

**Mémoire présenté devant le Centre d'Etudes Actuarielles  
pour l'obtention du diplôme  
du Centre d'Etudes Actuarielles  
et l'admission à l'Institut des Actuaire  
le : 22 juin 2011**

Par : Céline BOUQUET et Philippe MENARD

Titre : Assurance Automobile - Optimisation des ressources à l'échéance

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

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présent du jury de  
l'Institut des Actuaire :

signature

Entreprise :

Membres présents du jury du  
Centre d'Etudes Actuarielles :

Thomas BEHAR

Vincent DAMAS

Gérard CROSET

Arnaud COHEN

Jean-Pierre DIAZ

Brigitte DUBUS

Paul ESMEIN

Michel FROMENTEAU

Benoît HUGONIN

Christophe IZART

Pierre PETAUTON

Florence PICARD -

Christian-Yann ROBERT

Secrétariat :

Bibliothèque :

Nom : MMA

Signature :

Directeur de mémoire en entreprise :

14 bd Marie et Alexandre Oyon  
LE MANS

Nom : Hervé GANACHAUD

Signature :

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

Signature du responsable entreprise

Signature des candidats

Le travail de ce mémoire n'aurait pas été possible sans le soutien et la patience de notre entourage, nous les en remercions vivement.

Nous tenons tout particulièrement à remercier :

La société MMA pour la mise à disposition de ses données et logiciels, indispensables à la réalisation,

Messieurs Oliviers JARRY, Directeur Central IARD Grand Public, et Hervé GANACHAUD, Responsable de l'actuariat Grand Public, pour les réflexions que nous avons partagés sur ce sujet,

M. Olivier PETON, HDR à l'Ecole des Mines de Nantes, pour son aimable concours pédagogique sur les méthodes d'optimisation.

# Table des matières

<b>Résumé</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
1.1 Problématique . . . . .	5
1.1.1 Contexte général . . . . .	5
1.1.2 Contexte en assurance Automobile . . . . .	6
1.1.3 Bonus-Malus légal et Bonus MMA : exemple de mécanisme de fidélisation	9
1.1.4 Comportement de résiliation en assurance automobile . . . . .	13
1.1.5 Les enjeux de la résiliation en assurance automobile pour MMA . . . . .	15
1.2 Cadre de l'étude . . . . .	17
1.2.1 Objet de l'étude . . . . .	17
1.2.2 Périmètre de l'étude . . . . .	18
1.2.3 Détermination de la prime technique et de la cotisation à l'échéance . . .	19
1.2.4 Données explicatives . . . . .	22
<b>2 Analyse de la structure de portefeuille</b>	<b>32</b>
2.1 Introduction . . . . .	32
2.2 Portefeuille Auto 1ère catégorie entre 2004 et 2006 . . . . .	33
2.2.1 Evolution du portefeuille . . . . .	33
2.2.2 Evolutions marquantes de la structure de portefeuille . . . . .	33
2.3 Analyse exploratoire des critères influençant les résiliations . . . . .	35
2.3.1 Principe et annotations . . . . .	35
2.3.2 Application . . . . .	35
2.4 Analyse des composantes multiples . . . . .	46
<b>3 Modélisation des résiliations</b>	<b>50</b>
3.1 Introduction . . . . .	50
3.2 Analyse discriminante . . . . .	51
3.2.1 Détermination des critères les plus discriminants . . . . .	51
3.2.2 Résultats de l'analyse discriminante . . . . .	54

3.2.3	Estimation du modèle . . . . .	55
3.2.4	Analyse discriminante avec un nombre réduit de variables explicatives . . . . .	57
3.3	Modélisation logistique . . . . .	58
3.3.1	Détermination des critères les plus discriminants . . . . .	58
3.3.2	Régression logistique en réduisant le nombre de critères explicatifs . . . . .	62
3.3.3	Ajustement des coefficients . . . . .	63
<b>4</b>	<b>Optimisations des ressources à l'échéance</b>	<b>68</b>
4.1	Introduction . . . . .	68
4.1.1	Les techniques de tarification à l'échéance . . . . .	68
4.1.2	La fixation du prix en assurance automobile . . . . .	71
4.2	Différentes stratégies de détermination du prix à l'échéance . . . . .	75
4.2.1	Maximum du chiffre d'affaires probable . . . . .	75
4.2.2	Maximum de la marge probable . . . . .	80
4.2.3	Approfondissement . . . . .	84
<b>5</b>	<b>Mise en application simplifiée</b>	<b>86</b>
5.1	Introduction . . . . .	86
5.2	Objectif méthodologique . . . . .	87
5.3	Application . . . . .	92
5.3.1	Arbre de décision de variation . . . . .	93
5.3.2	Caractérisation des feuilles . . . . .	94
5.3.3	Simulation avec les encadrements obtenus . . . . .	103
	<b>Conclusions</b>	<b>105</b>
C.1.	Bilan méthodologique . . . . .	105
C.2.	Prolongements . . . . .	107
	<b>Bibliographie</b>	<b>108</b>
	<b>Annexes</b>	<b>109</b>
	Annexe n°1 - Analyse exploratoire . . . . .	109
	Annexe n°2 - Corrélation entre variables explicatives . . . . .	117
	Annexe n°3 - Analyse en composantes multiples . . . . .	118
	Théorie . . . . .	118
	Application . . . . .	123
	Annexe n°4 - Analyse discriminante "Stepwise" . . . . .	127
	Annexe n°5 - Régression logistique . . . . .	132

# Résumé

Depuis plusieurs décennies, le secteur de l'assurance automobile fait l'objet d'une concurrence de plus en plus soutenue. Aux principales causes identifiées à l'origine (émergence de nouveaux acteurs, conquêtes de parts de marché par l'invention du marketing de masse), s'est ajouté des progrès techniques (informatique massive avec segmentation du tarif, explosion de nouveaux canaux de commercialisation). Dernièrement, de nouveaux facteurs conjoncturels viennent encore renforcer cette concurrence (stagnation du parc automobile, effet "environnement" faisant diminuer la part de revenu consacré à l'automobile). Les acteurs sont nombreux et la matière assurable tend à se rétrécir.

Actuellement, le principal axe de maintien de la croissance repose sur l'acquisition de nouveaux clients, par le biais des campagnes promotionnelles. On constate une érosion de l'effet de ces opérations, du fait de la forte concurrence dans ce secteur. Dans le même temps, les analyses techniques réalisées montrent que la rentabilité automobile est intimement liée à une gestion précise et à la fidélisation du portefeuille.

Ce mémoire propose une méthodologie résolument pratique, visant à répondre conjointement à l'enjeu de rétention des contrats d'assurance automobile et de rentabilité du portefeuille automobile.

En premier lieu, nous nous proposons d'analyser le comportement de résiliation de nos assurés à l'échéance. Cette étude nous amène à établir une modélisation tenant compte, notamment de l'élasticité prix. Par la suite, cette modélisation contribuera à une analyse et une discussion sur la stratégie interne du secteur automobile. Pour conclure, nous chercherons à définir une application résolument pratique de l'établissement des évolutions de cotisation à l'échéance.

La méthodologie ainsi proposée présente l'originalité d'alterner des méthodes classiques de statistiques et modélisations actuarielles en environnement probabilisé, avec de la programmation sous contraintes, voie peu explorée par les actuaires. L'objectif sous-jacent est d'apporter au service d'actuariat-produit une solution d'optimisation des primes en portefeuille pouvant être mise en oeuvre aisément du fait de son aspect opérationnel.

# Abstract

For several decades, competition in the automobile insurance sector has become more and more severe. The main causes already indentified are : the emergence of new actors, the explosion of news of marketing. And the stagnation of the car fleet as well as the decline of purchasing power.

At present, the main axis of preservation of our growth is based on the acquisition of new customers, by means of advertising and promotional campaigns. We notice an erosion of all these operations due to strong competition in this sector.

At the same time, the technical analyses realised show that the profitability of the automobile sector is closely linked to precise management and portfolio loyalty.

This report offers a practical methodology aimed at answering both the retention of automobile insurance contracts and the profitability of portfolios.

Firstly we suggest analysing the ways in which our customers terminate their contracts at renewal date. This study will establish a model principally taking into account the impact of the price elasticity. It will also contribute to further analysis and debate on the internal strategy of the automobile sector. To conclude, we will try to define a practical policy to rate evolution when contract renewal is due.

The methodology presents the originality of alternate classics methods of statistics and actuarial model built in a probalised environment with programming under constraints, which is usually not yet used by the actuaries. The underlying objective is to bring in the actuariat-product department a solution of optimization of the premium in portfolio which could be implemented easily because of its pragmatic aspects.

# Chapitre 1

## Introduction

### 1.1 Problématique

#### 1.1.1 Contexte général

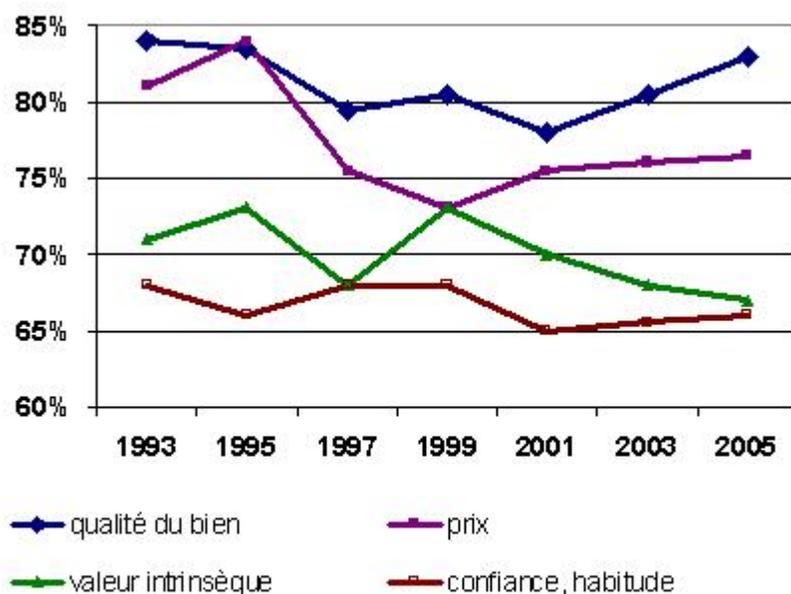
Depuis plusieurs années, nous sommes confrontés à un sentiment de fort ralentissement du pouvoir d'achat en France : après une période florissante à la fin des années 1990, les consommateurs français se trouvent de nouveau face à une conjoncture moins favorable. Afin de palier à l'érosion du pouvoir d'achat, les consommateurs réalisent des arbitrages entre les différents postes de consommation. Ainsi, les postes "incompressibles" tels que l'énergie, le logement et les dépenses de santé ne cessent de prendre de l'importance et ceci au détriment des services (hôtellerie, services récréatifs).

Dans le même temps, le prix des différents postes n'évolue pas de manière identique. Ces différences proviennent principalement des événements liés au marché et des objectifs fixés par les gouvernements :

- L'énergie connaît une forte augmentation du fait du cours du pétrole,
- l'augmentation du prix du tabac résulte de la volonté du gouvernement de réduire la consommation de ce produit,
- les transports aériens ont longtemps vu leurs prix se stabiliser du fait de la concurrence accrue sur ce secteur.
- les prix des équipements informatiques et audiovisuels baissent par l'évolution de la technologie et de la demande accrue,

Afin de contrer le sentiment de baisse du pouvoir d'achat et de hausse des prix, de nouveaux modèles économiques se mettent en place. On observe notamment la généralisation des modèles à "bas coûts", l'instauration des offres sans limitation d'utilisation (exemple : forfait internet illimité), ou encore l'apparition des produits à bas prix plus connus sous la dénomination de produits "low-cost". Face à ses nouveaux fondements économiques modifiant la lisibilité du marché, les consommateurs réagissent différemment selon leur profil et l'importance qu'ils donnent aux principaux critères de sélection que sont la qualité du bien, son prix, sa valeur intrinsèque et

l'habitude de consommation. La hiérarchie entre ces critères tend cependant à se modifier au cours des dernières années, le facteur prix prenant ainsi une plus grande importance :



Source : Crédoc, enquête consommation 2005

Figure n°1  
Evolution des critères d'achat

### 1.1.2 Contexte en assurance Automobile

Le secteur de l'assurance, en particulier celui de l'assurance Incendie, Accidents, Risques Divers (IARD), ne déroge pas à la morosité ambiante et aux modifications de comportement des consommateurs. Depuis deux décennies, nous constatons une concurrence soutenue sur le marché des produits d'assurance IARD et plus particulièrement dans le domaine de l'assurance automobile. Cette situation influence fortement le prix de marché.

Les causes d'une telle concurrence sont connues :

- En premier lieu, la concurrence importante sur ce marché avec près de 140 intervenants et l'émergence de nouveaux acteurs en quête de territoires de développement (la part de marché de la bancassurance passe de 6% à 8% entre 2006 et 2008),

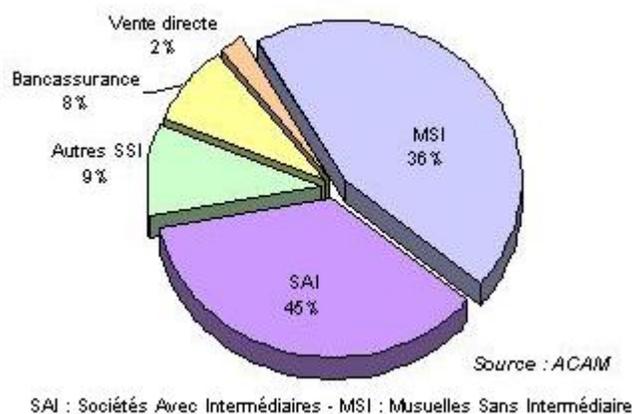
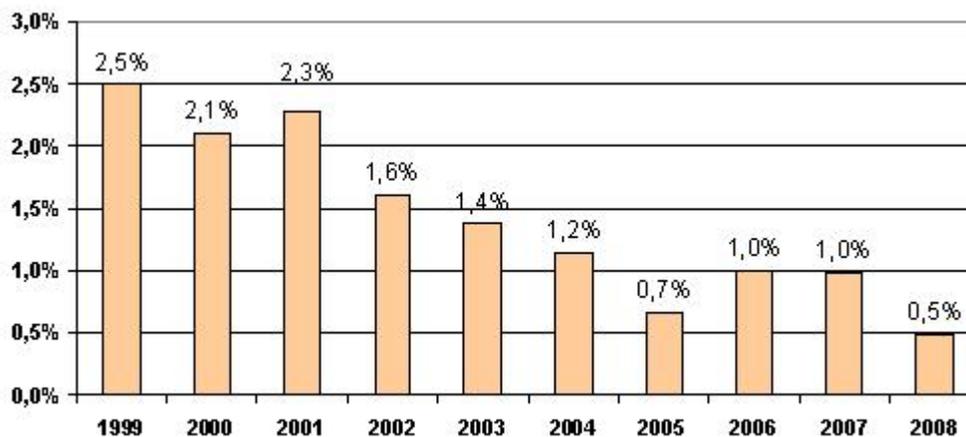


Figure n°2  
Part de marché 2008

- L'explosion de nouveaux canaux permettant d'aller au plus près du consommateur (exemple : vente par correspondance, téléphone et à présent Internet),
- Le désir des différents acteurs de prendre des parts de marché en augmentant sans cesse leurs moyens et investissements (exemple : budgets publicitaires télévisuels),
- Le ralentissement du pouvoir d'achat incitant les consommateurs à définir au mieux leurs besoins,
- La stagnation de l'évolution du parc automobile français sur les dernières années.



Source : Comité des Constructeurs Français d'Automobile (CCFA)

Figure n°3  
Evolution du parc automobile

Dans ce marché extrêmement affûté, donc fortement enclin à devenir élastique au prix, plusieurs voies stratégiques se distinguent pour réussir à développer le chiffre d'affaires des compagnies d'assurances, tout en respectant l'obligation de rentabilité imposée par l'actionnaire ou le législateur. Force est de constater que les différents acteurs du secteur rivalisent d'idées pour capter de nouveaux clients. Ainsi, les rabais commerciaux sont les premiers leviers utilisés (exemple : mois gratuits, garantie supplémentaire non facturée).

Paradoxalement, la rentabilité des compagnies provient, non pas des nouveaux entrants, mais bien des contrats en portefeuille. Dans le cadre de la société MMA, nous déterminons la rentabilité du portefeuille à partir du niveau de CRC<sup>1</sup> (Contribution aux Résultats de la Compagnie), également appelée "marge".

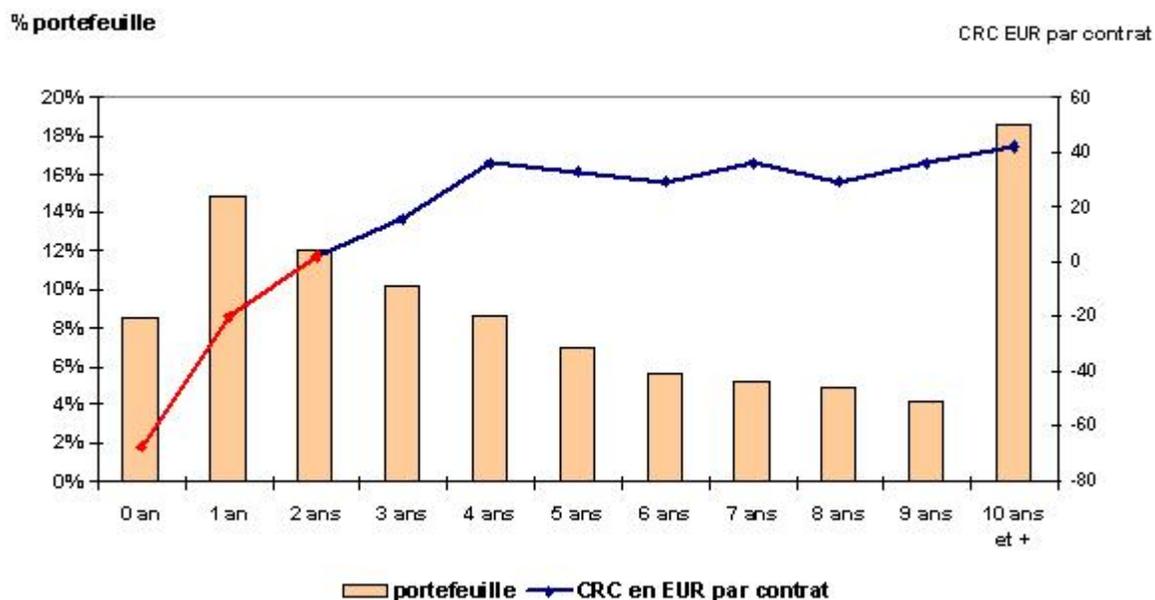


Figure n°4  
CRC 2008 (en EUR) par contrat selon l'ancienneté de contrat

Par conséquent, l'enjeu tend à se déplacer vers la rétention du portefeuille, afin de consolider les résultats. Il convient alors de se poser les questions suivantes :

- Quelles populations ont tendance à résilier ? Contribuent-elles à la politique de rentabilité de la compagnie ?
- Quelle stratégie mettre en place pour conserver les contrats ? En d'autres termes, doit-on miser sur une politique de conservation générale ou bien procéder à une sélection des risques à conserver ?

1. Le calcul de la CRC permet de prendre en considération l'ensemble des éléments affectant le compte d'exploitation de la société. La CRC est déterminée de la façon suivante :

$$\text{CRC} = \text{cotisations acquises} - \text{charge ultime} - \text{coût de réassurance} - \text{commissions} - \text{frais d'exploitation} - \text{variation de frais à reporter} + \text{produits financiers}$$

Ainsi, les leviers tarifaires initialement destinés aux affaires nouvelles se voient à présent proposés dans le cadre de la défense de portefeuille. L'objectif principal consiste en la conservation des populations rentables fragilisées par une modification de leur risque assurable (exemple : changement de véhicule). Dans le même temps, nous assistons à de réels efforts d'innovation sur les offres automobiles pour tenter de rendre attractif un produit initialement vendu du fait de son caractère obligatoire (Article L211-1 du Code des Assurances). L'assurance automobile est en passe de devenir un produit de consommation à part entière ainsi qu'un produit permettant de capter les clients. Les compagnies d'assurances ne cessent donc de trouver de nouvelles armes de vente :

- Création de nouvelles garanties : émergence des prestations d'assistance, de prévention et d'entretien,
- Mise en place de mécanismes de fidélisation : instauration du "bonus Matmut", création du "bonus à vie",
- Association de l'assurance avec d'autres services : un des meilleurs exemples s'avère être le crédit à taux préférentiel ou donnant des avantages sur l'assurance (exemple : réduction sur le contrat d'assurance, indemnisation améliorée).

Enfin, "last but not least", l'ensemble des acteurs du marché s'évertuent à segmenter toujours plus finement leur tarif, profitant des techniques statistiques développées sur des moyens informatiques de plus en plus puissants.

### 1.1.3 Bonus-Malus légal et Bonus MMA : exemple de mécanisme de fidélisation

#### • Le Bonus-Malus légal

Le système français de coefficient de réduction-majoration (CRM, Bonus-Malus ou BM) est apparu en juin 1976. Le principe est d'encadrer les primes en fonction de l'accidentologie selon les termes de la clause type de l'article A121-1 du Code des Assurances.

La règle du Bonus-Malus concerne l'ensemble des véhicules terrestres à moteur à l'exception :

- des cycles, tricycles, quadricycles à moteur (avec cylindrée  $\leq 80\text{cm}^3$ ),
- les véhicules et appareils agricoles, matériels forestiers, matériels de travaux publics et matériels spéciaux de service de lutte contre les incendies.

La période prise en compte pour l'application du CRM correspond à la période de 12 mois consécutifs, précédant de 2 mois l'échéance annuelle du contrat.

Actuellement, la clause de Bonus-Malus vérifie les hypothèses principales suivantes :

- **Il existe une classe initiale** : tous les assurés commencent avec un coefficient de 1.00,
- **le nombre de classe est fini** : le CRM est compris entre 0.50 et 3.50. De plus, il est arrêté à la deuxième décimale et arrondi par défaut,

- **il existe des règles de transition permettant de passer d'une classe à l'autre** : la détermination de la classe à un instant se fait en tenant compte de la classe précédente et de la survenance/non de sinistres responsables au cours d'une période annuelle d'assurance,

	<b>Absence de sinistre responsable</b>	<b>1 sinistre responsable total</b>	<b>1 sinistre responsable partiel</b>
<b>Clause "Normale"</b>	-5%	+25%	+12,5%
<b>Clause "Tournées"</b>	-7%	+20%	+10%

Figure n°5

*Evolutions de CRM en fonction de la sinistralité*

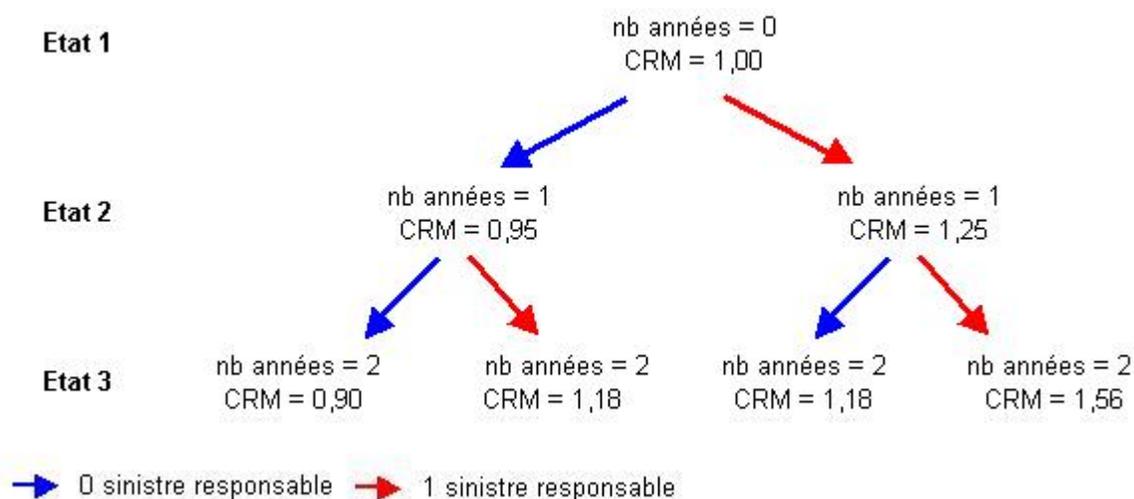


Figure n°6

*Exemple simplifié de l'évolution de CRM sur 3 années (clause "Normale")*

- **Il existe une classe dans laquelle l'ensemble des assurés doit se retrouver** : après un certain nombre d'années sans sinistre, le bonus maximal est fixé à 0.50. Le risque est alors considéré comme stable. Ainsi, au bout de 13 années sans sinistre responsable, un assuré en "clause Normale" bénéficie du coefficient de réduction-majoration à 0.50.

A partir de ces caractéristiques, le système de bonus-malus peut-être assimilé par une **chaîne de Markov**.

#### • Le Bonus MMA

D'après les chiffres communiqués par la FFSA en 2006, environ 25% du parc automobile (soit prêt de 9 millions de véhicules) a un bonus-malus compris entre 0.51 et 0.85.

A partir de ce constat, MMA décide de se constituer un axe de communication autour du bonus, permettant :

- De capter de nouveaux clients à fort potentiel : l'objectif est d'attirer des assurés plus jeunes, en voie d'équipement (cible des 25-35 ans),
- de fidéliser les assurés automobiles n'ayant pas encore atteint le bonus maximum : il s'agit de clients fragiles comme le montrait le taux de perte observé lors de l'étude initiale (20% contre 10.4% pour le reste du portefeuille).

L'objectif du Bonus MMA est d'octroyer un avantage tarifaire supplémentaire au client, tout en maintenant la gestion du Bonus-Malus légal. Le contrat peut bénéficier du Bonus MMA dès lors que le Bonus-Malus légal est inférieur ou égal à 0.85 et qu'il s'est déroulé 3 années sans sinistre. Le Bonus MMA permet une descente accélérée de 100% par rapport au Bonus-Malus légal.

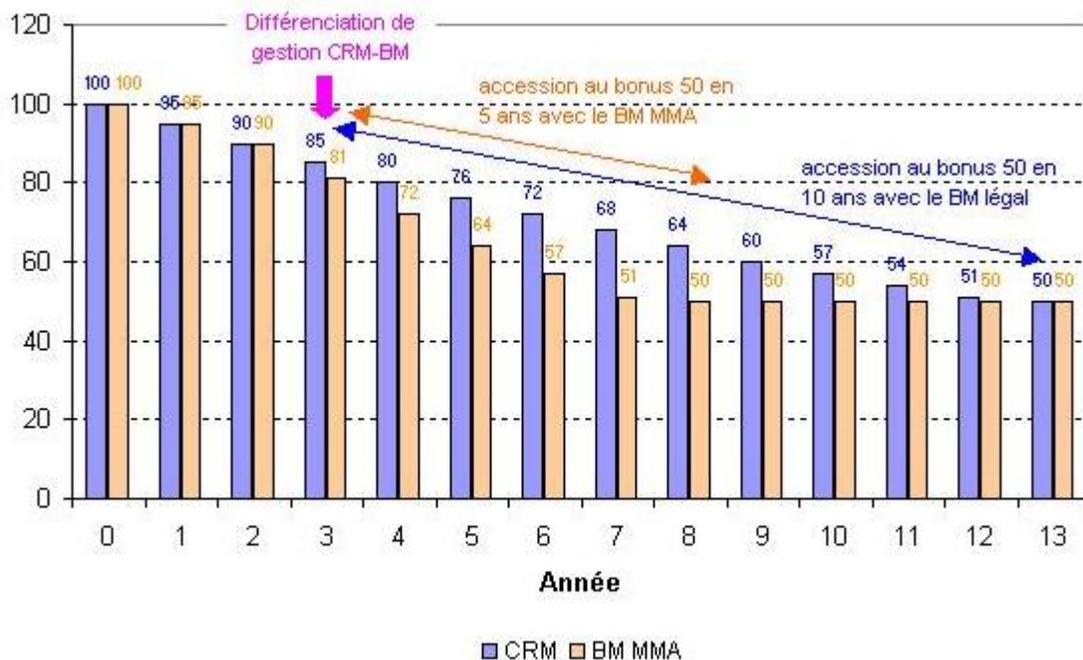


Figure n°7  
Comparaison des évolutions du CRM et du BM MMA

Cette descente accélérée de bonus et donc de cotisation génère :

- Une diminution du niveau de bonus plus importante sur les affaires nouvelles à venir que sur le portefeuille : en effet, le portefeuille bénéficie d'un niveau de Bonus-Malus inférieur aux nouveaux contrats, du fait notamment de la part plus importante des Bonus 0.50,
- une réduction importante du chiffre d'affaires au fil du temps du fait de l'accumulation de l'effet BM MMA.

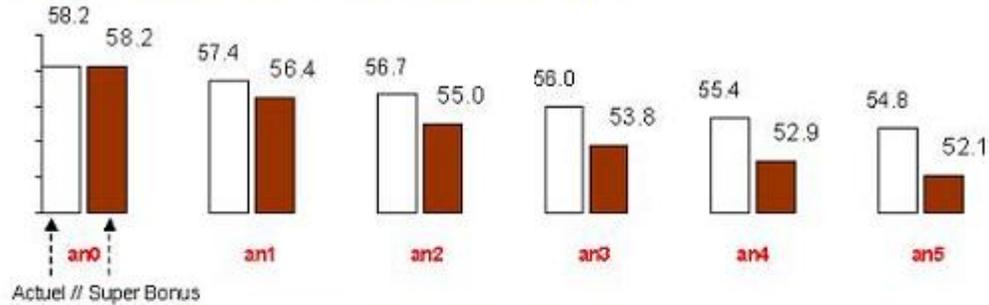
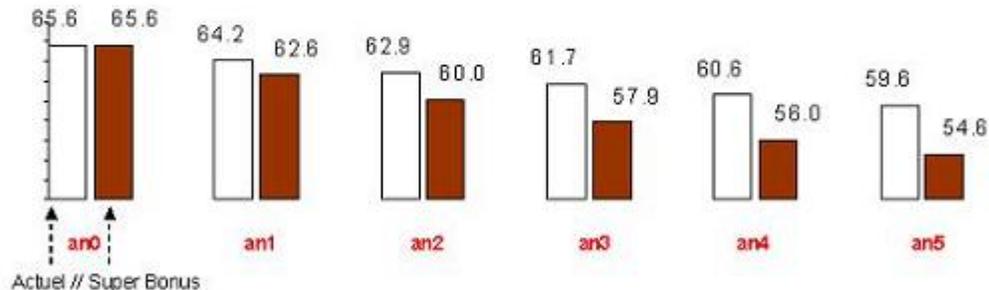
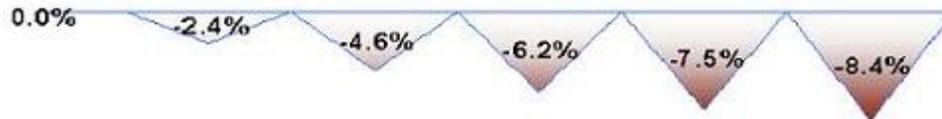
**Portefeuille fermé : Evolution du bonus moyen****Estimation de la baisse de chiffre d'affaires****Affaires Nouvelles : Evolution du bonus moyen en cas de non résiliation****Estimation de la baisse de chiffre d'affaires**

Figure n°8  
Prévisions de l'évolution de bonus et chiffre d'affaires

Forts de ces enseignements, il a été décidé de limiter la cible des assurés bénéficiant de ce mécanisme aux détenteurs de "Séries Spéciales"<sup>2</sup>. Ces formules, plus riches en terme de garanties et par conséquent avec un tarif supérieur, permettent en effet de limiter la perte de chiffre d'affaires.

2. la notion de "Séries Spéciales" est explicitée dans le paragraphe "Données explicatives"

### 1.1.4 Comportement de résiliation en assurance automobile

Le cadre législatif français de l'assurance automobile conditionne fortement la fidélité du client. Le contrat d'assurance est réputé à tacite reconduction (article L 113-15 du code des assurances, modifié par la Loi "Châtel" du 28 janvier 2005). Cette modification récente avait pour but de renforcer l'information du client sur son droit à résilier. Après quelques années, tous les acteurs de l'assurance ont observé un effet d'anticipation à l'échéance  $t$  des résiliations futures de l'exercice  $t + 1$ , notamment l'année ayant suivi le passage de la Loi "Châtel".

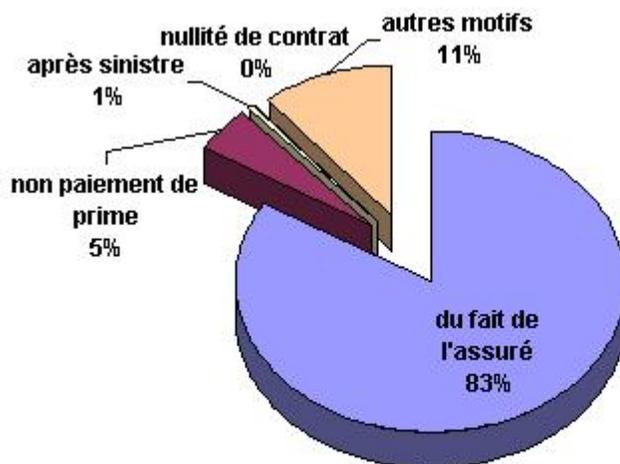
Depuis, le taux de résiliation contrat a repris son niveau classique, autour de 17% :

Exercice	Taux de résiliation	Taux d'affaires nouvelles
2006	16.9%	20.9%
2007	16.8%	20.4%
2008	16.6%	21.2%

source : FFSA, Assurance Automobile, rapport 2008 du 2 octobre 2009

Figure n°9  
Evolution des taux de résiliations et d'affaires nouvelles

Le code des assurances prévoit les motifs de rupture du contrat entre l'assureur et l'assuré : article L 113-9 à 113-15, L121-9 et 121-10, et, cas extrême, le retrait d'agrément de l'assureur article L326-12. Lorsque l'on étudie le volume de ces motifs, il ressort que la très grande majorité des résiliations sont du fait de l'assuré :



Source : Fichier des résiliations de l'Association pour la Gestion des Informations sur le Risque Automobile, paru par la FFSA, Assurance Automobile, rapport 2008 du 2 octobre 2009

Figure n°10  
Répartition des résiliation par motif

Des deux statistiques précédentes, nous pouvons établir le taux de résiliation de l'assuré automobile, au niveau national, à environ 14%. Ce taux paraît relativement faible si on compare l'assurance automobile à d'autres secteurs d'activité de service (exemples : téléphonie, internet) ou à d'autres pays n'ayant pas de tacite reconduction (exemples : pays anglo-saxons, Japon). Dans ces autres contextes, on parlera plutôt de "taux de rétention", tant l'enjeu est de conserver les contrats. L'encadrement législatif français permet donc de minimiser le taux de rotation du portefeuille de contrats, et rentabiliser sur la durée les investissements nécessaires à l'acquisition de nouveaux clients.

Si nous nous intéressons plus finement aux motifs individuels du "consommateur" de résilier son contrat d'assurance automobile, nous trouvons peu d'éléments généraux. En effet, les motifs avancés touchent souvent aux aspects stratégiques de gestion de chaque compagnie d'assurance. Dans de rares études de cabinet-conseils en marketing (exemple : Etude Kollen Partner, Assistance marketing), les principaux moteurs du comportement de résiliation sont souvent interdépendants et relatifs aux choix de la compagnie :

- la hausse de tarif non acceptée,
- l'insatisfaction dans la relation client,
- la promesse non tenue lors d'un règlement de sinistre,
- les gestes commerciaux inadaptés.

Ces moteurs s'exercent lors d'événements se révélant des opportunités de remettre en cause son assureur. On peut facilement mesurer sur un portefeuille d'assurance automobile la fréquence de ces événements, donnant lieu le plus souvent à avenant du contrat, ou éventuellement à résiliation :

- 20% à 25% de changements de véhicule,
- 10% à 15% de changements de lieu de résidence,
- 10% à 15% d'autres changements pouvant impacter le contrat d'assurance automobile (exemples : changements professionnels ou retraite, changement de la composition du foyer, nouveau conducteur).

### 1.1.5 Les enjeux de la résiliation en assurance automobile pour MMA

Chez MMA, le secteur automobile constitue un axe important du domaine des Particuliers IARD tant sur le plan du chiffre d'affaires que sur la rentabilité :

- plus de 60% du chiffre d'affaires des Particuliers IARD est réalisé par le secteur automobile (59% si on considère exclusivement le secteur automobile 1ère catégorie),
- la contribution aux risques de la compagnie (CRC) du secteur automobile tend à compenser les risques liés aux événements climatiques impactant fortement la CRC des Particuliers IARD (exemple : dégradation des résultats Habitation du fait des tempêtes Quinten et Klaus en 2009).

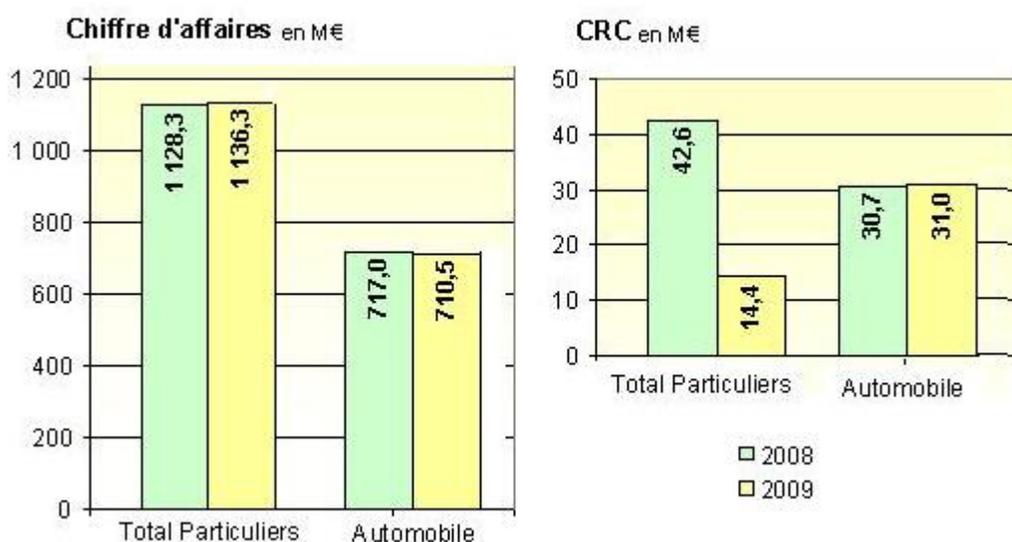


Figure n°11

Chiffre d'affaires et CRC des domaines Particuliers IARD et Automobile

Dans ce contexte, le secteur automobile doit poursuivre son développement tout en conservant son équilibre économique.

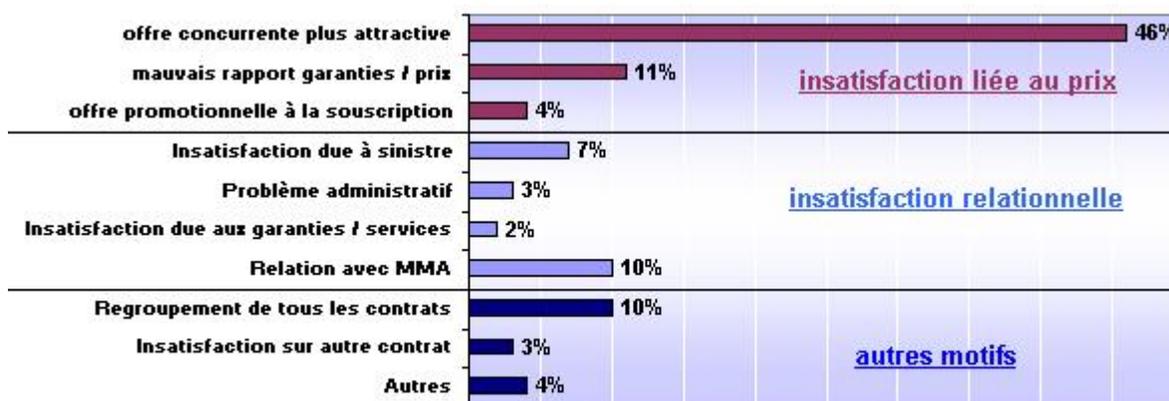
Au cours des dernières années, nous constatons un ralentissement de cette croissance, du fait de deux éléments combinés :

- le ralentissement de la production,
- la stabilisation du nombre de résiliations.



Figure n°12  
Evolution de la production nette Auto 1ère catégorie

Au même titre que la production ou la sensibilité des clients à souscrire leur assurance Auto chez MMA, l'analyse des résiliations joue un rôle primordial dans la volonté de la société d'augmenter sa part de marché.



Source : Etude "Actudes" réalisée en juin 2006 à partir de 560 clients ayant résilié leur contrat automobile

Figure n°13  
Motifs de résiliations évoqués par les clients

A partir des sondages réalisés sur nos clients ayant résilié, nous constatons que le prix est le principal critère de résiliation. Près de 61% des résiliations sont, en effet, expliquées par une notion de prix. La sensibilité au prix est un facteur pouvant être modélisé sur le plan statistique. Il est possible de s'appuyer sur des études de la sensibilité des résiliations au prix pour fidéliser le portefeuille ou une partie de celui-ci afin de contribuer au développement rentable.

Les autres motifs de résiliation sont qualitatifs et liés à des méthodes de gestion ou à l'aspect relationnel entre la société et le client. Ces critères peuvent difficilement être modélisés de manière statistique car ils dépendent de la sensibilité individuelle de chaque client. Ces aspects ne seront donc pas considérés à proprement parler dans ce mémoire.

## 1.2 Cadre de l'étude

### 1.2.1 Objet de l'étude

Nous considérons dans ce mémoire l'événement **résiliation du contrat du fait de l'assuré**. Certaines compagnies considèrent comme résiliation la disparition totale du panier du foyer client. Cet artifice permet d'afficher un taux de résiliation très bas, mais ne satisfait pas notre périmètre concernant le produit d'assurance automobile. Il s'agit bien ici d'étudier la résiliation de contrat sur un portefeuille produit, le client pouvant être toujours assuré par d'autres produits. Nous excluons d'autre part les motifs du fait de l'assureur, cités précédemment. Ces résiliations d'ordre "techniques" n'entrent pas dans le champ de la statistique incertaine, puisque ce sont des règles connues qui président à leur déclenchement certain (exemple : surveillance).

L'événement de résiliation est pris dans l'exercice immédiat qui suit l'échéance d'observation, sans contrainte de délai. Ce mémoire pose une hypothèse simplificatrice forte : **les évolutions à l'échéance conditionnent le comportement de résiliation sur l'exercice qui suit**. Cette hypothèse peut être contestée. En effet, la relation client est un processus continu dans le temps, accumulant ou annihilant l'insatisfaction. D'autres événements interviennent en cours d'année pouvant motiver la résiliation (exemples : déménagement, changement de conducteur ou de véhicule). Le système global de fidélisation qu'il conviendrait d'étudier s'avère, par conséquent, beaucoup plus complexe que le modèle analysé. Notre ambition se bornera à l'optimisation de la prime à l'échéance. Il s'agit donc de mesurer l'impact des décisions de l'échéance sur toute résiliation à posteriori, même si certaines n'auraient qu'un vague lien de cause à effet.

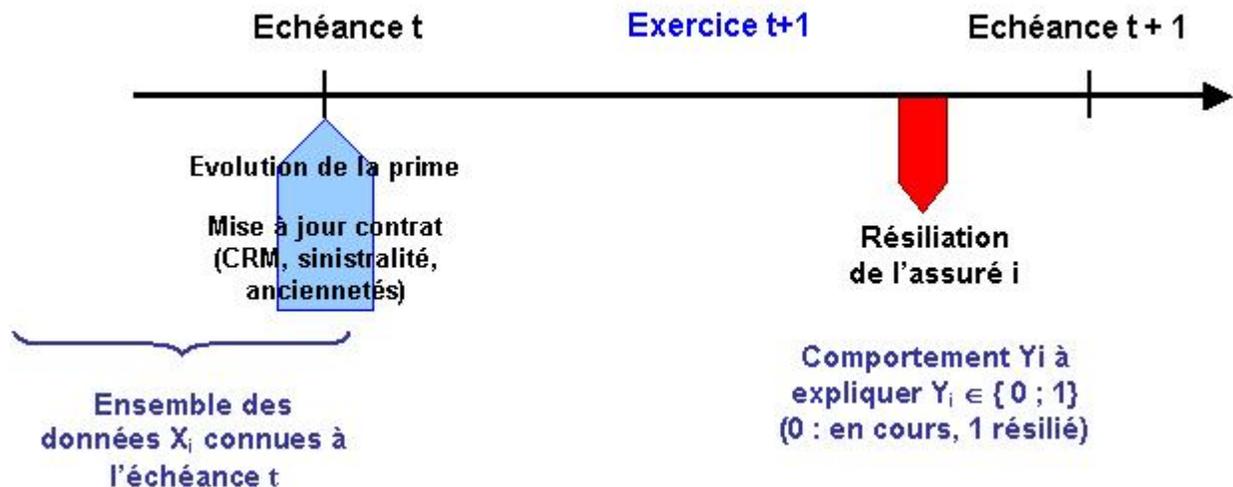


Figure n°14  
Schéma de l'étude proposée

Cette étude conduit à l'établissement d'une fonction de résiliation, définie à partir des caractéristiques du client et de l'élasticité au prix.

