

**Mémoire présenté devant l'Université Paris Dauphine  
pour l'obtention du diplôme du Master Actuariat  
et l'admission à l'Institut des Actuaires**

**le 16 octobre 2017**

Par : Margot Jaillard

Titre: L'AP/TP : un indicateur de rentabilité technique pour le pilotage d'une branche  
d'assurance IARD

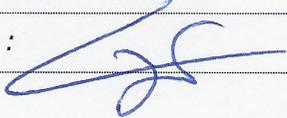
Confidentialité :  NON  OUI (Durée :  1 an  2 ans  5 ans)

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

Membre présent du jury de l'Institut  
des Actuaires :

Signature : Entreprise :

Nom : Allianz France IARD

Signature : 

Directeur de mémoire en entreprise :

Nom : Mme Emilie Roumégoux

Signature : 

Membres présents du jury du Master  
Actuariat de Dauphine :

**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 :

Secrétariat :



Bibliothèque :

Signature du candidat :

M. Jaillard



# Table des matières

Table des matières.....	1
Remerciements.....	3
Résumé.....	4
Summary.....	5
Note de synthèse.....	6
Executive summary.....	11
INTRODUCTION : Contexte économique de l'assurance IARD et enjeux de l'étude.....	16
CHAPITRE 1 : L'AP/TP, indicateur de rentabilité technique pertinent pour le pilotage de portefeuille.....	21
Partie 1 : Principes généraux de tarification en assurance IARD.....	22
a) Les diverses primes manipulées en assurance IARD.....	22
b) Le modèle collectif de risque.....	23
c) Le chargement de la prime pure.....	25
d) Les modulations tarifaires.....	26
Partie 2 : Détermination des différentes primes du portefeuille automobile de particuliers d'Allianz France	28
a) Détermination des primes à l'affaire nouvelle.....	28
b) Détermination des primes lors d'une échéance portefeuille.....	31
c) Détermination des primes lors d'un avenant.....	34
Partie 3 : Intérêt de l'AP/TP pour la mesure de rentabilité technique et prise de responsabilité de cet indicateur dans la gestion dynamique des contrats.....	37
a) L'AP/TP indicateur de pilotage technique complémentaire aux indicateurs classiques.....	37
b) Multiples utilisations de l'AP/TP dans le pilotage de portefeuille.....	39
CHAPITRE 2 : Analyse de la rentabilité d'un portefeuille automobile de particuliers au travers de l'AP/TP.....	42
Partie 1 : Présentation de la base et établissement du périmètre.....	43
a) Choix de la base de données utilisée.....	43
b) Présentation du support des variables descriptives.....	43
Partie 2 : Etude de la distribution et statistiques des AP/TP du portefeuille étudié.....	45
Partie 3 : Etude de la discrimination des variables sur l'AP/TP.....	49
a) Analyse uni-variée : influence des variables sur l'AP/TP.....	49
b) Analyse quantitative des liaisons entre variables.....	56
c) Test de significativité de l'analyse de la variance.....	61
d) Analyse multi-variées sur l'AP/TP et confortation des résultats uni-variés.....	61

CHAPITRE 3 : Modélisation de l'indicateur AP/TP par les Modèles Linéaires Généralisés .....	68
Partie 1 : Quelques notions théoriques utiles à la modélisation.....	69
a) Rappels sur certaines lois usuelles.....	69
b) Méthodes d'estimation des paramètres .....	70
c) Tests d'adéquations aux lois usuelles.....	72
d) Théorie sur les Modèles Linéaires Généralisés .....	73
Partie 2 : Application des Modèles Linéaires Généralisés à la modélisation de l'AP/TP .....	77
a) Sélection du modèle d'AP/TP .....	77
b) Implémentation du modèle .....	82
c) Validation du modèle.....	83
CHAPITRE 4 : Etude portée sur les AP/TP extrêmes .....	89
Partie 1 : Détermination du seuil des extrêmes par la théorie des valeurs extrêmes .....	90
a) La théorie des valeurs extrêmes.....	90
b) Application à la sélection du seuil des AP/TP extrêmes .....	92
Partie 2 : Caractérisation des extrêmes .....	93
a) Caractérisation en termes de primes .....	93
b) Caractérisation en termes de variables descriptives.....	94
Partie 3 : Détection de profils associés à une rentabilité extrême .....	95
a) Utilisation d'une régression logistique pour la détection de contrats extrêmes .....	96
b) Utilisation d'un arbre de décision pour la détection de contrats extrêmes .....	100
CONCLUSION.....	105
Références bibliographiques.....	106
Table des illustrations.....	107
Annexes.....	109
Annexe 1: Analyses uni-variées de l'AP/TP selon les variables.....	109
Annexe 2: Triangle de corrélations inter-variables.....	123
Annexe 3: Coordonnées et contribution des modalités de l'ACM sur les 10 premiers axes.....	124
Annexe 4: Tests d'adéquation de la loi empirique des AP/TP .....	126
Annexe 5: Résultats de la modélisation des AP/TP en loi unique.....	128
Annexe 6: Résultat de la régression logistique pour la détection des contrats associés à un AP/TP extrême .....	129
Annexe 7: Résultat graphique de l'arbre de classification pour la détection des contrats associés à un AP/TP extrême .....	130

# Remerciements

Je tiens tout particulièrement à remercier Emilie Roumégoux, ma tutrice au sein d'Allianz France, pour son encadrement tout au long de cette expérience, pour la richesse de ses conseils et pour la confiance qu'elle a pu m'accorder. J'ai vraiment apprécié travailler sous sa responsabilité.

Je remercie l'ensemble de la Direction Technique IARD et plus particulièrement mes collègues de l'équipe Business Steering pour leur accueil et pour leur bonne humeur quotidienne ayant contribué à mon épanouissement durant ce stage.

Je souhaite remercier Houssein, Médéric, Nicolas, Pierre et Yann pour leurs relectures attentives de ce rapport ainsi que pour la pertinence de leurs remarques. Un grand merci également à ceux qui ont pu m'apporter leurs riches connaissances lors de la réalisation de ce projet, notamment à l'équipe Pricing.

Je remercie enfin Vincent Rivoirard, mon tuteur académique de l'Université Paris Dauphine, pour la relecture de ce mémoire et pour ses commentaires avisés.

# Résumé

L'AP/TP est un indicateur de rentabilité technique récemment développé au sein de la Direction Technique IARD d'Allianz France et qui mesure, police à police, l'écart entre la cotisation payée par l'assuré et le coût probable de ses sinistres, ce dernier étant basé sur la prime technique estimée par des modélisations actuarielles tarifaires. Cet indicateur vient compléter le panel déjà existant des indicateurs classiques visant à mesurer la rentabilité, notamment les ratios sinistres à primes (S/P) ou les ratios combinés (CoR).

Le but de ce mémoire est de mieux appréhender ce tout nouvel indicateur en s'intéressant plus particulièrement au marché automobile quatre roues de particuliers (plus de 1.7 millions de contrats en portefeuille à fin 2016).

Pour cela, nous commencerons par présenter de manière synthétique les principes généraux de la tarification et du pilotage dynamique d'un portefeuille d'assurance IARD afin d'affiner la compréhension de cet indicateur et de mieux comprendre la rentabilité qu'il vise précisément à mesurer ainsi que la manière dont il peut être utilisé dans la gestion de l'activité d'assurance.

Nous analyserons ensuite cet indicateur sur l'ensemble du portefeuille automobile étudié à partir de méthodes statistiques descriptives et de modélisation de type GLM pour repérer certains profils particuliers présentant une sur ou sous-rentabilité.

Enfin, nous focaliserons l'analyse sur les profils les plus rentables afin d'identifier, via d'autres approches statistiques, de possibles caractéristiques pouvant enrichir notre compréhension du business automobile.

**Mots-clés :** Assurance IARD, Assurance automobile, Rentabilité technique, Indicateur police à police, Pilotage de portefeuille, Modèles linéaires généralisés, Analyse des Correspondances Multiples, Etudes des extrêmes, Scoring

# Summary

The AP/TP is a technical profitability indicator recently developed within the P&C Technical Department of Allianz France, which measures the ratio between the premium paid by the policyholder and the probable cost of his claims, the latter being based on the technical premium estimated by actuarial pricing models. This indicator refines the already existing panel of conventional indicators available to measure profitability, including loss ratios (S/P) or combined ratios (CoR).

The aim of this thesis is to better understand this brand-new indicator, with a particular focus on the retail motor insurance market (over 1.7 million policies in the insurance portfolio at the end of 2016).

To that end, we will first summarize the general principles of pricing and dynamic steering of a P&C portfolio in order to improve the understanding of this indicator and to better understand the profitability that it aims to measure and how it can be used in the management of the insurance business.

We will then analyze this indicator on the automobile portfolio using descriptive statistical methods and GLM modeling in order to identify particular profiles with over or under-profitability.

Finally, we will focus the analysis on the most profitable profiles in order to identify, thanks to other statistical approaches, possible characteristics that can enrich our understanding of the motor business.

**Key-words:** P&C Insurance, Motor Insurance, Technical profitability, Policy by policy indicator, Portfolio steering, Generalized Linear Model, Multiple Correspondence Analysis, Study of the extremes, Scoring

# Note de synthèse

Le contexte de concurrence vive, de mise en place de la loi Hamon et de taux financiers historiquement bas sur le marché de l'assurance IARD pousse les assureurs IARD à optimiser d'autant plus leur rentabilité technique. Ainsi, pour piloter la profitabilité de ses produits, la Direction Technique IARD d'Allianz France a souhaité compléter le panel des indicateurs de rentabilité classiques déjà existant en développant un nouvel indicateur, appelé AP/TP, qui permet de mesurer la rentabilité à l'échelle du contrat. Il est défini par :

$$AP/TP = \frac{\text{Actual Price}}{\text{Technical Price}}$$

, où l'Actual Price, ou Tarif Réel en français, correspond à la valeur de la cotisation payée par l'assuré, tandis que le Technical Price, ou Tarif Technique, correspond au coût du risque assuré, incluant les frais afférant à l'activité d'assurance (frais généraux, coûts de commission, coûts de la réassurance...).

Un contrat sera alors sur-tarifé dès lors que le montant engagé par l'assuré sera supérieur à la charge attendue de l'assureur sur ce même contrat, c'est à dire dès lors que l'AP/TP sera supérieur à 1. Lorsque l'indicateur sera inférieur à 1, le contrat sera au contraire sous-tarifé. Un contrat sera d'autant plus rentable que son AP/TP sera supérieur à 1 et d'autant moins rentable qu'il sera inférieur à 1.

Ce mémoire se base sur le portefeuille le plus important dont la Direction Technique IARD d'Allianz France a la charge et sur lequel l'utilisation de l'AP/TP est la plus développée, à savoir le portefeuille automobile quatre roues de particuliers distribué par le biais d'agents généraux (plus d'1.7 millions de polices vues à fin 2016).

Cette étude vise à mieux comprendre la rentabilité prise en compte par l'AP/TP ainsi que la pertinence de celui-ci dans le pilotage d'un portefeuille IARD, permettant ainsi de conforter la Direction Technique IARD d'Allianz France dans l'utilisation et la généralisation de cet indicateur.

La première partie est consacrée à la présentation de l'AP/TP de manière théorique ainsi qu'à la valeur ajoutée que confère cet indicateur par rapport aux indicateurs classiques utilisés par les compagnies d'assurance IARD, lui accordant alors dès à présent de multiples utilisations dans la gestion d'un portefeuille. Dans la deuxième partie, un certain nombre d'analyses descriptives de l'AP/TP sur le portefeuille automobile étudié sont réalisées, faisant ressortir de premières influences et discriminations des variables descriptives d'un contrat sur la rentabilité de celui-ci. Ces analyses sont basées sur des critères tarifaires et sur des modulations tarifaires étant donné l'influence que ces variables ont sur les valeurs du Tarif Technique et du Tarif Réel. La troisième partie a pour objet de confirmer les résultats obtenus dans la partie précédente en construisant un modèle prédictif de type GLM sur l'AP/TP. Enfin, dans la quatrième partie, un focus sera réalisé sur les extrêmes, profils se dégageant de l'ensemble du portefeuille, afin de tenter d'identifier les contrats associés à une telle rentabilité et ainsi de les favoriser lors de la gestion de portefeuille.

## L'AP/TP : un indicateur pertinent pour le pilotage d'un portefeuille IARD

L'AP/TP correspondant au rapport du Tarif Réel sur le Tarif Technique, il dépend fortement des différentes primes intervenant dans le calcul de l'AP/TP ainsi que de leurs processus de revalorisation respectifs. Lors de la souscription, le Tarif Réel se distingue du Tarif Technique en prenant en compte les modulations commerciales qui sont accordées au client. L'écart de ce Tarif Réel au Tarif Technique à la souscription est ainsi explicable majoritairement par les modulations commerciales qui sont appliquées à l'assuré. De ce fait, nous porterons un fort intérêt aux modulations tarifaires lors de l'analyse de caractères discriminants sur la valeur prise par

L'AP/TP. Néanmoins, lors des échéances le Tarif Réel et le Tarif Technique suivent des processus de revalorisation distincts : le Tarif Technique est revalorisé chaque année selon un taux global sur le portefeuille afin de tenir compte de l'inflation et des exigences de rentabilité sur le portefeuille, tandis que le Tarif Réel est revalorisé de manière plus fine, police à police, faisant intervenir un mécanisme d'une certaine complexité et tendant à faire augmenter le Tarif Réel plus vite que le Tarif Technique. Ceci justifiera plus tard la forte importance de l'ancienneté du contrat sur la discrimination de l'AP/TP. De plus, les taux de revalorisation du Tarif Réel à échéance dépendant de la formule souscrite et de la sinistralité passée, ces deux critères devraient également avoir un impact sur les écarts de rentabilité technique.

L'AP/TP est un indicateur pertinent dans la mesure de rentabilité puisque celui-ci combine un certain nombre d'aspects des indicateurs classiques tout en ayant une réelle valeur ajoutée.

	AP/TP	S/P vision dossier/dossier	S/P vision ultime	Ratio Combiné
Maille de reporting habituelle	Police à police ou sur des sous-segments	Sous-segment	Sous-segment	Sous-segment
Vision à l'ultime	Oui : Via la calibration du Tarif Technique (TP)	Non : Prise en compte uniquement du dossier/dossier*	Oui : Via la prise en compte des provisions pour sinistres (IBNR**) et pour recours	
Prise en compte des coûts de commission et frais généraux		Non	Non	Oui
Prise en compte du coût du capital		Non	Non	Non
Sinistralité prise en compte	Prospective : Espérance de la sinistralité du contrat basé sur les modèles de Tarif Technique	Rétrospective : Sinistralité observée (règlement + dossier/dossier) sur le segment	Rétrospective : Sinistralité observée (règlement + dossier/dossier) sur le segment et provisions pour sinistres (IBNR) et pour recours	
Modélisation actuarielle	Oui : Modélisation du TP	Non	Oui : Modélisation des provisions (IBNR)	

\* provision dossier/dossier : montant forfaitaire doté pour couvrir le sinistre selon les premières observations

\*\* IBNR (Incurred But Not Reported): Provision pour différence entre la charge dossier/dossier et charge ultime de la sinistralité est imée actuariellement

*Comparaison des indicateurs de rentabilité classique avec l'AP/TP*

L'AP/TP a un certain nombre d'avantages qui lui sont propres et qui lui permettent ainsi de compléter l'information fournie par les indicateurs de rentabilité classiques. En effet, l'AP/TP a la particularité que la sinistralité prise en compte dans cet indicateur est basée sur l'espérance du coût des sinistres, via la modélisation du Tarif Technique, et non sur l'observation de celle-ci comme dans les indicateurs classiques. La sinistralité prise en compte dans l'AP/TP ne présente donc pas l'aléa de l'observation comme avec le S/P et le Ratio Combiné lorsqu'il est étudié police à police, permettant alors d'observer l'AP/TP de manière individuelle. La prise en compte de la sinistralité par une espérance permet également de rendre l'AP/TP prospectif. L'AP/TP peut notamment être calculé dès l'entrée en portefeuille.

