résilient davantage), les écarts sont plus marqués en YZ puisque la déviation par rapport au portefeuille global y est plus importante.

Par ailleurs, les répartitions des polices par tranche d'âge ne sont pas les mêmes dans les deux portefeuilles. Ainsi, nous confirmons en regardant la répartition des polices par tranche d'âge que le portefeuille YZ est plus « jeune » que le portefeuille XW.

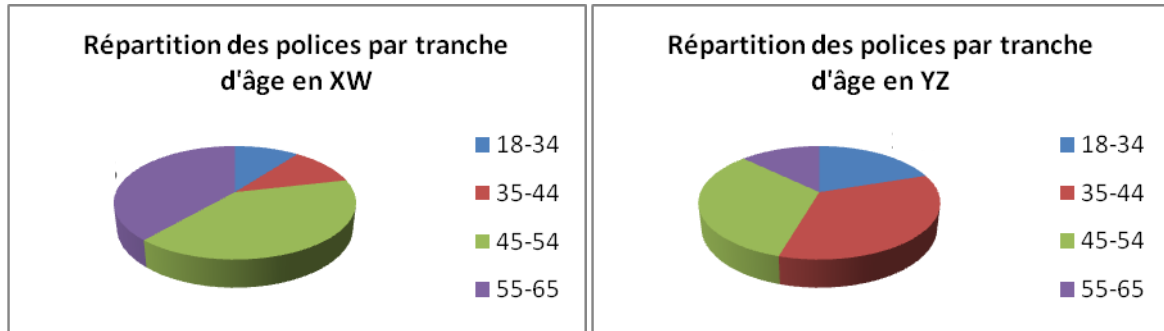


Figure 11 : Répartition des polices par tranche d'âge
 Source : Propres calculs à partir de données réelles

Le poids de chacune des tranches n'est donc pas le même dans les deux portefeuilles. Ceci explique notamment pourquoi, dans le portefeuille XW, les tranches d'âge 45-54 ans et 55-65 ans restent très proches de la tendance générale car elles représentent la grande majorité de la population totale. Toutefois, en YZ, la répartition des assurés ne peut expliquer seule, la dispersion des différentes tranches d'âge par rapport à la valeur de référence. La tranche d'âge à laquelle appartient l'assuré semble donc être une variable discriminante.

2.3.3. Etude des résiliations selon le sexe de l'assuré principal

Nous avons voulu ici déterminer l'influence du sexe de l'assuré principal sur la persistance des polices en portefeuille. Les hommes et les femmes ont donc été étudiés indépendamment.

2.3.3.1. Portefeuille XW

Nous comparons les niveaux de persistance des polices en fonction du sexe de l'assuré principal aux valeurs trouvées pour le portefeuille global. Les résultats sont donnés dans la figure ci-dessous.

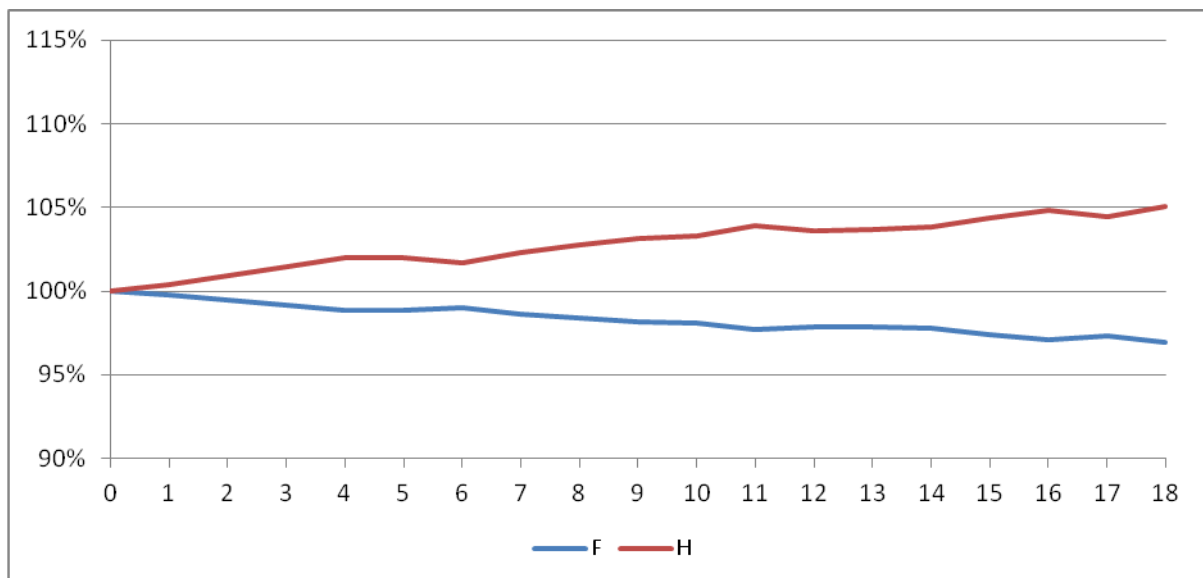


Figure 12 : Comparaison de la persistance après le premier paiement selon le sexe par rapport au portefeuille global en XW
 Source : Propres calculs à partir de données réelles

Il apparaît qu'en XW les femmes ont tendance à résilier plus que les hommes. Cette tendance est constante dans le temps : plus le nombre de mois écoulés est grand, plus la différence en termes de persistance

des polices est importante. Ainsi, après 18 mois, la persistance des hommes est de 105 % du portefeuille global alors que celle des femmes n'est que de 97 %.

2.3.3.2. Portefeuille YZ

La même étude est réalisée pour le portefeuille YZ. Les résultats sont présentés dans la figure ci-dessous.

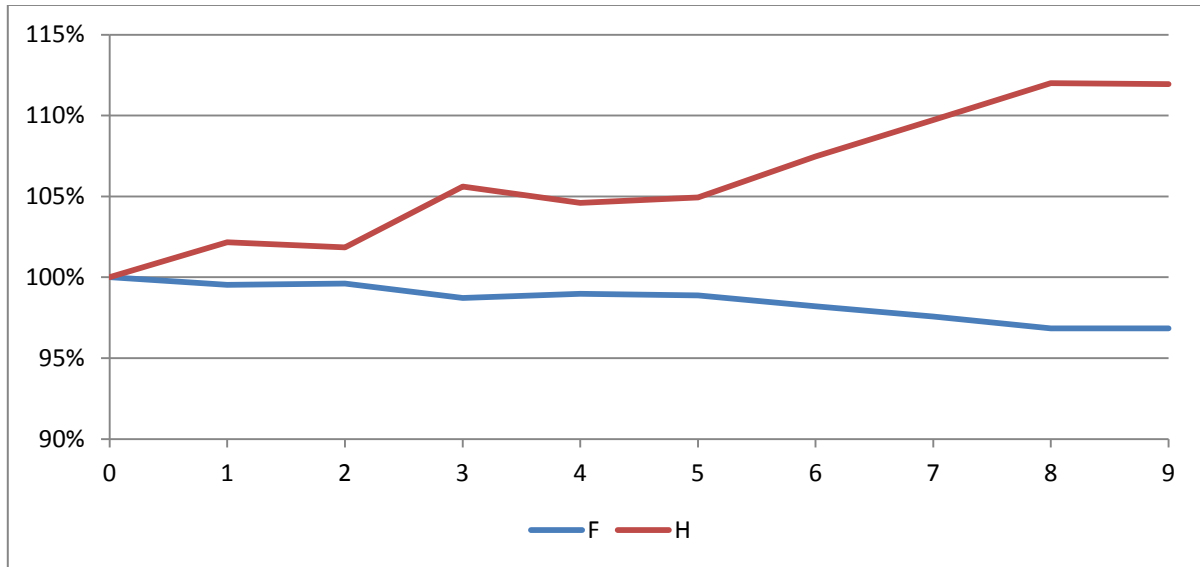


Figure 13 : Comparaison de la persistance après le premier paiement selon le sexe par rapport au portefeuille global en YZ
Source : Propres calculs à partir de données réelles

L'influence du sexe de l'assuré principal est la même qu'en XW : les femmes résilient plus que les hommes. Toutefois, la tendance semble moins homogène, particulièrement chez les hommes où des paliers alternent avec des périodes où la persistance par rapport au portefeuille global augmente rapidement. Par ailleurs, l'écart entre la persistance des hommes et des femmes se creuse plus rapidement en YZ : après 9 mois celle-ci est de 112 % chez les hommes contre 97 % pour les femmes.

2.3.3.3. Commentaires et comparaisons

Comme pour l'âge de l'assuré principal, l'influence du sexe fait apparaître des tendances similaires en XW et en YZ mais la déviation par rapport au portefeuille global est plus importante en YZ.

Intéressons nous à la répartition des polices selon le sexe de l'assuré principal en XW et en YZ.

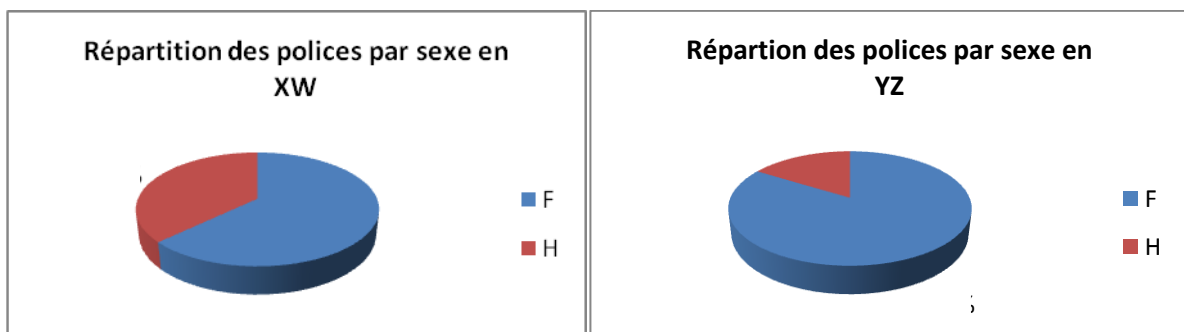


Figure 14 : Répartition des polices par sexe
Source : Propres calculs à partir de données réelles

Nous constatons que la proportion de femmes en portefeuille est nettement supérieure en YZ ce qui explique pourquoi la persistance des hommes s'éloigne beaucoup de la tendance générale en YZ. Nous comprenons aussi pourquoi les résiliations sur le portefeuille global sont plus importantes en YZ : les femmes qui présentent un risque de chute supérieur sont proportionnellement plus représentées en portefeuille qu'en XW.

2.3.4. Etude des résiliations selon la situation familiale

Tentons maintenant de déterminer l'influence de la situation familiale de l'assuré principal sur la persistance des polices en portefeuille. Nous distinguons ici les familles et les célibataires.

2.3.4.1. Portefeuille XW

Commençons par comparer les niveaux de persistance des polices en fonction de la situation familiale de l'assuré principal aux valeurs trouvées pour le portefeuille global. Les résultats sont donnés dans la figure ci-dessous.

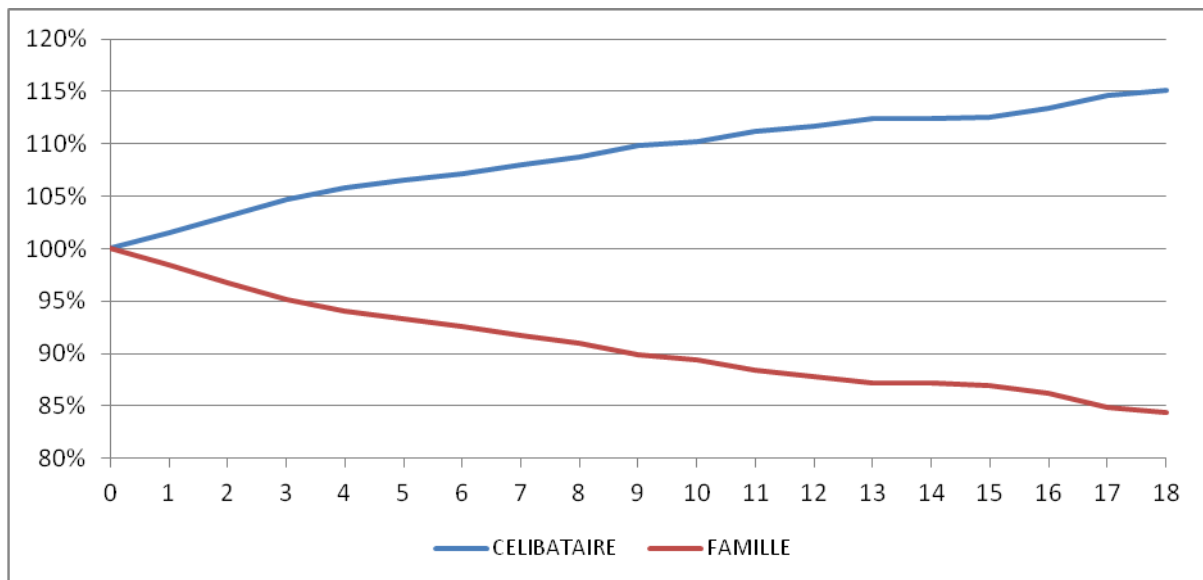


Figure 15 : Comparaison de la persistance après le premier paiement selon la situation familiale par rapport au portefeuille global en XW

Source : Propres calculs à partir de données réelles

En XW, les familles résilient davantage que les assurés célibataires. Cette tendance est constante dans le temps : plus le nombre de mois écoulés est grand, plus la différence en termes de persistance des polices est importante. Ainsi, après 18 mois, la persistance des célibataires est de 115 % du portefeuille global alors que celle des familles n'est que de 84 %.

2.3.4.2. Portefeuille YZ

A nouveau, réalisons une étude similaire pour le portefeuille YZ. Les résultats sont présentés dans la figure ci-dessous.

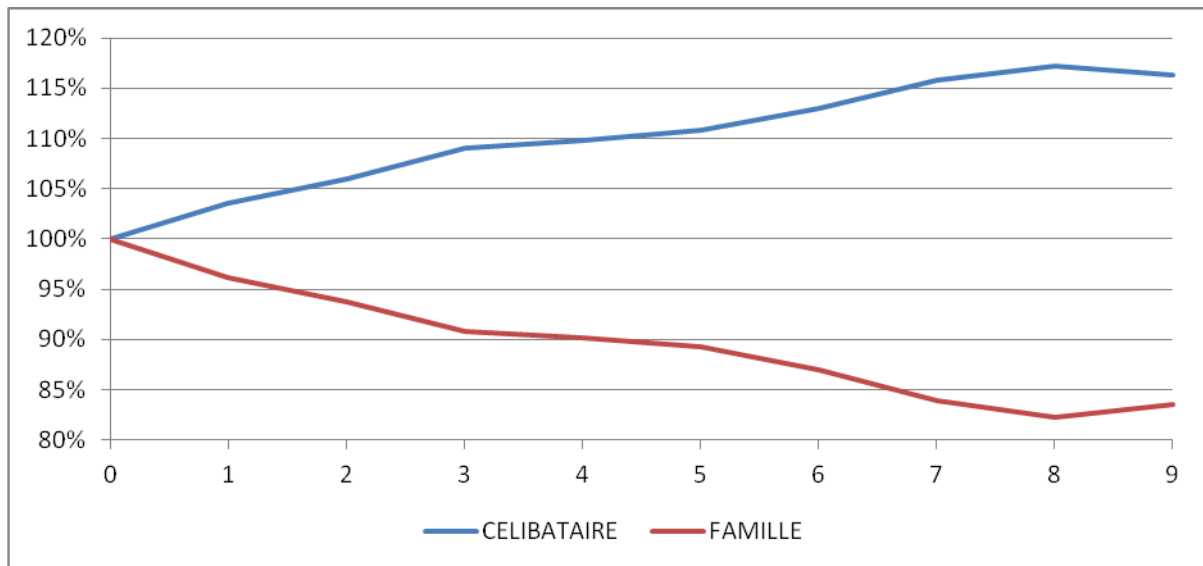


Figure 16 : Comparaison de la persistance après le premier paiement selon la situation familiale par rapport au portefeuille global en YZ

Source : Propres calculs à partir de données réelles

La tendance qui se dégage est la même que pour le portefeuille XW : les familles résilient plus que les célibataires. Néanmoins, l'évolution de cette tendance entre les deux groupes, caractérisée par des augmentations rapides entrecoupées de paliers, est moins régulière qu'en XW. A nouveau, l'écart entre la persistance des célibataires et des familles se creuse plus rapidement en YZ : après 9 mois celle-ci est de 116 % chez les célibataires contre 84 % pour les familles.

2.3.4.3. Commentaires et comparaisons

Comme pour l'âge et le sexe de l'assuré principal, l'influence de la situation familiale fait apparaître des tendances similaires en XW et en YZ, mais une déviation plus importante par rapport au portefeuille global est constatée en YZ.

D'autre part, la situation familiale semble être un meilleur critère discriminant que le sexe puisque les écarts en fin d'étude sont plus conséquents (de l'ordre de 30 % pour la situation familiale contre entre 8 % et 15 % selon le pays pour le sexe). Nous distinguons alors deux groupes de variables : un premier groupe est constitué de l'âge de l'assuré et de sa situation familiale, deux variables avec un fort pouvoir discriminant, tandis que le sexe de l'assuré appartient à un second groupe contenant des variables moins pertinentes.

La répartition des polices selon la situation familiale de l'assuré principal en XW et en YZ est donnée dans la figure ci-dessous.

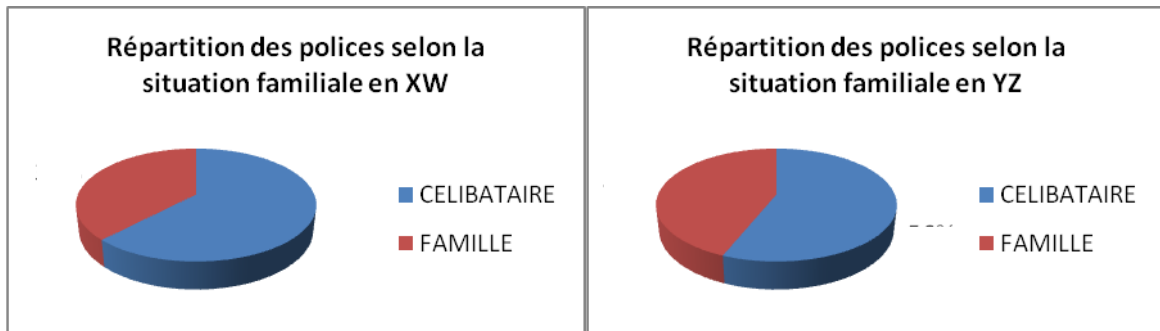


Figure 17 : Répartition des polices selon la situation familiale

Source : Propres calculs à partir de données réelles

Les compositions des portefeuilles XW et YZ sont assez proches. Il convient de faire attention à ces deux graphiques qui peuvent être trompeurs. En effet, si en proportion il y a davantage d'assurés principaux célibataires, en termes de nombre d'assurés et de montant de prime payée, la situation est différente car les assurés principaux en famille représentent plus d'assurés (l'assuré principal plus son conjoint et éventuellement ses enfants) et donc plus de prime qu'un assuré célibataire.

Regardons alors la répartition des primes dans les deux pays en fonction de la situation familiale.

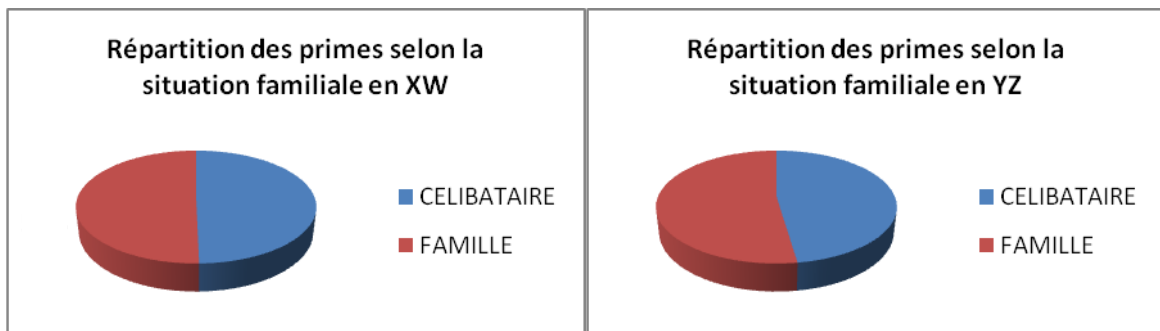


Figure 18 : Répartition des primes selon la situation familiale

Source : Propres calculs à partir de données réelles

Nous constatons que, dans les deux pays, les célibataires et les familles représentent chacun environ 50 % du montant de prime total. Il est donc cohérent que les niveaux de persistance des deux groupes s'éloignent de la même façon du niveau du portefeuille global dans cette étude.

2.3.5. Etude des résiliations selon le montant de garantie souscrit

Cette section vise à étudier l'influence du montant de garantie souscrit sur le niveau de persistance des polices.

2.3.5.1. Portefeuille XW

Les résultats de l'étude sur les niveaux de persistance des polices par rapport aux valeurs trouvées pour le portefeuille global en fonction du montant de garantie souscrit par l'assuré principal sont donnés dans la figure ci-dessous.

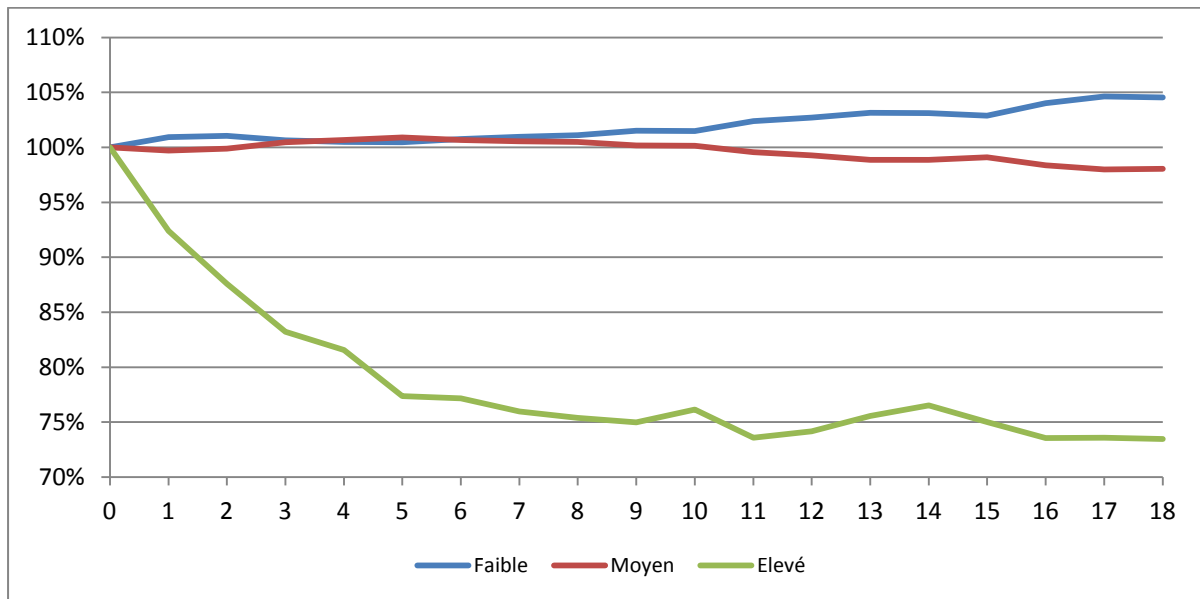


Figure 19 : Comparaison de la persistance après le premier paiement selon le montant de prestation par rapport au portefeuille global en XW

Source : Propres calculs à partir de données réelles

En XW, les assurés ayant souscrit un montant d'indemnité élevé résilient beaucoup plus que les assurés avec un montant d'indemnité moyen ou faible qui se comportent comme le portefeuille global. Nous remarquons tout de même que l'écart se creuse au cours des six premiers mois puis qu'il se stabilise autour de 25 % ensuite. Ainsi, entre le sixième et le dix-huitième mois, la persistance des assurés ayant souscrit un montant d'indemnité élevé ne chute que de 3 %. Après une première période pendant laquelle les assurés résilient beaucoup, il apparaît que ceux qui restent en portefeuille tiennent à garder leur garantie. Ceci est cohérent car les assurés ont, à ce moment, déjà payé un montant de prime important.

2.3.5.2. Portefeuille YZ

Les résultats de l'étude menée sur le portefeuille YZ sont présentés dans la figure ci-dessous.

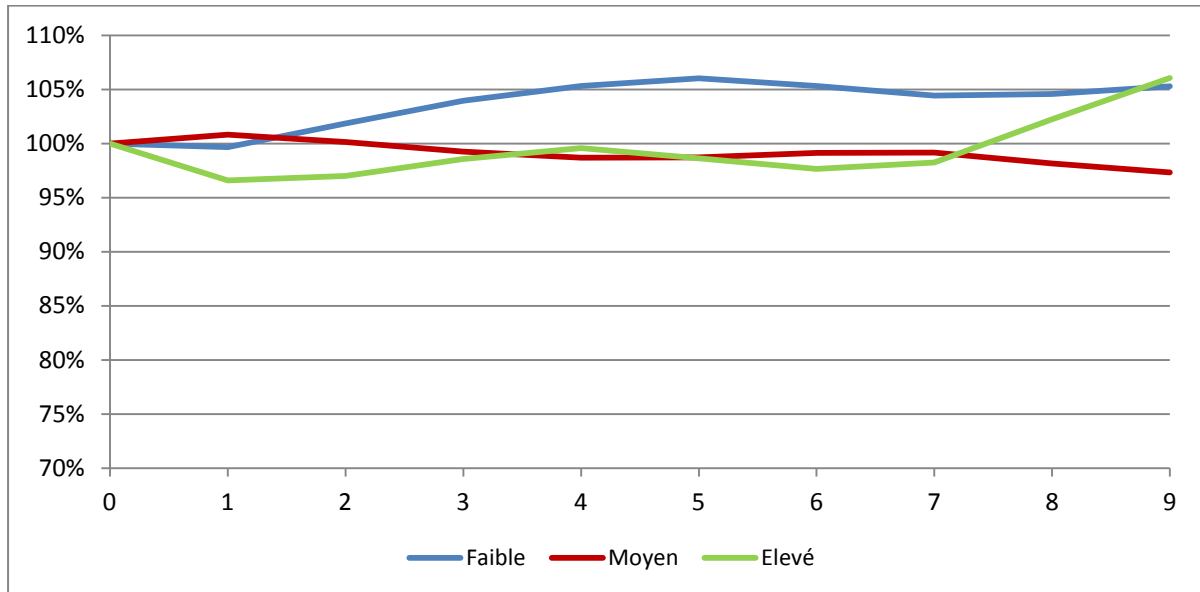


Figure 20 : Comparaison de la persistance après le premier paiement selon le montant de prestation par rapport au portefeuille global en YZ

Source : Propres calculs à partir de données réelles

Il est assez compliqué de dégager une tendance en YZ même s'il semble que les assurés souscrivant des montants d'indemnité faibles résilient moins. Toutefois, les écarts entre les trois groupes restent faibles quel que soit le nombre de mois depuis la souscription du contrat. Après 9 mois, la persistance des assurés avec une indemnité faible est de 105 % du portefeuille global, 97 % pour les indemnités intermédiaires et 106 % pour les indemnités élevées.

2.3.5.3. Commentaires et comparaisons

Le comportement des assurés par rapport à cette dernière variable d'étude est assez différent de ce qui a été observé auparavant. En effet, alors que pour l'âge, le sexe et la situation familiale de l'assuré principal les tendances sont similaires dans les deux pays, pour le montant d'indemnité de l'assuré principal deux profils très différents ressortent de l'étude. En XW, les assurés avec des montants d'indemnité importants résilient davantage au cours des premiers mois. Au contraire, en YZ, aucune tendance nette n'est observée : les taux de résiliation sont très volatils et les niveaux de persistance des polices se maintiennent autour de ceux du portefeuille global. Il est donc assez difficile de savoir si cette variable aura un fort pouvoir discriminant.

Intéressons nous à la répartition des polices selon le montant de l'indemnité en XW et en YZ.

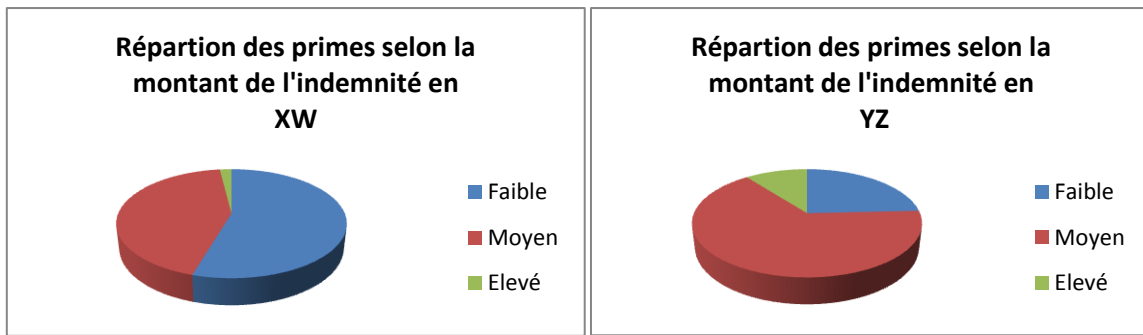


Figure 21 : Répartition des polices selon le montant de l'indemnité

Source : Propres calculs à partir de données réelles

En XW, les assurés ayant souscrit un montant d'indemnité faible sont majoritaires alors que les montants d'indemnité élevés sont marginaux, ce qui explique pourquoi ils s'éloignent considérablement du comportement du portefeuille global. En YZ, la majorité des assurés se dirige vers des montants d'indemnité intermédiaires dont la persistance reste ainsi très proche de celle du portefeuille global.

Il convient ici de prêter attention, comme pour la situation familiale de l'assuré, à la répartition des assurés en termes de primes et pas seulement de polices. Ainsi, les montants d'indemnité élevés ont beau être minoritaires en prenant en compte les polices, ils peuvent prendre un poids plus important en regardant les primes.

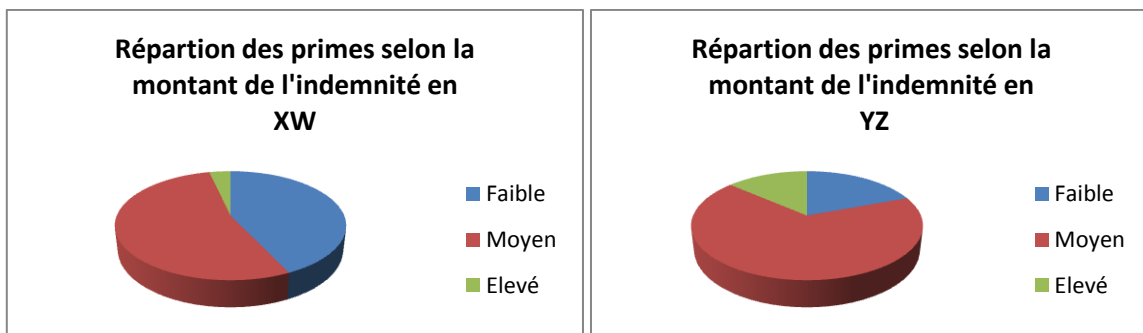


Figure 22 : Répartition des primes selon le montant de l'indemnité

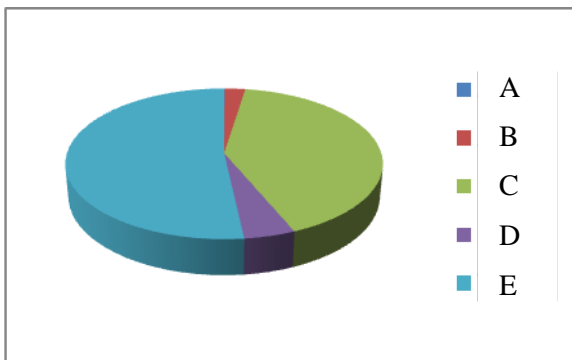
Source : Propres calculs à partir de données réelles

La part des assurés avec un montant d'indemnité élevé évolue finalement assez peu en étudiant les primes. Nous constatons plutôt un basculement en faveur des montants intermédiaires au dépend des montants faibles. Ce basculement est, par ailleurs, plus conséquent en XW. Toutefois, ces remarques ne modifient pas les conclusions de l'étude réalisée précédemment : les montants élevés s'éloignent de la tendance du portefeuille global du fait de leur faible poids au contraire des montants intermédiaires et, dans une moindre mesure, des montants faibles qui représentent la plus grande part des primes perçues.