Par la suite, l'objectif est de décliner, à partir de la fonction de résiliation définie précédemment, les variations individuelles optimales de prime à l'échéance suivante. Deux hypothèses sont alors testées :

- Maximiser la marge sous contrainte de conserver un chiffre d'affaires, au minimum équivalent au niveau actuel,
- Optimiser le chiffre d'affaires sous contrainte de maintenir une marge positive.

Nous nous proposons de résoudre ces deux problématiques duales par le biais de méthodes d'optimisation.

Pour conclure, parce qu'une société ne peut se satisfaire d'une "boîte noire" pour établir les tarifs à l'échéance, nous chercherons, par le biais d'un modèle d'arbre de décision, la solution optimale, à partir des critères habituellement utilisés dans la tarification automobile. Ce dernier modèle nous permettra ainsi de définir une application pratique à l'établissement des évolutions des cotisations à l'échéance.

La méthodologie proposée présente l'originalité d'alterner des méthodes classiques de statistiques et modélisations actuarielles en environnement probabilisé, avec de la programmation sous contraintes, voie peu explorée par les actuaires. L'objectif sous-jacent est d'apporter au service d'actuariat-produit une solution d'optimisation des primes en portefeuille pouvant être mise en oeuvre aisément du fait de son aspect pratique.

### 1.2.2 Périmètre de l'étude

Dans le cadre de ce mémoire, nous nous appuyerons sur l'échantillon d'un portefeuille automobile. Cet échantillon est constitué à partir des situations terme de 2004, 2005 et 2006. Afin d'analyser un portefeuille homogène, nous ne considérerons que le produit 4 roues "automobile - 1ère catégorie" actuellement commercialisé :

- Les produits annexes couvrant des véhicules tels que les 2 roues (appelés véhicules de 3ème catégorie), caravanes, remorques, camions (véhicules de 2ème catégorie) et tracteurs ne sont pas intégrés dans l'analyse : leur structure de risque est trop éloignée en comparaison à la structure "automobile" (exemple : critères tarifaires, sinistralité, frais gestion, contrats gérés en flotte de véhicules). De plus, leur fonctionnement à l'échéance diverge (exemple : produit majorable et non tarifable, mécanisme d'encadrement simplifié, produits parfois non soumis au coefficient de réduction-majoration (CRM),
- Les anciennes générations de produits "automobile - 1ère catégorie" commercialisés par le passé ou issus de fusions avec d'autres sociétés d'assurances, ne sont pas conservés dans la base : leur structure de portefeuille et les critères tarifaires s'avèrent différents, leur fonctionnement à l'échéance diverge.

Enfin, afin d'obtenir des caractéristiques homogènes sur l'évolution des primes entre deux situations, nous excluons de l'analyse, les contrats subissant des évolutions trop importantes de cotisation :

- Evolution à la hausse pour cause de sur-sinistralité,
- Evolution à la baisse en cas d'amélioration combinée de plusieurs critères (exemple : changements de tranche au niveau des grilles tarifaires du conducteur, du véhicule, de la sinistralité antérieure).

L'échantillon sera constitué de contrats ayant une évolution de cotisation entre la situation terme servant de référence à l'étude et la situation précédente, comprise entre -10% et +13%. Des niveaux d'évolutions inférieurs à -10% dénotent de l'évolution à la baisse de plusieurs critères alors que des évolutions de cotisation supérieures à +13% s'expliquent généralement par la réalisation de sinistres au cours de l'année précédant la situation terme analysée.

### 1.2.3 Détermination de la prime technique et de la cotisation à l'échéance

#### • Détermination de la prime technique

**La tarification de chaque garantie du produit Auto respecte le principe académique de décomposition de la prime pure en fréquence et coût moyen.**

Les fréquences sont modélisées par modèles linéaires généralisés (MLG) sur une loi de Poisson avec les critères du tarif utilisés pour notre étude. Pour approximer la loi des coûts moyens, la charge des sinistres est écrêtée à 80.000EUR pour les garanties Responsabilité Civile et accidents du conducteur. Chaque coût est modélisé sur une loi Gamma avec les mêmes critères du tarif.

La prime pure, obtenue pour chaque garantie, est ensuite chargée proportionnellement, en tenant compte :

- du coefficient de retour sur marge de solvabilité,
- des frais de gestion de sinistre, proportionnels à la prime pure,
- du solde de réassurance.

On ajoute un chargement fixe qui couvre l'amortissement des frais d'acquisition et les frais de structure et de gestion. Enfin on multiplie le tarif chargé par :

- le coefficient de couverture des budgets
- les commissions d'apporteur,
- les taxes.

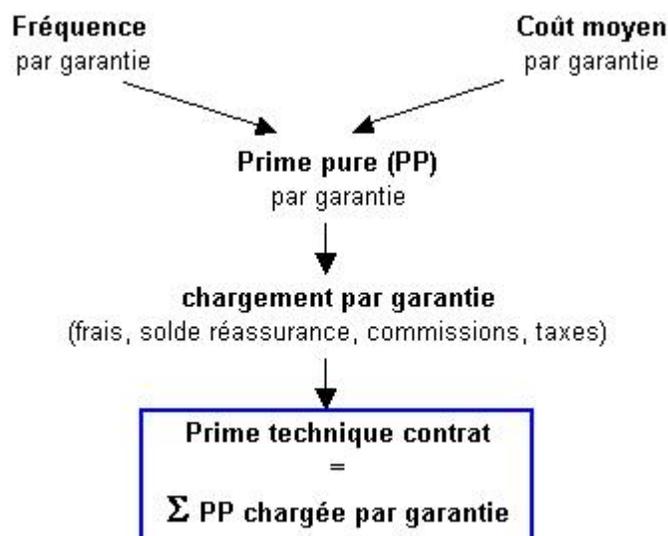


Figure n°15  
Détermination de la prime technique

Les résultats par garanties sont suivis tout au long de l'exercice. L'amélioration des fréquences en automobile de 2005 à 2009 a permis de faire bénéficier au portefeuille de baisses des tarifs substantielles. Depuis 2010, la reprise structurelle de la sinistralité, conjuguée à un poids d'événements naturels croissant, a imposé une réévaluation des garanties de dommages et bris de glaces.

#### • Détermination de la cotisation à l'échéance

La détermination de la cotisation au terme est un moment clé dans la vie du contrat, au même titre que la tarification d'un acte au comptant (affaire nouvelle, avenant). En effet, Le tarif donné à l'échéance contribue à la décision du client de résilier ou non son contrat pour l'année à venir.

Ainsi, le processus terme doit permettre d'obtenir un **tarif au plus près du risque** avec **une évolution de cotisation personnalisée** tout en restant un **mécanisme industrialisé** :

- *Tarif au plus près du risque* : Le principe de retarification complète s'impose sur un produit de masse comme l'auto 1ère catégorie, soumis au contexte concurrentiel. Le terme s'avère être le moment privilégié où tous les paramètres et les critères sont réévalués, pour profiter du meilleur tarif en vigueur et de la meilleure appréhension des risques,
- *Evolution personnalisée* : à l'occasion du terme, certains clients passent en situation de fragilité alors que d'autres entretiennent une relation plus solide. Ces éléments, certes probables, doivent influencer l'évolution de la cotisation à l'échéance afin de ne pas peser sur la fragilité des premiers, et maximiser les ressources apportées par les seconds.
- *Mécanisme industrialisé* : les évolutions doivent être communicables et généralisables de manière à éviter le phénomène de boîte noire et permettre une gestion informatique.

Pour répondre à ces attentes, MMA procède a un encadrement de la cotisation technique permettant le lissage de la cotisation dans le temps. Ce mécanisme s'appelle principe de "plancher-plafond" et son objectif est de faire converger la prime payée par l'assuré vers son tarif technique, sans lui provoquer les variations importantes lors de la retarification.

La prime client à l'échéance s'obtient avec la succession de 3 étapes principales :

- **Etape n°1 - Détermination des primes techniques et théoriques :** L'échéance anniversaire est l'occasion d'actualiser les critères susceptibles d'avoir évolué de manière légale (exemple : le CRM, la novicité) ou du fait de notre structure tarifaire (exemple : changement de classe d'ancienneté de véhicule ou d'âge du conducteur).

A partir de cette prime technique, il convient de prendre en considération la prime théorique :

$$\text{Prime théorique} = \text{Prime technique} \times \text{Déroptions octroyées par l'agent}$$

Cette prime théorique représente la prime qui serait payée par le client en l'absence de mécanisme d'encadrement.

- **Etape n°2 - Détermination des primes plancher et plafond :** Le client s'attend à voir sa prime à l'échéance évoluer (à la baisse ou à la hausse) de manière raisonnable en comparaison à la prime payée initialement. il convient donc de déterminer des primes minimale et maximale "acceptables", à partir de la prime précédente :

$$\text{Prime Plancher} = (1 + x\%).\text{Prime précédente}$$

$$\text{Prime Plafond} = (1 + y\%).\text{Prime précédente}$$

$$\text{avec } x \in \mathbb{R}, y \in \mathbb{R} \text{ et } x \leq y$$

- **Etape n°3 - Détermination de la prime client à l'échéance :** La prime client est déterminée en fonction des primes théorique, plancher et plafond définies précédemment :

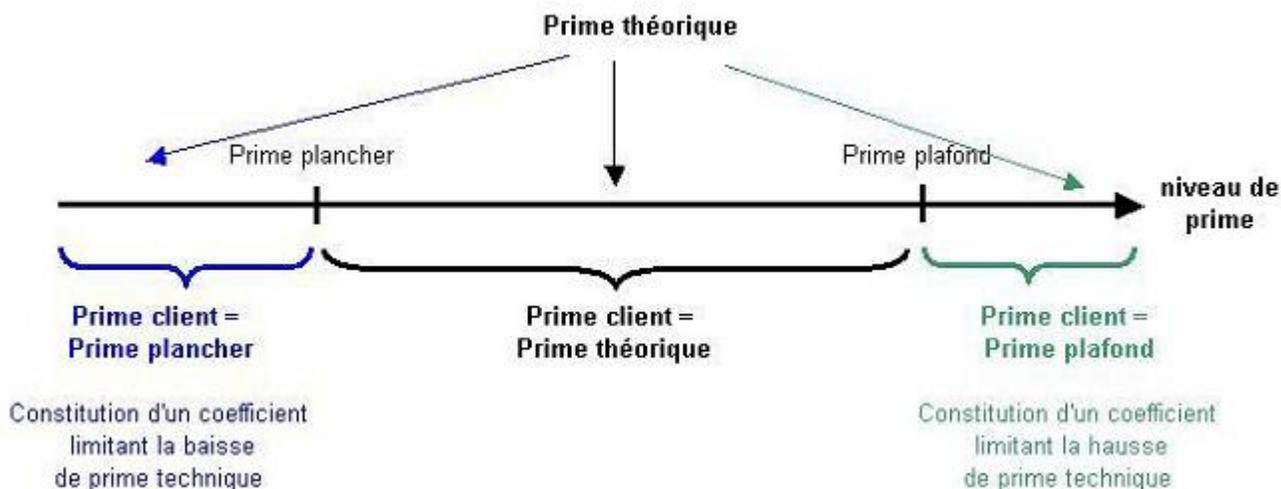


Figure n°16  
Détermination de la prime client

### 1.2.4 Données explicatives

Dans le cadre de l'analyse, trois typologies de variables sont prises en compte :

- les données "contrat", reprenant les données "conducteur(s)", "véhicule", "garanties" et "gestion du contrat",
- les données d'évolution du risque, liées à l'évolution de la "sinistralité", de "qualité de gestion du contrat" ainsi que de "l'évolution de la prime".
- les données annexes telles que la valeur du producteur et l'équipement du client.

#### • Données "contrat"

On définit comme faisant partie des données "contrat", l'ensemble des critères utilisés pour définir le risque assuré. Il s'agit notamment des critères tarifaires et des critères de gestion :

- Les critères tarifaires afférents au(x) conducteur(s),
- Les critères tarifaires associés au véhicule,
- Les garanties du risque,
- Les données de gestion.

#### *Les données "conducteur(s)"*

Elles reprennent l'ensemble des données tarifaires associées au(x) conducteur(s). La majorité des critères concerne le conducteur principal (exemples : son âge, son ancienneté de permis, son niveau de coefficient réduction-majoration (CRM)).

Certaines données sont relatives à l'ensemble des conducteurs désignés : nombre de conducteurs complémentaires, caractéristiques du plus jeune conducteur complémentaire (exemples : âge, ancienneté de permis).

Dans le cadre du produit étudié, il existe trois formules de conduite, définies en fonction du nombre de conducteurs désignés et de leur lien de parenté avec le conducteur principal :

- La formule de conduite dite "exclusive" : formule pouvant être souscrite lorsque le conjoint du conducteur principal est l'unique conducteur complémentaire désigné au contrat,
- La formule de conduite "famille" : cette formule peut être souscrite lorsque les conducteurs complémentaires désignés ont un lien de parenté avec le conducteur principal,
- La formule de conduite "tous conducteurs" : elle permet de désigner toute personne, comme conducteur complémentaire, même si elle n'a aucun lien de parenté avec le conducteur principal.

LIBELLE	FORMAT	COMMENTAIRES
Age du conducteur principal (CP)	numérique	0,01% des contrats ont un âge inférieur à 18 ans. Ils sont assimilés à "18 ans".
Ancienneté de permis du CP	numérique	
CP ayant fait la conduite accompagnée	booléen	Volumétries biaisées : la question est posée uniquement dans le cadre des sans antécédent d'assurance.
Sexe du CP	caractère	
Sexe du conducteur complémentaire (CC)	caractère	
CC ayant fait la conduite accompagnée	booléen	Volumétries biaisées : la question est posée uniquement dans le cadre des sans antécédent d'assurance.
Exemption de la majoration novicité	booléen	Non retenu par la suite : faibles volumétries
Présence de la réduction "agriculteur"	booléen	
Formule de conduite	caractère	
Coefficient de réduction-majoration (CRM)	numérique	
Ancienneté à CRM50	numérique	
Nombre de conducteurs complémentaires	numérique	Regroupement par la suite des "2 conducteurs complémentaires ou plus"
Usage tournées (gestion CRM)	booléen	

Figure n°17  
Liste des données conducteur(s)

#### Les données de "gestion"

Elles reprennent l'ensemble des données nécessaires à l'administration du contrat. Il s'agit principalement de définir le fractionnement de paiement, le mode de recouvrement (exemple : prélèvement bancaire), l'échéance du contrat ainsi que le niveau de cotisation de l'année.

DONNEE	LIBELLE	FORMAT	COMMENTAIRES
CGAR	Présence d'un garage	booléen	critère servant exclusivement en formules médianes ou tous risques
FRACT	Fractionnement de paiement	caractère	
INSEE	Code INSEE du lieu de garage	numérique	Nombre de modalités importantes : non retenu par la suite
FSOUS	Forfait kilométrique souscrit	caractère	
ANCCONT	Ancienneté de contrat	numérique	

Figure n°18  
Liste des données de gestion

*Les données "véhicule"*

Il existe deux typologies de données liées au véhicule :

- les données du modèle assuré
- les données particulières du véhicule assuré.

Les données du modèle assuré correspondent à l'ensemble des caractéristiques définissant le modèle "constructeur". Ces données sont centralisées dans la nomenclature dite "SRA" fournie par l'association Sécurité et Réparation Automobile<sup>3</sup>. Il s'agit :

- des données permettant d'identifier le véhicule (exemples : marque, modèle, version),
- des données techniques telles que le poids, la puissance, la vitesse maximale,
- de données calculées par le SRA et permettant de hiérarchiser les véhicules (exemple : classes de risque, classe de réparation).

En plus des données de type "constructeur", on observe des données spécifiques au véhicule assuré : son lieu de circulation, son stationnement habituel (garage ou non), son ancienneté, sa fréquence d'utilisation. Cette dernière caractéristique est définie, dans l'offre étudiée, par le forfait kilométrique souscrit (moins de 8 000 km, moins de 20 000 km et sans limitation).

LIBELLE	FORMAT	COMMENTAIRES
Marque du véhicule	caractère	Regroupement par type de véhicule (françaises, asiatiques, luxe...)
Modèle du véhicule	caractère	Non utilisé par la suite : trop de modalités
Ancienneté de véhicule (en années)	numérique	
Carrosserie	caractère	
Classe véhicule (SRA)	caractère	
Groupe véhicule (SRA)	numérique	
Genre véhicule (utilitaire / particulier)	caractère	
Energie	caractère	
Longueur du véhicule	numérique	Non retenu : 68% non alimenté
Largeur du véhicule	numérique	Non retenu : 68% non alimenté
Poids total en charge	numérique	Non retenu : 68% non alimenté
Puissance réelle	numérique	Non retenu : 68% non alimenté
Vitesse	numérique	Non retenu : 20% non alimenté

*Figure n°19  
Liste des données véhicule*

3. Cette association regroupe toutes les entreprises d'assurance automobile membre de la FFSA (Fédération Française des Sociétés d'Assurance) et du GEMA (Groupement des Entreprises Mutuelles d'Assurances). Elle est en charge de réaliser des études visant un meilleure maîtrise du prix de revient de l'assurance automobile (exemples : informations sur les véhicules, analyse des coûts de réparation)

*Les données "garanties"*

Le produit analysé se constitue d'une montée en gamme et d'options. La montée en gamme se décompose en trois niveaux de formules :

- La formule au tiers : souvent surnommée "droit de rouler", elle se compose des garanties indispensables à la circulation, et notamment de la responsabilité civile (RC).
- Les formules dites médianes : elles permettent également de couvrir le véhicule contre le vol et l'incendie. Il existe deux formules médianes dans l'offre actuellement commercialisée : la formule standard et la "série spéciale", qui constitue un package de garanties plus complet.
- Les formules tous risques : elles assurent également le véhicule contre tout dommage subi. Cinq formules tous risques sont proposées dont deux "séries spéciales". Ces formules diffèrent principalement par les périmètres garantis en inclusion.

	<b>Formule 4</b>	<b>Série Spéciale 4</b>
<b>Responsabilité Civile</b>		
<b>Défense Pénale et Recours</b>		
<b>Vol</b>		
<b>Incendie</b>		
<b>Dommages tous accidents</b>		
<b>Accessoires / Contenu</b>		périmètre étendu
<b>Panne 0 kilomètre</b>		
<b>Assurance conducteur</b>		
<b>Véhicule de remplacement</b>		
<b>Indemnisation majorée</b>	option	
<b>Taxi</b>		

Figure n°20

*Comparaison d'une formule tous risques et de sa série spéciale associée*

Outre le choix de formule, le client peut opter pour des options de garanties (exemples : garantie du conducteur sur une formule au tiers, objets contenus en formule médiane) ainsi que des niveaux de garanties particuliers (exemple : garantie du conducteur étendue aux véhicules loués ou empruntés).

<b>LIBELLE</b>	<b>FORMAT</b>	<b>COMMENTAIRES</b>
Formule de garanties	caractère	Formules classiques et séries spéciales
Franchise bris de glaces	numérique	
Franchise vol-incendie	numérique	
Franchise dommages	numérique	Non retenue par la suite : peut être déterminée à partir des variables FORMULE et FRVO

Figure n°21

*Liste des données associées aux garanties*

### • Données d'évolution du risque

Nous considérons, en plus des caractéristiques statiques du contrat, des éléments évolutifs tels que :

- Les primes,
- La sinistralité,
- Les indices de "qualité de gestion" (exemple : incidents de paiement).

#### *Evolution de la cotisation*

Dans le cadre de la modélisation des résiliations, nous souhaitons intégrer l'impact de l'évolution de la prime,  $\Pi$ , sur les résiliations. Il convient donc de mettre en place un indicateur d'évolution de cotisation  $\Delta_{\Pi_t}$ , à l'instant  $t$ , définie par :

$$[1] \forall t > 0, \Delta_{\Pi_t} = \frac{\Pi(t)}{\Pi(t-1)} - 1$$

$t$  représente la situation qui sert à l'analyse soit la situation terme de référence. La prime  $\Pi(t-1)$  représente la dernière cotisation connue avant le terme analysé. Il peut s'agir :

- de la cotisation de la situation terme de l'année précédente si le contrat n'a pas été mouvementé entre les deux échéances,
- de la cotisation de l'affaire nouvelle s'il s'agit d'une affaire nouvelle non mouvementée entre l'entrée en portefeuille et la date de première échéance,
- de la cotisation du dernier avenant réalisé avant la situation terme analysée, lorsque le contrat a été mouvementé.

La donnée  $\Delta_{\Pi}$  s'exprime en pourcentage et peut être aussi bien

- positive : il s'agit alors d'une majoration de la prime,
- nulle : la prime est alors stable à l'échéance,
- négative : il y a alors minoration de la cotisation.

L'évolution de cotisation est le fruit de plusieurs éléments complémentaires :

- Le contexte concurrentiel : les primes ont en effet tendance à moins fluctuer sur les secteurs à forte concurrence afin de conserver leurs parts de marché,
- La politique de la société d'assurances : le niveau des prix dépend de la stratégie de la société sur le secteur. Un secteur prioritaire verra des évolutions de cotisations à la hausse moins importantes,
- L'évolution du risque : le principal critère de majoration des cotisations reste la sinistralité propre du contrat. Pour un contrat n'évoluant pas entre deux échéances (pas d'avenant), les autres critères tarifaires sont des éléments généralement baissiers (exemple : âge du

- conducteur, ancienneté de véhicule) ou bien stables (exemple : formule de conducteur, classement véhicule),
- Le mécanisme d'encadrement au terme permettant de limiter les fluctuations trop importantes de cotisation d'une année à l'autre,
  - L'évolution des coûts et des fréquences sinistres.

	Fréquence 2008 <sup>(1)</sup>		Rappel 2007
	Niveau	Variation	Variation
RC :	43,7 %	- 4,9 %	- 2,2 %
<i>dont - RC corporels</i>	6,2 %	- 5,0 %	- 2,3 %
<i>- RC matériels</i>	37,5 %	- 4,9 %	- 2,2 %
Vol	8,1 %	-13,3 %	-11,5 %
Bris de glaces	87,5 %	+ 2,9 %	- 4,3 %
Domages tous acc.	77,1 %	- 2,4 %	- 3,0 %

(1) La fréquence représente le nombre de sinistres avec suite pour 1000 véhicules assurés pour cette garantie.

	Coûts moyens 2008		Rappel 2007
	En euros	Variation	Variation
RC matériels	1 454€	- 2,6 %	+ 0,2 %
Vol	2 178€	+ 3,7 %	+ 2,0 %
Bris de glaces	327€	+ 4,8 %	+ 3,0 %
Domages tous acc.	1 912€	+ 0,1 %	+ 0,2 %

Figure n°22  
Fréquences et coût moyens (FFSA)

### La sinistralité

L'assurance consiste à couvrir le client contre un événement aléatoire tant en fréquence qu'en coût. La sinistralité est constamment étudiée au niveau des équipes actuarielles. Elle constitue un élément essentiel dans le cadre de la prévision de la sinistralité à venir. Dans le cadre de la modélisation des résiliations du fait de l'assuré, les éléments de sinistralité sur les trois dernières années en portefeuille seront prises en compte afin de déterminer un lien éventuel entre la pension à résilier et le fait d'avoir des sinistres.

Afin d'obtenir une bonne vision des sinistres tout en conservant des volumétries significatives, nous distinguerons le nombre des sinistres sur trois années, en distinguant plusieurs typologies de sinistres :

- Les sinistres responsables matériels et corporels,
- Les sinistres circulation non responsables,
- Les autres sinistres (exemple : bris de glaces).

DONNEE	LIBELLE	FORMAT	COMMENTAIRES
EVOLPRIME	Evolution de la prime	numérique	Evolution entre la situation "terme" et la situation précédente (terme précédent, affaire nouvelle, dernier avenant)
MED	Présence de mises en demeure	booléen	
NRC <sub>xx</sub>	Nombre de sinistres responsables corporels	numérique	xx correspond au nombre de mois observés : 12, 24 ou 36 Donnée non retenue par la suite (1 seule donnée regroupant les sinistres des 36 derniers mois)
NRM <sub>xx</sub>	Nombre de sinistres responsables matériels	numérique	xx correspond au nombre de mois observés : 12, 24 ou 36 Donnée non retenue par la suite (1 seule donnée regroupant les sinistres des 36 derniers mois)
CIRC <sub>xx</sub>	Nombre de sinistres circulation, non responsables	numérique	xx correspond au nombre de mois observés : 12, 24 ou 36 Donnée non retenue par la suite (1 seule donnée regroupant les sinistres des 36 derniers mois)
AUT <sub>xx</sub>	Nombre de sinistres hors circulation (hors assistance)	numérique	xx correspond au nombre de mois observés : 12, 24 ou 36 Donnée non retenue par la suite (1 seule donnée regroupant les sinistres des 36 derniers mois)
TOT <sub>xx</sub>	Nombre total de sinistres (hors assistance)	numérique	xx correspond au nombre de mois observés : 12, 24 ou 36 Donnée non retenue par la suite (1 seule donnée regroupant les sinistres des 36 derniers mois)

*Figure n°23  
Liste des données liées à l'évolution du risque*

• **Les données annexes**

Deux données sont à considérer :

- la valeur du producteur,
- la valeur du client.

*La valeur du producteur*

L'objectif de MMA est de réaliser un développement rentable. Tous les acteurs sont mis à contribution. A ce titre, la société réalise de nombreuses études pour essayer d'améliorer la compétitivité et la technicité des intermédiaires (les agents).

Dans le cadre de l'analyse des résiliations, nous tiendrons compte de l'indicateur "Performance réseau". Cet indicateur s'appuie sur deux critères :

- la rentabilité déterminée par le rapport  $R = \frac{CRC_{n-1} + CRC_n}{cot.acquise_{n-1} + cot.acquise_n}$ ,
- le développement défini par le rapport  $D = \frac{App_{n-1} + App.net_n}{portefeuille_{n-2}}$  ou *App* désigne l'apport net.

La rentabilité est calculée sur 5 segments : le marché automobile, la multirisque habitation (MRH), le marché des professionnels, le marché de l'entreprise et le regroupement des autres marchés.

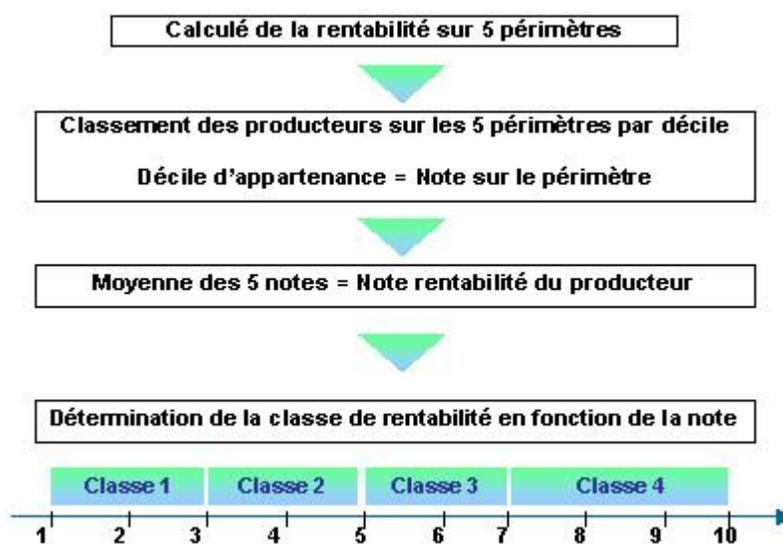
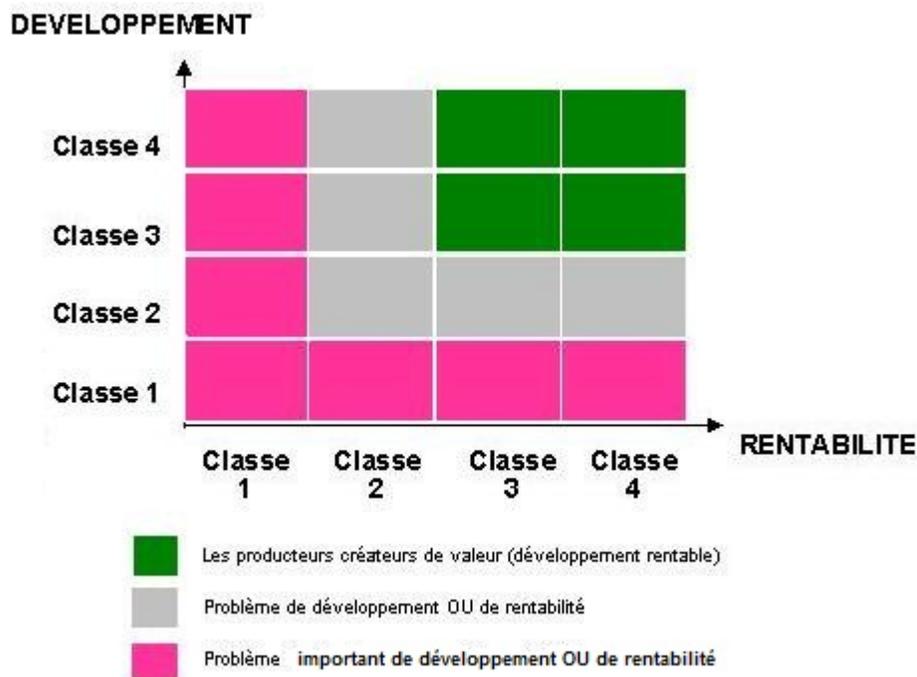


Figure n°24  
Détermination de la classe rentabilité

La note Développement est déterminée selon le processus, en s'appuyant sur 9 segments : les 4 premiers segments utilisés pour la rentabilité auxquels sont ajoutés la santé, la prévoyance, la protection juridique du particulier, l'épargne et la banque.

Une matrice de performance est déterminée à partir du classement Rentabilité et Développement. Cette matrice permet de définir 3 profils de producteurs :



*Figure n°25*  
 Matrice "Développement - Rentabilité"

### *La valeur du client*

Dans ce mémoire, la "valeur" du client sera déterminée par le principe de multi-détention. Nous nous appuierons sur :

- les gammes "pro" : présence de multi-risque professionnelle ou agricole,
- la présence de contrats automobiles (2 roues, tracteurs...) autres que le contrat analysé,
- la présence de contrats multi-risque habitation,
- les contrats épargne,
- les contrats vie,
- l'assurance prévoyance,
- l'assurance santé.

DONNEE	LIBELLE	FORMAT	COMMENTAIRES
VALAGT	Valeur agent	caractère	Valeur agent déterminée à partir des exercices 2007 et 2008 (principe inexistant avant). 8% des agences non rapprochés : la valeur "non alimentée" retenue
MULTIDET	Multidétention du client	numérique	Donnée déterminée à partir du profil client à fin décembre de chaque année. 4% de clients non rapprochés : considérés comme non équipés

*Figure n°26*  
*Liste des données annexes*

## Chapitre 2

# Analyse de la structure de portefeuille

### 2.1 Introduction

Ce chapitre a pour objet d'appréhender la structure de portefeuille de l'offre "automobile - 1ère catégorie" actuellement commercialisée par une importante compagnie d'assurance. Il s'appuie sur la base de travail constituée à partir des années d'observation 2004, 2005 et 2006.

L'objectif du mémoire étant d'analyser les résiliations et d'appréhender les leviers (notamment liés à l'évolution de la prime) susceptibles de canaliser les résiliations, il convient d'étudier, dans un premier temps, l'influence de chaque variable issue de la base sur la propension à résilier. Cette étude porte sur l'intégralité du portefeuille et peut nous amener à réaliser des regroupements, notamment sur les variables numériques et qualitatives ayant un nombre important de modalités (exemples : âge du conducteur, classe du véhicule, définie par le SRA). Toutes ces analyses sont réalisées à partir du logiciel SAS.

Par la suite, nous analyserons les liaisons entre les différentes variables explicatives, en nous appuyant sur l'analyse en composantes multiples (ACM).

## 2.2 Portefeuille Auto 1ère catégorie entre 2004 et 2006

### 2.2.1 Evolution du portefeuille

	Contrats		Cotisations Acquisées	
	Nombre	Evolution	en MEUR	Evolution
<b>2004</b>	828 121	18,8%	336,9	16,3%
<b>2005</b>	935 691	13,0%	376,3	11,7%
<b>2006</b>	1 025 240	9,6%	412,9	9,7%

*Figure n°27*  
Evolution du portefeuille entre 2004 et 2006

Le portefeuille de l'offre commercialisée connaît une forte croissance entre 2004 et 2006 du fait :

- de la production des affaires nouvelles : autour de 190000 affaires nouvelles par an,
- du transfert des anciennes gammes vers le produit commercialisé : cette conversion représente entre 30000 et 35000 contrats par an.

Les primes acquises évoluent dans des proportions similaires à l'évolution du portefeuille. La prime moyenne par contrat évolue peu au cours de cette période et est restée légèrement supérieure à 400EUR.

### 2.2.2 Evolutions marquantes de la structure de portefeuille

#### • L'ancienneté de contrat

	2004	2005	2006
<b>&lt; 1 an</b>	22,7%	18,1%	15,9%
<b>1 an</b>	15,6%	17,1%	15,4%
<b>2 ans</b>	10,8%	11,9%	13,1%
<b>3-4 ans</b>	18,3%	15,9%	15,9%
<b>5-9 ans</b>	15,2%	18,7%	20,3%
<b>10-19 ans</b>	10,2%	11,2%	12,0%
<b>&gt; 19 ans</b>	7,2%	7,2%	7,4%

*Figure n°28*  
Répartition du portefeuille par formule de garanties

La part des contrats récents devient moins importante. Ce phénomène s'explique par l'augmentation régulière du portefeuille alors que, dans le même temps, le nombre d'affaires nouvelles reste stable. L'impact "affaire nouvelle" s'avère donc de moins en moins important.

- La formule de garanties

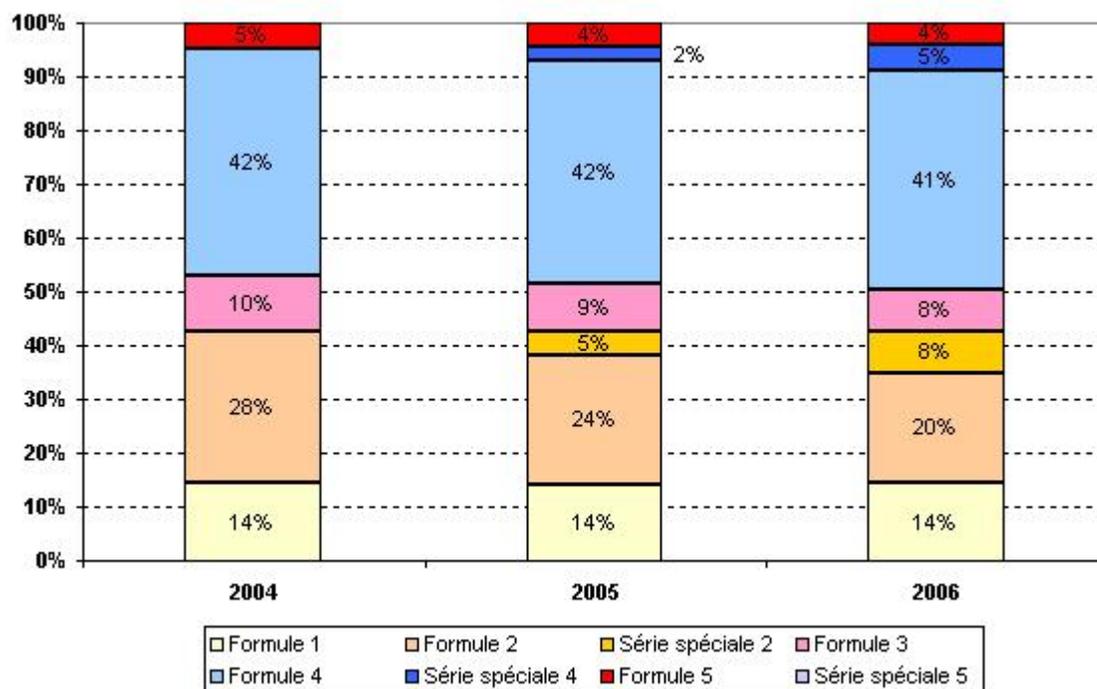


Figure n°29

Répartition du portefeuille par formule de garanties

La répartition du portefeuille selon la formule de garanties reste relativement homogène. Les quelques différences (exemple : réduction de la part des formules 3) s'explique par la mise en place des Séries spéciales au début de l'année 2005 et notamment de la Série spéciale n°2. Le transfert de portefeuille s'effectue également des formules 4 et 5 vers la Série spéciale n°4, mais dans des proportions de portefeuilles beaucoup plus faibles.

- Le forfait kilométrique

	2004	2005	2006
< 8000km	24,5%	35,4%	36,6%
< 20000km	56,9%	52,0%	51,5%
km illimité	18,6%	12,6%	12,0%

Figure n°30

Répartition du portefeuille par forfait kilométrique

La structure de portefeuille évolue vers des forfaits kilométriques plus faibles. Cette évolution peut s'expliquer par notre bon positionnement tarifaire sur le petits forfaits kilométriques en comparaison aux compagnies n'utilisant pas cette segmentation. Cet avantage nous permet d'acquérir une part importante des petits forfaits kilométriques en affaires nouvelles (exemple : en 2006, 41% des affaires nouvelles étaient souscrites avec un forfait "moins de 8000km").

## 2.3 Analyse exploratoire des critères influençant les résiliations

### 2.3.1 Principe et annotations

Les contrats automobiles étant de plus en plus segmentés et les environnements informatiques permettant de faire toujours plus de liens entre les différentes caractéristiques d'un client, le nombre de variables pouvant être utilisées dans le cadre de la modélisation des résiliations s'avère important. L'analyse des critères nous permet de réaliser une première sélection des critères influençant le plus les résiliations. L'analyse univariée des données présentes dans la base initiale nous permet d'obtenir une liste plus restreinte des variables pertinentes pour la suite de l'analyse et de recoder les variables ayant un nombre important de modalités.

*Notations :*

- R constitue l'identifiant de résiliation : R=1 lorsque le contrat est résilié à l'échéance et 0 s'il est encore en cours,
- $n$  l'effectif total analysé,
- $n_i$  l'effectif de la modalité  $i$  d'une variable,
- $r_i$  le nombre de résiliés à l'échéance, de la modalité  $i$ .

Dans le cadre de l'analyse descriptive, on considère les deux rapports suivants :

- $\%PTF = \frac{n_i}{n}$  : la part du portefeuille appartenant à la modalité  $i$ ,
- $tx_R = \frac{r_i}{n_i}$  : le taux de résiliation associé à la modalité  $i$ .

### 2.3.2 Application

L'analyse univariée laisse apparaître des disparités importantes dans les résiliations en fonction de certaines variables. En plus des données liées à la valeur (agent et client), trois effets ressortent de cette première étude :

- le facteur prix,
- Le vieillissement,
- le facteur sinistralité.

Nous structurerons les résultats selon la description des types de données explicatives recensées au premier chapitre. L'ensemble des résultats se retrouve en **Annexe 1**. Pour faciliter l'analyse de la description univariée, les corrélations entre variables explicatives sont recensées en **Annexe 2**.

### • Les données conducteur(s)

On constate en premier lieu que le temps joue un rôle important dans les résiliations.

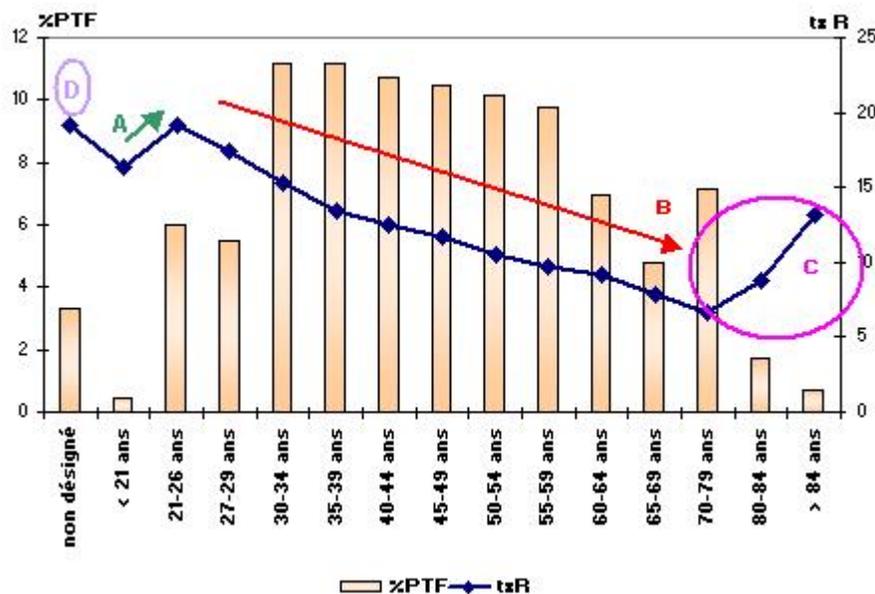


Figure n°31  
Taux de résiliation en fonction de l'âge du conducteur

Le taux de résiliation évolue, en fonction de l'âge du conducteur principal, selon quatre effets :

- un taux de résiliation qui se dégrade fortement lorsque l'assuré atteint plus de 21 ans (A). La population des moins de 21 ans est généralement fortement majorée et cherche rapidement à optimiser son budget. Cette hypothèse se trouve confortée lorsque l'on observe le taux de résiliations des contrats ayant une majoration novice : il se trouve supérieur de 3 points au reste de la population,
- un baisse linéaire du taux de résiliation sur les conducteurs âgés entre 21 et 79 ans (B),
- une recrudescence du taux de résiliation à partir de 80 ans (C), probablement influencée par l'arrêt de la conduite voire la disparition sur cette population,
- un taux de résiliation très important (19%) lorsque le conducteur n'est pas désigné (D).

L'évolution des résiliations liée à l'âge est similaire à l'évolution liée à l'ancienneté de permis du fait de la corrélation de ces deux données (93%). La constitution de la donnée "âge au moment du permis" permet d'illustrer cette corrélation : plus de 54% de nos assurés a, en effet, obtenu son permis entre 18 et 19 ans.

Le taux de résiliation élevé pour les conducteurs non désignés se retrouve de manière similaire sur les données :

- sexe du conducteur et ancienneté de permis,
- formule de conduite : la formule de conduite "tout conducteur", représentée majoritairement par les sociétés ne souhaitant pas désigner de conducteur. Le taux élevé de résiliation pour cette population peut notamment s'expliquer par des tarifs plus élevés du fait de la formule de conduite à proprement parler mais également du forfait kilométrique plus élevé et donc moins bien positionné sur le plan tarifaire (42% des contrats souscrits sans désignation de conducteurs se trouvent en forfait illimité contre 12% de la population).