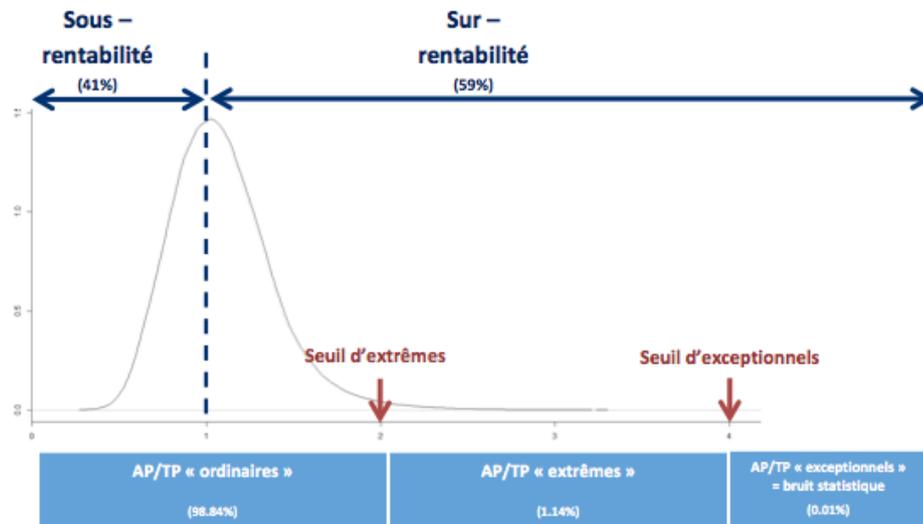
La pertinence de l'AP/TP lui a permis de prendre une place importante dans le pilotage de portefeuille. Cet indicateur intervient désormais notamment lors de revalorisations tarifaires à échéances et lors de la surveillance de portefeuille.

Il est tout de même important de relever que l'AP/TP est un indicateur qui repose fortement sur la qualité du Tarif Technique. Or un Tarif Technique ne pourra en pratique jamais être en parfaite adéquation avec le risque assuré pour un contrat donné et sera toujours perfectible.

## Etude de rentabilité du portefeuille automobile étudié à travers l'AP/TP

La base qui sert de support pour cette étude contient l'ensemble des polices en portefeuille à fin décembre 2016. Les bases à fin décembre 2014 et fin décembre 2015 seront également utilisées exceptionnellement pour tester la robustesse des modèles.

Tout d'abord, une étude est portée sur la distribution de l'AP/TP afin de faire ressortir une image de la rentabilité du portefeuille.



On constate une certaine asymétrie de la distribution des AP/TP, avec la présence d'une queue de distribution à droite. La distribution des AP/TP fait apparaître trois segments :

- Les AP/TP « ordinaires », caractérisés par des valeurs d'AP/TP inférieures à 2 et représentant l'essentiel du portefeuille.
- Les AP/TP « extrêmes », caractérisés par des valeurs d'AP/TP comprises entre 2 et 4 et qui constituent la queue de contrats sur-rentables à droite.
- Les AP/TP « exceptionnels », caractérisés par des valeurs d'AP/TP supérieures à 4. Ces contrats ayant des valeurs d'AP/TP non explicables seront assimilés à du bruit statistique et ne seront pas pris en compte dans la suite du mémoire.

Le choix des seuils d'AP/TP égaux à 2 et 4 a tout d'abord été réalisé avec une logique économique. En effet, par exemple les contrats caractérisés par des AP/TP supérieurs ou égaux à 2 sont les assurés payant une cotisation au moins deux fois supérieure au risque qu'ils assurent. Les seuils seront dans la suite du mémoire revus et déterminés par des méthodes mathématiques.

Ensuite, diverses méthodes statistiques d'analyse descriptive sont réalisées sur l'AP/TP, permettant de souligner les facteurs ayant une influence sur le niveau de rentabilité mesuré par notre indicateur. Celles-ci consistent en :

- une analyse uni-variée représentant les valeurs d'AP/TP en fonction des différents critères observés
- une étude des corrélations des différentes variables entre elles, permettant de prendre en compte les effets de corrélations
- une étude des corrélations des variables considérées avec l'AP/TP
- une analyse de la variance (ANOVA), permettant de confirmer l'importance de la discrimination d'une variable sur l'AP/TP à un seuil de 5%
- une Analyse en Composantes Multiples (ACM) permettant de confirmer les résultats obtenus précédemment en tenant compte des corrélations inter-variables

L'ensemble de ces méthodes permet de montrer l'influence notable de variables tarifaires telles que la note de sécurité, de l'ancienneté du contrat, des antécédents sinistres ainsi que de la formule souscrite. Les modulations tarifaires sont également révélées fortement discriminantes sur l'AP/TP.

Par la suite, un modèle de prévision faisant appel aux modèles linéaires généralisés est réalisé. Seules les variables permettant d'expliquer au mieux la valeur de l'AP/TP ont été conservées. Le modèle intègre tout d'abord l'ancienneté du contrat, étant donné l'importance qu'il a sur les revalorisations. Un certain nombre de modalités relatives aux modulations tarifaires sont également prises en compte dans le modèle. On note enfin l'importance de la formule souscrite sur le niveau de rentabilité mesuré par l'AP/TP, les formules les moins couvrantes étant les plus rentables. Ce modèle s'avère robuste et stable dans le temps, après application du modèle sur les années 2014 et 2015. Les plus importants écarts d'estimation sont cependant réalisés au niveau des AP/TP extrêmes, la queue lourde est sous-estimée.

### **Etude portée sur les extrêmes**

Une meilleure compréhension de ces contrats caractérisés par une sur-rentabilité nous permettrait de les identifier à la souscription ainsi que de piloter le portefeuille de manière à favoriser ces profils au détriment de ceux caractérisés par un AP/TP faible.

Dans cette partie, un nouveau seuil d'extrêmes à 1,5 a été fixé à partir d'outils de la théorie des valeurs extrêmes. Les AP/TP supérieurs à 4 sont toujours retirés de notre base d'étude.

Une première étude des extrêmes a été réalisée. Celle-ci compare les répartitions des critères tarifaires et des modulations tarifaires observées sur les extrêmes par rapport à celles observées sur l'ensemble du portefeuille. Cette étude nous permet de noter notamment la sur-représentation, par rapport à l'ensemble du portefeuille, des coefficients bonus-malus supérieur à 100, des contrats anciens, des formules moins couvrantes et des véhicules de classe de prix élevées ou classés « Hors Classe ».

Une fois avoir mieux compris de quoi étaient composés les contrats associés à une rentabilité extrême, des outils permettant d'accompagner l'aide à la décision dans des problématiques de détection de contrats sur-profitables ont été mis en place, faisant appel à deux modèles de prévision statistique binaire : les régressions logistiques et les arbres de classification binaires. Ils permettent alors de pouvoir capter un maximum de contrats caractérisés par une valeur d'AP/TP extrême. On retrouve les variables concernant les modulations tarifaires, l'ancienneté du contrat et les formules souscrites parmi les variables segmentant les contrats ayant un AP/TP ordinaire des contrats ayant un AP/TP extrême pour les deux modèles construits. Des actions commerciales ciblées pourraient ainsi être réalisées en faveur de ces contrats détectés comme apportant une certaine profitabilité à l'entreprise, au moment de leur souscription et permettant ainsi de capter ces bons assurés du marché. Il faudrait cependant au préalable tester ces modèles sur la durée afin de s'assurer de leur fiabilité dans le temps.

Une piste d'amélioration aurait été d'utiliser des modèles de Machine Learning plus élaborés de type forêt aléatoire ou gradient boosting pour la détection d'AP/TP extrêmes.

Ce mémoire propose une première étude approfondie du portefeuille automobile le plus important d'Allianz France à travers ce tout nouvel indicateur qu'est l'AP/TP. Ses caractères individuels et prospectifs nous ont ainsi confortés dans l'idée qu'il s'agissait d'un indicateur pertinent pour le pilotage de portefeuille d'assurance IARD.

Un certain nombre de critères se sont avérés être favorables à la sur-rentabilité du contrat ou au contraire à sa sous-rentabilité. Nous avons notamment souligné l'importance de l'ancienneté du contrat, de la formule souscrite ou encore des modulations tarifaires.

Cette étude nous encourage davantage à développer l'utilisation de l'AP/TP sur d'autres périmètres tels que l'assurance habitation.

## Executive summary

The context of keen competition, implementation of the Hamon law and historically low financial rates in the P&C insurance market has pushed P&C insurers to optimize their technical profitability. In order to manage the profitability of its products, the P&C Technical Department of Allianz France has completed the panel of existing profitability indicators by developing a new indicator called AP/TP, which enables to measure profitability at the scale of an insurance contract. It is defined by:

$$AP/TP = \frac{\text{Actual Price}}{\text{Technical Price}}$$

, where the Actual Price corresponds to the value of the premium paid by the insured, while the Technical Price corresponds to the actuarially estimated cost of the insured risk, including the costs related to the insurance activity (overhead costs, commission costs, reinsurance costs, etc.).

A contract will then be overrated if the amount invested by the insured is greater than the expected charge of the insurer on this same contract, that is to say if the AP/TP is greater than 1. When the indicator is lower than 1, the contract will then be underrated. A contract will be all the more profitable that its AP/TP will be greater than 1 and the less profitable that it will be lower than 1.

This thesis is based on the largest portfolio the Allianz France P&C Technical Department is in charge of and on which the use of the AP/TP is the most developed, namely the tied agents four-wheel portfolio (over 1.7 million policies at the end 2016).

This study aims to better understand the profitability taken into account by the AP/TP as well as its relevance in the management of a P&C insurance portfolio, thus will strengthen the Allianz France P&C Technical Department conviction in the use and generalization of this indicator.

The first part of the thesis is devoted to the theoretical presentation of the AP/TP, as well as the added value of this indicator compared to the conventional indicators used in P&C insurance companies, giving it already multiple uses in the management of a portfolio. In the second part, a number of descriptive analyzes of the studied motor portfolio are carried out, showing first influences and discriminations of the descriptive variables of a contract on the profitability of the latter. These analyzes are based on pricing criteria and pricing modulations given the influence that these variables have on the values of the Technical Price and the Actual Price. The third part aims to confirm the results obtained in the previous section by constructing a predictive GLM model on the AP/TP indicator. Finally, in the fourth part, a focus will be realized on the highest values of AP/TP, profiles emerging from the portfolio as a whole, in order to try to identify the contracts associated with such a profitability and thus to favor them in portfolio management.

### **The AP/TP: a relevant indicator for the steering of a P&C portfolio**

The AP/TP corresponding to the ratio of the Actual Price over the Technical Price, depends strongly on these different premiums and their respective tariff reevaluation processes. At underwriting, the Actual Price is distinguished from Technical Price by taking into account the commercial modulations that are granted to the customer. The difference between the Actual Price and the Technical Price at underwriting is thus mainly explained by the commercial modulations applied to the insured. As a result, we will be very interested in pricing modulations when analyzing discriminating characteristics of the values taken by the AP/TP. Nevertheless, during the portfolio annual terms the Actual Price and the Technical Price follow separate reevaluation processes: the Technical Price is revalued every year according to an overall rate on the portfolio

in order to take inflation rate and profitability requirements into account, while the Actual Price is reevaluated more finely, policy by policy, involving a mechanism of a certain complexity and tending to increase the Actual Price faster than the Technical Price. This will undoubtedly give high priority to the age of the contract on discrimination in the AP/TP. In addition, the rate of reevaluation of Actual Price at term depends on the insurance formula subscribed and the claim history, these two criteria should also have an impact on the technical profitability differences.

The AP/TP is a relevant indicator in the measure of profitability since it combines a number of aspects of the conventional indicators while having real added value.

	AP/TP	Case Loss Ratio	Ultimate vision Loss Ratio	Combined Ratio
Usual Reporting reference	Policy by policy or sub-segment	Sub-segment	Sub-segment	Sub-segment
Ultimate vision	Yes : By calibration of the Technical Price (TP)	No : Only case reserve* supported	Yes: By taking into account IBNR** and subrogation	
Taking account of commission and overhead costs		No	No	Yes
Taking account of costs of capital		No	No	No
Claim history taken into account	Forward-looking: Expectancy of contract claims based on Technical Price Models	Backward-looking Observed claim performance (payment + case reserve) on the segment	Backward-looking: Observed claim performance (payment + case reserve) on the segment and IBNR and subrogation	
Actuarial modeling	Yes: Modeling of the Technical Price	No	Yes: Modeling of the IBNR reserves	

\* Case reserve: lump sum provided to cover the incident according to initial observations

\*\* IBNR (Incurred But Not Reported): Difference between the case reserve and ultimate actuarially estimated loss

*Comparison of traditional profitability indicators with AP / TP*

The AP/TP has a number of distinct advantages which enables to complement the information provided by the traditional profitability indicators. Indeed, the AP/TP has the specificity that the cost of the losses taken into account in this indicator is based on the expectation, through the modeling of the Technical Price, and not on the observation of the cost of losses as in the conventional indicators. The loss taken into account in the AP/TP thus does not present the hazard of the observation as with the Loss Ratio and the Combined Ratio when it is studied policy by policy, and then allows observing the AP/TP on an individual basis. The estimation of the loss also makes the AP/TP forward-looking. In particular, the AP/TP can be calculated as soon as it enters the portfolio.

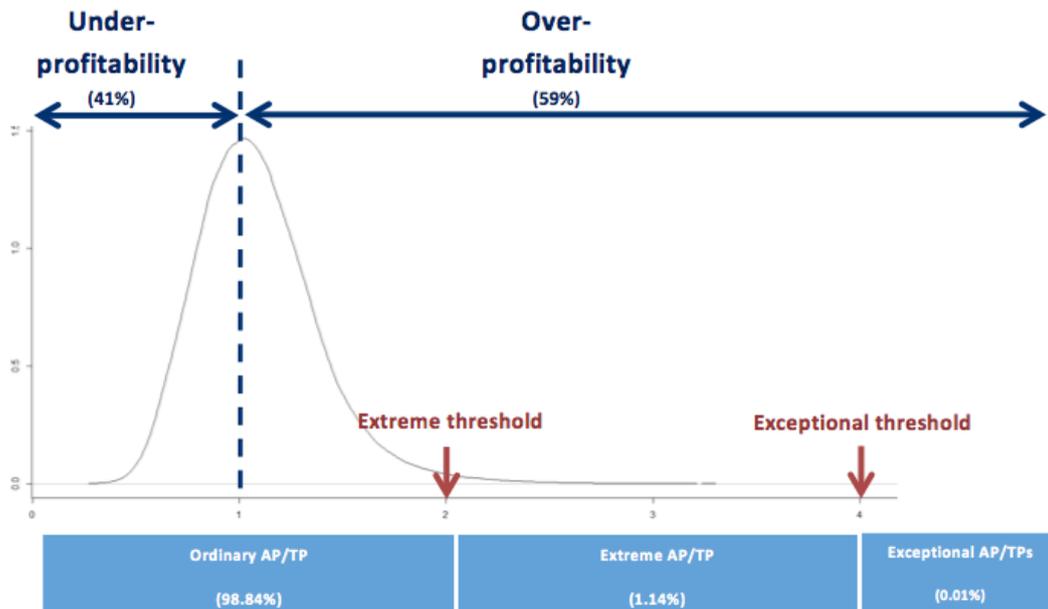
The relevance of the AP/TP allowed it to take a major role in the steering of the portfolio. This indicator is now being used in particular for pricing reevaluation at annual terms and during portfolio monitoring.