2.3.6. Etude des résiliations selon la campagne de télémarketing

Nous disposons pour le portefeuille YZ d'une variable d'analyse supplémentaire. Il s'agit de la campagne de télémarketing, c'est-à-dire le listing téléphonique utilisé, sous laquelle le produit a été vendu. Sur la période, quatre fichiers ont été utilisés : « E », « D », « C » et « B » qui forme un panel varié de listes utilisées pour ce type de produit. Les assurés avec la mention « A » sont non identifiés.

La figure ci-dessous donne la proportion d'assurés par option pour le portefeuille YZ.



Deux campagnes, « E » et « C » représentent l'essentiel des polices en portefeuille. Nous allons donc étudier le comportement des assurés pour ces deux campagnes ainsi que pour la campagne « D » qui constitue 5 % des polices en portefeuille.

Figure 23 : Répartition des polices selon la campagne en YZ

Source : Propres calculs à partir de données

Les résultats de l'étude sont présentés dans le graphique ci-dessous.

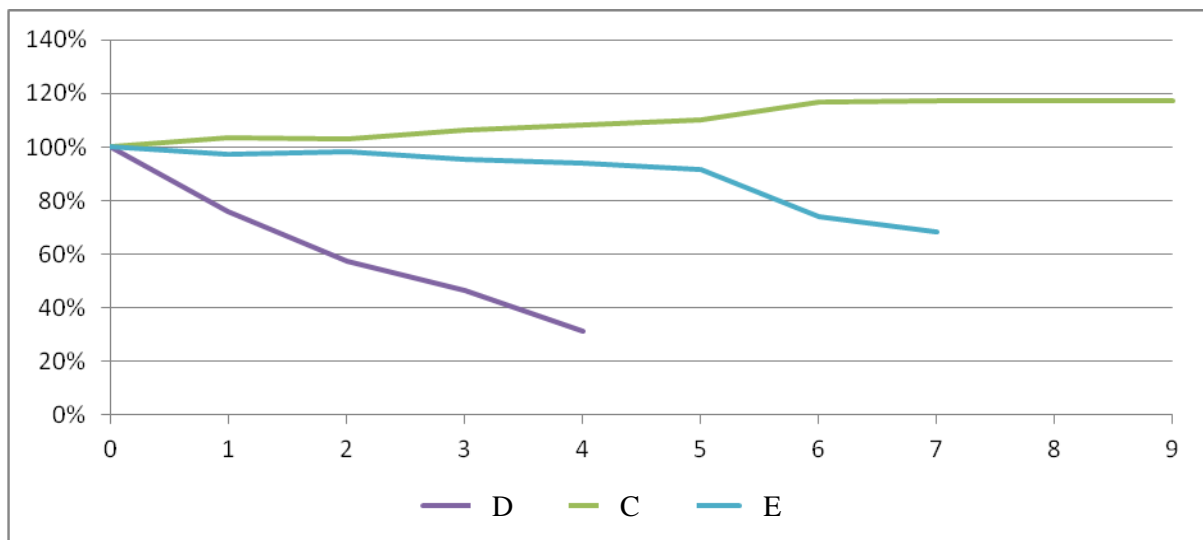


Figure 24 : Comparaison de la persistance après le premier paiement selon la campagne par rapport au portefeuille global en YZ

Source : Propres calculs à partir de données réelles

Notons d'abord que l'historique de données étant plus récent pour les campagnes « E » et « D » (premières polices vendues respectivement en janvier 2010 et avril 2010), l'étude des résiliations porte sur moins de mois que pour la campagne « C ».

Il apparaît que les assurés issus de la campagne « D » résilient davantage que les autres campagnes puisqu'après seulement 4 mois il ne reste plus que 16 % des assurés en portefeuille. Par ailleurs, les persistance des assurés des campagnes « C » et « E » restent proches du niveau du portefeuille global au cours des cinq premiers mois avant de commencer à s'en éloigner : les assurés de la campagne « E » résilient alors plus, et ceux de la campagne « C » moins, que l'ensemble du portefeuille.

Gardons à l'esprit que ces hypothèses sur les profils des assurés selon la campagne de télémarketing sont émises a priori. Une analyse plus rigoureuse de ces profils est menée dans les chapitres suivants.

2.3.7. Etude des résiliations selon le mois de vente

Du fait de l'existence d'un grand nombre de modalités pour cette variable, (18 pour XW et 9 pour YZ) et de la différence d'historique des taux de résiliation pour chaque mois (en effet, plus le mois de vente est proche de la date de fin d'observation, moins l'historique est important), il est, en pratique, difficile de représenter et d'interpréter les résultats obtenus de la même façon que les autres variables explicatives. Toutefois, en notant la croissance du niveau de persistance des primes pour chacun des mois de vente, et en considérant que les résiliations surviennent principalement au cours des premiers mois suivant la souscription, il semble que les résiliations soient plus sévères durant les premiers mois de vente sur les deux portefeuilles.

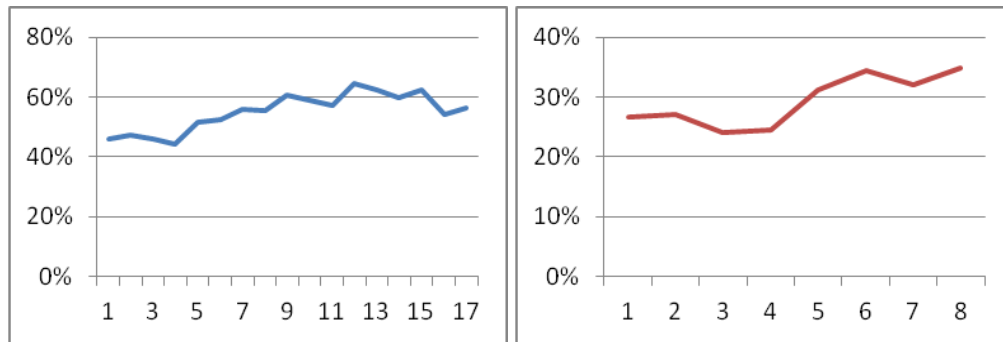


Figure 25 : Comparaison de la persistance à la fin de la période d'observation en fonction du mois de vente pour les portefeuilles XW (à gauche) et YZ (à droite)

Source : Propres calculs à partir de données réelles

Il convient aussi de remarquer, qu'à nouveau, les résiliations semblent plus sévères sur le portefeuille YZ quel que soit le mois de vente de la police.

A l'issue de cette étude descriptive des résiliations, certaines tendances se dégagent sur les deux portefeuilles. Les « jeunes », les femmes, les familles ou bien encore les assurés avec un montant de prime élevé résilient davantage. Nous nous demandons alors si ces variables ne sont pas corrélées entre elles et, si oui, dans quelle mesure. En effet, un effet prix semble apparaître lorsque nous mixons les caractéristiques de ces assurés à risque : les « jeunes » sont souvent la population avec le moins de capacité financière, les femmes souscrivent plus souvent des garanties pour plusieurs membres de la famille ce qui implique un niveau de prime plus élevés. Il convient désormais de modéliser un profil de risque afin de pouvoir orienter par la suite les centres d'appels vers les « meilleures » cibles d'assurés potentiels. Ainsi, pour aller plus loin que la simple analyse descriptive, et pour acquérir une intuition plus poussée, ce mémoire s'intéresse dans les deux prochains chapitres, d'abord de façon théorique puis en développant un modèle spécifique aux portefeuilles étudiés, aux méthodes d'arbres décisionnels et de régression logistique.

Chapitre 3. Les arbres de décision et l'algorithme CART

Dans les domaines de l'aide à la décision et de l'exploration de données (*data mining*), certains algorithmes produisent des arbres de décision, utilisés pour répartir une population d'individus (ici les personnes assurées) en groupes homogènes, selon un ensemble de variables discriminantes (les caractéristiques de la police et de l'assuré) en fonction d'un objectif fixé et connu (ici la modélisation d'une probabilité de résiliation). Il s'agit de prédire avec le plus de précision possible les valeurs prises par la variable de sortie à partir d'un ensemble de variables d'entrée.

L'algorithme CART (*Classification And Regression Trees*) vise à construire un arbre de décision en classifiant un ensemble d'enregistrements. Cette méthode a été développée par Breiman et al. (1984)³⁶ afin de segmenter une population en divisant de façon itérative une base de données en deux groupes à l'aide de lois binaires. Cet outil de décision non paramétrique permet de construire de façon flexible des arbres binaires facilement interprétables dans le cadre de problèmes de classification. L'algorithme CART permet ainsi de déterminer quels aspects sont déterminants dans la décision de l'assuré de résilier son contrat.

Dans ce chapitre, les fondements théoriques de cet algorithme sont présentés afin de pouvoir ensuite appréhender les résultats obtenus par cette méthode dans le cadre du modèle développé pour notre étude.

Section 3.1. Généralités

3.1.1. Position du problème

Prenons l'exemple de notre étude et de notre base de données : nous disposons de N polices pour lesquelles nous avons les valeurs des variables X_1, X_2, \dots, X_m et l'information sur l'état de la police Y . Nous voulons expliquer Y en fonction de X_1, X_2, \dots, X_m . Y est appelée la variable de sortie, les $X_i, i = 1 \dots m$ sont appelées les variables d'entrée. Par exemple X_1 peut représenter le sexe de l'assuré, X_2 sa situation familiale, etc. Le statut de la police est fonction de ces quantités.

En d'autres termes, soit Y et X_1, X_2, \dots, X_m des variables aléatoires. Nous notons $X = (X_1, X_2, \dots, X_m)$. La base de données est la donnée de N $(m + 1)$ -uplets $(X_{j,1}, X_{j,2}, \dots, X_{j,m}, Y_j)$ indépendants et de même loi que (X, Y) . Les variables $\{X_{j,i}, Y_j | 1 \leq j \leq N, 1 \leq i \leq m\}$ sont appelées les observations. Bien que leur loi ne soit pas connue, la notion de variable aléatoire est commode pour les appréhender théoriquement.

Nous savons qu'il existe une fonction \bar{f} mesurable telle que :

$$E[(Y|X_1, X_2, \dots, X_m)] = \bar{f}(X_1, X_2, \dots, X_m)$$

En ce sens, la donnée des variables d'entrée permet théoriquement de prédire la sortie. Si toutes les variables aléatoires considérées appartiennent à l'espace $\mathcal{L}^2(\Omega)$, l'espérance conditionnelle $E(Y|X_1, X_2, \dots, X_m)$ est la projection orthogonale de Y sur le sous-espace $\mathcal{L}^2(\mathcal{F}(X_1, X_2, \dots, X_m))$, où $\mathcal{F}(X_1, X_2, \dots, X_m)$ désigne la tribu engendrée par les variables X_1, X_2, \dots, X_m .

³⁶ Breiman L., Friedman J., Olshen R. A., Stone C. J. (1984), « *Classification and Regression Trees* », Chapman and Hall

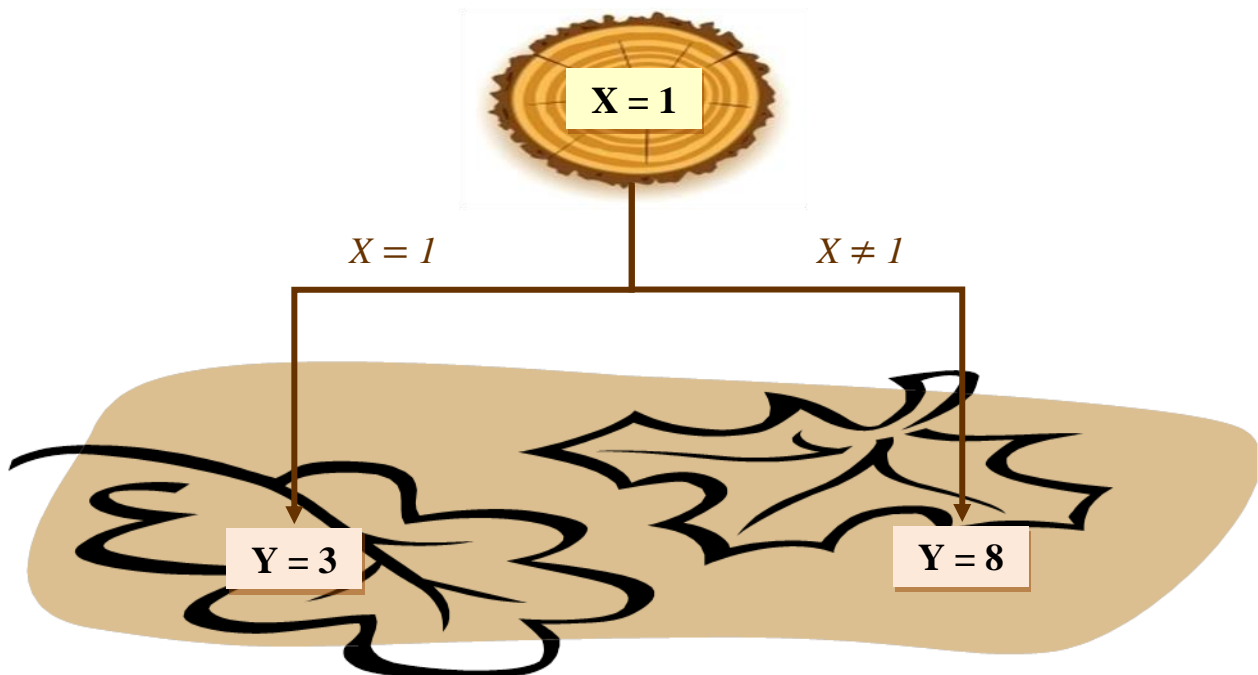
Nous avons donc :

$$\bar{f} = \operatorname{argmin}_{f \in \mathcal{M}(\mathbb{R}^m, \mathbb{R})} E[(Y - f(X))^2]$$

Nous cherchons donc \bar{f} mais la distribution de (X, Y) est inaccessible : nous ne connaissons que les observations. Ne pouvant calculer \bar{f} à partir d'un nombre fini d'observations, nous proposons alors d'estimer \bar{f} , c'est-à-dire de construire à partir des observations une fonction \hat{f} assez proche de \bar{f} selon un certain critère. Ici, ce n'est d'ailleurs pas tant l'estimation de \bar{f} que nous recherchons que la construction explicite, à partir des données, d'une fonction \hat{f} ayant de bonnes capacités prédictives au sens où $E[(Y - f(X))^2]$ n'est pas trop élevée. Ne pouvant en pratique considérer toutes les fonctions mesurables, nous sommes amenés à rechercher \hat{f} sous une certaine forme. L'algorithme CART recherche \hat{f} sous forme d'un arbre.

3.1.2. Arbres de décision

Un arbre binaire est un ensemble de nœuds desquels partent des branches. Une feuille est un nœud dont ne part aucune branche. Considérons par exemple l'arbre ci-après possédant un nœud et deux branches menant à deux feuilles :



Un tel arbre se lit de la façon suivante : si $X = 1$, il faut emprunter la branche de gauche qui mène à une feuille à laquelle est rattachée la valeur 3, on obtient donc $Y = 3$. Si, au contraire, $X \neq 1$, la branche de droite est empruntée et on lit $Y = 8$. L'arbre établit donc un lien fonctionnel entre Y et X , que nous pourrions d'ailleurs expliciter au moyen de fonctions indicatrices.

Le principe de lecture est le même pour des arbres plus complexes. Ainsi, à chaque fois, la valeur de la variable Y est lue au niveau d'une feuille en suivant le chemin découvert au fur et à mesure des questions posées au niveau des nœuds. L'intérêt d'une telle représentation réside dans la facilité de lecture qu'elle procure grâce à la hiérarchisation de l'effet des variables aléatoires d'entrée et de la sélection de leur influence sur la variable aléatoire Y : la première question devrait donc être la plus importante, la deuxième l'être un peu moins et ainsi de suite.

De façon synthétique, un arbre n'est ici qu'une fonction de (X_1, X_2, \dots, X_m) constante par morceaux sur des rectangles du type $\{a_1 \leq X_1 \leq b_1; a_2 \leq X_2 \leq b_2; \dots; a_m \leq X_m \leq b_m\}$.

On notera \mathcal{T} l'ensemble des arbres binaires.

Revenons à la méthode CART. Il s'agit d'un algorithme visant à rechercher le meilleur arbre prédictif $\hat{T} \in \mathcal{T}$. A l'issue de l'algorithme, un arbre prédisant la variable de sortie Y inscrite au niveau des feuilles est obtenu en fonction des variables d'entrée X_i . Le modèle CART donne les variables explicatives les plus discriminantes concernant les résiliations (les variables les plus discriminantes étant les plus proches du sommet de l'arbre). Toutefois, si ce modèle permet de classer un assuré comme « risqué » au moment de la souscription (ou plus tard), la réponse prédite est binaire (nous pouvons obtenir la probabilité de résiliation d'une classe donnée mais pas une probabilité individuelle de résiliation). En cela, l'algorithme CART se distingue des méthodes de modèles linéaires généralisés (MLG) que nous étudions dans un prochain chapitre et qui permettent de donner directement la probabilité de résiliation d'un assuré en fonction de ses caractéristiques et de mesurer l'effet d'un changement de classe (variable catégorielle) ou de valeur (variable continue) d'une variable explicative grâce à la technique des *odd ratios*.

Section 3.2. L'algorithme CART

Cette section présente la façon dont est construit l'arbre de classification. Pour ce faire, la partie théorique et les notations de l'article de X. Milhaud, S. Loisel et V. Maume-Deschamps intitulé « *Surrender triggers in life insurance : classification and risk predictions* » publié en janvier 2010³⁷ sont largement reprises.

3.2.1. Principes

Nous rappelons que l'objectif de l'algorithme CART est de construire un arbre de décision binaire par des divisions successives de l'échantillon de données en deux sous-ensembles.

L'algorithme commence par séparer la base de données en deux groupes. Pour ce faire, il choisit parmi l'ensemble des variables explicatives, la variable et la valeur de cette variable qui permettent de maximiser un critère de séparation. Une fois cette première séparation effectuée, l'algorithme effectue de manière récursive l'étape de séparation en deux groupes dans les nœuds. Notons qu'un des avantages de la méthode CART est l'absence de restrictions quant au type de données qui peuvent être catégorielles ou numériques.

Par ailleurs, contrairement aux autres méthodes de segmentation, elle n'impose aucune règle d'arrêt de division des segments basée sur une approche statistique. Elle fournit, à partir de l'arbre binaire complet, la séquence des sous-arbres obtenue en utilisant une procédure d'élagage. Celle-ci est basée sur la suppression successive des branches les moins informatives et la conservation du nœud qui assure lors du passage à ses segments descendants la plus grande réduction de l'« impureté ».

³⁷ Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance : classification and risk predictions* », Hal, p2-8

Les étapes successives de l'algorithme CART sont données dans la figure ci-dessous.

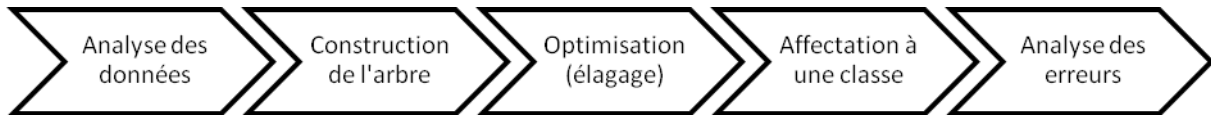


Figure 26 : Étapes successives de l'algorithme CART

3.2.2. Définitions et propriétés

Avant de rentrer plus en détail dans la description des aspects théoriques et pratiques de l'algorithme CART, nous allons énoncer certaines définitions et propriétés mathématiques utilisées plus loin.

Notation 1

Soit $\epsilon = (x_n, j_n)_{1 \leq n \leq N}$ un échantillon de taille N , où j_n est une observation de la variable de sortie Y ($Y \in \mathcal{C} = \{1, 2, \dots, J\}$) et $x_n = \{x_{n,1}, x_{n,2}, \dots, x_{n,m}\}$ une observation de $X = (X_1, X_2, \dots, X_m)$ dans \mathbb{X} ($\mathbb{X} = \prod_{i=1}^m \mathbb{X}_i$ où \mathbb{X}_i est l'ensemble des valeurs pouvant être prises par une variable catégorielle ou continue).

Définition 1

Soit,

- $class$, la fonction de classification telle que :

$$class : \mathbb{X} \rightarrow \mathcal{C}$$

$$x \mapsto class(x, \epsilon) = j$$

- $\pi_j = \frac{N_j}{N}$ où $N_j = card\{j_n | j_n = j\}$, la probabilité a priori d'appartenir au groupe j ,

- $N_t = card\{(x_n, j_n) \in \epsilon, x_n \in t\}$ où $t \subset \mathbb{X}$ (t est un sous-ensemble fini de \mathbb{X}),

- $N_j(t) = card\{(x_n, j_n) \in \epsilon, j_n = j \text{ sachant que } x_n \in t\}$

- $P(j, t)$, la probabilité d'appartenir au sous-ensemble t et au groupe j ,

- $p(j, t) = \pi_j \times \frac{N_j(t)}{N(t)}$ est un estimateur de $P(j, t)$,

- $P(t)$, la probabilité d'appartenir au sous-ensemble t ,

- $p(t) = \sum_{j=1}^J p(j, t)$ est un estimateur de $P(t)$,

- $p(j|t)$, la probabilité a posteriori d'appartenir au groupe j est égale à $\frac{p(j, t)}{p(t)} = \frac{N_j(t)}{N(t)} = \frac{p(j, t)}{\pi_j}$

3.2.3. Construction de l'arbre maximal

La première étape de l'algorithme CART consiste à construire le plus gros arbre possible : T_{Max} . Le point crucial pour cette construction est de déterminer quelle question placer au niveau de chaque nœud. L'algorithme CART commence par construire le premier nœud, ou racine, c'est-à-dire celui qui est le plus haut dans l'arbre. Pour déterminer quelle question placer sur ce nœud, il pose toutes les questions binaires possibles à partir des variables d'entrée. Chaque question produit alors deux sous-ensembles qui partitionnent la base de données : celui pour lequel la réponse à la question est oui et celui pour lequel la réponse est non.

Ils correspondent à deux nœuds. Pour chaque question, l'algorithme CART calcule la partition associée. Il obtient ainsi différentes partitions de la base de données parmi lesquelles il faut choisir la partition optimale.

Plus formellement, ceci consiste à diviser \mathbb{X} en q classes où q n'est pas connu a priori. L'algorithme CART construit alors une séquence de partition de \mathbb{X} . Le passage d'une classe à l'autre est donné par une règle binaire (ou règle de division) du type :

$$x \in t, \text{ pour } t \subset \mathbb{X}$$

Par exemple, la première partition de \mathbb{X} pourrait être le sexe de l'assuré. Considérons t comme la modalité « femme ». L'assuré, dont les caractéristiques sont x , est alors un homme si $x \notin t$ et une femme si $x \in t$.

Critère 1

La règle dépend uniquement d'un « palier » μ et d'une variable $x_l, 1 \leq l \leq d$:

- $x_l \leq \mu, \mu \in \mathbb{R}$ dans le cas d'une variable ordinale,
- $x_l \in \mu$ où μ est un sous-ensemble de $\{\mu_1, \mu_2, \dots, \mu_M\}$ et μ_m ($m = 1, \dots, M$) les modalités d'une variable catégorielle.

Nous partons de la racine \mathbb{X} qui est divisée en deux sous ensembles disjoints, les nœuds, que nous notons t_g et t_d . Chacun des nœuds est alors divisé de la même façon. Nous obtenons finalement la partition T_{Max} de \mathbb{X} en q groupes, les feuilles.

Notation 2

Soit \tilde{T} , l'ensemble des feuilles de l'arbre T .

Afin de déterminer à chaque nœud la meilleure division de t à t_G et t_D , l'algorithme CART évalue l'hétérogénéité par rapport à la variable de sortie Y dans les deux nœuds inférieurs obtenus. Si t est un nœud de l'arbre T on note :

$$R(t) = g(p(1|t), p(2|t), \dots, p(J|t))$$

où g est une fonction d'« impureté ».

Ainsi, l'hétérogénéité de l'arbre T est égale à :

$$\mathcal{R}(T) = \sum_{t \in \tilde{T}} \mathcal{R}(t)$$

où $\mathcal{R}(t) = p(t) \times R(t)$.

Une règle binaire Δ (ou règle de division) d'un nœud t donne $p_G = \frac{p(t_G)}{p(t)}$ observations dans t_G et $p_D = \frac{p(t_D)}{p(t)}$ observations dans t_D . Nous voulons maximiser la « réduction de l'impureté » :

$$\delta R(\Delta, t) = R(t) - p_G \times R(t_G) - p_D \times R(t_D)$$

Propriété 1

Si la fonction d'« impureté » g est concave, alors la « réduction de l'impureté » $\delta R(\Delta, t)$ est positive. Il convient alors de trouver la règle binaire Δ_t^* qui maximise $\delta R(\Delta, t)$:

$$\Delta_t^* = \operatorname{argmax}_{\Delta \in D} (\delta R(\Delta, t))$$

où D est l'ensemble des règles binaires.

A chaque nouvelle division, la « pureté » de l'arbre doit augmenter. L'algorithme CART identifie la partition pour laquelle l'hétérogénéité est minimale dans les deux nœuds. Cette partition optimise le classement des données, puisque celles-ci sont regroupées de la façon la plus homogène possible dans chaque nœud. L'algorithme CART retient donc cette partition ainsi que la question qui lui est associée.

Une fois ceci effectué pour le premier nœud, il faut recommencer le processus avec les deux nouveaux nœuds obtenus : chacun correspond en effet à un nouvel ensemble de données. Nous obtenons alors par récurrence un arbre maximal noté T_{Max} . Cet arbre semble être un bon estimateur de Y puisque à chaque étape nous avons essayé de trier au mieux les données. Les choses sont toutefois plus compliquées.

Problème 1

Pour mieux comprendre pourquoi l'arbre maximal T_{Max} que nous avons construit n'est pas l'arbre optimal, étudions le problème suivant.

Nous simulons 100 tirages d'une gaussienne $Z = \mathcal{N}(0,1)$. Notons $a = (a_1, a_2, \dots, a_{100})$ le résultat. Supposons connu que Z est absolument continue par rapport à la mesure de Lebesgue mais que nous ne connaissons pas sa loi. Uniquement à partir des données, nous cherchons à estimer la densité de Z par un histogramme.

Dans la définition de cet histogramme, le seul paramètre que nous pouvons choisir est le nombre de classes k . Nous allons discuter la valeur de k choisie.

Imaginons pour commencer que $k = 1$. L'histogramme est alors une fonction indicatrice d'intervalle. La densité obtenue a plus l'air de celle d'une variable aléatoire de loi uniforme que d'une gaussienne, elle est en fait trop simple par rapport à la densité de Z . Il faudrait choisir k plus grand, de façon à mieux exploiter l'information apportée par la répartition des nombres $(a_1, a_2, \dots, a_{100})$. L'histogramme obtenu est donc un mauvais prédicteur.

A l'inverse, prenons une valeur de k très élevée par rapport au nombre de données, par exemple $k = 200$. L'histogramme obtenu présente de nombreuses petites bosses ; nous n'avons pas non plus une gaussienne clairement visible. La densité obtenue est plus complexe qu'une gaussienne. Ici, l'histogramme exploite toute l'information apportée par les observations, et plus encore, puisqu'il s'attache aux fluctuations des données : les bosses observées loin de la moyenne 0 sont dues au caractère aléatoire de Z . L'histogramme obtenu reflète donc le vecteur a mais ne parvient pas à se détacher de celui-ci pour estimer la vraie loi. Il est donc aussi un mauvais prédicteur.

La valeur de k doit être raisonnable afin d'éviter les deux écueils constatés. Il faut donc prendre k assez grand pour avoir un modèle assez complexe mais pas trop grand pour se détacher des fluctuations.

Par analogie avec la situation précédente, nous comprenons donc que dans le cas des arbres, il existe également un compromis à établir entre précision du modèle et fluctuation des données. En effet, un arbre trop peu ramifié montre un biais très important et, dans le cas extrême d'un arbre à une feuille, la valeur de Y est toujours la moyenne des données. Toutefois, un arbre trop profond se perd dans les détails de la base et présente une variance importante. C'est en particulier le cas de T_{Max} .

Un algorithme d'élagage a pour fonction de résoudre ce problème. Avant de présenter la procédure d'élagage, étudions comment la qualité de la classification de l'arbre construit est évaluée.

3.2.4. Fonction de classification et estimation de la qualité de l'arbre

Nous rappelons que l'objectif de l'algorithme CART est de construire la fonction *class* telle que :

$$\begin{aligned} \text{class} : \mathbb{X} &\rightarrow \mathcal{C} \\ x &\mapsto \text{class}(x, \epsilon) = j \end{aligned}$$

L'idée sous-jacente est qu'avec cette fonction nous serons capables de classer un assuré selon ses caractéristiques x dans un ensemble B_j pour prédire le résultat. La fonction *class* est ainsi une formalisation mathématique de l'arbre de décision T .

Notation 3

$$B_j = \{x \in \mathbb{X}; \text{class}(x, \epsilon) = j\}$$

Considérons que l'arbre optimal a été construit. Pour savoir à quelle classe les nœuds terminaux correspondent, nous pouvons utiliser le critère suivant.

Critère 2

$$\text{class}(x, \epsilon) = \underset{j \in \mathcal{C}}{\text{argmax}} p(j|t)$$

Ce critère correspond à la règle de Bayes : il faut maximiser la probabilité a posteriori d'appartenir à la classe j sachant que l'on se situe dans le nœud t .

Définition 2

$$r(t) = 1 - \text{class}(x, \epsilon) = 1 - \underset{j \in \mathcal{C}}{\text{max}} p(j|t)$$

Définition 3

Soit,

- $\hat{t}(t) = p(t) \times r(t)$ le taux de mauvaise classification au nœud t . Pour chaque nœud de l'arbre, il représente la probabilité d'être dans le nœud t multiplié par la probabilité de mal classer une observation sachant que nous nous situons dans le nœud t .

- $\hat{t}(T) = \sum_{t \in \bar{T}} \hat{t}(t)$ le taux de mauvaise classification de l'arbre entier T .

Définition 4

Soit,

- $\tau(\text{class}) = \mathbb{P}(\text{class}(X, \epsilon) \neq Y)$ l'erreur de prédiction égale à la probabilité qu'un assuré soit classé dans la mauvaise classe par la fonction $\text{class}(\cdot, \epsilon)$

- $\hat{t}(\text{class}) = \frac{1}{N} \times \sum_{(x_n, j_n) \in \mathcal{E}} \mathbb{1}\{\text{class}(x_n, \epsilon) \neq j_n\}$, un estimateur du taux de mauvaise classification de l'arbre

Problème 2

L'utilisation de l'ensemble de la base de données pour construire l'arbre et pour mesurer l'erreur de classification conduit à sous-estimer cette erreur car nous nous servons des mêmes données pour construire

l'arbre et la fonction de classification et pour tester l'efficacité de la procédure. Il convient alors de découper la base en deux : un échantillon d'apprentissage et un échantillon de validation. Habituellement, l'échantillon d'apprentissage représente $\frac{2}{3}$ de la base totale et l'échantillon de validation $\frac{1}{3}$.

Une deuxième solution est la validation croisée. Cette méthode consiste à découper la base de départ en K sous-groupes disjoints $(\epsilon_k)_{1 \leq k \leq K}$ de même taille et de définir alors K échantillons d'apprentissage tels que $\epsilon^k = \epsilon - \epsilon_k$. Il est alors possible de définir une fonction de classification sur chaque échantillon ϵ^k telle que $class^k(.) = class(., \epsilon^k)$.

Définition 5

Soit,

$$\begin{aligned}
 - \hat{\tau}^{val}(class) &= \frac{1}{N'} \times \sum_{(x_n, j_n) \in W} \mathbb{1}\{class(x_n, \epsilon) \neq j_n\}, & \text{où } N' = \frac{N}{3} \text{ et } W \text{ est le sous-ensemble de } \epsilon \\
 & & \text{correspondant à l'échantillon de validation.} \\
 - \hat{\tau}^{vc}(class) &= \frac{1}{N} \times \sum_{k=1}^K \sum_{(x_n, j_n) \in \epsilon^k} \mathbb{1}\{class(x_n, \epsilon^k) \neq j_n\}
 \end{aligned}$$

Les concepts utiles à l'évaluation de la qualité de classification d'un arbre ayant été présentés, étudions désormais la procédure d'élagage de l'arbre T_{Max} qui doit mener à la construction de l'arbre optimal en termes de minimisation de l'erreur de classification.