FORMULE DE CONDUITE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Exclusive	51,49	13,14	54,78
Famille	43,65	10,81	38,21
Tout conducteur	4,86	17,83	7,02

Figure n°32

Taux de résiliation en fonction la formule de conduite

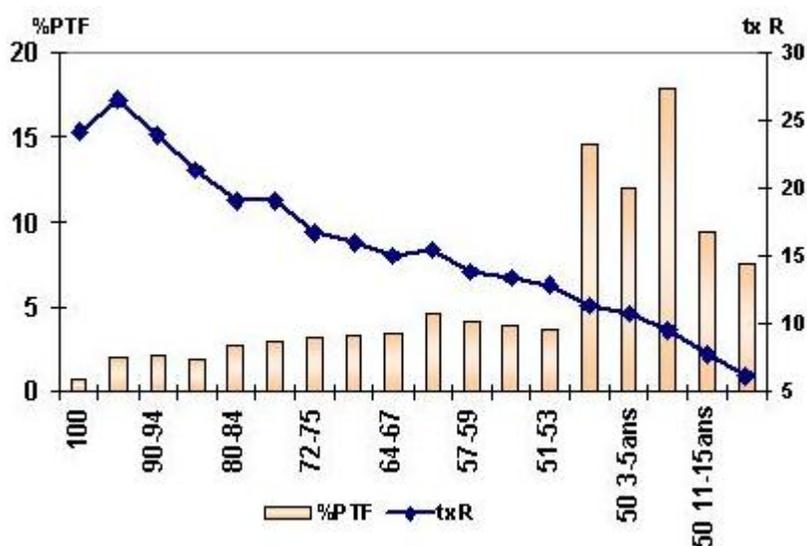


Figure n°33

Taux de résiliation en fonction du coefficient de réduction-majoration

De la même manière que pour l'âge ou l'ancienneté de permis, le niveau de CRM paraît être un critère discriminant dans le cadre des résiliations de portefeuille. L'ensemble de ces données est fortement corrélé, du fait même de la construction du coefficient de réduction-majoration, succinctement expliquée au chapitre précédent.

Les assurés "agricoles" bénéficient de taux de résiliations bien inférieurs au reste de l'échantillon (7.1% contre 12.6%). Cette rétention importante peut s'expliquer par plusieurs phénomènes :

- l'implantation de nos agences, facilitant les relations avec cette population traditionnellement plus conservatrice dans ses choix de consommation,
- la réduction tarifaire qui est octroyée sur cette cible et qui s'accompagne généralement de la souscription des petits forfaits kilométriques (51% des agriculteurs souscrivent un forfait "moins de 8000km" contre 37% de l'ensemble du portefeuille).

#### • Les données véhicules

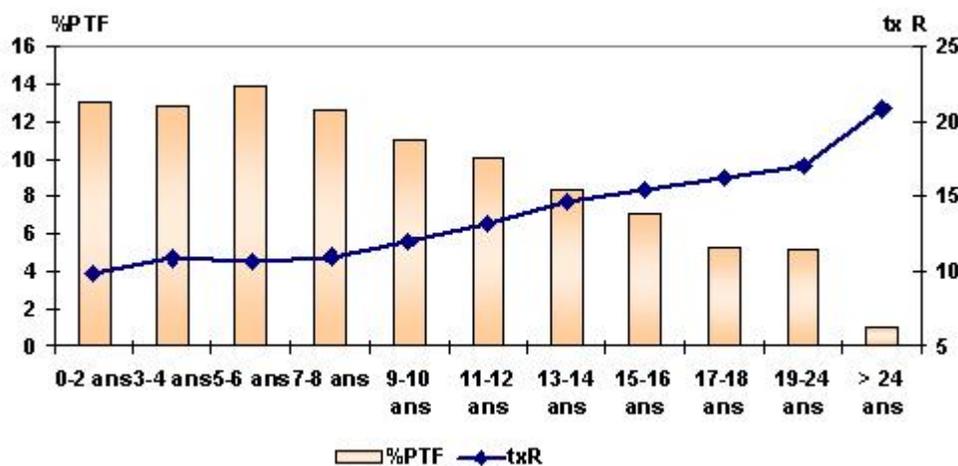


Figure n°34

Taux de résiliation en fonction de l'ancienneté du véhicule

La variable "ancienneté du véhicule" semble fortement impactée par le facteur temps, et ceci de manière négative. Le renouvellement du véhicule étant un moment propice pour résilier son contrat d'assurance automobile le taux de résiliation augmente sur les véhicules anciens, plus enclins à être remplacés.

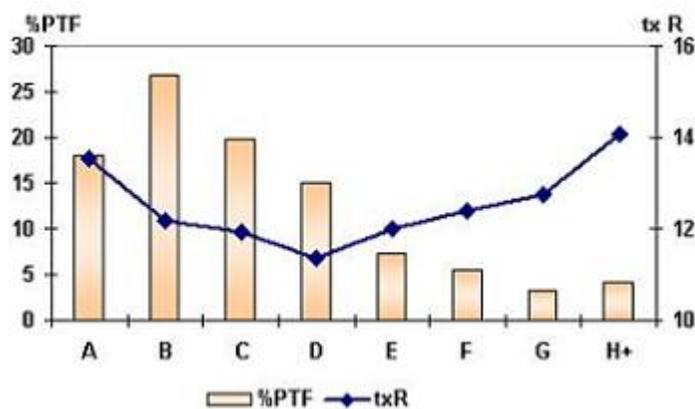


Figure n°35

Taux de résiliation en fonction de la classe du véhicule

Bien que l'amplitude du taux de résiliation soit inférieure par rapport à d'autres données explicatives (exemple : ancienneté de véhicule), la classe du véhicule peut être considérée comme un critère explicatif des résiliations. Deux populations semblent plus exposées aux résiliations :

- les véhicules à classe faible : il s'agit de véhicules plus anciens (ancienneté moyenne de 14 ans sur cette cible contre 10 ans sur l'ensemble du portefeuille) donc plus exposés aux résiliations lors des changements de véhicules,
- les véhicules de classe élevée : il s'agit des véhicules dits familiaux. Ils sont généralement utilisés pour réaliser des trajets plus importants. Le forfait kilométrique souscrit s'avère moins avantageux (20% en forfait kilométrique illimité contre 12.6% pour l'ensemble du portefeuille) entraînant une concurrence tarifaire plus importante.

Cette même analyse peut être faite à partir de la donnée "Groupe". Ce phénomène s'explique par la forte corrélation entre les données "Groupe" et "Classe" (86%).

	A	B	C	D	E	F	G+
8 et -	75%	19%	5%	1%	0%	0%	0%
9	44%	42%	14%	0%	0%	0%	0%
10	30%	59%	7%	3%	1%	0%	0%
11	7%	56%	29%	6%	1%	0%	0%
12	1%	31%	41%	19%	7%	2%	0%
13	0%	12%	34%	28%	17%	6%	3%
14	0%	2%	17%	31%	30%	14%	7%
15	0%	0%	2%	12%	31%	31%	24%
16	0%	0%	0%	0%	8%	31%	60%
17	0%	0%	0%	0%	0%	3%	97%

Figure n°36

Véhicules classés par le SRA : Répartition des Classes pour un Groupe donné

L'énergie du véhicule ne semble pas avoir aucune incidence sur le taux de résiliation.

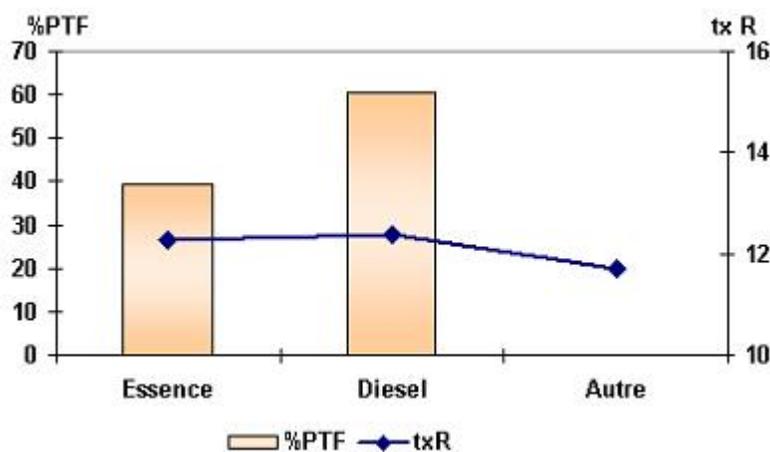


Figure n°37

Taux de résiliation en fonction de l'énergie du véhicule

### • Les garanties

On constate que les contrats riches en garanties sont moins soumis aux résiliations en comparaison aux formules au tiers et médianes.

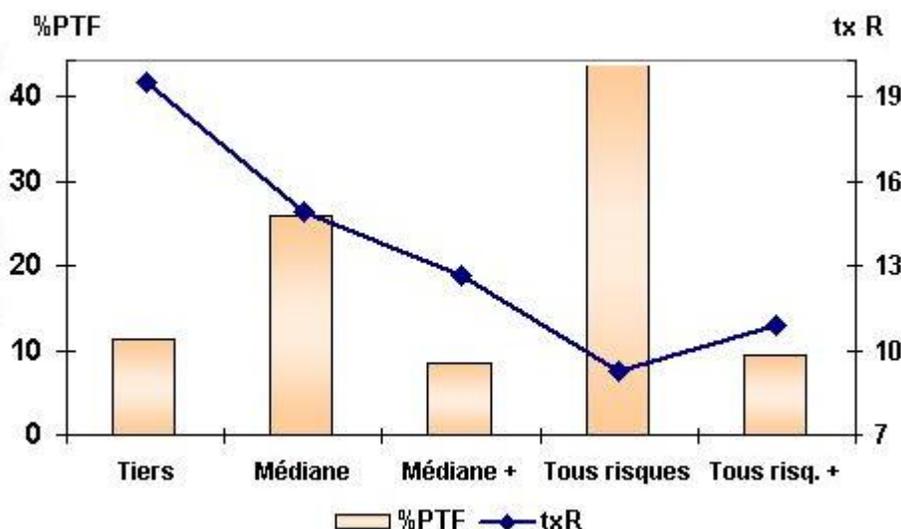


Figure n°38

Taux de résiliation en fonction de la formule de garanties

Plusieurs hypothèses peuvent être émises pour expliquer ce phénomène :

- l'ancienneté moyenne des véhicules sur les formules "pauvres" est beaucoup plus élevée que sur l'ensemble du portefeuille (17 ans contre 10 ans) : nous avons vu précédemment que le changement de véhicule (plus fréquent pour les véhicules anciens) était un moment propice à la résiliation,
- la présence de populations fragilisées sur le plan financier, en formules au tiers ou médiane, cherchant régulièrement à optimiser leur budget.

Les contrats avec franchise "vol" à 0EUR sont plus fortement résiliés. Ce taux élevé peut résulter d'une exigence accrue de ces assurés au moment de l'indemnisation d'un sinistre. On retrouve également, dans le tableau ci-dessous, le taux élevé de résiliation pour les contrats "sans franchise" c'est à dire les contrats en formule au tiers.

FRANCHISE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0 EUR	7,81	22,72	14,36
152 EUR	62,59	9,86	49,97
304+ EUR	18,27	11,98	17,72
SANS	11,33	19,56	17,95

Figure n°39

Taux de résiliation en fonction du niveau de franchise vol

- Les données de gestion

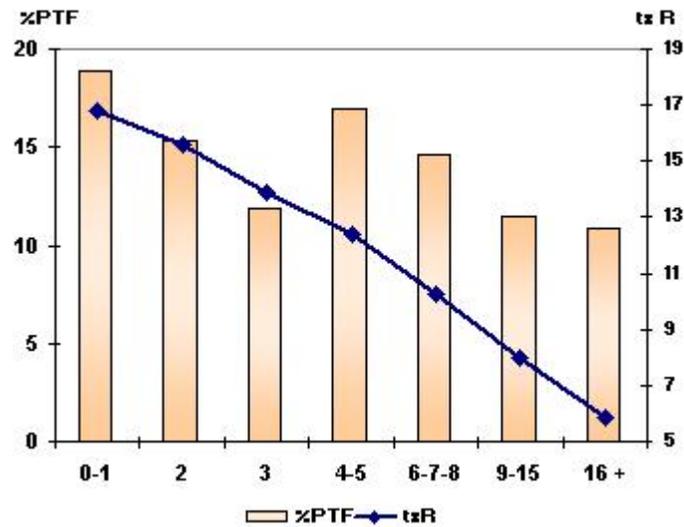


Figure n°40

Taux de résiliation en fonction de l'ancienneté du contrat

Le taux de résiliation décroît fortement avec l'ancienneté du contrat. Ceci peut s'expliquer par la corrélation entre cette donnée et l'âge du conducteur. En effet, bien que les changements de conducteurs puissent être réalisés, l'intégration des jeunes conducteurs dans notre portefeuille se fait principalement par le biais des affaires nouvelles.

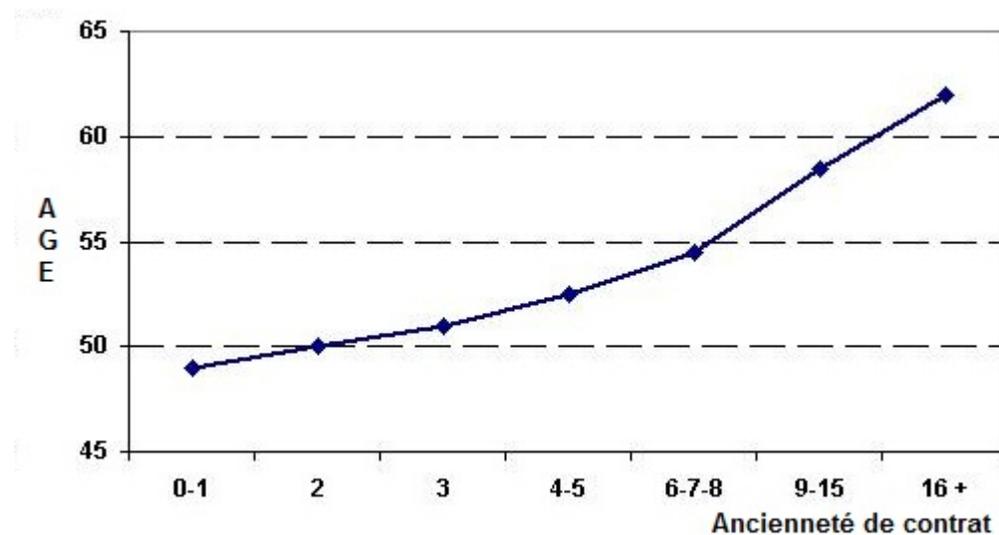


Figure n°41

Age moyen du conducteur en fonction de l'ancienneté de contrat

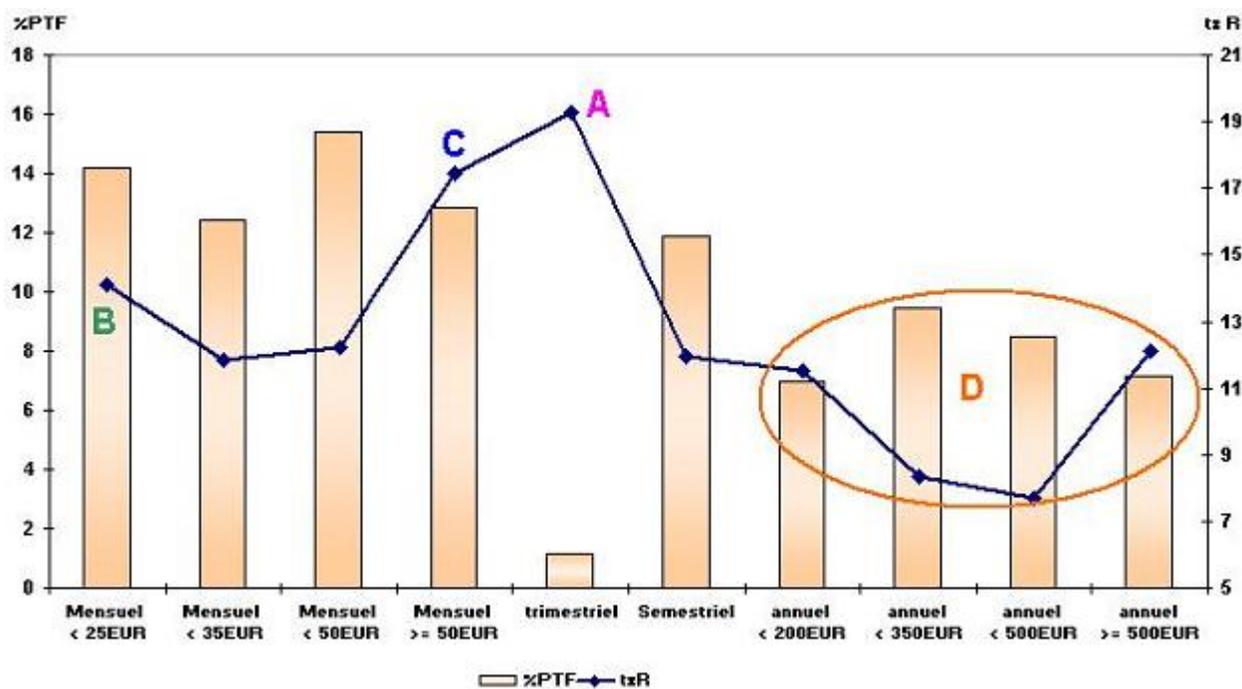


Figure n°42

Taux de résiliation en fonction du fractionnement

Les contrats en prélèvement trimestriel (A) ont un taux de résiliation bien plus élevé que le reste du portefeuille. Ce comportement s'explique notamment par une forte proportion de formules de garanties au tiers (21% contre 11% sur l'ensemble du portefeuille) au détriment des formules riches (à peine 45% de formules tous risques et tous risques + contre 54% pour l'ensemble du portefeuille).

Les contrats en prélèvement mensuel sont également impactés de manière négative (taux de résiliation élevé) sur deux cibles particulières :

- les primes fractionnées faibles (B) : les contrats sont dans 24% des cas des contrats en formule au tiers.
- les primes fractionnées élevées (C) : on constate une sur-représentation des jeunes conducteurs sur cette cible (12% ont moins de 25 ans contre 5% pour l'ensemble du portefeuille).

Enfin, le taux de résiliation constaté s'avère plus faible pour les contrats en fractionnement annuel (moins de 10%). Les primes annuelles faibles (41% de formule au tiers) et élevées (véhicules plus puissants avec un groupe moyen de 14 contre 12 pour l'ensemble du portefeuille) ont des résultats toutefois plus dégradés.

La donnée "garage" se trouve fortement corrélée, négativement (-59%), avec la "formule de garanties". En effet, la modalité "sans objet" correspond aux contrats en formule au tiers. Le taux de résiliation de cette modalité est, par conséquent, bien plus élevé que pour les autres modalités.

MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
sans garage	29,63	13,73	32,95
avec garage	59,03	10,27	49,10
sans objet	11,33	19,56	17,95

Figure n°43

*Taux de résiliation en fonction de la présence ou non de garage*

Les contrats en forfait kilométrique illimité ont un taux de résiliation plus élevé que les contrats en forfaits dits limités. La principale hypothèse pouvant être avancée face à ce phénomène est la segmentation de MMA sur cette donnée, aux détriment du forfait illimité, alors que dans le même temps, la concurrence mutualise le risque sur ce critère.

MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
K08	30,88	12,13	30,34
K20	53,68	11,79	51,23
K99	15,44	14,74	18,43

Figure n°44

*Taux de résiliation en fonction du forfait kilométrique*

- **Les données liées à l'évolution du risque (prime, sinistre)**

En plus du facteur temps, on constate que les résiliations semblent fortement liées au facteur prix. En effet, les contrats potentiellement à fortes primes (exemples : conducteurs non désignés, contrats avec application de majoration novice) ont des taux de résiliation bien supérieurs à la moyenne du portefeuille. Dans le même temps, les contrats bénéficiant de réductions (exemple : les agriculteurs) voient leur taux de résiliation plus faible que la moyenne.

Cette adéquation entre les résiliations et le niveau de prime se confirme lorsque l'on observe les résultats liés à l'évolution de la prime.

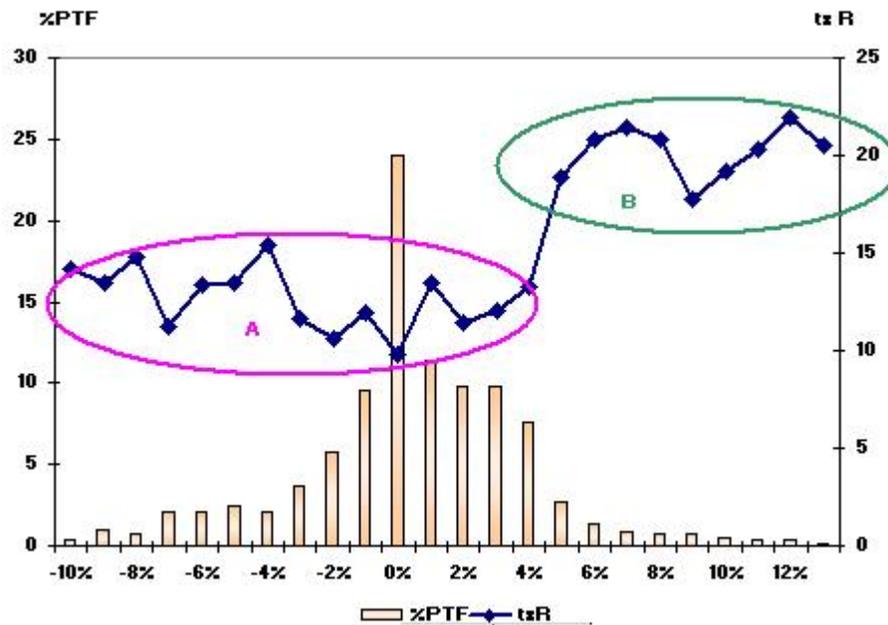


Figure n°45

Taux de résiliation en fonction de l'évolution de la prime

Deux populations sont identifiées :

- les contrats dont la prime n'évolue pas plus de 4% (A) : sur la période de 2004 à 2006, le secteur de l'automobile connaît une guerre des prix importantes. Ainsi, MMA réussit à conserver son portefeuille (taux de résiliation avoisinant les 12%) en octroyant à ses clients "privilegiés" des évolutions proches de celles proposées par le marché.
- les contrats dont la prime est supérieure à 4% (B) : la cotisation évolue alors trop fortement, incitant le client à prospecter à la concurrence.

On notera la forme logistique du taux de résiliation, en fonction de la variation du prix. Cette observation motivera notre modélisation par la suite.

Le nombre de résiliations augmente de manière importante pour les contrats sinistrés, notamment dans le cadre de sinistres responsables. Cette évolution peut résulter d'un principe de résiliation amiable pour des clients en sur-risque ou une gestion sinistre difficilement acceptée. Ce comportement tend à se confirmer avec l'analyse de la donnée "coefficient de réduction-majoration" vue précédemment.

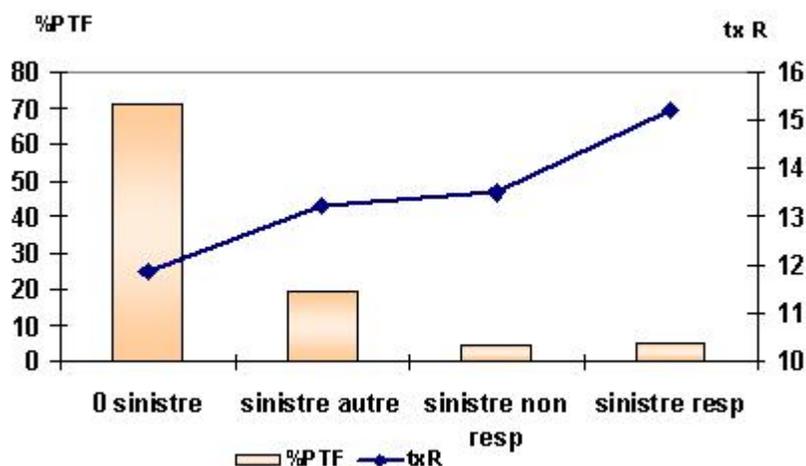


Figure n°46

Taux de résiliation en fonction de la sinistralité du contrat

#### • Valeur client et agent

Les clients multi-équipés sont moins enclin à résilier leur contrat automobile. L'importance de l'équipement du client, dans le comportement de résiliation est régulièrement étudiée dans le cadre des analyses commerciales.

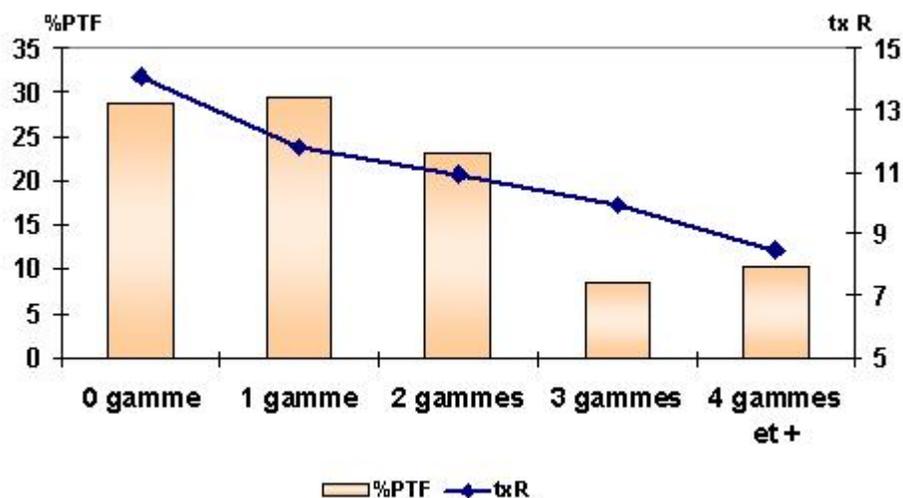


Figure n°47

Taux de résiliation en fonction du multi-équipement du client

MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Niveau 1	10,68	14,20	12,91
Niveau 2	46,17	11,75	46,18
Niveau 3	35,09	10,87	32,47
Non Renseigné	8,05	12,32	8,44

Figure n°48  
Taux de résiliation en fonction de l'apporteur

Le comportement du producteur semble avoir une incidence sur les résiliations. Les agents ayant une bonne notation "résultats techniques / production" ont tendance à avoir également une meilleure conservation de leur portefeuille.

## 2.4 Analyse des composantes multiples

Dans le cadre de notre analyse, nous considérons la variable "résilié oui/non" comme une variable supplémentaire.

### • Décomposition de l'inertie

L'analyse est réalisée à partir de 25 variables, soit 118 modalités. L'inertie totale est donc de :

$$I = \frac{c}{p} - 1 = \frac{118}{25} - 1 = 3,72$$

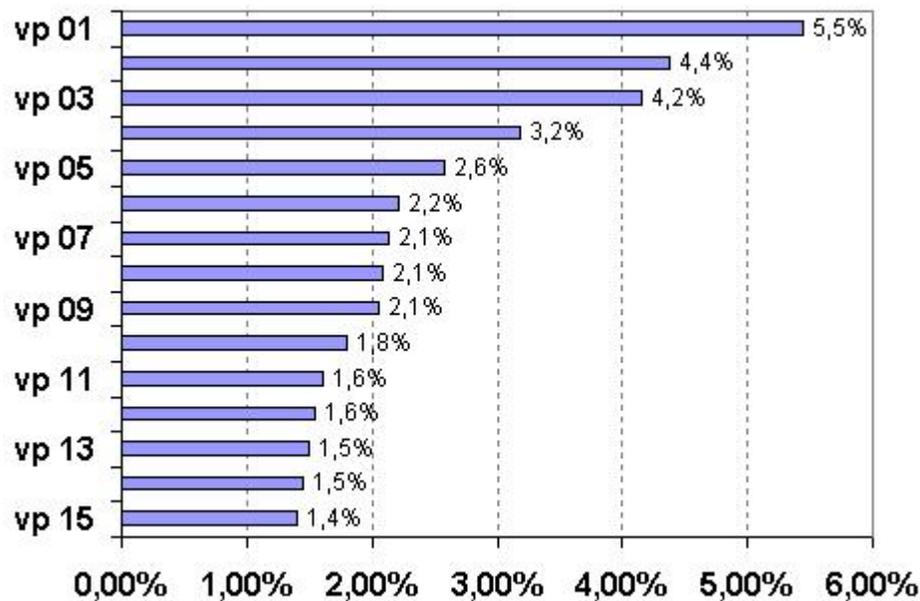


Figure n°49  
Valeurs propres et contributions associées

Les valeurs propres, ainsi que l'inertie associée, sont très faibles du fait du nombre élevé de variables et de modalités. De plus, le nombre de modalités varie fortement entre les variables, du fait notamment de la présence de variables binaires.

Le nombre de valeurs propres différentes, associées à des axes factoriels est de :

$$c - p = 118 - 25 = 93$$

Dans le cadre de l'analyse, nous ne retiendrons que les 3 premiers axes qui impliquent 14% de l'inertie. Il serait tout à fait possible d'étudier des axes supplémentaires (les 2 axes suivants impliquent 5.75% d'information supplémentaire). L'analyse de ses axes est cependant moins aisée.

• **1er axe : niveau de cotisation**

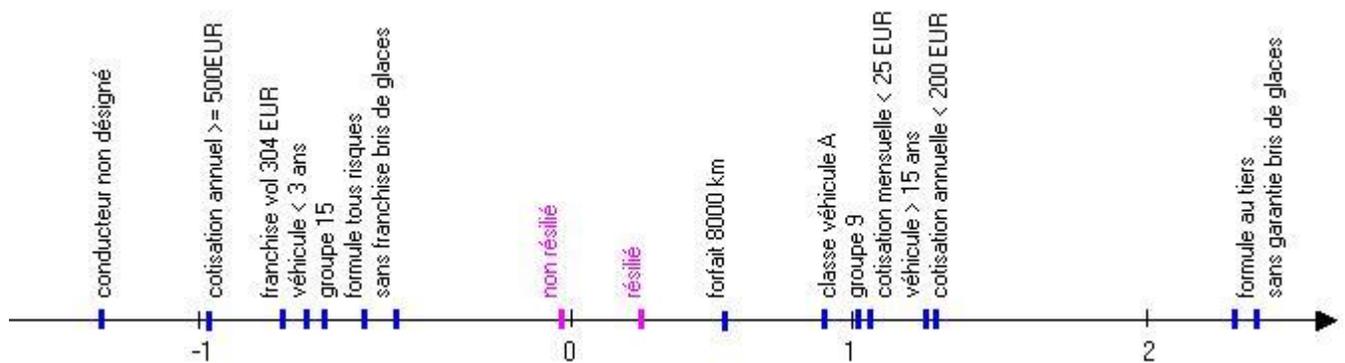


Figure n°50  
Modalités contribuant à l'interprétation de l'axe 1

On constate que deux variables ont des coordonnées positives importantes sur cet axe :

- les formules de conduite au tiers,
- les contrats sans garantie bris de glaces.

Les contributions associées (0.12 et 0.11) correspondent aux deux contributions les plus importantes de cet axe.

D'autres données ont des coordonnées positives, mais contribuent moins fortement à l'axe (exemple : cotisations faibles véhicules anciens).

Les données avec des coordonnées négatives ont des contributions moindres mais permettent d'expliquer la signification de l'axe, en s'appuyant sur le principe d'opposition :

- les contrats sans conducteur désigné,
- les contrats à cotisations élevées,
- les véhicules récents et puissants (groupe élevé).

Ainsi, on constate que le premier axe oppose les contrats ayant des niveaux de cotisations importants, à gauche de l'axe, et les contrats à faible cotisation. Les contrats à forte cotisation sont caractérisés par des véhicules récents, de grande capacité (groupe 15 et plus), assurés en tous risques. Les contrats à faible cotisation contribuent positivement à l'axe et se caractérisent par de faibles niveaux de garanties et des véhicules de petite valeur (équipement, ancienneté).

La représentation de la variable supplémentaire "résilié" n'est pas optimale sur cet axe. Il convient toutefois de noter que les contrats peu équipés paraissent les plus exposés aux résiliations.

#### • 2ème axe : le temps

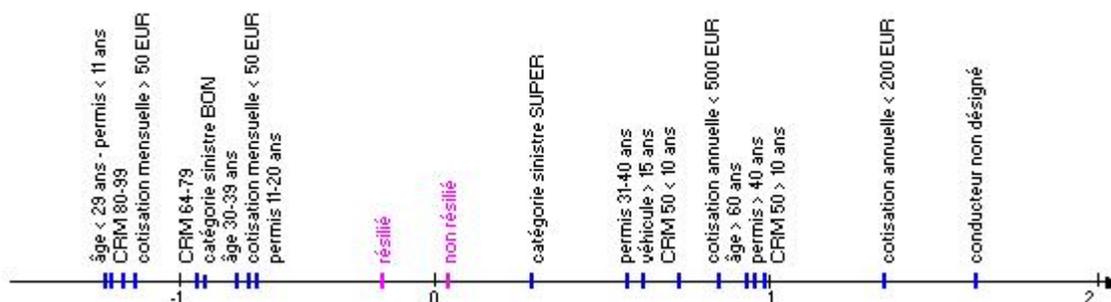


Figure n°51  
Modalités contribuant à l'interprétation de l'axe 2

Il oppose les jeunes conducteurs, dont les coordonnées sont négatives, aux conducteurs expérimentés qui contribuent positivement à l'axe :

- confrontation au niveau de l'âge du conducteur principal : les "moins de 29 ans" ont une coordonnée à -1.16 alors que les "plus de 60 ans" ont comme coordonnée 0.85,
- confrontation similaire sur l'ancienneté de permis.

On constate que le niveau de cotisation contribue également à l'analyse de cet axe. Dans la mesure où le niveau de cotisation s'avère plus élevé pour les conducteurs avec peu d'expérience, l'hypothèse de la contribution de la cotisation à l'axe s'avère, par conséquent, justifiée.

La représentation de la variable supplémentaire "résilié" est peu représentative sur cet axe. Nous constatons toutefois que les conducteurs ayant peu d'expérience sont plus enclins à résilier leur contrat automobile.

• **3ème axe : les conducteurs non désignés**

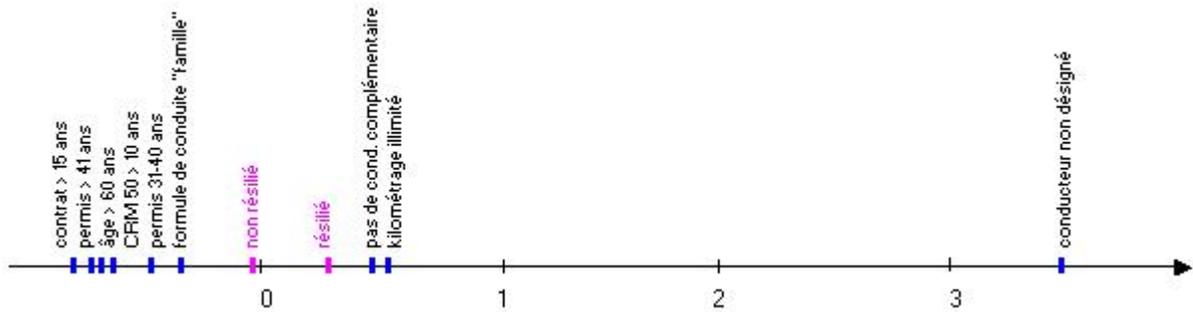


Figure n°52  
Modalités contribuant à l'interprétation de l'axe 3

**Le troisième axe met en évidence les modalités "extrêmes" :** il oppose le portefeuille des "non désignés", au comportement atypique, du reste du portefeuille :

- les données "âge non désigné", "ancienneté de permis non désigné" et "personne morale" correspondent aux données contribuant le plus fortement : leur coordonnée et leur contributions sont identiques (4.6 et 0.18) car elles représentent exactement les mêmes populations,
- la coordonnée et la contribution (3.3 et 0.14) de la donnée "formule de conduite prêt de volant" sont légèrement plus faibles car elles comprennent, en plus des contrats "personnes morales", des contrats avec désignation de multiples conducteurs sans liens de parentés.

Bien que la représentation de la variable supplémentaire ne soit pas optimale, tout tend à indiquer que les contrats extrêmes, notamment les contrats sans désignation de conducteur, sont plus exposés aux résiliations.

## Chapitre 3

# Modélisation des résiliations

### 3.1 Introduction

La régression constitue une méthode incontournable en traitement des données dans le cadre de la démarche de modélisation d'une variable. Il s'agit alors de déterminer une relation entre une variable dépendante  $Y$  et des variables explicatives  $X_i$ . La méthode la plus utilisée est la régression linéaire multiple. Elle s'applique lorsque la variable à expliquer et les variables prédictives s'avèrent continues. Dans le cadre d'une variable à expliquer  $Y$  qualitative et en particulier binaire, il convient de réaliser une régression logistique ou, dans certains cas, une analyse discriminante.

Ce chapitre propose, dans un premier temps, de réaliser une analyse discriminante binaire afin de définir une fonction de résiliation permettant d'appréhender, de manière mathématique, le comportement des clients vis à vis de la conservation de leur contrat automobile.

Après avoir réalisé une analyse discriminante et interprété les résultats qui en découlent, nous serons amenés à réaliser une régression logistique afin d'analyser l'événement "contrat résilié". Nous analyserons par la suite la robustesse de notre modèle.

## 3.2 Analyse discriminante

La part des résiliés dans la base initiale étant relativement faible (autour de 12% de la base étudiée), nous allons réaliser l'analyse discriminante sur une base, issue de la base initiale, réalisée en sur-pondérant la proportion des résiliés. Pour y parvenir, il convient de répartir la base initiale en 2 bases : les "résiliés" et les "non résiliés". Sur la base des "non résiliés", nous réalisons un tirage aléatoire, basé sur une loi uniforme afin d'obtenir un nombre de "non résiliés" équivalent au nombre de résiliés.

A partir de cette base sur-pondérée en contrats résiliés, nous constituons deux bases complémentaires, par tirage aléatoire sans remise et s'appuyant sur une loi uniforme :

- Le premier échantillon, constitué d'environ 80% de la base sur-pondérée en contrats résiliés, sert de base de référence pour réaliser l'analyse discriminante,
- le second échantillon permet quant à lui d'analyser la solidité du modèle réalisé.

### 3.2.1 Détermination des critères les plus discriminants

Dans un premier temps, nous allons hiérarchiser les variables, pour leur effet discriminant, en utilisant la méthode "pas à pas" ou "Stepwise". Cette méthode de hiérarchisation permettra éventuellement, par la suite, de réduire le nombre de variables explicatives afin de gagner en stabilité sur le modèle et d'éliminer les variables redondantes. Le fait de réduire le nombre de variables explicatives permet de gagner en stabilité, mais ne donne pas forcément les meilleurs pourcentages de "correctement classés".

Le principe consiste à partir avec un modèle sans variable. A chaque étape, l'ensemble du modèle est examiné :

- Si la variable contribuant le moins au modèle ne discrimine pas suffisamment le modèle, alors elle est supprimée,
- Si une variable est absente du modèle, elle peut être ajoutée s'il s'agit de la plus discriminante parmi toutes les variables absentes et qu'elle remplit les conditions de discrimination fixées.

Le modèle réalisé est obtenu avec des niveaux d'entrée et de sortie fixés à 15%. Ces taux sont comparés à la valeur F de chaque variable.

L'ordre d'entrée des critères explicatifs dans le modèle est le suivant :

N° classement	Variable	N° classement	Variable
1	Coefficient de Réduction-Majoration	6	Zone du risque circulation
2	Niveau de Garanties	7	Ecart avec le barème MMA
3	Ancienneté de contrat	8	Antécédents sinistres
4	Mises en demeure	9	Forfait kilométrique souscrit
5	Evolution de la cotisation à l'échéance	10	Ancienneté véhicule

Figure n°53

10 variables les plus discriminantes de l'Analyse Discriminante

En hiérarchisant les variables explicatives en fonction de leur effet discriminant, nous constatons que les variables "Age du conducteur principal" et "Ancienneté de permis du conducteur principal" sont peu discriminantes (20 et 21ème positions). Comme nous l'avons constaté au chapitre précédent, les variables "Age du conducteur principal" et "Ancienneté de permis du conducteur principal" sont fortement corrélées à la variable CRM, du fait même de la construction de cette variable (corrélation supérieur à 60%).

Lors de l'analyse univarié, nous avons déjà constaté la forte relation linéaire entre le taux de résiliation et le niveau de CRM.

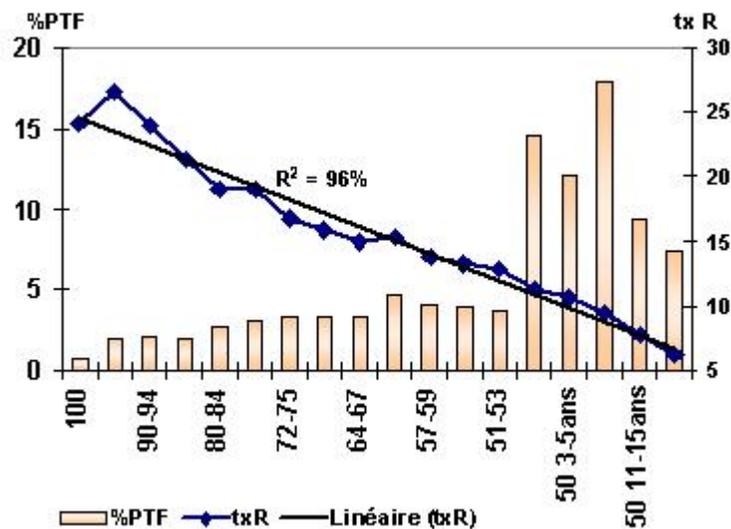


Figure n°54

Taux de résiliation selon le niveau de CRM

Tout comme l'âge et l'ancienneté de permis, les données "niveau de franchise vol" et "présence d'un garage" sont peu discriminantes. Par construction, elles sont corrélées avec "le niveau de garanties" :

- pour le niveau de franchise vol : sur les 4 modalités, la modalité "sans vol" correspond exclusivement aux formules aux tiers (11% du portefeuille) alors que la franchise 0EUR ne peut être souscrite qu'en formule tous risques,
- pour le garage : la modalité "sans objet" concerne exclusivement les formules au tiers.

A l'exception des formules "tous risques +" qui ont un comportement assez atypiques (taux de résiliation de 11% contre 9% pour une formule "tous risques classiques"), il existe une relation linéaire très importante (corrélation de 98%) entre le taux de résiliation et la formule de garanties. Les formules "tous risques +" font baisser ce taux à 82%.

Comme le montre la *figure n°40* au chapitre précédent, il existe une forte relation linéaire entre le taux de résiliation et l'ancienneté de contrat. Cette relation est "négative" dans le sens où plus le contrat est ancien, plus la propension à résilier est faible.

Bien que la donnée "évolution de la prime" nous paraisse avoir un lien important avec les résiliations, cette donnée ne ressort qu'en cinquième position lors de la modélisation.

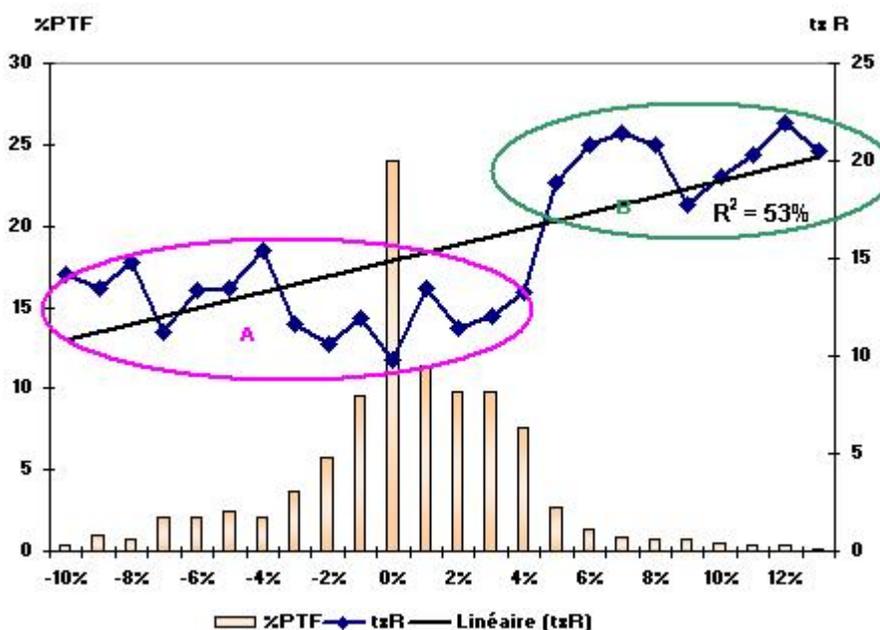


Figure n°55  
Taux de résiliation selon l'évolution de la prime

Cette mauvaise position peut notamment s'expliquer par l'évolution du taux en fonction de cette donnée : plus qu'une relation linéaire, le lien peut être assimilé à deux relations linéaires (une pour la partie A, l'autre pour le B) ou plus exactement deux paliers.