It is important to note that the AP/TP is an indicator that relies heavily on the quality of the Technical Price. However, in practice, a Technical Price can never perfectly match the risk insured for a given contract and will always be perfectible.

## Profitability study of the automotive portfolio studied through the AP/TP

The base which supports this study contains all the policies in portfolio at the end of December 2016. The bases at the end of December 2014 and at the end of December 2015 will also be used exceptionally to test the robustness of the models.

First, a study is conducted on the distribution of the AP/TP in order to give a picture of the portfolio's profitability.



We notice that there is some asymmetry in the distribution of AP/TP, with the presence of a right heavy-tailed distribution. The distribution of AP/TPs reveals three segments:

- "Ordinary" AP/TPs, characterized by AP/TP values lower than 2 and representing the bulk of the policies of the portfolio
- "Extreme" AP/TPs, characterized by AP/TP values between 2 and 4 and which constitute the right tail of contracts that are over-profitable.
- "Exceptional" AP/TPs, characterized by AP/TP values greater than 4. These contracts with non-explicable AP/TP values will be treated as statistical noise and will not then be taken into account for the rest of the thesis.

The choice of AP/TP thresholds equal to 2 and 4 was first done with an economic logic. Indeed, for instance contracts characterized by an AP/TP greater than or equal to 2 are insured persons paying a contribution that represent at least twice the risk they insure. The thresholds will later in the thesis be redetermined by mathematical methods.

Then, various statistical methods of descriptive analysis are carried out on the AP/TP, highlighting factors that influence the level of profitability measured by our indicator. These consist of:

- a one-way analysis representing the AP/TP values according to the different observed criteria
- a study of the correlations between the different variables, allowing to take into account the effects of correlations
- a study of the correlations of the variables with the AP/TP
- an analysis of variance (ANOVA), confirming the importance of the discrimination of a variable on the AP/TP at a 5% threshold

- a Multiple Correspondence Analysis to confirm the results obtained previously taking into account inter-variable correlations

All these methods make it possible to show the significant influence of pricing variables such as the safety note, the age of the contract, the claims history and the subscribed formula. The pricing modulations are also strongly discriminating on the AP/TP.

Subsequently, a forecasting model using generalized linear models is carried out. Only the variables that best explain the value of the AP/TP have been kept. The model integrates the age of the contract, given the importance it has on portfolio monitoring and reevaluation. A number of modalities relating to pricing modulations are also taken into account in the model. Finally, we note the importance of the formula subscribed in the level of profitability measured by the AP/TP, the least covering formula being the most profitable. This model proves to be robust and stable over time, after applying the model to the portfolios of 2014 and 2015. However, the largest deviations of the estimates are made on the extreme AP/TP, the right tail is under-considered in the GLM model.

### **Study of the extremes**

A better understanding of these contracts characterized by over-profitability would enable us to identify them at underwriting as well as to steer the portfolio in such a way as to favor these profiles to the detriment of those characterized by a low AP/TP.

In this part, a new threshold of extremes of 1.5 has been established using extreme value theory tools. The values of AP/TPs those are greater than 4 are still removed from our study base.

A first study of the extremes is carried out. It consists in comparing the distribution of pricing criteria and pricing modulations observed at the extremes compared to those observed across the whole portfolio. This study allows us to note in particular the over-representation of bonus-malus coefficients that are greater than 100, of old contracts, of less covering formulas and of high-priced or "Out of Class" vehicles.

Once we have a better understanding of what the contracts associated with extreme profitability are, decision support tools that help in the detection of over-profitable contracts were set up, using two models of binary statistical prediction: logistic regressions and binary classification trees. They allow us to be able to capture a maximum of contracts characterized by an extreme AP/TP value. Variables relating to pricing modulations, age of the contract and subscribed formulas are revealed to be the most appropriated to segment contracts with an ordinary AP/TP and contracts with extreme AP/TP for both of the models. Targeted commercial actions could thus be carried out in favor of these contracts, which are detected as bringing a certain profitability to the company, at the time of their underwriting and thus making it possible to collect these good insured from the market. However, it would be necessary to test these models over time to ensure their reliability over time.

An idea of improvement would be to use more elaborate Machine Learning models, such as random forest or gradient boosting in order to detect extreme values of AP/TPs.

This thesis was the subject of a first in-depth study of Allianz France's largest automotive portfolio through this newly developed AP/TP indicator. Its individual and prospective characteristics have reinforced us in the idea that it was a relevant indicator for the steering of a P&C insurance portfolio.

A certain number of criteria proved to be favorable to the over-profitability of the contract or, on the contrary, to its under-profitability. In particular, we highlighted the importance of the age of the contract, of the subscribed formula, or of pricing modulations.

This study encourages us to develop the use of AP/TP on other perimeters such as home insurance.

# INTRODUCTION : Contexte économique de l'assurance IARD et enjeux de l'étude

On mesure couramment la rentabilité d'une activité d'assurance de biens et de responsabilité à partir du Ratio Combiné ou Combined ratio (CoR) en anglais. Celui-ci permet de juger de la rentabilité globale d'un segment en exprimant le rapport entre l'ensemble des charges afférentes à l'activité d'assurance (indemnisation des sinistres et frais de fonctionnement), et les encaissements (cotisations payées par les assurés).

Lorsque l'on considère cet indicateur sur les dix dernières années, on obtient le graphique suivant :

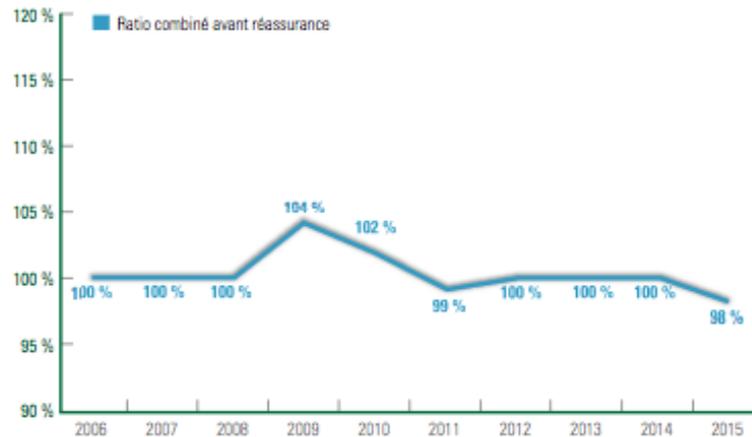


Figure 1: Evolution du ratio combiné en assurance de biens et de responsabilité (Source: FFA)

On constate que, au-delà de la sinistralité exceptionnelle qui a eu lieu en 2009 et 2010, la rentabilité globale du marché de l'assurance de biens et de responsabilité en France est globalement très proche de 100%. La rentabilité est ainsi très serrée car une telle valeur de l'indicateur correspond au cas où il y a un équilibre total entre le montant des recettes et le montant des coûts.

Nous porterons dans ce mémoire un intérêt particulier à l'assurance automobile de particuliers, qui représente la part de marché la plus importante de l'assurance de biens et responsabilité :

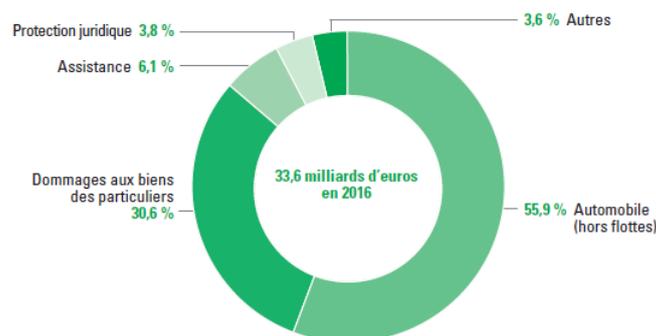


Figure 2: Répartition des cotisations des particuliers par branches en 2016 (Source : FFA)

A l'échelle de la ligne de business qu'est l'assurance automobile, la pression de la rentabilité est d'autant plus importante que sur l'ensemble du marché de l'assurance de biens et de responsabilité puisque l'on constate un Ratio Combiné toujours supérieur à 100%.

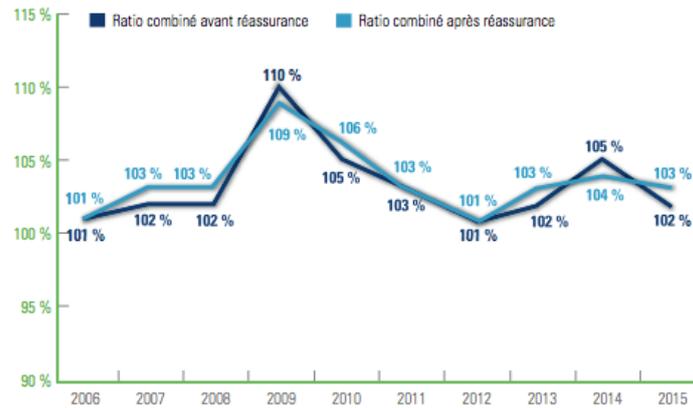


Figure 3: Evolution du ratio combiné en assurance automobile (Source : ACPR, FFA)

Au-delà de cette rentabilité serrée de l'activité de l'assurance de biens et de responsabilité, il existe deux autres effets d'actualité venant intensifier la nécessité de maîtrise de la rentabilité du marché.

Tout d'abord, les taux financiers sont aujourd'hui historiquement bas. Depuis 2010, les rendements financiers des compagnies d'assurance non-vie sont en baisse.

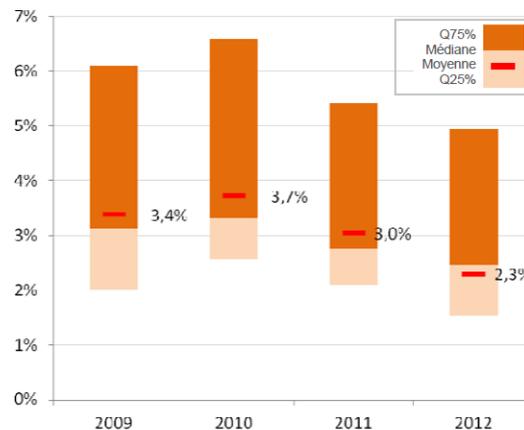


Figure 4: Evolution des produits financiers nets de charges sur le total des placements (Source : ACPR)

Dès lors que les compagnies d'assurance de biens et de responsabilité ne peuvent plus compenser leurs éventuelles pertes techniques par les résultats de leurs produits financiers, celles-ci se doivent plus encore d'assurer une réelle excellence technique.

Par ailleurs, le passage de la loi Hamon en 2014 sur les marchés automobile et habitation, accordant la flexibilité aux assurés de résilier leur contrat à tout moment une fois avoir acquis un an d'ancienneté, a provoqué une stimulation de la concurrence déjà vive sur ces marchés. Les taux de résiliations et d'affaires nouvelles ont par conséquent été fortement impactés depuis cette date.

En assurance automobile, les taux de résiliations et d'affaires nouvelles ont connu une croissance de respectivement +0.9 point et +1.1 points entre 2014 et 2015 alors qu'on observait jusque-là une quasi-stabilité du turn-over.

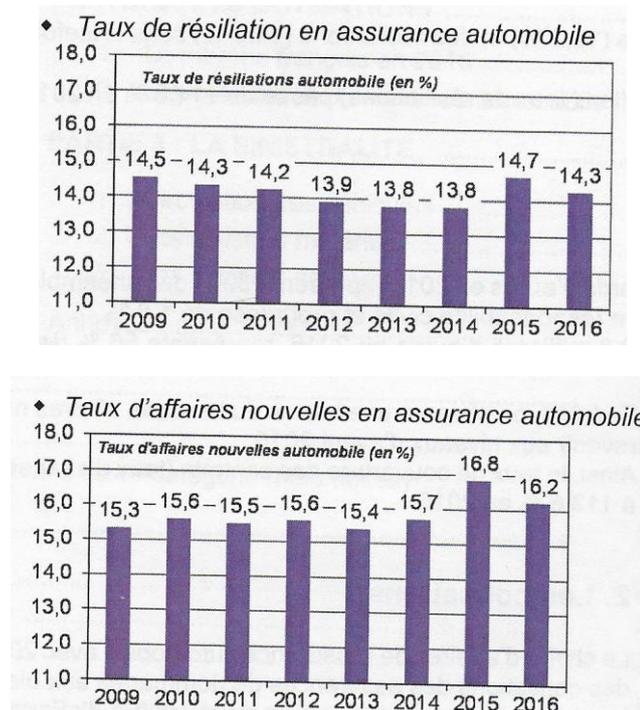


Figure 5: Evolutions du taux de résiliations et du taux d'affaires nouvelles sur le marché de l'assurance automobile (Source : FFA)

La forte rotation des portefeuilles induite par l'entrée en vigueur de la loi Hamon accentue ainsi deux défis pour les compagnies d'assurance IARD :

- les compagnies d'assurance IARD doivent se montrer attractives et commercialement intéressantes face à la concurrence afin d'assurer la capture de nouvelles affaires. Pour cela, elles sont amenées à accorder aux nouveaux assurés un certain nombre de réductions commerciales, engendrant une diminution de la prime. Ces opérations commerciales doivent donc être réalisées de manière contrôlée afin de pouvoir assurer malgré tout une maîtrise de la rentabilité.

- les compagnies d'assurance IARD doivent réaliser un réel pilotage du fond de portefeuille. Ce sont, en effet, les polices déjà en portefeuille qui permettent de financer les opérations commerciales accordées aux nouvelles affaires, notamment à travers les revalorisations annuelles des primes. Ces dernières doivent alors être pilotées finement pour cibler en priorité les polices les moins rentables et prévenir les résiliations de celles les plus rentables.

Ce bilan sur l'état du marché de l'assurance de biens et de responsabilité permet de mettre en valeur la nécessité pour les assureurs d'évoluer vers un processus de plus en plus individualisé de l'étude et du pilotage de la rentabilité.

C'est dans ce contexte là qu'Allianz France a développé, au sein de sa Direction Technique IARD, un nouvel indicateur de rentabilité qui s'observe à l'échelle de la police et venant compléter le panel déjà existant des indicateurs de rentabilité classiques (S/P, CoR) qui seront présentés dans la suite de ce mémoire (Partie 3

du Chapitre 1) et qui sont observés de manière plus globale. Ce nouvel indicateur, dénommé AP/TP, est défini par :

$$AP/TP = \frac{\text{Actual Price}}{\text{Technical Price}}$$

, soit par le rapport entre la prime effectivement payée par l'assuré à l'assureur et l'estimation actuarielle du coût réel de son risque.