3.2.5. Elagage de l'arbre maximal

Le problème de la complexité trop importante de l'arbre T_{Max} peut facilement être résolu en utilisant un bon estimateur du taux de mauvaise classification pour déterminer la bonne profondeur de l'arbre parmi l'ensemble des arbres élagués.

Désormais, si T et T' sont deux arbres, nous noterons $T \preceq T'$ si T est un sous-arbre de T' c'est-à-dire si tous les nœuds et branches de T sont des nœuds et branches de T' et que T et T' ont la même racine.

Pour obtenir le meilleur arbre possible, nous sommes donc tentés de prendre :

$$\hat{T} = argmin_{T \preceq T_{Max}} \tau(T)$$

Nous utilisons pour cela les techniques d'échantillon d'apprentissage/échantillon de validation et de validation croisée. En effet, nous rappelons que $\tau(T)$ est l'erreur de prédiction (ou taux de mauvaise classification) de l'arbre T et que $\hat{\tau}(T)$, $\hat{\tau}^{val}(T)$ et $\hat{\tau}^{vc}(T)$ sont des estimateurs de cette erreur. Nous allons donc chercher l'arbre qui minimise l'estimateur de l'erreur de prédiction.

Le choix de cet arbre permet de maximiser la précision de l'arbre mais conduit à des fluctuations importantes. L'arbre obtenu est T_{Max} et nous sommes confrontés au problème de complexité trop importante de l'arbre. Une solution, pour réaliser méthodiquement l'élagage de l'arbre T_{Max} est de modifier $\tau(T)$. Nous allons introduire dans le terme d'erreur un coût de complexité représentant cette idée. Le nouveau taux de mauvaise classification (ou mesure du coût de complexité) est alors :

$$\hat{\tau}_\alpha(T) = \hat{\tau}(T) + \alpha \times Card(\tilde{T})$$

où $\alpha > 0$, $Card(\tilde{T})$ est le nombre de nœuds terminaux de T et $\alpha \times Card(\tilde{T})$ est appelé le terme de complexité.

Nous cherchons donc le sous-arbre $T(\alpha) \leq T_{Max}$ qui minimise $\hat{\tau}_\alpha(T)$:

$$\tau_\alpha(T(\alpha)) = \min_{T \leq T_{Max}} \tau_\alpha(T)$$

Pour les problèmes d'existence et d'unicité de l'arbre $T(\alpha)$, le lecteur pourra se référer au livre de référence de Breiman et al. (1984)³⁸.

Par ailleurs, le paramètre α , appelé par la suite coût de complexité, est lié à la taille de l'arbre final élagué : si α est petit, alors la pénalité pour avoir beaucoup de feuilles est faible et l'arbre $T(\alpha)$ est grand. Deux cas remarquables sont à signaler :

- $\alpha = 0$: chaque feuille contient seulement une observation (T_{Max} est très grand). Chaque assuré est correctement classifié et $\tau(T_{Max}) = 0$. T_{Max} minimise $\tau_0(T)$.
- $\alpha \rightarrow \infty$: la pénalité est importante et l'arbre minimisant l'erreur ne contiendra qu'un seul nœud, la racine de l'arbre.

Nous ne détaillerons pas ici l'algorithme d'élagage (le lecteur intéressé pourra se reporter à l'article de X.Milhaud, S.Loisel et V.Maume-Deschamps³⁹ évoqué précédemment) mais l'idée générale est la suivante :

- calculer effectivement pour chaque α l'arbre $T(\alpha)$ minimum,
- déterminer parmi ce premier choix d'arbres, lequel est le meilleur en termes de minimisation de l'erreur de prédiction.

Remarquons que la valeur de α est choisie une fois connus les différents minima pour chaque valeur de α . Dans une certaine mesure, ce sont donc les données elles-mêmes qui nous indiquent quel modèle retenir.

3.2.6. Limites et améliorations

L'utilisation d'arbres de classification offre certains avantages comme ne pas avoir à se restreindre à un certain type de données (les variables catégorielles ou numériques sont toutes deux acceptées) ou disposer d'une classification finale sous une forme simple et facilement compréhensible.

Dans sa recherche de la meilleure division possible, l'algorithme réalise une sélection progressive des variables en essayant de réduire le coût de complexité. Par ailleurs, il s'agit d'une méthode non-paramétrique qui ne requiert donc pas une spécification particulière de la relation entre les variables de sortie et d'entrée.

Toutefois, l'effet d'une variable d'entrée peut être masqué dans l'arbre final. Aussi, il peut exister des divisions presque aussi bonnes, pouvant jouer un rôle important au niveau des interprétations et qui conduiraient à la construction d'un arbre différent. Pour éviter ceci, il existe des solutions comme la classification des variables en fonction de leur potentiel : il s'agit de la technique des divisions concurrentes (*secondary splits*) qui assurent après \hat{T} , la division optimale, les plus fortes réductions de l'impureté. Par

³⁸ Breiman L., Friedman J., Olshen R. A., Stone C. J. (1984), « *Classification and Regression Trees* », Chapman and Hall

³⁹ Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance : classification and risk predictions* », Hal, p24-25

ailleurs, un des avantages supplémentaires de l'algorithme CART est sa capacité à gérer les valeurs manquantes dans la base de données. En effet, en présence de valeurs manquantes, la majorité des algorithmes d'apprentissage ou des méthodes de modèles linéaires généralisés proposent soit de supprimer l'observation, soit de remplacer la valeur manquante par la valeur la plus probable. L'algorithme CART propose une alternative à ces solutions en offrant la possibilité de créer des divisions suppléantes (*surrogate splits*) qui fournissent les répartitions les plus proches de la division \hat{T} . Elles permettent de gérer l'existence de données manquantes dans l'affectation d'un nouvel individu à une classe. Quand, pour un individu, une donnée est manquante pour une variable divisant un segment, l'algorithme cherche la variable la remplaçant au mieux. Par exemple, si la séparation du nœud a lieu à 60 ans et que la seconde variable qui explique le mieux cette séparation est le sexe féminin alors, si l'âge n'est pas renseigné pour un assuré, l'algorithme regarde la valeur du sexe et classe l'observation dans un nœud en fonction du sexe de l'assuré. Ainsi, toutes les observations sont utilisées pour paramétrer l'arbre ce qui permet d'améliorer la qualité de la prédiction. Toutefois, ces divisions de réserve (concurrentes et suppléantes) ne sont pas pleinement satisfaisantes. En effet, l'approche par segmentation perd alors en simplicité et l'utilisateur peut être déconcerté par l'instabilité des arbres obtenus. De plus, cette méthode de segmentation nécessite des échantillons de grande taille ce qui est le cas dans cette étude. En outre, la sélection des variables peut présenter un léger biais du fait que les variables explicatives ayant un plus grand nombre de modalités, en offrant plus de divisions possibles, sont plus souvent sélectionnées.

Par ailleurs, les divisions ne se font que sur des variables seules et, lorsque la structure d'une classe dépend de combinaisons de variables, l'algorithme standard ne pourra pas rendre compte de ce type de structure.

Il existe d'autres difficultés, en particulier :

- La difficulté de lire l'arbre final à cause de ses nombreuses ramifications. Ce problème est lié à la trop grande variance de l'arbre T_{Max} due à sa trop grande précision par rapport à la base de données. Ainsi, ce n'est pas parce que l'arbre est très ramifié qu'il sera meilleur. Ce dilemme entre précision et variabilité peut-être réduit en utilisant une procédure d'élagage et en introduisant un coût de complexité.
- L'algorithme CART fournit un moyen d'avoir une idée de l'importance de chacune des variables explicatives. Ainsi, parcourir l'arbre de la racine aux feuilles donne un classement décroissant de l'importance des variables. Cependant, cette méthode est relativement instable puisque qu'une petite modification des données peut donner des classifications finales assez différentes. Cette instabilité se révèle très problématique si nous souhaitons faire des prédictions : cela signifierait qu'une variable peut-être considérée comme importante avec un jeu de données et absente avec un autre quasi-similaire. Ce problème peut toutefois être résolu en utilisant la validation croisée, un échantillon d'apprentissage et un échantillon de validation ou des méthodes de ré-échantillonnage telles que les forêts aléatoires (*random forest*).

3.2.6.1. Principes de la méthode des forêts aléatoires

Cette sous-section ne rentre pas dans les détails théoriques de la méthode des forêts aléatoires. Le lecteur intéressé pourra se reporter aux ouvrages⁴⁰ et au site internet⁴¹ de Breiman. Toutefois, nous allons présenter les grands principes des forêts aléatoires qui sont ensuite utilisées dans l'étude réalisée.

La faible robustesse de l'algorithme CART lors du changement du jeu de données original a déjà été évoquée. Ceci conduit à essayer d'autres classifications optimales mais ce problème peut aussi être résolu par des méthodes de ré-échantillonnage dont la plus célèbre est le *bootstrap* (c'est-à-dire un tirage aléatoire de N individus avec remise dans un échantillon de taille N). Le principe des forêts aléatoires repose sur un cas particulier de la méthode *bagging* qui est une variante du *bootstrap*.

Le *bagging* est une agrégation de classificateurs obtenus à l'aide des échantillons *bootstrap*. En pratique, le classificateur *bagging* affecte à une observation une classe qui est la classe majoritaire parmi les classificateurs *bootstrap*. La décision de classification finale s'assimile donc à un « vote » à la majorité sur l'ensemble des arbres de classification construits à partir des échantillons *bootstrap* : la forêt choisit la classification la plus représentée parmi les arbres de la forêt. Signalons qu'ici le classificateur ne peut être représenté sous la forme d'un arbre puisqu'il s'agirait d'un arbre qui emprunterait les caractéristiques majoritaires d'un ensemble d'arbres.

L'outil de forêts aléatoires est une variante du *bagging* appliquée à des arbres de décision dans laquelle est introduit un tirage aléatoire parmi les variables explicatives. Nous avons donc une *randomisation* supplémentaire par rapport à un simple *bagging* (d'où le nom de forêt aléatoire). Avec une telle méthode, l'intérêt est d'avoir un grand nombre d'arbres pour que l'algorithme ait une fiabilité satisfaisante. L'erreur de la forêt (*out-of-bag error* dans la littérature) correspond au taux de mauvaise classification des valeurs prédites à l'aide des arbres construits pour les observations non utilisées dans le processus de *bagging*. Cette erreur se stabilise rapidement lorsque le nombre d'arbres de la forêt augmente.

En pratique, la première étape consiste à « bootstraper » l'échantillon de données. Il s'agit ensuite de diviser l'échantillon à chacun des nœuds en utilisant la meilleure variable entre les m' variables choisies aléatoirement parmi les m variables explicatives de départ pour réduire au maximum le terme d'« impureté » (ici $m' < m$ car le cas $m' = m$ correspond à un *bagging* simple). On construit alors le plus gros arbre possible sans élaguer. Le taux d'erreur de la forêt dépend de la capacité de chacun des arbres à bien classer les individus et de la corrélation entre chaque paire d'arbres de la forêt. Quand la force de l'arbre augmente, l'erreur de la forêt décroît mais quand la corrélation augmente, l'erreur devient plus importante. Le seul paramètre ajustable est m' , le nombre de variables explicatives sélectionnées aléatoirement pour créer les arbres. Ainsi une réduction de ce paramètre réduit à la fois la corrélation et la force des arbres de la forêt. Il s'agit alors de trouver le paramètre m' optimal qui réduit l'erreur de la forêt. Il n'est techniquement pas possible de représenter la forêt, mais nous allons constater qu'il est possible d'obtenir de meilleurs résultats avec cette méthode. Enfin, avec cette méthode de ré-échantillonnage, il n'y a pas besoin de séparer la base de données en échantillons d'apprentissage et de validation (ceci n'aurait pas de sens dans le cadre de l'utilisation du *bootstrap*).

⁴⁰ Breiman L. (2001), « *Random forests* », Machine Learning (45), p5-32

⁴¹ <http://www.stat.berkeley.edu/users/breiman/RandomForests/>

3.2.7. Implémentation de l'algorithme CART avec le logiciel de statistique R

Plusieurs packages existent sous R⁴² pour construire des arbres de décision avec l'algorithme CART. Le package « *rpart* »⁴³ de T.M.Therneau et B.Atkinson a été retenu. Les résultats produits par l'algorithme dépendent fortement du choix des paramètres du modèle rassemblés dans ce vecteur :

```
rpart.control(minbucket=1,maxcompete=2, maxsurrogate=0, xval=10,cp=0)
```

minbucket désigne le nombre minimum d'individus dans une feuille de T_{Max} . Ce paramètre a été fixé à 1 afin de pouvoir créer l'arbre T_{Max} le plus précis possible.

maxcompete est le nombre maximum de divisions concurrentes implémentées. Nous avons choisi de prendre ce paramètre égal à 2.

maxsurrogate est le paramètre qui permet de définir le nombre de variables suppléantes. Ici, ce paramètre est fixé à 0 car il a été fait en sorte de ne pas avoir de données manquantes dans la base.

xval désigne le nombre de validations croisées effectuées par le modèle. Le nombre d'échantillons pour la validation croisée est fixé à 10. Ainsi, initialement, la base de données est divisée aléatoirement en 10 parties. Ensuite, le modèle est ajusté sur la base sans la première partie (qui sert de base de validation), puis il est ajusté sur la base sans la seconde partie, etc. Le meilleur des 10 modèles ajustés est retenu. Cette technique de validation croisée nous permet d'obtenir des résultats plus précis.

cp est le coût de complexité de l'arbre, proportionnel au nombre de nœuds de l'arbre et à l'erreur d'apprentissage. Le fixer à zéro permet de ne pas retenir ce critère d'arrêt de l'algorithme et donc de produire l'arbre T_{Max} le plus complet avant élagage.

La fonction *prune()* permet ensuite d'élaguer l'arbre T_{Max} obtenu avec la fonction *rpart()*.

Par ailleurs, la fonction d'« impureté » *g* utilisée par défaut par la fonction *rpart()* est l'indice de Gini. Ainsi, *g* est égale à :

$$g(t) = \sum_{j \neq k} P(j|t) \times P(k|t)$$

L'indice de Gini est une fonction concave. La propriété 1 (« réduction de l'impureté » positive à chaque nœud) est donc bien vérifiée.

D'autre part, pour consolider les résultats obtenus par la fonction *rpart()*, nous utilisons le package « *randomForest* »⁴⁴ implémenté sur le logiciel R par A.Liaw et M.Wiener.

Les principaux paramètres à choisir dans la fonction *randomForest()*, en plus de la spécification du modèle (qui est celui testé avec la fonction *rpart()*) et du choix des données (il s'agit ici de la base entière car il n'est pas nécessaire de séparer la base en un échantillon d'apprentissage et de validation avec cette méthode), sont le nombre d'arbres de la forêt et le nombre *m'* de variables choisies aléatoirement parmi les *m* variables explicatives de départ dont dépend l'erreur (*out-of-bag error*) commise par la forêt. Deux variables explicatives sont choisies à chacune des divisions. Par ailleurs, théoriquement, l'erreur se stabilise rapidement

⁴² Librement téléchargeable sur <http://www.r-project.org/>

⁴³ Documentation disponible sur <http://cran.r-project.org/web/packages/rpart/rpart.pdf>

⁴⁴ Documentation disponible sur <http://cran.r-project.org/web/packages/randomForest/randomForest.pdf>

lorsque le nombre d'arbres de la forêt augmente. Nous essaierons de vérifier cette propriété graphiquement en fixant le nombre d'arbres à 100.

Section 3.3. Application de l'algorithme CART aux deux portefeuilles

Cette section a pour objectif d'expliquer puis de modéliser le taux de résiliation des contrats d'hospitalisation distribués par télémarketing en fonction des caractéristiques de la police et de l'assuré.

3.3.1. Les données

Les fichiers de données sont obtenus à partir des mêmes bases que celles de l'étude réalisée précédemment.

Les variables explicatives entrant dans le cadre de l'étude sont l'âge, le sexe, la situation familiale de l'assuré principal, le niveau de garantie de la police, le montant de la prime mensuelle, le mois de vente de la police et le nombre de primes payées (ainsi que la campagne de télémarketing pour le portefeuille YZ). Le tableau ci-dessous donne une description complète de celles-ci.

Code Variable	Description
STAT	Statut de la police : « CHUTE » ou « NON »
AGE	Tranche d'âge de l'assuré principal sur la police : « 18-34 », « 35-44 », « 45-54 », « 55-65 »
SEXE	Sexe de l'assuré principal : Homme (« H ») ou Femme (« F »)
SF	Situation familiale de l'assuré principal : « CELIBATAIRE » ou « FAMILLE »
GAR	Niveau de garantie de la police : « Faible », « Moyenne » ou « Elevée »
MV	Mois de vente de la police
NMP	Nombre de primes payées : compris entre 0 et 23(XW) ou 0 et 11 (YZ)
PMM	Prime mensuelle moyenne : « <15 € », « 15 €-30 € » ou « >30 € »
CAMP	Campagne de télémarketing : « C », « D », « E »

Tableau 7: Description des variables utilisées dans l'implémentation de l'algorithme CART

Un échantillon d'apprentissage et un échantillon de validation ont été créés pour les deux pays avec respectivement $\frac{2}{3}$ et $\frac{1}{3}$ de l'échantillon total comme recommandé par la théorie.

Par ailleurs, les mois de vente des polices ont été regroupés par période de 6 mois pour le portefeuille XW et de trois mois pour le portefeuille YZ. La variable explicative correspondant à ce regroupement est notée « MV2 ». De même, les nombres de primes payées sont amalgamés au sein des classes « 0 », « 1→3 », « 4→10 », « 11 » et « 12→23 » pour le portefeuille XW et « 0 », « 1→3 », « 4→10 » et « 11 » pour le portefeuille YZ constituant ainsi la variable « NMP2 ».

Finalement, nous obtenons empiriquement des échantillons d'apprentissage de 18 733 polices pour XW et de 11 482 polices pour YZ et des échantillons de validation de 9 292 polices pour XW et 5 665 pour YZ.

3.3.2. Portefeuille XW

L'arbre T_{Max} obtenu à l'aide de la fonction $rpart()$ paramétrée par le vecteur $rpart.control()$ décrit précédemment compte 111 feuilles.

	Prédit $Y = 0$ (résiliation)	Prédit $Y = 1$ (non résiliation)
Observé $Y = 0$	1 738	433
Observé $Y = 1$	234	6 887

Tableau 8 : Matrice d'erreur pour T_{Max} sur l'échantillon de validation

Source : Propres calculs à partir de données réelles

L'erreur de prédiction calculée pour l'arbre T_{Max} sur l'échantillon de validation est égale à 7,2 %. L'arbre est donc un prédicteur précis mais il possède trop de branches pour être lisible. Nous allons donc élaguer.

L'algorithme d'élagage dépend du paramètre de complexité choisi dans la fonction $prune()$ de R. Il s'agit ici de trouver un bon compromis entre la taille finale de l'arbre et le taux de mauvaise classification le plus faible possible.

La fonction $printcp()$ du package $rpart()$ permet de visualiser la table des coûts de complexité.

CP	nsplit	rel error	xerror	xstd	CP	nsplit	rel error	xerror	xstd
2.0530e-01	0	1.00000	1.00000	0.0133054	1.1521e-04	32	0.30530	0.31544	0.0082079
7.7304e-02	2	0.58940	0.58940	0.0108288	9.8749e-05	47	0.30323	0.31959	0.0082574
5.9677e-02	4	0.43479	0.43479	0.0094916	7.6805e-05	54	0.30253	0.32258	0.0082929
5.5300e-02	5	0.37512	0.38134	0.0089500	6.5833e-05	69	0.30115	0.32327	0.0083011
1.2903e-03	6	0.31982	0.31982	0.0082602	5.7604e-05	76	0.30069	0.32350	0.0083038
1.0945e-03	11	0.31336	0.31359	0.0081858	4.6083e-05	94	0.29954	0.32350	0.0083038
2.3041e-04	15	0.30899	0.30968	0.0081385	3.8402e-05	99	0.29931	0.32512	0.0083228
2.0161e-04	18	0.30829	0.31475	0.0081996	0.0000e+00	111	0.29885	0.32627	0.0083363
1.5361e-04	29	0.30576	0.31544	0.0082079					

Tableau 9 : Table des paramètres de complexité

Source : Propres calculs à partir de données réelles

Chaque paramètre de complexité (CP) correspond à un arbre optimal dont la taille (nsplit) est spécifiée. Par défaut, ces arbres sont construits en utilisant 10 validations croisées. La quantité *rel error* est l'erreur d'apprentissage (erreur de prédiction ou taux de mauvaise classification). Cette erreur diminue avec la taille de l'arbre construit et mesure la fiabilité de l'arbre. *xerror* correspond au taux d'erreur sur les 10 échantillons issus de la validation croisée. Enfin *xstd* est l'écart-type de *xerror*. L'arbre optimal est celui minimisant $err = xstd + xerror$. Si deux arbres ont la même erreur *err*, celui avec le moins de feuilles est sélectionné. Par ailleurs, le package $rpart()$ offre aussi l'opportunité de construire le graphe du taux d'erreur en fonction du paramètre de complexité et de la taille de l'arbre.

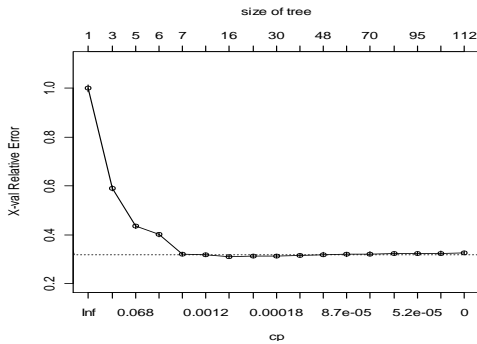


Figure 27 : Taux d'erreur de la validation croisée en fonction de la taille de l'arbre et du coût de complexité

Source : Propres calculs à partir de données réelles

Nous constatons que l'arbre T_{Max} possède bien 112 divisions (la racine plus les 111 feuilles de l'arbre) et correspond à un coût de complexité nul. Le graphique ci-contre fait également apparaître une chute rapide de l'erreur de validation croisée sur les premières divisions suivie d'un plateau caractérisé par une croissance faible de l'erreur.

Un arbre possédant 15 feuilles et un coût de complexité égal à $2,3041e-04$ permet de minimiser l'erreur de validation croisée. Cet arbre, présentant de bonnes propriétés et facilement interprétable, est conservé.

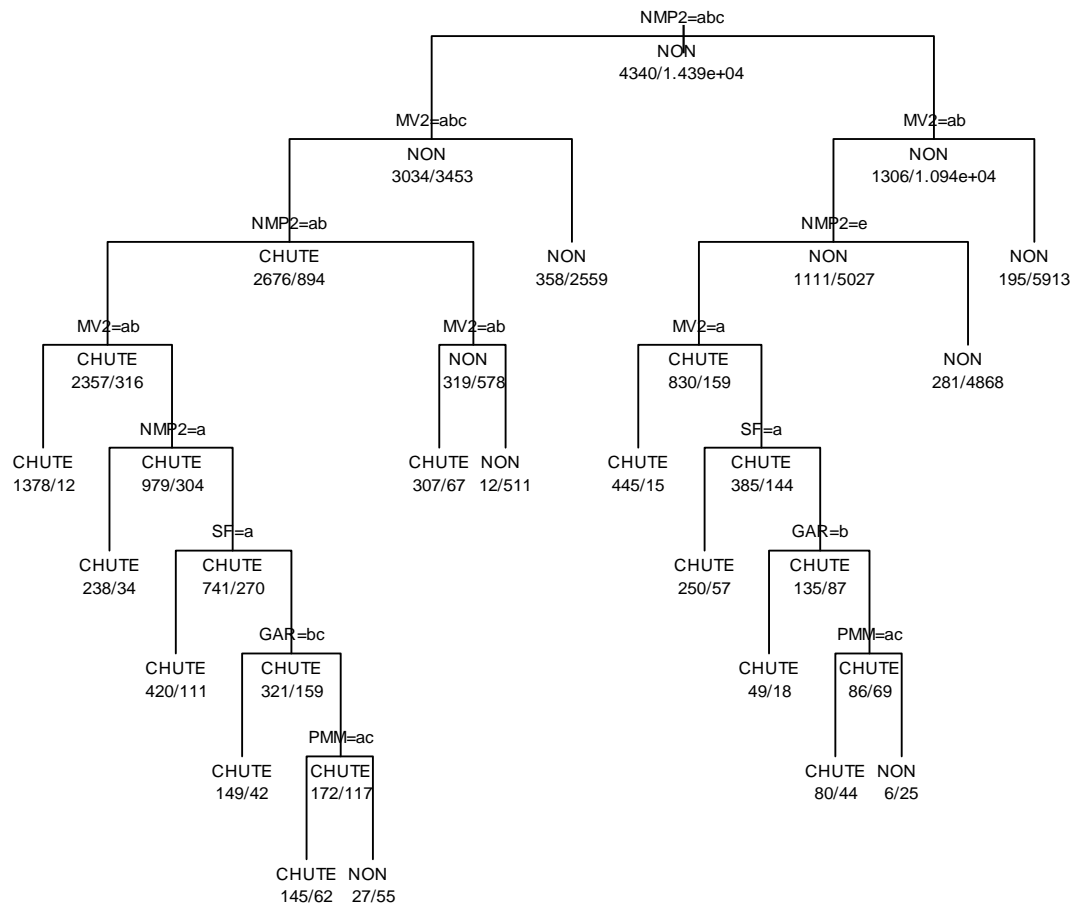


Figure 28 : Arbre de classification élagué

Source : Propres calculs à partir de données réelles

L'arbre ainsi obtenu donne à chaque division une réponse à la question binaire suivante : l'assuré résilie-t-il son contrat ? A chaque nœud, le nombre de gauche correspond à la réponse « OUI » et celui de droite à la réponse « NON ». La classe prédictive (libellée « CHUTE » ou « NON ») correspond à la réponse majoritaire à la question posée.

Le premier nœud correspond à l'ensemble du portefeuille avant toute segmentation. La probabilité de résilier est de $\frac{4\,340}{18\,733} = 23,17\%$ et la classe prédictive est donc « NON » puisque moins de 50 % de l'échantillon d'apprentissage a chuté. L'échantillon est alors divisé en deux selon le nombre de primes payées (« NMP2 ») par l'assuré. S'il possède la modalité a, b ou c (les modalités prises par les variables explicatives étant classées par ordre alphabétique⁴⁵, cela correspond aux assurés ayant payés 0, 1→3 ou 11 primes), il passe dans la feuille de gauche, sinon il passe à droite. Ainsi, si la police possède une de ces modalités, la probabilité de résiliation de l'assuré est égale à $\frac{3\,034}{3\,034 + 3\,453} = 46,77\%$ et la classe prédictive est « NON ». Toutefois, la proportion d'assurés qui résilient leur contrat est plus grande que dans l'échantillon pris dans son ensemble. Nous retrouvons ici un phénomène déjà aperçu : les polices sont davantage résiliées lors des premiers mois du contrat et au moment du renouvellement.

Notons que les variables explicatives continues ont été initialement transformées en variables catégorielles pour l'étude.

Au vu de l'arbre obtenu, les principales variables discriminantes semblent être le nombre de primes payées et le mois de vente (« MV2 »). La situation familiale (« SF »), le niveau de garantie (« GAR ») et la prime mensuelle (« PMM ») ont également une importance dans la décision de l'assuré et apparaissent dans les feuilles inférieures de l'arbre élagué. En effet, il semble que les individus célibataires, avec un niveau de garantie moyen voire élevé et payant une prime mensuelle faible ou moyenne tendent davantage à résilier que les individus ne présentant pas ces caractéristiques. Cependant, les résultats obtenus ne reflètent pas toujours les tendances observées dans les études uni-variées et il convient donc de nuancer les conclusions sur ces trois variables dans l'algorithme CART. En effet, celles-ci n'apparaissent que dans les branches inférieures de l'arbre et subissent l'influence des premières divisions opérées au niveau de l'échantillon. Il est donc nécessaire de regarder l'ensemble de l'arbre, en particulier les branches supérieures, pour pouvoir appréhender l'importance réelle d'une variable dans la décision de l'assuré ainsi que les interactions possibles avec d'autres variables.

Le sexe (« SEXE ») et l'âge (« AGE ») de l'assuré principal ne figurent pas dans cet arbre. En réalité, la représentation de l'arbre occulte les divisions moins informatives (les divisions concurrentes) à chacun des nœuds. Une analyse sur l'arbre complet ferait néanmoins apparaître ces variables. Il s'agit donc ici d'un problème de présentation des résultats lié à la nature des arbres de classification.

Donnons un second exemple de lecture des résultats de l'arbre élagué et d'interaction entre les modalités des différentes variables. Considérons une police dont les caractéristiques sont les suivantes : un nombre de primes payées appartenant à la catégorie b (qui correspond de 1 à 3 primes payées), un mois de vente de la modalité c (c'est-à-dire entre décembre 2009 et mai 2010), souscrite dans le cadre d'une adhésion familiale pour un montant de garantie faible et une prime mensuelle moyenne faible ou moyenne. Un tel assuré est dans une situation risquée puisque $\frac{145}{145+62} = 70,04\%$ des individus possédant ces caractéristiques ont résilié.

D'un point de vue prédictif, l'arbre élagué obtenu permet de réaliser une bonne prédiction puisque seulement 7,0 % des prédictions sont erronées selon la matrice d'erreur sur l'échantillon de validation que nous avons représentée ci-dessous.

⁴⁵ Voir Annexe 4 : Table des correspondances pour l'algorithme CART sur le portefeuille XW

	Prédit Y = 0 (résiliation)	Prédit Y = 1 (non résiliation)
Observé Y = 0	1 756	415
Observé Y = 1	235	6 886

Tableau 10 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation

Source : Propres calculs à partir de données réelles

L'élagage de l'arbre est donc une opération intéressante puisqu'elle permet d'obtenir un arbre plus facilement interprétable tout en diminuant de 0,2 % l'erreur de prédiction sur l'échantillon de validation. Cependant, l'information la plus pertinente à analyser, au-delà de l'erreur globale, est l'erreur en termes de polices à risque ayant été considérées par l'algorithme comme des polices non résiliées. Pour l'arbre T_{Max} , $\frac{433}{433+1\ 738} = 19,9\%$ des polices risquées ont été mal classées contre 19,1 % pour l'arbre élagué. L'opération d'élagage a donc aussi été profitable dans l'amélioration de la prédiction des résiliations.

Analysons maintenant les résultats obtenus avec la fonction `randomForest()`. L'erreur de la forêt est de 7,2 % et l'erreur en termes de détection des polices à risque est de 20,3 %. Nous obtenons des résultats proches de ceux de l'arbre T_{Max} et de l'arbre élagué. La matrice d'erreur du classificateur « `randomForest` » est donnée ci-dessous.

	Prédit Y = 0 (résiliation)	Prédit Y = 1 (non résiliation)
Observé Y = 0	5 190	1 321
Observé Y = 1	696	20 818

Tableau 11 : Matrice d'erreur du classificateur « randomForest »

Source : Propres calculs à partir de données réelles

L'importance des différentes variables explicatives dans la réduction de la fonction d'« impureté » (c'est-à-dire l'indice de Gini dans cette étude) ainsi que l'influence du nombre d'arbres dans la forêt pour stabiliser l'erreur sont données ci-dessous.