Lorsque l'on observe le modèle avec 5 variables, nous constatons que les corrélations des variables absentes avec le modèle sont toutes supérieures ou égales à 25%. Cette corrélation passe à 45% lorsque le modèle possède les 10 variables explicatives. Nous constatons donc que les variables sont fortement corrélées entre elles.

Même si certaines variables s'avèrent corrélées à d'autres, plus discriminantes, l'ensemble des variables initialement présentes peut être conservé dans l'analyse discriminante (aucune donnée n'est rejetée à l'entrée du modèle). Chacune de ces variables apporte donc une information complémentaire, non portée par les précédentes, sur l'analyse des "contrats résiliés" (exemple : l'âge du conducteur principal permet, par exemple, d'isoler la population des "plus de 70 ans", qui voient leur taux fortement augmenter du fait de la mortalité).

### 3.2.2 Résultats de l'analyse discriminante

Pour réaliser l'analyse discriminante, nous allons utiliser la procédure DISCRIM de SAS. Elle permet d'obtenir une analyse discriminante probabiliste afin de classer chaque individu, à partir des valeurs prises par les variables prédictives, dans la catégorie à laquelle il a le plus de chances d'appartenir.

La procédure DISCRIM s'appuie sur la méthode bayésienne qui consiste à calculer, à posteriori, les probabilités d'appartenance à un groupe :

$$P(Y^i = a/X^i) = \frac{P(X^i/Y^i = a).P(Y^i = a)}{\sum_{k=1}^x P(X^i/Y^i = k).P(Y^i = k)}$$

Les coefficients issus de la modélisation dénotent des difficultés pour scinder proprement la base selon la variable de classe. en effet, nous constatons que les coefficients du modèle "résilié" et "non résilié" sont assez proches, pour chaque variable, ce qui laisse présager des difficultés de scission de la population.

L'effet des données "âge du conducteur" et "ancienneté de permis", fortement corrélées, s'anulent. Comme nous l'avons constaté précédemment, il s'agit des données les moins discriminantes du modèle. En plus d'être corrélées entre elles, elles sont également étroitement liées au niveau de coefficient de réduction majoration (CRM), première variable explicative.

Le facteur temps contribue aussi bien de manière positive que négative en fonction des variables :

- le temps s'avère être un critère fidélisant dans le cadre du conducteur et de l'ancienneté de contrat. Les CRM 50 sont en règle générale des cibles privilégiées pour les assureurs et bénéficient, par conséquent, de mesures spécifiques.
- à contrario, il a un effet négatif (augmentation des résiliations) dans le cadre de l'ancienneté du véhicule. Cet effet peut s'expliquer par le fait que le changement de véhicule est

généralement propice à la prospection et au changement d'assureur, comme nous l'avons constaté au chapitre précédent.

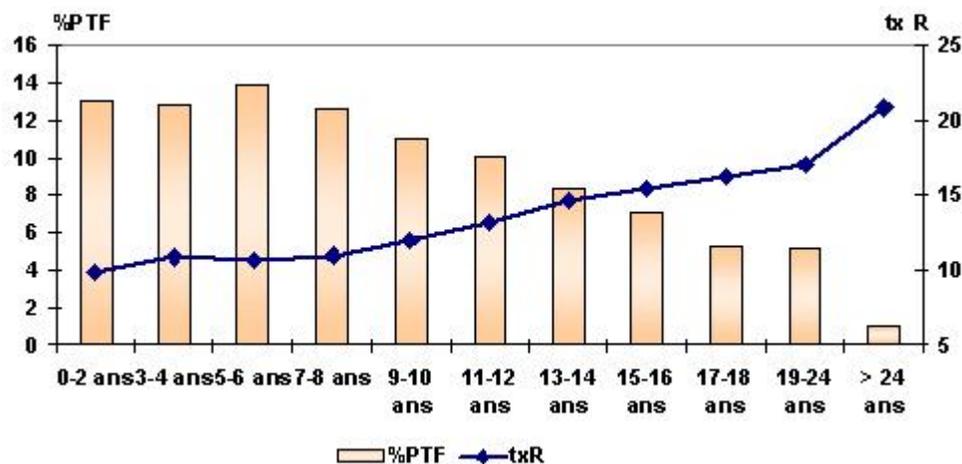


Figure n°56  
Taux de résiliation selon l'ancienneté du véhicule

Le risque de résiliation semble plus élevé dans les zones urbaines. Cette hypothèse est confortée par une plus grande stabilité, dans notre portefeuille, des agriculteurs.

### 3.2.3 Estimation du modèle

L'estimation de la fiabilité du modèle s'appuie sur deux principales estimations :

- l'estimation de re-substitution,
- l'estimation de l'échantillon test.

L'estimation par re-substitution est obtenue en appliquant la règle de décision à la base qui a été utilisée pour constituer le modèle. Cette restitution nous permet d'obtenir un taux apparent d'erreur, correspondant au taux des individus mal classés.

L'estimation sur l'échantillon test consiste à appliquer le modèle sur une base n'ayant pas servi à son obtention. Ce résultat est généralement moins bon que le résultat par re-substitution. Il permet ainsi d'avoir une estimation de la fiabilité du modèle.

Dans les deux cas, l'estimation repose sur l'analyse du tableau suivant :

	$\hat{Y}^1$	$\hat{Y}^2$	
$Y^1$	$n_{11}$	$n_{12}$	$n_{1.}$
$Y^2$	$n_{21}$	$n_{22}$	$n_{2.}$
	$n_{.1}$	$n_{.2}$	$n$

où  $\hat{Y}^k$  représente le résultat du modèle pour la modalité  $k$ .

Le taux d'erreur général du modèle est défini par  $E = \frac{n_{21} + n_{12}}{n}$  alors que le taux d'erreur de la modalité  $i$  de la variable de classe est défini par  $E_i = \frac{n_{i\bar{i}}}{n_i}$  où  $\bar{i}$  représente la seconde modalité de la variable de classe.

Dans le cadre de notre étude, les résultats de la re-substitution donne les résultats suivants :

	non résilié (NR)	résilié (R)	
non résilié (NR)	64%	36%	50%
résilié (R)	40%	60%	50%
	52%	48%	

$$E = 38\%, E_{NR} = 36\% \text{ et } E_R = 40\%$$

Les taux d'anomalies constatés, dans le cadre de la méthode de re-substitution sont relativement importants. Le modèle linéaire ne paraît pas être le modèle le plus adéquat pour modéliser les résiliations en assurance automobile.

Les résultats en appliquant le modèle sur la base de test sont les suivants :

	non résilié (NR)	résilié (R)	
non résilié (NR)	43%	57%	50%
résilié (R)	23%	77%	50%
	33%	67%	

$$E = 40\%, E_{NR} = 57\% \text{ et } E_R = 23\%$$

Les taux d'anomalies constatés sur la base de test sont supérieurs à ceux observés sur la base de modélisation. Les résultats obtenus montrent que le modèle obtenu ne sépare pas suffisamment les groupes déterminés par la variable de classe. De plus, nous constatons une sur-représentation des contrats "résiliés" dans la population observée des "non résiliés".

La modélisation par analyse discriminante ne permet pas d'avoir des résultats satisfaisants et il convient de réaliser une modélisation non linéaire du problème.

### 3.2.4 Analyse discriminante avec un nombre réduit de variables explicatives

Lors de la hiérarchisation du pouvoir discriminant des données, nous avons constaté que certaines variables étaient peu discriminantes car fortement corrélées à d'autres. Ce constat a été conforté par l'analyse des résultats de l'analyse discriminante. Il convient donc de réaliser une analyse simplifiée, utilisant les variables les plus significatives.

Le modèle que nous allons étudier se voit amputé des variables :

- âge et permis du conducteur principal, trop corrélées avec le niveau de coefficient réduction-majoration,
- franchise vol, corrélée avec la souscription des formules de garanties,
- groupe du véhicule, corrélée, du fait de sa construction, à la classe du véhicule.

les 5 dernières données étant fortement corrélées avec des données déjà présentes dans le modèle, nous pouvons réaliser l'analyse discriminante en supprimant les données "âge", "ancienneté de permis", "groupe du véhicule", "franchise bris de glaces" et "franchise dommages du modèle".

Nous constatons ainsi les résultats suivants en appliquant les nouveaux coefficients sur la base ayant servi à constituer le nouveau modèle :

	non résilié (NR)	résilié (R)	
non résilié (NR)	64%	36%	50%
résilié (R)	40%	60%	50%
	52%	48%	

$$E = 38\%, E_{NR} = 36\% \text{ et } E_R = 40\%$$

Les résultats sont proches de ceux obtenus à partir du modèle complet (intégration de toutes les variables explicatives). Cela tend à confirmer que les 5 variables non intégrées étaient peu discriminantes quant au risque de résiliation.

Les résultats en appliquant le modèle sur la base de test sont les suivants :

$$E = 38\%, E_{NR} = 36\% \text{ et } E_R = 40\%$$

Le modèle obtenu s'avère plus stable que le modèle réalisé à partir de l'exhaustivité des variables explicatives. Dans le cas présent, il est préférable d'utiliser un modèle ayant un plus faible nombre de variables.

### 3.3 Modélisation logistique

Le risque de résiliation étant sous représenté dans la population (inférieur à 15%) l'analyse s'appuie sur une base constituée par tirage aléatoire simple, permettant de sur-pondérer la part des contrats résiliés. La méthode d'échantillonnage utilisée a été décrite succinctement dans le cadre du chapitre sur l'analyse discriminante.

Nous connaissons la valeur des  $p$  variables explicatives et nous cherchons à estimer la probabilité que l'événement "résiliation" se réalise.

#### 3.3.1 Détermination des critères les plus discriminants

Tout comme dans le cadre de l'analyse discriminante, il est possible de hiérarchiser les variables explicatives par leur contribution dans la modélisation. Nous utilisons, de la même manière que précédemment, la méthode "Stepwise".

N° classement	Variable	N° classement	Variable
1	Coefficient de Réduction-Majoration	6	Présence de mise en demeure
2	Niveau de Garanties	7	Antécédents sinistres
3	Ancienneté de contrat	8	Forfait kilométrique souscrit
4	Evolution de la cotisation à l'échéance	9	Ancienneté véhicule
5	Age du conducteur principal	10	Zone du risque circulation

*Figure n°57*  
10 variables les plus discriminantes de la modélisation logistique

#### • Variables explicatives dans les modélisations discriminante et logistique

On constate que certaines variables ressortent comme explicatives aussi bien dans le cadre de l'analyse discriminante que de la régression logistique. C'est le cas notamment du coefficient de réduction-majoration, du niveau de garanties et de l'ancienneté de contrat. Ces données s'avèrent être les trois principaux critères discriminants tant en analyse discriminante qu'en régression logistique.

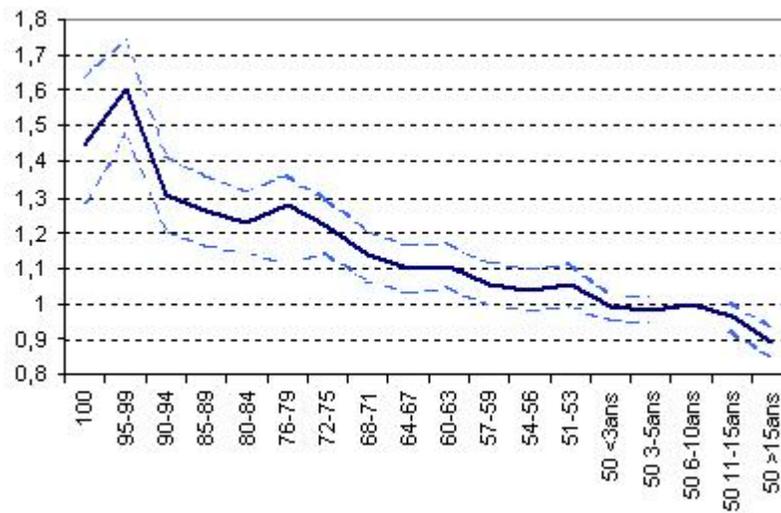


Figure n°58  
Coefficient issu de la logistique pour le CRM et intervalle de confiance à  $\pm 5\%$

La donnée "évolution de la prime à l'échéance" s'avère plus discriminante dans le cadre de la régression logistique. Cette donnée devient ainsi le quatrième critère explicatif. Un rapide ajustement (non optimisé) du taux de résiliation en fonction de cette donnée nous permet d'illustrer la plus-value apportée par le modèle logistique :

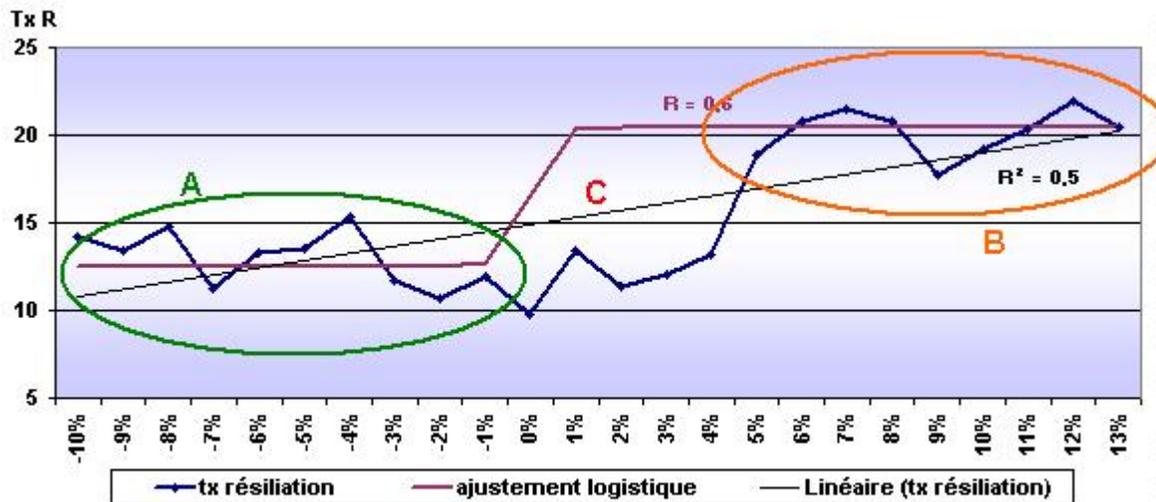


Figure n°59  
Ajustement du taux de résiliation en fonction de l'évolution de prime à l'échéance

Cette amélioration de l'ajustement est principalement liée à l'interprétation plus fine des valeurs extrêmes, de la variable explicative (zones A et B). Les évolutions de primes élevées, à la hausse ou à la baisse, bénéficient d'un taux de résiliation relativement stable. Ce constat ne peut pas

être modélisé proprement par un ajustement linéaire. La modélisation doit, en effet, tenir compte des deux paliers.

On constate également qu'un ajustement logistique optimisé sur la zone C (de -3% à +3%) permettrait de s'approcher fortement du comportement réel.

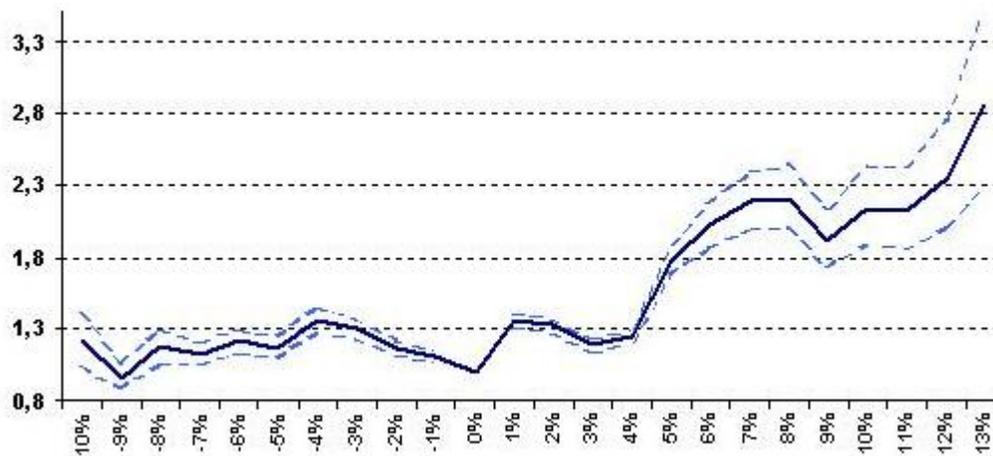


Figure n°60

Coefficient issu de la logistique pour l'évolution de cotisation à l'échéance et intervalle de confiance à  $\pm 5\%$

On remarque un intervalle de confiance beaucoup plus large sur les évolutions de cotisation élevées du fait de la répartition de la population (faible densité sur ces tranches).

• Nouvelle donnée explicative dans la modélisation logistique

La donnée "âge du conducteur principal", bien que fortement corrélée à d'autres données explicatives (notamment le CRM), ressort cependant parmi les dix variables les plus explicatives du modèle.

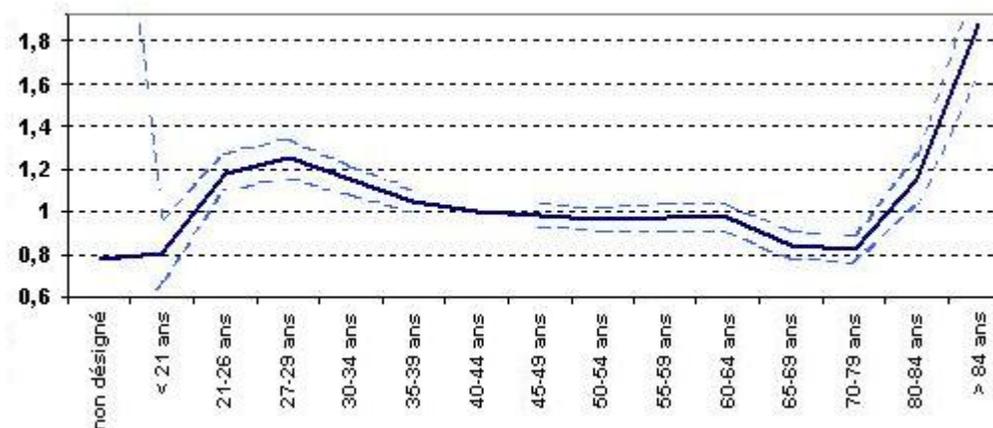


Figure n°61

Coefficient issu de la logistique pour l'âge du conducteur et intervalle de confiance à  $\pm 5\%$

En observant les résultats afin de déterminer l'information complémentaire par rapport à la donnée CRM, on observe que cette donnée permet :

- d'identifier les contrats sans désignation de conducteur : l'intervalle de confiance est cependant extrêmement étendu du fait de la faible proportion de ces contrats dans le portefeuille.
- de cibler les jeunes conducteurs,
- d'observer l'homogénéité de la tranche des 22 - 65 ans,
- d'identifier un comportement spécifique au delà de 65 ans.

#### • Les autres données explicatives du modèle

La classe du véhicule (11ème donnée explicative) a une amplitude faible entre les différents coefficients. Elle s'avère explicative uniquement dans le cadre des classes de véhicules supérieures à E. Pour les autres valeurs, l'intervalle de confiance englobe généralement les valeurs précédentes et suivantes, ce qui dénote un problème de fiabilité dans les coefficients obtenus et un manque de stabilité du modèle. Dans le même temps, la donnée "groupe du véhicule" (20ème donnée explicative) n'est pas significative car trop fortement corrélée avec la "classe du véhicule".

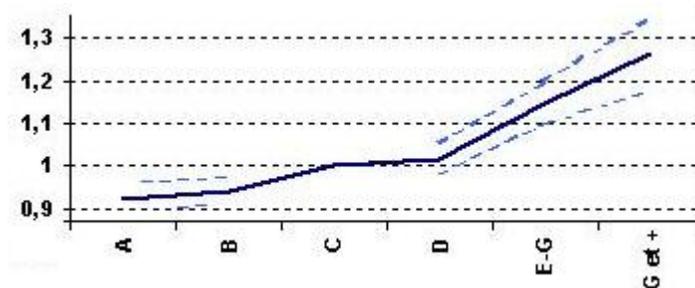


Figure n°62

Coefficient issu de la logistique pour la classe du véhicule et intervalle de confiance à  $\pm 5\%$

La "formule de conduite" (13ème donnée explicative) accentue l'écart entre contrat avec conducteur désigné et contrat sans conducteur désigné. Elle n'apporte de l'information que pour une faible population, tout comme la donnée "agriculteur", ce qui entraîne son classement au delà de la 10ème place des critères explicatifs.

	Coeff.	Intervalle +/- 95%	
<b>Exclusive</b>	1,008	0,986	1,031
<b>Famille</b>	1,000		
<b>Tout conducteur</b>	1,373	1,334	1,413

Figure n°63

Coefficient issu de la logistique pour la formule de conduite et intervalle de confiance à  $\pm 5\%$

### 3.3.2 Régression logistique en réduisant le nombre de critères explicatifs

Afin de faciliter l'intégration de la modélisation logistique dans la suite de l'étude, nous préférons modéliser la loi de survie et non de résiliation.

Nous retenons les critères suivants, dans le cadre de la modélisation logistique :

- Le niveau de CRM,
- l'ancienneté de contrat,
- l'ancienneté de véhicule,
- l'âge du conducteur principal : 3 modalités, pour gérer la recrudescence des résiliations sur les extrêmes,
- les antécédents sinistres,
- le forfait kilométrique souscrit,
- la formule de garanties,
- la présence de mises en demeure,
- l'évolution de la cotisation à l'échéance.

#### Adéquation du modèle

Dans le cas présent, nous allons considérer la statistique  $\hat{C}$  de Hosmer-Lermeshow (1989). Cette statistique se calcule en triant les probabilités estimées et en les répartissant en déciles.

Group	Non résilié			Résilié		
	Réel (R)	Modélisé (M)	(R-M) <sup>2</sup> / M	Réel (R)	Modélisé (M)	(R-M) <sup>2</sup> / M
1	4 774	4 692	1,42	15 226	15 308	0,44
2	6 765	6 696	0,70	13 236	13 305	0,35
3	7 920	7 886	0,15	12 082	12 116	0,10
4	8 893	8 819	0,62	11 107	11 181	0,49
5	9 469	9 662	3,87	10 532	10 339	3,62
6	10 367	10 500	1,70	9 637	9 504	1,88
7	11 255	11 352	0,83	8 747	8 650	1,09
8	12 128	12 229	0,84	7 871	7 770	1,32
9	13 265	13 206	0,26	6 734	6 793	0,51
10	15 015	14 806	2,94	4 977	5 186	8,41

Figure n°64  
Calcul de la statistique Hosmer-Lemeshow

$$\hat{C} = \sum \frac{(R - M)^2}{M}$$

$$\hat{C} = 31.53 \text{ et } P(\chi^2(8ddl) > \hat{C}) = 0.01\%$$

L'hypothèse d'adéquation du modèle est donc acceptée.

#### Résultats du modèle

	non résilié (NR)	résilié (R)	
non résilié (NR)	63%	37%	50%
résilié (R)	39%	61%	50%
	51%	49%	

$$E = 38\%, E_{NR} = 37\% \text{ et } E_R = 39\%$$

Les résultats sont proches de ceux obtenus dans le cadre de l'analyse discriminante. On constate toutefois une meilleure répartition dans les classements (résultats équivalents entre "non résiliés" et "résiliés").

En appliquant le modèle logistique à notre échantillon test, nous constatons que le modèle est stable.

	non résilié (NR)	résilié (R)	
non résilié (NR)	62%	38%	50%
résilié (R)	38%	62%	50%
	51%	49%	

$$E = 38\%, E_{NR} = 38\% \text{ et } E_R = 38\%$$

### 3.3.3 Ajustement des coefficients

Comme le montre l'exemple ci-dessous, les coefficients obtenus lors de la régression logistique génèrent deux biais principaux :

- les intervalles de confiance englobent généralement la valeur des modalités juste inférieures et supérieures du critère.
- les coefficients peuvent entraîner des fluctuations contraires, au fil du temps, difficilement compréhensibles (exemple : CRM compris entre 70 et 80).

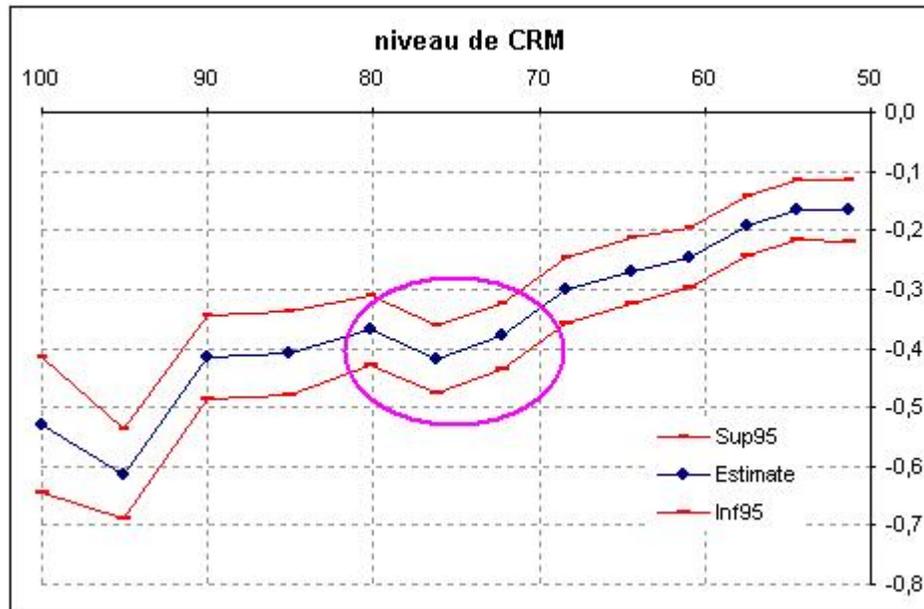


Figure n°65  
Coefficients liés au niveau de CRM

Ces deux objections nous amènent à modifier les coefficients des données liées au temps, afin de consolider les résultats dans la suite de ce mémoire. Les coefficients subissant une évolution sont les suivants :

- Le niveau de CRM,
- l'ancienneté de contrat,
- l'ancienneté de véhicule,
- l'évolution de la cotisation.

*Ancienneté de contrat*

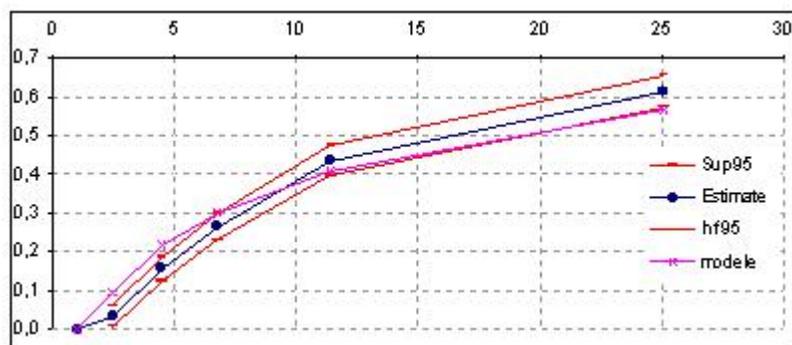


Figure n°66  
Ajustement des coefficients liés à l'ancienneté de contrat

Dans le cadre de l'ancienneté de contrat, nous considérons un ajustement logarithmique pour se rapprocher à plus près du comportement des coefficients : augmentation de la rétention des clients, au fil du temps, avec cependant une pente de plus en plus faible.

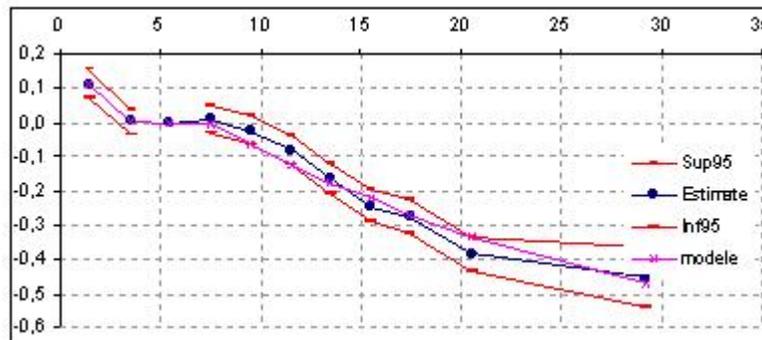
Le taux de corrélation est alors de 93%, l'écart le plus important se situant sur les tranches d'ancienneté élevées (ajustement en limite de l'intervalle de confiance 95%).

#### *Ancienneté de véhicule*

Les coefficients issus de la modélisation permettent de déterminer trois phases importantes :

- pour les véhicules récents : faible probabilité de résiliation. La résiliation se fait à l'avenant de changement de véhicule et non pas au terme,
- pour les durées comprises entre 3 et 7 ans : il s'agit généralement de la durée de détention du véhicule. Ces tranches d'ancienneté ont un comportement relativement similaire. Elles sont donc ajustées par une droite horizontale,
- au delà de 7 ans : le taux de conservation du contrat diminue, du fait notamment du changement de parc. Nous ajusterons cette partie de manière polynomiale (le taux de détention des contrats s'assimile à la disparition du risque sur les véhicules anciens).

Le taux de corrélation est alors de 99%.



*Figure n°67*

*Ajustement des coefficients liés à l'ancienneté du véhicule*

## Niveau de coefficient de réduction-majoration (CRM)

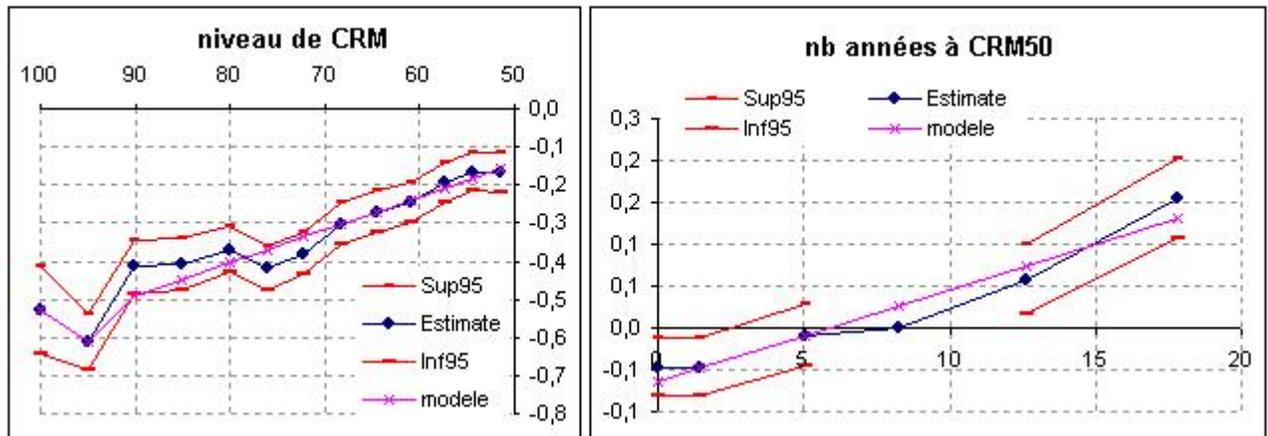


Figure n°68  
Ajustement des coefficients liés au CRM

Trois effets se combinent :

- une forte chute de la conservation des contrats lorsque le CRM avoisine les 95 : il s'agit des conducteurs qui ne subissent plus la majoration novice. Cette période est alors propice à la recherche d'un meilleur tarif et donc à la résiliation,
- entre 95 et 50, le taux de rétention augmente de manière linéaire, au fur et à mesure que le coefficient s'améliore.
- la conservation des contrats CRM 50 est accélérée avec la durée de détention du CRM 50. Cette observation se concrétise par un ajustement linéaire, dont la pente est supérieure à l'ajustement des CRM 90 à 50.

Le taux de corrélation est alors de 99%.

## Evolution de la cotisation à l'échéance

Trois effets sont observés :

- en dessous de -2%, on observe une légère diminution de la détention. Cette partie paraissant très irrégulière, sera ajustée à notre niveau, par un ajustement linéaire afin de limiter l'effet des fluctuations,
- entre -2% et 4%, la conservation des contrats diminue fortement et tend à se stabiliser entre 3 et 4%. Il s'agit d'un seuil d'augmentation semblant être accepté par une partie de la population. Cet intervalle sera ajusté de manière polynomiale.
- au delà du seuil de majoration acceptable (3 - 5%), nous constatons une nouvelle fois une forte augmentation des résiliations. L'ajustement sera également réalisé de manière polynomiale.

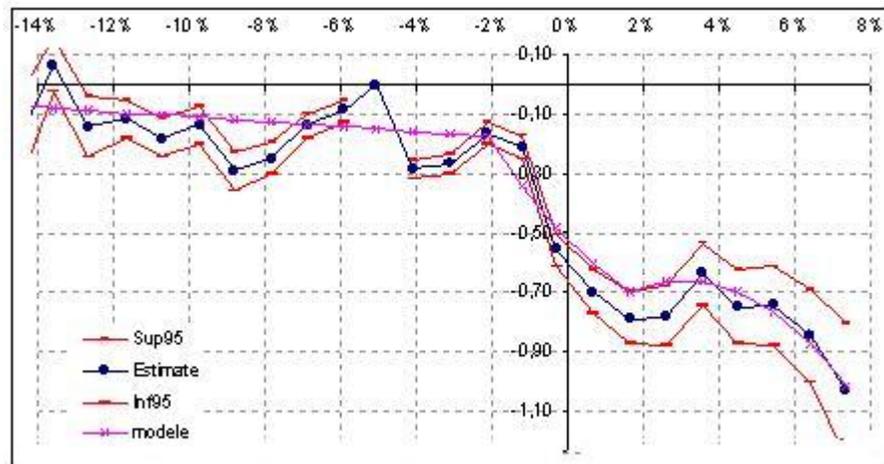


Figure n°69  
Ajustement des coefficients liés à l'évolution de la cotisation

Le taux de corrélation est alors de 96%.

# Chapitre 4

## Optimisations des ressources à l'échéance

### 4.1 Introduction

#### 4.1.1 Les techniques de tarification à l'échéance

L'échéance du contrat est l'occasion pour l'assureur de remettre à jour tous les critères déterminant la tarification, ainsi que les montants de garanties et de franchises si ils sont indexés. Il y a obligation de faire évoluer le CRM du contrat, conformément à l'article A 121-1 du code des assurances. Citons les éléments de tarification classiques pour l'assurance auto :

- l'évolution de base suivant l'anticipation de prime pure,
- la sinistralité antérieure sur un passé variable (de 12 à 36 mois),
- l'âge et l'ancienneté de permis des conducteurs,
- l'ancienneté du véhicule assuré,
- obligatoirement, le CRM ou l'ancienneté à bonus 50.

La segmentation poussée des tarifs en assurance automobile et les travaux actuariels de tarification à posteriori peuvent provoquer des évolutions importantes sur d'autres critères :

- modification de l'influence des informations techniques du véhicule,
- modification de l'influence des informations de zonage, le lieu du risque,
- modifications des frais pris en compte dans le tarif
- autres critères propres à chaque assureur.

Ces évolutions contribuent à une meilleure appréhension du risque à l'entrée, et ainsi permettent de proposer le meilleur barème technique, sur un marché où le prix est prépondérant.

Pour certaines cases de risque, toutes ces variations peuvent se cumuler ou se compenser. Pour un contrat en portefeuille, appliquer directement le barème à chaque échéance peut se révéler extrêmement dangereux pour l'assuré et l'assureur. Prenons pour illustration un cas réel de contrat d'assurance auto en portefeuille MMA.

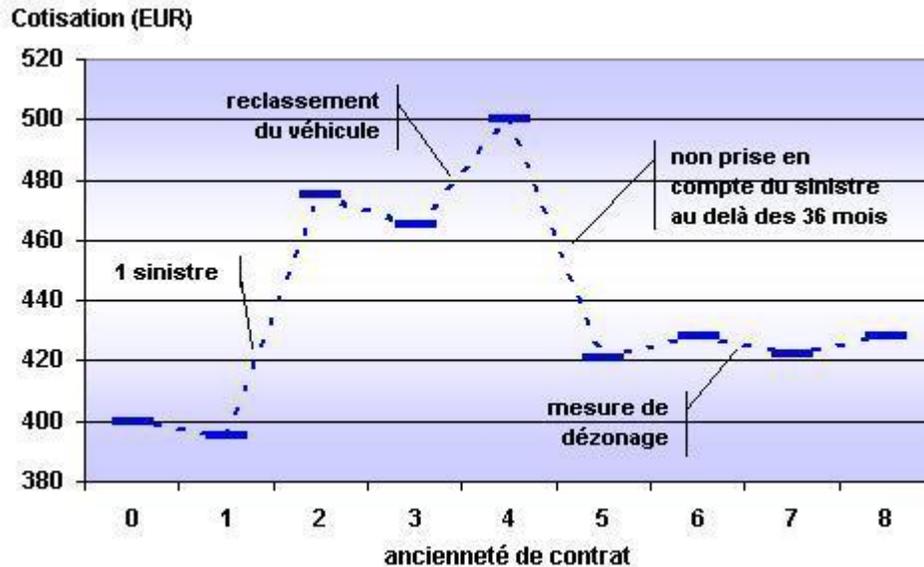


Figure n°70  
Evolution de la cotisation théorique

La sinistralité antérieure sur les 36 derniers mois est un critère de tarification assez fort. On constate, dans le cadre de l'application du barème :

- un saut de +20% à l'échéance 2 par la prise en compte d'un sinistre enregistré en année 2,
- un saut de +6.5% à l'échéance 4 occasionné par une remise à jour du classement du véhicule de l'assuré,
- une chute de -15% à l'échéance 5 par blanchiment de la période de sinistralité antérieure.

Autant les hausses subites de tarif se révèlent inacceptables pour le client, et favorisent nettement le comportement de résiliation, autant les chutes de tarif se révèlent complètement inutiles pour l'assureur : l'effet "cadeau" sans accompagnement est une perte sèche de chiffre d'affaires sans gain sur la fidélité.

Face à ces effets de tarification, les assureurs proposent deux réponses pour la tarification à l'échéance :

- la prime majorable : l'assureur décrète le taux de variation à appliquer à la prime précédente, en abstraction de la position de la prime barème,
- la retarification encadrée : l'assureur construit un encadrement "raisonnable" de variation, construit à partir de la prime précédente [plancher ; plafond]. Il tarifie au barème la nouvelle

prime, et fait payer la prime la plus proche du barème dans cet intervalle, selon la règle :

$$\text{Prime payée} = \text{Mediane} (\text{plancher} ; \text{barème} ; \text{plafond})$$

Si nous reprenons l'exemple précédent, nous pouvons illustrer le mécanisme d'encadrement de manière schématique :

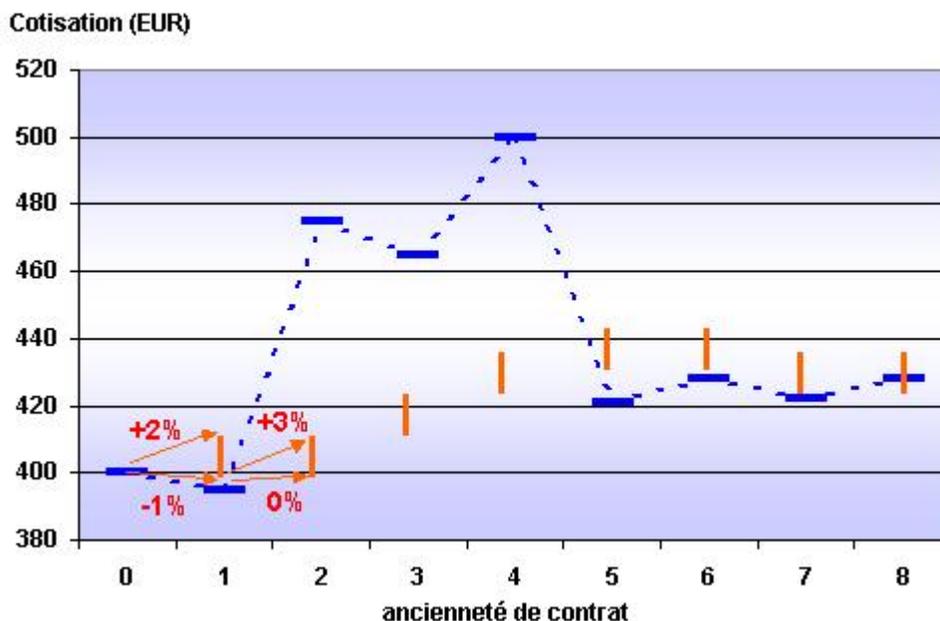


Figure n°71  
Evolution de la cotisation

On constate visuellement que :

- il existe souvent un écart entre tarif barème et tarif payé à l'échéance (sauf en exercice 8). Cet écart se mesure par la distance entre la borne de l'intervalle et le barème,
- l'encadrement réalise une sorte de lissage dans le temps de la prime. La prime payée tend vers le barème. Il tend d'autant plus vite que les intervalles sont ouverts.

L'écart obtenu entre le barème et la prime payée est un supplément de marge. Dans notre exemple sur les exercices 1, 5, 6 et 7, la prime payée est supérieure à la prime barème. A la marge normalisée et constante issue de la tarification, on peut ainsi ajouter une marge positive de l'échéance. A l'inverse, pendant les exercices 2 à 4, c'est à dire la période où le sinistre est pris en compte pour le contrat, le contrat est sous-tarifé, la marge d'échéance est négative.

Les deux prochains chapitres ont pour objectif d'explorer ces deux techniques, sous l'angle de l'optimisation du chiffre d'affaires probabilisé puis de la marge probabilisée.

### 4.1.2 La fixation du prix en assurance automobile

Traitée dans de nombreux mémoires d'actuariat, la fixation du prix trouve des réalisations intéressantes sur le marché de l'assurance automobile, considéré comme le plus élastique au prix. Pour étayer nos hypothèses, nous nous appuyons sur les importants travaux de BOULANGER et TRAN VAN LIEU (1994) concernant la détermination du tarif commercial par l'élasticité au prix.

L'élasticité de la demande par rapport au prix est un concept micro-économique bien connu du marketing actuel et qui remonte au XIX-ème siècle. La littérature économique s'entend à dater la naissance du concept d'élasticité en 1838, par les travaux du mathématicien français Augustin Cournot (1801-1877). Il est le premier à utiliser l'analyse mathématique pour construire les raisonnements de l'économie moderne. Il pose notamment la fonction de demande et sa relation avec le prix, justifiées par l'axiome "chacun cherche à tirer de sa chose ou de son travail la plus grande valeur possible".

La demande annuelle dépend du prix unitaire et, le plus souvent, de façon inverse : plus le prix augmente, et plus la quantité demandée risque de diminuer. Pour Cournot, cette dépendance n'est d'habitude pas de type proportionnel. Cette non-proportionnalité est la clé de voûte que nous allons utiliser.

L'économiste français Léon Walras (1834-1910) reprend ce concept dans le développement de "l'utilité marginale" en 1870, et une approche marginale de la consommation, donc de la dérivée première de l'expression précédente. L'économiste anglais Alfred Marshall (1842-1924) modélise, quant à lui, les phénomènes de variations de l'offre et de la demande en fonction du prix et donc de la quantité de biens échangés sur un marché concurrentiel. Ces deux auteurs chercheront à établir le point d'équilibre maximisant l'équilibre partiel des prix.

Leurs travaux ouvrent la voie à des formulations différentes de l'élasticité, suivant les écoles de micro-économie et les statistiques employées.

Les économistes américains Houttacker et Taylor (1970) posent l'équation la plus communément employée, par leurs travaux sur l'offre. Cette définition classique est reprise par l'économiste américain Lindon (1981), qui définit l'élasticité comme le rapport entre la variation de la demande et la variation correspondante du prix.

Nous appliquons cette définition à notre objectif : nous ne cherchons pas à augmenter la demande des prospects, mais à diminuer la fuite de clients en portefeuille, par une variation du prix à l'échéance.