L'analyse de l'AP/TP permet d'apprécier de la rentabilité technique à l'échelle d'un contrat. Celle-ci est à l'équilibre lorsque l'AP/TP est égal à 1. Supérieur à ce seuil, le contrat est techniquement sur-tarifé pour l'assureur, et au deçà de ce seuil il est sous-tarifé. Plus l'AP/TP est supérieur à 1 et plus la rentabilité technique du contrat considéré est élevée.

Ainsi, le but de ce mémoire est de permettre à la Direction Technique IARD d'Allianz France de mieux appréhender cet indicateur de rentabilité qui a été récemment été développé en interne. Cette étude se basera sur le portefeuille le plus important dont la Direction Technique est en charge, à savoir le portefeuille automobile quatre roues de particuliers. Ce portefeuille compte plus de 1.7 millions de contrats en portefeuille à fin décembre 2016.

Pour ce faire, nous commencerons par présenter de manière synthétique les principes généraux de tarification et d'établissement des primes de ce portefeuille aux différents moments de la vie d'un contrat, permettant d'affiner la compréhension de la construction de cet indicateur et de mieux comprendre la rentabilité qu'il vise précisément à mesurer ainsi que la manière dont il peut être intégré dans la gestion de l'activité d'assurance.

Par la suite, nous analyserons cet indicateur sur l'ensemble de portefeuille automobile considéré à travers diverses méthodes statistiques descriptives et de modélisation de type GLM, pour repérer certains critères particuliers favorables à la sur ou sous-rentabilité du contrat.

Enfin, nous réaliserons une étude plus approfondie sur les profils les plus rentables afin d'identifier, via d'autres approches statistiques, de possibles caractéristiques pouvant enrichir notre compréhension du marché de l'automobile de particuliers.

# CHAPITRE 1 : L'AP/TP, indicateur de rentabilité technique pertinent pour le pilotage de portefeuille

Ce chapitre portera un intérêt particulier à l'indicateur de rentabilité technique qu'est l'AP/TP ainsi qu'à sa pertinence dans le pilotage de portefeuille, justifiant alors sa récente expansion au sein de la Direction Technique IARD d'Allianz France.

Pour ce faire, nous reviendrons tout d'abord sur certains principes de tarification, ceci nous permettant de définir les différentes primes manipulées au sein de la Direction Technique et intervenant dans le calcul de l'AP/TP. Nous nous attarderons par la suite sur l'établissement et la revalorisation des primes dans le cas particulier du portefeuille étudié, l'assurance automobile classique de particuliers. La mise à plat de ces différents mécanismes nous permettra de comprendre comment évolue cet écart au Tarif Technique ou AP/TP. Cela nous permettra également de nous faire une idée des critères qui, par la structure des mécanismes mis en place, auront une influence sur les écarts de rentabilité des contrats. Nous expliquerons enfin en quoi il s'agit d'un outil dont la pertinence est certaine pour la mesure de rentabilité d'un contrat et lors du pilotage de portefeuilles.

## Partie 1 : Principes généraux de tarification en assurance IARD

Pour mieux comprendre les différentes notions de primes utilisées notamment dans le calcul de l'AP/TP, il est nécessaire d'exposer au préalable quelques notions générales d'assurance non vie.

### a) Les diverses primes manipulées en assurance IARD

On compte quatre définitions de primes en assurance IARD. Elles font toutes référence au coût de l'assurance qui a été souscrite, mais intègrent chacune des éléments divers et ont alors chacune une utilité qui leur est propre. Voici un schéma présentant globalement ces différentes primes et ce qu'elles incluent :

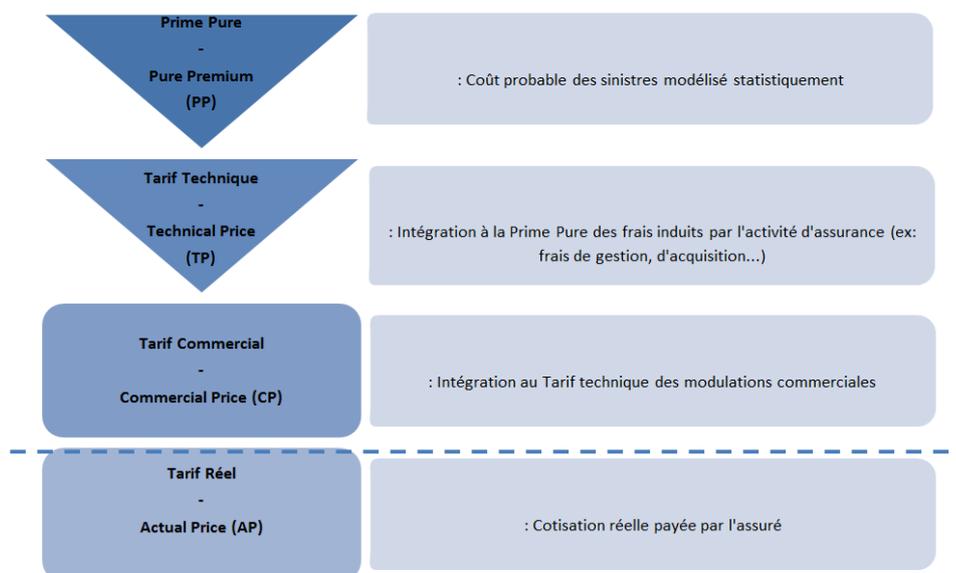


Figure 6: Présentation des différentes primes assurantielles

La Prime Pure est un élément intégré dans le calcul du Tarif Technique. Son montant sera dans ce mémoire peu étudié. On portera un intérêt plus particulier aux trois autres primes : le Tarif Technique, le Tarif Commercial et le Tarif Réel.

Afin d'alléger les notations, on notera parfois par la suite PP pour Prime Pure (Pure Premium en anglais), TP pour Tarif Technique (Technical Price), CP pour Tarif Commercial (Commercial Price) et AP pour Tarif Réel (Actual Price).

## **b) Le modèle collectif de risque**

Les compagnies d'assurance suivent le principe de mutualisation, c'est-à-dire que l'ensemble des primes collectées par l'assureur permettent de couvrir l'intégralité des sinistres subis par les assurés ainsi que les coûts que le fonctionnement de la compagnie requiert durant la période de couverture. C'est pourquoi, lors de la tarification d'un portefeuille, un assureur cherche au préalable à évaluer la charge de sinistre probable d'un contrat, lui permettant ainsi de faire payer à l'assuré une cotisation se rapprochant du risque que celui-ci représente. On en déduira également un certain montant afin de prendre en compte les retours sur investissement générés par les primes placées.

La Prime Pure est, pour chaque assuré, le coût probable des sinistres pour la période d'assurance considérée. C'est la meilleure estimation des coûts de sinistres futurs du contrat.

- *Sinistres attritionnels et sinistres graves*

L'estimation du coût des sinistres se fait sur la base d'historiques de sinistres dont le montant correspond à la charge ultime. En effet, étant donné que le coût final d'un sinistre n'est en général atteint que tardivement, on projette sa valeur finale par des méthodes de provisionnement (Chain Ladder, fréquence X coût moyen, ...).

On distingue les sinistres « attritionnels » des sinistres « graves ». Les sinistres attritionnels constituent l'essentiel de la sinistralité du portefeuille en termes de nombre. Les sinistres graves diffèrent des sinistres attritionnels par leur faible occurrence mais dont les très forts coûts ont un impact beaucoup plus lourd sur la sinistralité globale du portefeuille considéré.

Par manque d'historique des sinistres extrêmes, la modélisation des coûts se fait en général en écrêtant les sinistres, c'est à dire en ne considérant que l'historique des sinistres ayant un montant inférieur à un seuil de grave, lui-même déterminé par des méthodes empiriques. On ajoute ensuite à l'estimation des coûts une certaine charge, en général à travers un coefficient de chargement, afin de prendre en compte les sinistres extrêmes

- *Prime pure décomposable par garantie*

Pour chaque contrat, la Prime Pure est décomposée selon les différentes garanties incluses dans la formule qui a été souscrite. On compte les garanties Responsabilité Civile matérielle et corporelle, Bris de Glace, Vol et Incendie, Dommage Tous Accidents et Assistance. Chacune de ces composantes est estimée à partir de l'historique de sinistres attritionnels et graves du portefeuille et permet d'établir une prime pure propre à chaque garantie. Certaines primes pures dont les garanties ont un historique de sinistralité plus réduit, sont calculées en pourcentage d'autres garanties. C'est le cas de la garantie catastrophes naturelles notamment. Le

coût des sinistres pour une garantie donnée est calculé via un modèle statistique, qui est couramment basé sur un modèle collectif.

- *Le modèle collectif de risque*

Le modèle collectif est le modèle de base de l'assurance automobile. Il permet d'estimer la sinistralité d'un ensemble de contrats, appelé « classe », dont le risque est similaire. Une classe est en général formée par une « case tarifaire », c'est-à-dire l'ensemble des assurés présentant la même combinaison de modalités pour chacun des critères tarifaires considérés. Le nombre et le coût des sinistres ne sont pas rattachés à un contrat de manière individuelle mais globalement à une classe de contrats.

Pour une case tarifaire donnée, soient  $M$  le nombre d'assurés,  $N$  le nombre total de sinistres et  $Y_1, Y_2, \dots, Y_N$  les montants des  $N$  sinistres.

On suppose de plus que :

- $N$  est une variable aléatoire à valeurs entières
- $Y_1, Y_2, \dots, Y_N$  sont des variables aléatoires indépendantes, identiquement distribuées et à valeurs positives
- $N$  et  $Y_i$  sont indépendantes pour tout  $i = 1, 2, \dots, N$

La charge totale des sinistres est donnée par  $S = \sum_{i=1}^N Y_i$ . La prime pure est alors égale à l'espérance de cette charge totale, notée  $E(S)$ .

$$\begin{aligned}
 E(S) &= E(E(S|N)) = E(E(\sum_{i \leq N} Y_i | N)) = E(E(\sum_{i \geq 1} Y_i \cdot \mathbb{1}_{i \leq N} | N)) = E\left(\sum_{i \geq 1} E(Y_i \cdot \mathbb{1}_{i \leq N} | N)\right) \\
 &= E(\sum_{i \geq 1} \mathbb{1}_{i \leq N} E(Y_i | N)) \text{ par } N\text{-mesurabilité de } \mathbb{1}_{i \leq N} \\
 &= E(\sum_{i \geq 1} \mathbb{1}_{i \leq N} E(Y_i)) \text{ par indépendance de } N \text{ avec } Y_1, Y_2, \dots, Y_N \\
 &= E(\sum_{i \geq 1} \mathbb{1}_{i \leq N} E(Y)) \text{ car les } Y_i, i = 1, \dots, N \text{ sont identiquement distribués} \\
 &= E(Y) \cdot E\left(\sum_{i \geq 1} \mathbb{1}_{i \leq N}\right) = E(Y) \cdot E(N)
 \end{aligned}$$

D'où

$E(S) = E(N) \cdot E(Y)$
$\underbrace{\hspace{10em}} \quad \underbrace{\hspace{10em}}$
Fréquence      Coût moyen des sinistres    des sinistres

Pour obtenir la prime pure d'une case tarifaire, on multiplie la fréquence par le coût moyen des sinistres estimés sur cette case. On parle ainsi de modèle collectif « Fréquence \* Coût Moyen ».

### c) Le chargement de la prime pure

Nous avons vu précédemment que la totalité des primes récoltées doit permettre de compenser le risque de sinistralité de l'ensemble du portefeuille selon le principe de mutualisation.

Cependant, des frais supplémentaires déployés par la compagnie d'assurance pour permettre le fonctionnement de son activité vont également être inclus dans le Tarif Technique. Ceux-ci font référence aux frais d'acquisition et de gestion, aux frais induits par la réassurance et aux frais permettant l'atteinte des exigences de rentabilité imposées par les actionnaires.

- *Prise en compte des frais de gestion et d'administration*

Un certain nombre de frais doivent être déployés afin de permettre le bon fonctionnement de la compagnie d'assurance. Il existe différents frais, rattachables soit à l'administration et à la gestion d'un contrat soit à la gestion des sinistres. Ces frais sont de différentes natures : commissions versées principalement aux apporteurs d'affaires (typiquement les agents) ou frais généraux (salaires, IT, location des locaux...)

- *Prise en compte de la réassurance*

De plus, la compagnie se réassure au global afin de pouvoir faire face aux lourdes conséquences que pourrait avoir la survenance de risques extrêmes. Cette réassurance représente un certain coût pour la compagnie et est alors également supportée par l'ensemble des contrats.

- *Prise en compte des retours sur investissements*

Les cotisations payées par les assurés sont, jusqu'à versement des prestations, placées sur le marché financier. Elles permettent alors de dégager des retours sur investissement. Pour tenir compte de cet apport financier permis par les assurés, on « décharge » la prime pure.

- *Prise en compte des exigences de rentabilité imposées par le groupe*

Des contraintes de solvabilité sont imposées aux compagnies d'assurance, actuellement sous le régime de Solvabilité II. Les contrats souscrits nécessitent notamment l'immobilisation d'un certain capital minimum afin de pouvoir faire face aux risques de déviation de la sinistralité. Une certaine rentabilité sur ce capital immobilisé est alors imposée par le groupe Allianz. Le Return on Required Capital (RoRC) est défini par :

$$RoRC = \frac{Resultat\ Net}{Capital\ Requis}$$

La Direction Financière d'Allianz France évalue la rentabilité nécessaire pour chaque branche sur la base de l'indicateur RoRC et permet de définir un coefficient de chargement de la prime pure afin de répondre à ces exigences de rentabilité.

- *De la Prime Pure au Tarif Technique*

Le Tarif Technique permet ainsi de couvrir le risque qu'un contrat représente mais également l'intégralité des frais engendrés par l'activité de la compagnie. Un certain chargement est ainsi ajouté à la prime pure pour intégrer ces coûts supplémentaires au tarif rattaché à l'assuré. Chacun de ces éléments est alors représenté par un coefficient de chargement qui vient majorer la Prime Pure et définissant ainsi le Tarif Technique. Un coefficient viendra enfin minorer la prime pure pour la prise en compte des retours sur investissements.

$$\text{Tarif Technique} = \text{Prime Pure} \cdot \left(1 + \rho_{\substack{\text{frais de gestion} \\ \text{et d'administration}}}\right) \cdot \left(1 + \rho_{\text{frais de réassurance}}\right) \cdot \left(1 - \rho_{\text{investissements}}\right) \cdot \left(1 + \rho_{\text{exigence de rentabilité}}\right)$$

Chargement de la Prime Pure

#### d) Les modulations tarifaires

La Prime Commerciale d'un assuré diffère de la Prime Pure chargée associée à son risque. En effet, des décalages de rentabilité peuvent être temporairement acceptés auprès de certains risques afin de se développer, pour tenir compte de modalités plus ou moins rentables ou pour compenser des pertes sur d'autres critères. Ces modulations créent des modifications de tarif en acceptant des poches de sous-tarification et au contraire sur-tarification du Tarif Commercial par rapport au Tarif Technique. Il existe alors deux types de modulations tarifaires, permettant de moduler à la hausse ou à la baisse la prime de l'assuré. On compte les modulations de type technique et celles de type commercial. L'ensemble de ces modulations tarifaires sont décidées et contrôlées au sein de la Direction Technique.