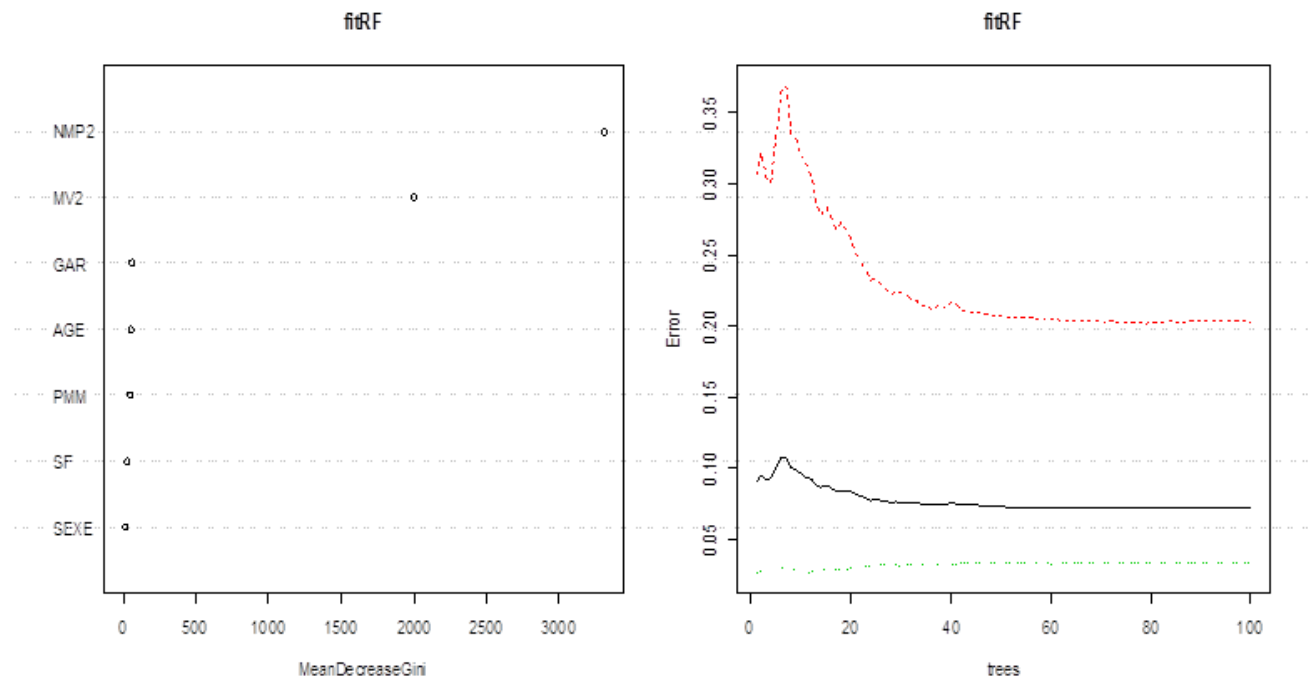


Figure 29 : Résultats de l'utilisation de la fonction « randomForest »

Source : Propres calculs à partir de données réelles

A gauche, la représentation de l'importance de chaque variable explicative dans la réduction de la fonction d'« impureté » permet de classer le pouvoir discriminant de chacune. Ces résultats confirment les premiers résultats : le nombre de primes payées (« NMP2 ») et le mois de vente (« MV2 ») sont les variables les plus discriminantes pour expliquer le comportement des assurés. Viennent ensuite le niveau de garantie (« GAR »), l'âge (« AGE »), la prime mensuelle (« PMM »), la situation familiale (« SF ») de l'assuré puis le sexe (« SEXE »). A droite, la ligne noire donne l'évolution de l'erreur globale en fonction du nombre d'arbres dans la forêt, la ligne verte l'erreur de la catégorie « NON » et la ligne rouge l'erreur de la catégorie « CHUTE ». L'erreur globale tend à se stabiliser rapidement (pour environ quarante arbres). Notons néanmoins une différence dans la stabilisation de l'erreur de la catégorie « NON » par rapport à la catégorie « CHUTE ». Pour la catégorie « CHUTE », l'erreur ne se stabilise que pour une soixantaine d'arbres tandis que pour la catégorie « NON », l'erreur est stable quasi-instantanément. Ce phénomène semble logique car il est plus facile d'identifier correctement un assuré restant en portefeuille plutôt qu'un assuré résiliant sa police. En effet, le second type d'assurés, moins représenté dans l'échantillon utilisé pour cette étude, nécessite alors plus d'arbres pour être repéré convenablement.

Dans cette étude sur le portefeuille XW, nous avons également souhaité tester la sensibilité de l'algorithme CART au retrait des polices avec aucune prime payée de l'échantillon. En effet, d'une part, ces polices sont d'un intérêt limité pour l'assureur car correspondant à des assurés n'ayant jamais « réellement » appartenu au portefeuille et, d'autre part, il est apparu que cette modalité était particulièrement discriminante dans la décision de l'assuré. Cette nouvelle étude pourrait donc nous permettre de faire émerger d'autres modalités importantes. Nous veillerons néanmoins à ne pas faire augmenter par ce biais l'erreur de prédiction.

Nous ne représentons pas dans le corps du mémoire les différentes sorties obtenus sous R⁴⁶. Toutefois, un certain nombre de remarques peuvent être faites sur les résultats observés. Tout d'abord, concernant l'arbre élagué obtenu, ce dernier est quasiment identique à celui de l'échantillon complet. En effet, bien que le nouveau modèle donne moins d'importance aux variables « MV2 » et « NMP2 » (d'après le graphique des résultats de l'utilisation de la fonction « *randomForest* » pour le nouveau modèle), celles-ci restent de façon assez nette les plus discriminantes pour expliquer les résiliations. Surtout, le taux d'erreur de prédiction pour des polices risquées augmente. Il passe ainsi de 20,3 % à 21,6 % en utilisant la fonction « *randomForest* » alors que l'erreur globale tend à diminuer (6,7 % contre 7,2 % auparavant). Le nouveau modèle prédit mieux les polices non risquées mais moins bien les polices risquées. Ce constat paraît cohérent puisque des enregistrements possédant une modalité très discriminante pour les résiliations ont été retirés. Enfin, l'erreur de prédiction semble se stabiliser à peu près suivant le même rythme pour les deux modèles, quel que soit le type d'erreur.

Les résultats ne sont donc pas ceux espérés car nous n'avons pas réussi à faire émerger de nouvelles variables discriminantes. Le problème mis en avant dans cette utilisation de l'algorithme CART est que les résiliations du portefeuille ne sont pas distribuées de façon homogène suivant les modalités des différentes variables. Certaines modalités des variables « MV2 » et « NMP2 » possèdent un très fort pouvoir discriminant sur la variable « CHUTE » et biaisent donc en partie l'étude en classant très rapidement les assurés dès les premières feuilles de l'arbre. Les feuilles inférieures ne procèdent alors plus qu'à des ajustements sur les groupes déjà constitués en amont.

⁴⁶ Voir Annexe 6 : Sorties obtenues avec le logiciel R pour l'algorithme CART dans l'étude du portefeuille XW sans la modalité « 0 » prime payée de la variable « NMP2 »

3.3.3. Portefeuille YZ

Dans cette sous-section, nous utilisons la même méthodologie que pour le portefeuille XW.

L'arbre T_{Max} obtenu à l'aide de la fonction $rpart()$ paramétrée par le vecteur $rpart.control()$ compte 126 feuilles.

	Prédit $Y = 0$ (résiliation)	Prédit $Y = 1$ (non résiliation)
Observé $Y = 0$	1 631	259
Observé $Y = 1$	339	3 436

Tableau 12 : Matrice d'erreur pour T_{Max} sur l'échantillon de validation
Source : Propres calculs à partir de données réelles

L'erreur de prédiction calculée pour l'arbre T_{Max} sur l'échantillon de validation est égale à 10,6 %. Nous obtenons à nouveau un bon prédicteur mais il reste préférable d'élaguer l'arbre afin d'obtenir des résultats facilement lisibles et interprétables.

Comme pour le portefeuille XW, la table des coûts de complexité est représentée.

CP	nsplit	rel error	xerror	xstd	CP	nsplit	rel error	xerror	xstd
2.7337e-01	0	1.00000	1.00000	0.0129996	1.7072e-04	43	0.29193	0.31549	0.0084925
4.9296e-02	2	0.45327	0.45327	0.0099086	1.2804e-04	67	0.28784	0.31626	0.0085016
2.5736e-02	4	0.35467	0.35467	0.0089370	8.5361e-05	75	0.28681	0.31780	0.0085198
7.6825e-04	6	0.30320	0.30320	0.0083449	6.4020e-05	87	0.28579	0.31831	0.0085258
3.1518e-04	7	0.30243	0.30499	0.0083667	5.1216e-05	103	0.28476	0.32036	0.0085498
2.5608e-04	21	0.29731	0.31370	0.0084713	3.2010e-05	118	0.28399	0.32138	0.0085618
2.0487e-04	30	0.29501	0.31626	0.0085016	0.0000e+00	126	0.28374	0.32138	0.0085618
1.9206e-04	39	0.29270	0.31472	0.0084834					

Tableau 13 : Table des paramètres de complexité
Source : Propres calculs à partir de données réelles

Le graphe du taux d'erreur en fonction du paramètre de complexité et de la taille de l'arbre est donné ci-dessous.

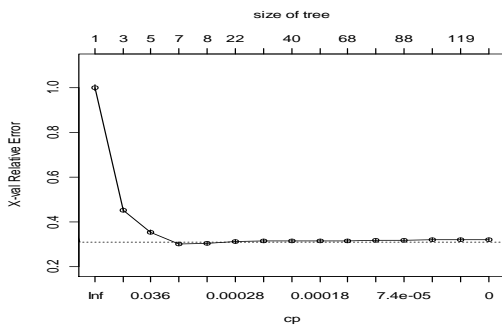


Figure 30 : Taux d'erreur de la validation croisée en fonction de la taille de l'arbre et du coût de complexité
Source : Propres calculs à partir de données réelles

L'arbre T_{Max} possède bien 127 divisions (la racine plus les 126 feuilles de l'arbre) et correspond à un coût de complexité nul. Comme pour le portefeuille XW, l'erreur de validation croisée chute rapidement avant de se stabiliser et de croître faiblement par la suite.

Un arbre possédant 6 feuilles et un coût de complexité égal à 7,6825e-04 permet de minimiser l'erreur de validation croisée. Cet arbre, présentant de bonnes propriétés et pouvant être facilement interprétable, est conservé.

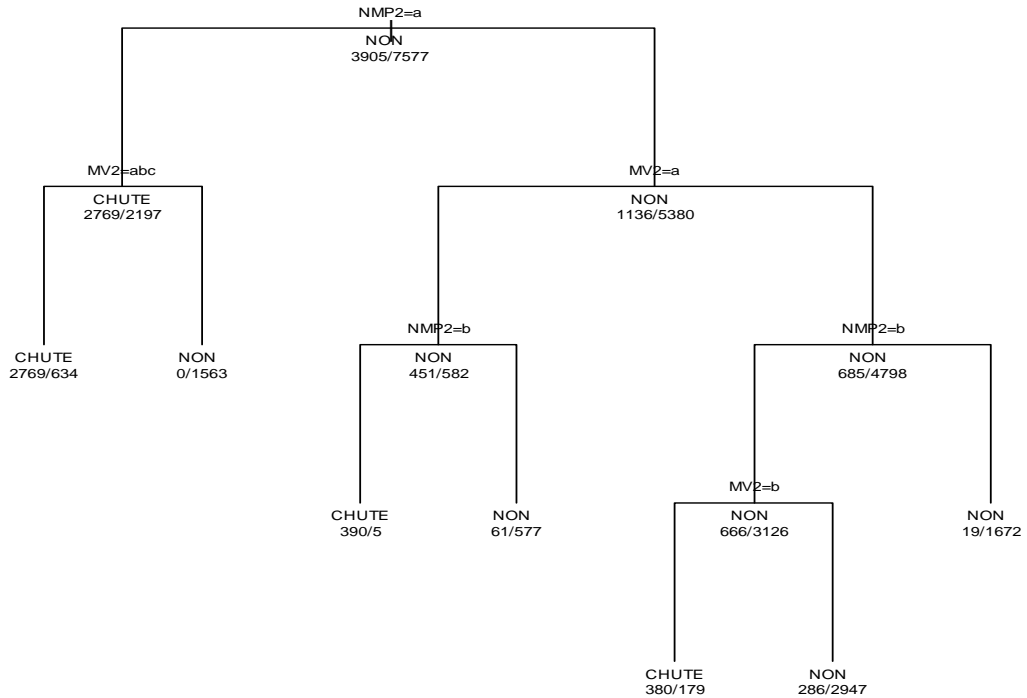


Figure 31 : Arbre de classification élagué

Source : Propres calculs à partir de données réelles

Le principe de fonctionnement de l'arbre obtenu est le même que pour le portefeuille XW. A chaque division une réponse est donnée à la question binaire suivante : l'assuré résilie-t-il son contrat ? Le nombre de gauche correspond à la réponse « OUI » et celui de droite à la réponse « NON ». La classe prédictive (libellée « CHUTE » ou « NON ») correspond à la réponse majoritaire à la question posée.

Le premier nœud correspond à l'ensemble du portefeuille avant toute segmentation. La probabilité de résiliation est de $\frac{3\,905}{11\,482} = 34,01\%$ et la classe prédictive est donc « NON ». Le portefeuille est alors divisé en deux selon le nombre de primes payées (« NMP2 ») par l'assuré. S'il possède la modalité a (les modalités prises par les variables explicatives étant classées par ordre alphabétique⁴⁷, cela correspond aux assurés ayant payés 0 prime), alors il passe dans la feuille de gauche, sinon il passe à droite. Ainsi, si la police possède cette modalité, la probabilité de résiliation est égale à $\frac{2\,759}{2\,759+2\,197} = 55,67\%$ et la classe prédictive est « CHUTE ». Nous retrouvons le phénomène de résiliation précoce des polices au cours du premier mois de couverture. En effet, de très nombreux assurés résilient leur contrat avant même d'avoir payé la moindre prime tandis que, passé ce premier mois, les assurés encore en portefeuille résilient dans une proportion plus raisonnable à hauteur de $\frac{1\,136}{1\,136+5\,380} = 17,43\%$. Ce phénomène, déjà existant pour le portefeuille XW, mais apparaissant de façon flagrante sur le portefeuille YZ, est assez symptomatique des contrats distribués en télémarketing, encore plus particulièrement de ce type de contrat où le premier mois de couverture est gratuit. Ainsi, les assurés résilient très souvent leur contrat après cette période gratuite.

Les deux variables discriminantes présentes dans l'arbre sont le nombre de primes payées (« NMP2 ») et le mois de vente (« MV2 »). L'âge (« AGE »), la prime mensuelle (« PMM »), la campagne de télémarketing (« CAMP »), le sexe (« SEXE »), le niveau de garantie (« GAR ») et la situation familiale (« SF ») de l'assuré

⁴⁷ Voir Annexe 5 : Table des correspondances pour l'algorithme CART sur le portefeuille YZ

principal ne figurent pas dans cet arbre. Ainsi, que cela soit à travers la variable « NMP2 » ou la variable « MV2 », la composante temporelle permet de classer les assurés de façon très rapide. Dans ce modèle, les autres variables explicatives ne sont plus que des variables d'ajustement qui n'interviennent que postérieurement aux deux variables apparaissant dans l'arbre élagué. A nouveau, il paraît raisonnable de penser que les résultats obtenus sont, dans une certaine mesure, biaisés du fait de l'importance exagérée des variables « NMP2 » et « MV2 » dans l'explication des résiliations. Concernant la deuxième variable, un autre phénomène apparaît lié à des problèmes de gestion au début de la période d'observation puisque, sans raison apparente, les résiliations sont nettement plus fréquentes au cours des premiers mois sur le portefeuille.

Donnons un second exemple de lecture des résultats de l'arbre élagué pour illustrer ce phénomène. Prenons une police dont les caractéristiques sont les suivantes : un mois de vente a (c'est-à-dire au cours de la période novembre 2009-janvier 2010) et un nombre de primes payées compris entre 1 et 3. Un tel assuré est dans une situation particulièrement risquée puisque $\frac{390}{390+5} = 98,73\%$ des individus avec ces caractéristiques ont résilié. Un tel chiffre est révélateur des problèmes posés par l'utilisation de l'algorithme CART avec l'échantillon de données utilisé dans cette étude : les autres variables explicatives dont nous souhaitons connaître l'influence sont totalement marginalisées par le pouvoir discriminant des variables « NMP2 » et « MV2 ».

Malgré les limites de l'utilisation de l'algorithme CART pour ces données évoquées dans les paragraphes précédents, l'arbre élagué permet de réaliser une bonne prédiction puisque seulement 10,0 % des prédictions sont erronées d'après la matrice d'erreur sur l'échantillon de validation représentée ci-dessous.

	Prédit Y = 0 (résiliation)	Prédit Y = 1 (non résiliation)
Observé Y = 0	1 719	171
Observé Y = 1	394	3 381

Tableau 14 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation

Source : Propres calculs à partir de données réelles

L'élagage de l'arbre est intéressant puisqu'il a permis d'obtenir un arbre plus facilement interprétable tout en diminuant de 0,6 % l'erreur de prédiction sur l'échantillon de validation. Quant à l'erreur en termes de polices risquées, classées à tort en polices non risquées, nous observons un taux d'erreur de 13,7 % avec l'arbre T_{Max} contre 9,0 % pour l'arbre élagué. Cette différence assez notable peut se comprendre par le fait que l'arbre élagué ne sélectionne que les variables les plus « pertinentes » pour expliquer les résiliations et ne s'égare pas dans les variables secondaires, moins discriminantes, pouvant créer des erreurs de classification plus fréquentes.

Analysons maintenant les résultats obtenus avec la fonction *randomForest()*. L'erreur de la forêt est de 10,3 % et l'erreur en termes de détection des polices à risques est de 10,5 %. Nous constatons des résultats proches de ceux trouvés pour T_{Max} et l'arbre élagué. La matrice d'erreur du classificateur « *randomForest* » est donnée ci-dessous.

	Prédit Y = 0 (résiliation)	Prédit Y = 1 (non résiliation)
Observé Y = 0	5 187	608
Observé Y = 1	1 160	10 192

Tableau 15 : Matrice d'erreur du classificateur « randomForest »

Source : Propres calculs à partir de données réelles

L'importance des différentes variables explicatives dans la réduction de la fonction d'« impureté » (l'indice de Gini dans cette étude) ainsi que l'influence du nombre d'arbres dans la forêt pour stabiliser l'erreur sont données ci-dessous.

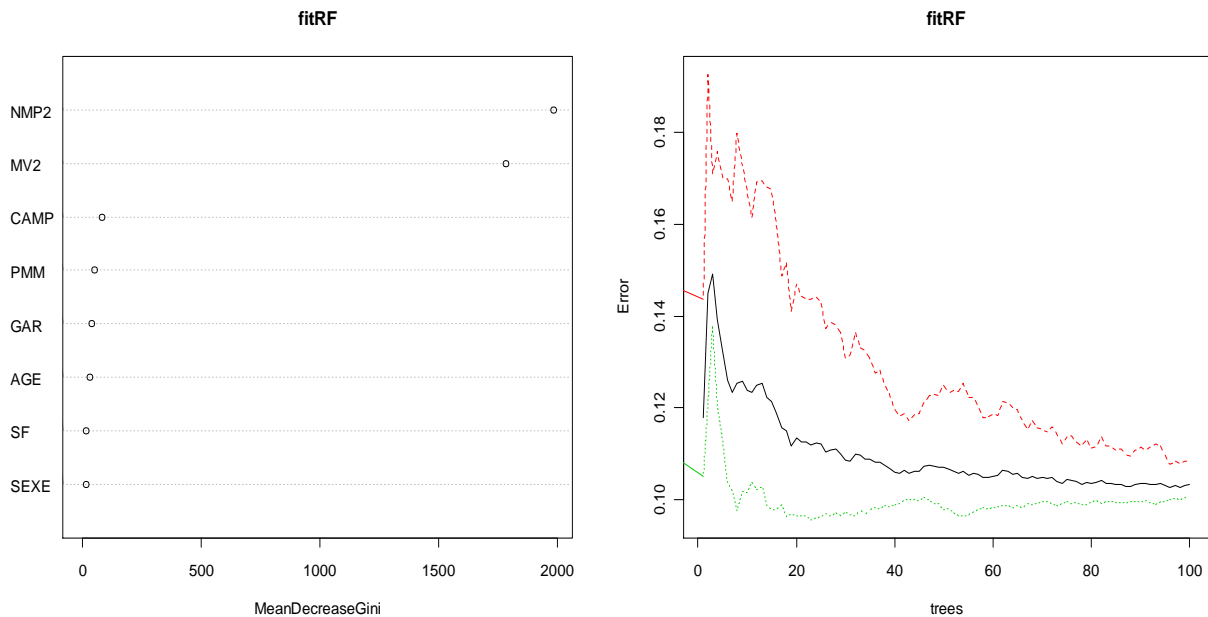


Figure 32 : Résultats de l'utilisation de la fonction « randomForest »

Source : Propres calculs à partir de données réelles

Le nombre de primes payées (« NMP2 ») et le mois de vente (« MV2 ») sont, sans surprise, les deux variables les plus importantes dans la réduction de la fonction d'« impureté ». Viennent ensuite la campagne de télémarketing (« CAMP »), la prime mensuelle (« PMM »), le niveau de garantie (« GAR »), l'âge (« AGE »), la situation familiale (« SF ») de l'assuré et le sexe (« SEXE »). A droite, la ligne noire donne l'évolution de l'erreur globale en fonction du nombre d'arbres dans la forêt, la ligne verte l'erreur de la catégorie « NON » et la ligne rouge l'erreur de la catégorie « CHUTE ». L'erreur tend à nouveau à se stabiliser pour environ quarante arbres. L'erreur globale et l'erreur sur les assurés qui ne résilient pas se stabilisent autour de 40 arbres dans la forêt tandis que l'erreur sur la catégorie « CHUTE » semble encore devoir baisser au-delà de 100 arbres dans la forêt. Par rapport au portefeuille XW, le niveau d'erreur est, de façon générale, plus volatil et plus long à se stabiliser. Ce constat, assez surprenant, nous a semblé difficile à expliquer de façon pertinente.

L'utilisation de l'algorithme CART a permis de faire ressortir l'importance de certaines variables dans la décision de l'assuré de résilier ou non son contrat. Le nombre de primes payées et le mois de vente de la police sont, de façon nette, les variables les plus discriminantes. Il est logique de trouver la première variable car hormis les premiers mois où sont constatées de nombreuses rétractions et les dates anniversaires de la police où l'assuré a la possibilité de résilier, plus un contrat reste en portefeuille, moins il a tendance à être résilié. L'influence du mois de vente peut néanmoins paraître discutable. Cette variable semble apparaître pour deux raisons. Tout d'abord, elle interagit avec la variable « nombre de primes payées » car, pour une police ayant été souscrite au cours des premiers mois de la période d'observation, un assuré avec très peu de primes payées a plus de chances de correspondre à une police résiliée qu'à une police souscrite récemment. Ainsi, le premier effet de la variable « mois de vente » est un effet corrélé au temps. Par ailleurs, il s'avère que pour les deux portefeuilles, les premiers mois d'observation correspondent à une période où la proportion de résiliation a été importante comme constaté au cours de l'étude uni-variée. Ce phénomène de « sur-

résiliation » au cours des premiers mois est dû à des problèmes de gestion dans la mise en place des portefeuilles qui ont été résolus au bout de quelques mois dans les deux cas.

Si les autres variables explicatives utilisées semblent être moins déterminantes, elles ne doivent pas être négligées car elles peuvent confirmer ou infirmer une tendance et interagir avec d'autres variables. A ce titre, les études uni-variées mettent en avant l'influence particulière de ces « variables oubliées » par l'algorithme CART. Il convient donc de nuancer les résultats obtenus et de rester prudent quant aux interprétations sur les facteurs de risque.

Néanmoins, malgré les limites évoquées plus haut, l'algorithme CART se révèle assez performant pour réaliser des modèles prédictifs, en particulier pour détecter les assurés à risque en fonction de leurs caractéristiques.

Si ce modèle permet de classer un assuré comme « risqué » au moment de la souscription ou après, la réponse prédite reste binaire (nous pouvons obtenir la probabilité de résiliation d'une classe donnée mais pas une probabilité individuelle de résiliation). En cela, l'algorithme CART se distingue des méthodes de modèles linéaires généralisés (MLG) que nous allons désormais étudier et qui permettent de donner directement la probabilité de résiliation d'un assuré en fonction de ses caractéristiques.

Une fois cette nouvelle étude réalisée, il conviendra de recouper les résultats obtenus par les deux méthodes afin de tenter de déterminer un profil de risque des individus en portefeuille selon le pays d'origine.

Chapitre 4. Le modèle de régression logistique

La régression logistique est un cas particulier des modèles linéaires généralisés et une technique statistique qui a pour objectif, à partir d'un fichier d'observations, de produire un modèle permettant de prédire les valeurs prises par une variable catégorielle, le plus souvent binaire, à partir d'une série de variables explicatives continues et/ou binaires.

Par rapport aux techniques connues en régression, notamment la régression linéaire, la régression logistique se distingue essentiellement par le fait que la variable expliquée est catégorielle. Elle est souvent utilisée par les actuaires pour modéliser des événements binaires tels que la mortalité ou la morbidité d'un portefeuille et leur permet ainsi de segmenter en fonction du niveau de risque. Dans le cadre de cette étude, l'objectif est de prédire la décision de l'assuré de résilier ou non son contrat d'assurance.

Dans ce chapitre, nous nous attachons à présenter les fondements théoriques de la régression logistique, les outils statistiques mis à disposition par le logiciel R puis les résultats obtenus en implémentant concrètement cette méthode sur les deux portefeuilles d'assurés afin de compléter les résultats obtenus grâce à l'algorithme CART.

Section 4.1. Présentation de la fonction logistique

L'intérêt de la fonction logistique est qu'elle permet, à partir d'une variable d'entrée pouvant varier sur l'espace des réels, d'obtenir une variable de sortie dont les valeurs sont comprises entre zéro et un. En effet, l'expression de la fonction logistique Φ est :

$$\Phi(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}$$

L'utilisation de cette fonction pour modéliser une probabilité semble donc tout à fait pertinente. En pratique, z représente une « exposition au risque » modélisée par l'équation de régression suivante :

$$z = \beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m$$

où $\forall i = 1 \dots m, \beta_i$ représente le coefficient de la régression associé au facteur de risque i (la situation familiale, le montant de la prime, etc.).

Par ailleurs, si nous posons $\Phi(z) = p$, la fonction inverse de la fonction logistique étant la fonction *logit* (c'est-à-dire $z = \Phi^{-1}(p) = \text{logit}(p)$), nous avons de façon équivalente à l'équation précédente :

$$\text{logit}(p) = \beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m$$

Cette écriture permet ainsi de se placer dans le cadre des modèles linéaires généralisés. Nous y reviendrons dans la suite de ce chapitre.

Section 4.2. La régression logistique

Revenons au cadre de ce mémoire. Nous cherchons à modéliser Y (la résiliation ou non de la police) en fonction des variables explicatives X_1, X_2, \dots, X_m . Nous rappelons le vocabulaire suivant : Y est appelée la variable de sortie, les $X_i, i = 1 \dots m$ sont les variables d'entrée. Par exemple, X_1 peut représenter le sexe de l'assuré, X_2 sa situation familiale, etc. Le statut de la police est fonction de ces quantités. Nous ne revenons pas en détail sur la formulation du problème⁴⁸ mais gardons à l'esprit que l'objectif de la modélisation est de trouver la fonction \bar{f} telle que :

$$\bar{f} = \underset{f \in \mathcal{M}(\mathbb{R}^m, \mathbb{R})}{\text{argmin}} E[(Y - f(X))^2]$$

Dans le cadre de la régression logistique nous cherchons \hat{f} , un estimateur de \bar{f} sous la forme d'un prédicteur linéaire à l'aide des modèles linéaires généralisés.

4.2.1. Définition d'un modèle linéaire généralisé

L'objectif de cette sous-section n'est pas de présenter en détail la théorie des modèles linéaires généralisés vue au cours de la formation ISFA mais de rappeler certaines définitions et résultats pour pouvoir mieux appréhender la suite de l'étude.

Les modèles linéaires généralisés, présentés de façon complète par P.McCullagh et J.A.Nelder (1989)⁴⁹, ont pour but de relier des variables explicatives regroupées dans un vecteur $X = (X_1, X_2, \dots, X_m)$ à une variable à expliquer Y à travers une fonction lien. En se plaçant dans le cas simplifié de la régression linéaire classique, le problème s'écrit de façon matricielle :

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X_{1,1} & \dots & X_{1,m} \\ \vdots & \ddots & \vdots \\ X_{N,1} & \dots & X_{N,m} \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

⁴⁸ Pour rappel se reporte à la sous-section Position du problème

⁴⁹ **McCullagh P., Nelder J. A. (1989)**, « *Generalized linear models, 2nd ed.* », Chapman and Hall

où :

- $Y = (Y_1, \dots, Y_N)'$ représente la décision de chacun des assurés de résilier ou non,
- la matrice $X = \begin{pmatrix} X_{1,1} & \cdots & X_{1,m} \\ \vdots & \ddots & \vdots \\ X_{N,1} & \cdots & X_{N,m} \end{pmatrix}$ est telle qu'une ligne représente les caractéristiques d'un assuré et une colonne les valeurs observées pour un facteur de risque.

Par rapport à ce modèle linéaire classique, il faut, schématiquement, faire deux choix pour construire un modèle linéaire généralisé. Le premier choix concerne la loi de la variable à expliquer qui doit appartenir à la famille exponentielle. Ce choix peut être orienté par le type de la variable à modéliser. Dans notre cas, la résiliation de la police est une variable binaire et semble logiquement devoir être modélisée par une loi de Bernoulli (qui appartient à la famille exponentielle). Le second choix porte sur la fonction lien. Typiquement, pour une variable binaire, le choix se porte de préférence sur les liens *logit*, *probit* ou *cloglog*. Nous ne présentons pas la forme de ces fonctions que le lecteur peut facilement trouver dans la littérature.

Nous connaissons désormais les éléments nécessaires à la mise en place d'un modèle linéaire généralisé. Revenons donc désormais au cas particulier du mémoire.

4.2.2. Introduction à la régression logistique

La régression logistique⁵⁰ est un cas particulier des modèles linéaires généralisés. Dans ce modèle, l'espérance conditionnelle de Y suivant X , $E[(Y|X_1, X_2, \dots, X_m)]$, peut être définie de deux façons :

$$(i) \quad \mu = E[(Y|X_1, X_2, \dots, X_m = x_1, x_2, \dots, x_m)] = E[(Y|X = x)] = \Phi(z)$$

ou

$$(ii) \quad \text{logit}(\mu) = \ln\left(\frac{\mu}{1-\mu}\right) = z$$

avec $z = \beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m$.

La deuxième écriture correspond à un modèle de régression linéaire à ceci près que la fonction identité a été remplacée par une fonction continue, dérivable et inversible : la fonction *logit* qui est donc la fonction lien.

Il reste cependant à définir la loi de Y . Cherchant à modéliser un événement binaire, la loi la plus naturelle est la loi binomiale $\beta(N, p)$ où N est le nombre d'expériences de Bernoulli réalisées et p est la probabilité de « succès » (c'est-à-dire ici de résiliation). Nous avons donc :

$$Y = \begin{cases} 1, & \text{si résiliation} \\ 0, & \text{sinon} \end{cases}$$

Ce qui donne, adapté à la régression logistique avec $p = \mu$:

$$(iii) \quad \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{\mathbb{P}(Y = 1|X = x)}{\mathbb{P}(Y = 0|X = x)}\right) = \beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m$$

⁵⁰ Balakrishnan N. (1991), « Handbook of the Logistic Distribution », Marcel Dekker, Inc.

Nous déduisons alors :

$$(iv) \quad \left. \begin{aligned} \Phi(\text{logit}(p)) &= \Phi(\Phi^{-1}(p)) = p \\ \Phi(\text{logit}(p)) &= \Phi(\beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m) \end{aligned} \right\} \Rightarrow p = \Phi(\beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m)$$

Un modèle linéaire généralisé d'une variable qualitative Y , ou, plus exactement, de la transformation par la fonction *logit* de l'espérance conditionnelle d'une variable qualitative $\text{logit}(E[(Y|X = x)])$, a donc été mis en place. Cette modélisation après transformation est linéaire en x .

Section 4.3. Aspects théoriques de la régression logistique

Nous allons présenter dans cette section les fondements théoriques de la régression logistique que nous utilisons ensuite dans l'étude. Pour ce faire, nous reprenons dans une large mesure, la partie théorique et les notations de l'article de X. Milhaud, S. Loisel et V. Maume-Deschamps intitulé « *Surrender triggers in life insurance : classification and risk predictions* » publié en janvier 2010⁵¹.

4.3.1. Détermination du maximum de vraisemblance et des coefficients de la régression

Soit N , le nombre d'observations (c'est-à-dire de polices en portefeuille). Par définition, la vraisemblance L d'une loi binomiale est de la forme suivante :

$$(v) \quad L(X, \beta) = \prod_{i=1}^N \phi(X_i \times \beta')^{Y_i} \times (1 - \phi(X_i \times \beta'))^{1-Y_i}$$

où $\beta' = (\beta_0, \beta_1, \dots, \beta_m)$ est la transposée du vecteur des coefficients de la régression.