Afin d'adapter le concept de Lindon, nous posons les hypothèses suivantes :

1. transformation du prospect en client : l'hypothèse préalable est que sur ce marché, le client ait connaissance de l'offre et de l'environnement concurrentiel. Cette hypothèse est largement vérifiée pour notre client d'assurance automobile, puisqu'il s'agit d'un marché fortement concurrentiel (cf Chapitre 1 - § 1.1.2), pour lequel nous lui indiquons le prix sur son avis d'échéance en début de période,

2. Sur le marché de l'assurance automobile en France, l'élasticité globale de la demande doit être distinguée entre fonction d'élasticité à l'entrée et à la sortie. En effet, BOULANGER et TRAN VAN LIEU affirment qu'un consommateur est plus sensible aux conditions de l'offre qu'un consommateur déjà en portefeuille, globalement satisfait de son fournisseur. Ce phénomène est rendu plus aigu sur le marché français, par la tacite reconduction. Nous nous concentrons donc sur l'élasticité à la sortie, en construisant notre raisonnement sur l'effet d'une variation de tarif à l'échéance pour les contrats en portefeuille à cette échéance.
3. Le volume de la demande observé, qui devient dans notre application le volume de contrats restés en portefeuille après l'échéance, nous paraît réducteur pour une décision d'entreprise. Si l'on se place purement du point de vue marketing, on cherchera effectivement à maximiser la fidélité, donc le nombre de contrats. Nous nous plaçons résolument sur le compte de résultat de l'entreprise, et à ce niveau, nous chercherons plutôt à maximiser soit le chiffre d'affaires restant probable, soit la marge probable à l'issue de l'exercice suivant, comme nous le verrons plus loin.

*Notations :*

- $CA_0$ , le chiffre d'affaires obtenu avec les contrats en cours à l'échéance,
- $CA_1$ , le chiffre d'affaires suivant l'échéance,
- $\Delta_i$ , la variation de tarif que nous appliquerons à chaque contrat  $i$  du portefeuille,
- $\nu_i$ , la probabilité que le contrat  $i$  soit encore en cours sur l'exercice suivant l'échéance,
- $\pi_i$ , la prime associée au contrat  $i$  après l'échéance.

En utilisant le modèle logistique de probabilité de rétention d'un contrat obtenu (Chapitre 3) [2], fonction de  $\Delta_i$ , nous pouvons exprimer l'espérance de la variable aléatoire  $CA_1$  par :

$$E(CA_1) = \sum_{i=1}^n \pi_i(1) \cdot \nu(\delta_i)$$

$$E(CA_1) = \sum_{i=1}^n \pi_i(0) \cdot (1 + \Delta_i) \cdot \nu(\delta_i) \quad [3]$$

La fonction logistique issue du chapitre 3 peut être représentée en fonction du  $\Delta_i$ , en prenant pour hypothèse sur ce graphique que tous les  $\Delta_i$  sont égaux et appliqués uniformément à tous les contrats :

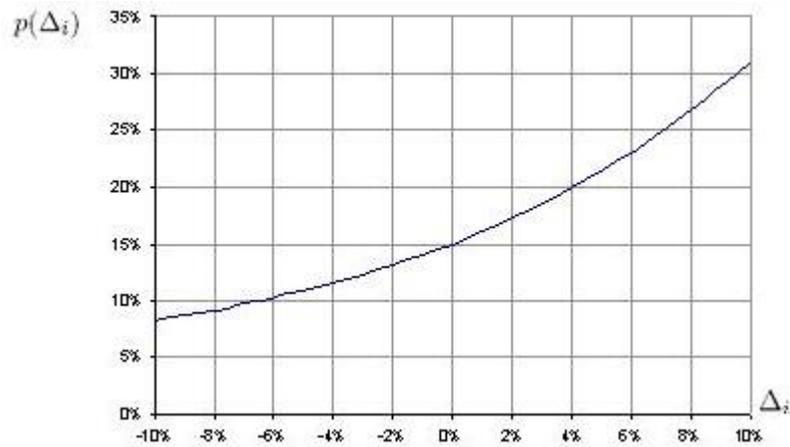


Figure n°72

Fonction logistique estimant la probabilité à résilier

On peut représenter de la probabilité complémentaire de rétention à l'issue de l'échéance,  $\nu(\Delta_i)$  :

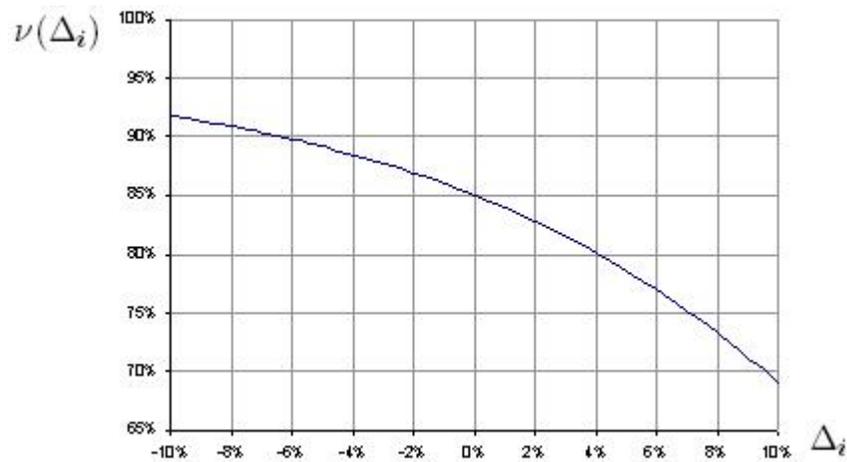


Figure n°73

Fonction logistique estimant la probabilité de rétention

En appliquant alors la relation [3], nous pouvons représenter l'espérance unitaire du contrat, à optimiser :

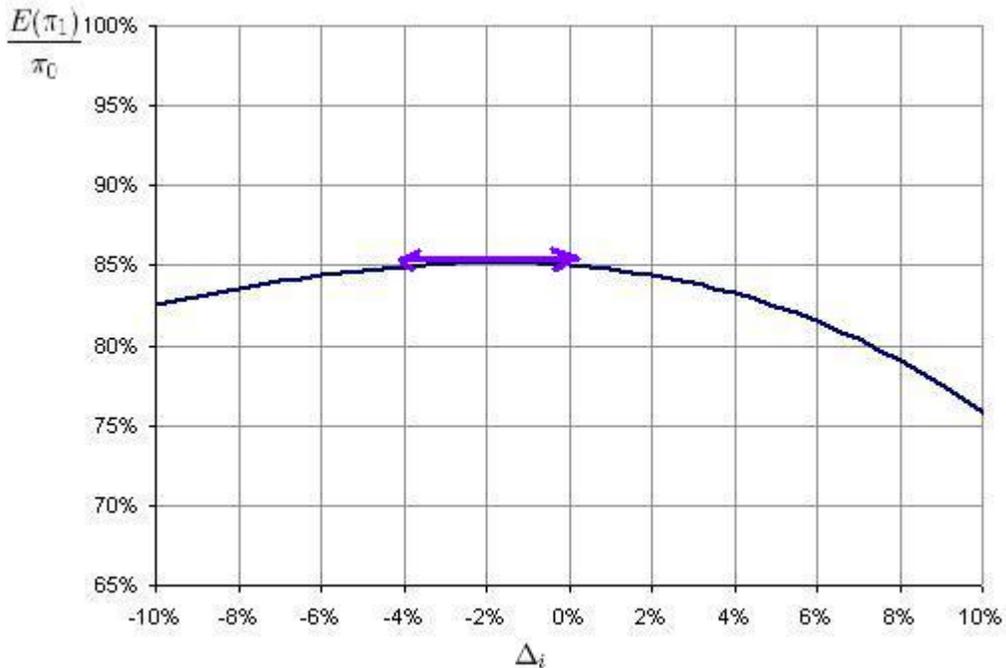


Figure n°74  
Variation de prime probable sur l'exercice suivant

Cet exemple illustre que :

- l'espérance de chiffre d'affaires à optimiser est une fonction concave de la variation de prix. Elle admet donc à priori un maximum, pour une variation de prix optimale.
- Il ne sert à rien de trop baisser les prix du portefeuille : on perd inutilement du chiffre d'affaires.
- Il est dangereux d'appliquer de fortes hausses des prix : l'élasticité étant assez forte en assurance automobile, la destruction de chiffre d'affaires s'accélère.

## 4.2 Différentes stratégies de détermination du prix à l'échéance

Dans cette partie, nous allons prendre plusieurs scénarii pour tenter de trouver la variation de tarif optimale à appliquer à l'échéance. Face à son portefeuille d'assurance automobile, l'entreprise peut avoir plusieurs ambitions :

- Maximiser le chiffre d'affaires probable, en appliquant une variation unique à l'échéance (cas simple de la prime majorable), sans contrainte de marge.
- Maximiser la marge probable. L'entreprise souhaite conserver les "meilleurs risques", les contrats qui procurent le maximum de marge, au sens défini au §4.1.1.

### 4.2.1 Maximum du chiffre d'affaires probable

#### • Objectif

Reprenons l'espérance du chiffre d'affaires probable pour la période qui suit l'échéance [3], sachant que nous appliquerons **une variation identique à tous les contrats**  $\Delta$ , avant application éventuelle de l'évolution du Coefficient de Réduction-Majoration (CRM) (obligatoire) : Cette évolution du CRM est notée par le rapport entre le CRM après l'échéance ( $CRM_1$ ) et le CRM avant l'échéance ( $CRM_0$ ). Cette évolution est réglementaire selon la sinistralité responsable constatée sur le contrat au cours des 12 mois précédant de 2 mois l'échéance.  $CRM_1$  n'est donc pas une variable aléatoire, car elle peut être connue au moment de la détermination de la prime suivante.

$$E(CA_1) = \sum_{i=1}^n \pi_i(0) \cdot \frac{CRM_1}{CRM_0} \cdot (1 + \delta) \cdot \nu_i(\delta)$$

La recherche d'un optimum par démonstration analytique n'est pas triviale et s'avère lourde en expression. Nous passerons donc par un algorithme d'optimisation sans contrainte, proposé par le professeur J.P. DUSSAULT. Il s'agit de découper notre fonction de chiffre d'affaires probables par segments de plus en plus fin, convergeant vers le maximum de cette fonction concave.

Cet algorithme est optimisé en découpant les segments de convergence par  $\frac{\sqrt{5}-1}{2}\sqrt{2} \simeq 0,618$ , c'est à dire le nombre d'or - 1 (ou l'inverse du nombre d'or), permettant d'accélérer les itérations.

Nous forcerons l'algorithme à trouver une solution pour un  $\Delta \in [-10\%; +10\%]$ , ce qui paraît une amplitude de variation assez vaste.

**Etape 0 d'initialisation :**  $c$  et  $d$  sont placés à 0,618 et  $(1-0,618)$  des extrémités de notre espace de départ.  $E(c)$  étant supérieur à  $E(d)$ , on conserve  $E(c)$ .

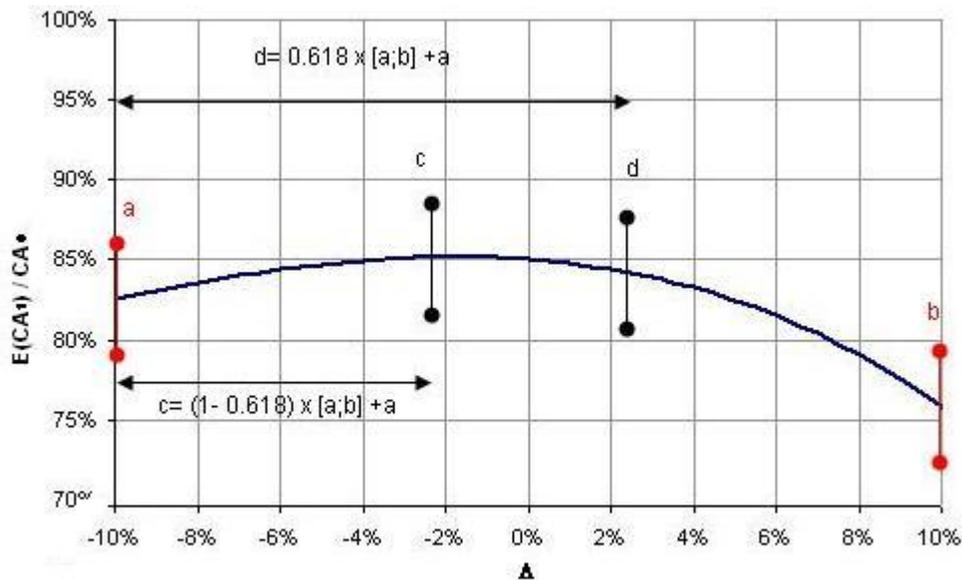


Figure n°75

Etape 0 : variation de la prime probable sur l'exercice suivant

**Etape 1 :**  $d$  prend alors la place de  $c$ ,  $b$  la place de  $d$  et  $c$  est positionné à  $(1-0,618)$  du nouvel intervalle de recherche.  $E(d)$  étant supérieur à  $E(c)$ , on conserve  $E(d)$ .

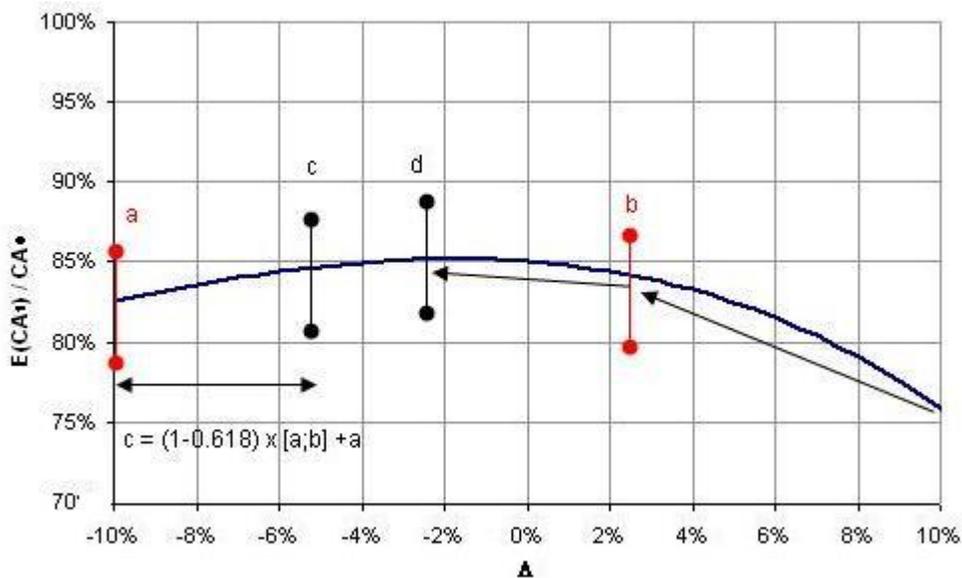


Figure n°76

Etape 1 : variation de la prime probable sur l'exercice suivant

**Etape 2 :**  $c$  prend alors la place de  $d$ ,  $a$  la place de  $c$  et  $d$  est positionné à 0,618 du nouvel intervalle de recherche.  $E(c)$  étant supérieur à  $E(d)$ , on conserve  $E(c)$ .

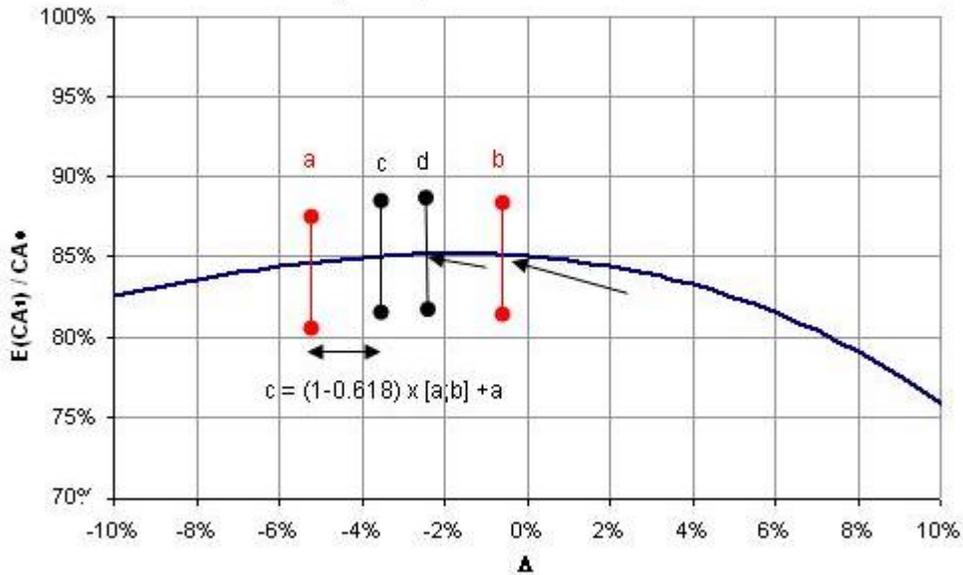


Figure n°77

Etape 2 : variation de la prime probable sur l'exercice suivant

**Etape 3 :**  $d$  prend alors la place de  $c$ ,  $b$  la place de  $d$  et  $c$  est positionné à (1-0,618) du nouvel intervalle de recherche :

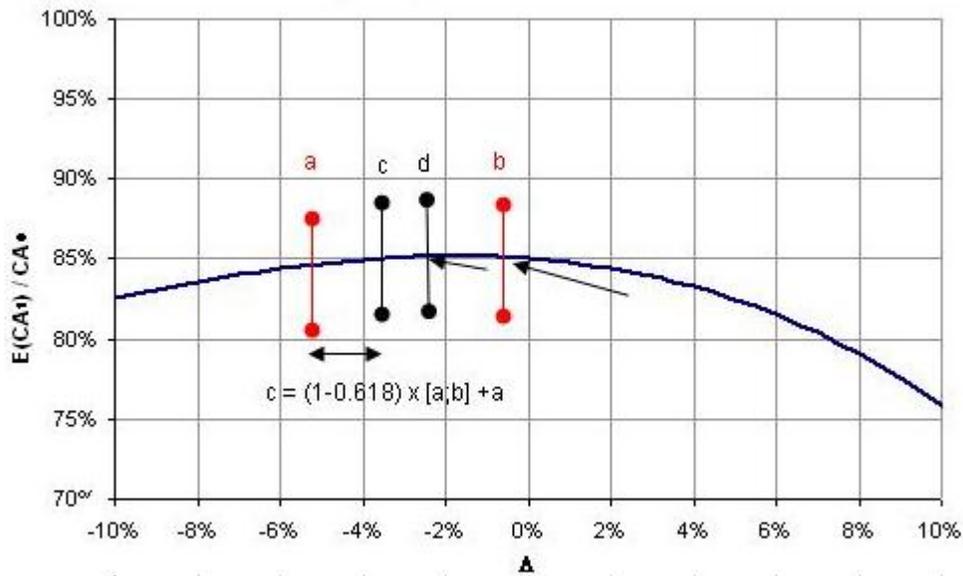


Figure n°78

Etape 3 : variation de la prime probable sur l'exercice suivant

L'encadrement  $[c; d]$  a dépassé la solution optimale. En fait, la solution se trouve systématiquement dans  $[a; b]$ . La condition d'arrêt s'effectue donc sur la longueur de l'intervalle  $[a; b]$   $E(d)$  étant supérieur à  $E(c)$ , on conserve  $E(d)$ .

**Etape 4 :**  $c$  prend alors la place de  $d$ ,  $a$  la place de  $c$  et  $d$  est positionné à 0,618 du nouvel intervalle de recherche :

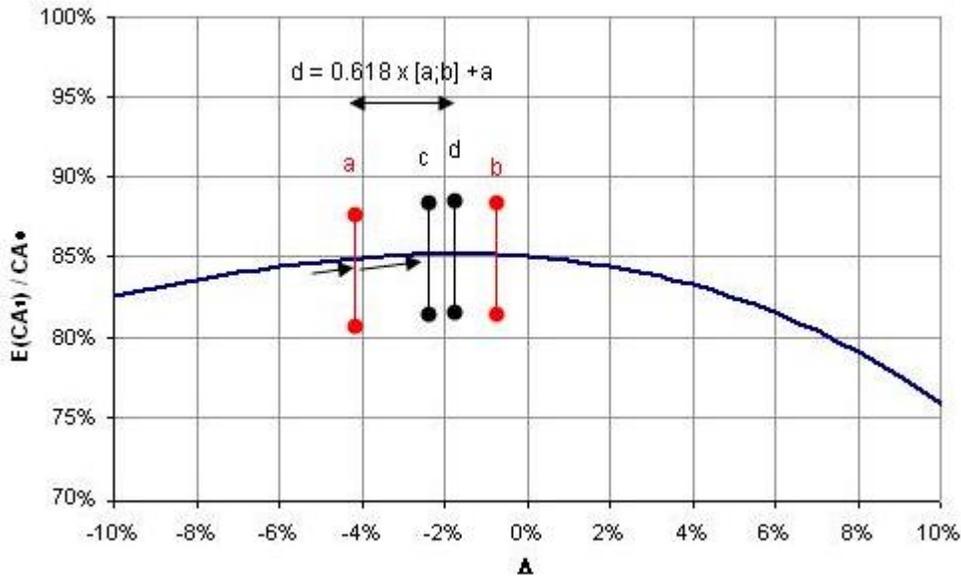


Figure n°79

Etape 4 : variation de la prime probable sur l'exercice suivant

Cette fois  $[c; d]$  encadre la solution optimale à moins de 0.5%. On voit donc que dans 2 itérations,  $[a; b]$  l'aura remplacé et aura convergé.

#### • Application

Nous avons repris la fonction logistique obtenue au chapitre 3 comme estimateur de la probabilité pour chaque contrat de se maintenir en cours à l'échéance, suite à la variation de tarif appliquée. Cette fonction dépend des paramètres du contrat évalués au moment de l'échéance :

- âge du conducteur principal,
- antécédent de sinistralité,
- formule de garanties souscrites,
- incident de paiement dans les 36 derniers mois,
- coefficient de réduction/majoration,
- ancienneté du contrat,
- forfait kilométrique,
- ancienneté du véhicule.

Cette fonction dépend également de la variation de tarif à appliquer au terme,  $\Delta$ . Dans cette application, l'algorithme fait varier ce paramètre unique pour tout le portefeuille, afin d'obtenir notre optimum de chiffre d'affaires probable. Tous les autres paramètres restent soit constants à l'échéance (formule, forfait kilométrique) soit vieillissent d'1 an (âge, CRM, ancienneté de contrat et de véhicule).

Pour obtenir une simulation réaliste, nous faisons quelques adaptations à la fonction  $\nu$  obtenue au chapitre 3 :

- la constante est modifiée afin d'obtenir le taux de survie observé à l'échéance :

Variable	N	Mean	Std Dev	Min	Max
Taux de survie observé	1 522 200	88%	32%	0	1
$\nu$ modifiée	1 522 200	88%	8.5%	43%	93%

- la "réponse" de survie ou de résiliation pour chaque contrat est adaptée pour tenir compte du résultat observé à l'issue du terme :
  - Si  $\Delta <$  écart réel observé ET contrat observé est en survie ALORS  $\nu_i(\Delta) = 100\%$ .  
En effet, si nous observons que le contrat n'a pas été résilié pour un écart, et que nous simulons avec un écart inférieur (plus favorable), alors il est certain que le contrat va rester en cours.
  - Si  $\Delta >$  écart réel observé ET contrat observé est en résilié ALORS  $\nu_i(\Delta) = 0\%$ .  
En effet, si nous observons que le contrat a été résilié pour un écart, et que nous simulons avec un écart supérieur (moins favorable), alors il est certain que le contrat sera résilié.

Avec ces corrections, nous obtenons notre optimum en 13 itérations de l'algorithme, pour une base de 1 522 200 contrats terminés. Les résultats sont résumés dans le tableau ci-dessous :

$\Delta_{optimal} = -2.02\%$  avant application du coefficient de réduction-majoration

	Avant terme	Après 1 an	Variation
Chiffre d'affaires observé	614 533 280EUR	531 190 146EUR	-13.56%
Chiffre d'affaires simulé (optimum recherché)	614 533 280EUR	565 306 895EUR	-8.01%
Nombre de contrats observés	1 522 200	1 345 145	-11.63%
Nombre de contrats simulés	1 522 200	1 391 570	-8.58%
Marge observée	28 839 046EUR	27 806 814EUR	-3.58%
Marge simulée finale	28 839 046EUR	28 841 176EUR	+0.01%

Ce résultat de simulation invite à plusieurs commentaires :

- Notre fonction logistique a une pente forte sur la variation de tarif inférieure à 0. Cela impacte fortement le résultat de l'optimisation, trouvant l'optimum à -2%.
- Notre base de simulation est prise sur une période de cycle baissier de l'assurance automobile, avec une forte pression sur les prix par une concurrence accrue. Ce contexte peut influencer aussi le résultat à la modération tarifaire, voire à la baisse.

En voulant optimiser notre CA de fin de période (un gain de 6.4% soit 34MEUR), nous avons réussi à optimiser aussi le taux de résiliation de 3%, et la marge probable de 3.7%.

Le tableau de simulation pas à pas sur les 12 itérations est en annexe.

## 4.2.2 Maximum de la marge probable

### • Objectif

La cotisation de chaque contrat de notre portefeuille est soit au barème (cas classique : les contrats de 1ère année), soit en écart par rapport au barème en fonction des encadrements des périodes précédentes. En effet, MMA applique sur son portefeuille auto la retarification encadrée.

Posons  $b_i$  la prime théorique calculée au barème pour chaque contrat  $i$  de notre portefeuille, avec ces données mises à jour.

Il est possible de construire  $M_i$  la marge probabilisée induite par l'écart entre le barème et la prime que paiera le client à l'échéance :

$$\forall i \in [N], M_i = \left( \pi_i(0) \cdot \frac{CRM_1}{CRM_0} \cdot (1 + \Delta_i) - b_i \right) \cdot \nu_i(\Delta_i) \quad [6]$$

$M_i$  est donc une variable aléatoire dont l'espérance peut être estimée à l'aide de notre modèle logistique de probabilité de rétention.  $M_i$  peut être positive ou négative, suivant la position relative du barème.

Sous cet objectif à maximiser, nous accentuons l'impact de la probabilité de rétention sur la détermination du delta de tarif optimal. Nous utiliserons donc un résultat de BOULANGER et TRAN VAN LIEU concernant la maximisation de la marge par segment. Ils démontrent que la maximisation de la marge calculée au segment, et dans notre cas, au contrat, est supérieure à la maximisation de la marge globale. Cela implique la recherche du  $\Delta$  individuel à chaque contrat, maximisant la relation [5].

Là encore, la recherche d'un optimum par démonstration analytique est lourde. Nous reprenons donc notre algorithme d'optimisation sans contrainte précédent, cette fois appliqué à chaque

contrat. Nous forcerons l'algorithme à trouver une solution dans l'intervalle  $[-10\%; +10\%]$ , ce qui paraît une amplitude de variation assez vaste, avant application de la variation du coefficient de réduction-majoration.

### • Application

Reprenons notre base d'expérience d'un portefeuille de 1 522 200 contrats avant terme. L'optimum pour chaque contrat a nécessité 12 itérations maximum : nombre identique à l'optimisation globale du chiffre d'affaires.

La distribution des  $\Delta_i$  (variation de tarif optimale au terme pour chaque contrat, avant application de l'évolution CRM), se fait comme suit :

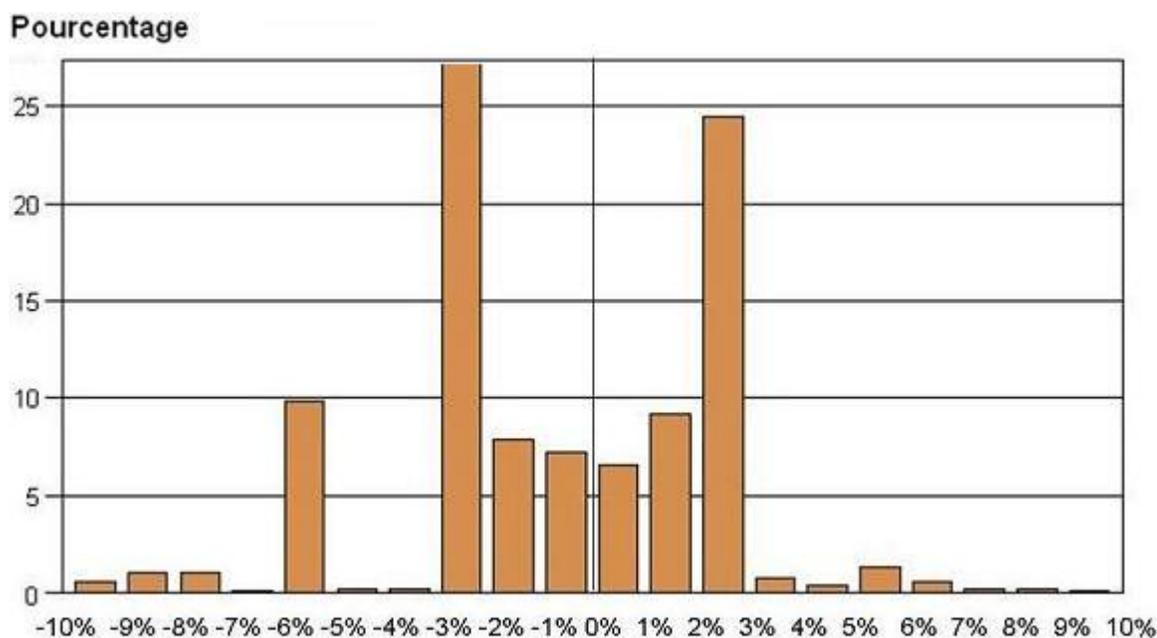


Figure n°80  
Distribution des variations individuelles optimales

$\Delta$  moyen (variation du CA au terme) = **-0.006%** avant application du CRM

$\Delta$  médian = **-0.005%** avant application du CRM

1er quartile = -2.3%, 3ème quartile = +2.4%, écart-type = 3.0%

La distribution de l'optimisation est assez centrée et concentrée entre +2.5% et -2.5%, avec une distribution en double modale : l'algorithme tranche assez nettement avec la pratique. Il provoque une évolution plutôt systématique et marquée, dans un sens comme dans l'autre, plutôt qu'une distribution classique en cloche, proche de la moyenne.

On remarquera le pic à -6% (avant CRM) : cible pour laquelle la marge est importante alors que notre modèle estime une élasticité au tarif tendue. Il est donc nécessaire d'abandonner une partie de cette marge sur ces contrats pour tenter d'en sauver plus.

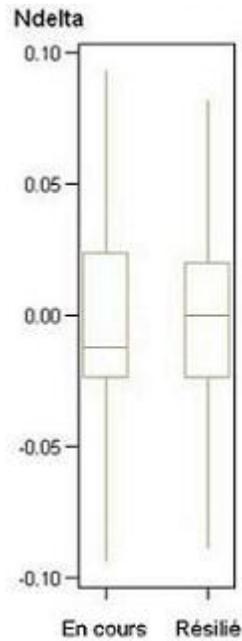


Figure n°81

*Distribution des écarts suivant l'état REEL du contrat 1 an plus tard*

L'écart entre les deux populations de contrats est faible. Paradoxalement, la moyenne des variations appliquées sur les résiliations observées est un peu plus grande que sur les contrats restés en cours. La recherche de marge pousse l'algorithme à être plus exigeant dans la variation de contrats ayant peu de marge, en recherche de tarif optimal et donc plus enclin à la résiliation.

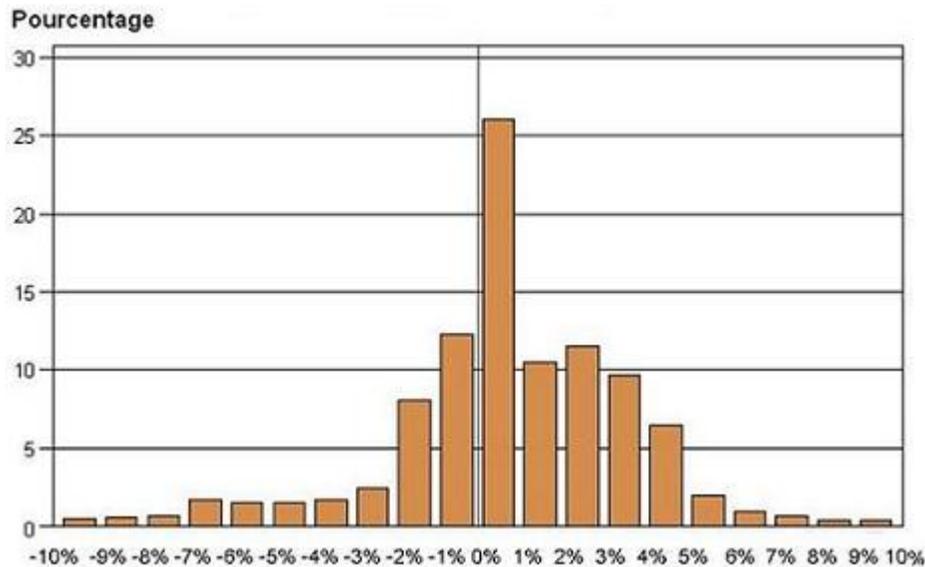


Figure n°82

*Distribution de la variation réelle observée (avant application de l'évolution de CRM)*

Pour comparaison, il paraît intéressant de comparer la distribution précédente des variations optimales à celle des variations réellement appliquées sur notre base d'expérience. On retrouve une distribution classique en cloche, relativement modeste (+0,5% de variation moyenne avant CRM, écart type 2,8%), mais peu de baisses marquées.

De ces deux variations, il paraît intéressant d'examiner l'écart d'application entre le théorique et le réalisé :

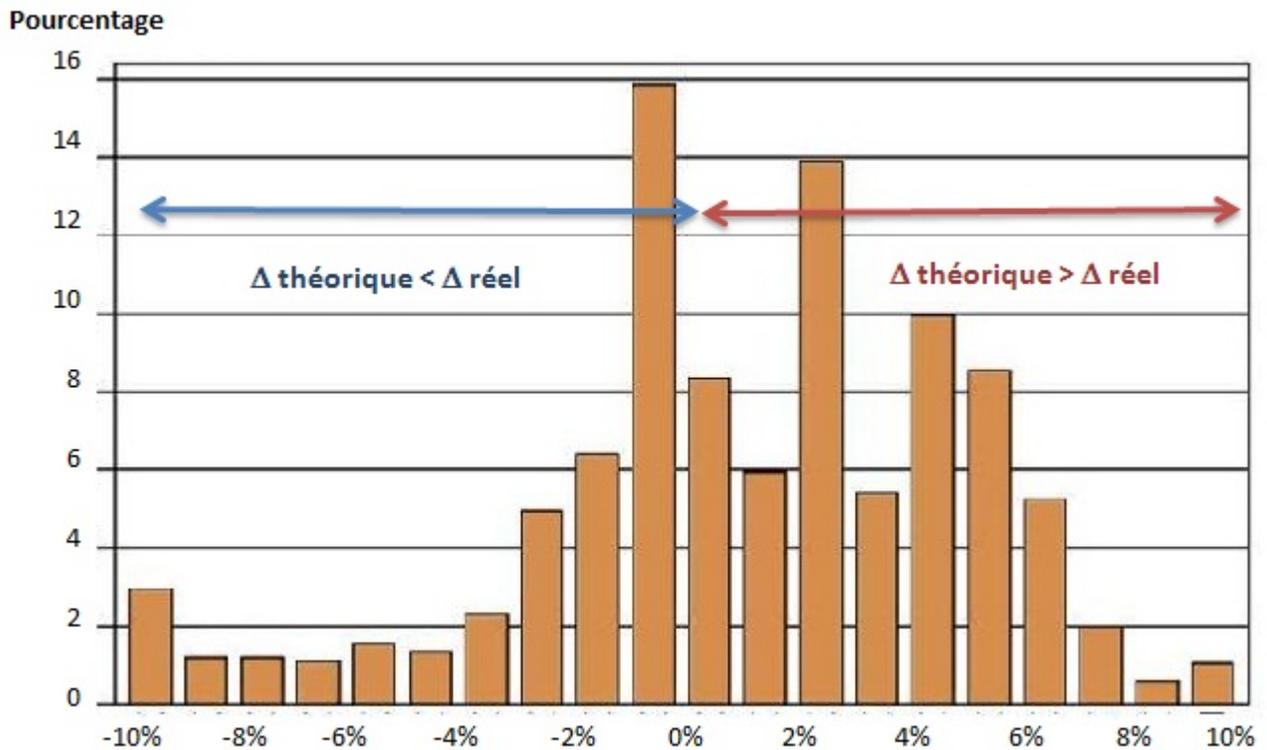


Figure n°83

*Distribution des écarts entre  $\Delta_i$  et la variation réelle appliquée au portefeuille (avant application de l'évolution de CRM)*

L'écart de variation moyen est de +1.1%, l'écart type à +4%.

L'algorithme propose des variations assez différentes de celle réalisées. La recherche de marge décale fortement la mesure que l'on appliquerait de façon plus générale.

Notre méthode d'optimisation de marge conduit donc à une variation du prix à l'échéance beaucoup plus marquée, à la hausse comme à la baisse. Ces variations optimales individuelles produisent une progression, certes minime, de la marge probable :

	Avant terme	Après 1 an	Variation
Chiffre d'affaires (CA) observé	614 533 280EUR	531 190 146EUR	-13.56%
CA simulé ( <b>optimum recherché</b> ) (1)	614 533 280EUR	<b>565 306 895EUR</b>	<b>-8.01%</b>
CA simulé méthode individuelle (2)	614 533 280EUR	536 306 423EUR	-12.77%
Nombre de contrats observés	1 522 200	1 345 145	-11.63%
Nombre de contrats simulés (1)	1 522 200	1 391 570	-8.58%
Nombre de contrats simulés (2)	1 522 200	1 357 613	-10.81%
Marge observée (M)	28 839 046EUR	27 806 814EUR	-3.58%
M simulée finale (1)	28 839 046EUR	28 841 176EUR	+0.01%
M simulée finale ( <b>optimum recherché</b> )(2)	28 839 046EUR	<b>29 083 864EUR</b>	<b>+0.85%</b>

On notera aussi que cette stratégie d'optimisation de la marge probable n'a pas dégradé nos autres indicateurs :

- le taux de résiliation s'améliore de 0.8% par rapport à l'observé,
- le chiffre d'affaires probable limite sa décroissance aussi de 0.8% par rapport à l'observé.

Si l'on compare à la méthode précédente (optimisation du chiffre d'affaires probable), on obtient une baisse du chiffre d'affaires de -5%, pour une hausse de la marge probable de 0.8%. Le taux de marge du portefeuille s'améliore faiblement entre les 2 méthodes (il passe de 5.1% à 5.4%), avec une baisse du chiffre d'affaires conséquente. En terme de décision économique, l'optimisation de la marge ne semble pas suffisamment satisfaisante pour compenser une perte du chiffre d'affaires probable.

### 4.2.3 Approfondissement

L'optimisation du chiffre d'affaires probable pour un  $\Delta$  unique (cf § 4.2.1) dispose d'un avantage notable : la méthode est facilement opérationnelle et utilisable pour terminer un portefeuille et paraît satisfaisante économiquement.

La seconde méthode unitaire, cherchant à optimiser la marge individuelle probable au contrat (cf § 4.2.2) procure le maximum de marge à court terme. En effet, le champ d'étude de ce mémoire est de mesurer l'impact sur T+1 d'une décision tarifaire à une échéance T. Il nous paraît plus prudent néanmoins de penser que cette décision peut avoir des conséquences à plus long terme. D'autre part, le décalage que l'on crée entre le barème et prime payée (ce que nous avons baptisé marge) ne se rattrape pas en un seul exercice.

En conséquence, une voie d'approfondissement serait de mesurer sur le long terme l'impact de cette décision tarifaire unitaire. BOULANGER et TRAN VAN LIEU proposent cette voie

---

d'analyse au niveau macroscopique, par la valeur intrinsèque du portefeuille (embedded value). On peut transposer le raisonnement au niveau contrat, en calculant la valeur individuelle du contrat (CMV). Mais cela suppose de connaître les élasticités au prix sur toutes les périodes actualisées... un champ d'investigation s'avère extrêmement vaste mais qui peut révéler une stratégie gagnante sur le long terme pour la seconde méthode.

La seconde méthode présente un inconvénient : l'antithèse de la première méthode. Il s'agit de mettre à chaque contrat une variation individualisée, difficile de mettre en oeuvre cette solution, sauf à la générer mois par mois par un service d'actuariat et la transposer dans l'informatique de production, ce qui peut paraître coûteux. Dans le prochain chapitre, le plus appliqué, nous proposerons une solution plus aisée d'implémentation de variation de tarif au terme, par des règles d'encadrement par segment, à partir des résultats de cette seconde méthode.

# Chapitre 5

## Mise en application simplifiée

### 5.1 Introduction

Au chapitre précédent nous avons mis en lumière une méthode classique de tarification à l'échéance : la retarification encadrée. L'assureur construit un encadrement "raisonnable" de variation, construit à partir de la prime précédente.

Cette méthode peut être déclinée sur un certain nombre de segments identifiables dans le portefeuille. Il est ainsi possible de fournir un algorithme facilement implémentable dans le système d'informations d'une compagnie d'assurances, lors de l'établissement des cotisations au terme (avis d'échéance).

Les données fournies par l'actuariat sont décrites dans une table des segments, associant pour chaque cas leurs variations minimale (plancher) et maximale (plafond). On utilise cette description dans le processus suivant, à mettre en oeuvre et à contrôler au terme, pour chaque contrat dont la cotisation à l'échéance est à émettre :

1. identification du segment du contrat dans la liste des segments,
2. associer des variations minimale et maximale de son segment d'appartenance, pour chaque contrat,
3. calcul de la prime barème du contrat : évolution les critères et retarification totale,
4. calcul de l'encadrement de variation par rapport à la prime payée l'année précédente,
  - en multipliant la prime précédente par la variation minimale = plancher,
  - en multipliant la prime précédente par la variation maximale = plafond.
5. détermination de la prime à payer pour l'échéance, par la relation :

$$\text{cotisation} = \text{MEDIANE} (\text{plancher}, \text{barème}, \text{plafond}),$$

6. calcul du "coefficient de marge"  $M$  correspondant à l'écart relatif entre la prime à payer et la prime barème :

$$M = \frac{PaP}{PB} \text{ (avec } PaP = \text{ prime à payer et } PB = \text{ prime barème)}$$

cf. §4.1.1. ce coefficient de marge peut être supérieur ou inférieur à 1, suivant l'historique d'évolution du contrat et de la politique tarifaire passée.

Dans cette partie, nous allons préciser les variations minimale et maximale par segment, afin d'obtenir un processus optimisé. Dans la partie précédente, nous avons identifié ces variations optimales **au contrat**. Ce chapitre utilise une méthodologie d'arbre de regroupement afin de donner les variations optimales au segment. Nous perdons ainsi en précision, mais nous gagnons en opérationnalité sur le processus décrit.

## 5.2 Objectif méthodologique

Notre objectif est de réaliser un regroupement des  $\Delta$  les plus proches, par une partition de notre portefeuille en groupes homogènes et distincts.

On utilisera à cet effet une technique d'arbre de décision, encore appelée classification descendante supervisée. Cette méthode consiste à segmenter un ensemble d'individus (le portefeuille) selon les variables les plus discriminantes (exemples : niveau de CRM, âge, sinistralité antérieure) en fonction d'un objectif : les  $\Delta_i$  les plus proches.

Dans chaque "feuille" ainsi obtenue, on pourra alors associer :

- comme plancher, une valeur minimale prise par les  $\Delta_i$  de la feuille,
- comme plafond, une valeur maximale prise par les  $\Delta_i$  de la feuille.

Cette structure se prête ainsi à un algorithme de type retarification encadrée, tout en garantissant d'appliquer sur des segments relativement homogènes.

La segmentation au mieux de chaque feuille peut être réalisée à partir de différents indices :

- la métrique du  $\chi^2$ ,
- l'indice de Gini,
- l'entropie de Shannon...

Notre préférence ira à la méthode CHAID<sup>1</sup> ( $\chi^2$  Automatic Interaction Detection), proposée par KASS (1980), disponible sous SPAD®. Cette méthode présente l'avantage de produire une segmentation N-aire la plus fine, en regroupant des feuilles les plus fines.

---

1. Algorithme proposé par JP. NAKACHE et J. CONFAIS

### • Algorithme de la méthode CHAID

Soit  $\Delta$  la variable à expliquer. Il s'agit d'une variable continue. L'ensemble à segmenter est donc trié par  $\Delta$  croissant. On cherche alors la meilleure variable de segmentation à  $m$  modalités  $X_m$  :

1. dans l'ensemble de départ, pour chaque point de coupure de  $\Delta$ , on crée la pseudo-variable qualitative  $\Delta'$  à 2 modalités. On lui associe pour chaque variable de segmentation candidate  $X_m$  le tableau de contingence (2,m),
2. Regroupement des modalités de  $X_m$  : On détermine la paire de modalité de  $X_m$  dont le sous-tableau (2,2) fournit le plus petit  $\chi^2$ . S'il est non significatif, on regroupe les deux modalités en une seule. On répète ce pas jusqu'à ce que toutes les modalités soient regroupées, donc exclusion de  $X_m$ , ou que les modalités restantes soient toutes significatives au sens du  $\chi^2$ .
3. On calcule la significativité de chaque variable de segmentation  $X_m$ , pour chaque coupure de  $\Delta$ . On retient le  $\chi^2$  le plus grand. On divise alors l'ensemble des individus de départ en autant de feuilles que  $X_m$  a de modalités regroupées.
4. Pour chaque feuille divisible (c'est-à-dire que son cardinal permet un découpage autorisé par l'utilisateur), on reprend au point 1.