- *Les modulations techniques*

Les modulations techniques sont en fait directement intégrées au Tarif Technique d'un contrat. Elles visent à favoriser les contrats dont la rentabilité est élevée en leur réduisant le Tarif Technique. Cela permet de corriger des sous-tarifications ou sur-tarifications qui n'ont pas été identifiées par le Tarif Technique.

- *Les modulations commerciales*

Une fois le Tarif Technique défini, on doit y ajouter certaines composantes de stratégie commerciale et marketing permettant ainsi d'animer les ventes. Celles-ci prennent en compte toutes les variations de primes explicables de manière commerciale associées aux produits d'assurance. Ces modulations commerciales permettent alors de passer d'un Tarif Technique, qui reflète le coût du risque assuré, à un tarif de commercialisation.

Voici un tableau expliquant les différentes modulations tarifaires intervenant sur le portefeuille automobile :

<i>Les modulations techniques</i>	
Le protocole	Réduction accordée à l'assuré mais dont la valeur dépend de la rentabilité technique de l'agence dans laquelle il a souscrit son contrat. Cette modulation encourage l'agent à avoir une certaine qualité de souscription puisqu'elle lui permet à terme de proposer des tarifs plus compétitifs à ses clients. Le protocole va donc par exemple favoriser les ventes de formules et options rentables et va veiller au bon respect des règles de souscription lors de la saisie des données (antécédent sinistre, bonus-malus...). Une même personne se rendant dans deux agences Allianz ne se verra pas proposer le même tarif d'une agence à l'autre, la prime dépendant du taux protocole propre à l'agence.
Les codes firmes	Ce sont des réductions qui permettent d'accompagner sur certaines cibles de rentabilité. Ces codes de réduction sont créés et surveillés au sein de la Direction Technique, en forte interaction avec la direction en charge de la distribution des produits. Il existe par exemple des réductions réservées aux détenteurs de véhicules semi-autonomes, aux reprises auprès de bancassureurs ou aux contrats ayant souscrit un contrat en formule Intermédiaire ou Tous Risques (appelée couramment respectivement formule C2 et C3 au sein de la Direction Technique IARD d'Allianz France) et en paiement mensuel de la prime.
<i>Les modulations commerciales</i>	
L'avantage client	Permet de favoriser la multi-possession de contrats toute branche confondue (IARD, Entreprise, Santé, Vie, Banque) et donc d'encourager la fidélisation des clients. Il est accordé dès le deuxième contrat souscrit chez Allianz dans la mesure où un des contrats n'appartient pas à la branche auto. L'avantage client dépend du nombre de contrats détenus et de la valeur des encours de l'assuré.
L'avantage prospect	Réduction accordée à l'assuré en affaire nouvelle uniquement. Il est accordé lorsque l'agent considère que l'assuré peut potentiellement souscrire un autre contrat Allianz d'ici la fin de la première échéance portefeuille. C'est une anticipation de l'avantage client.
Budget Développement *	Accorde des réductions à de potentiels assurés afin de capter des affaires nouvelles (reprise de contrats à la concurrence et la souscription de contrats d'enfants ou conjoints d'assurés) et ne sont pas accordés uniquement lors de la première année du contrat. A partir du lendemain du premier anniversaire du contrat, une part croissante au fil des échéances de la réduction n'est plus rattachée au Tarif Commercial mais uniquement au Tarif Réel à chaque échéance. Ce processus vient alors diminuer le Tarif Commercial sans changer la valeur du Tarif Réel. Nous allons voir plus tard dans ce mémoire que l'indicateur majeur utilisé lors des majorations de primes, appelé ETP et correspondant au rapport du Tarif Réel sur le Tarif Commercial, est alors réduit, ce qui viendra augmenter la revalorisation lors de l'échéance suivante. On parle de « grignotage » du Budget Développement.
Budget Défense *	Permet de défendre les contrats déjà en portefeuille en leur accordant une réduction lors de leur terme ou lors d'un avenant et ainsi que de les dissuader de réaliser une résiliation (en général les plus sur-tarifés).
Réduction Affaire Nouvelle (REDAN)	Permet d'accorder un abattement sur la prime commerciale lors de la première année dans le cas où l'assuré vérifie certaines conditions à la souscription.

Mois gratuits	Des offres ponctuelles permettent aux nouveaux assurés en paiement mensuel d'obtenir des mois d'assurance gratuits. Au maximum, deux mois gratuits seront accordés : un lors de la souscription et un différé d'un an. Les mois gratuits ne viennent cependant pas impacter le Tarif Réel visible dans nos bases portefeuille, les primes en portefeuille étant nettes de mois gratuits.
---------------	---

\*Le budget Défense et le budget Développement sont contingentés par une enveloppe commune. L'agent est alors limité à la distribution d'un certain budget. Le fait que les deux enveloppes soient mises en commun permet aux agents de réaliser un meilleur arbitrage du budget entre les affaires nouvelles et les contrats déjà en portefeuille.

Suite à l'application de ces modulations tarifaires le Tarif Commercial se retrouve finalement décalé du tarif défini sur des bases techniques et de même, le Tarif Réel peut se retrouver décalé du Tarif Technique.

Toute modification de la politique stratégique tarifaire d'Allianz aura donc un impact direct sur le Tarif Commercial et sur le Tarif Réel.

## Partie 2 : Détermination des différentes primes du portefeuille automobile de particuliers d'Allianz France

Dans cette partie, nous allons nous focaliser sur le portefeuille le plus important d'Allianz France IARD, à savoir le portefeuille automobile quatre roues de particuliers distribué par l'intermédiaire d'agents généraux et qui fait l'objet d'une étude particulière dans ce mémoire. Ce portefeuille est communément nommé « 4RP Agents » en interne. Il n'inclut pas les conducteurs novices, qui correspondent aux conducteurs ne pouvant justifier d'une période d'assurance effective de 36 mois, étant donné que ces derniers d'admettent pas la même structure tarifaire que celle qui sera présentée par la suite.

Pour mieux comprendre l'intérêt de l'étude de l'AP/TP et ce que cet indicateur reflète au niveau du portefeuille considéré, il est essentiel d'expliquer au préalable comment sont calculées au sein de la Direction Technique IARD d'Allianz France les différentes primes du produit « 4RP Agents ». Le calcul des différentes primes intervient à l'affaire nouvelle lors de la souscription d'un contrat, à l'échéance principale à chaque anniversaire de la souscription ou à l'avenant lors de la modification du risque assuré par un système de revalorisation des primes de l'année précédente.

### a) Détermination des primes à l'affaire nouvelle

- *Le Tarif Technique : coût assurantiel du risque*

Dû à l'importance des ressources techniques et informatiques que cela impliquerait, le Tarif Technique actuel du portefeuille Allianz considéré ne peut être revu chaque année. C'est ainsi un tarif par garantie datant de quelques années que l'on fait « vieillir » chaque année en lui appliquant des revalorisations successives et

en ajustant quelques relativités du modèle. Pour des raisons de confidentialité, notons x l'année de la modélisation du Tarif Technique de référence.

A l'origine, la Prime Pure de laquelle est issue le Tarif Technique avait été définie par un modèle linéaire généralisé Fréquence \* Coût moyen multiplicatif pour chacune des garanties. Les coefficients obtenus définissant alors la Prime Pure pour la garantie j pour chaque contrat i :

$$PP_i^j = PP_{de\ base}^j \cdot coef_{1(i)}^j \cdot coef_{2(i)}^j \cdot \dots \cdot coef_{n(i)}^j$$

Pour un contrat donné, les primes pures modélisées par garantie sont ensuite additionnées pour former la prime pure globale du contrat.

Comme expliqué dans la partie précédente, un certain chargement doit ensuite être ajouté à la Prime Pure de base de chaque contrat pour prendre en compte les retours sur investissement et les dépenses qui ont été effectuées par la compagnie afin de permettre le déroulement de son activité d'assurance. Les dépenses comprennent à la fois les frais d'acquisition et de gestion, mais également les frais de réassurance et les exigences de rentabilité imposées par le groupe. La Prime Pure chargée définit ainsi le Tarif Technique.

Initialement, lors de la création du Tarif Technique, pour prendre en compte le chargement de la Prime Pure et calibrer le modèle, le Tarif Technique avait été calé, au global sur le portefeuille, sur le Tarif Commercial. On parle de Tarif Technique iso-chiffre d'affaire.

Voici un schéma synthétique et théorique permettant d'expliquer le principe de calage du Tarif Technique sur le Tarif Commercial :

$$coef_{calibration} = \frac{500\ 000}{350\ 000} = 1.4$$

Police	Tarif Commercial	Tarif Technique	Tarif Technique calé
1	200	TP de base x coef(gprsa) x coef(ageCond) ...= 250	1,40 x TP de base x coef(gprsa) x coef(ageCond) ...= 357 €
2	150	TP de base x coef(gprsa) x coef(ageCond) ...= 170	1,40 x TP de base x coef(gprsa) x coef(ageCond) ...= 243€
...	...	...	...
n	...	...	...
Total portefeuille	500 000 €	350 000 €	500 000 €

Figure 7: Explication théorique du principe de calibration du Tarif Technique

Cela revient finalement à modifier le TP de base en posant :

$$TP_{de\ base\ calé\ en\ x} = \left(1 + \rho_{frais\ de\ gestion\ et\ d'administration}\right) \cdot \left(1 + \rho_{frais\ de\ réassurance}\right) \cdot \left(1 - \rho_{investissements}\right) \cdot \left(1 + \rho_{exigence\ de\ rentabilité}\right) \cdot PP_{de\ base\ en\ x} \cdot coef_{calibration}$$

$$TP_{de\ base\ calé\ en\ x} = \underbrace{\left(1 + \rho_{\substack{\text{frais de gestion} \\ \text{et d'administration}}}\right) \cdot \left(1 + \rho_{\substack{\text{frais de} \\ \text{réassurance}}}\right) \cdot (1 - \rho_{\text{investissements}}) \cdot \left(1 + \rho_{\substack{\text{exigence} \\ \text{de rentabilité}}}\right)}_{\text{Chargement = Coefficient de calage}} \cdot PP_{de\ base\ calé\ en\ x}$$

Chargement = Coefficient de calage

A la fin de l'année x, pour un contrat défini par ses modalités de risques, on a :

$$TP_{x(i)} = TP_{de\ base\ calé\ en\ x} \cdot coef_{1(i)} \cdot coef_{2(i)} \cdot \dots \cdot coef_{n(i)}$$

Le Tarif Technique défini en année x est revalorisé chaque année afin de prendre en compte l'inflation et les évolutions de sinistralité. Les coefficients de revalorisation sont chaque année décidée suite à un certain nombre de discussions entre les différentes directions d'Allianz France. Ces taux de revalorisation sont appelés « taux de revalorisation Affaire Nouvelle » car ce sont eux qui permettent de définir le Tarif Technique revalorisé à considérer lors d'une affaire nouvelle à partir du Tarif Technique de l'année précédente. Ce choix est très stratégique car plusieurs effets sont à prendre en compte lors de la détermination du taux de revalorisation, faisant pression à la hausse ou à la baisse sur ce taux. En effet, ils dépendent de la rentabilité souhaitée par les actionnaires, mais également d'un point de vue plus commercial de l'élasticité des assurés au prix de la cotisation afin d'éviter les potentielles résiliations des clients puisque ce taux est également directement celui qui est appliqué lors de la revalorisation du Tarif Commercial pour les contrats en portefeuille.

Pour déterminer le Tarif Technique d'une année donnée, il faut donc le faire vieillir en le multipliant successivement par les coefficients de revalorisation des années écoulées depuis l'année x.

Le Tarif Technique final inclut enfin toutes les modulations tarifaires de type technique qui comprennent le taux protocole de l'agence auprès de laquelle a été souscrit le contrat ainsi que les codes firmes applicables (voir modulations techniques).

- *Le Tarif Commercial : Animation des ventes*

Le Tarif Commercial est le tarif qui est censé prendre en compte à la fois le risque de l'assuré et l'ensemble des modulations commerciales qui lui sont appliquées, venant s'ajouter ou se déduire du Tarif Technique.

Lors de la souscription, ce n'est pas le Tarif Technique qui sert de base pour le Tarif Commercial, mais un autre tarif appelé « Tarif Machine ». Celui-ci est issu d'un Tarif Technique ayant été retravaillé pour répondre aux contraintes commerciales (pas de prise en compte du sexe, le tarif doit être proportionnel au coefficient bonus-malus) ainsi que pour se rapprocher des tarifs proposés par la concurrence, à travers des ajustements appelés parfois « cross-subsidies ».

De même, ce Tarif Machine n'est pas remis à jour chaque année, il faut donc le revaloriser en lui appliquant successivement les coefficients de revalorisation Affaire Nouvelle des années écoulées depuis l'établissement du Tarif Machine.

On applique ensuite au Tarif Machine toutes les modulations tarifaires de type commercial afin de former le Tarif Commercial. On compte notamment la réduction affaire nouvelle (REDAN) qui est une modulation exclusivement réservée aux affaires nouvelles sous certaines conditions. De l'avantage prospect peut également être déployé lorsque l'agent anticipe l'ouverture de nouveaux contrats, de l'avantage client en cas

de multi-possession ou du budget développement afin d'aider la rétention de contrats. Il s'agit du tarif à l'affaire nouvelle, qui est payé l'année de la date de souscription.

- *Le Tarif Réel en affaire nouvelle*

Un troisième tarif est alors défini, faisant référence à la cotisation effectivement payée par l'assuré : c'est le Tarif Réel, aussi bien appelé Actual Price (AP) ou Tarif Portefeuille.

Comme expliqué précédemment, à la souscription, la cotisation versée par l'assuré est donnée par le Tarif Commercial en affaire nouvelle.

### **b) Détermination des primes lors d'une échéance portefeuille**

Les trois tarifs étudiés doivent subir des modifications dans le temps afin de rester en adéquation avec le risque supporté par l'assureur. C'est ainsi que lors de chaque anniversaire du contrat, on doit ajuster les primes afin de prendre en compte à la fois l'inflation et les nouveaux objectifs de résultats financiers de l'année fixés par Allianz France, mais également le fait que le risque de l'assuré a évolué.

- *Mise à jour du Tarif Technique lors de l'échéance portefeuille*

A chaque terme, c'est-à-dire à chaque anniversaire du contrat, le Tarif Technique par garantie est recalculé. Celui-ci doit prendre en compte l'évolution du profil de risque de l'assuré depuis l'année précédente. En effet, le conducteur, le véhicule et le contrat ont ainsi gagné un an et le conducteur peut avoir eu un accident, ayant ainsi une répercussion sur sa sinistralité des 36 derniers mois et sur son coefficient Bonus-Malus. La mise à jour de l'ensemble des coefficients concernés ainsi que la revalorisation du Tarif Technique par le taux de revalorisation affaire nouvelle décidé par la direction pour l'année en question permet ainsi de définir le Tarif Technique après terme par garantie par la formule suivante :

$$TP_{(i)} = TP_{de\ base\ révalorisé} \cdot coef_{1(i)} \cdot coef_{2(i)} \cdot \dots \cdot coef_{n(i)}$$

Le Tarif Technique final inclus enfin les modulations tarifaires de type technique qui comprennent le taux protocole de l'agence auprès de laquelle a été souscrit le contrat ainsi que le code firme potentiellement applicable à l'assuré.