La log-vraisemblance l est alors telle que :

$$(vi) \quad \begin{aligned} l(X, \beta) = \ln(L(X, \beta)) &= \sum_{i=1}^N Y_i \times \ln(\phi(X_i \times \beta')) + \sum_{i=1}^N (1 - Y_i) \times \ln(1 - \phi(X_i \times \beta')) = \\ &= \sum_{i=1}^N Y_i \times \ln\left(\frac{\exp(X_i \times \beta')}{1 + \exp(X_i \times \beta')}\right) + \sum_{i=1}^N (1 - Y_i) \times \ln\left(1 - \left(\frac{\exp(X_i \times \beta')}{1 + \exp(X_i \times \beta')}\right)\right) = \\ &= \sum_{i=1}^N Y_i \times X_i \times \beta' - \ln(1 + \exp(X_i \times \beta')) \end{aligned}$$

Afin d'obtenir la valeur de β qui maximise la vraisemblance (ce qui est équivalent à maximiser la log-vraisemblance), il faut trouver $\hat{\beta}$ tel que :

$$(vii) \quad \left. \frac{\partial l}{\partial \beta} \right|_{\hat{\beta}} = 0$$

Ceci conduit à résoudre un système d'équations ouvert. La façon la plus courante de résoudre ce type de système est d'utiliser l'algorithme de Newton-Raphson⁵²

⁵¹ Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance : classification and risk predictions* », Hal, p8-11

⁵² Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance : classification and risk predictions* », Hal, p25-27

4.3.2. Estimation d'un intervalle de confiance

Une fois $\hat{\beta}$ déterminé, il est possible de calculer, pour chaque individu, une estimation de la probabilité de résiliation. En effet d'après (iv):

$$\hat{p} = \Phi(\widehat{\beta}_0 + \widehat{\beta}_1 \times X_1 + \dots + \widehat{\beta}_m \times X_m)$$

où $\widehat{\beta}_i, 1 \leq i \leq m$ sont les coefficients de la régression logistique estimés par maximum de vraisemblance.

Pour chaque assuré, une estimation de la probabilité de résiliation selon ses caractéristiques est obtenue.

L'étape suivante est de déterminer des intervalles de confiance pour cette probabilité sur l'ensemble du portefeuille en utilisant l'approximation de la loi binomiale, c'est-à-dire en considérant que le nombre de résiliations parmi N polices suit une loi normale. Pour cela, il convient de s'assurer que :

- Il y a homogénéité du portefeuille quant à la probabilité p de résiliation de la police,
- $N \rightarrow \infty$, c'est-à-dire la taille du portefeuille est suffisante (ce qui est le cas sur un portefeuille de polices, de taille importante par nature).

Cette approximation est une conséquence directe du théorème central limite. En effet, le nombre de résiliation suit une loi binomiale qui est une somme de loi de Bernoulli. Par conséquent, si ces lois de Bernoulli sont indépendantes et identiquement distribuées et que la deuxième hypothèse est vérifiée, le théorème central limite peut s'appliquer et l'hypothèse de normalité du nombre de résiliations est pertinente. Toutefois, la première hypothèse pose problème car un portefeuille d'assurés est hétérogène par nature. Imaginons alors que le portefeuille est divisé en groupes homogènes d'assurés, alors chacun des groupes est distribué selon une loi normale. Or, une somme de lois normales est encore une loi normale. Ainsi, il est possible d'utiliser malgré tout l'approximation normale.

Déterminons maintenant la forme de l'intervalle de confiance de la probabilité de résiliation sur l'ensemble du portefeuille en utilisant l'approximation normale. Considérons que la décision de l'assuré suit une loi de Bernoulli, alors le nombre de résiliations N_i^S dans un groupe homogène i comprenant n_i assurés est indépendamment et identiquement distribué selon une loi binomiale $\beta(n_i, p_i)$. Ainsi, nous pouvons écrire :

$$\begin{aligned} \mathbb{E}[N_i^S] &= \sum_{i=1}^{n_i} p_i = n_i \times p_i \\ \text{Var}[N_i^S] &= \sum_{i=1}^{n_i} p_i \times (1 - p_i) = n_i \times p_i \times (1 - p_i) \\ \sigma[N_i^S] &= \sqrt{\text{Var}[N_i^S]} = \sqrt{n_i \times p_i \times (1 - p_i)} \end{aligned}$$

Notons $\hat{p}_i = \frac{N_i^S}{n_i}$, le taux de résiliation du groupe i , nous avons alors :

$$\begin{aligned} (viii) \quad \mathbb{E}[\hat{p}_i] &= \frac{\mathbb{E}[N_i^S]}{n_i} = p_i \\ (ix) \quad \sigma[\hat{p}_i] &= \frac{\sigma[N_i^S]}{n_i} = \sqrt{\frac{p_i \times (1 - p_i)}{n_i}} \end{aligned}$$

Des équations (viii) et (ix), se déduit l'intervalle de confiance d'une loi normale pour un groupe homogène i . Or, la somme de lois normales indépendantes est toujours une loi normale et le nombre total de résiliations est la somme des résiliations des différents groupes homogènes (nous posons $N^S = \sum_i N_i^S$). Il est

donc juste d'étendre ces résultats à l'ensemble du portefeuille et d'obtenir l'intervalle de confiance à 95 % de $\hat{p} = \frac{N^s}{N}$:

$$IC_{95\%}(p) = \left[\hat{p} - 1,96 \times \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{N}}, \hat{p} + 1,96 \times \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{N}} \right]$$

4.3.3. Tests de significativité

Cette sous-section étudie les différents moyens de juger de la qualité du modèle construit et de la pertinence de chacune des variables explicatives à l'intérieur de celui-ci. Pour rappel, nous nous plaçons dans le modèle spécifique de régression logistique où Y suit une loi binomiale et où la fonction lien est la fonction logistique. Les définitions présentées par la suite ne sont donc pas générales mais s'appliquent à ce cas précis.

4.3.3.1. Evaluation statistique de la régression logistique

Afin de valider la pertinence du modèle, le test du ratio de log-vraisemblance peut être utilisé.

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_M = 0 \\ H_1 : \exists i \in 1 \dots m \mid \beta_i \neq 0 \end{cases}$$

Notons $l(\beta)$, la log-vraisemblance de la régression logistique à $M + 1$ coefficients et $l(\beta_0)$ celle du modèle de régression logistique le plus simple possible (c'est-à-dire réduit à la constante β_0). La statistique du ratio de log-vraisemblance est :

$$\Lambda = 2 \times (l(\beta) - l(\beta_0))$$

Cette statistique suit une loi du χ^2 à M degrés de liberté. Si la p -value du test est inférieure au seuil fixé à 5 %, H_0 est rejetée et le modèle est pertinent.

Le logiciel R utilise la déviance D du modèle plutôt que la statistique Λ pour valider la pertinence du modèle. L'expression de la déviance est :

$$D = 2 \times (l(\beta_{saturé}) - l(\beta))$$

où $l(\beta_{saturé})$ est la log-vraisemblance du modèle saturé, c'est-à-dire le modèle basé sur la même loi de probabilité et la même fonction lien mais contenant autant de variables explicatives indépendantes que de données (ce modèle permet donc de reproduire parfaitement les données).

Lorsqu'un modèle est adéquat, il paraît logique que $l(\beta)$ soit proche du modèle saturé, et donc que la déviance soit proche de 0. Inversement, plus la déviance d'un modèle est importante, moins le modèle en question est pertinent. Cette notion est exploitée afin de tester la significativité d'un modèle. L'hypothèse nulle de test est donc :

$$H_0 : \text{« Le modèle à } p \text{ variables explicatives est significatif »}$$

Sous H_0 , la déviance suit asymptotiquement une loi du χ^2 à $n - p$ degrés de liberté (avec n le nombre d'observations distinctes et p le nombre de variables du modèle). Ainsi pour que l'hypothèse nulle soit retenue au seuil de 5 %, il faut vérifier :

$$D < \chi_{n-p}^2(5\%)$$

où $\chi_{n-p}^2(5\%)$ est le quantile à 95 % d'une loi du χ^2 à $n - p$ degrés de liberté.

Or, pour une telle loi, nous avons toujours $n - p < \chi_{n-p}^2(5\%)$, donc une condition suffisante pour accepter l'hypothèse nulle est :

$$D < n - p$$

Un exemple concret de ce test de significativité du modèle est donné lorsque nous étudions les données des portefeuilles d'étude.

4.3.3.2. Pertinence d'une variable aléatoire donnée

Ce test consiste à comparer la valeur du coefficient estimé β_j qui est associé à la variable explicative X_j à sa variance.

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

La statistique du test est la statistique de Wald qui suit une loi du χ^2 à un degré de liberté :

$$\frac{\widehat{\beta}_j^2}{\text{Var}(\widehat{\beta}_j)}$$

Si la p -value du test est inférieure au seuil fixé à 5 %, H_0 est rejetée et la variable explicative est pertinente pour le modèle.

4.3.4. Choix du modèle adéquat

Pour choisir le meilleur modèle à retenir pour nos données, il faut décider quelles variables explicatives retenir. Pour prendre cette décision, plusieurs indicateurs existent dont les plus connus sont les critères AIC (acronyme de *Akaike Informative Criterion*) et BIC (acronyme de *Bayesian Informative Criterion*).

L'idée de ces critères repose sur le principe que plus la vraisemblance (ou de manière équivalente la log-vraisemblance) est grande, meilleur est le modèle. De plus, afin de faire en sorte que le rajout de variables explicatives ne signifie pas nécessairement une augmentation du critère, et d'obtenir un modèle de taille raisonnable, ces critères contiennent un facteur pénalisant fonction du nombre de paramètres p . Les expressions de ces deux indicateurs sont les suivantes :

$$AIC = -2 \times L + 2 \times p$$

$$BIC = -2 \times L + \log(N) \times p \text{ où } N \text{ est le nombre d'observations}$$

Le logiciel R utilise par défaut le critère AIC dans sa procédure de sélection du modèle optimal. Nous choisissons donc le modèle minimisant le critère AIC par la suite.

Section 4.4. Interprétation du modèle obtenu

Les coefficients de la régression fournissent des informations sur l'effet de chacun des facteurs de risque. Le coefficient β_0 (nommé *intercept*) est la valeur de z pour le profil de référence lorsque les différentes variables explicatives prennent leur modalité de référence respective. Dès lors, les coefficients β_i décrivent l'influence des modalités de chacune des variables explicatives par rapport à la modalité de référence : si β_i est positif (respectivement négatif), alors posséder la modalité correspondante fait augmenter (respectivement diminuer) la probabilité de résiliation par rapport à une police possédant la modalité de référence.

Par ailleurs, le rapport $\frac{\beta_i}{\sigma(\beta_i)}$ (où $\sigma(\beta_i)$ est l'écart-type de l'estimateur du coefficient) renseigne sur l'influence du risque i quant à la probabilité de survenance de résiliation de la police. En effet, si ce rapport est grand, alors le risque i a une influence importante.

Par la suite, nous étudierons les *odd-ratios* du modèle de régression logistique obtenu avec nos données. Par définition, l'indicateur *odd-ratio* est égal à $\frac{p}{1-p}$, c'est-à-dire au rapport de la probabilité de succès (c'est-à-dire, dans cette étude, la probabilité de résiliation de la police) sur la probabilité d'échec.

Pour mieux appréhender cette notion, prenons un exemple concret. Soit $p = \mathbb{P}(Y = 1|X) = 0,75$, la probabilité de succès d'un évènement sachant X (par complémentarité $1 - p = \mathbb{P}(Y = 0|X) = 0,25$). Les chances de succès sont définies par le *odd-ratio* et sont égales à $\frac{0,75}{0,25} = 3$. En d'autres termes, cela signifie que pour des caractéristiques similaires (le vecteur X), le succès a trois fois plus de chances de survenir que l'échec. Réciproquement, les chances d'échec sont égales à 0,33.

Or, dans le cas de la régression logistique, d'après (ii) :

$$\begin{aligned} \frac{p}{1-p} &= \exp(\text{logit}(p)) = \exp(\beta_0 + \beta_1 \times X_1 + \dots + \beta_m \times X_m) \\ &= \exp(\beta_0) \times \exp(\beta_1 \times X_1) \times \dots \times \exp(\beta_m \times X_m) \end{aligned}$$

Ainsi, pour une variable explicative binaire (ce qui est le cas des variables catégorielles utilisées qui valent 1 si l'assuré possède la modalité et 0 sinon) nous avons :

$$\begin{cases} \text{si } x_i = 0 \text{ alors } \exp(0) = 1 \text{ et le terme disparaît} \\ \text{si } x_i = 1 \text{ alors } \exp(\beta_i \times x_i) = \exp(\beta_i) \end{cases}$$

Puisque toutes les variables x_i sont binaires dans le modèle, si un assuré possède la modalité d'une variable (c'est-à-dire si $x_i = 1$), alors la probabilité de résiliation change : par exemple, si $x_1 = 1$ et $x_k = 0 \forall k > 1$, alors le *odd-ratio* est égal à $\exp(\beta_1)$.

De façon synthétique, les *odd-ratios* représentent la différence en termes de probabilité pour l'évènement modélisé lorsque la modalité de la variable explicative étudiée est changée. Des exemples concrets sont donnés lors de la présentation des résultats de l'étude.

La régression logistique semble être un outil très performant pour modéliser une variable de sortie connaissant les caractéristiques des variables d'entrée. Outre la possibilité de pouvoir comparer le niveau de risque des assurés en portefeuille, le principal avantage de cette méthode est de pouvoir également faire de la prédiction. Toutefois, la régression logistique présente aussi un inconvénient car son utilisation repose sur l'hypothèse de l'indépendance des variables explicatives pour pouvoir avoir l'indépendance des variables ($Y_i|X_i$). En effet, en pratique, les variables explicatives ne sont pas indépendantes. Il est cependant possible d'implémenter le modèle en considérant que les modalités d'une variable catégorielle sont indépendantes, ce qui est généralement le cas. Une autre limite possible pour la régression logistique est que celle-ci réclame une quantité importante de données pour être fiable. Néanmoins, ce dernier problème n'en est pas réellement un dans le cadre de cette étude sur des portefeuilles d'assurance qui contiennent par définition suffisamment de données.

Section 4.5. Application de la régression logistique aux deux portefeuilles

L'objectif de cette section est d'expliquer puis de modéliser le taux de résiliation des contrats d'hospitalisation distribués en télémarketing en fonction des caractéristiques de la police et de l'assuré.

4.5.1. Les données

Les fichiers de données sont les mêmes que ceux utilisés pour la modélisation à l'aide de l'algorithme CART. Ainsi, nous avons à nouveau un échantillon d'apprentissage et un échantillon de validation pour chacun des portefeuilles. De plus, comme précédemment, le modèle (c'est-à-dire les coefficients de la régression logistique dans ce cas) est ajusté à l'aide de l'échantillon d'apprentissage tandis que les prédictions sont effectuées sur l'échantillon de validation.

Les échantillons d'apprentissage sont composés de 18 733 polices pour XW et de 11 482 polices pour YZ tandis que les échantillons de validation contiennent 9 292 polices pour XW et 5 665 pour YZ.

Les variables explicatives entrant dans le cadre de l'étude sont les mêmes que dans l'application de l'algorithme CART à savoir, l'âge, le sexe, la situation familiale de l'assuré principal, le niveau de garantie de la police, le montant de prime mensuelle moyenne, le mois de vente de la police et le nombre de primes payées (et également la campagne de télémarketing pour le portefeuille YZ). Le tableau ci-dessous donne une description plus complète de celles-ci.

Code Variable	Description
STAT	Statut de la police : « CHUTE » ou « NON »
AGE	Tranche d'âge de l'assuré principal sur la police : « 18-34 », « 35-44 », « 45-54 », « 55-65 »
SEXE	Sexe de l'assuré principal : Homme (« H ») ou Femme (« F »)
SF	Situation familiale de l'assuré principal : « CELIBATAIRE » ou « FAMILLE »
GAR	Niveau de garantie de la police : « Faible », « Moyenne » ou « Elevée »
MV	Mois de vente de la police
NMP	Nombre de primes payées : compris entre 0 et 23(XW) ou 0 et 11 (YZ)
PMM	Prime mensuelle moyenne : « <15 € », « 15 €-30 € » ou « >30 € »
CAMP	Campagne de télémarketing : « C », « D », « E »

Tableau 16 : Description des principales variables explicatives pour la régression logistique

Les regroupements effectués pour les variables « MV » et « NMP » sont identiques pour l'application de la régression logistique et de l'algorithme CART. Les variables issues des regroupements sont à nouveau appelées « MV2 » et « NMP2 ».

4.5.2. Implémentation de la régression logistique avec le logiciel R

Le package utilisé dans le cadre de cette étude est le package « MASS »⁵³ implémenté sous R par B.Ripley.

⁵³Documentation disponible sur <http://cran.r-project.org/web/packages/MASS/MASS.pdf>

Comme pour l'algorithme CART, les variables explicatives continues ont été initialement transformées en variables catégorielles pour l'étude.

Pour réaliser une régression logistique sous R, il faut utiliser la fonction *glm()*. Le modèle permet de visualiser en sortie l'effet de chacune des modalités des variables par rapport à une modalité de référence ainsi que les intervalles de confiance des estimateurs et la déviance du modèle. Ainsi, pour réaliser une modélisation logistique avec toutes les variables explicatives, il faut lancer le code suivant :

```
RL1=glm  
(STAT~AGE+SEXE+SF+GAR+NMP+CAMP+MV+PMM,family=binomial(link="logit"),data=app)
```

Comme précisé dans la partie théorique, la variable de sortie est la probabilité de résiliation de la police modélisée par une loi binomiale avec une fonction lien *logit*.

Nous ne souhaitons cependant garder que les variables les plus pertinentes dans le modèle. Pour ce faire, nous utilisons une régression logistique hiérarchique dont le principe est de comparer un modèle contenant m_0 variables explicatives sur les m du modèle original à tous les « sous-modèles » (c'est-à-dire contenant au moins une variable explicative en moins) et à tous les « sur-modèles » (c'est-à-dire contenant au moins une variable explicative en plus) possibles. Pour cela, la fonction *stepAIC()*, qui permet d'enlever ou d'ajouter des variables explicatives au modèle, est utilisée. Nous obtenons finalement le meilleur modèle avec le minimum de variables pertinentes. Le code correspondant est simplement *RL2= stepAIC(RL1)*

Les autres fonctions utiles sont :

- *summary()* qui permet de contrôler la pertinence du modèle et des variables explicatives,
- *drop1()* qui est utilisée pour contrôler la significativité des variables explicatives en testant le modèle complet contre le modèle sans une variable.

Des exemples d'utilisation de ces fonctions ainsi que d'interprétation des résultats obtenus sont donnés lors de l'analyse concrète des résiliations sur les deux portefeuilles.

4.5.3. *Portefeuille XW*

Après avoir réalisé la modélisation avec toutes les variables explicatives (modèle *RL1*), nous avons gardé les variables les plus pertinentes à l'aide de la fonction *stepAIC()* pour obtenir le modèle *RL2*.

Afin de mieux comprendre cette procédure, analysons les différentes sorties fournies par le logiciel R.

```

> step <- stepAIC(glm, trace = TRUE)
Start:  AIC=8642.98
STATUT ~ AGE + SEXE + SF + GAR + MV2 + NMP2 + PMM

      Df Deviance    AIC
- SEXE 1   8609.2 8641.2 }
<none>      8609.0 8643.0 } Baisse de l'AIC
- AGE  3   8616.6 8644.6 }
- SF   1   8614.7 8646.7 } AIC du modèle testé
- PMM  2   8618.1 8648.1 }
- GAR  2   8637.6 8667.6 } Hausse de l'AIC
- MV2  3 16540.2 16568.2 }
- NMP2 4 18579.2 18605.2 }
Step:  AIC=8641.18
STATUT ~ AGE + SF + GAR + MV2 + NMP2 + PMM

      Df Deviance    AIC
<none>      8609.2 8641.2
- AGE  3   8616.8 8642.8
- SF   1   8615.0 8645.0
- PMM  2   8618.3 8646.3
- GAR  2   8637.7 8665.7
- MV2  3 16541.6 16567.6
- NMP2 4 18580.5 18604.5
    
```

Figure 33 : Principe de la fonction « *stepAIC()* »

Comme son nom l'indique, la fonction « *stepAIC()* » procède par étape. La première étape consiste à calculer l'AIC du modèle de référence (ici RL1), ainsi que l'impact du retrait de chaque variable sur l'AIC. Nous enlevons ainsi la variable dont le retrait permet une baisse maximale de l'AIC jusqu'à ce que le retrait d'une des variables restantes soit synonyme d'augmentation de l'AIC. Ici la variable « SEXE » est la seule à être retirée pour obtenir le modèle RL2, optimal en termes de réduction de l'AIC. Notons que les variables « MV2 » et « NMP2 » semblent avoir une grande importance dans le modèle car leur retrait entrainerait une forte hausse de l'AIC.

Une fois le critère AIC optimisé, il convient de contrôler la pertinence du modèle et des variables explicatives à l'aide des fonctions « *summary()* » et « *drop1()* ». Les sorties R sont données ci-dessous avec, pour chaque fonction, une explication des résultats obtenus.

```

> summary(step)

Call:
glm(formula = STATUT ~ AGE + SF + GAR + MV2 + NMP2 + PMM, family = binomial(link = "logit"),
    data = app)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6547  -0.4170  -0.2695  -0.0599   3.6151

Coefficients:
(Intercept)      7.76127    0.20618  37.642 < 2e-16 ***
AGE35-44         0.12356    0.12598   0.981 0.326667
AGE45-54         0.06068    0.10880   0.558 0.577046
AGE55-65         0.24275    0.11971   2.028 0.042578 *
SFFAMILLE       0.21270    0.08790   2.420 0.015535 *
GAR40-60        0.42379    0.08311   5.099 3.41e-07 ***
GAR75-100       0.77134    0.21843   3.531 0.000414 ***
MV2200906-200911 -1.77638    0.12959 -13.707 < 2e-16 ***
MV2200912-201005 -5.99128    0.14862 -40.313 < 2e-16 ***
MV2201005-201010 -9.01530    0.17097 -52.731 < 2e-16 ***
NMP21-3         -1.14469    0.10393 -11.014 < 2e-16 ***
NMP211          -5.31997    0.18051 -29.472 < 2e-16 ***
NMP211-23      -10.33542    0.18655 -55.404 < 2e-16 ***
NMP24-10       -5.12827    0.13135 -39.044 < 2e-16 ***
PMM>30         -0.42395    0.14064  -3.014 0.002575 **
PMM15-30       -0.20634    0.09040  -2.282 0.022466 *

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 20280.1 on 18732 degrees of freedom
Residual deviance: 8609.2 on 18717 degrees of freedom
AIC: 8641.2

Number of Fisher Scoring iterations: 6
    
```

Figure 34 : Eléments de la fonction « *summary()* » d'une régression logistique

En ce qui concerne le contrôle de la légitimité du modèle, il convient de s'intéresser à la déviance et de la comparer au nombre de degrés de liberté. Nous trouvons la déviance du modèle (*Residual deviance*) de même que le nombre de degrés de liberté sur la deuxième ligne au niveau du cercle A. Ici la déviance est égale à 8 609,2 ce qui est inférieur au nombre de degrés de liberté. Nous pouvons donc admettre que le modèle est pertinent (en effet, le modèle est acceptable si la déviance est inférieure au nombre de degrés de liberté), sans prendre le soin de réaliser un test de déviance puisque le quantile du χ^2 est automatiquement supérieur à ce nombre.

La fonction « summary » fournit aussi le critère AIC (cercle B) ainsi que les coefficients et écarts types estimés des paramètres associés à chaque occurrence des prédicteurs (cercle C). Les deux colonnes suivantes (cercle D) sont issues d'un test de Student visant à tester l'hypothèse de nullité d'un coefficient. Lorsque la *p - value* est faible, le coefficient associé est significativement non nul et un certain nombre d'astérisques apparaissent. Ici, la plupart des coefficients de la régression semblent significatifs à l'exception des tranches d'âges « 35-44 » et « 45-54 ». Ces modalités ne donneront pas lieu à des interprétations dans la suite de l'étude. Cette information ne permet cependant pas de juger la significativité d'une variable car le test mis en œuvre porte sur les occurrences de la variable et non sur la variable elle-même. Par ailleurs, toutes les occurrences des variables ne sont pas apparentes. En effet, pour chaque variable, une occurrence correspond à une modalité de référence (la première dans l'ordre alphabétique) et l'ensemble forme l'individu de référence représenté par l'*intercept*. Ici l'individu de référence est donc un célibataire âgé de 18 à 34 ans avec un niveau de garantie entre 15 et 25 €, ayant souscrit son contrat en début de période d'observation, ayant payé 0 prime et dont le niveau de prime mensuelle est inférieur à 15 €.

En ce qui concerne l'interprétation des coefficients, la fonction lien étant la fonction logistique, un coefficient négatif indique un risque plus faible que pour l'individu de référence et, inversement, un coefficient positif indique un risque plus fort. Ainsi, à titre d'exemple, les assurés ayant payé 0 prime sont les assurés les plus risqués par rapport à cette variable. Par ailleurs, beaucoup de coefficients étant proches de 0, en particulier pour les variables « AGE », « SF », « GAR » et « PMM », ceci signifie que, dans le modèle, ces variables ont peu d'importance dans l'explication des résiliations. Nous retrouvons donc la prépondérance des variables « MV2 » et « NMP2 » que nous avons déjà mis en avant en utilisant l'algorithme CART. Enfin, l'individu de référence a un niveau de risque élevé puisque l'estimation du coefficient β_0 est égale à 7,76127.