Avec notre portefeuille de contrats d'assurance automobile, nous disposons pour chacun la variation optimale établie par notre méthode définie au paragraphe 4.2. Nous allons réaliser ensemble les premiers pas de l'algorithme CHAID dans un but d'illustration, avec le logiciel SPAD®. A chaque pas, les modalités des variables qualitatives sont regroupées par les étapes 1 et 2. Les variables continues sont découpées au point où l'inertie INTER entre les deux groupes obtenus soit la plus forte. L'étape 3 consiste alors à calculer les statistiques du  $\chi^2$  pour chacune des pseudo-qualitatives obtenues. Au premier pas, SPAD propose le tableau de  $\chi^2$  suivant :

Variable	Impact	
MargeAV	0.2534	
CCATANT	0.0523	
NB50anc	0.0216	
ANCCONT	0.0190	
FORMULE	0.0147	
NBM	0.0140	
RapCRM	0.0064	
PERM	0.0060	
MED	0.0052	
ANVEH	0.0051	
CL	0.0044	
AGE	0.0041	
FSOUS	0.0024	
ZRC	0.0013	
prime	0.0008	
ENE	0.0008	
SEXE	0.0003	

Figure n°84  
Premier pas

C'est la variable continue MargAV (indicateur de marge avant l'échéance) qui obtient le meilleur indice de coupure, l'inertie est maximale pour la valeur 0. C'est-à-dire qu'on obtient la meilleure discrimination du  $\Delta$  à appliquer au terme en fonction de la marge précédente, si elle était positive ou négative :

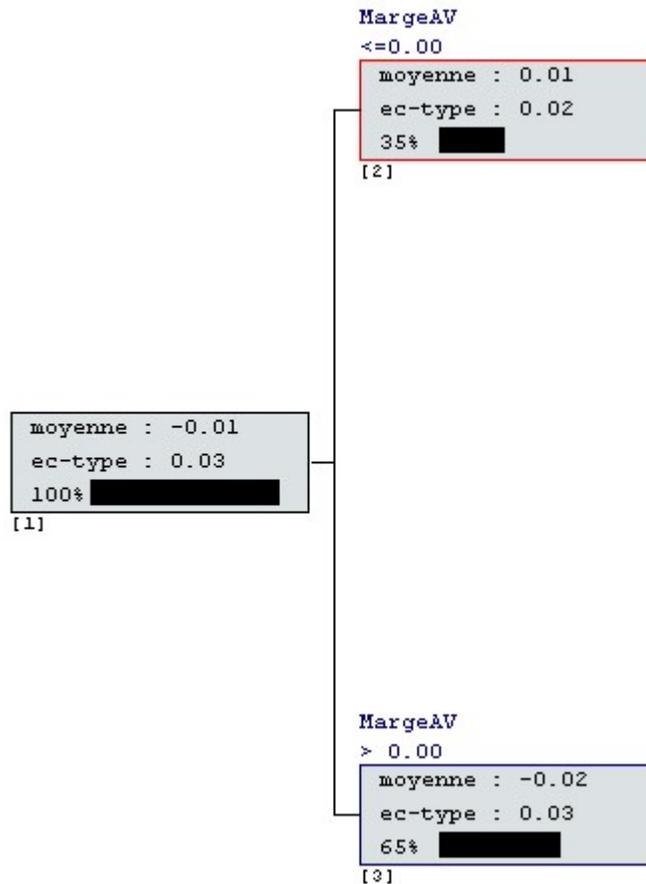


Figure n°85  
Interprétation du premier pas

Le groupe des marges précédentes négatives se voit appliquer une variation au terme positive, en moyenne de +1% (logique). Le groupe des marges précédentes positives se voit appliquer une variation au terme négative afin de modérer l'effet résiliation, en moyenne de -2% (résultat logique toujours). SPAD fournit alors une statistique de chaque noeud, où l'on peut contrôler les statistiques internes au noeud obtenu. Ici le noeud des marges précédentes négatives (34.9% de notre portefeuille), où l'on voit que la variation optimale s'étend de -9.7% à +5.3%, en moyenne +1.34%.

Var. à expliquer	Ndelta
Min.	-0.0967
Max.	0.0531
Moyenne	0.0134
Ecart-type	0.0185

Figure n°86  
Interprétation du premier noeud

Sur ce noeud, l'algorithme reprend les étapes 1 à 3 en construisant le tableau d'indice de coupure suivant :

Variable	Impact
FORMULE	0.0042
ANVEH	0.0030
prime	0.0022
CCATANT	0.0021
ANCCONT	0.0018
MargeAV	0.0015
RapCRM	0.0011
NB50anc	0.0009
FSOUS	0.0009
CL	0.0009
NBM	0.0007
PERM	0.0002
AGE	0.0002
MED	0.0002
ENE	0.0001
SEXE	0.0000
ZRC	0.0000

Figure n°87  
Indices de coupure

On constate que la meilleure variable candidate pour segmenter le groupe des marges précédentes négatives est alors la FORMULE de garanties du contrat :

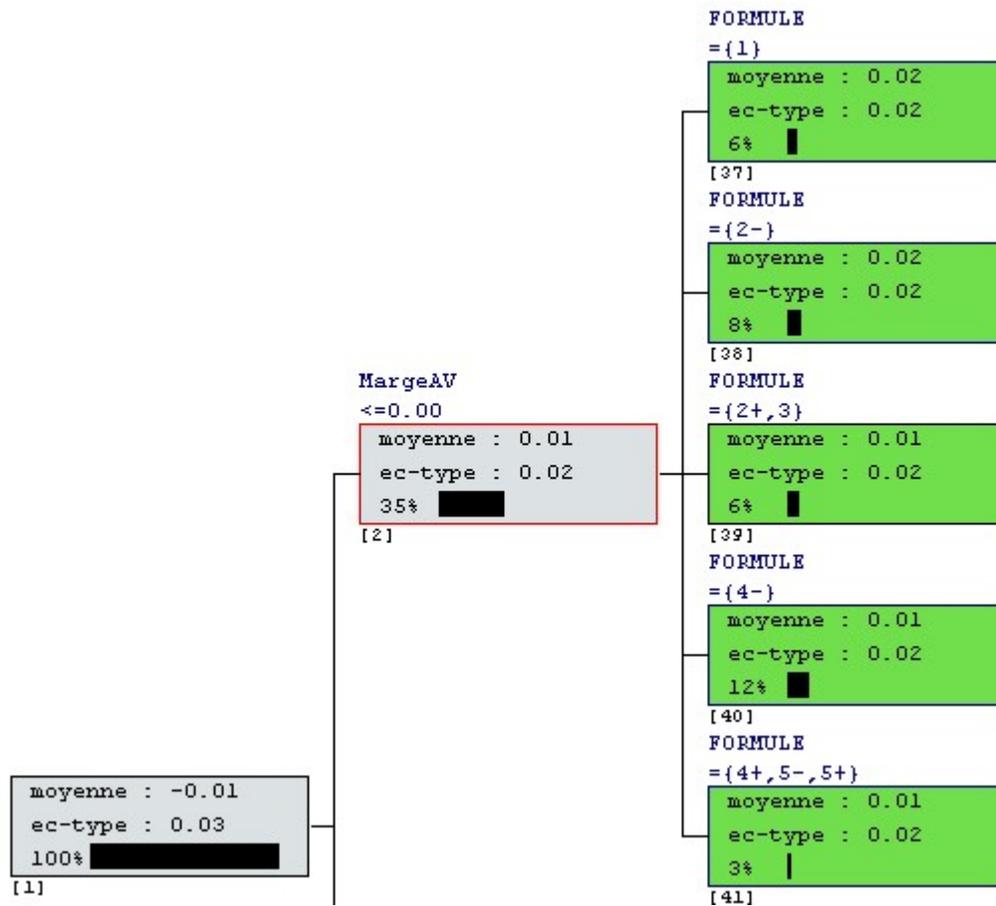


Figure n°88  
2ème étape

Les formules minimum (1 et 2-) se voient affecter la variation la plus forte, en moyenne +2%. Ce sont les primes les plus basses, ce qui rend compatible une hausse en pourcentage importante car moins impactante en euro pour le client. Les formules médianes et tous risques (2+ à 5+) ont la variation la moins forte, de l'ordre de +1%, pour les raisons symétriques.

SPAD propose alors un comparatif des tailles relatives des sous-noeuds obtenus. Ici on compare le poids des populations de chaque formule pour les marges précédentes négatives, avec le poids de chaque formule au niveau global portefeuille :

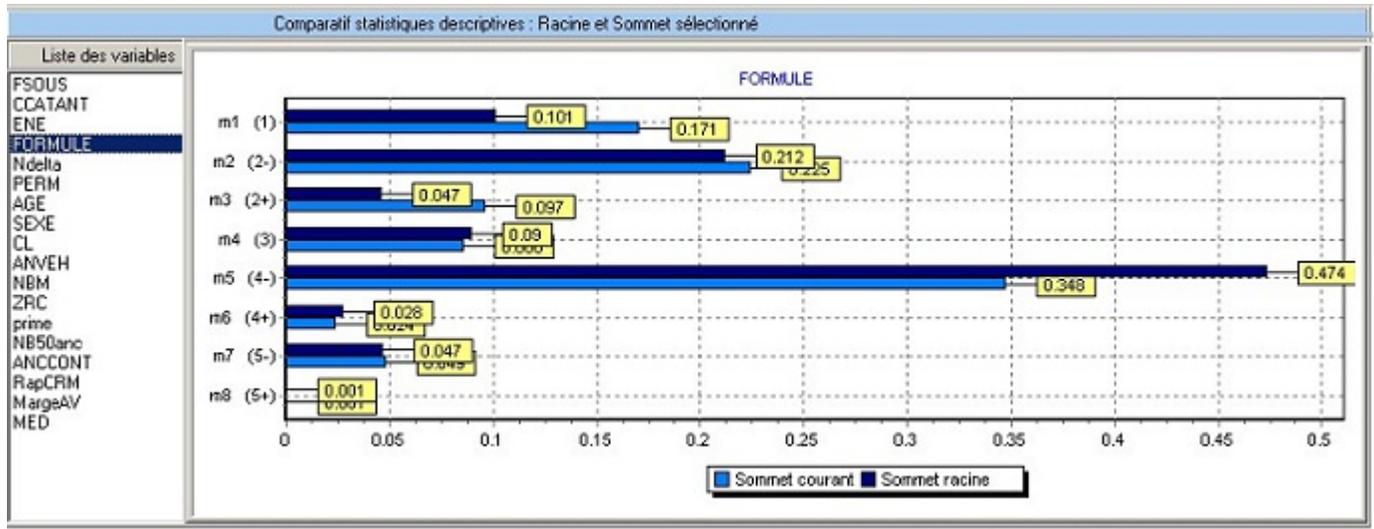


Figure n°89  
Analyse du 2ème noeud

On constate une sur-représentation des formules tiers 1 et médiane 2+ dans les marges précédentes négatives, contre une sous-représentation de la formule tous risques 4-. Le besoin de reconstitution de marge n'est donc pas uniforme pour chaque formule.

L'algorithme reproduit ce comportement itérativement sur chaque noeud obtenu, selon les conditions d'arrêt paramétrées : soit limite d'effectif à diviser, soit profondeur de l'arbre à atteindre.

### 5.3 Application

Avec notre portefeuille de contrats d'assurance automobile, nous lançons l'algorithme CHAID décrit, en paramétrant une limite de profondeur à obtenir : Rappelons que notre objectif est une mise en oeuvre simplifiée des mesures, inutile de complexifier à l'extrême par une succession de noeuds à définir dans l'informatique du terme.

### 5.3.1 Arbre de décision de variation

Nous obtenons l'arbre de décision suivant avec le logiciel SPAD® :

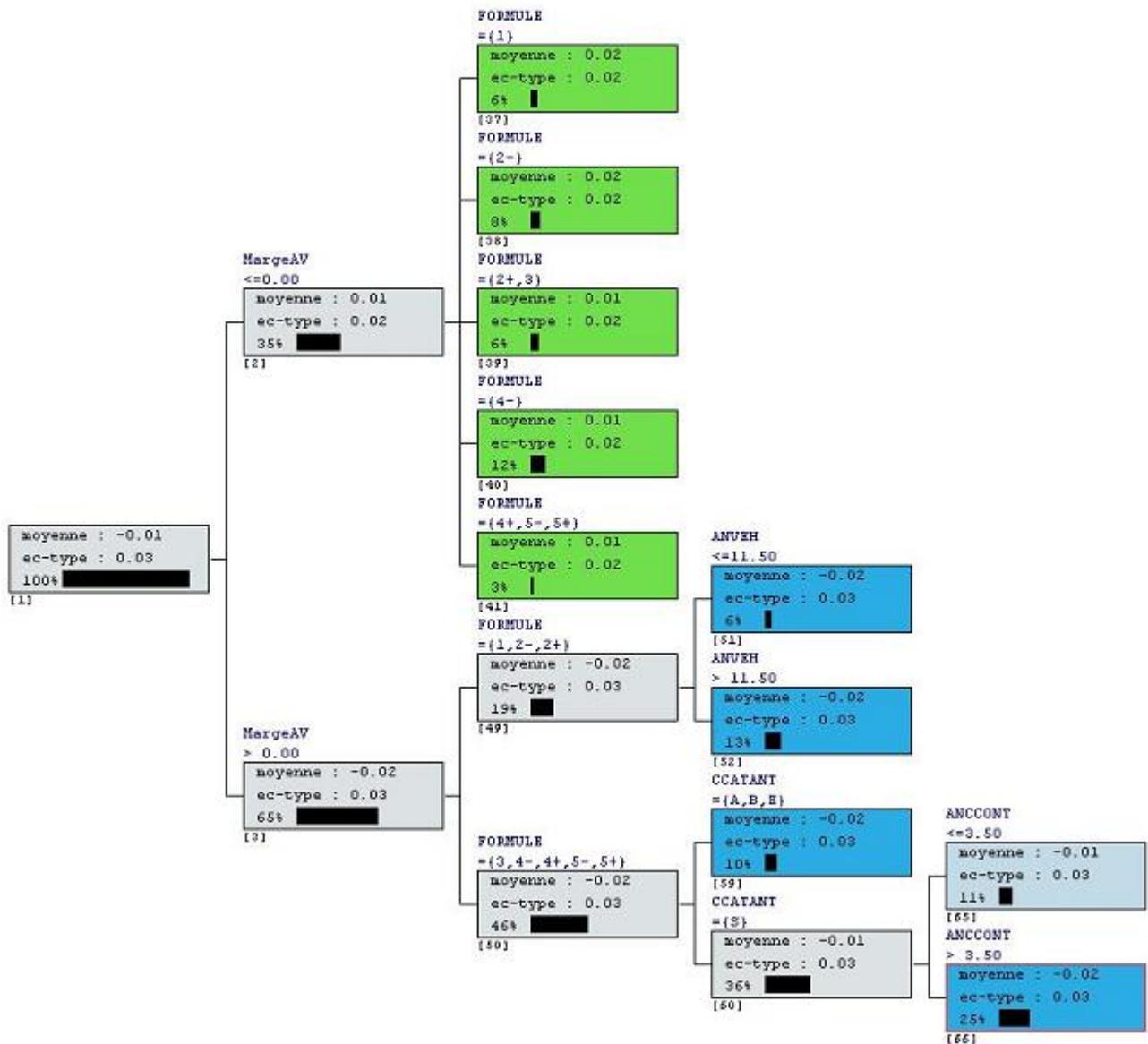


Figure n°90  
Arbre de décision

Quelques commentaires sur l'arbre de décision obtenu :

- le premier critère, et de loin, qui identifie une coupure en deux groupes est la marge. Ce qui peut paraître tout à fait normal étant l'objectif de notre algorithme. On retrouve ainsi classiquement que les contrats à marge positive peuvent être baissés, et les contrats à marge négative doivent être augmentés,

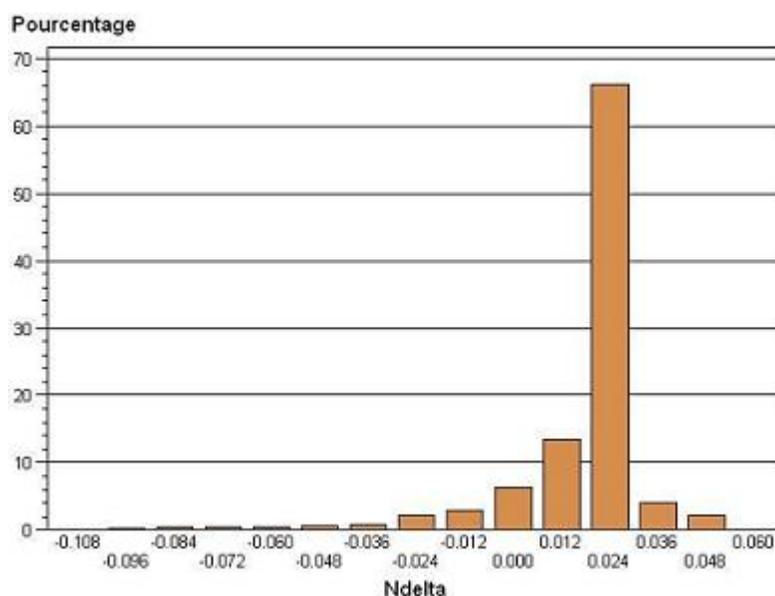
- pour les marges négatives, l’algorithme dégage simplement les formules de garanties comme critère discriminant : plus la couverture est grande, moins l’augmentation relative est forte.
- pour les marges positives, on trouve une logique inverse : il convient de baisser les petites couvertures (au tiers et médiane) et plutôt majorer les formules Tous Risques. Il faut là plutôt aller sur notre fonction de résiliation qui laisse apparaître une plus forte tension sur les contrats à faible couverture, plus sensibles au prix,
- pour les petites couvertures, c’est plutôt l’ancienneté du véhicule qui guide la variation : plus le véhicule est ancien, plus la baisse doit être forte. On retrouve un élément de prime,
- pour les formules Tous Risques, c’est plutôt la sinistralité antérieure qui guide l’arbre, puis l’ancienneté du contrat.

### 5.3.2 Caractérisation des feuilles

#### *Segment 1 : Marge négative en Formule au tiers*

Ce segment est caractérisé par de petites primes, et une sinistralité passée qui a produit une marge négative. Une variation à la hausse s’impose.

Poids du segment	6%
Variation moyenne	+1.75%
Variation médiane	+2.39%
Ecart-type de variation	1.77%



*Figure n°91*  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

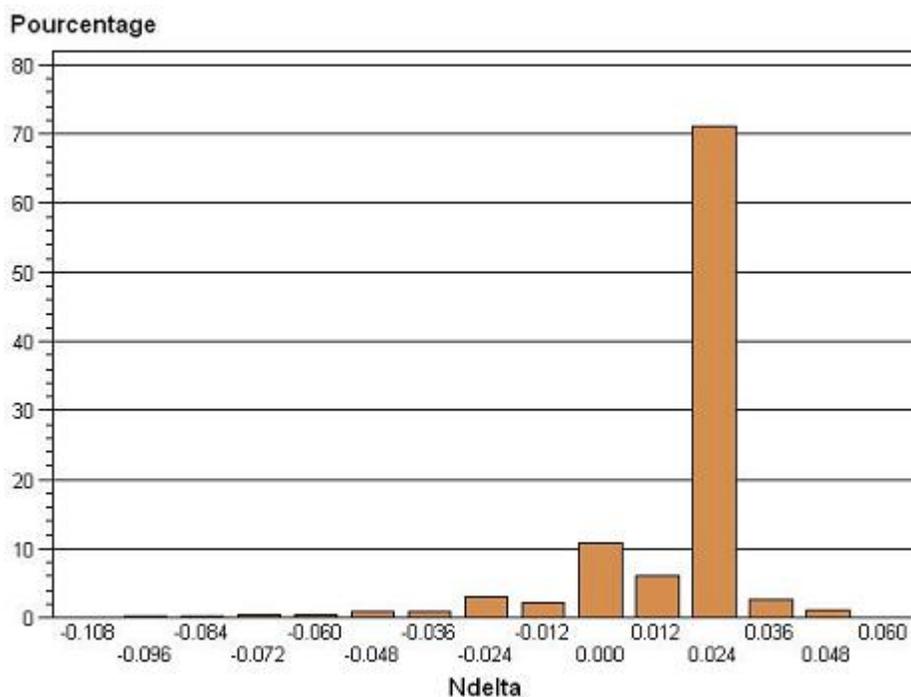
L'essentiel des variations se concentre sur +2.4%. Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : +1.5%
- Plafond de variation : +5%

*Segment 2 : Marge négative en Formule médiane sans option*

Il s'agit d'un segment caractérisé également par de petites primes et une sinistralité passée. Là encore une variation à la hausse s'impose, légèrement inférieure à la précédente.

Poids du segment	8%
Variation moyenne	+1.66%
Variation médiane	+2.39%
Ecart-type de variation	1.75%



*Figure n°92*  
*distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2*

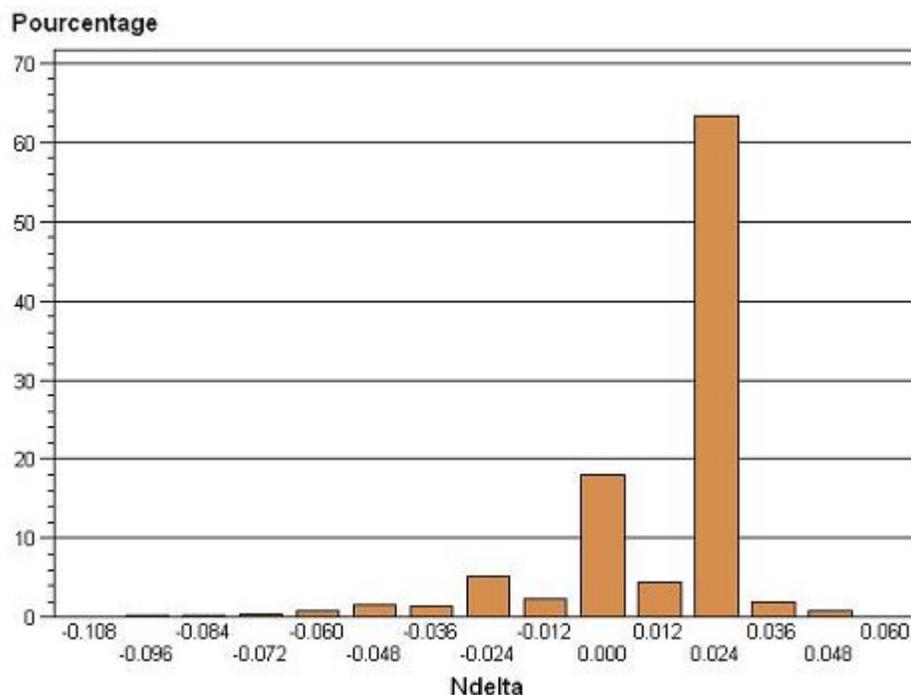
Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : +1%
- Plafond de variation : +4%

*Segment 3 : Marge négative en Formule médiane avec options ou tous risques sans option*

Ce segment contient des primes plus élevées que les précédentes, une variation à la hausse plus modérée est déterminée.

Poids du segment	7%
Variation moyenne	+1.32%
Variation médiane	+2.39%
Ecart-type de variation	1.96%



*Figure n°93*  
*distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2*

Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : +0%
- Plafond de variation : +3%

*Segment 4 : Marge négative en Formule tous risques avec options*

Ce segment contient des primes plus élevées que les précédentes, la tension d'élasticité est plus forte malgré une marge négative : la variation à la hausse est encore plus modérée.

Poids du segment	12%
Variation moyenne	+1.10%
Variation médiane	+2.02%
Ecart-type de variation	1.74%

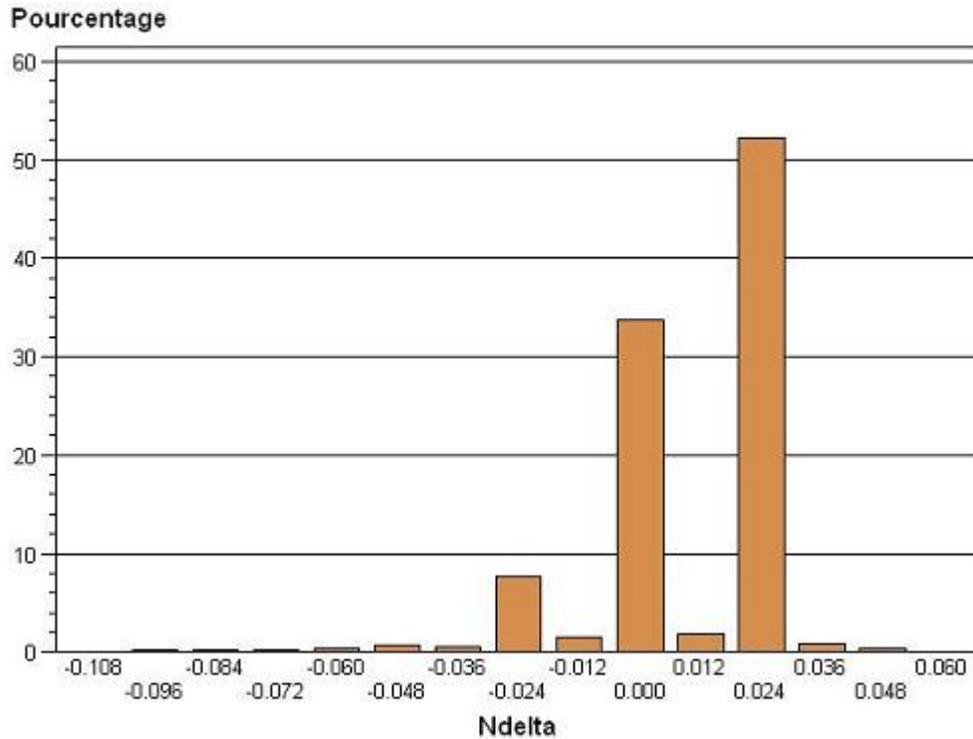


Figure n°94  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : +0%
- Plafond de variation : +3%

*Segment 5 : Marge négative en Formule tous risques étendue*

Ce segment se caractérise par des primes plus élevées et une tension importante : la variation à la hausse est encore plus modérée.

Poids du segment	3%
Variation moyenne	+0.07%
Variation médiane	+0.06%
Ecart-type de variation	1.84%

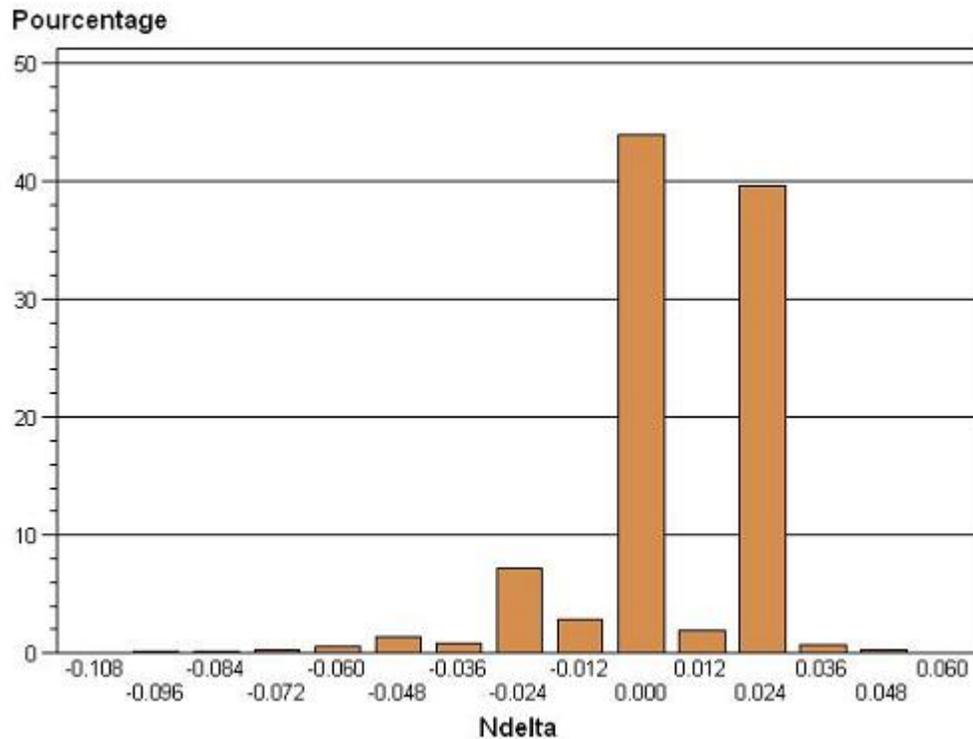


Figure n°95  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : +0%
- Plafond de variation : +2.5%

*Segment 6 : Marge positive en Formule au tiers ou médiane, pour des véhicules de moins de 12 ans*

Même si la marge est positive, les primes sont faibles : l'encadrement de variation est centré sur 0.

Poids du segment	6%
Variation moyenne	-1.80%
Variation médiane	-2.33%
Ecart-type de variation	3.10%

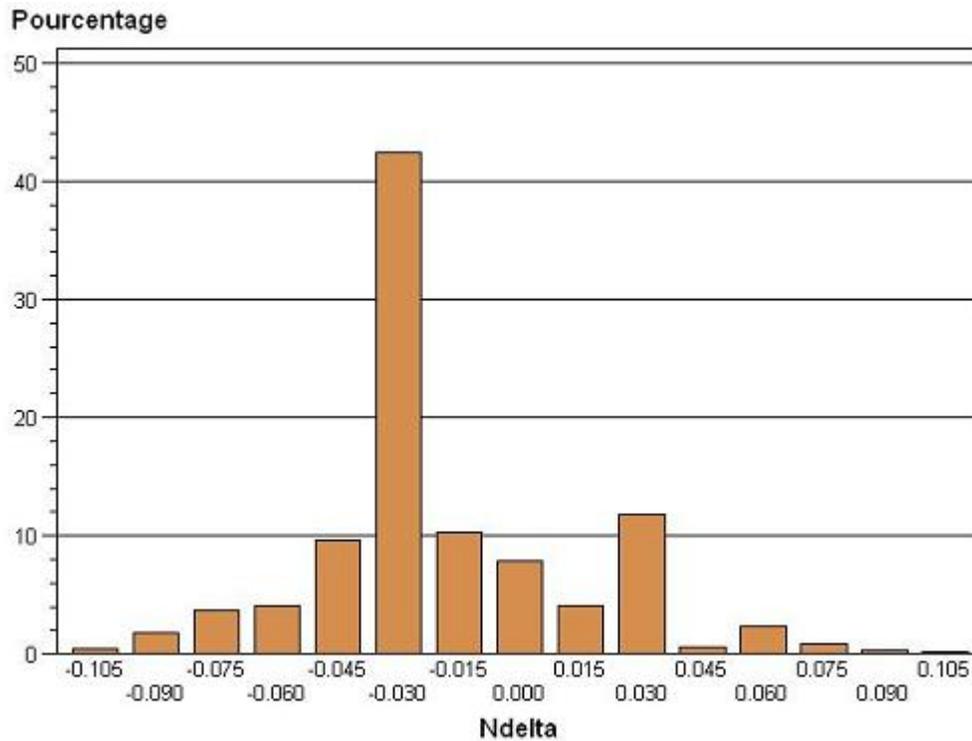


Figure n°96  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : -3%
- Plafond de variation : +3%

*Segment 7 : Marge positive en Formule au tiers ou médiane, pour des véhicules de plus de 12 ans*

Ce segment est à forte résiliation. Cela génère une élasticité importante, qui impose des variations de primes à la baisse.

Poids du segment	13%
Variation moyenne	-2.20%
Variation médiane	-2.33%
Ecart-type de variation	2.86%

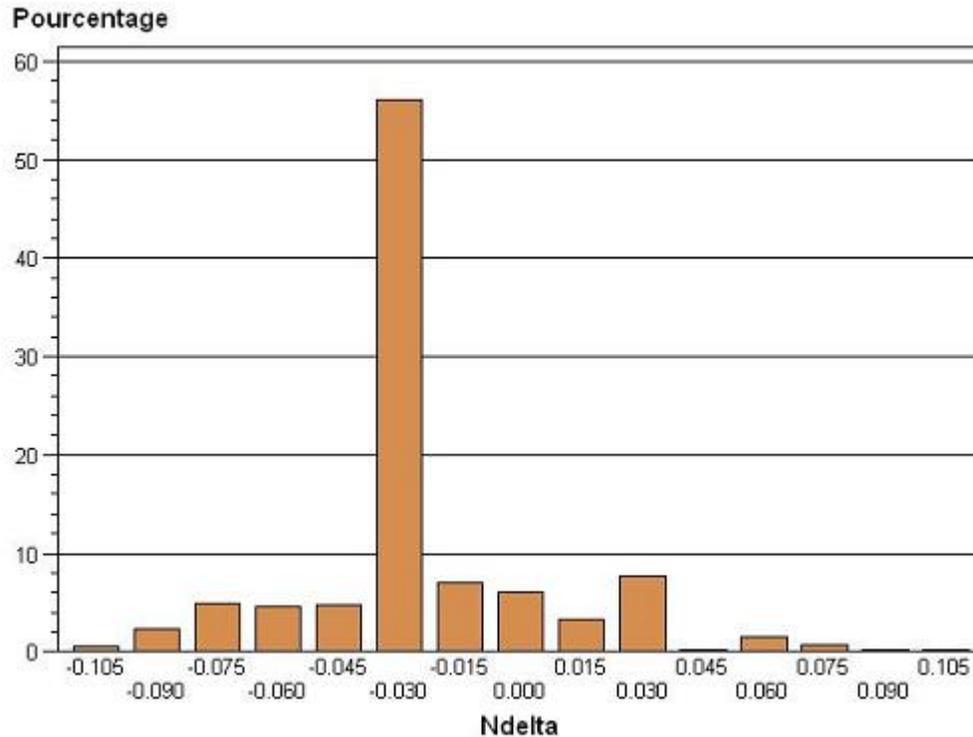


Figure n°97  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant, qui concentre 80% des variations optimales du segment :

- Plancher de variation : -3%
- Plafond de variation : +1%

*Segment 8 : Marge positive en Formule tous risques, avec sinistralité antérieure dans les 36 derniers mois*

Ce segment est à forte résiliation, mais avec une marge insuffisante : les variations optimales sont très étendues sur ce segment.

Poids du segment	10%
Variation moyenne	-1.20%
Variation médiane	-2.33%
Ecart-type de variation	3.06%

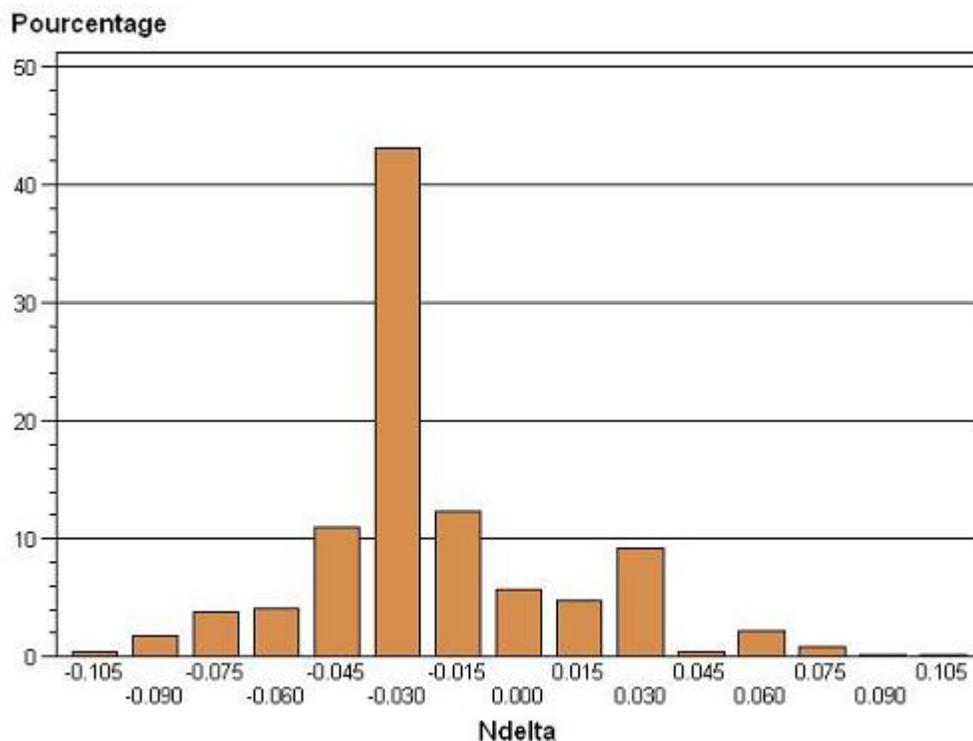


Figure n°98  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant :

- Plancher de variation : -4%
- Plafond de variation : +3%

*Segment 9 : Marge positive en Formule tous risques, sans sinistralité antérieure dans les 36 derniers mois et contrat de moins de 4 ans*

Il s'agit du segment comportant le plus de disparités au niveau des variations optimales. C'est pourtant un segment précieux à forte valeur potentielle. Nous imposons plutôt une variation à la baisse.

Poids du segment	12%
Variation moyenne	-1.34%
Variation médiane	-2.33%
Ecart-type de variation	3.18%

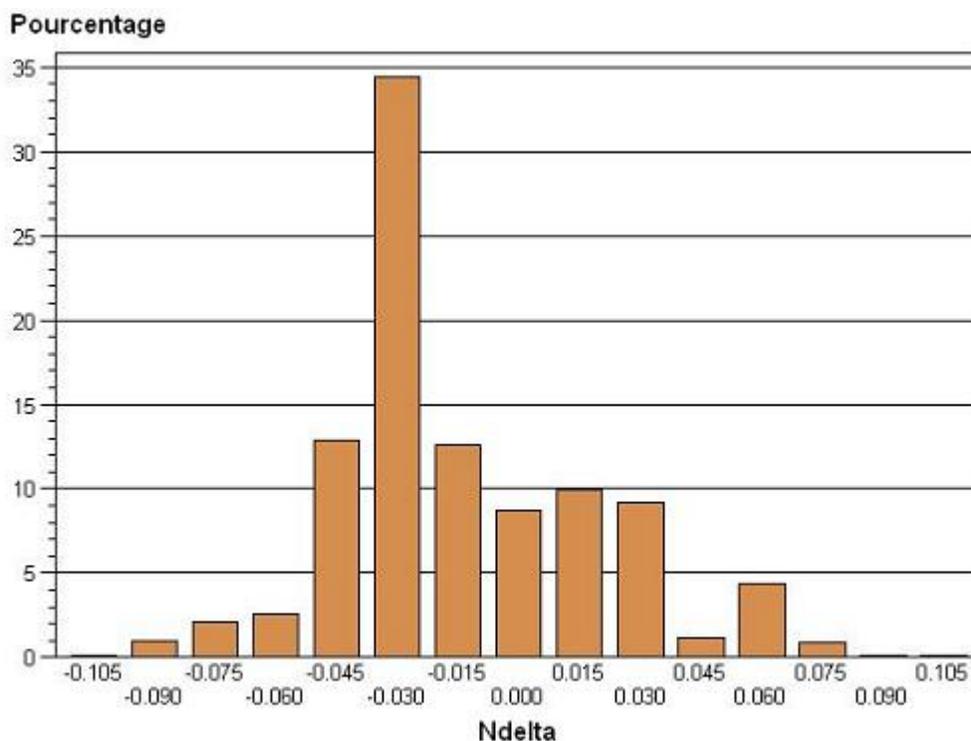


Figure n°99  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant :

- Plancher de variation : -5%
- Plafond de variation : +2%

*Segment 10 : Marge positive en Formule tous risques, sans sinistralité antérieure dans les 36 derniers mois et contrat de plus de 4 ans*

Il s'agit du segment le plus important en terme de volume (portefeuille) et de chiffre d'affaires. Ce segment est précieux car à forte valeur. L'ancienneté du portefeuille limite l'effet d'élasticité. Dans ce cadre, nous imposons donc une variation à la baisse moins importante que pour le segment 9.

Poids du segment	23%
Variation moyenne	-1.60%
Variation médiane	-2.33%
Ecart-type de variation	2.67%

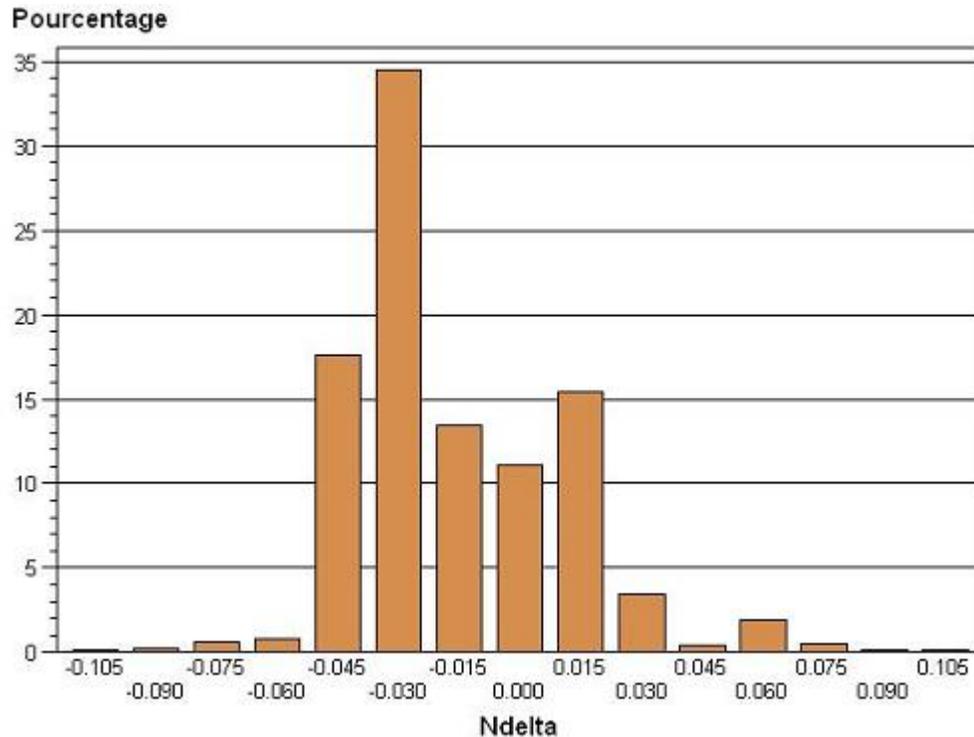


Figure n°100  
distribution des variations optimales  $\Delta_i$  obtenues par la méthode 2

Nous proposons l'encadrement suivant :

- Plancher de variation : -4%
- Plafond de variation : +2%

### 5.3.3 Simulation avec les encadrements obtenus

Les encadrements déterminés précédemment nous permettent d'obtenir un algorithme plus simple à implémenter pour les variations de tarif au terme. Nous allons vérifier que cette troisième méthode ne donne pas de résultats trop éloignés des deux premières, en terme de variation de chiffre d'affaires et de marge obtenue, et qu'elle améliore toujours les résultats réels observés. Pour chaque contrat, nous encadrons la variation optimale par le plancher et le plafond de son segment, déterminés dans la partie précédente.