- *Mise à jour du Tarif Commercial lors de l'échéance portefeuille*

Lors de chaque échéance portefeuille, de même que le Tarif Technique, le Tarif Machine doit être mis à jour de manière à inclure dans sa valeur les évolutions de sinistralité et les changements d'ancienneté du contrat, du véhicule et de l'assuré. Il est également revalorisé par le taux de revalorisation affaire nouvelle. On déduit enfin du Tarif Machine l'ensemble des modulations commerciales applicables qui peuvent inclure de l'avantage client en cas de multi-possession et du budget développement. Cependant, on rappelle qu'à partir du lendemain de la première échéance portefeuille, le budget défense qui a été accordé à la souscription est peu à peu grignoté, c'est-à-dire qu'une partie de ce budget n'est plus intégrée dans le Tarif Commercial.

- *Mise à jour du Tarif Réel lors de l'échéance portefeuille*

#### - Enjeux de la détermination du Tarif Réel

La valeur de la cotisation ne peut pas être définie à tout moment par le Tarif Commercial, la refonte du Tarif Technique ou toute modification de la politique stratégique tarifaire d'Allianz engendrerait une variation trop importante de la valeur de la cotisation payée. Un mécanisme tarifaire a alors été mis en place, permettant de gérer les politiques de majorations lors des termes et des avenants pour faire converger sans heurt le portefeuille vers le tarif cible qu'est le Tarif Commercial.

De plus, la détermination de la cotisation à échéance est un élément clé de la vie d'un contrat, puisque celle-ci influe sur la décision de résiliation ou non du client pour l'année à venir. En effet, le choix de la résiliation peut se faire suite à une revalorisation qui lui paraît trop excessive ou par comparaison aux tarifs proposés par la concurrence. Le choix de la cotisation payée doit être d'autant plus optimisé que le client peut désormais résilier avec moins de contraintes depuis le passage de la loi Hamon en 2014. L'assuré est encore plus maître de sa situation puisque cette loi lui permet de mettre un terme à son contrat lorsqu'il le souhaite une fois la première échéance passée.

Par conséquent, il existe une véritable optimisation du comportement client face aux évolutions de tarification, dont un service de la Direction Technique est en charge. Le processus de revalorisation à terme doit permettre d'obtenir un tarif au plus près du risque avec une évolution de cotisation personnalisée tout en restant un processus automatisé.

#### -Processus de revalorisation et mécanisme d'ETP

La détermination des montants des différents tarifs ayant un rôle clé dans le suivi de la rentabilité des produits ainsi que dans la souscription et la fidélisation des clients, divers mécanismes complexes interviennent lors de l'établissement de leur montant. La technique de revalorisation portefeuille s'appuie essentiellement sur un outil de pilotage de portefeuille appelé « Ecart au Tarif Portefeuille » (ETP). Son nom vient du fait qu'il mesure l'écart entre la cotisation payée et le Tarif Commercial. Il est défini par :

$$ETP = \frac{\text{Actual Price}}{\text{Commercial Price}}$$

L'ETP permet principalement de piloter la valeur des primes commerciales avec finesse lors des échéances portefeuille et des avenants car il mesure l'écart entre la prime payée et la prime définie sur des bases concurrentielles et de commercialisation.

Avant chaque échéance, on recalcule l'ETP à partir du Tarif Réel de l'année qui vient de s'écouler sur la valeur du Tarif Commercial revalorisé au taux de revalorisation affaire nouvelle de la nouvelle année. On a ainsi :

$$ETP_{\text{avant échéance}} = \frac{\text{Actual Price}_{\text{année } N-1}}{\text{Commercial Price}_{\text{année } N-1 \text{ revalorisé}}}$$

Le Tarif Commercial étant considéré comme le tarif de référence puisqu'il est notamment calé sur la concurrence, l'ETP vise à faire converger le Tarif Réel vers le Tarif Commercial lors des revalorisations à terme. En effet après chaque échéance, on va appliquer au Tarif Réel un certain taux de revalorisation, fonction de la formule souscrite ainsi que de la tranche d'ETP avant terme et de la sinistralité des 36 derniers mois. Pour chaque formule, les taux de revalorisation sont donnés par un tableau à double entrée de la forme suivante :

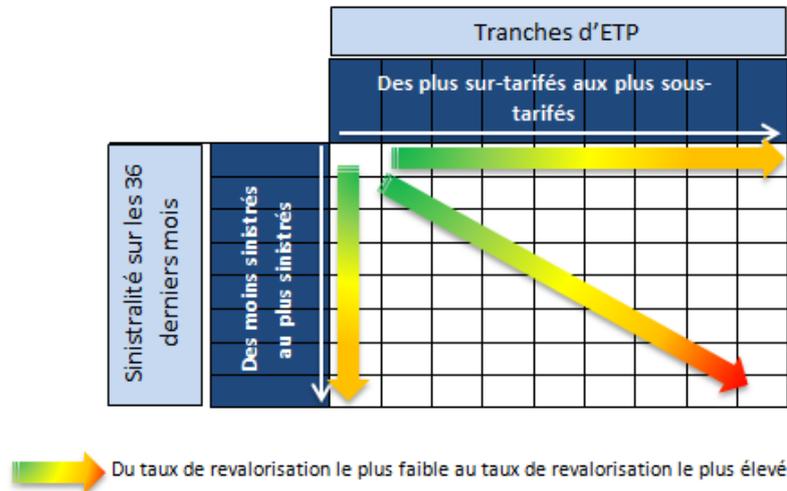


Figure 8: Grille de revalorisation du Tarif Réel à partir de l'ETP

Lorsque l'on constate une valeur d'ETP avant échéance inférieure à 1, c'est-à-dire que la cotisation payée est inférieure au Tarif Commercial, l'écart entre le taux de revalorisation du Tarif Réel et celui du Tarif Commercial va permettre d'augmenter l'ETP à chaque échéance. Par application successive de ce mécanisme au moment des échéances portefeuille, l'ETP converge progressivement vers 1 et donc le Tarif Réel va se rapprocher du Tarif Commercial.

- Revalorisations additionnelles et AP/TP :

Bien que le Tarif Commercial soit la principale référence du Tarif Réel car celui-ci est calé au marché, le Tarif Technique, qui représente le coût réel du risque assuré, doit également être pris en compte lors des revalorisations. Pour ce faire, des revalorisations supplémentaires sur le Tarif Réel vont être appliquées lors de l'échéance dépendant de l'AP/TP. Celles-ci seront présentées par la suite lors de la mise en valeur des multiples utilisations de l'AP/TP dans le processus de gestion du portefeuille automobile.

Pour conclure, la revalorisation à échéance est essentiellement pilotée par l'ETP qui vise à faire converger le Tarif Réel vers le Tarif Commercial. Le Tarif Commercial représente la cible vers laquelle doit converger le montant des cotisations de chaque police. Le fait que le Tarif Réel payé par l'assuré ne suive pas directement le Tarif Commercial, mais en dépendant par un mécanisme complexe, permet de réviser le Tarif Affaire Nouvelle sans pour autant impacter brutalement le portefeuille. En cas d'établissement d'un nouveau Tarif Technique, on ferait ainsi converger progressivement le Tarif Réel vers le nouveau Tarif Commercial par le mécanisme d'ETP. De plus, des revalorisations additionnelles interviennent lors des échéances portefeuille, permettant ainsi d'intégrer l'écart entre le montant de la cotisation payée et le coût réel assuré. La cotisation payée est par conséquent ramenée progressivement à un niveau suffisamment proche du risque assuré. Cette prise en

compte de l'écart au tarif Technique et donc l'intégration de l'AP/TP au processus de revalorisation lors des échéances est ainsi primordiale pour le pilotage de la rentabilité d'un portefeuille.

### c) Détermination des primes lors d'un avenant

En cas d'avenant, le risque que représente l'assuré est modifié suite au changement de véhicule assuré ou suite au changement de la formule souscrite par exemple. Les trois tarifs étudiés doivent chacun subir des modifications avec le temps afin de rester en adéquation avec le risque supporté par l'assureur, tout en restant attrayant pour l'assuré afin d'éviter sa résiliation. La réalisation d'un avenant engendrera ainsi directement un impact sur la valeur de l'AP/TP.

- *Mise à jour du Tarif Technique et du Tarif Commercial lors de l'avenant*

Lors d'un avenant, l'ensemble des coefficients concernés doivent être modifiés, définissant alors un nouveau Tarif Technique ainsi qu'un nouveau Tarif Commercial.

- *Mise à jour du Tarif Réel lors de l'avenant*

La détermination de la cotisation suite à un avenant est un moment clé de la vie d'un contrat. En effet, il faut veiller à ce que celui-ci puisse couvrir le nouveau risque assuré, mais également d'éviter toute résiliation induite par le nouveau Tarif Technique. Il existe alors, tout comme lors des échéances portefeuille, un processus de revalorisation du Tarif Réel en cas d'avenant, faisant intervenir l'ETP.

Pour déterminer la cotisation qui devra être payée par l'assuré après l'avenant, quatre étapes sont nécessaires:

-Etape 1 : On calcule tout d'abord l'ETP associé à l'ancien risque, qui est défini par :

$$ETP_{\text{ancien risque}} = \frac{\text{Actual Price}_{\text{ancien risque}}}{\text{Commercial Price}_{\text{ancien risque}}}$$

-Etape 2 : Si la valeur de l'ETP avant avenant est supérieure à 1, alors on conserve la valeur de l'ETP. Si la valeur de l'ETP est inférieure à 1, alors on va le faire converger vers 1 plus ou moins rapidement selon que l'évolution du risque  $\frac{\text{Commercial Price}_{\text{nouveau risque}}}{\text{Commercial Price}_{\text{ancien risque}}}$  est importante ou non. La valeur de la revalorisation est définie à partir d'une grille dépendant de la tranche d'ETP avant avenant et du pourcentage d'évolution du risque.

-Etape 3 : On calcule enfin la valeur de la cotisation payée par l'assuré à partir de la relation suivante :

$$\text{Tarif Réel}_{\text{nouveau risque}} = ETP_{\text{nouveau risque}} * \text{Tarif Commercial}_{\text{ancien risque}}$$

-Etape 4 : En cas de sur-tarification avant le changement de risque, l'agent peut potentiellement réduire la valeur du nouveau Tarif Réel en utilisant une partie de l'enveloppe budget défense. La valeur de la réduction diffère selon s'il y a changement de véhicule, de l'ancienneté du contrat ou de la sinistralité passée.

Voici un schéma récapitulatif du processus de fonctionnement de l'ETP en cas d'avenant :

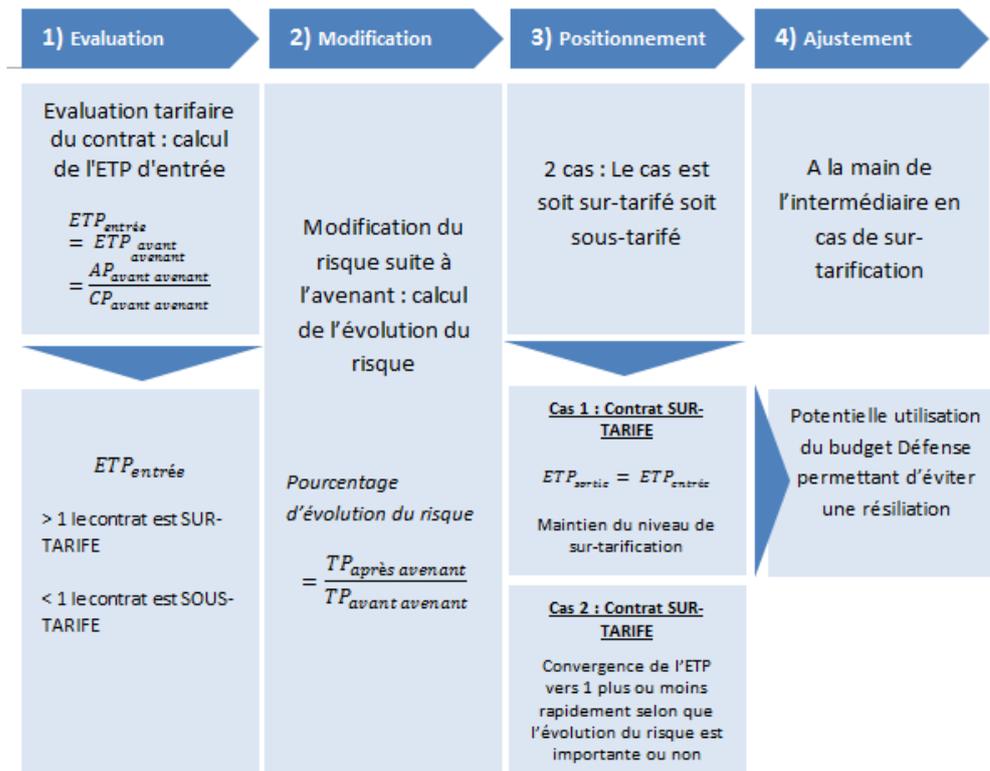


Figure 9: Récapitulatif du fonctionnement de l'ETP en cas d'avenant

Pour illustrer ce processus de revalorisation de primeréel en cas d'avenant, nous prendrons l'exemple du contrat suivant subissant différents scénarios d'évolution de risque :

Risque d'origine :

- Tarif commercial = 1 000€
- Tarif Réel = 700€

- Premier exemple de mécanisme d'ETP : Contrat sous-tarifé et risque qui augmente

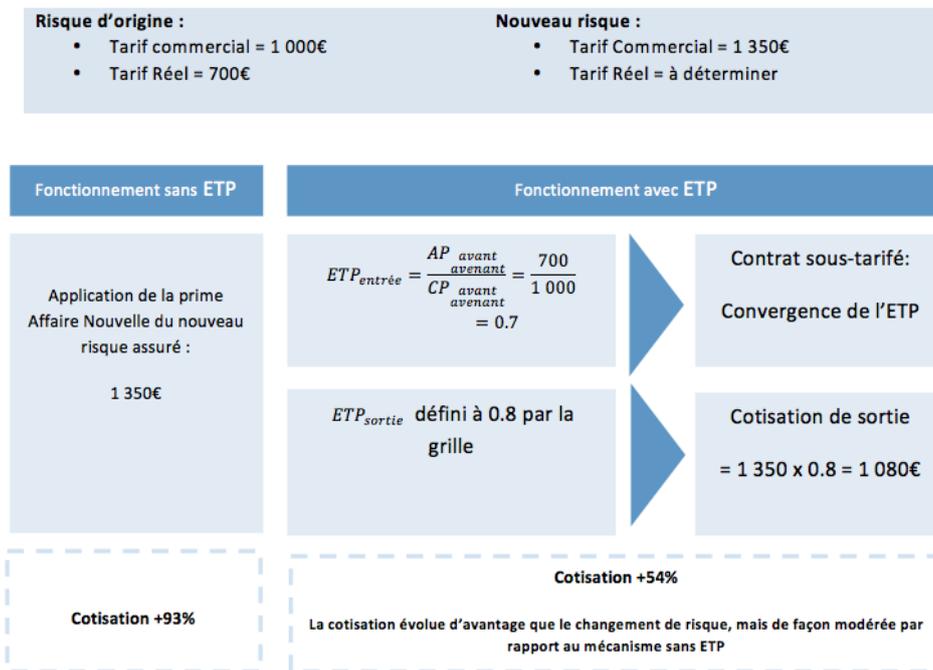


Figure 10: Exemple de mécanisme d'ETP avec un contrat sous-tarifé et un risque qui augmente