Afin de juger de la significativité des variables et non plus seulement des occurrences, il est possible d'utiliser la fonction « *drop1()* »

```
> drop1(step, test="Chisq")
Single term deletions ←

Model:
STATUT ~ AGE + SF + GAR + MV2 + NMP2 + PMM
Df Deviance   AIC    LRT Pr(Chi)
<none>      8609.2 8641.2
AGE      3   8616.8 8642.8   7.6 0.05423 .
SF       1   8615.0 8645.0   5.8 0.01565 *
GAR      2   8637.7 8665.7  28.5 6.5e-07 ***
MV2      3 16541.6 16567.6 7932.5 < 2e-16 ***
NMP2     4 18580.5 18604.5 9971.4 < 2e-16 ***
PMM      2   8618.3 8646.3   9.1 0.01047 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 35 : Résultats issus de la fonction « *drop1()* »

La fonction « *drop1()* » permet de réaliser un test de significativité des variables explicatives. Selon la même logique que précédemment, une *p - value* (dernière colonne) faible est synonyme de la significativité

de la variable. Il est à noter que la fonction « *drop1()* » teste le modèle complet contre le modèle avec une variable en moins. Enfin, pour les modélisations de données binaires ou de comptage, l'option « *test=Chisq* » est retenue. Toutes les variables sont ici significatives, bien que la significativité soit moins prononcée pour les variables « AGE », « SF » et « PMM ».

Au vu des différents contrôles de pertinence, le modèle *RL2* semble convenir aux données du portefeuille XW. Essayons maintenant d'interpréter ce modèle à l'aide de la technique des *odd-ratios*. Le tableau représenté ci-dessous fait la synthèse des *odd-ratios* obtenus de façon empirique (c'est-à-dire en utilisant les statistiques descriptives sur les données de l'échantillon d'apprentissage) et grâce au modèle *RL2*.

Odd-ratio(OR)	Référence	Autres Modalités			
« AGE »	18-34	35-44	45-54	55-65	
<i>Nombre de résiliations</i>	627	506	1673	1534	
<i>OR empiriques</i>		0,69	0,61	0,58	
<i>OR modélisés</i>		1,13	1,06	1,27	
« SF »	CELIBATAIRE	FAMILLE			
<i>Nombre de résiliations</i>	2586	1754			
<i>OR empiriques</i>		1,17			
<i>OR modélisés</i>		1,24			
« GAR »	15-25	40-60	75-100		
<i>Nombre de résiliations</i>	2058	2147	135		
<i>OR empiriques</i>		1,40	2,28		
<i>OR modélisés</i>		1,53	2,16		
« MV2 »	200812-200905	200906-200911	200912-201005	201006-201010	
<i>Nombre de résiliations</i>	1727	1067	1158	386	
<i>OR empiriques</i>		0,60	0,32	0,14	
<i>OR modélisés</i>		0,17	0,0025	≅0	
« NMP2 »	0	1 → 3	4 → 10	11	12 → 23
<i>Nombre de résiliations</i>	914	1801	1025	319	281
<i>OR empiriques</i>		0,32	0,075	0,24	0,025
<i>OR modélisés</i>		0,32	0,0059	0,0049	≅0
« PMM »	<15	>30	15-30		
<i>Nombre de résiliations</i>	1246	1008	2086		
<i>OR empiriques</i>		1,47	1,02		
<i>OR modélisés</i>		0,65	0,81		

Tableau 17 : Synthèse des *odd-ratios* obtenus de façon empirique et en utilisant le modèle *RL2*

Donnons une illustration du calcul des *odd-ratios* pour chacune des techniques évoquées. Considérons par exemple un assuré possédant la modalité « FAMILLE » pour la variable « SF ». La technique des *odd-ratios* permet de comparer les chances de résiliation de cet assuré par rapport à un assuré avec la modalité de référence (c'est-à-dire « CELIBATAIRE ») toutes choses égales par ailleurs.

Pour faire un calcul empirique, il faut d'abord représenter la table qui donne le nombre d'assurés ayant résilié ou non pour les deux modalités qui nous intéressent :

	« CELIBATAIRE »	« FAMILLE »
« NON »	9 114	5 279
« CHUTE »	2 586	1754

Tableau 18 : Nombre de polices en fonction du statut de la police et de la tranche d'âge

Source : Propres calculs à partir de données réelles

Le *odd-ratio* empirique est alors égal à $\frac{9\,114 \times 1\,754}{5\,279 \times 2\,586} \cong 1,17$.

La valeur de l'*odd-ratio* modélisé est égal à $\exp(\beta_{FAMILLE}) = \exp(0,2127) \cong 1,24$. Le coefficient de la modalité de référence étant égal à 0, il convient de comparer ces deux valeurs à 1 qui est la valeur de l'*odd-ratio* pour la modalité de référence. Ainsi, un *odd-ratio* plus grand que 1 signifie un risque de résiliation plus important que la modalité de référence alors qu'inversement un *odd-ratio* plus petit que 1 signifie un risque moindre.

Nous nous apercevons que, pour les variables « SF », « GAR », « MV2 » et « NMP2 », les *odd-ratios* modélisés représentent plutôt bien la réalité donnée par les *odd-ratios* empiriques. L'exemple précédent illustre d'ailleurs ce constat : alors que le modèle *RL2* nous indique que les assurés souscrivant en famille ont 1,24 fois plus de chances de résilier leur contrat qu'un assuré célibataire, l'expérience montre qu'ils résilient 1,17 fois plus souvent (c'est également ce qui ressortait des études uni-variées) toutes choses égales par ailleurs. Cependant, pour les variables « AGE » et « PMM », le modèle *RL2* n'est pas idéalement ajusté aux données. Ce constat peut être en partie expliqué par le fait que ces deux variables ne disposent pas d'une aussi bonne significativité dans le modèle que les autres variables (à nuancer toutefois car la variable « SF » est à peu près aussi significative que la variable « PMM »).

Même s'il est dommageable que la régression logistique ne soit pas parfaitement ajustée aux données, nous obtenons malgré tout de bons résultats pour les variables les plus importantes dans la construction du modèle. Par ailleurs, il convient de vérifier la précision des prédictions faites par le modèle sur un échantillon de validation.

Étudions donc la matrice d'erreur du modèle *RL2* sur l'échantillon de validation qui donne le nombre d'assurés non correctement classés et représente le pouvoir prédictif du modèle. Afin de pouvoir classer les assurés, nous considérons que $Y = 1$ si la probabilité de résiliation modélisée est supérieure à 0,5 et que $Y = 0$ sinon.

	Prédit $Y = 1$ (résiliation)	Prédit $Y = 0$ (non résiliation)
Observé $Y = 1$	#succès = 1 775	#échecs = 396
Observé $Y = 0$	#assurés risqués à tort = 285	#rejets corrects = 6 836

Tableau 19 : Matrice d'erreur pour le modèle *RL2* sur l'échantillon de validation

Source : Propres calculs à partir de données réelles

Le modèle *RL2* permet de réaliser une bonne prédiction puisque seulement 7,3 % des prédictions sont erronées d'après la matrice d'erreur sur l'échantillon de validation. Cette erreur est très proche de celle donnée par l'algorithme CART sur l'arbre élagué (7,0 %) ou à l'aide des forêts aléatoires (7,2 %). Par ailleurs, l'erreur en termes de polices à risque non détectées par le modèle est de 18,2 % ce qui représente une légère amélioration par rapport à l'algorithme CART pour lequel cette erreur est supérieure à 19 %.

Il existe d'autres critères pour mesurer et comparer la qualité de prédiction de deux modèles. Nous définissons deux de ces critères : la sensibilité (Se) et la spécificité (Sp). Un succès est défini comme la situation où la valeur observée et la valeur prédite sont égales à 1 et un échec comme le cas où la valeur prédite vaut 0 et celle observée 1. Un rejet correct est comptabilisé lorsque l'observation et la prédiction valent 0. Enfin, un assuré est « risqué à tort » lorsque la prédiction vaut 1 tandis que l'observation est égale à 0.

Ainsi nous avons :

$$Se = \frac{\text{Nombre de succès}}{\text{Nombre de contrats pour lesquels une chute est observée}}$$

$$Sp = \frac{\text{Nombre de rejets corrects}}{\text{Nombre de contrats pour lesquels une chute n'est pas observée}}$$

Le tableau suivant synthétise les valeurs des critères de performance pour chacune des méthodes. Nous cherchons à minimiser la proportion d'échec (qui est égale à $1 - Se$). Cette proportion d'échec est ce que nous avons précédemment appelé l'erreur en termes de polices à risque non détectées.

	T_{Max}	$T_{\text{élagué}}$	$T_{\text{forêt aléatoire}}$	<i>RL2</i>
Se	80,06 %	80,88 %	79,71 %	81,76 %
Sp	96,74 %	96,70 %	96,76 %	96,00 %
1-Se	19,94 %	19,12 %	20,29 %	18,24 %

Tableau 20 : Critères de performance de la prédiction

Source : Propres calculs à partir de données réelles

Les prédictions obtenues avec le modèle *RL2* ont moins d'échecs mais légèrement plus d'assurés risqués à tort que les prédictions réalisées avec l'algorithme CART. De manière générale, la spécificité est meilleure que la sensibilité : les modèles ont plus de facilités à bien classer un assuré non risqué plutôt qu'un assuré risqué.

Pour conclure sur le modèle de régression logistique appliqué au portefeuille XW, plusieurs observations sont à faire :

- Le modèle utilisé n'est pas toujours adapté pour établir des profils de risque pour un assuré en fonction de ses caractéristiques. En effet, lorsque la technique des *odd-ratios* est utilisée, des conclusions contradictoires ont parfois été obtenues entre les valeurs prédites par le modèle et les valeurs calculées empiriquement. Toutefois, pour les variables les plus pertinentes du modèle (qui sont aussi les variables les plus discriminantes pour expliquer les résiliations dans l'algorithme CART), de bons résultats ont été obtenus.
- Le modèle donne de bons résultats en termes de prédiction sur l'échantillon de validation. Ces résultats sont très proches de ceux obtenus pour l'algorithme CART voire meilleurs en ce qui concerne la proportion d'échec.

Si ce modèle semble intéressant, il convient malgré tout de faire preuve de prudence dans son utilisation.

4.5.4. Portefeuille YZ

Après avoir réalisé la modélisation avec toutes les variables explicatives (modèle *RLE1*), seules les variables les plus pertinentes ont été conservées pour obtenir le modèle *RLE2* à l'aide de la fonction *stepAIC()*

Afin de mieux comprendre comment le modèle *RLE2* a été obtenu, les différentes sorties fournies par le logiciel R sont données ci-dessous.

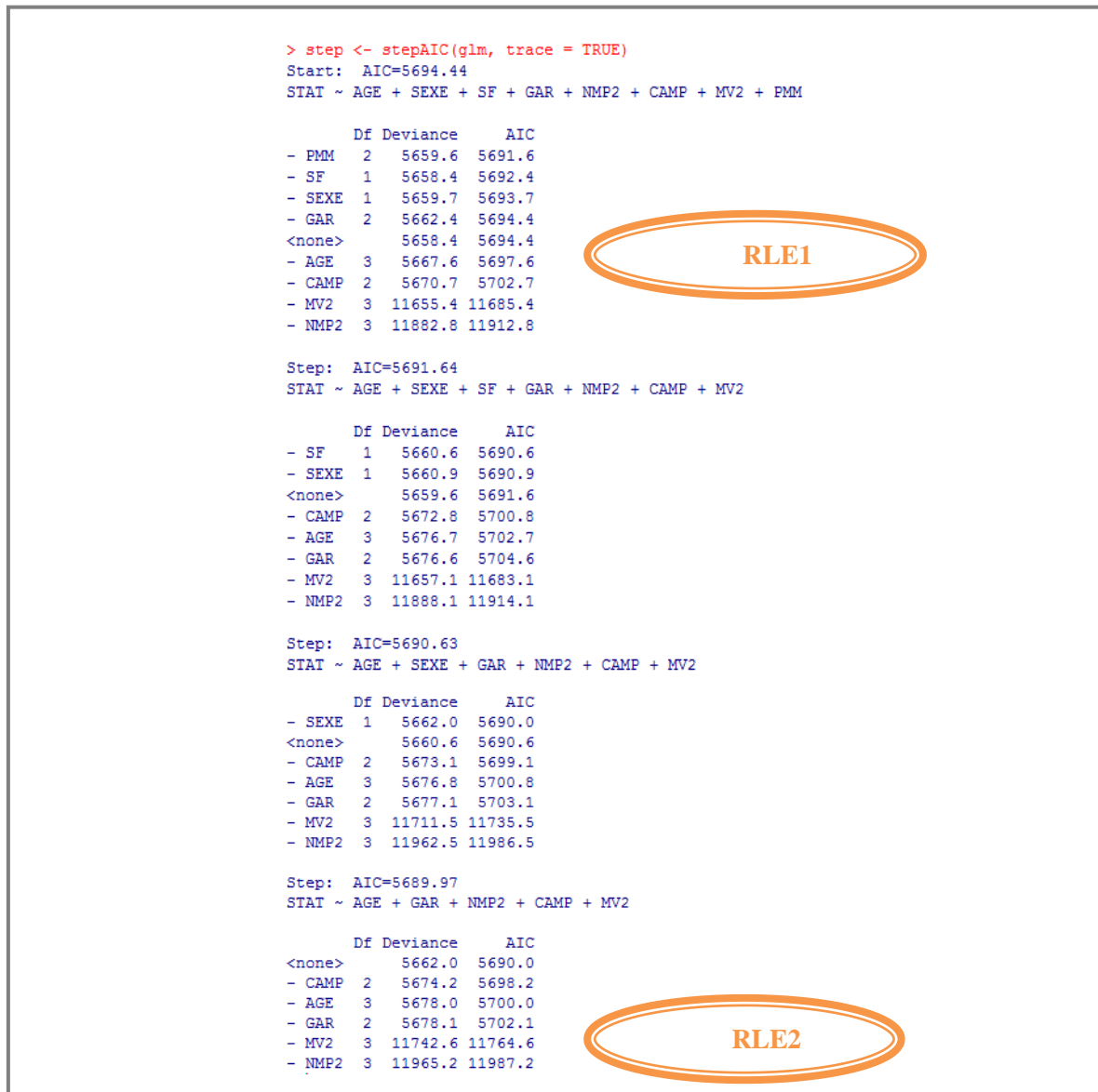


Figure 36 : Eléments de la fonction « *stepAIC()* »

Pour rappel, la procédure « *stepAIC()* » permet de sélectionner le meilleur modèle selon un critère de minimisation de l'AIC par des retraits successifs de variables explicatives en partant du modèle complet. La première étape consiste à calculer l'AIC du modèle de référence (ici *RLE1*), ainsi que l'impact du retrait de chaque variable sur l'AIC. La variable dont le retrait permet une baisse maximale de l'AIC est enlevée, jusqu'à ce que le retrait d'une des variables restantes soit synonyme de l'augmentation du critère. Ici les variables « PMM », « SF » et « SEXE » ont été successivement retirées pour obtenir le modèle *RLE2* optimal. Comme pour le portefeuille XW, nous remarquons l'importance des variables « MV2 » et « NMP2 » dans la réduction de l'AIC.

Une fois le critère de sélection optimisé, il convient de contrôler la pertinence du modèle et des variables explicatives à l'aide des fonctions « *summary()* » et « *drop1()* ». Les sorties R sont représentées ci-dessous avec, pour chaque fonction, une explication des résultats obtenus.

```

> summary(step)

Call:
glm(formula = STAT ~ AGE + GAR + NMP2 + CAMP + MV2, family = binomial(link = "logit"),
    data = app)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6160  -0.4486  -0.0001   0.2165   3.1294

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    5.83566    0.24046  24.269 < 2e-16 ***
AGE35-44      -0.08759    0.09814  -0.892  0.37214
AGE45-54       0.01310    0.10700   0.122  0.90253
AGE55-65       0.38300    0.13207   2.900  0.00373 **
GAR50-60       0.18535    0.08796   2.107  0.03511 *
GAR90-120      0.55850    0.13949   4.004  6.23e-05 ***
NMP21-3        -2.81450    0.07652 -36.780 < 2e-16 ***
NMP211        -25.60576   6187.70494 -0.004  0.99670
NMP24-10       -8.21624    0.21114 -38.914 < 2e-16 ***
CAMPSFER      -0.32183    0.15738  -2.045  0.04086 *
CAMPVENC       0.14199    0.08001   1.775  0.07596 .
MV2201002-201004 -2.42085    0.20494 -11.812 < 2e-16 ***
MV2201005-201007 -5.46509    0.20999 -26.026 < 2e-16 ***
MV2201008-201009 -25.25445   217.43063 -0.116  0.90753
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14722  on 11481  degrees of freedom
Residual deviance:  5662  on 11468  degrees of freedom
AIC: 5690

Number of Fisher Scoring iterations: 18

```

Figure 37 : Eléments de la fonction « *summary()* » d'une régression logistique

Nous étudions à nouveau la déviance afin de juger de la pertinence du modèle. Celle-ci est égale à 5 662 ce qui est inférieur au nombre de degrés de liberté (11 468). Nous pouvons donc admettre que le modèle est pertinent (pour rappel, le modèle est acceptable si la déviance est inférieure au nombre de degrés de liberté), sans prendre le soin de réaliser un test de déviance puisque le quantile du χ^2 est automatiquement supérieur à ce nombre.

En regardant le test de Student pour tester l'hypothèse de nullité d'un coefficient, nous nous apercevons qu'un certain nombre de coefficients ne sont pas significatifs. C'est le cas des modalités « 35-44 » et « 45-54 » de la variable « AGE », « 11 » de la variable « NMP2 » et « 201008→201009 » de la variable « MV2 ». Nous déduisons de la non-significativité des deux dernières modalités que nous avons soit un manque de données, soit des données incohérentes pour les deux derniers mois étudiés. L'ensemble des coefficients associés à ces modalités peuvent donc être considérés comme nuls et ne sont pas étudiés dans la suite de l'étude. Il convient alors de se demander si la nullité des coefficients a une influence sur la significativité des variables elles-mêmes. Cette question est étudiée lors de l'analyse de la sortie de la fonction « *drop1()* » dans la suite de cette section.

Par ailleurs, l'individu de référence utilisé pour calculer l'*intercept* a le profil suivant : il s'agit d'un assuré âgé de 18 à 34 ans avec un niveau de garantie entre 15 et 25 €, ayant souscrit son contrat en début de période d'observation au cours de la campagne de télémarketing « C » et ayant payé 0 prime.

En ce qui concerne l'interprétation des coefficients, la fonction lien étant la fonction logistique, un coefficient négatif indique un risque plus faible que pour l'individu de référence et inversement un coefficient positif indique un risque plus fort. Ainsi nous constatons, à titre d'exemple, que les assurés ayant payé 0 prime

sont les assurés les plus risqués par rapport à cette variable. Par ailleurs, pour le portefeuille YZ, beaucoup de coefficients sont proches de 0, en particulier pour les variables « AGE », « GAR » et « CAMP ». Dès lors, il semble que ces variables ont peu d'importance dans l'explication des résiliations. Les variables « MV2 » et « NMP2 » sont donc prépondérantes comme nous l'avons déjà vu à plusieurs reprises. Enfin, l'individu de référence a un niveau de risque élevé puisque l'estimation du coefficient β_0 est égale à 5,83566.

Afin de juger de la significativité des variables et non plus seulement des occurrences, la fonction « *drop1()* » est utilisée :

```
> drop1(step, test="Chisq")
Single term deletions

Model:
STAT ~ AGE + GAR + NMP2 + CAMP + MV2
Df Deviance   AIC    LRT   Pr(Chi)
<none>          5662.0 5690.0
AGE      3    5678.0 5700.0  16.0 0.0011261 **
GAR      2    5678.1 5702.1  16.1 0.0003222 ***
NMP2     3   11965.2 11987.2 6303.2 < 2.2e-16 ***
CAMP     2    5674.2 5698.2  12.2 0.0022708 **
MV2      3   11742.6 11764.6 6080.6 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 38 : Résultats issus de la fonction « *drop1()* »

Le principe de la fonction « *drop1()* » ayant été donné dans la section sur le portefeuille XW, nous invitons le lecteur à s'y reporter pour rappel. Les variables « GAR », « NMP2 » et « MV2 » sont pertinentes pour la modélisation par régression logistique des résiliations du portefeuille YZ. Les variables « AGE » et « CAMP » sont également acceptables à un niveau supérieur à 99 %.

Au vu des différents contrôles de pertinence, le modèle *RLE2* convient aux données du portefeuille YZ. Interprétons maintenant ce modèle à l'aide de la technique des *odd-ratios*. Le tableau représenté ci-dessous fait la synthèse des *odd-ratios* obtenus de façon empirique.

Odd-ratio (OR)	Référence	Autres modalités		
		35-44	45-54	55-65
« AGE »	18-34			
Nombre de résiliations	772	1489	1190	454
OR empiriques		1,09	0,96	0,90
OR modélisés		0,91	1,01	1,46
« GAR »	15-25	50-60	90-120	
Nombre de résiliations	755	2588	562	
OR empiriques		1,38	2,24	
OR modélisés		1,20	1,75	
« MV2 »	200911-201001	201002-201004	201005-201007	201008-201009
Nombre de résiliations	1000	1502	1403	0
OR empiriques		0,83	0,76	0,58
OR modélisés		0,61	0,24	0
« NMP2 »	0	1 → 3	4 → 10	11
Nombre de résiliations	2769	1056	80	0
OR empiriques		0,41	0,21	0,58
OR modélisés		0,27	0,028	0

« <i>CAMP</i> »	C	D	E	
<i>Nombre de résiliations</i>	1892	137	1876	
<i>OR empiriques</i>		0,54	0,65	
<i>OR modélisés</i>		0,74	1,15	

Tableau 21 : Synthèse des odd-ratios obtenus de façon empirique et en utilisant le modèle RLE2

Pour rappel, un *odd-ratio* plus grand que 1 signifie un risque de résiliation plus important que la modalité de référence alors qu'inversement un *odd-ratio* plus petit que 1 signifie un risque moindre.

Nous nous apercevons que les *odd-ratios* modélisés représentent mal la réalité donnée par les *odd-ratios* empiriques pour les variables « AGE » et « CAMP ». Néanmoins, pour les variables « GAR », « MV2 » et « NMP2 », la comparaison des *odd-ratios* empiriques et modélisés donne des résultats intéressants. Ainsi, si le modèle *RLE2* n'est pas idéalement ajusté, nous constatons une bonne adéquation du modèle aux données pour les variables les plus pertinentes du modèle. Il convient alors d'étudier la précision des prédictions afin d'évaluer le pouvoir prédictif du modèle.

L'analyse de la matrice d'erreur du modèle *RLE2* sur l'échantillon de validation donne le nombre d'assurés non correctement classés et représente le pouvoir prédictif du modèle. Afin de pouvoir classer les assurés, nous considérons que $Y = 1$ si la probabilité de résiliation modélisée est supérieure à 0,5 et que $Y = 0$ sinon.

	Prédit $Y = 1$ (résiliation)	Prédit $Y = 0$ (non résiliation)
Observé $Y = 1$	#succès = 1 711	#échecs = 179
Observé $Y = 0$	#assurés risqués à tort = 389	#rejets corrects = 3 386

Tableau 22 : Matrice d'erreur pour le modèle RLE2 sur l'échantillon de validation

Source : Propres calculs à partir de données réelles

Le modèle *RLE2* permet de réaliser une bonne prédiction puisque seulement 10,1 % des prédictions sont erronées d'après la matrice d'erreur sur l'échantillon de validation. Cette erreur est très proche de celle donnée par l'algorithme CART sur l'arbre élagué (10,0 %) ou à l'aide des forêts aléatoires (10,3 %). Par ailleurs, l'erreur en termes de polices à risque non détectées par le modèle est de 9,5 %, ce qui est sensiblement égal à ce qui a été obtenu pour l'arbre élagué avec l'algorithme CART (9,1 %).

Comparons maintenant les critères de sensibilité et de spécificité, définis dans l'étude du portefeuille XW, pour les différents modèles.

	T_{Max}	$T_{\text{élagué}}$	$T_{\text{forêt aléatoire}}$	<i>RLE2</i>
Se	86,30 %	90,95 %	89,51 %	90,53 %
Sp	91,02 %	89,56 %	89,78 %	89,70 %
1-Se	13,70 %	9,05 %	10,49 %	9,47 %

Tableau 23 : Critères de performance de la prédiction

Source : Propres calculs à partir de données réelles

Les prédictions obtenues avec le modèle *RLE2* semble offrir un bon compromis entre un faible taux d'échec et un bon niveau de spécificité. Les résultats obtenus avec ce modèle sont très proches de ceux obtenus avec l'arbre élagué de l'algorithme CART.

Les conclusions quant au modèle de régression logistique appliqué au portefeuille YZ sont très proches de celles du portefeuille XW, à savoir :

- Le modèle utilisé n'est pas toujours adapté pour établir des profils de risque pour un assuré en fonction de ses caractéristiques. En effet, lorsque la technique des *odd-ratios* est utilisée, des conclusions contradictoires ont parfois été obtenues entre les valeurs prédites par le modèle et les valeurs calculées empiriquement. Toutefois, pour les variables les plus pertinentes du modèle (qui sont aussi les variables les plus discriminantes pour expliquer les résiliations dans l'algorithme CART), de bons résultats ont été obtenus.
- Le modèle donne de bons résultats en termes de prédiction sur l'échantillon de validation. Ces résultats sont très proches de ceux obtenus pour l'algorithme CART voire meilleurs en ce qui concerne la proportion d'échec.

Il convient donc, là encore, de faire preuve de prudence en utilisant ce modèle.

Par ailleurs, bien que les niveaux de résiliation soient sensiblement plus élevés en YZ qu'en XW, nous ne constatons pas de réelles différences quant aux facteurs de risque des deux portefeuilles. L'aspect temporel joue un rôle crucial dans chacun des cas et éclipse, dans une certaine mesure, le poids des autres variables.

A l'issue de cette partie sur l'analyse du risque de chute, il apparaît que les prédictions peuvent aussi bien être faites à l'aide de l'algorithme CART que d'une régression logistique, les deux méthodes donnant des résultats similaires en termes de qualité de prédiction sur les échantillons de validation des deux portefeuilles. Toutefois, ces modèles sont plus performants pour détecter les polices non risquées plutôt que risquées. Ceci est, dans une certaine mesure, la conséquence d'un manque de données sur les individus ayant résilié leur police mais, bien que ce point soit dommageable, la fiabilité des modèles ne semble toutefois pas devoir être remise en cause.

Quant à la détermination de profils de risque en fonction des caractéristiques de l'assuré, il est apparu que chaque méthode possède ses forces et ses faiblesses. Les études descriptives issues de l'application de la méthode de Kaplan Meier permettent de comparer facilement l'impact d'une modalité d'une variable par rapport au portefeuille global. Pour autant, les sensibilités étudiées ne concernent qu'une seule variable et il n'est pas possible d'étudier certaines interactions ou de déterminer les variables les plus discriminantes dans la décision de l'assuré de résilier. De ce point de vue, l'algorithme CART est très intéressant car il permet de hiérarchiser de façon quantitative l'importance de chacune des variables. De même, l'utilisation de l'algorithme CART offre la possibilité d'étudier un assuré à travers plusieurs variables explicatives par la lecture de l'arbre de décision élagué. Cette dernière représentation présente en plus l'avantage d'une simplicité d'interprétation et de représentation pour l'utilisateur. Pour autant, lorsque nous avons utilisé cette méthode, nous nous sommes aperçus que celle-ci n'est peut-être pas totalement adaptée à notre étude des résiliations sur un portefeuille de polices issues de campagnes de télémarketing. En effet, à la lecture des résultats, il apparaît que l'aspect temporel est déterminant pour expliquer les résiliations (en particulier les premiers mois après la souscription) reléguant les autres variables à un niveau secondaire. Ce phénomène semble avoir plusieurs sources. Tout d'abord, la nature même du processus de télémarketing est propice à des résiliations précoces : les assurés se rétractent beaucoup au cours des premiers mois car d'une part, la vente n'est pas toujours réellement souhaitée par l'assuré et, d'autre part, des gestes commerciaux sont souvent pratiqués lors des premiers mois de couverture. Une autre raison pouvant expliquer les résiliations précoces est la nature des données elles-mêmes. En effet, il est apparu que des problèmes de gestion survenus au cours des premiers mois de mise en place des portefeuilles, en particulier pour YZ, permettent de comprendre la proportion très importante de résiliations précoces lors des mois qui ont suivi le lancement des produits. Enfin,

il s'agit aussi d'un phénomène naturel : plus un assuré reste dans le portefeuille de l'assureur, moins il a tendance à résilier. Il en ressort que l'algorithme CART n'est peut-être pas totalement adapté à nos données pour pouvoir établir des profils de risque. Toutefois, cette méthode reste intéressante car elle a permis de faire ressortir certains aspects non détectés pendant l'étude descriptive et de réaliser des prédictions de bonne qualité. La régression logistique présente elle aussi des avantages et des inconvénients. Elle permet de réaliser des prédictions de qualité équivalente à ce que nous obtenons avec l'algorithme CART. Par ailleurs, nous avons réussi à l'aide de cette méthode à déterminer les variables les plus informatives en regardant leur contribution dans la baisse du critère AIC et à comparer, pour chaque variable, le niveau de risque d'une modalité par rapport à une modalité de référence. Grâce à ce modèle, il est donc possible de déterminer les variables les plus discriminantes pour expliquer les résiliations et d'évaluer le risque porté par des individus possédant des caractéristiques différentes. Pour autant, il est aussi apparu, à travers la comparaison des *odds ratios* empiriques et modélisés, que, pour les variables les moins informatives, le modèle ne correspond pas forcément à la réalité.

Il apparaît donc que les trois modèles sont complémentaires et peuvent être utilisés conjointement pour recouper des informations et faire ressortir certains aspects plutôt que d'autres. Ainsi, il convient d'être prudent lors de l'utilisation de chacun des modèles pris séparément, chacun présentant des lacunes internes ou inhérentes aux données utilisées qui ne sont pas toujours adaptées au modèle ou de qualité non optimale (malgré des contrôles préalables). Une démarche intéressante est donc la confrontation des méthodes que nous avons tenté de réaliser dans cette étude. Ainsi, concrètement, l'étude de ces résultats nous permet de saisir l'intérêt de bien comprendre le profil des assurés du portefeuille et d'en identifier rapidement les variables discriminantes afin de pouvoir adapter/piloter le profil de risque au travers de la distribution si besoin.

Enfin, avec le recul, il semble que la démarche idéale aurait plutôt consisté à sélectionner les variables explicatives les plus pertinentes à l'aide de l'algorithme CART et des forêts aléatoires puis, confirmer ces hypothèses à l'aide d'études uni-variées, avant d'implémenter le modèle de régression logistique pour utiliser la technique des *odd-ratios* qui présente un avantage de simplicité d'interprétation.

PARTIE IV : RISQUE DE CREDIT LIE A DES STRUCTURES DE FINANCEMENT

Le risque de contrepartie (également appelé risque de crédit) est défini⁵⁴ comme le risque de perte résultant d'une défaillance imprévue ou d'une dégradation de la note de crédit des contreparties de l'entreprise. Il est fonction de trois paramètres : le montant de la créance, la probabilité de défaut et la proportion de la créance qui ne sera pas recouvrée en cas de défaut.

Dans le cadre de la directive européenne Solvabilité II, ce risque doit être quantifié et constitue un des modules de risque de la formule standard du SCR (*Solvency Capital Requirement*) dont le calcul est exigé pour les organismes assureurs.

L'étude de ce risque pour les contrats d'assurance issus de la distribution directe se révèle particulièrement intéressante car, contrairement à une structure de réassurance classique impliquant uniquement la cédante et le réassureur, il est possible d'avoir pour de telles affaires des relations financières entre le réassureur et le courtier. Ainsi, cette dernière partie a pour objectif de mettre en évidence l'impact de différentes structures de réassurance sur la quantification du risque de crédit dans le calcul du SCR.

Le premier chapitre de cette dernière partie explique la problématique du risque de contrepartie pour un réassureur. Nous présentons ensuite les spécifications techniques de la dernière étude quantitative d'impact (QIS5) pour le calcul du SCR du module du risque de contrepartie (Chapitres 2 et 3). Le dernier chapitre illustre les concepts définis à travers un exemple numérique développé spécifiquement pour ce mémoire et comparant deux structures de réassurance : avec, puis sans risque de crédit du courtier pour le réassureur dans le cas d'une clause de reprise sur commission initiale en cas de résiliation de l'assuré.

Chapitre 1. Risque de contrepartie chez le réassureur

Initialement, les deux parties concernées par la vente du produit d'assurance sont le courtier qui place la police auprès de l'assuré et l'assureur qui reçoit les primes, paie les sinistres et verse une commission au courtier.

Les affaires de distribution directe prévoient en particulier le versement par l'assureur d'une commission d'acquisition au courtier pour chaque police placée par ce dernier. Cette commission d'acquisition, payée au commencement du contrat, peut être importante (par exemple de l'ordre de 100 % de la prime annuelle) et représente parfois un investissement trop lourd à supporter pour l'assureur qui préfère alors se décharger de cette obligation en demandant au réassureur, qui a une plus grande capacité financière, de s'acquitter de ce paiement initial.

La possibilité d'un mécanisme de protection de l'assureur (ici du réassureur) qui souhaite que les polices restent suffisamment en portefeuille afin de rentabiliser cet investissement initial a déjà été évoquée. Il s'agit de la clause de reprise de commission qui doit inciter le courtier à avoir une politique de souscription de qualité.

⁵⁴ Vernimmen P., Quiry P., Le Fur Y. (2011), « Finance d'entreprise 2011 », Dalloz, p377, 438, 441

Pour rappel, le principe de la reprise de commission est le suivant : pour chacune des polices résiliées avant une période donnée (généralement un an), le courtier doit rembourser la commission d'acquisition perçue à la souscription *prorata temporis* de la durée pendant laquelle l'assuré n'est plus dans le portefeuille. Par exemple, si l'assuré résilie au bout de trois mois, le courtier devra rembourser à l'assureur l'équivalent de 3/4 de la commission d'acquisition. Ainsi, la commission d'acquisition n'appartient réellement au courtier qu'au bout d'un an.

Le risque de crédit encouru par le réassureur est directement lié à l'existence des commissions et de la reprise de commission. En effet, en payant « d'avance » le financement de la police, le réassureur s'expose au risque de la disparition ou de la dégradation de la note de crédit de la contrepartie qui doit le rembourser en cas de résiliation du contrat d'assurance.

Deux cas sont alors à envisager pour la caractérisation et l'évaluation du risque de contrepartie.

- Le réassureur paie la commission d'acquisition ainsi que les commissions de renouvellement à l'assureur et est protégé par une clause de reprise de commission en cas de résiliation précoce. L'assureur verse ensuite les différentes commissions au courtier. Il n'y a aucun flux financier entre le réassureur et le courtier.
- Le réassureur paie la commission d'acquisition ainsi que les commissions de renouvellement au courtier et est protégé par une clause de reprise de commission, versée par le courtier, en cas de résiliation précoce.

Quelle que soit la situation, dans le cadre d'un traité en quote-part, le réassureur paie x % des sinistres et reçoit x % des primes.

Dès lors, le paiement de la commission d'acquisition crée une exposition du réassureur au risque de contrepartie représenté par la cédante (dans le cas où le réassureur n'a pas de relation avec le courtier) ou par le courtier (dans le cas où le réassureur paie directement celui-ci) et la cédante.

L'exposition maximale au risque de contrepartie pour les polices souscrites d'un mois donné est alors égale à :

- La commission d'acquisition si le réassureur n'est exposé qu'au risque de la cédante,
- La reprise de commission pour l'exposition vis-à-vis du courtier plus la commission d'acquisition déduction faite de la reprise de commission vis-à-vis de la cédante dans le cas où il existe une relation contractuelle entre le courtier et le réassureur.

Analysons désormais les conséquences pratiques de ces deux structures de financement à la lumière des spécifications techniques du dernier questionnaire quantitatif d'impact (QIS5).

Chapitre 2. Calcul du SCR pour le module risque de contrepartie

Avant d'évoquer la prise en compte et l'évaluation du risque de contrepartie dans le projet Solvabilité II, introduisons les éléments nécessaires à la bonne compréhension de cette problématique.

Section 2.1. La directive Solvabilité II : objectifs et organisation en piliers

La directive Solvabilité II est le résultat d'un projet européen démarré à la fin des années 90. Mis en place afin de créer un régime de solvabilité plus sensible aux risques encourus par les compagnies d'assurance, ce

projet a permis de réaliser une revue en profondeur des règles prudentielles du monde de l'assurance. A l'issue de ce processus, la directive cadre (2009/138/CE) a été adoptée le 5 mai 2009 par le conseil ECOFIN après avoir été votée par le parlement européen le 22 avril 2009 dans l'objectif d'une application à partir du 1^{er} janvier 2013.

La réforme Solvabilité II est fondée sur une structure en trois piliers :

- Le pilier 1 détermine des exigences quantitatives à respecter, notamment sur l'harmonisation des provisions techniques et l'instauration de minima de fonds propres,
- Le pilier 2 impose la mise en place de dispositifs de gouvernance des risques (processus, responsabilités, production et suivi d'indicateurs...),
- Le pilier 3 fixe les exigences en termes de reporting et de transparence.

Dans le cadre de ce mémoire, nous nous intéressons à une partie du pilier 1. Ce pilier introduit notamment la notion de « Capital de Solvabilité Requis » (ou SCR : *Solvency Capital Requirement*). Il s'agit du minimum de fonds propres à détenir par la compagnie pour se protéger contre des situations exceptionnelles, mais également d'un premier seuil d'alerte pour une intervention du superviseur. Ce SCR pourra être calculé soit avec une formule standard, soit au travers d'un modèle interne, soit à l'aide d'une combinaison des deux.

Section 2.2. Le processus consultatif dans Solvabilité II: les études d'impacts quantitatives

Tout au long de l'élaboration de la directive Solvabilité II, des échanges ont eu lieu entre la Commission Européenne et les acteurs du marché de l'assurance afin de préparer le marché à la réforme en mettant en place les méthodes de prise en compte et d'évaluation des risques et d'obtenir des informations sur la faisabilité des calculs ainsi que sur l'impact potentiel en termes de besoin de capital. Pour ce faire, différentes études d'impacts quantitatives (QIS) ont été réalisées depuis 2005.

La dernière étude en date, le QIS5, a pris fin en octobre 2010 et avait pour but de mesurer l'impact du dernier calibrage de la formule standard, d'évaluer la faisabilité des calculs ainsi que l'état d'avancement des acteurs du marché et de collecter des données.

Pour réaliser cette étude, les acteurs du marché de l'assurance ont eu à leur disposition un certain nombre de spécifications techniques⁵⁵ devant servir au calcul et au calibrage des différents éléments exigés par le pilier 1 de la réforme. La détermination de la valeur du SCR est une de ces exigences.

Section 2.3. Calcul du SCR dans les spécifications techniques du QIS5 (Modèle Standard)

Le SCR, ou capital de solvabilité requis, correspond au niveau de fonds propres nécessaire à une entreprise afin d'être en mesure d'absorber des pertes imprévues importantes. Il est calibré de manière à représenter la *Value at Risk* (VaR) des pertes de la société sur un horizon temporel d'un an avec un niveau de confiance de 99,5 %. Autrement dit, le SCR doit couvrir une probabilité de ruine de 0,5 % sur un horizon d'un an ce qui correspond à une faillite tous les 200 ans.