Nous obtenons le tableau comparatif suivant :

	Avant terme	Après 1 an	Variation
Chiffre d'affaires (CA) observé	614 533 280EUR	531 190 146EUR	-13.56%
CA simulé (optimum recherché) (1)	614 533 280EUR	<b>565 306 895EUR</b>	<b>-8.01%</b>
CA simulé méthode individuelle (2)	614 533 280EUR	536 306 423EUR	-12.77%
CA simulé par encadrement segmenté (3)	614 533 280EUR	537 290 010EUR	-12.57%
Nombre de contrats observés	1 522 200	1 345 145	-11.63%
Nombre de contrats simulés (1)	1 522 200	1 391 570	-8.58%
Nombre de contrats simulés (2)	1 522 200	1 357 613	-10.81%
Nombre de contrats simulés (3)	1 522 200	1 355 657	-10.94%
Marge observée (M)	28 839 046EUR	27 806 814EUR	-3.58%
M simulée finale (1)	28 839 046EUR	28 841 176EUR	+0.01%
M simulée finale (optimum recherché)(2)	28 839 046EUR	<b>29 083 864EUR</b>	<b>+0.85%</b>
M simulée finale (3)	28 839 046EUR	28 607 997EUR	-0.80%

Cette méthode d'encadrement segmenté ne réalise aucune amélioration notable sur les indicateurs.

# Conclusions

## 1. Bilan méthodologique

Dans un environnement très concurrentiel et tiré par les prix, l'assurance automobile se prête généralement bien aux études d'élasticité au prix, et à la modélisation de la résiliation.

En effet, l'analyse exploratoire de notre portefeuille de contrats automobile révèle trois effets majeurs explicatifs de la résiliation après le terme :

- le prix, composante importante de la décision du client,
- le vieillissement : l'élasticité s'amointrit avec le temps en portefeuille et avec l'âge. A contrario, le vieillissement de l'objet du risque, c'est à dire le véhicule, pousse plutôt à la résiliation lors de son changement,
- la sinistralité responsable, dans une moindre mesure : il peut s'agir de résiliation amiable pour des clients en sur-risque, une gestion de sinistre difficilement acceptée, voire une majoration du prix trop forte à la dernière échéance (malus).

L'analyse multivariée confirme l'influence prépondérante du tarif dans la décision de résilier : l'axe principal est uniquement basé sur les cotisations. Les axes suivants retracent le temps et le vieillissement, puis opposent des segments peu représentés (conducteur non désigné, plus de 15 ans d'ancienneté dans la même compagnie...).

Pour mener à bien notre objectif d'optimisation de la prime à l'échéance, il apparaît nécessaire de prendre en considération le niveau et la variation du prix, et donc de déterminer de façon statistique un modèle d'élasticité. Cependant, d'autres critères peuvent influencer la décision de résilier après l'échéance.

Nous avons mis en oeuvre deux méthodes concurrentes afin de cerner les meilleurs effets. Dans l'analyse discriminante, basée sur une linéarité de l'influence du critère, la masse est relativement bien discriminée, mais les comportements extrêmes posent problème : variation de cotisation importante, mises en demeure de paiement (faible volume mais forte suspicion de résilier), non-linéarité de l'effet ancienneté de véhicule dans la résiliation. L'enseignement des résultats obtenus par la discriminante nous a orienté sur la deuxième méthode : la régression logistique.

Le modèle de résiliation, s'appuyant sur la régression logistique, est alors expliqué par des effets plus conformes à la pratique observée, en gardant notamment :

- le coefficient de réduction/majoration (bonus/malus), synthèse de l'ancienneté d'assurance automobile du conducteur principal et de son âge,
- le niveau de couverture des garanties du contrat,
- l'ancienneté du contrat et du véhicule,
- l'évolution de la cotisation lors de la dernière échéance,
- les antécédents de sinistralité,
- le forfait kilométrique souscrit,
- la présence de mise en demeure de paiement dans les 36 derniers mois.

La présence du critère "évolution de la cotisation" permet d'assimiler ce modèle à une élasticité au prix à l'échéance. Il peut donc être mis à profit dans une stratégie d'optimisation de la prime à émettre au terme. Il faut néanmoins passer par une phase de lissage des coefficients obtenus par le biais de la régression logistique afin notamment de se placer dans l'espace continu des variations que nous pouvons appliquer à la cotisation.

La quatrième partie place ce mémoire sur deux choix stratégiques des plus courants :

1. L'entreprise d'assurance cherche-t-elle à maximiser son chiffre d'affaires ? Dans ce cas, il s'agit plutôt d'un objectif commercial qui guide alors notre recherche de l'optimum.
2. L'entreprise d'assurance cherche-t-elle à maximiser sa marge ou son profit ? La démarche s'avère alors plus classique au niveau actuariel. Nous visons la rentabilité à court terme (sur une échéance) de notre variation du prix.

Nous avons fait le choix de répondre à ces deux problématiques, par deux méthodes originales inspirées de l'optimisation sous contraintes. Les résultats de ces méthodes sont probants puisque, pour chaque objectif :

1. un gain de 5.5 points en chiffre d'affaires peut être obtenu par la méthode de variation globale et unique appliquée au portefeuille,
2. un gain de 4 points en marge peut être obtenu par la méthode d'optimisation individuelle (selon les caractéristiques de chaque contrat).

Ces deux résultats d'optimisation (globale ou individuelle) ne conduisent à satisfaire qu'un seul des objectifs fixés. Ce qui rend d'autant plus crucial le choix stratégique préalable : veut-on optimiser le chiffre d'affaires ou la marge ?

Ce mémoire se voulant résolument orienté application, nous proposons ensuite une méthode simplifiée facilement implémentable dans un système d'informations. Nous avons tenté de segmenter les solutions individuelles de l'optimum de marge par un arbre de décision. L'objectif était de produire quelques règles simples de détermination de la variation à appliquer.

## 2. Prolongements

En premier lieu, nous regrettons l'absence du passage de la théorie à la pratique : il aurait été intéressant de confronter les résultats de la simulation avec la réalité d'un échantillon du portefeuille. En effet, les gains simulés nous paraissent exceptionnels à atteindre pour le domaine de l'assurance automobile. Il est possible que les gains réels soient en deçà de ce que nous affichons en simulation. De plus il paraît difficile commercialement de tester en réel des variations différentes entre un échantillon témoin et le reste du portefeuille.

Ce mémoire se concentre sur l'optimisation du prix à un moment privilégié du contrat : son échéance. On pourrait avoir pour ambition d'optimiser le prix à chaque émission de prime sur un contrat : sa souscription, ses avenants, son échéance.

Cependant, nos conclusions ne s'appliquent pas forcément à tous ces actes. La recherche de l'optimum globale de marge (ou de chiffre d'affaires) pourrait conduire à une segmentation différente de la politique tarifaire, en affaire nouvelle, en avenant et au terme.

Par exemple, on pourrait être conduit à augmenter le tarif en entrée des contrats à faible espérance de vie, afin de gagner la marge sur le court terme. A l'inverse en portefeuille, l'effet fidélisant de l'ancienneté du contrat pourrait conduire à diminuer la prime des faibles espérances de vie, donc conserver le contrat pour un gain à plus long terme.

Plus généralement, notre mémoire traite d'une optimisation à court terme de la marge. Nous nous plaçons à une échéance  $T$  et regardons le comportement du portefeuille à l'instant  $T+1$ . L'optimisation de la marge mériterait d'être calculée sur une projection plus longue, comme la valeur actuelle des marges futures à optimiser, afin de prendre en compte l'allongement de l'espérance de vie du contrat, apportée par la variation appliquée à l'échéance  $T$ .

Cela pose le problème complexe d'évaluation des variations futures : la variation optimale en  $T$  est-elle à reproduire sur tous les exercices futurs ?

# Bibliographie

- [1] BOULANGER, TRAN VAN LIEU, *Tarif, Concurrence et Rentabilité : le cas de l'assurance automobile*, Mémoire CEA, 1994.
- [2] CHOVINO M., *Les statistiques : l'analyse de la variance, modélisation*, Europstat Editions, 1996.
- [3] DUSSAULT J.P., *Programmation Linéaire*, Université de Sherbrooke CA, 2005.
- [4] DUYME L., CLAUSTRIAUX J.J., *Note : La régression logistique binaire*, 2006.
- [5] HEBEL P., FAUCONNIER N., DAVID M., *Cahier de recherche n°215 : la nouvelle sensibilité des consommateurs au prix*, CREDOC, 2005.
- [6] KERMORGANT C., *Modélisation du risque de résiliations en assurance automobile*, Mémoire EURIA, 1999.
- [7] LEBART L., PIRON M., MORINEAU A., *Statistique exploratoire multidimensionnelle*, DUNOD, 2006.
- [8] NAKACHE J.P., CONFAIS J., *sStatistique explicative appliquée*, Editions Technip, 2003.
- [9] NAKACHE J.P., CONFAIS J., *Approche pragmatique de la classification*, Editions Technip, 2005.
- [10] SAPORTA G., *Probabilités, analyse de données et statistiques*, Edition DUNOD, 1973.
- [11] TREILHOU A., *Elaboration d'un modèle de résiliation des contrats auto*, Mémoire ISFA, 2000.
- [12] Etude Kollen Partners / Assistance marketing citée par F. DUPONT-METAYER, *Secteur de l'assurance face à un nouvel enjeu de fidélisation*, 2008.
- [13] *Le marché de l'assurance automobile en 2008 - analyse des résultats détaillés*, FFSA/DESSI, 2009.

# Annexes

## Annexe n°1 - Analyse exploratoire

- Le conducteur

AGE DU CONDUCTEUR PRINCIPAL (CP)			
MODALITE	REPARTITION PORTEFEUILLE (%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
CP non désigné	3,30	19,21	5,13
< 21 ans	0,45	16,33	0,59
21-26 ans	6,02	19,20	9,36
27-29 ans	5,48	17,44	7,74
30-34 ans	11,16	15,32	13,84
35-39 ans	11,15	13,39	12,09
40-44 ans	10,70	12,50	10,83
45-49 ans	10,45	11,74	9,94
50-54 ans	10,16	10,55	8,68
55-59 ans	9,79	9,75	7,73
60-64 ans	6,95	9,23	5,20
65-69 ans	4,77	7,87	3,04
70-79 ans	7,15	6,63	3,84
80-84 ans	1,75	8,74	1,24
> 84 ans	0,73	13,12	0,77

NOMBRE DE CONDUCTEUR(S) COMPLEMENTAIRE(S)			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0	37,89	14,71	45,12
1	50,00	11,15	45,15
2	12,10	9,92	9,72

SEXE CONDUCTEUR PRINCIPAL			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
homme	62,61	11,98	60,74
femme	34,10	12,36	34,13
non désigné	3,30	19,21	5,13

<b>AGE CP AU MOMENT DU PERMIS</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
18 ans	29,35	13,16	31,30
19 ans	25,14	12,03	24,50
20 ans	11,05	12,45	11,14
21 ans	7,26	11,99	7,05
22 ans	5,08	11,27	4,64
23 ans	3,83	11,40	3,54
24-25 ans	4,75	11,74	4,52
26-27 ans	3,26	12,43	3,29
28-29 ans	2,46	12,38	2,47
30-32 ans	2,67	12,18	2,64
33-35 ans	1,84	11,83	1,76
36-40 ans	1,77	12,23	1,75
41-50 ans	1,23	11,54	1,15
>50 ans	0,30	11,32	0,28

<b>ANCIENNETE DE PERMIS DU CP</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0-2 ans	0,98	19,99	1,59
3-5 ans	3,30	21,16	5,66
6-10 ans	10,15	17,60	14,46
11-15 ans	11,97	14,96	14,50
16-20 ans	11,18	12,89	11,66
21-25 ans	11,33	12,30	11,28
26-30 ans	12,45	11,09	11,18
31-35 ans	9,73	9,78	7,70
36-40 ans	9,68	8,83	6,92
41-55 ans	14,22	7,58	8,73
> 55 ans	1,71	8,52	1,18
Non désigné	3,30	19,21	5,13

<b>FORMULE DE CONDUITE</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Exclusive	51,49	13,14	54,78
Famille	43,65	10,81	38,21
Tout conducteur	4,86	17,83	7,02

<b>CONDUCTEUR PRINCIPAL AGRICULTEUR (ou assimilé)</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
agri	4,46	7,06	2,55
non agri	95,54	12,60	97,45

<b>MAJORATIONS NOVICE</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
sans majo	98,35	12,30	97,92
avec majo	1,65	15,52	2,08

- Le véhicule

ANCIENNETE DU VEHICULE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0-2 ans	13,00	9,84	10,32
3-4 ans	12,83	10,80	11,17
5-6 ans	13,90	10,67	11,96
7-8 ans	12,56	10,99	11,14
9-10 ans	10,98	12,02	10,65
11-12 ans	10,04	13,18	10,67
13-14 ans	8,32	14,68	9,85
15-16 ans	7,10	15,46	8,85
17-18 ans	5,23	16,19	6,83
19-24 ans	5,08	16,99	6,96
> 24 ans	0,96	20,79	1,61

CLASSE DU VEHICULE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
A	18,12	13,54	19,87
B	26,77	12,19	26,42
C	19,83	11,92	19,13
D	15,00	11,36	13,79
E	7,30	12,02	7,10
F	5,52	12,40	5,54
G	3,27	12,76	3,38
H+	4,19	14,05	4,77

ENERGIE DU VEHICULE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Essence	39,29	12,28	39,08
Diesel	60,64	12,39	60,85
Autre	0,07	11,70	0,07

GROUPE DU VEHICULE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
8 et -	4,88	14,45	5,71
9	10,39	13,35	11,23
10	15,16	12,80	15,72
11	13,03	12,76	13,47
12	13,56	12,43	13,65
13	15,63	10,92	13,82
14	13,06	11,19	11,83
15	6,75	11,77	6,43
16	3,59	12,38	3,59
17	3,96	14,22	4,56

MARQUE VEHICULE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Audi	1,93	14,55	2,27
Asiatiques	5,42	11,81	5,18
BMW	2,25	15,77	2,87
Citroën	13,38	10,75	11,65
Fiat	3,17	13,99	3,60
Ford	5,42	12,94	5,68
Luxe/US	4,04	13,23	4,33
Opel	4,38	12,73	4,51
Peugeot	20,13	11,87	19,34
Renault	28,91	12,12	28,36
Semi-luxe	2,11	13,29	2,27
Volkswagen	5,95	13,65	6,58
Divers	2,91	14,26	3,36

PRESENCE D'UN GARAGE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
sans garage	29,63	13,73	32,95
avec garage	59,03	10,27	49,10
sans objet	11,33	19,56	17,95

- La sinistralité antérieure

JOKER			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
sans joker	15,50	11,43	14,34
avec joker	45,87	8,84	32,81
non impacté	38,63	16,90	52,84

SINISTRALITE 12 DERNIERS MOIS			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0 sinistre	71,32	11,84	68,39
sinistre autre	18,90	13,21	20,21
sinistre non resp	4,45	13,48	4,86
sinistre resp	5,33	15,18	6,55

CATEGORIE ANTECEDENTS			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Acceptable	4,35	20,39	7,18
Bon	6,50	22,30	11,74
Excellent	31,82	14,01	36,10
Super	57,33	9,69	44,99

<b>COEFFICIENT DE REDUCTION MAJORATION</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
100	0,72	24,10	1,41
95-99	1,94	26,57	4,18
90-94	2,09	23,93	4,04
85-89	1,94	21,31	3,34
80-84	2,69	19,01	4,13
76-79	2,99	19,07	4,61
72-75	3,23	16,72	4,38
68-71	3,32	15,92	4,28
64-67	3,36	14,90	4,05
60-63	4,61	15,40	5,75
57-59	4,11	13,78	4,59
54-56	3,92	13,28	4,21
51-53	3,71	12,88	3,86
50 <3ans	14,55	11,31	13,33
50 3-5ans	12,08	10,69	10,45
50 6-10ans	17,94	9,47	13,75
50 11-15ans	9,38	7,79	5,92
50 >15ans	7,42	6,16	3,70

<b>RAPPORT BONUS MALUS VS THEORIE</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
< 1	4,03	10,22	3,34
1	16,54	13,37	17,90
1,1 - 1,3	21,75	13,51	23,79
1,4 - 1,8	19,87	12,68	20,39
1,9 - 3	20,16	10,92	17,82
> 3	17,65	11,73	16,76

- Le contrat

<b>FORFAIT KILOMETRIQUE SOUSCRIT</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
K08	30,88	12,13	30,34
K20	53,68	11,79	51,23
K99	15,44	14,74	18,43

<b>FRACTIONNEMENT DE PAIEMENT</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Annuel	32,09	9,70	25,19
Mensuel	54,86	13,85	61,50
Trimestriel	1,17	19,25	1,82
Semestriel	11,88	11,94	11,49

<b>ANCIENNETE DE CONTRAT</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0-1	18,89	16,82	25,74
2	15,30	15,55	19,27
3	11,87	13,88	13,34
4-5	16,99	12,36	17,00
6-7-8	14,61	10,24	12,12
9-15	11,43	7,99	7,40
16 +	10,91	5,82	5,15

<b>ZONE DE CIRCULATION</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
1	19,30	9,52	14,88
2	11,09	10,81	9,70
3	13,46	11,34	12,36
4	17,97	11,99	17,44
5-6	15,23	13,52	16,67
7	7,97	14,37	9,27
8-9	12,44	16,02	16,14
10	2,54	17,23	3,54

<b>COTISATION PAR FRACTIONNEMENT</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
Mensuel < 25	14,21	14,12	16,24
Mensuel < 35	12,43	11,85	11,92
Mensuel < 50	15,39	12,20	15,20
Mensuel >= 50	12,84	17,45	18,14
trimestriel	1,17	19,25	1,82
Semestriel	11,88	11,94	11,49
annuel >= 500	7,17	12,10	7,02
annuel < 200	6,97	11,50	6,49
annuel < 350	9,48	8,36	6,42
annuel < 500	8,47	7,67	5,26

<b>USAGE TOURNEES</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
courant	98,24	12,26	97,55
tournées	1,76	17,20	2,45

<b>PRESENCE DE MISE EN DEMEURE</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
pas de MED	93,42	11,65	88,13
avec MED	6,58	22,29	11,87

EVOLUTION DE LA PRIME			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
-10%	0,37	14,21	0,42
-9%	1,00	13,43	1,08
-8%	0,75	14,75	0,89
-7%	2,13	11,27	1,95
-6%	2,10	13,35	2,27
-5%	2,40	13,48	2,62
-4%	2,03	15,36	2,52
-3%	3,63	11,67	3,43
-2%	5,77	10,66	4,97
-1%	9,54	11,93	9,21
0%	23,95	9,78	18,97
1%	11,42	13,44	12,43
2%	9,78	11,41	9,03
3%	9,74	12,03	9,49
4%	7,57	13,24	8,11
5%	2,68	18,89	4,10
6%	1,36	20,78	2,29
7%	0,89	21,44	1,55
8%	0,79	20,78	1,32
9%	0,75	17,72	1,08
10%	0,50	19,21	0,78
11%	0,40	20,29	0,65
12%	0,32	21,96	0,57
13%	0,16	20,47	0,27

- Les garanties

FORMULE DE GARANTIE			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
1	11,33	19,56	17,95
2	25,95	14,92	31,34
3	8,42	12,68	8,64
4	44,77	9,28	33,65
5	9,53	10,92	8,42

FRANCHISE VOL			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0	7,81	22,72	14,36
152	62,59	9,86	49,97
304+	18,27	11,98	17,72
SANSVOL	11,33	19,56	17,95

<b>FRANCHISE BRIS DE GLACES</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
0	59,01	10,06	48,07
76	31,31	14,42	36,55
SANS BG	9,68	19,62	15,38

<b>ZERO DECOTE</b>			
MODALITE	REPARTITION PORTEFEUILLE(%)	TAUX DE RESILIATION (%)	REPARTITION RESILIES (%)
sans ZD	86,60	12,65	88,72
avec ZD	13,40	10,40	11,28



## Annexe n°3 - Analyse en composantes multiples

### Théorie

#### • Principe

L'analyse factorielle des composantes multiples (ACM) constitue une généralisation de l'analyse factorielle des correspondances, lorsque la base étudiée comporte plus de deux variables explicatives. Il s'agit d'une technique de description des données qualitatives. Elle a pour objet de mettre en évidence les liaisons non linéaires entre les modalités des différentes variables du modèle. On considère un modèle portant sur  $n$  points et  $p$  variables. Le principe de l'ACM consiste à considérer un espace vectoriel  $\mathbb{R}^p$ , muni d'une métrique, ainsi que la représentation d'un nuage de  $n$  points, muni chacun d'une masse. L'ACM permet de calculer l'inertie totale de ce nuage, de déterminer les axes d'inertie et de repérer les points dans la base formée par ces axes.

*Notations :*

- $I_a$  l'individu  $a$ ,
- $p$ , le nombre de variables,
- $X_i$  le vecteur dont les composantes sont des modalités prises par la  $i$ -ème variable explicative sur les  $n$  individus,
- $c_i$  le nombre de modalités de la variable  $i$ ,
- $X_i(k)$  la modalité  $k$  de la variable  $i$  et  $x_i(k)$  son observation associée  $X_i(k)$ ,
- $X_i^a(k)$  vaut 1 lorsque  $X_i^a = k$ , 0 sinon,
- $X_i^a$  correspond à la valeur de la variable  $i$  pour l'individu  $a$ .
- $n_i(k)$  l'effectif associé à la modalité  $k$  de la variable  $i$ .

Le nombre total des modalités est défini par  $c = \sum_{i=1}^p c_i$ .

Afin d'illustrer la constitution des tables d'analyse, nous utiliserons comme exemple une base de trois individus et de deux variables, ayant chacune deux modalités.

	$X_1$	$X_2$
$I_1$	1	1
$I_2$	1	2
$I_3$	2	1

*Base d'étude servant aux illustrations*

L'analyse des composantes multiples peut être réalisée à partir :

- D'un tableau disjonctif complet  $X$  : celui-ci est bâti en écrivant les  $c$  réponses des  $n$  individus par un codage binaire. Il est constitué en concaténant les différentes variables  $X_i$ ,  $X = (X_i^a(k))_{aik} = |X_1| \dots |X_p|$ .

	$X_1(1)$	$X_1(2)$	$X_2(1)$	$X_2(2)$
$I_1$	1	0	1	0
$I_2$	1	0	0	1
$I_3$	0	1	1	0

*Exemple de tableau disjonctif complet*

- D'une table de Burt  $B$  : elle comporte une ligne et une colonne pour chaque modalité des variables étudiées. Au croisement de la  $k$ -ème modalité de la variable  $i$ ,  $x_i(k)$ , et de la  $q$ -ème modalité de la variable  $j$ ,  $x_j(q)$ , le table de Burt contient le nombre  $\alpha_{qk}$  d'individus vérifiant  $x_i(k)$  et  $x_j(q)$ . La table de Burt est construite à partir du tableau disjonctif complet  $B = X'X$ .

	$X_1(1)$	$X_1(2)$	$X_2(1)$	$X_2(2)$
$X_1(1)$	2	0	1	1
$X_1(2)$	0	1	1	0
$X_2(1)$	1	1	2	0
$X_2(2)$	1	0	0	1

*Exemple de table de Burt*

L'analyse des correspondances multiples consiste à faire un analyse des correspondances simples sur le tableau  $X$ . Il convient de représenter les  $n$  individus et les  $c$  modalités dans un espace  $\mathbb{R}^q$  où  $q = c - p$ .

### • Calcul des composantes principales

*Notations :*

- $Y_m^j$  la coordonnée de l'individu  $m$  sur l'axe  $j$ ,
- $X_{ik}^j$  la coordonnée de la  $k$ -ième modalité de la variable  $i$  sur l'axe  $j$ ,
- $n_i(k)$  le nombre d'individus prenant la modalité  $k$  de la variable  $i$ .

On note :

$$P = \begin{pmatrix} n_1(1) & & 0 \\ & \dots & \\ 0 & & n_p(k) \end{pmatrix}$$

P est une matrice diagonale de dimension "c x c".

Deux matrices sont à diagonaliser :  $C_M = \frac{1}{p} P^{-1} \cdot {}^t X \cdot X$  pour l'analyse des profils lignes (individus) et  $C_{IK} = \frac{1}{p} X \cdot P^{-1} \cdot {}^t X$ .

$C_M$  et  $C_{IK}$  ont les mêmes valeurs propres  $\lambda_j$ . On considère les vecteurs propres normés  $v^1 \dots v^q$  de  $C_{IK}$  associés au  $q$  plus grandes valeurs propres différentes et  $w^1 \dots w^q$  les valeurs propres normées de  $C_M$  associées aux mêmes valeurs propres.

On obtient alors :

$$X_{ik}^j = \frac{1}{n_i(k) \cdot \sqrt{\lambda_j}} \sum_m X_i^m(k) \cdot Y_m^j \text{ avec } Y_m^j = \frac{1}{n \cdot \sqrt{\lambda_j}} \sum_{ik} X_i^m(k) \cdot X_i^j(k)$$

#### • Inertie des axes

L'inertie totale  $I$  est égale au rapport entre le nombre de modalités et le nombre de variables diminué de 1 :

$$I = \frac{c}{p} - 1 = \frac{q}{p}$$

Comme elle n'est pas liée au  $\chi^2$ , elle ne dépend pas des liaisons entre les variables et n'a donc pas de signification statistique particulière.

L'inertie  $I_i$  de la variable  $X_i$  est égale à :

$$I_i = \frac{c_i - 1}{p}$$

Elle est définie à partir de son nombre de modalités. Il est donc conseillé de ne pas avoir de trop fortes disparités sur le nombre de modalités entre les variables afin de ne pas biaiser les résultats.

L'inertie  $I_i(k)$  de modalité  $k$  de la variable  $X_i$ , c'est à dire sa contribution à l'inertie totale, est définie par :

$$I_i(k) = \frac{1}{p} \cdot \left(1 - \frac{n_i(k)}{n}\right)$$

Le fait que l'inertie des modalités de variables soit inversement proportionnelle à leur effectif nous montre l'intérêt de ne pas constituer de modalité avec un trop petit effectif, sous peine de déséquilibrer les résultats.

#### • Variable supplémentaire

L'analyse des composantes multiples permet d'intégrer des variables supplémentaires dans la représentation du modèle. Les centres de groupes d'individus définis par les modalités des variables supplémentaires sont alors projetés dans l'espace des variables.

#### • Eléments d'interprétation

En analyse des composantes multiples, l'inertie de chaque axe est faible. Il est, par conséquent, difficile de déterminer le nombre optimum d'axes factoriels retenus dans le cadre de l'interprétation des résultats. Il convient, en effet, d'étudier un nombre d'axes suffisant pour obtenir une assez bonne visualisation de la base d'étude tout en synthétisant l'information mais en limitant au maximum le nombre d'axes d'étude. Les valeurs propres différentes, associées aux axes factoriels sont au nombre de  $c - p = q$ . En tenant compte de la définition de l'inertie totale, on en déduit que la valeur moyenne des valeurs propres est  $\frac{1}{p}$ .

Dans le cadre de l'analyse des composantes principales, on détermine le nombre d'axes factoriels en utilisant le critère de Kaiser. Ce dernier consiste à ne conserver que les axes ayant une valeur propre supérieure à la moyenne. Cet argument peut aussi être retenu dans le cadre de l'analyse des composantes multiples. Il convient cependant de conforter la décision prise en vérifiant la présence d'un « coude » sur le diagramme des valeurs propres.

*Notations :*

- $v_j(x_i(k))$  la coordonnée de la modalité  $x_i(k)$  sur l'axe factoriel  $j$ ,
- $\lambda_j$  la valeur propre de l'axe  $j$ .

L'interprétation des axes factoriels se fait en déterminant les modalités  $x_i(k)$  ayant la plus forte contribution. La contribution de la modalité  $x_i(k)$  est définie par :

$$C_j(x_i(k)) = \frac{v_j^2}{\lambda_j} \cdot \frac{n_i(k)}{n \cdot p}$$

En pratique, on s'intéresse uniquement aux modalités  $x_i(k)$  ayant une contribution supérieure à leur poids  $\frac{n_i(k)}{n \cdot p}$  c'est à dire dont la coordonnée est supérieure à  $\sqrt{\lambda_j}$ .

La liaison entre un axe factoriel  $j$  et une variable supplémentaire  $\alpha$  est déterminée en calculant la valeur-test pour chaque modalité de la variable. Pour chaque modalité  $i$  de la variable supplémentaire  $\alpha$ , notée  $\alpha_i$ , on considère :

- $v_j(\alpha_i)$  la coordonnée de la modalité  $\alpha_i$  sur l'axe factoriel  $j$ ,
- $n_{\alpha_i}$  l'effectif de la modalité  $i$  de la variable  $\alpha$ .

la valeur-test  $VF(\alpha_i)$  de la modalité  $\alpha_i$  est définie par :

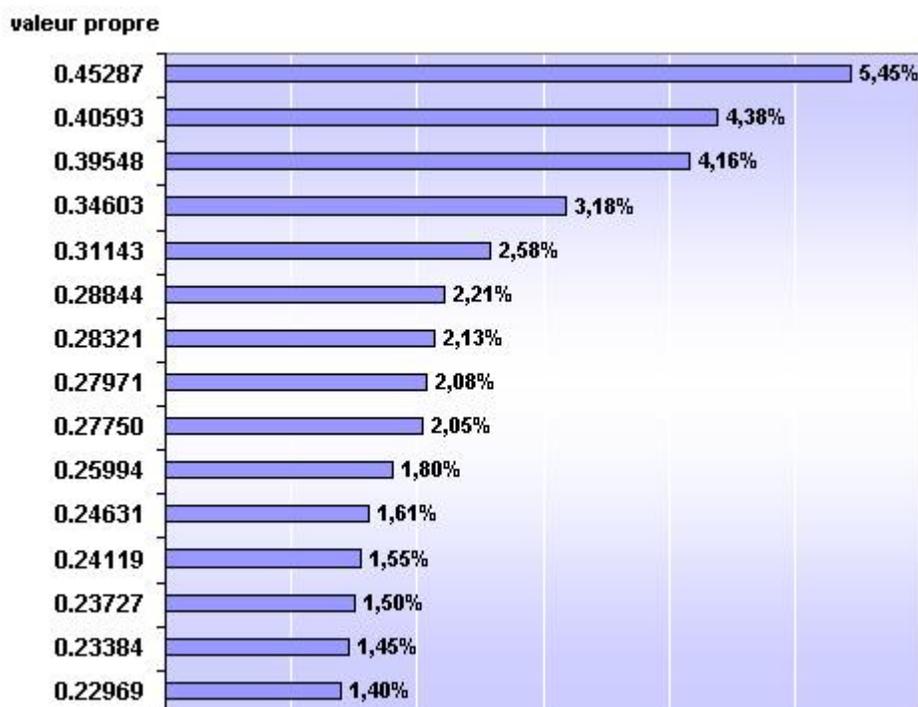
$$VF(\alpha_i) = v_j(\alpha_i) \cdot \sqrt{n_{\alpha_i}} \cdot \sqrt{\frac{n-1}{n-n_{\alpha_i}}}$$

Sous l'hypothèse que la valeur-test suit une loi Normale centrée réduite, la modalité  $\alpha_i$  est considérée comme significative dès lors que  $|VT(\alpha_i)| > 1,96$ .

Il convient également de prendre en compte la qualité de représentation d'une modalité sur un axe factoriel. Celle-ci est déterminée par le carré du cosinus de l'angle du vecteur de la modalité par l'axe dans le plan. Si le cosinus est grand, la modalité est voisine de l'axe. Elle est alors analysable par projection de la modalité sur l'axe, en comparaison aux autres modalités. Dans le cadre de l'analyse, il convient de ne conserver que les modalités permettant une interprétation significative. On ne retiendra que les modalités pour lesquelles le carré du cosinus est supérieur à 0,10.

## Application

- Premières valeurs propres et contributions associées



- Contribution des variables

VAR	NAME_	Quality	Mass	Inertia	Dim1	Contr1	SqCos1	Dim2	Contr2	SqCos2	Dim3	Contr3	SqCos3
AGE COND	≤ 29 ans	0,67	0,01	0,01	0,49	0,01	0,03	- 1,16	0,04	0,18	0,46	0,01	0,03
AGE COND	≤ 39 ans	0,71	0,01	0,01	0,16	0,00	0,01	- 0,76	0,03	0,17	0,18	0,00	0,01
AGE COND	≤ 49 ans	0,78	0,01	0,01	0,02	0,00	0,00	- 0,09	0,00	0,00	- 0,16	0,00	0,01
AGE COND	≤ 59 ans	0,71	0,01	0,01	- 0,10	0,00	0,00	0,39	0,01	0,04	- 0,40	0,01	0,04
AGE COND	> 60 ans	0,66	0,01	0,01	- 0,16	0,00	0,01	0,85	0,04	0,20	- 0,62	0,02	0,10
AGE COND	Non désigné	0,96	0,00	0,01	- 1,37	0,01	0,06	2,07	0,03	0,15	4,61	0,18	0,72
ANC VEH	≤ 3 ans	0,22	0,01	0,01	- 0,72	0,02	0,12	0,02	0,00	0,00	0,19	0,00	0,01
ANC VEH	≤ 6 ans	0,12	0,01	0,01	- 0,54	0,01	0,07	- 0,16	0,00	0,01	- 0,01	0,00	0,00
ANC VEH	≤ 10 ans	0,04	0,01	0,01	- 0,19	0,00	0,01	- 0,19	0,00	0,01	- 0,09	0,00	0,00
ANC VEH	≤ 15 ans	0,14	0,01	0,01	0,50	0,01	0,07	- 0,03	0,00	0,00	- 0,06	0,00	0,00
ANC VEH	> 15 ans	0,35	0,01	0,01	1,25	0,04	0,27	0,54	0,01	0,05	- 0,01	0,00	0,00
GARAGE	Avec	0,30	0,03	0,00	- 0,38	0,02	0,21	0,06	0,00	0,00	- 0,23	0,01	0,08
GARAGE	sans	0,20	0,01	0,01	- 0,13	0,00	0,01	- 0,30	0,01	0,04	0,26	0,01	0,03
GARAGE	non défini	0,97	0,00	0,01	2,32	0,12	0,69	0,49	0,01	0,03	0,53	0,01	0,04

VAR	NAME_	Quality	Mass	Inertia	Dim1	Contr1	SqCos1	Dim2	Contr2	SqCos2	Dim3	Contr3	SqCos3
CLASSE VEH	A	0,34	0,01	0,01	0,88	0,03	0,17	0,32	0,00	0,02	- 0,06	0,00	0,00
CLASSE VEH	B	0,14	0,01	0,01	0,11	0,00	0,00	0,06	0,00	0,00	0,07	0,00	0,00
CLASSE VEH	C	0,12	0,01	0,01	- 0,14	0,00	0,00	- 0,14	0,00	0,00	0,01	0,00	0,00
CLASSE VEH	D	0,11	0,01	0,01	- 0,33	0,00	0,02	- 0,17	0,00	0,01	- 0,06	0,00	0,00
CLASSE VEH	E-F	0,18	0,01	0,01	- 0,52	0,01	0,04	- 0,13	0,00	0,00	0,04	0,00	0,00
CLASSE VEH	G et +	0,46	0,00	0,01	- 0,62	0,01	0,03	- 0,06	0,00	0,00	- 0,08	0,00	0,00
FRACTIONNEMENT	annuel	0,64	0,01	0,01	- 0,14	0,00	0,01	0,89	0,06	0,37	- 0,14	0,00	0,01
FRACTIONNEMENT	mensuel	0,61	0,02	0,01	0,09	0,00	0,01	- 0,59	0,05	0,42	0,10	0,00	0,01
FRACTIONNEMENT	semestriel	0,95	0,01	0,01	- 0,06	0,00	0,00	0,35	0,00	0,02	- 0,11	0,00	0,00
FRACTIONNEMENT	trimestriel	0,99	0,00	0,01	0,43	0,00	0,00	- 0,31	0,00	0,00	0,45	0,00	0,00
ANC CONT	≤ 1 an	0,10	0,01	0,01	0,23	0,00	0,01	- 0,38	0,01	0,03	0,35	0,01	0,03
ANC CONT	2 ans	0,04	0,01	0,01	0,10	0,00	0,00	- 0,29	0,00	0,02	0,29	0,00	0,02
ANC CONT	3 ans	0,01	0,01	0,01	0,04	0,00	0,00	- 0,15	0,00	0,00	0,19	0,00	0,00
ANC CONT	≤ 5 ans	0,02	0,01	0,01	0,04	0,00	0,00	- 0,14	0,00	0,00	0,07	0,00	0,00
ANC CONT	≤ 8 ans	0,02	0,01	0,01	- 0,09	0,00	0,00	0,00	0,00	0,00	- 0,14	0,00	0,00
ANC CONT	≤ 15 ans	0,06	0,00	0,01	- 0,20	0,00	0,01	0,50	0,01	0,03	- 0,40	0,00	0,02
ANC CONT	> 15 ans	0,24	0,00	0,01	- 0,31	0,00	0,01	0,93	0,02	0,11	- 0,72	0,01	0,06
ANTECEDANTS	Acceptable	0,07	0,00	0,01	0,30	0,00	0,00	- 0,63	0,00	0,02	0,22	0,00	0,00
ANTECEDANTS	Bon	0,16	0,00	0,01	0,53	0,00	0,02	- 0,83	0,01	0,05	0,36	0,00	0,01
ANTECEDANTS	Excellent	0,07	0,01	0,01	- 0,10	0,00	0,00	- 0,32	0,01	0,05	0,09	0,00	0,00
ANTECEDANTS	Super	0,24	0,02	0,00	- 0,03	0,00	0,00	0,32	0,01	0,13	- 0,11	0,00	0,02
ZONE CIRCULATION	1	0,06	0,01	0,01	- 0,11	0,00	0,00	0,26	0,00	0,02	- 0,20	0,00	0,01
ZONE CIRCULATION	2	0,01	0,00	0,01	- 0,04	0,00	0,00	0,16	0,00	0,00	- 0,13	0,00	0,00
ZONE CIRCULATION	3	0,00	0,01	0,01	- 0,01	0,00	0,00	0,07	0,00	0,00	- 0,10	0,00	0,00
ZONE CIRCULATION	4	0,00	0,01	0,01	- 0,00	0,00	0,00	- 0,02	0,00	0,00	- 0,02	0,00	0,00
NB COND COMPL	0	0,24	0,02	0,01	0,12	0,00	0,01	0,05	0,00	0,00	0,48	0,02	0,14
NB COND COMPL	1	0,20	0,02	0,01	- 0,06	0,00	0,00	- 0,15	0,00	0,02	- 0,23	0,01	0,05
NB COND COMPL	2 et +	0,26	0,01	0,01	- 0,13	0,00	0,00	0,46	0,01	0,03	- 0,56	0,01	0,04
MISE EN DEMEURE	oui	0,05	0,00	0,01	0,34	0,00	0,01	- 0,48	0,00	0,02	0,31	0,00	0,01
MISE EN DEMEURE	non	0,05	0,04	0,00	- 0,02	0,00	0,01	0,03	0,00	0,02	- 0,02	0,00	0,01
FORFAIT KM	8000 km	0,31	0,01	0,01	0,53	0,02	0,12	0,48	0,02	0,10	- 0,21	0,00	0,02
FORFAIT KM	20000 km	0,13	0,02	0,01	- 0,14	0,00	0,02	- 0,25	0,01	0,07	- 0,04	0,00	0,00
FORFAIT KM	illimité	0,15	0,01	0,01	- 0,56	0,01	0,06	- 0,09	0,00	0,00	0,55	0,01	0,06
FORMULE COND	Exclusive	0,29	0,02	0,01	0,13	0,00	0,02	- 0,34	0,01	0,12	- 0,01	0,00	0,00
FORMULE COND	Famille	0,28	0,02	0,01	- 0,04	0,00	0,00	0,21	0,00	0,04	- 0,36	0,01	0,10
FORMULE COND	Non désigné	0,78	0,00	0,01	- 0,96	0,01	0,05	1,65	0,03	0,14	3,36	0,14	0,58
ANC PERMIS	≤ 10 ans	0,74	0,01	0,01	0,56	0,01	0,05	- 1,15	0,05	0,22	0,47	0,01	0,04
ANC PERMIS	≤ 20 ans	0,72	0,01	0,01	0,12	0,00	0,00	- 0,70	0,03	0,15	0,14	0,00	0,01
ANC PERMIS	≤ 30 ans	0,80	0,01	0,01	- 0,02	0,00	0,00	0,02	0,00	0,00	- 0,22	0,00	0,02
ANC PERMIS	≤ 40 ans	0,70	0,01	0,01	- 0,14	0,00	0,00	0,53	0,01	0,07	- 0,48	0,01	0,06
ANC PERMIS	> 40 ans	0,67	0,01	0,01	- 0,20	0,00	0,01	0,94	0,03	0,17	- 0,67	0,02	0,08
ANC PERMIS	Non désigné	0,96	0,00	0,01	- 1,37	0,01	0,06	2,07	0,03	0,15	4,61	0,18	0,72
GRPE VEHICULE	9 et -	0,33	0,01	0,01	1,01	0,03	0,18	0,44	0,01	0,04	- 0,03	0,00	0,00
GRPE VEHICULE	10	0,09	0,01	0,01	0,30	0,00	0,02	0,11	0,00	0,00	0,06	0,00	0,00
GRPE VEHICULE	11	0,06	0,01	0,01	0,14	0,00	0,00	- 0,01	0,00	0,00	0,07	0,00	0,00
GRPE VEHICULE	12	0,06	0,01	0,01	- 0,04	0,00	0,00	- 0,03	0,00	0,00	0,10	0,00	0,00
GRPE VEHICULE	13	0,12	0,01	0,01	- 0,34	0,00	0,02	- 0,18	0,00	0,01	- 0,09	0,00	0,00
GRPE VEHICULE	14	0,10	0,01	0,01	- 0,50	0,01	0,04	- 0,19	0,00	0,01	- 0,05	0,00	0,00
GRPE VEHICULE	15 et +	0,65	0,01	0,01	- 0,65	0,01	0,07	- 0,17	0,00	0,00	- 0,04	0,00	0,00
FRANCHISE VOL	0 EUR	0,10	0,00	0,01	- 0,64	0,00	0,02	0,51	0,00	0,01	- 0,54	0,00	0,01
FRANCHISE VOL	Franchise mini	0,51	0,03	0,00	- 0,15	0,00	0,04	- 0,06	0,00	0,01	- 0,13	0,00	0,03
FRANCHISE VOL	Franchise max	0,33	0,01	0,01	- 0,74	0,02	0,12	- 0,21	0,00	0,01	0,26	0,00	0,02
FRANCHISE VOL	sans gar vol	0,97	0,00	0,01	2,32	0,12	0,69	0,49	0,01	0,03	0,53	0,01	0,04