- Deuxième exemple de mécanisme d'ETP : Contrat sous-tarifé et risque qui diminue

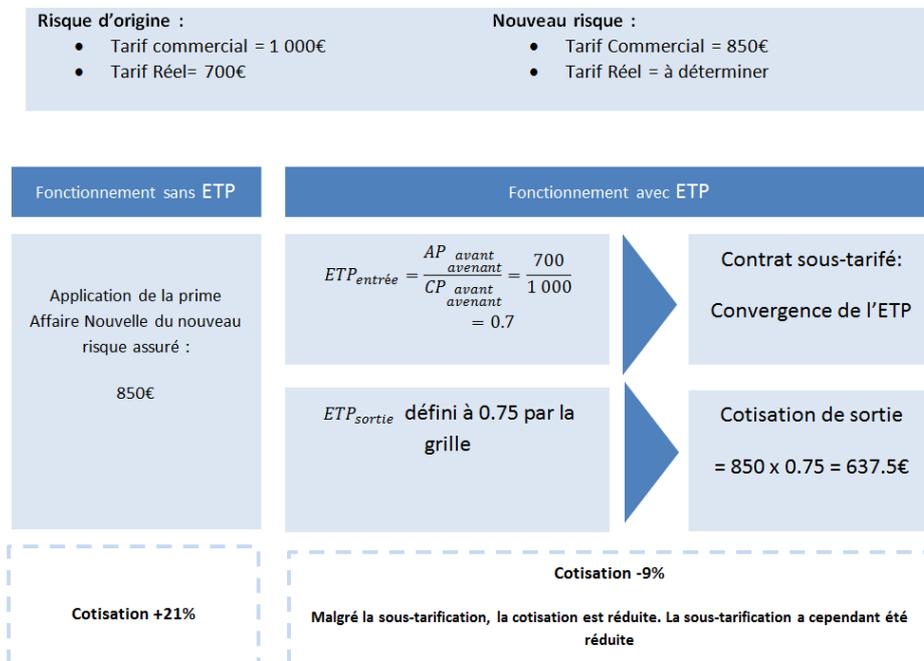


Figure 11: Exemple de mécanisme d'ETP avec un contrat sous-tarifé et un risque qui diminue

Cette partie nous a permis de nous familiariser avec les évolutions des différentes notions de primes intervenant dans la construction de l'AP/TP au cours de la vie du contrat. Nous avons vu que lors de la souscription, le Tarif Réel, égal au tarif Commercial, se retrouve proche du Tarif Technique. Il prend cependant en compte les modulations commerciales qui sont accordées au client. L'écart au Tarif Technique à l'affaire nouvelle est ainsi explicable majoritairement par les modulations commerciales qui sont appliquées à l'assuré. De par ce fait, nous porterons un fort intérêt aux modulations tarifaires lors de l'analyse de caractères discriminants sur l'AP/TP. Néanmoins, lors des échéances portefeuille le Tarif Réel et le Tarif Technique suivent des processus de revalorisation distincts, avec une certaine complexité pour le Tarif Réel. Ceci fera sans doute de l'ancienneté du contrat un critère fortement discriminant sur l'AP/TP. De plus, il était souligné que les taux de revalorisation à échéances dépendent de la formule souscrite et des antécédents sinistres. Ces deux critères devraient également sans doute ressortir lors de l'étude des écarts de rentabilité technique.

## Partie 3 : Intérêt de l'AP/TP pour la mesure de rentabilité technique et prise de responsabilité de cet indicateur dans la gestion dynamique des contrats

### a) L'AP/TP indicateur de pilotage technique complémentaire aux indicateurs classiques

L'AP/TP est un indicateur de rentabilité interne nouvellement introduit et désormais fréquemment utilisé au sein de la Direction Technique IARD d'Allianz France. Son utilisation a vocation à être développée et vulgarisée au sein de l'ensemble de la compagnie. Il correspond au rapport du Tarif Réel (Actual Price) payé par l'assuré sur le Tarif Technique (Technical Price) de son risque. Cet indicateur est, par conséquent, également appelé « Ecart au Tarif Technique ».

$$AP/TP = \frac{\text{Actual Price}}{\text{Technical Price}}$$

L'AP/TP permet, par construction, de rendre compte de la rentabilité de manière individuelle d'un contrat en mesurant l'écart entre la valeur de la cotisation effectivement payée et ce qui devrait être payé par l'assuré pour couvrir son risque assurantiel. Cet indicateur doit être comparé avec la valeur 1. En effet, si l'indicateur est supérieur à 1 alors la cotisation payée couvre plus que le risque assuré et s'il est inférieur à 1 alors ce contrat est sous-tarifé et constitue une fuite de profitabilité. Un contrat est d'autant plus rentable que l'indicateur est supérieur à 1 et d'autant moins rentable que son AP/TP est inférieur à 1.

Les indicateurs de rentabilité techniques classiques fréquemment utilisés par les compagnies d'assurance IARD et qui mesurent l'écart entre ce qui est apporté à la compagnie par les assurés et le montant engagé par la compagnie envers ces assurés sont :

- Le ratio sinistre à prime (S/P) : qui correspond au rapport du coût des sinistres sur les primes perçues par l'assureur sur un segment donné. Deux types de ratio S/P sont analysés, selon la valeur de la charge de sinistralité prise en compte. En effet, étant donné que l'on ne connaît le coût final d'un sinistre que tardivement, on peut :

- soit considérer la charge dossier/dossier correspondant à la somme des règlements et des provisions dossier/dossier. On parle alors de ratio S/P dossier/dossier.

- soit considérer la charge ultime qui est estimée par des méthodes de provisionnement (Chain Ladder, fréquence X coût moyen, ...). On parle dans ce cas de ratio S/P à l'ultime

- Le Ratio Combiné (CoR) : qui correspond au ratio S/P à l'ultime auquel on ajoute les frais d'administration et de gestion des sinistres supportés par l'entreprise sur la même période.

Voici un tableau présentant les principales différences et similitudes entre les indicateurs de rentabilité classiques et l'AP/TP :

	AP/TP	S/P vision dossier/dossier	S/P vision ultime	Ratio Combiné
Maille de reporting habituelle	Police à police ou sur des sous-segments	Sous-segment	Sous-segment	Sous-segment
Vision à l'ultime	Oui : Via la calibration du Tarif Technique (TP)	Non : Prise en compte uniquement du dossier/dossier*	Oui : Via la prise en compte des provisions pour sinistres (IBNR**) et pour recours	
Prise en compte des coûts de commission et frais généraux		Non	Non	Oui
Prise en compte du coût du capital		Non	Non	Non
Sinistralité prise en compte	Prospective : Espérance de la sinistralité du contrat basé sur les modèles de Tarif Technique	Rétrospective : Sinistralité observée (règlement + dossier/dossier) sur le segment	Rétrospective : Sinistralité observée (règlement + dossier/dossier) sur le segment et provisions pour sinistres (IBNR) et pour recours	
Modélisation actuarielle	Oui : Modélisation du TP	Non	Oui : Modélisation des provisions (IBNR)	

\* provision dossier/dossier : montant forfaitaire doté pour couvrir le sinistre selon les premières observations

\*\* IBNR (Incured But Not Reported): Provision pour différence entre la charge dossier/dossier et charge ultime de la sinistralité estimée actuariellement

Figure 12: Comparaison des indicateurs de rentabilité classique avec l'AP/TP

Ce tableau nous permet ainsi de souligner la pertinence de l'AP/TP puisque celui-ci combine un certain nombre d'aspects des indicateurs classiques tout en ayant une réelle valeur ajoutée.

En effet, l'AP/TP intègre les coûts de commission, les frais généraux et le coût du capital à travers le Tarif Technique. Il tient aussi compte de la sinistralité du contrat à travers la modélisation du Tarif Technique.

De plus, par construction, l'AP/TP a un certain nombre d'autres avantages qui lui sont propres et lui permettent ainsi de compléter l'information fournie par les indicateurs de rentabilité classiques. En effet, l'AP/TP a la particularité que la sinistralité prise en compte dans cet indicateur est basée sur l'espérance du coût des sinistres, via la modélisation du Tarif Technique, et non sur l'observation de celle-ci comme dans les indicateurs classiques. La sinistralité prise en compte dans l'AP/TP ne présente donc pas l'aléa de l'observation comme avec le S/P et le Ratio Combiné lorsqu'il est étudié police à police, permettant alors d'observer l'AP/TP de manière individuelle. La prise en compte de la sinistralité à travers une espérance permet également de rendre l'AP/TP prospectif. L'AP/TP peut notamment être calculé dès l'entrée en portefeuille.

Ainsi, les dimensions individuelles et prospectives de l'AP/TP lui ont permis de prendre une place importante dans le pilotage de portefeuille et dont les diverses utilisations seront développées dans le paragraphe suivant.

### b) Multiples utilisations de l'AP/TP dans le pilotage de portefeuille

L'indicateur permet de rendre compte de combien la cotisation de l'assuré s'écarte du risque assuré, créant ainsi un certain degré de sous ou sur rentabilité contrat par contrat. Cette indication a un rôle primordial pour le management du portefeuille et l'AP/TP est ainsi intégré à différents moments de la vie du contrat :

- Des reportings d'AP/TP sont réalisés de manière trimestrielle sur l'ensemble des portefeuilles selon le type de produit assurantiel et le canal de distribution. Ces rapports permettent ainsi de suivre la rentabilité technique globale du portefeuille selon s'il s'agit d'un renouvellement ou d'une affaire nouvelle. Des reportings segmentés sur certains critères de tarification ou de modulation tarifaire pourraient être mis en place pour suivre de plus près cette rentabilité mesurée par l'AP/TP.

- L'AP/TP intervient également lors de la surveillance des portefeuilles. En effet, de manière récurrente, les polices ayant été sinistrées dans les 12 derniers mois sont surveillées : On attribue à chacun de ces contrats un score de sinistralité sur 5 ans dépendant de l'ancienneté de la survenance de ces sinistres ainsi que de leur sévérité (responsable ou non responsable, corporel, matériel, vol ou bris de glace). Certaines polices se voient alors soit fortement revalorisées ou soit résiliées en fonction de la valeur de leur score de sinistralité et de leur tranche d'AP/TP. Cette surveillance permet alors d'écarter les contrats risqués et sous-tarifés.

Claims score on the last 5 years							
		-				+	
AP/TP	↑ +	not detected ✓	not detected ✓	not detected ✓	increase of $x_1\%$ ⚠	increase of $x_2\%$ ⚠	increase of $x_3\%$ ⚠
		not detected ✓	not detected ✓	increase of $x_1\%$ ⚠	increase of $x_2\%$ ⚠	increase of $x_3\%$ ⚠	increase of $x_4\%$ ⚠
		not detected ✓	increase of $x_1\%$ ⚠	increase of $x_2\%$ ⚠	increase of $x_3\%$ ⚠	increase of $x_4\%$ ⚠	increase of $x_5\%$ ⚠
		increase of $x_1\%$ ⚠	increase of $x_2\%$ ⚠	increase of $x_3\%$ ⚠	increase of $x_4\%$ ⚠	increase of $x_5\%$ ⚠	Cancellation ✖
		increase of $x_2\%$ ⚠	increase of $x_3\%$ ⚠	increase of $x_4\%$ ⚠	increase of $x_5\%$ ⚠	Cancellation ✖	Cancellation ✖
		increase of $x_3\%$ ⚠	increase of $x_4\%$ ⚠	increase of $x_5\%$ ⚠	Cancellation ✖	Cancellation ✖	Cancellation ✖
	↓ -						

Figure 13: Fonctionnement de la surveillance de portefeuilles faisant intervenir l'AP/TP

- Le suivi de la valeur de l'AP/TP permet de rendre compte de l'évolution de la rentabilité d'un contrat au cours de sa vie. À la souscription et lors des premières années d'assurance, on accorde aux assurés un certain nombre d'avantages tarifaires et notamment commerciaux, créant alors un fort écart entre le Tarif Technique et la cotisation payée. L'AP/TP est alors faible et en général inférieur à 1. Les contrats jeunes sont alors non-rentables. Cependant, au cours de la vie d'un contrat, les avantages commerciaux sont peu à peu réduits, notamment par le processus complexe de revalorisation du Tarif Réel, faisant ainsi augmenter peu à peu l'AP/TP. Les contrats anciens deviennent de cette manière sur-profitables, finançant ainsi la perte faite au niveau des contrats jeunes. On pourrait alors faire une analogie avec la valeur contrat, qui s'avère en général être non rentable durant les premières années suivant la souscription dû aux forts coûts d'acquisition qui ont été déployés et doivent être compensés, puis devient ensuite rentable et dont la rentabilité est à piloter afin d'assurer conjointement la rétention des clients et leur surveillance.

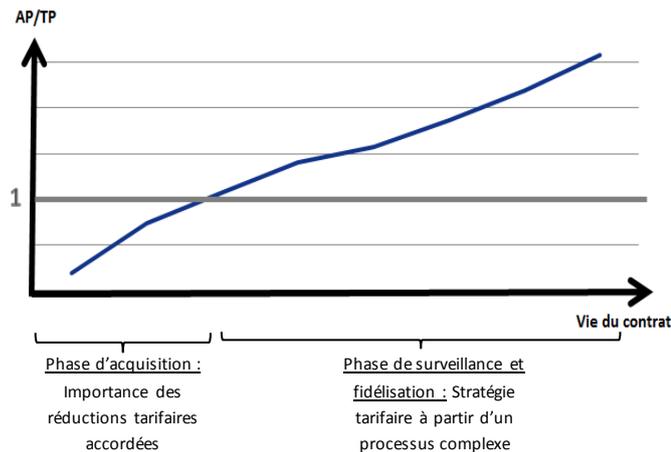


Figure 14: AP/TP et suivi de la rentabilité technique au cours de la vie d'un contrat

- Enfin, nous avons vu que l'AP/TP intervenait dans le processus de revalorisation, lors du choix des revalorisations additionnelles, permettant de pénaliser les contrats dont la cotisation se retrouvait sous-estimée par rapport au risque qu'ils représentent. Bien que le Tarif Commercial soit la principale référence du Tarif Réel car celui-ci est calé au marché, le Tarif Technique, qui représente le coût réel du risque assuré, doit également être pris en compte lors des revalorisations. En effet, on ne veut pas que la cotisation payée se retrouve trop loin du risque assuré. Pour prendre en compte cet écart entre la valeur de la cotisation et le Tarif Technique au moment des échéances portefeuille, des revalorisations supplémentaires peuvent être appliquées au Tarif Réel dépendant de l'AP/TP, de l'index de compétitivité ainsi que de la sensibilité du client au prix. L'AP/TP est ainsi totalement intégré dans le processus de revalorisation.