⁵⁵ « QIS5 Technical Specifications » consultable sur http://ec.europa.eu/internal_market/insurance/docs/solvency/qis5/201007/technical_specifications_en.pdf

Le calcul du SCR suivant la formule standard est décomposé en plusieurs modules de risque eux-mêmes découpés en différents sous-modules. La structure du SCR, donnée dans les spécifications techniques du QIS5, comprend trois modules de risque de souscription (vie, non-vie et santé), un module de risque de contrepartie, un module de risque de marché, un module de risque pour les actifs incorporels ainsi qu'un module pour le risque opérationnel. La structure générale du SCR selon le QIS5 est donnée ci-dessous.

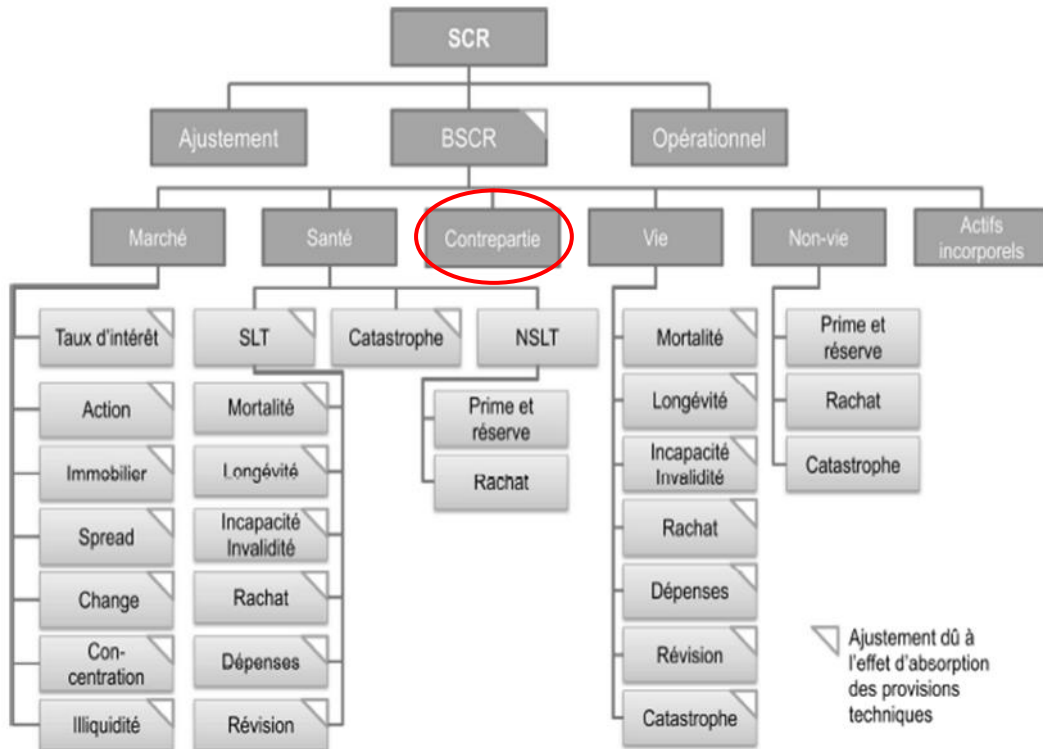


Figure 39 : Structure générale du SCR
 Source : Spécifications techniques du QIS5, CEIOPS

Après cette introduction assez générale sur le calcul du SCR dans le QIS5, intéressons-nous désormais au module du risque de contrepartie.

Chapitre 3. Spécifications du QIS5 pour le calcul du SCR du module du risque de contrepartie

Dans la méthodologie proposée dans le QIS5, le CEIOPS (*Committee of European Insurance and Occupational Pensions Supervisors*), l'ancien organisme regroupant les représentants des autorités de contrôle des Etats membres remplacé depuis le 1^{er} janvier 2011 par l'EIOPA (*European Insurance and Occupational Pensions Authority*), tendait à séparer les contreparties en deux types d'exposition :

- D'une part, les contreparties de type 1 peu diversifiées (moins de 15 parties) et qui font ou pourraient faire l'objet d'une notation. Elles comprennent, par exemple, les programmes de réassurance, les produits dérivés, les titrisations, les lettres de crédit et les espèces en banques.
- D'autre part, les contreparties de type 2 qui sont des contreparties plus diversifiées (excèdent 15 parties) mais non notées, comme les comptes à payer aux intermédiaires, les arriérés de primes des assurés et les dépôts.

Section 3.1. Calcul du capital requis pour une exposition de type 1

Dans la formule proposée dans les spécifications techniques du QIS5, la probabilité de défaut de la contrepartie est fonction de la notation de cette dernière :

<i>Notation</i>	<i>Probabilité de défaut</i>
AAA	0,002 %
AA	0,01 %
A	0,05 %
BBB	0,24 %
BB	1,20 %
B	6,04 %
CCC	30,41 %

Tableau 24: Probabilités de défaut selon la notation
Source : Spécifications techniques du QIS5, CEIOPS

Pour les contreparties non notées soumises à Solvabilité II et disposant d'un montant minimum de fonds propres (ou MCR : *Minimum Capital Requirement*), la probabilité de défaut dépend du ratio de solvabilité⁵⁶. Ainsi, à titre d'exemple, une contrepartie dont le ratio de solvabilité est compris entre 175 % et 200 % correspond à une probabilité de défaut de 0,05 % (c'est-à-dire équivalente à une contrepartie notée A). Pour les contreparties soumises à Solvabilité II n'ayant pas atteint leur MCR, la probabilité de défaut est fixée à 30 %. Enfin, pour les contreparties non notées et non soumises à Solvabilité II, la probabilité de défaut est égale à 10 %.

Dès lors, le capital requis pour des expositions de type 1 est calculé ainsi :

$$SCR_{def,1} = \begin{cases} 3 \times \sqrt{V} & \text{si } \sqrt{V} \leq 5\% \times \sum_i LGD_i \\ \min\left(\sum_i LGD_i; 5 \times \sqrt{V}\right) & \text{sinon} \end{cases}$$

où :

- LGD_i (Loss Given Default) est le montant de perte en cas de défaut de la contrepartie i , c'est-à-dire l'exposition au risque, dont le calcul, proposé par le CEIOPS, s'écrit :

$$LGD_i = \max\left((1 - RR_{re}) \times (Recoverables_i + RM_{re,i} - Collateral); 0\right)$$

avec :

- RR_{re} est le taux de recouvrement de la créance en cas de défaut,
- $Recoverable_i$ est la valeur de la créance de la contrepartie i ,
- $RM_{re,i}$ est l'effet d'atténuation du risque grâce à la diversification,
- $Collateral$ matérialise l'atténuation du risque de contrepartie par des nantissements (c'est-à-dire par des outils qui permettent au réassureur d'avoir une certaine garantie en cas de défaut des cédantes dans notre cas).

⁵⁶ Le ratio de solvabilité est défini dans les spécifications techniques du QIS5 comme le rapport $\frac{\text{Fonds propres}}{SCR}$

- V représente la variance d'une distribution de la perte liée au risque de contrepartie, déterminée par la formule suivante :

$$V = \sum_j \sum_k u_{j,k} \times y_j \times y_k + \sum_j v_j \times z_j$$

où pour chaque classe de notation j , y_j et z_j sont définis comme des sommes sur toutes les contreparties indépendantes i pour le niveau de notation j .

$$y_j = \sum_i LGD_i \text{ et } z_j = \sum_i (LGD_i)^2$$

où $u_{j,k}$ et v_j sont deux facteurs fixes qui ne dépendent que de la notation :

$$u_{j,k} = \frac{p_j \times (1 - p_j) \times p_k \times (1 - p_k)}{(1 + \gamma)(p_j + p_k) - p_j \times p_k}, v_j = \frac{(1 + 2 \times \delta) \times p_j \times (1 - p_j)}{2 + 2 \times \delta - p_j}$$

et où p_j est la probabilité de défaut d'une contrepartie de classe j et $\delta = 0,25$.

Section 3.2. Calcul du capital requis une exposition de type 2

Dans le cas des contreparties de type 2, le capital nécessaire est beaucoup plus simple à calculer et consiste en la somme du produit de chaque exposition par un facteur de risque, indépendant de la probabilité de défaut de contrepartie individuelle et de la taille ou du nombre de contreparties. La seule différence quant au facteur de risque appliqué réside dans le temps de recouvrement des actifs. Le montant du SCR pour le risque de contrepartie est alors égal à :

$$SCR_{def,2} = 15 \% \times E + 90 \% \times E_{+3}$$

où E est la somme de toutes les expositions de type 2 à l'exception des intermédiaires mettant plus de 3 mois à payer les créances dues et E_{+3} représente l'exposition des intermédiaires mettant plus de 3 mois à payer les créances dues.

Section 3.3. Calcul du capital requis pour l'ensemble des contreparties

Finalement, après avoir calculé les capitaux requis pour les contreparties de type 1 et 2, la valeur du SCR pour le module du risque de crédit est égale à :

$$SCR_{def} = \sqrt{SCR_{def,1}^2 + 1,5 \times SCR_{def,1} \times SCR_{def,2} + SCR_{def,2}^2}$$

Chapitre 4. Comparaison de deux structures de réassurance : avec/sans risque de crédit du courtier pour le réassureur dans le cas d'une clause de reprise sur commission initiale en cas de résiliation

Dans ce dernier chapitre, nous cherchons à quantifier, à travers un modèle simplifié développé dans le cadre du mémoire, l'impact pour un réassureur de la structure contractuelle d'un traité de réassurance en quote-part sur l'évaluation du SCR du module du risque de contrepartie. Dans cette optique, les résultats obtenus sont comparés pour deux structures :

- Dans la première structure, le réassureur est exposé uniquement au risque de défaut de la cédante.
- Dans la deuxième structure, le réassureur est exposé au risque de défaut du courtier.

Section 4.1. Hypothèses retenues

4.1.1. Exposition au risque

En théorie, l'exposition totale du réassureur au risque de crédit est égale à la commission d'acquisition qu'il finance au titre des profits futurs de la police et de la clause de reprise de commission, si une telle clause a été mise en place. Ainsi, selon la structure contractuelle qui lie le réassureur aux autres parties, l'exposition théorique au risque pour le réassureur, pour un mois M est :

- la commission d'acquisition s'il est exposé uniquement à la cédante.
- la reprise de commission, par rapport au courtier, plus la commission d'acquisition diminuée de la reprise de commission, par rapport à la cédante, s'il est exposé aux deux parties.

Dans cette application, seule l'exposition liée à la clause de reprise de commission est prise en compte. En effet, notre objectif consiste uniquement à analyser l'impact de la clause de reprise de commission initiale en fonction de l'hypothèse de résiliation retenue et du type de structure contractuelle. Dès lors, l'exposition au risque retenue pour l'application est :

- la reprise de commission par rapport à la cédante dans le premier cas,
- la reprise de commission par rapport au courtier dans le second cas.

Par ailleurs, la durée de la clause de reprise de commission est de 12 mois et, passée cette période, le réassureur n'est plus en risque.

En pratique, l'exposition au risque (ou LGD selon la dénomination adoptée dans les spécifications techniques du QIS5) est égale à la valeur actuelle probable sur 12 mois des flux de reprise de commission attendus par le réassureur. En effet, nous avons décidé, dans un souci de clarté, de considérer que :

- le taux de recouvrement est égal à 100 %,
- l'effet d'atténuation par diversification est nul car nous ne considérons qu'une seule contrepartie pour le réassureur,
- le nantissement n'existe pas.

Un taux d'actualisation annuel, i_{annuel} , de 3 % a été retenu.

Enfin, la commission d'acquisition est payée par le réassureur au mois « 0 », c'est-à-dire avant même que les primes de réassurance ne lui soient versées. Dès lors, les premiers flux de reprise de commission sont attendus dès le mois « 0 ».

Dans le modèle, l'exposition au risque au mois M est obtenue par l'application de la formule suivante :

$$LGD = \begin{cases} \frac{\sum_{j=M}^{11} \text{Reprise de commission}_j}{(1 + i_{mensuel})^j} & \text{si } 0 \leq M \leq 11 \\ 0 & \text{sinon} \end{cases}$$

où :

- $Reprise\ de\ commission_j = Prime\ mensuelle\ moyenne \times 12$
 $\times Commission\ d'acquisition$
 $\times (persistence_j - persistence_{j+1})$
 $\times \frac{\sum_{j=M}^{11} Prime\ mois_j}{\sum_{j=0}^{11} Prime\ mois_j}$
 $\times (\% \text{ de la quote} - part)$
- $Prime\ mois_j = Prime\ mensuelle\ moyenne \times (\% \text{ de la quote} - part) \times persistence_j$
- $persistence_j = \prod_{k=0}^j (1 - \text{taux de résiliation mensuel}_k)$

4.1.2. Hypothèses techniques

Le tableau ci-dessous donne les hypothèses techniques de l'application numérique :

% du réassureur dans la quote-part	70 %
Prime mensuelle moyenne	15 €
Commission d'acquisition	120 % de la prime de première année

Par ailleurs, deux hypothèses concernant la loi de résiliation ont été retenues :

Mois	Taux de résiliation mensuel (en %)	
	Hypothèse A	Hypothèse B
0	20	15
1	20	7
2	10	5
3	4	3
4	4	3
5	4	3
6	4	3
7	4	3
8	4	3
9	4	3
10	4	3
11	4	3
12	4	3

Dans les deux cas, l'hypothèse d'un taux de résiliation élevé les trois premiers mois a été retenue. En effet, ce choix a semblé cohérent avec la dynamique des résiliations observée sur les contrats souscrits en distribution directe où des taux de résiliation élevés ont été constatés au cours des premiers mois suivant la conclusion du contrat.

Il convient de noter qu'intuitivement, le risque de contrepartie semble corrélé au risque de résiliation car, plus les assurés renoncent à leur protection, plus le montant de la reprise de commission est important et le réassureur exposé au risque de défaut.

4.1.3. Hypothèses sur la nature des contreparties

Afin de déterminer le montant de fonds propres à immobiliser, pour un assuré, afin de faire face au risque de crédit, nous considérons des contreparties de qualités différentes. Ainsi, le modèle est appliqué à une cédante dont la notation est AAA, AA, A, BBB, BB, B ou CCC puis à un courtier dont les créances sont recouvrées sous un délai inférieur ou supérieur à 3 mois et avec une proportion égale de créances recouvrées avant et après 3 mois.

Section 4.2. Résultats de l'application

Dans cette section, nous présentons deux graphiques :

- L'évolution mensuelle de l'exposition au risque du réassureur en fonction de l'hypothèse de résiliation des contrats afin d'illustrer la façon dont la valeur du SCR évolue au cours du temps.
- Le SCR exprimé en pourcentage de l'exposition au risque (ou LGD) afin de comparer le niveau de risque représenté par les différentes contreparties.

Pour une contrepartie de qualité donnée, et quelle que soit l'hypothèse de résiliation des polices retenue, ce pourcentage reste le même au cours du temps. En effet, seul le SCR en valeur évolue car, le réassureur n'étant plus en risque au bout de 12 mois, l'exposition au risque diminue progressivement au cours du temps.

La confrontation des deux graphiques dévoilés dans cette section permet donc d'avoir une vision complète du risque de contrepartie pour le réassureur en fonction de l'hypothèse de résiliation des polices et de la nature de la contrepartie.

4.2.1. Evolution mensuelle de l'exposition au risque en fonction de l'hypothèse de résiliation des contrats

Le graphique ci-dessous illustre les résultats de la première étude :

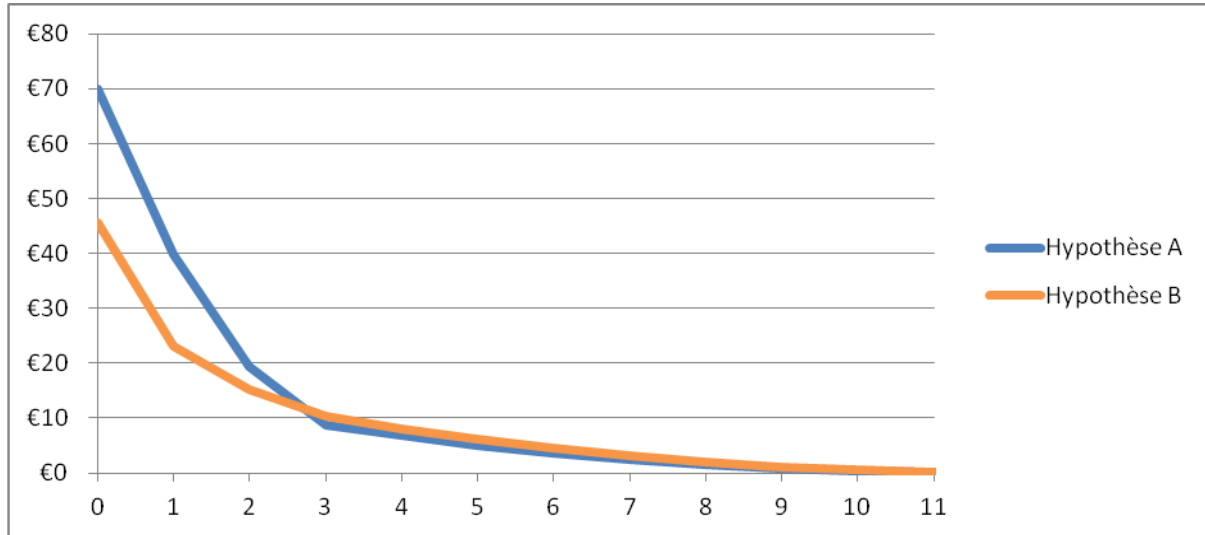


Figure 37: Evolution mensuelle de l'exposition au risque en fonction de l'hypothèse de résiliation

Plusieurs remarques peuvent être faites. Tout d'abord, l'exposition au risque diminue rapidement pour les deux hypothèses de résiliation testées. En effet, au début du troisième mois, la valeur du LGD passe en dessous de la barre de 10 € alors que les niveaux de départ étaient de 70,0 € pour l'hypothèse A et 45,7 € pour l'hypothèse B. Par ailleurs, au cours de ces trois premiers mois, l'exposition au risque pour l'hypothèse A est supérieure à celle de l'hypothèse B, ce qui confirme l'intuition selon laquelle un risque de résiliation précoce plus élevé est corrélé à une exposition au risque plus importante. Passés les trois premiers mois, les taux de résiliation se stabilisent à un niveau bas dans les deux hypothèses. Dès lors, nous notons une décroissance régulière de l'exposition au risque jusqu'à la fin de la période pendant laquelle le réassureur est en risque. De plus, l'exposition au risque devient plus faible dans l'hypothèse de résiliation A après le troisième mois. Toutefois, les niveaux d'exposition étant faibles, ce constat n'a pas de réelles conséquences pour le réassureur.

Pour conclure sur l'exposition au risque de contrepartie, deux idées sont à retenir. En premier lieu, le niveau de risque supporté par le réassureur devient faible à partir du troisième mois après la vente. Ensuite, un taux de résiliation plus sévère se traduit par une exposition au risque plus importante au cours des trois premiers mois.

4.2.2. Niveau du SCR en pourcentage du LGD pour différentes contreparties

Concernant la seconde étude, les résultats sont représentés dans le graphique suivant :

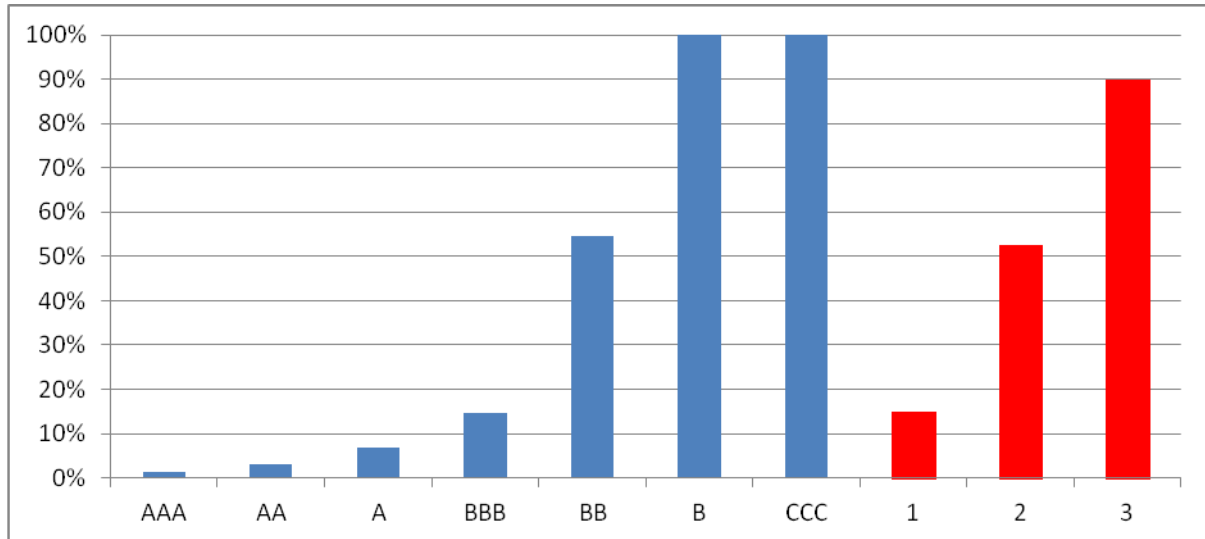


Figure 40 : Niveau du SCR en pourcentage du LGD selon la nature de la contrepartie

Dans ce graphique, les contreparties de type « cédante » faisant l'objet d'une notation sont en bleu alors que les contreparties de type « courtier » sont en rouge. Pour ces dernières contreparties, trois niveaux de qualité ont été définis. Le tableau suivant donne le détail de cette classification :

<i>Qualité du courtier</i>	<i>Signification de la notation</i>
1	Courtier dont les créances sont recouvrées sous un délai inférieur à 3 mois
2	Courtier dont les créances sont recouvrées pour moitié sous un délai inférieur à 3 mois et pour moitié sous un délai supérieur à 3 mois
3	Courtier dont les créances sont recouvrées sous un délai supérieur à 3 mois

Ce dernier graphique nous apporte plusieurs informations. Tout d'abord, les cédantes possédant une notation supérieure ou égale à A permettent au réassureur d'immobiliser une faible proportion du LGD pour couvrir le risque de contrepartie à leur égard (6,7 % du LGD à immobiliser pour une cédante notée A par exemple). Il est donc particulièrement intéressant pour le réassureur d'avoir des partenaires assureurs bien notés. En effet, à partir de la notation BBB, la proportion de LGD à immobiliser pour couvrir le SCR du module du risque de contrepartie croît très nettement pour atteindre 100 % pour les assureurs notés B ou CCC. En ce qui concerne le courtier, le niveau de SCR à mettre en place dépend beaucoup de sa capacité à honorer ses engagements sous un délai inférieur à 3 mois. Si cette condition est vérifiée, alors le courtier équivaut à une cédante de qualité BBB. Dans le cas contraire, le réassureur doit immobiliser nettement plus de fonds propres. Le réassureur prend en général un risque de crédit plus important si la structure le met directement en contact avec le courtier.

Section 4.3. Quelques mots pour conclure

L'illustration proposée avait pour but de mettre en avant un certain nombre de phénomènes relatifs au niveau de SCR requis pour le module de risque de contrepartie dans le cas d'un réassureur en fonction de différents critères. A l'issue de l'étude menée, il apparaît que la réforme Solvabilité II permet de prendre

efficacement en compte la nature de la contrepartie dans le calcul d'exigence de fonds propres pour financer une affaire de distribution directe de produits d'assurance.

Toutefois, le modèle présenté est une simplification de la réalité puisque nous avons uniquement analysé le risque de crédit lié à une unique contrepartie dans l'hypothèse d'une clause de reprise de commission initiale en cas de résiliation de l'assuré. Certains faits ont donc pu être occultés et une étude plus approfondie serait sans doute intéressante dans le cadre d'un mémoire traitant principalement de la question du risque de contrepartie. Néanmoins, nous avons réussi à confirmer des intuitions et à tirer certaines conclusions sur l'influence de la qualité de la contrepartie et de l'hypothèse de résiliation des polices retenue dans l'évaluation de ce risque pour le réassureur dans le cas d'une clause de reprise de commission initiale.

Ces informations seront particulièrement importantes pour le réassureur et ses partenaires lors de la détermination du choix de la structure de financement pour des contrats d'assurance issus de la distribution directe.

CONCLUSION DU MEMOIRE

Encore marginale aujourd'hui en France, la distribution directe de produits d'assurance connaît malgré tout une croissance progressive. Pour permettre l'essor de ce marché, le rôle du réassureur est particulièrement important. En effet, de par la nature de son métier et l'environnement international dans lequel il évolue, le réassureur possède une expérience précieuse en gestion du risque et des capitaux qu'il peut mettre au profit de partenaires souhaitant développer cette ligne d'affaire. La distribution directe, notamment à travers l'utilisation des techniques de marketing direct, représente un relais de croissance pour le réassureur.

Dans ce contexte, Transamerica Reinsurance développe cette ligne d'affaire, largement développée dans les pays anglo-saxons depuis la fin des années 1990, sur le marché français. A travers ce mémoire, nous avons cherché à montrer quels pouvaient être les risques et les opportunités inhérents à la distribution directe.

Nous avons tout d'abord présenté les caractéristiques de la distribution directe de produits d'assurance en France. Les techniques de marketing direct et télémarketing sont aujourd'hui bien connues et utilisées par l'ensemble des acteurs de la distribution directe. Se développe désormais chez les assureurs l'utilisation d'Internet et des nouveaux outils de communication dans des stratégies de distribution multi-canal qui permettent à l'assuré de choisir son mode de communication avec l'assureur en fonction de ses goûts et besoins. Toutefois, l'acceptation de la distribution directe n'est pas encore totalement entrée dans les mœurs et il reste un certain nombre de freins (aspects légaux, sélection des risques, réticence de l'assuré à souscrire sans contact direct avec son conseiller...) à un essor réel de la souscription à distance.

Nous nous sommes ensuite penchés sur l'étude de trois problématiques du réassureur lors de la mise en place d'une affaire de distribution directe. En effet, nous avons expliqué que ce mode de distribution sans intermédiaire soulève de nombreuses questions relatives à la conception, au financement et au suivi du risque des produits. En particulier, la prise en charge du financement de coûts d'acquisition importants à la souscription du contrat représente pour le réassureur la principale problématique.

L'étude de la rentabilité sous différents référentiels comptables a fait apparaître l'importance de l'utilisation de coûts d'acquisition différés pour obtenir un profit régulier dans le temps. La méthode comptable imposée par la législation peut donc considérablement changer la mesure de la rentabilité d'une affaire et la reconnaissance des profits au cours du temps. Le réassureur, de par son envergure internationale, peut assumer le financement des coûts d'acquisition et trouver les meilleurs arbitrages comptables en fonction de la législation des différents pays dans lesquels il opère.

Nous avons ensuite étudié le risque de chute pour des contrats distribués en utilisant le marketing direct. En effet, l'importance du financement initial à la souscription d'un nouvel assuré impose au réassureur de rechercher des clients « fidèles ». A travers la sélection et l'étude de variables explicatives, nous avons tenté de faire ressortir les meilleures cibles d'assurés afin de contrôler les coûts d'acquisition et de piloter la rentabilité du portefeuille. L'utilisation de modèles prédictifs tels que l'algorithme CART et la régression logistique sur deux portefeuilles réels a permis d'obtenir des résultats satisfaisants dans la prédiction des résiliations et la mise en avant de profils d'assurés à risque. De plus, ces modèles se sont révélés plus efficaces en termes de temps et de qualité qu'une analyse uni-variée réalisée de manière intuitive. Il semble alors plus pertinent de développer et d'utiliser conjointement l'algorithme CART et la régression logistique avant d'entreprendre une analyse descriptive des résiliations. Cependant, dans le cadre de ce mémoire, ces modèles ont montré certaines limites en partie dues au manque de données sur les individus ayant résilié.

Enfin, l'étude du risque de contrepartie dans le cadre de la directive européenne Solvabilité II a montré comment le réassureur pouvait prendre en compte efficacement la nature de la contrepartie dans le calcul d'exigence de fonds propres pour financer une affaire de distribution directe de produits d'assurance. Cette information est particulièrement importante pour le réassureur lors de la détermination du choix de la structure de financement.

En conclusion, ce mémoire a permis de mettre en lumière le rôle particulièrement important du réassureur dans le développement de la distribution directe. En effet, il constitue souvent un support financier et stratégique pour les acteurs, assureurs ou distributeurs.

BIBLIOGRAPHIE

- « *Etude des stratégies de distribution multi-canal des assureurs français* », **Accenture**.
Consultable sur <http://www.accenture.com/fr-fr/Documents/PDF/EnqueteMulticanal.pdf>
- « *Marketing direct, le plébiscite des assureurs* », **L'Argus de l'assurance**, N°6 826, 21/03/2003
- « *La vente à distance soumise à ordonnances* », **L'Argus de l'assurance**, N°6 962, 27/01/2006
- « *Vente directe, l'Europe du Nord montre la voie au Sud* », **L'Argus de l'assurance**, N°7 076, 23/05/2008
- « *L'assurance à l'heure du multi-canal et d'Internet* », **L'Argus de l'assurance**, N°7 166, 30/04/2010
- « *Un nouvel entrant sur le marché des comparateurs d'assurance en ligne* », **L'Argus de l'assurance**, Article Web, 07/05/2010
- « *Les services financiers sur Internet séduisent* », **L'Argus de l'assurance**, Article Web, 11/05/2010
- « *Assurance en ligne : le Web n'est encore qu'une vitrine* », **L'Argus de l'assurance**, N°7 173, 11/06/2010
- « *L'avenir des agents passe par Internet* », **L'Argus de l'assurance**, N°7 134, 11/09/2009
- « *La souscription en ligne monte en puissance* », **L'Argus de l'assurance**, N°7 183, 17/09/2010
- « *La M-assurance fait rêver, mais des freins demeurent* », **L'Argus de l'assurance**, N°7 182, 10/09/2010
- « *L'affinitaire a le vent en poupe* », **L'Argus de l'assurance**, N°7 206, 04/03/2011
- « *Rapport annuel FFSA 2010 - L'essentiel de la FFSA – Tableaux de bord* », **FFSA**, 30/06/2011, p32.
Consultable sur http://www.ffsa.fr/sites/upload/docs/application/pdf/2011-06/raffsa2010fessentielffsa2010-3_2011-06-30_17-16-49_557.pdf
- « *New Accounting Rule – To DAC or not to DAC* », **The Messenger**, Mars 2012
- « *Les défis de la télé-souscription* », **Newsletter Scor**, Juin 2009
- American Council of Life Insurance/Health Insurance Association of America (1986)**, « *Introduction to Direct Response Marketing in the Life and Health Insurance Industry* », Washington (DC), p. 33-34
- Balakrishnan N. (1991)**, « *Handbook of the Logistic Distribution* », Marcel Dekker, Inc
- Bernheim Y. (2000)**, « *L'essentiel des US GAAP, 2ème édition* », Maxima, p19-25
- Breiman L., Friedman J., Olshen R. A., Stone C. J. (1984)**, « *Classification and Regression Trees* », Chapman and Hall
- Breiman L. (2001)**, « *Random forests* », Machine Learning (45), p5-32
- Herget R. (2006)**, « *US GAAP for life insurers* », Society of Actuaries, p81-122
- McCullagh P., Nelder J. A. (1989)**, « *Generalized linear models, 2nd ed.* », Chapman and Hall

Milhaud X., Loisel S., Maume-Deschamps V., (2010), « *Surrender triggers in life insurance: classification and risk predictions* », Hal

Planchet F., Théron P., (2006), « *Modèles de durée, Applications actuarielles* », Economica

Rubin H. (2000), « *Dictionary of Insurance Terms, 4th edition* », Barron's Educational Series, p257

Vernimmen P., Quiry P., Le Fur Y. (2011), « *Finance d'entreprise 2011* », Dalloz, p377, 438, 441

SITES INTERNET DE REFERENCES

Site de l'Argus de l'Assurance : <http://www.argusdelassurance.com/>

Site de Breiman sur les forêts aléatoires : <http://www.stat.berkeley.edu/users/breiman/RandomForests/>

Site de communiqués de presse relatifs à l'univers des assurances : <http://assurances.annuairecommuniques.com/>

Site permettant l'accès au droit de l'Union Européenne et à d'autres documents considérés comme publics : <http://eur-lex.europa.eu/fr/index.htm>

Site de la Fédération du E-commerce et de la Vente à Distance : <http://www.fevad.com/>

Site de la Fédération Française des Sociétés d'Assurances : <http://www.ffsa.fr/>

Site de l'International Accounting Standards Board : www.ifrs.org/

Site Légifrance : www.legifrance.gouv.fr/

Site de Molitor Consult, spécialistes du marketing et de la communication en assurance : www.molitorconsult-assurance.com

Site de Pierre Théron contenant de nombreuses ressources sur les normes IFRS : <http://www.pierretherond.fr/>

Site du projet R : <http://www.r-project.org/>

ANNEXES

Annexe 1 : Articles de loi concernant la commercialisation à distance de contrats d'assurance

Article L112-2-1 [En savoir plus sur cet article...](#)

Modifié par [Ordonnance n°2010-76 du 21 janvier 2010 - art. 18 \(V\)](#)

I.-1° La fourniture à distance d'opérations d'assurance à un consommateur est régie par les dispositions du présent livre et par celles des dispositions des sous-sections 2 et 3 de la section 2 du chapitre Ier du titre II du livre Ier du code de la consommation, à l'exception des articles L. 121-20-10, L. 121-20-12 et L. 121-20-17, ci-après reproduites :

" Sous-section 2 : Dispositions particulières aux contrats portant sur des services financiers

" Art.[L. 121-20-8](#)

" La présente sous-section régit la fourniture de services financiers à un consommateur dans le cadre d'un système de vente ou de prestation de services à distance organisé par le fournisseur ou par un intermédiaire qui, pour ce contrat, utilise exclusivement une ou plusieurs techniques de communication à distance jusqu'à, et y compris, la conclusion du contrat. "

" Elle s'applique aux services mentionnés aux livres Ier à III et au titre V du livre V du code monétaire et financier ainsi que les opérations pratiquées par les entreprises régies par le code des assurances, par les mutuelles et unions régies par le livre II du code de la mutualité et par les institutions de prévoyance et unions régies par le titre III du livre IX du code de la sécurité sociale sans préjudice des dispositions spécifiques prévues par ces codes. "

" Art.[L. 121-20-9](#)

" Pour les contrats portant sur des services financiers comportant une première convention de service suivie d'opérations successives ou d'une série d'opérations distinctes, de même nature, échelonnées dans le temps, les dispositions de la présente sous-section ne s'appliquent qu'à la première convention de service. Pour les contrats renouvelables par tacite reconduction, les dispositions de la présente sous-section ne s'appliquent qu'en vue et lors de la conclusion du contrat initial.