VAR	NAME	Quality	Mass	Inertia	Dim1	Contr1	SqCos1	Dim2	Contr2	SqCos2	Dim3	Contr3	SqCos3
FRANCHISE BDG	sans	0,55	0,03	0,00	- 0,47	0,03	0,32	0,03	0,00	0,00	- 0,08	0,00	0,01
FRANCHISE BDG	avec	0,57	0,01	0,01	0,15	0,00	0,01	- 0,22	0,00	0,02	- 0,01	0,00	0,00
FRANCHISE BDG	sans gar BDG	0,88	0,00	0,01	2,42	0,11	0,63	0,53	0,01	0,03	0,55	0,01	0,03
DEROGATION	≤ 97	0,03	0,01	0,01	0,05	0,00	0,00	- 0,35	0,00	0,02	0,07	0,00	0,00
DEROGATION	≤ 99	0,00	0,01	0,01	- 0,00	0,00	0,00	- 0,00	0,00	0,00	- 0,05	0,00	0,00
DEROGATION	≤ 100	0,02	0,01	0,01	- 0,18	0,00	0,01	0,11	0,00	0,00	- 0,05	0,00	0,00
DEROGATION	≤ 101	0,01	0,00	0,01	0,18	0,00	0,00	0,04	0,00	0,00	- 0,03	0,00	0,00
DEROGATION	≤ 103	0,01	0,01	0,01	- 0,09	0,00	0,00	0,20	0,00	0,01	- 0,06	0,00	0,00
DEROGATION	≤ 110	0,03	0,01	0,01	0,21	0,00	0,01	- 0,12	0,00	0,00	0,17	0,00	0,01
SEXE COND PRIN	Femme	0,11	0,01	0,01	0,10	0,00	0,01	- 0,28	0,01	0,04	- 0,09	0,00	0,00
SEXE COND PRIN	Homme	0,14	0,03	0,00	0,02	0,00	0,00	0,04	0,00	0,00	- 0,19	0,01	0,06
SEXE COND PRIN	Non désigné	0,96	0,00	0,01	- 1,37	0,01	0,06	2,07	0,03	0,15	4,61	0,18	0,72
RED AGRI	oui	0,11	0,00	0,01	- 0,14	0,00	0,00	1,09	0,01	0,06	- 0,58	0,00	0,02
RED AGRI	Non désigné	0,11	0,04	0,00	0,01	0,00	0,00	- 0,05	0,00	0,06	0,03	0,00	0,02
CRM	50 de +10 ans	0,30	0,01	0,01	- 0,18	0,00	0,01	0,95	0,04	0,18	- 0,60	0,02	0,07
CRM	50 6-10 ans	0,14	0,01	0,01	- 0,14	0,00	0,00	0,59	0,01	0,05	- 0,25	0,00	0,01
CRM	50 4-5 ans	0,07	0,01	0,01	- 0,13	0,00	0,00	0,30	0,00	0,01	- 0,06	0,00	0,00
CRM	50 0-1 ans	0,06	0,01	0,01	- 0,23	0,00	0,01	0,06	0,00	0,00	0,02	0,00	0,00
CRM	51-63	0,21	0,01	0,01	0,01	0,00	0,00	- 0,46	0,01	0,04	0,22	0,00	0,01
CRM	64-79	0,19	0,01	0,01	0,32	0,00	0,01	- 0,86	0,02	0,11	0,37	0,00	0,02
CRM	80-99	0,47	0,00	0,01	0,71	0,01	0,05	- 1,11	0,03	0,13	0,59	0,01	0,04
FORMULE GAR	FG1	0,97	0,00	0,01	2,32	0,12	0,69	0,49	0,01	0,03	0,53	0,01	0,04
FORMULE GAR	FG2	0,64	0,01	0,01	0,32	0,01	0,04	- 0,15	0,00	0,01	- 0,12	0,00	0,01
FORMULE GAR	FG3	0,06	0,00	0,01	- 0,24	0,00	0,01	- 0,36	0,00	0,01	0,10	0,00	0,00
FORMULE GAR	FG4	0,47	0,02	0,01	- 0,59	0,03	0,28	0,10	0,00	0,01	- 0,10	0,00	0,01
FORMULE GAR	FG5	0,12	0,00	0,01	- 0,67	0,01	0,05	- 0,32	0,00	0,01	0,09	0,00	0,00
SINISTRALITE	sans sinistre	0,10	0,03	0,00	0,14	0,00	0,05	0,12	0,00	0,03	- 0,05	0,00	0,01
SINISTRALITE	non resp	0,01	0,00	0,01	- 0,19	0,00	0,00	- 0,35	0,00	0,01	0,13	0,00	0,00
SINISTRALITE	resp	0,02	0,00	0,01	- 0,28	0,00	0,00	- 0,10	0,00	0,00	0,22	0,00	0,00
SINISTRALITE	autre	0,07	0,01	0,01	- 0,40	0,01	0,04	- 0,32	0,00	0,02	0,10	0,00	0,00
FRACT / COTIS	men -25EUR	0,36	0,01	0,01	1,04	0,03	0,18	- 0,10	0,00	0,00	0,00	0,00	0,00
FRACT / COTIS	men -35EUR	0,13	0,01	0,01	0,05	0,00	0,00	- 0,49	0,01	0,03	- 0,09	0,00	0,00
FRACT / COTIS	men -50EUR	0,18	0,01	0,01	- 0,29	0,00	0,02	- 0,73	0,02	0,10	0,05	0,00	0,00
FRACT / COTIS	men +51EUR	0,37	0,01	0,01	- 0,47	0,01	0,03	- 1,06	0,04	0,17	0,43	0,01	0,03
FRACT / COTIS	Semestriel	0,95	0,01	0,01	- 0,06	0,00	0,00	0,35	0,00	0,02	- 0,11	0,00	0,00
FRACT / COTIS	an -200EUR	0,35	0,00	0,01	1,27	0,02	0,12	1,37	0,03	0,14	- 0,29	0,00	0,01
FRACT / COTIS	an -350EUR	0,22	0,00	0,01	- 0,10	0,00	0,00	0,93	0,02	0,09	- 0,51	0,01	0,03
FRACT / COTIS	an -500EUR	0,18	0,00	0,01	- 0,65	0,01	0,04	0,72	0,01	0,05	- 0,28	0,00	0,01
FRACT / COTIS	an +501EUR	0,30	0,00	0,01	- 0,97	0,01	0,07	0,56	0,01	0,02	0,65	0,01	0,03
FRACT / COTIS	trimestriel	0,99	0,00	0,01	0,43	0,00	0,00	- 0,31	0,00	0,00	0,45	0,00	0,00

<b>VAR</b>	<b>_NAME_</b>	<b>Quality</b>	<b>Mass</b>	<b>Inertia</b>	<b>Dim1</b>	<b>Contr1</b>	<b>SqCos1</b>	<b>Dim2</b>	<b>Contr2</b>	<b>SqCos2</b>	<b>Dim3</b>	<b>Contr3</b>	<b>SqCos3</b>
CLIENT	0 gamme	0,06	0,00	0,01	- 0,28	0,00	0,01	0,42	0,00	0,02	- 0,23	0,00	0,01
CLIENT	1 gamme	0,12	0,01	0,01	0,21	0,00	0,02	- 0,40	0,01	0,06	0,02	0,00	0,00
CLIENT	2 gammes	0,02	0,01	0,01	0,03	0,00	0,00	- 0,00	0,00	0,00	- 0,04	0,00	0,00
CLIENT	3 gammes	0,05	0,01	0,01	- 0,10	0,00	0,00	0,22	0,00	0,01	0,13	0,00	0,01
CLIENT	4 gammes et +	0,01	0,00	0,01	- 0,20	0,00	0,00	0,25	0,00	0,01	- 0,01	0,00	0,00
AGENT	Valeur 1	0,02	0,00	0,01	0,15	0,00	0,00	- 0,13	0,00	0,00	0,13	0,00	0,00
AGENT	Valeur 2	0,00	0,02	0,01	- 0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00
AGENT	Valeur 3	0,01	0,01	0,01	- 0,05	0,00	0,00	0,03	0,00	0,00	- 0,04	0,00	0,00
AGENT	non renseigné	0,00	0,00	0,01	0,04	0,00	0,00	0,00	0,00	0,00	- 0,02	0,00	0,00
RESILIE	non	0,03			- 0,04		0,01	0,03		0,01	- 0,03		0,01
RESILIE	oui	0,03			0,33		0,01	- 0,20		0,01	0,26		0,01

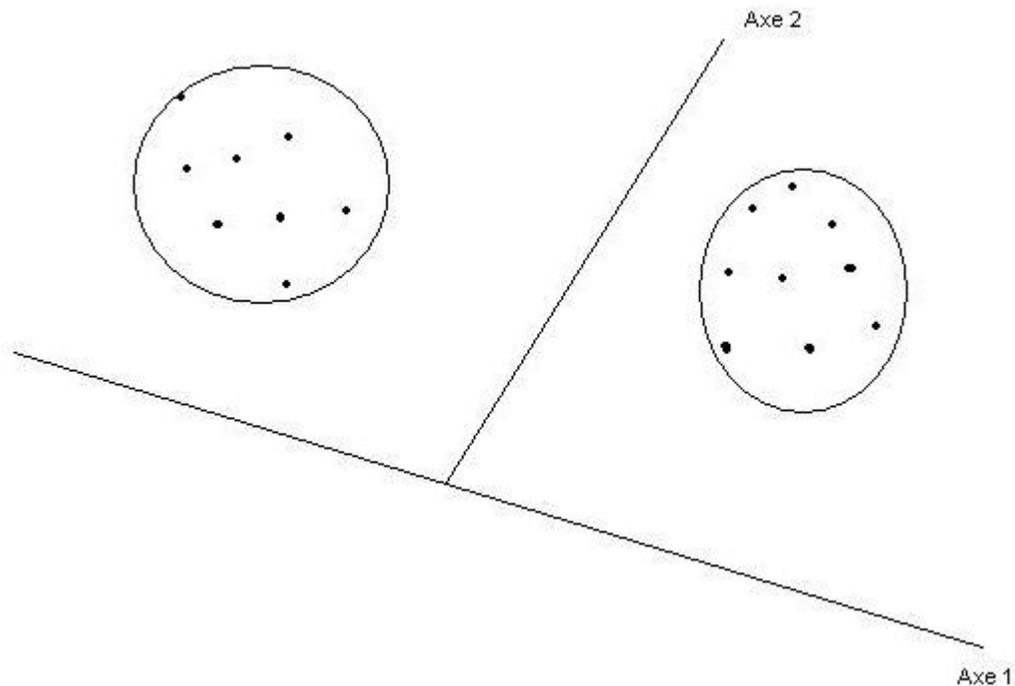
## Annexe n°4 - Analyse discriminante "Stepwise"

### Théorie

#### • Principe général

L'objectif est d'étudier une population de  $n$  individus répartis en  $x$  classes, pour une variable donnée. L'analyse discriminante est une méthode permettant de classer les individus, en se basant sur  $p$  variables quantitatives de décrire chaque individu.

Dans le cadre de l'analyse discriminante, nous recherchons donc les combinaisons linéaires pouvant exister entre les  $p$  variables explicatives et la variable qualitative d'appartenance. La méthode consiste à séparer au mieux les groupes d'observation définis par les modalités de la variable de classe.



*Principe de l'analyse discriminante*

Dans la suite de ce chapitre théorique, nous considérons un tableau de données, constitué de  $n$  individus, représentés en ligne, des  $p$  variables explicatives et de la variables de classement, représentées en colonne.

Notations :

- $X$  représente la matrice de dimension  $np$ ,
- $X_j^i$  représente la  $j$ -ème variable associée à l'individu  $i$ ,
- $X_j$ , le vecteur de  $\mathbb{R}^n$  représentant les valeurs de la  $j$ -ème variable sur l'ensemble des individus,
- $X^i$ , le vecteur sur  $\mathbb{R}^p$  déterminant, pour l'individu  $i$ , les modalités prises pour les  $p$  variables explicatives,
- $Y^i$  représente la valeur de la variable de classe pour l'individu  $i$  et  $Y$  le vecteur de  $\mathbb{R}^n$  associé.  $Y$  peut prendre  $x$  valeurs.
- $N = \{X^i \mid i = 1, \dots, n\}$  constitue la représentation de l'ensemble des individus sur  $\mathbb{R}^p$ ,
- $N_1 \dots N_x$  constituent des sous-nuages de  $N$  :  $N_k = \{X^i \mid Y^i = k\}$ ,
- $p_i = \frac{1}{n}$  le poids associé à l'individu  $i$  et  $P_k = \sum_{i|Y^i=k} p_i$ ,
- $\tilde{X}_j^i$  la variable  $j$  centrée, associée à l'individu  $i$ .

Les centres de gravité des nuages  $N_k$  et  $N$  sont alors désignés par :

$$g_k = \frac{1}{P_k} \sum_{i|Y^i=k} p_i \cdot X^i \text{ et } g = \sum_{i=1}^n p_i \cdot X^i = \sum_{k=1}^x P_k \cdot g_k$$

De plus, on considère les matrices de covariance suivantes :

Matrice de covariance totale	$V = \sum_{i=1}^n p_i \cdot (X^i - g) \cdot {}^t(X^i - g)$
Matrice de covariance intra-classe	$W = \sum_{k=1}^x \sum_{i Y^i=k} p_i \cdot (X^i - g_k) \cdot {}^t(X^i - g_k)$
Matrice de covariance inter-classe	$B = \sum_{k=1}^x P_k \cdot (g_k - g) \cdot {}^t(g_k - g)$

D'après le théorème de Huygens,  $V = W + B$ .

L'objectif de l'analyse discriminante est de déterminer  $s$  variables  $(a_1, \dots, a_s)$ , définies comme combinaisons linéaires des  $p$  variables explicatives initiales, et qui s'avèrent être les plus discriminantes possible.

$$a_k = \sum_{j=1}^p \alpha_{jk} (X_j - \bar{X}_j) = \tilde{X} \cdot \alpha_k \text{ avec :}$$

$$\tilde{X} = \begin{pmatrix} X_1^1 - \bar{X}_1 & \dots & X_p^1 - \bar{X}_p \\ \vdots & & \vdots \\ X_1^n - \bar{X}_1 & \dots & X_p^n - \bar{X}_p \end{pmatrix} \text{ et } \alpha_k = \begin{pmatrix} \alpha_{1p} \\ \vdots \\ \alpha_{pk} \end{pmatrix}$$

On a alors,  $\text{var}(a_k) = {}^t\alpha_k \cdot V \cdot \alpha_k$ , qui se décompose en une variance intra-classe et une variance inter-classe de la façon suivante :  ${}^t\alpha_k \cdot V \cdot \alpha_k = {}^t\alpha_k \cdot W \cdot \alpha_k + {}^t\alpha_k \cdot B \cdot \alpha_k$ .

L'analyse discriminante consiste à déterminer la combinaison linéaire  $a$  qui permet, dans un même temps, de séparer au mieux les différentes classes tout en définissant des classes homogènes. Le problème revient donc à déterminer  $\alpha \in \mathbb{R}^p$  permettant de maximiser le rapport  $R(a) = \frac{{}^t\alpha \cdot B \cdot \alpha}{{}^t\alpha \cdot V \cdot \alpha}$  sous la contrainte  ${}^t\alpha \cdot V \cdot \alpha = 1$ .

La dérivée du quotient, par rapport aux différentes composantes de  $\alpha$  doit être nulle :

$$2 \cdot ({}^t\alpha \cdot V \cdot \alpha) \cdot B \cdot \alpha - 2 \cdot ({}^t\alpha \cdot B \cdot \alpha) \cdot V \cdot \alpha = 0.$$

On obtient alors :

$$B \cdot \alpha = \frac{{}^t\alpha \cdot B \cdot \alpha}{{}^t\alpha \cdot V \cdot \alpha} \cdot V \cdot \alpha = \lambda \cdot V \cdot \alpha \text{ et } V^{-1} \cdot B \cdot \alpha = \lambda \cdot \alpha$$

#### • Cas de l'analyse discriminante binaire

Dans le cadre d'une analyse discriminante binaire, la variable discriminante ne possède que 2 modalités. Il n'existe donc qu'un critère discriminant  $(x - 1)$ , noté  $a$  et un facteur discriminant noté  $\alpha$ . Nous considérons alors  $g_1$  et  $g_2$  les centres de gravité de 2 groupes. Nous pouvons alors noter :

$$\alpha = V^{-1}(g_1 - g_2) \text{ et } a = \tilde{X} \cdot \alpha$$

$$B = \sum_{k=1}^2 P_k (g_k - g)^t (g_k - g)$$

Précédemment, nous avons déterminé que  $g = \sum_{k=1}^2 P_k \cdot g_k = \frac{n_1}{n} \cdot g_1 + \frac{n_2}{n} \cdot g_2$ .

A partir de cette équation, nous pouvons déterminer

$$g_1 - g = \frac{n_2}{n} \cdot (g_1 - g_2) \text{ et } g_2 - g = \frac{n_1}{n} \cdot (g_2 - g_1).$$

Nous obtenons  $B = \frac{n_1 \cdot n_2}{n^2} \cdot (g_1 - g_2) \cdot t(g_1 - g_2)$ .

Comme le vecteur propre  $\alpha$  associé à  $V^{-1} \cdot B$  est défini par  $\alpha = V^{-1} \cdot (g_1 - g_2)$  et que la valeur propre  $\lambda$  est définie par  $V^{-1} \cdot B \cdot \alpha = \lambda \cdot \alpha$ , on en déduit que :

$$\lambda = \frac{n_1 \cdot n_2}{n^2} \cdot t(g_1 - g_2) \cdot V^{-1} \cdot (g_1 - g_2)$$

## Résultats

- Etape 1 : Aucune variable dans le modèle

Variable	R-Square	F Value	Pr > F
AGE CP	0.0138	2803.31	<.0001
ANCIENNETE VEHICULE	0.0131	2650.52	<.0001
PRESENCE GARAGE	0.0204	4156.89	<.0001
ANCIENNETE PERMIS CP	0.0153	3107.94	<.0001
ZONE CIRCULATION	0.0081	1630.52	<.0001
FORMULE DE CONDUITE	0.0002	37.98	<.0001
ANTECEDENTS SINISTRES	0.0199	4063.05	<.0001
SEXE CONDUCTEUR PRINCIPAL	0.0013	266.41	<.0001
ANCIENNETE CONTRAT	0.0274	5629.13	<.0001
FRACTIONNEMENT	0.0077	1543.59	<.0001
REDUCTION "AGRICOLE"	0.0037	739.21	<.0001
SINISTRALITE	0.0002	33.89	<.0001
NOMBRE DE CONDUCTEURS COMPL.	0.0066	1335.97	<.0001
FORMULE DE GARANTIES	0.0290	5965.92	<.0001
EVOLUTION DE LA PRIME	0.0029	581.58	<.0001
GROUPE VEHICULE	0.0014	288.58	<.0001
MISE EN DEMEURE	0.0110	2232.04	<.0001
CRM	0.0336	6961.17	<.0001
CLASSE VEHICULE	0.0002	45.70	<.0001
FORFAIT KILOMETRIQUE	0.0004	89.18	<.0001
FRANCHISE VOL	0.0137	2779.64	<.0001
FRANCHISE BRIS DE GLACES	0.0237	4849.67	<.0001
DEROGATION TARIFAIRE	0.0042	848.47	<.0001
MULTI-EQUIPEMENT CLIENT	0.0070	1412.22	<.0001
VALEUR AGENT	0.0008	156.96	<.0001

La variable CRM est la mieux représentée : elle entre dans le modèle.

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.966365	6961.17	1	199998	<.0001
Pillai's Trace	0.033635	6961.17	1	199998	<.0001
Average Squared Canonical Correlation	0.033635				

- Etape 2 : Variable CRM intégrée au modèle

Variable	Partial R-Square	F Value	Pr > F
NIVEAU DE GARANTIES	0.0190	3870.02	<.0001
CRM	0.0237	4855.14	<.0001

Variable	R-Square	F Value	Pr > F	Tolérance
AGE CP	0.0000	6.64	0.0100	0.6196
ANCIENNETE VEHICULE	0.0118	2383.47	<.0001	0.9982
PRESENCE GARAGE	0.0115	2317.32	<.0001	0.9524
ANCIENNETE PERMIS CP	0.0001	12.81	0.0003	0.5884
ZONE CIRCULATION	0.0040	806.08	<.0001	0.9760
FORMULE DE CONDUITE	0.0004	76.05	<.0001	0.9683
ANTECEDENTS SINISTRES	0.0042	833.71	<.0001	0.7853
SEXE CONDUCTEUR PRINCIPAL	0.0006	113.51	<.0001	0.9949
ANCIENNETE CONTRAT	0.0117	2366.79	<.0001	0.8691
FRACTIONNEMENT	0.0006	114.03	<.0001	0.8721
REDUCTION "AGRICOLE"	0.0015	295.00	<.0001	0.9839
SINISTRALITE	0.0000	1.11	0.2919	0.9966
NOMBRE DE CONDUCTEURS COMPL.	0.0021	420.02	<.0001	0.9584
FORMULE DE GARANTIES	0.0190	3870.02	<.0001	0.9579
EVOLUTION DE LA PRIME	0.0036	727.24	<.0001	0.9992
GROUPE VEHICULE	0.0008	162.11	<.0001	0.9970
MISE EN DEMEURE	0.0065	1303.97	<.0001	0.9787
CLASSE VEHICULE	0.0000	0.37	0.5412	0.9944
FORFAIT KILOMETRIQUE	0.0000	0.02	0.8882	0.9864
FRANCHISE VOL	0.0085	1709.33	<.0001	0.9773
FRANCHISE BRIS DE GLACES	0.0148	2996.33	<.0001	0.9596
DEROGATION TARIFAIRE	0.0014	277.76	<.0001	0.9753
MULTI-EQUIPEMENT CLIENT	0.0019	382.47	<.0001	0.9477
VALEUR AGENT	0.0004	87.53	<.0001	0.9983

La variable "formule de garanties" est la mieux représentée : elle entre dans le modèle.

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.948020	5482.93	2	199997	<.0001
Pillai's Trace	0.051980	5482.93	2	199997	<.0001
Average Squared Canonical Correlation	0.051980				

## Annexe n°5 - Régression logistique

### Théorie de la régression logistique

#### • Principe général

Lorsque la variable dépendante est qualitative, elle n'admet pas de valeur numérique naturelle et le modèle normal n'est pas approprié. La régression logistique consiste à étudier une variable qualitative bivariable  $Y$  à partir de  $s$  variables explicatives et quantitatives  $X_i$ .

La variable aléatoire qualitative est caractérisée par les probabilités des différentes valeurs qu'elle peut prendre. Le modèle de régression logistique a pour objectif d'évaluer l'influence de différents facteurs sur cette variable et ainsi de relier les probabilités d'observer les différentes modalités de la variable  $Y$  aux variables explicatives.

#### • Théorie

La régression logistique binaire fait partie de la catégorie des modèles linéaires généralisés. Un modèle linéaire généralisé est défini par une fonction de lien  $g$ , mettant en relation la variable à expliquer  $Y$  et les  $p$  variables prédictives définies par la matrice  $X = (X_1, \dots, X_p)$ .

*Notations :*

- $Y$ , la variable cible binaire, pouvant prendre comme valeurs 1 lorsque l'événement se réalise c'est-à-dire lorsque le contrat est résilié et 0 dans le cas contraire.
- $p$  variables explicatives, binaires, continues ou qualitatives, notées  $(X_1, \dots, X_p)$  et  $X$  le vecteur associé.

L'objectif de la régression logistique consiste à modéliser l'espérance conditionnelle de la variable  $Y$ ,  $E(Y|X = x)$ , c'est-à-dire définir la valeur moyenne prise par  $Y$  pour chacune des valeurs de  $X$ . Dans le cadre de la régression logistique binaire, l'espérance conditionnelle est assimilée à la probabilité que  $Y = 1$ ,  $Y$  suivant une loi de Bernouilli. On définit alors l'espérance conditionnelle par :

$$E(Y|X = x) = P(Y = 1|X = x) = \pi(x)$$

Lorsque  $X$  parvient à discriminer  $Y$ , la relation en  $Y$  et  $X$  est une relation sigmoïdale dont la représentation graphique est présentée ci-dessous :

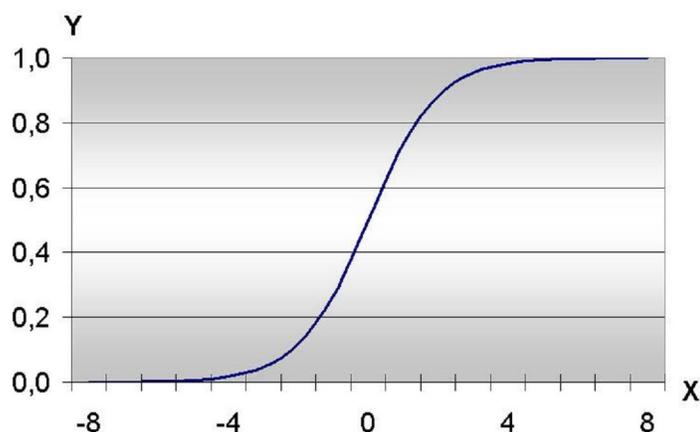


Figure n°28 : représentation de la relation sigmoïdale

Dans le cadre d'une relation sigmoïdale ou logistique, la relation entre  $Y$  et  $X$  peut s'écrire sous la forme :

$$\pi(x) = \frac{e^{\beta_0 + \sum_j \beta_j x_j}}{1 + e^{\beta_0 + \sum_j \beta_j x_j}} \text{ ou encore } \text{Log}\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \sum_{j=1}^p \beta_j x_j$$

où les  $\beta_i$  sont les coefficients du modèle.

Ainsi, la probabilité  $\pi(x)$  peut être définie par la fonction de lien  $f$ , appelée fonction logit :

$$f(t) = \text{Log}\left(\frac{t}{1-t}\right)$$

$\text{Log}$  représentant la fonction logarithme.

L'estimation de  $\beta = (\beta_1, \dots, \beta_p)$  est réalisée à partir de la méthode du maximum de vraisemblance, l'objectif étant de fournir une estimation des paramètres permettant de maximiser la probabilité d'obtenir les valeurs réellement observées. Elle nécessite de définir une fonction de vraisemblance notée  $\mathcal{L}(\beta)$ .

On considère les  $n$  observations indépendantes du modèle  $(y_i, x_i)$  (le vecteur  $x_i$  étant défini par  $(x_{i1}, \dots, x_{ip})$ ). Les  $y_i$  suivent chacun une loi de Bernouilli. La fonction de vraisemblance  $\mathcal{L}(\beta, y_i)$  pour une observation donnée  $i$  se note :

$$\mathcal{L}(\beta, y_i) = \pi(x_i)^{y_i} \cdot (1 - \pi(x_i))^{1-y_i}$$

Pour l'ensemble des observations, on obtient, par conséquent, la fonction de vraisemblance :

$$\mathcal{L}(\beta, y) = \prod_{i=1}^n \mathcal{L}(\beta, y_i) = \prod_{i=1}^n \pi(x_i)^{y_i} \cdot (1 - \pi(x_i))^{1-y_i}$$

Il convient alors de rechercher la valeur  $\hat{\beta}$ , estimateur de  $\beta$ , permettant de maximiser le logarithme de la vraisemblance :

$$\text{Log}(\mathcal{L}(\beta, y)) = \sum_{i=1}^n \left[ y_i \cdot \text{Log} \frac{\pi(x_i)}{1 - \pi(x_i)} + \text{Log}(1 - \pi(x_i)) \right]$$

Cette équation dépend des probabilités  $\pi(x_i)$  inconnues, qui dépendent à leur tour des coefficients  $\beta_j$  également inconnus. Afin de déterminer une estimation des  $\beta_j$ , il convient de dériver l'expression  $\text{Log}(\mathcal{L}(\beta, y_i))$  en fonction des  $\beta_j$  par rapport à  $j$ . Cela nous amène à un système de deux équations dont la résolution ne peut se faire numériquement qu'au moyen d'une méthode itérative : il s'agit de la procédure des moindres carrés pondérés.

### Analyse du modèle des coefficients

Dans le modèle de régression apparaissent deux sortes de coefficients  $\beta_0$  et  $\beta_j$  appelés respectivement constante ou ordonnée à l'origine et pente ou coefficient de régression. La valeur numérique de ces coefficients n'ont pas d'interprétation directe. Seul leur signe permet de savoir si la probabilité de réussite est une fonction croissante ou non de la variable  $X$ .

Le coefficient  $\hat{\beta}_j$  sert à définir le rapport de chance ou *odds ratio* (*OR*). En effet, on peut montrer que :

$$\hat{\beta}_j = f(x+1) - f(x) = \text{Log} \left( \frac{\frac{\hat{\pi}_{x+1}}{1 - \hat{\pi}_{x+1}}}{\frac{\hat{\pi}_x}{1 - \hat{\pi}_x}} \right)$$

cette dernière quantité étant souvent utilisée pour définir le *odds ratio* :

$$\hat{OR} = \exp(\hat{\beta}_j)$$

*Exemple d'interprétation du OR :*

Soit  $OR = 1,014$ . Cette valeur signifie que si on augmente la valeur, alors la chance de voir la probabilité augmenter s'accroît de 1,4%.

### • Tests de significativité du modèle

Le test de significativité consiste à déterminer si les coefficients  $\beta_i$  sont significativement liés à la variable  $Y$  que l'on cherche à analyser. Il revient donc à vérifier que chaque variable  $X_i$  peut figurer parmi les variables explicatives du modèle.

Soit l'hypothèse  $H_0$  définie par :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Cette hypothèse est testée en utilisant le rapport des maximums de vraisemblance  $G$  :

$$G = -2 \text{Log} \left( \frac{\mathcal{L}(\hat{\beta}_{H_0}, y)}{\mathcal{L}(\hat{\beta}, y)} \right)$$

$\hat{\beta}_{H_0}$  correspond à l'estimateur des  $\beta_i$  sous l'hypothèse  $H_0$  et  $\hat{\beta}$  l'estimateur sous l'hypothèse alternative.

Sous l'hypothèse  $H_0$ ,  $G$  suit une loi du  $\chi^2$  à  $p$  degrés de liberté. Le modèle est jugé significatif si l'hypothèse " $\forall i \in [1, p], \beta_i = 0$ " est rejeté ce qui se traduit par :

$$P = P(\chi^2(p) \geq G) \leq \text{seuil de significativité}$$

### • Tests de significativité des variables du modèle

Lors de l'analyse des résultats de la régression logistique, il convient de tester que chaque variable explicative  $X_i$  retenue est significativement liée à la variable  $Y$  et par conséquent d'analyser la pertinence de conserver  $X_i$  comme variable explicative dans le modèle.

*Notations :*

- Soit  $\mathcal{L}$ , la fonction de vraisemblance,
- $p$ , le nombre de paramètres ajustés,
- $n$ , le nombre d'observations du modèle.

Trois méthodes peuvent être utilisées pour évaluer la significativité de la liaison entre la variable dépendante et chaque variable explicative :

- La méthode des rapports de vraisemblance,
- le test de Wald,
- le test de Score.

Le rapport de vraisemblance est défini par la fonction suivante :

$$G_i = -2 \log \left( \frac{\mathcal{L}(\hat{X}_i)}{\mathcal{L}(X_i)} \right).$$

où  $\mathcal{L}(\hat{X}_i)$  représente la vraisemblance sans la variable  $X_i$  et  $\mathcal{L}(X_i)$  la vraisemblance avec la variable  $X_i$ .

Sous l'hypothèse  $\beta_i = 0$ , la statistique  $G_i$  suit une loi du  $\chi^2$  à 1 degré de liberté. La variable  $X_i$  est jugée significative si l'hypothèse  $\beta_i = 0$  est rejetée c'est à dire si :

$$P(\chi^2(1) \geq G) \leq \text{seuil de significativité.}$$

On détermine l'apport d'une variable explicative  $X_i$  par la statistique de Wald :

$$W = \frac{\hat{\beta}_i}{\sigma(\hat{\beta}_i)}$$

où  $\sigma(\hat{\beta}_i)$  représente l'écart-type de  $\hat{\beta}_i$ , déterminé à partir d'approximations.

Cette statistique suit une loi normale réduite, sous l'hypothèse  $\hat{\beta}_i = 0$ . On considère que la variable  $X_i$  est significative, lorsque l'hypothèse  $\hat{\beta}_i = 0$  est rejetée c'est-à-dire lorsque :

$$P = P(|\mathcal{N}(0, 1)| \geq |W|) \leq \text{seuil de significativité}$$

*Remarque :*

Du fait de l'approximation de l'écart-type de  $\hat{\beta}_i$ , il est souvent préférable d'utiliser d'autres méthodes pour analyser la significativité des variables, comme le test du rapport de vraisemblance, défini précédemment.

Le test du Score est défini à partir de la statistique :

$$SCORE = \left[ \frac{\delta \mathcal{L}}{\delta \beta} \right]'_{\hat{\beta}_{H_0}} \left[ 1(\hat{\beta}_{H_0}) \right]^{-1} \left[ \frac{\delta \mathcal{L}}{\delta \beta} \right]_{\hat{\beta}_{H_0}}$$

où  $\hat{\beta}_{H_0}$  correspond au vecteur des paramètres estimés avec la contrainte que le paramètre associé à la variable  $X_i$  en particulier soit nul ( $\hat{\beta}_i = 0$ ) et  $\mathcal{L}$  représente la vraisemblance de l'échantillon.

Sous l'hypothèse  $\hat{\beta}_i = 0$ , la statistique du score suit une loi du  $\chi^2$  à 1 degré de liberté.

### • Adéquation du modèle

Le modèle peut s'avérer significatif, compte tenu des statistiques précédemment citées, mais ne pas être correctement ajusté. Il convient donc d'utiliser des outils de validation supplémentaires

permettant d'analyser la qualité du modèle et d'étudier les résidus de ce modèle. Dans le cas de la modélisation d'une variable binaire, il est également possible d'étudier la corrélation des rangs et de comparer les prédictions réalisées sur la variable de classe avec les valeurs réelles observées.

Pour mesurer l'efficacité globale du modèle, il existe trois indices :

- $-2\text{Log}(\mathcal{L})$  : cette statistique ne tient pas compte du nombre de variables explicatives ni du nombre d'observations. Elle peut, par conséquent, être biaisée (modèle préféré du fait du nombre important de variables  $X_i$  en comparaison au nombre d'observations du modèle) et les deux autres indices lui sont souvent préférés,
- Le critère d'information d'Akaïké,  $AIC = -2\text{Log}(\mathcal{L}) + 2p$  : lorsque deux modèles possèdent la même valeur pour le premier critère, le modèle ayant le moins de variables explicatives est alors privilégié.
- Le critère de Schwartz,  $SC = -2\text{Log}(\mathcal{L}) + p \cdot \log(n)$  : ce critère pénalise les modèles complexes. Il est préféré au critère d'information d'Akaïké, dans le cadre des modèles sur des populations élevés.

L'analyse des résidus peut s'appuyer sur deux méthodes :

- le résidu de Pearson,
- l'analyse de la déviance.

dans le cadre de l'analyse du résidu de Pearson, le modèle peut s'écrire sous la forme :

$$y_i = \pi_i + \varepsilon_i$$

avec  $\pi_i = P(y_i = 1)$  et la variable  $\varepsilon_i$  correspond au résidu pour l'observation  $i$ .

$\varepsilon_i$  suit une loi de bernoulli :

- $\varepsilon_i = 1 - \pi_i$  avec probabilité  $\pi_i$
- $\varepsilon_i = -\pi_i$  avec probabilité  $1 - \pi_i$ .

Comme  $\varepsilon_i$  suit une loi de bernoulli, nous en déduisons que  $E(\varepsilon_i) = 0$  et  $V(\varepsilon_i) = \pi_i \cdot (1 - \pi_i)$ .

Le modèle de régression logistique permettant d'estimer  $P(y_i = 1)$  ( $\hat{\pi}_i$ ), nous pouvons écrire le résidu de Pearson  $r_i$  de la manière suivante :

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

La statistique permettant de juger de la qualité du modèle est définie par :

$$X^2 = \sum_{obs} r_i^2$$

lorsque le modèle ajusté est correct, alors  $X^2$  suit approximativement une loi du  $\chi^2$  à n-p-1 degrés de liberté.

Dans le cadre de l'analyse de la déviance, nous définissons la déviance par

$$D = -2\log\left(\frac{\mathcal{L}_s}{\mathcal{L}_e}\right)$$

où  $\mathcal{L}_s$  représente la vraisemblance du modèle saturé et  $\mathcal{L}_e$  celle du modèle étudié.

On considère qu'un modèle est saturé lorsqu'il modélise parfaitement les données. Le modèle est alors caractérisé par  $\hat{\pi}_i = 1$  si  $y_i = 1$  et  $\hat{\pi}_i = 0$  si  $y_i = 0$ . Si le modèle étudié est ajusté, la déviance suit approximativement une loi du KHI-2 avec n-p-1 degrés de liberté.

Dans le cadre de la modélisation d'une variable de classe binaire, La méthode de corrélation des rangs consiste à étudier les corrélations entre fréquences observées et probabilités prédites. Pour réaliser cette analyse, on utilise le principe des paires concordantes c'est à dire une paire d'individus  $(y_i, x_i)$  et  $(y_j, x_j)$  telle que  $y_i > y_j$ .

On dit que la paire est concordante si  $P(Y = 1|X_{1i}, \dots, X_{pi}) > P(Y = 1|X_{1j}, \dots, X_{pj})$ , c'est-à-dire si la probabilité, estimée en fonction des valeurs des variables explicatives, d'observer  $Y = 1$  pour l'observation  $i$  est supérieure à celle estimée pour l'observation  $j$ .

Si  $P(Y = 1|X_{1i}, \dots, X_{pi}) < P(Y = 1|X_{1j}, \dots, X_{pj})$ , alors la paire est discordante. Si la paire n'est ni concordante, ni discordante, elle est dite liée.

*Notations :*

- $t$ , le nombre de paires avec  $y_i$  différents,
- $n_c$ , le nombre de paires concordantes parmi les  $t$ ,
- $n_d$ , le nombre de paires discordantes parmi les  $t$ .

On en déduit qu'il y a  $t - n_c - n_d$  paires liées.

La qualité du modèle est déterminé à partir des statistiques suivantes :

- $C = n_c + 0.5 \cdot \frac{t - n_c - n_d}{t}$ , variant entre 0 et 1,
- $D = \frac{n_c - n_d}{t}$ , dénommé  $D$  de Sommer et variant -1 et 1,
- $G = \frac{n_c - n_d}{n_c + n_d}$ , appelé  $G$  de Goodman-Kruskal et variant entre -1 et 1
- le  $T$  de Kendall, variant entre -1 et 1 et défini par  $T = \frac{n_c - n_d}{0.5 \cdot n \cdot (n - 1)}$ .

## Résultats

Criterion	Intercept only	Intercept & covariates
AIC	277260.43	259156.05
SC	277270.63	259860.27
-2 Log L	277258.43	259018.05

## Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18240.3736	68	<.0001
Score	17461.0623	68	<.0001
Wald	16044.4725	68	<.0001

## Type 3 Analysis of Effects

	DF	Wald Chi-Square	Pr > ChiSq
AGE COND. PRINCIPAL	3	821.3848	<.0001
ANTECEDENTS SINISTRE	3	654.7889	<.0001
FORMULE DE GARANTIES	4	701.1508	<.0001
NB MISE EN DEMEURE	1	767.6188	<.0001
NIVEAU DE CRM	17	943.2911	<.0001
ANCIENNETE CONTRAT	5	1293.7984	<.0001
FORFAIT KM SOUSCRIT	2	496.5619	<.0001
EVOLUTION COTISATION	23	1576.0201	<.0001
ANCIENNETE VEHICULE	10	486.6845	<.0001

## Association of Predicted Probabilities and Observed Responses

Percent Concordant	66.7	Somers' D	0.338
Percent Discordant	32.9	Gamma	0.339
Percent Tied	0.4	Tau-a	0.169
Pairs	9999977799	c	0.669

## Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
AGE CP ND vs CP<65a	0.533	0.507	0.560
AGE CP<22a vs CP<65a	1.983	1.732	2.271
AGE CP>64a vs CP<65a	1.168	1.133	1.204
ANTECED Acceptable vs Super	0.643	0.610	0.677
ANTECED Bon vs Super	0.624	0.598	0.651
ANTECED Excellent vs Super	0.842	0.823	0.862
FGAR FG1 vs FG4	0.625	0.602	0.649
FGAR FG2 vs FG4	0.729	0.708	0.751
FGAR FG3 vs FG4	0.809	0.781	0.838
FGAR FG5 vs FG4	0.990	0.952	1.029
MEDcont MED vs MED0	0.609	0.588	0.631

Effect	Point Estimate	95% Wald Confidence Limits	
BONUS BM50m3 vs BM50m10	0.954	0.922	0.987
BONUS BM50m6 vs BM50m10	0.992	0.956	1.028
BONUS BM50p16 vs BM50m10	1.166	1.112	1.223
BONUS BM5153 vs BM50m10	0.846	0.802	0.893
BONUS BM5456 vs BM50m10	0.848	0.805	0.893
BONUS BM5759 vs BM50m10	0.825	0.783	0.869
BONUS BM6063 vs BM50m10	0.782	0.743	0.823
BONUS BM6467 vs BM50m10	0.764	0.723	0.808
BONUS BM6871 vs BM50m10	0.739	0.700	0.781
BONUS BM7275 vs BM50m10	0.684	0.647	0.724
BONUS BM7679 vs BM50m10	0.658	0.621	0.697
BONUS BM8085 vs BM50m10	0.691	0.651	0.734
BONUS BM8590 vs BM50m10	0.665	0.620	0.713
BONUS BM9094 vs BM50m10	0.661	0.616	0.709
BONUS BM9599 vs BM50m10	0.542	0.503	0.584
BONUS BM9999 vs BM50m10	0.588	0.525	0.660
ANCONT CONTle15 vs CONTle1	1.544	1.488	1.603
ANCONT CONTle3 vs CONTle1	1.035	1.007	1.063
ANCONT CONTle5 vs CONTle1	1.169	1.134	1.206
ANCONT CONTle8 vs CONTle1	1.303	1.260	1.347
ANCONT CONTle99 vs CONTle1	1.849	1.773	1.928
FSOUS K08 vs K20	0.987	0.966	1.009
FSOUS K99 vs K20	0.728	0.707	0.748
ANCVEH VEH<03a vs VEH<07a	1.121	1.077	1.166
ANCVEH VEH<05a vs VEH<07a	1.001	0.963	1.041
ANCVEH VEH<09a vs VEH<07a	1.008	0.970	1.048
ANCVEH VEH<11a vs VEH<07a	0.977	0.938	1.017
ANCVEH VEH<13a vs VEH<07a	0.923	0.885	0.962
ANCVEH VEH<15a vs VEH<07a	0.850	0.813	0.888
ANCVEH VEH<17a vs VEH<07a	0.782	0.747	0.819
ANCVEH VEH<19a vs VEH<07a	0.759	0.723	0.798
ANCVEH VEH<25a vs VEH<07a	0.681	0.647	0.717
ANCVEH VEH>24a vs VEH<07a	0.637	0.583	0.696

Effect	Point Estimate	95% Wald Confidence Limits	
deltap EcartP091 vs EcartP100	1.068	0.976	1.169
deltap EcartP092 vs EcartP100	0.866	0.781	0.961
deltap EcartP093 vs EcartP100	0.889	0.832	0.949
deltap EcartP094 vs EcartP100	0.833	0.780	0.889
deltap EcartP095 vs EcartP100	0.871	0.819	0.926
deltap EcartP096 vs EcartP100	0.746	0.698	0.797
deltap EcartP097 vs EcartP100	0.778	0.738	0.821
deltap EcartP098 vs EcartP100	0.869	0.831	0.908
deltap EcartP099 vs EcartP100	0.915	0.883	0.948
deltap EcartP101 vs EcartP100	0.750	0.725	0.775
deltap EcartP102 vs EcartP100	0.764	0.737	0.792
deltap EcartP103 vs EcartP100	0.849	0.819	0.880
deltap EcartP104 vs EcartP100	0.808	0.778	0.840
deltap EcartP105 vs EcartP100	0.571	0.541	0.604
deltap EcartP106 vs EcartP100	0.497	0.462	0.536
deltap EcartP107 vs EcartP100	0.455	0.416	0.498
deltap EcartP108 vs EcartP100	0.457	0.414	0.505
deltap EcartP109 vs EcartP100	0.528	0.474	0.587
deltap EcartP110 vs EcartP100	0.472	0.417	0.535
deltap EcartP111 vs EcartP100	0.474	0.414	0.542
deltap EcartP112 vs EcartP100	0.428	0.366	0.500
deltap EcartP113 vs EcartP100	0.356	0.284	0.446

Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.000	99 851	-	100 000	-	49,9	100,0	0,0	50,1	
0.050	99 851	-	100 000	-	49,9	100,0	0,0	50,1	
0.100	99 833	102	100 000	18	50,0	100,0	0,1	50,1	15,0
0.150	99 666	1 087	99 062	185	50,4	99,8	1,1	49,8	14,5
0.200	99 079	3 738	96 411	772	51,4	99,2	3,7	49,3	17,1
0.250	97 587	8 597	91 552	2 264	53,1	97,7	8,6	48,4	20,8
0.300	94 826	15 760	84 389	5 025	55,3	95,0	15,7	47,1	24,2
0.350	90 411	24 692	75 457	9 440	57,6	90,5	24,7	45,5	27,7
0.400	83 555	35 830	64 319	16 296	59,7	83,7	35,8	43,5	31,3
0.450	74 136	48 435	51 714	25 715	61,3	74,2	48,4	41,1	34,7
0.500	62 904	61 181	38 968	36 947	62,0	63,0	61,1	38,3	37,7
0.550	50 495	72 695	27 454	49 356	61,6	50,6	72,6	35,2	40,4
0.600	37 558	82 528	17 621	62 293	60,0	37,6	82,4	31,9	43,0
0.650	24 222	90 762	9 387	75 629	57,5	24,3	90,6	27,9	45,5
0.700	13 210	95 860	4 289	86 641	54,5	13,2	95,7	24,5	47,5
0.750	5 848	98 657	1 492	94 003	52,3	5,9	98,5	20,3	48,8
0.800	1 302	99 866	283	98 549	50,6	1,3	99,7	17,9	49,7
0.850	-	100 000	1	99 851	50,1	0,0	100,0	100,0	49,9
0.900	-	100 000	-	99 851	50,1	0,0	100,0		49,9
0.950	-	100 000	-	99 851	50,1	0,0	100,0		49,9
1.000	-	100 000	-	99 851	50,1	0,0	100,0		49,9