" En l'absence de première convention de service, lorsque les opérations successives ou distinctes, de même nature, échelonnées dans le temps, sont exécutées entre les mêmes parties, les dispositions de l'article L. 121-20-10 ne sont applicables qu'à la première opération. Cependant, lorsqu'aucune opération de même nature n'est effectuée pendant plus d'un an, ces dispositions s'appliquent à l'opération suivante, considérée comme une première opération. "

" Art.[L. 121-20-11](#)

" Le consommateur doit recevoir, par écrit ou sur un autre support durable à sa disposition et auquel il a accès en temps utile et avant tout engagement, les conditions contractuelles ainsi que les informations mentionnées à l'article L. 121-20-10. Le fournisseur peut remplir ses obligations au titre de l'article L. 121-20-10 et du présent article par l'envoi au consommateur d'un document unique, à la condition qu'il s'agisse d'un support écrit ou d'un autre support durable et que les informations mentionnées ne varient pas jusqu'à et y compris la conclusion du contrat.

" Le fournisseur exécute ses obligations de communication immédiatement après la conclusion du contrat, lorsque celui-ci a été conclu à la demande du consommateur en utilisant une technique de communication à distance ne permettant pas la transmission des informations précontractuelles et contractuelles sur un support papier ou sur un autre support durable.

" A tout moment au cours de la relation contractuelle, le consommateur a le droit, s'il en fait la demande, de recevoir les conditions contractuelles sur un support papier. En outre, le consommateur a le droit de changer les techniques de communication à distance utilisées, à moins que cela ne soit incompatible avec le contrat à distance conclu ou avec la nature du service financier fourni. "

" Art.[L. 121-20-13](#)

" I.-Les contrats pour lesquels s'applique le délai de rétractation mentionné à l'article L. 121-20-12 ne peuvent recevoir de commencement d'exécution par les parties avant l'arrivée du terme de ce délai sans l'accord du consommateur. Lorsque celui-ci exerce son droit de rétractation, il ne peut être tenu qu'au paiement proportionnel du service financier effectivement fourni, à l'exclusion de toute pénalité.

" Le fournisseur ne peut exiger du consommateur le paiement du service mentionné au premier alinéa que s'il peut prouver que le consommateur a été informé du montant dû, conformément à l'article L. 121-20-10. Toutefois, il ne peut pas exiger ce paiement s'il a commencé à exécuter le contrat avant l'expiration du délai de rétractation sans demande préalable du consommateur.

" Pour les contrats de crédit à la consommation prévus au chapitre Ier du titre Ier du livre III, même avec l'accord du consommateur, ils ne peuvent recevoir de commencement d'exécution durant les sept premiers jours, sauf s'agissant des contrats de crédit affecté mentionnés au IV de l'article L. 121-20-12, qui ne peuvent recevoir de commencement d'exécution durant les trois premiers jours.

" II.-Le fournisseur est tenu de rembourser au consommateur dans les meilleurs délais et au plus tard dans les trente jours toutes les sommes qu'il a perçues de celui-ci en application du contrat, à l'exception du montant mentionné au premier alinéa du I. Ce délai commence à courir le jour où le fournisseur reçoit notification par le consommateur de sa volonté de se rétracter. Au-delà du délai de trente jours, la somme due est, de plein droit, productive d'intérêts au taux légal en vigueur.

" Le consommateur restitue au fournisseur dans les meilleurs délais et au plus tard dans les trente jours toute somme et tout bien qu'il a reçus de ce dernier. Ce délai commence à courir à compter du jour où le consommateur communique au fournisseur sa volonté de se rétracter. "

" Art.[L. 121-20-14](#)

" Les dispositions de l'article L. 34-5 du code des postes et communications électroniques, reproduites à l'article L. 121-20-5, sont applicables aux services financiers.

" Les techniques de communication à distance destinées à la commercialisation de services financiers autres que celles mentionnées à l'article L. 34-5 du code des postes et communications électroniques ne peuvent être utilisées que si le consommateur n'a pas manifesté son opposition.

" Les mesures prévues au présent article ne doivent pas entraîner de frais pour le consommateur. "

" Sous-section 3 : Dispositions communes

" Art.[L. 121-20-15](#)

" Lorsque les parties ont choisi la loi d'un Etat non membre de la Communauté européenne pour régir le contrat, le juge devant lequel est invoquée cette loi est tenu d'en écarter l'application au profit des dispositions plus protectrices de la loi de la résidence habituelle du consommateur assurant la transposition de la directive 97 / 7 / CE du Parlement européen et du Conseil du 20 mai 1997 concernant la protection des consommateurs en matière de contrats à distance et de la directive 2002 / 65 / CE du Parlement européen et du Conseil du 23 septembre 2002, concernant la commercialisation à distance de services financiers auprès des consommateurs, lorsque le contrat présente un lien étroit avec le territoire d'un ou plusieurs Etats membres de la Communauté européenne ; cette condition est présumée remplie si la résidence des consommateurs est située dans un Etat membre. "

" Art.[L. 121-20-16](#)

" Les dispositions de la présente section sont d'ordre public ; "

2° Pour l'application du 1°, il y a lieu d'entendre :

a) " Le souscripteur, personne physique, qui agit à des fins qui n'entrent pas dans le cadre de son activité commerciale ou professionnelle " là où est mentionné " le consommateur " ;

b) " L'assureur ou l'intermédiaire d'assurance " là où est mentionné " le fournisseur " ;

- c) " Le montant total de la prime ou cotisation " là où est mentionné " le prix total " ;
- d) " Droit de renonciation " là où est mentionné " le droit de rétractation " ;
- e) " Le II de l'article L. 112-2-1 du code des assurances " là où est mentionné " l'article L. 121-20-12 " ;
- f) " Le III de l'article L. 112-2-1 du code des assurances " là où est mentionné " l'article L. 121-20-10 " ;

3° Pour l'application de l'article L. 121-20-11 du code de la consommation, les conditions contractuelles doivent comprendre, outre les informations prévues selon les cas à l'article L. 112-2 ou à l'article L. 132-5-1, un modèle de lettre destiné à faciliter l'exercice du droit de renonciation lorsque ce droit existe.

II.-1° Toute personne physique ayant conclu à des fins qui n'entrent pas dans le cadre de son activité commerciale ou professionnelle un contrat à distance dispose d'un délai de quatorze jours calendaires révolus pour renoncer, sans avoir à justifier de motif ni à supporter de pénalités. Ce délai commence à courir :

- a) Soit à compter du jour où le contrat à distance est conclu ;
- b) Soit à compter du jour où l'intéressé reçoit les conditions contractuelles et les informations, conformément à l'article L. 121-20-11 du code de la consommation, si cette dernière date est postérieure à celle mentionnée au a ;

2° Toutefois, en ce qui concerne les contrats d'assurance vie, le délai précité est porté à trente jours calendaires révolus. Ce délai commence à courir :

- a) Soit à compter du jour où le contrat à distance est conclu ;
- b) Soit à compter du jour où l'intéressé reçoit les conditions contractuelles et les informations, conformément à l'article L. 121-20-11 du code de la consommation, si cette dernière date est postérieure à celle mentionnée au a ;

3° Le droit de renonciation ne s'applique pas :

- a) Aux polices d'assurance de voyage ou de bagage ou aux polices d'assurance similaires à court terme d'une durée inférieure à un mois ;
- b) Aux contrats d'assurance mentionnés à l'article L. 211-1 du présent code ;
- c) Aux contrats exécutés intégralement par les deux parties à la demande expresse du consommateur avant que ce dernier n'exerce son droit de renonciation.

III.-En temps utile avant la conclusion à distance d'un contrat le souscripteur reçoit les informations suivantes :

1° La dénomination de l'entreprise d'assurance contractante, l'adresse de son siège social, lorsque l'entreprise d'assurance est inscrite au registre du commerce et des sociétés, son numéro d'immatriculation, les coordonnées de l'autorité chargée de son contrôle ainsi que, le cas échéant, l'adresse de la succursale qui propose la couverture ;

2° Le montant total de la prime ou cotisation ou, lorsque ce montant ne peut être indiqué, la base de calcul de cette prime ou cotisation permettant au souscripteur de vérifier celle-ci ;

3° La durée minimale du contrat ainsi que les garanties et exclusions prévues par celui-ci ;

4° La durée pendant laquelle les informations fournies sont valables, les modalités de conclusion du contrat et de paiement de la prime ou cotisation ainsi que l'indication, le cas échéant, du coût supplémentaire spécifique à l'utilisation d'une technique de commercialisation à distance ;

5° L'existence ou l'absence d'un droit à renonciation et, si ce droit existe, sa durée, les modalités pratiques de son exercice notamment l'adresse à laquelle la notification de la renonciation doit être envoyée. Le souscripteur doit également être informé du montant de prime ou de cotisation que l'assureur peut lui réclamer en contrepartie de la prise d'effet de la garantie, à sa demande expresse, avant l'expiration du délai de renonciation ;

6° La loi sur laquelle l'assureur se fonde pour établir les relations précontractuelles avec le consommateur ainsi que la loi applicable au contrat et la langue que l'assureur s'engage à utiliser, avec l'accord du souscripteur, pendant la durée du contrat ;

7° Les modalités d'examen des réclamations que le souscripteur peut formuler au sujet du contrat y compris, le cas échéant, l'existence d'une instance chargée en particulier de cet examen, sans préjudice pour lui d'intenter une action en justice ainsi que, le cas échéant, l'existence de fonds de garantie ou d'autres mécanismes d'indemnisation.

Les informations sur les obligations contractuelles communiquées en phase précontractuelle doivent être conformes à la loi applicable au contrat.

Ces informations, dont le caractère commercial doit apparaître sans équivoque, sont fournies de manière claire et compréhensible par tout moyen adapté à la technique de commercialisation à distance utilisée.

IV.-L'assureur doit également indiquer, pour les contrats d'assurance vie les informations mentionnées à l'article L. 132-5-1, notamment le montant maximal des frais qu'il peut prélever et, lorsque les garanties de ces contrats sont exprimées en unités de compte, les caractéristiques principales de celles-ci. Dans ce dernier cas, il doit en outre préciser qu'il ne s'engage que sur le nombre des unités de compte et non sur leur valeur qui peut être sujette à des fluctuations à la hausse comme à la baisse.

V.-Un décret en Conseil d'Etat fixe les informations communiquées au souscripteur en cas de communication par téléphonie vocale.

VI.-Les infractions aux dispositions du présent article sont constatées et sanctionnées par l'Autorité de contrôle prudentiel dans les conditions prévues à la section 2 du chapitre unique du titre Ier du livre III.

Les infractions constituées par l'absence matérielle des éléments d'information prévus au III du présent article, ainsi que le refus de l'assureur de rembourser le souscripteur personne physique dans les conditions fixées à l'article L. 121-20-13 du code de la consommation peuvent également être constatées et poursuivies dans les conditions prévues à l'article L. 121-20-17 du même code.

Les conditions d'application du présent article sont définies en tant que de besoin par décret en Conseil d'Etat.

Annexe 2 : Exemple de grille de sélection emprunteur

Âge à l'adhésion	Montants assurés & Conditions				
< 45	< 200 000 Questionnaire Simplifié	200 001 à 350 000 Questionnaire Complet	350 001 à 500 000 Questionnaire Complet Rapport Médical & Bilan Sanguin	> 500 000 < 1 000 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et ECG repos	> 1 000 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et épreuve ECG d'effort
45 à 55	< 125 000 Questionnaire Simplifié	125 001 à 250 000 Questionnaire Complet	250 001 à 400 000 Questionnaire Complet Rapport Médical & Bilan Sanguin	400 001 à 750 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et ECG repos	> 750 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et épreuve ECG d'effort
56 à 65	< 75 000 Questionnaire Simplifié	75 001 à 175 000 Questionnaire Complet	175 001 à 300 000 Questionnaire Complet Rapport Médical & Bilan Sanguin	300 001 à 500 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et ECG repos	> 500 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et épreuve ECG d'effort
> 65	< 50 000 Questionnaire Simplifié "Senior"	50 001 à 100 000 Questionnaire Complet Rapport Médical & Bilan Sanguin	100 001 à 250 000 Questionnaire Complet Rapport Médical & Bilan Sanguin (PSA pour les hommes).	250 001 à 400 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et ECG repos	> 400 000 Questionnaire Complet Rapport Médical & Bilan Sanguin et épreuve ECG d'effort
Précisions	Test VIH : > 400 000€ jusqu'à 60 ans Rapport médical : Etabli par le Médecin Généraliste Bilan sanguin : Cholestérol HDL, LDL, triglycérides, vitesse de sédimentation (VS), glycémie à jeun, HbA1c, créatinine, acide urique, transaminases (SGPT et SGOT), gamma GT, antigène HBs, sérologie HCV et test de la cotinine. Epreuve d'effort : Rapport cardiovasculaire et épreuve d'effort avec compte rendu Test Cotinine : > 400 000 € non fumeur. > 45 ans : PSA avec bilan sanguin pour les hommes Dès 1 000 000 € : Questionnaire financier				

Annexe 3 : Articles de loi concernant les commissions et frais d'acquisition à amortir

Article R.332-33 [En savoir plus sur cet article...](#)

Modifié par [Décret n° 95-153 du 7 février 1995 - art. 4 JORF 14 février 1995](#)

Les entreprises pratiquant les opérations mentionnées au 2° ou au 3° de l'article L. 310-1 inscrivent à l'actif du bilan la fraction non imputable à l'exercice des frais d'acquisition des contrats constatés en charge de l'exercice. La période d'imputation des frais d'acquisition ne peut s'étendre au-delà de la date à laquelle le souscripteur peut exercer son droit de résiliation ou de non-reconduction ni, lorsque les frais à reporter sont des commissions payables à chaque échéance de prime, au-delà de la prochaine échéance de prime.

Le montant reporté est calculé contrat par contrat ou sur la base de méthodes statistiques, dans les mêmes conditions et selon les mêmes méthodes que la provision pour primes non acquises ; il ne peut faire l'objet d'une augmentation ultérieure ; il est amorti linéairement sur la durée restant à courir entre la date de l'inventaire et la fin de la période d'imputation des frais, et au maximum sur cinq exercices ; il est amorti en totalité en cas de résiliation anticipée, d'annulation ou de transfert du contrat.

La fraction non imputable à l'exercice des commissions des réassureurs est également inscrite au bilan ; le montant reporté est calculé et repris en compte de résultat selon les mêmes méthodes que celles retenues pour les frais d'acquisition des affaires brutes correspondantes.

Article R.332-35 [En savoir plus sur cet article...](#)

Modifié par [Ordonnance n° 2010-76 du 21 janvier 2010 - art. 18 \(V\)](#)

Les entreprises pratiquant les opérations mentionnées au 1° de l'article L. 310-1 inscrivent à l'actif du bilan les frais d'acquisition à reporter en fonction de la durée de vie résiduelle des contrats. La méthode retenue est décrite dans l'annexe mentionnée au premier alinéa de l'article R. 341-3. Le montant des frais d'acquisition ainsi reportés est au plus égal à l'écart entre les montants de provisions mathématiques inscrites au bilan conformément à l'article L. 331-1 et le montant des provisions mathématiques qui seraient à inscrire si les chargements d'acquisition n'étaient pas pris en compte dans les engagements des assurés. Le montant de cet écart ainsi que le calcul des frais d'acquisition reportés doivent pouvoir être justifiés à tout moment auprès de l'Autorité de contrôle prudentiel. Les frais d'acquisition reportés sont admis en représentation des provisions techniques.

Annexe 4 : Table des correspondances pour l'algorithme CART sur le portefeuille XW

AGE	SEX	SF	GAR	MV	NMP	PMM	lettre correspondante
18-34	F	CELIBATAIRE	15-25	200812-200905	0	<15	a
35-44	H	FAMILLE	40-60	200906-200911	1->3	>30	b
45-54			75-100	200912-201005	10	15-30	c
55-65				201006-201010	11		d
					12->23		e

Tableau 25 : Table des correspondances de lecture des arbres de décision binaires (portefeuille XW)

Annexe 5 : Table des correspondances pour l'algorithme CART sur le portefeuille YZ

AGE	SEX	SF	GAR	MV	NMP	PMM	CAMP	lettre correspondante
18-34	F	CELIBATAIRE	20-30	200911-201001	0	<15	CIRC	a
35-44	H	FAMILLE	50-60	201002-201004	1->3	>30	SFER	b
45-54			90-120	201005-201007	4->10	15-30	VENC	c
55-65				201008-201009	11			d

Tableau 26 : Table des correspondances de lecture des arbres de décision binaires (portefeuille YZ)

Annexe 6 : Sorties obtenues avec le logiciel R pour l'algorithme CART dans l'étude du portefeuille XW sans la modalité « 0 » prime payée de la variable « NMP2 »

	Prédit $Y = 0$ (résiliation)	Prédit $Y = 1$ (non résiliation)
Observé $Y = 0$	1 339	356
Observé $Y = 1$	203	6 711

Tableau 27 : Matrice d'erreur pour T_{Max} sur l'échantillon de validation
Source : Propres calculs à partir de données réelles

CP	nsplit	rel error	xerror	xstd	CP	nsplit	rel error	xerror	xstd
1.5163e-01	0	1.00000	1.00000	0.0153123	2.5540e-04	17	0.34063	0.34764	0.0097227
9.7928e-02	2	0.69673	0.69673	0.0132474	1.9459e-04	28	0.33742	0.34880	0.0097378
7.5598e-02	4	0.50088	0.50088	0.0114801	1.4594e-04	31	0.33684	0.34968	0.0097491
7.0053e-02	5	0.42528	0.43637	0.0107906	9.7295e-05	39	0.33567	0.35435	0.0098092
2.0432e-03	6	0.35522	0.35522	0.0098204	7.2971e-05	54	0.33392	0.35581	0.0098278
1.3865e-03	10	0.34705	0.34705	0.0097151	5.8377e-05	68	0.33275	0.35785	0.0098539
2.9189e-04	14	0.34151	0.34267	0.0096581	0.0000e+00	78	0.33217	0.35931	0.0098724

Tableau 28 : Table des paramètres de complexité
Source : Propres calculs à partir de données réelles

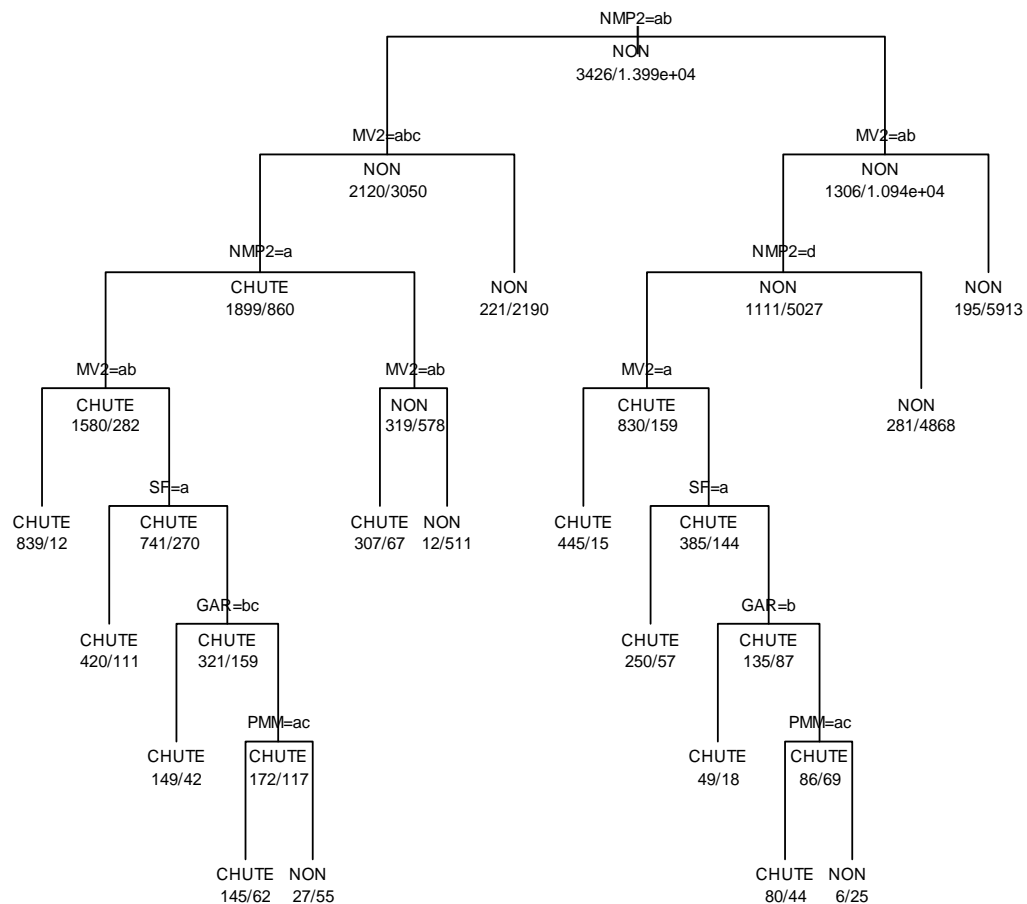


Figure 41 : Arbre de classification élagué
Source : Propres calculs à partir de données réelles

	Prédit $Y = 0$ (résiliation)	Prédit $Y = 1$ (non résiliation)
Observé $Y = 0$	1 363	332
Observé $Y = 1$	210	6 704

Tableau 29 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation
Source : Propres calculs à partir de données réelles

	Prédit $Y = 0$ (résiliation)	Prédit $Y = 1$ (non résiliation)
Observé $Y = 0$	4 014	1 107
Observé $Y = 1$	631	20 273

Tableau 30 : Matrice d'erreur du classificateur « randomForest »
Source : Propres calculs à partir de données réelles

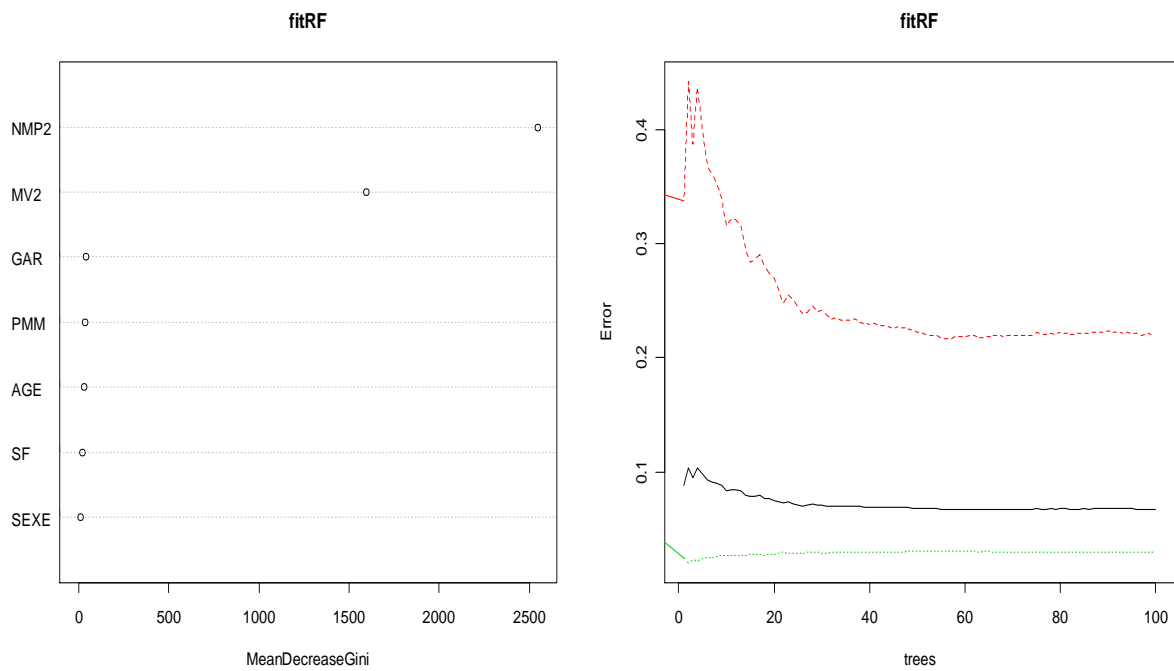


Figure 42 : Résultats de l'utilisation de la fonction « randomForest »
Source : Propres calculs à partir de données réelles

TABLE DES FIGURES

Figure 1 : Structure de distribution en marketing direct	23
Figure 2 : Relations entre les différentes parties dans un contrat d'assurance conclu à distance	35
Figure 3 : Approche par les flux du calcul des provisions techniques proposée par l'exposure draft	41
Figure 4: Résultats après impôt.....	48
Figure 5: Résultats en pourcentage des primes	49
Figure 6: Evolution mensuelle du nombre de polices des portefeuilles XW (à gauche) et YZ (à droite) ..	54
Figure 7 : Résiliations sur le portefeuille XW	58
Figure 8 : Résiliations sur le portefeuille YZ.....	58
Figure 9 : Comparaison de la persistance après le premier paiement par tranche d'âge	59
Figure 10 : Comparaison de la persistance après le premier paiement par tranche d'âge	60
Figure 11 : Répartition des polices par tranche d'âge	61
Figure 12 : Comparaison de la persistance après le premier paiement selon le sexe par rapport au portefeuille global en XW.....	61
Figure 13 : Comparaison de la persistance après le premier paiement selon le sexe par rapport au portefeuille global en YZ.....	62
Figure 14 : Répartition des polices par sexe	62
Figure 15 : Comparaison de la persistance après le premier paiement selon la situation familiale.....	63
Figure 16 : Comparaison de la persistance après le premier paiement selon la situation familiale	64
Figure 17 : Répartition des polices selon la situation familiale	65
Figure 18 : Répartition des primes selon la situation familiale.....	65
Figure 19 : Comparaison de la persistance après le premier paiement selon le montant de prestation	66
Figure 20 : Comparaison de la persistance après le premier paiement selon le montant de prestation	67
Figure 21 : Répartition des polices selon le montant de l'indemnité.....	68
Figure 22 : Répartition des primes selon le montant de l'indemnité	68
Figure 23 : Répartition des polices selon la campagne	69
Figure 24 : Comparaison de la persistance après le premier paiement selon la campagne.....	69
Figure 25 : Comparaison de la persistance à la fin de la période d'observation en fonction du mois de vente.....	70
Figure 26 : Etapes successives de l'algorithme CART.....	74
Figure 27 : Taux d'erreur de la validation croisée en fonction de la taille de l'arbre et du coût de complexité.....	85
Figure 28 : Arbre de classification élagué	85
Figure 29 : Résultats de l'utilisation de la fonction « randomForest »	87
Figure 30 : Taux d'erreur de la validation croisée en fonction de la taille de l'arbre et du coût de complexité.....	89
Figure 31 : Arbre de classification élagué	90
Figure 32 : Résultats de l'utilisation de la fonction « randomForest »	92
Figure 33 : Principe de la fonction « <i>stepAIC()</i> »	103
Figure 34 : Eléments de la fonction « <i>summary()</i> » d'une régression logistique.....	103
Figure 35 : Résultats issus de la fonction « <i>drop1()</i> »	104
Figure 36 : Eléments de la fonction « <i>stepAIC()</i> ».....	108
Figure 37 : Eléments de la fonction « <i>summary()</i> » d'une régression logistique.....	109
Figure 38 : Résultats issus de la fonction « <i>drop1()</i> »	110
Figure 39 : Structure générale du SCR	117

Figure 40 : Niveau du SCR en pourcentage du LGD selon la nature de la contrepartie.....	124
Figure 41 : Arbre de classification élagué	135
Figure 42 : Résultats de l'utilisation de la fonction « randomForest »	136

TABLE DES TABLEAUX

Tableau 1 : Les cotisations selon la forme de distribution (en %)	9
Tableau 2 : Principales hypothèses techniques retenues pour l'application numérique	42
Tableau 3 : Structure générale du modèle comptable annuel	45
Tableau 4 : Comparaison des scenarii étudiés dans l'application	47
Tableau 5 : Principales caractéristiques des tables « contrats »	53
Tableau 6 : Synthèse sur les portefeuilles XW et YZ	55
Tableau 7: Description des variables utilisées dans l'implémentation de l'algorithme CART	83
Tableau 8 : Matrice d'erreur pour TMax sur l'échantillon de validation	84
Tableau 9 : Table des paramètres de complexité	84
Tableau 10 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation	87
Tableau 11 : Matrice d'erreur du classificateur « randomForest »	87
Tableau 12 : Matrice d'erreur pour TMax sur l'échantillon de validation	89
Tableau 13 : Table des paramètres de complexité	89
Tableau 14 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation	91
Tableau 15 : Matrice d'erreur du classificateur « randomForest »	91
Tableau 16 : Description des principales variables explicatives pour la régression logistique	101
Tableau 17 : Synthèse des <i>odd-ratios</i> obtenus de façon empirique et en utilisant le modèle RL2.....	105
Tableau 18 : Nombre de polices en fonction du statut de la police et de la tranche d'âge	105
Tableau 19 : Matrice d'erreur pour le modèle RL2 sur l'échantillon de validation.....	106
Tableau 20 : Critères de performance de la prédiction	107
Tableau 21 : Synthèse des <i>odd-ratios</i> obtenus de façon empirique et en utilisant le modèle RLE2.....	111
Tableau 22 : Matrice d'erreur pour le modèle RLE2 sur l'échantillon de validation	111
Tableau 23 : Critères de performance de la prédiction	111
Tableau 24: Probabilités de défaut selon la notation	118
Tableau 25 : Table des correspondances de lecture des arbres de décision binaires (portefeuille XW)...	134
Tableau 26 : Table des correspondances de lecture des arbres de décision binaires (portefeuille YZ)	134
Tableau 27 : Matrice d'erreur pour TMax sur l'échantillon de validation	135
Tableau 28 : Table des paramètres de complexité	135
Tableau 29 : Matrice d'erreur pour l'arbre élagué sur l'échantillon de validation	136
Tableau 30 : Matrice d'erreur du classificateur « randomForest »	136