Voici un tableau schématisant le principe de fonctionnement de la revalorisation additionnelle faisant intervenir l'AP/TP:

		Index de compétitivité			
		Mauvais positionnement		Bon positionnement	
AP / TP	Sous-tarifé (AP/TP < 1)	Sensible +	Non sensible ++	Sensible +	Non sensible ++
	Sur-tarifé (AP/TP > 1)	Sensible --	Non sensible =	Sensible -	Non sensible =

Figure 15: Fonctionnement des revalorisations additionnelles faisant intervenir l'AP/TP

Par exemple, un contrat ayant un AP/TP inférieur à 1, c'est-à-dire que l'assuré paye une cotisation (Actual Price) inférieure au coût de son risque (Technical price), et ayant un bon positionnement par rapport à la concurrence se verra attribuer une revalorisation additionnelle positive, pénalisant ainsi le fait que sa cotisation n'est pas alignée au risque qui est assuré.

Ce premier chapitre nous a permis de présenter ce nouvel indicateur de rentabilité technique développé au sein de la Direction Technique IARD d'Allianz France. Celui-ci a la particularité de se calculer sur la maille très fine qu'est le contrat d'assurance puisqu'il mesure, police par police, l'écart entre la cotisation payée et le risque assuré. Nous avons tout d'abord expliqué comment cet indicateur était construit en présentant les différents mécanismes d'établissement des primes intervenant dans le calcul de cet indicateur. Par la suite nous avons démontré le fait qu'il s'agissait d'un indicateur adéquat pour la mesure de rentabilité technique à l'échelle du contrat. Enfin, nous avons souligné que sa réelle valeur ajoutée lui avait alors déjà permis de se voir attribuer un rôle primordial dans la gestion de portefeuille d'assurance sur le périmètre de l'automobile de particuliers.

Il est tout de même important de relever que l'AP/TP est un indicateur qui repose fortement sur la qualité du Tarif Technique. En effet, il considère que ce dernier est en parfaite adéquation avec le risque qui est assuré pour un contrat donné. Or en pratique, le Tarif Technique sera en effet très représentatif du risque assuré, mais il sera tout de même toujours perfectible.

## CHAPITRE 2 : Analyse de la rentabilité d'un portefeuille automobile de particuliers au travers de l'AP/TP

A partir de ce chapitre et pour la suite de ce mémoire, nous nous focaliserons sur le portefeuille le plus important de la Direction Technique IARD d'Allianz France, à savoir le portefeuille automobile classique de particulier distribué par le biais d'agents généraux.

Dans ce chapitre nous présenterons tout d'abord la base qui servira de support pour la suite de nos études sur l'indicateur de rentabilité technique qu'est l'AP/TP sur le périmètre du marché de l'assurance automobile. Ensuite, nous réaliserons une première analyse statistique visant à dresser un état des lieux de la rentabilité technique du portefeuille transcrite par l'AP/TP. Cette première approche nous permettra également de mieux appréhender la rentabilité par segments d'assurés, dégageant ainsi des critères typiquement associés à une sur ou sous-rentabilité des contrats.

## Partie 1 : Présentation de la base et établissement du périmètre

### a) Choix de la base de données utilisée

De manière trimestrielle, la Direction Technique IARD d'Allianz France livre au groupe, pour chaque produit IARD de particuliers et de professionnels, un reporting sur l'AP/TP les informant de la valeur de l'indicateur sur chacun des canaux de distribution et selon s'il s'agit de contrat en affaire nouvelle ou déjà en portefeuille.

La base qui sert de support pour cette étude est donc la base de données utilisée lors la réalisation du reporting AP/TP du portefeuille quatre-roues de particuliers ayant été intermédié par le biais d'agents généraux d'Allianz France, couramment abrégé « 4RP Agents » en interne. Celui-ci est vu à fin décembre 2016 et permet de dresser un état des lieux de la rentabilité technique des contrats en portefeuille à une date donnée relevée par l'AP/TP. On rappelle que les contrats novices ne font pas partie du périmètre étudié, étant donné qu'ils ne présentent pas la même structure de Tarif Technique que le reste du portefeuille. Les considérer dans cette étude viendrait biaiser la significativité de l'indicateur AP/TP. Le portefeuille étudié comporte plus d'1.7 millions de contrats.

Les bases d'AP/TP à fin décembre 2014 et à fin décembre 2015 seront aussi exceptionnellement utilisées lors de la réalisation de tests de stabilité de modèles dans le temps.

### b) Présentation du support des variables descriptives

Chaque ligne de la base correspond à un contrat en portefeuille à fin décembre 2016, identifié par un numéro de police. La base étudiée comporte le Tarif Technique total associé à l'ensemble des garanties souscrites et le Tarif Réel annuel par ligne, nous permettant ainsi d'obtenir l'AP/TP pour chacune des polices. On dispose également d'un ensemble de données décrivant chacun des contrats en portefeuille. Pour cette étude, il a été jugé pertinent de prendre en compte l'ensemble des variables tarifaires utilisées lors de la tarification d'un produit et dont nous disposons dans la base initiale, étant donné que le Tarif Technique et le Tarif Réel en découlent. Elles prennent en compte un certain nombre de données décrivant le risque assuré. Parmi les variables tarifaires, certaines sont extraites de données externes fournies par la société de Sécurité et Réparation Automobile (SRA). Elles comprennent le Groupe de véhicule, la Classe de prix, la Classe de réparation, l'Energie, la Carrosserie et la Note de sécurité. Les données disponibles sur la base concernant les modulations commerciales accordées à l'assuré vont également être considérées étant donné l'impact qu'ils ont sur le Tarif Commercial et par conséquent sur le Tarif Réel. Ainsi, nous disposons à la fois de données sur le

conducteur principal, sur le véhicule, sur la zone géographique du risque, sur le contrat souscrit ainsi que sur les modulations tarifaires qui leur sont accordées.

Voici les variables retenues :

Intitulé de la variable	Signification
<b>Données sur le conducteur principal</b>	
Age	Age du conducteur
CSP	Catégorie Socio-Professionnelle
CoefBM	Coefficient Bonus-Malus
AgeP	Age d'obtention du permis
AntSin	Score de sinistralité sur les 36 derniers mois
Sexe	Sexe : Homme, Femme ou Personne Morale
<b>Données sur le véhicule assuré</b>	
Marque	Marque du véhicule
ClRep*	Classe de réparation
AncVehi	Vétusté du véhicule
Vehprop	Véhicule propre : Véhicule de moins de 2 ans et bénéficiant du bonus écologique à la date de souscription (à ce jour émission inférieure ou égale à 60 g/km de CO2) ou véhicule électrique
Clprix*	Classe de prix
TauxCO2	Taux d'émission de CO2
GPSRA*	Groupe de véhicule
Ancacq	Année d'acquisition
ClasseVol	Classe de vol : les véhicules ayant une fréquence de vol élevée sont répertoriés par Allianz
Energie*	Energie: Essence, Diesel, GPL, Autre
Carrosserie*	Type de carrosserie
NoteSecu*	Note de sécurité : dépend du matériel de prévention et de sécurité présent dans le véhicule. Les modalités possibles sont -2, -1, 0, et 1, la note la plus faible étant la meilleure
<b>Données géographiques</b>	
ZoneRC	Zone de risque pour la garantie Responsabilité Civile
ZoneBDG	Zone de risque pour la garantie Bris de Glace
ZoneVI	Zone de risque pour la garantie Vol/Incendie
ZoneDOM	Zone de risque pour la garantie Dommage Tous Accidents
Garage	Présence ou absence de garage
<b>Données sur le contrat</b>	
Formule	Formule souscrite : 5 formules sont proposées (Formule aux Tiers, Formule Intermédiaire, Formule Tous Risques et deux autres formules dérivées)
AgeContrat	Ancienneté de la souscription du contrat Allianz France
Fractionnement	Fractionnement de la prime (Annuel, Semestriel, Trimestriel, Mensuel)

KiloMetre	Usage kilométrique : il existe divers forfaits kilométriques avec ou sans relevés de compteurs par la SOFCA
DoubleAuto	Offre « Double Auto » : Client assurant au moins deux véhicules quatre roues dont l'un en forfait kilométrique
ValeurPlus	Option « Valeur Plus » : permet à l'assuré de protéger son véhicule de la dépréciation durant ses premières années (36 mois). Indemnisation à valeur d'achat en cas de sinistre
PackRéparation	« Pack Réparation » : lorsqu'il est souscrit avant les 7 ans du véhicule, permet la prise en charge des réparations suite à une panne de toute catégorie (mécanique, électrique, électronique, hydraulique)
FranchiseBDG	Montant de la franchise de la garantie Bris de Glace
FranchiseVol	Montant de la franchise de la garantie Vol
FranchiseDom	Montant de la franchise de la garantie Dommage Tous Accidents
ClauseTari	Clause tarifaire : conduite couple ou autre
<b>Données sur les modulations tarifaires</b>	
AvClient	Avantage client
AvProsp	Avantage prospect
TauxProto	Taux protocole
TauxBudget	Taux budget
TauxFirme	Taux firme

\* données fournies par la SRA

Figure 16: Ensemble de variables de la base initiale sélectionnées pour notre étude

## Partie 2 : Etude de la distribution et statistiques des AP/TP du portefeuille étudié

Il est rappelé que l'AP/TP est indicateur particulier puisqu'il permet de rendre compte de la rentabilité technique de manière individuelle, c'est-à-dire contrat par contrat. Représentant le montant des primes encaissées sur les dépenses engagées, celui-ci est à l'équilibre lorsqu'il est égal à 1. Supérieur à ce seuil, le contrat est techniquement sur-rentable pour l'assureur, et en deçà de ce seuil il est sous-rentable. La rentabilité technique mesurée sera d'autant meilleure que l'AP/TP sera supérieur à 1.

Nous allons tout d'abord porter une attention particulière à la distribution de l'AP/TP de notre portefeuille. Ceci nous permettra d'avoir une idée de la répartition de la rentabilité technique sur le portefeuille et de constater à quel point l'AP/TP est dispersé de 1, valeur d'AP/TP qui est associé à une rentabilité équilibrée (ni de sous-rentabilité, ni de sur-rentabilité). On se rendra ainsi compte des proportions de contrats sur-profitables et au contraire sous-profitables.

La distribution de l'indicateur étudié sur l'ensemble de notre portefeuille quatre roues de particuliers vu à fin décembre 2016 est la suivante :

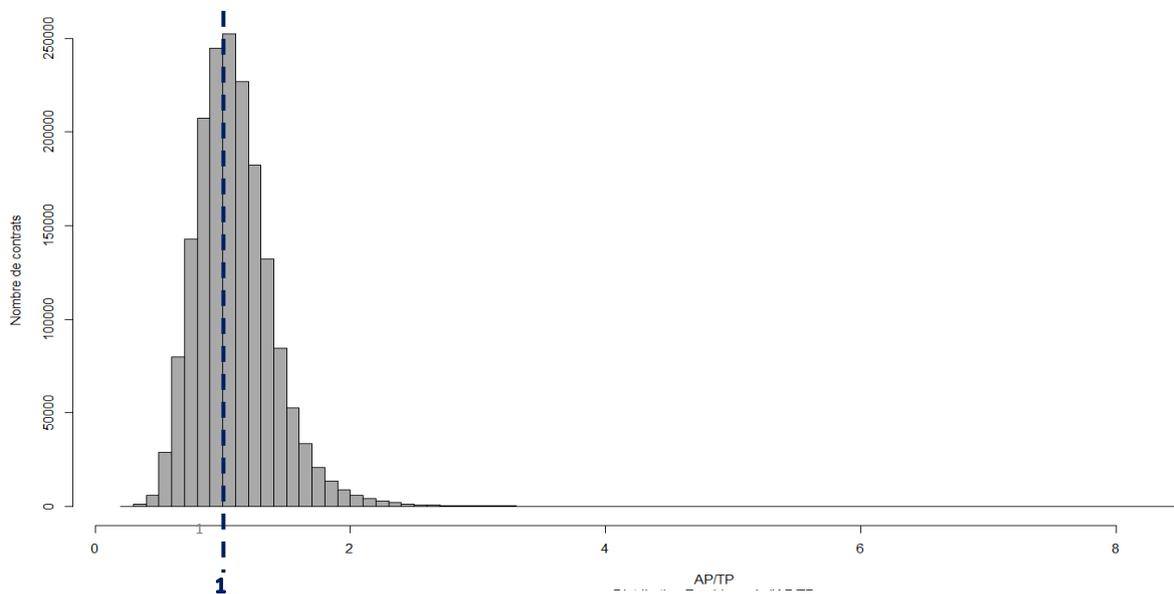


Figure 17: Distribution empirique de l'AP/TP du portefeuille « 4RP Agents »

La distribution de l'AP/TP nous expose une distribution globale concentrée autour de 0.5 et 1.5, avec la présence d'une queue de distribution à droite.

Les statistiques suivantes sur l'AP/TP ont également été obtenues grâce à la procédure « univariate » du logiciel SAS :

Moments				Quantile	Estimate
<b>N</b>	1738508	<b>Sum Weights</b>	1738508	<b>100% Max</b>	9.454477
<b>Mean</b>	1.09531507	<b>Sum Observations</b>	1904214	<b>99%</b>	2.040837
<b>Std Deviation</b>	0.30785969	<b>Variance</b>	0.09477759	<b>95%</b>	1.625096
<b>Skewness</b>	1.36648877	<b>Kurtosis</b>	7.2800176	<b>90%</b>	1.465934
<b>Uncorrected SS</b>	2250485.78	<b>Corrected SS</b>	164771.496	<b>75% Q3</b>	1.259337
<b>Coeff Variation</b>	28.1069525	<b>Std Error Mean</b>	0.00023349	<b>50% Median</b>	1.062225
				<b>25% Q1</b>	0.886229
				<b>10%</b>	0.746206
				<b>5%</b>	0.671298
				<b>1%</b>	0.551423
				<b>0% Min</b>	0.201901

Figure 18: Statistiques sur l'AP/TP du portefeuille « 4RP Agent s »

On remarque que la moyenne des AP/TP observés est de 1.09 sur le portefeuille étudié. Le portefeuille dégage donc globalement une rentabilité de 9% meilleure que la cible. Il est important de rappeler que notre étude se fait hors conducteurs novices dont la rentabilité est bien moindre comparée au reste du portefeuille « 4RP Agents » et pensant ainsi à la baisse sur la rentabilité globale de l'ensemble du portefeuille.

Les quantiles de la distribution empirique de l'AP/TP nous indiquent que l'indicateur prend des valeurs comprises entre 0.20 et 9.45. De plus, le quantile à 99%, égal à 2.04, nous confirme que l'essentiel des AP/TP du portefeuille se trouve concentré entre 0.20 et 2.04.

Les valeurs de l'AP/TP doivent être acceptables car celles-ci dépendent de l'écart des revalorisations entre le Tarif Réel et le Tarif Technique, qui doit être fait de manière raisonnable afin d'éviter l'insatisfaction du client qui résulterait en la résiliation de son contrat. Lorsque l'on tient compte de ce fait, on peut expliquer la raison pour laquelle les AP/TP sont bornés et peu dispersés, ayant un écart type de 0.30.

Le Skewness égale à 1.37 nous fait part d'une certaine asymétrie à droite. On a ainsi plus de contrats ayant un AP/TP supérieur à la valeur moyenne sur le portefeuille que de contrats ayant un AP/TP inférieur à la valeur moyenne. Le coefficient de Kurtosis de 7.28 témoigne de la présence d'une queue de distribution étalée à droite. On a ici affaire à une distribution ayant une queue plus lourde que la loi normale qui est caractérisée par une Kurtosis de 3.

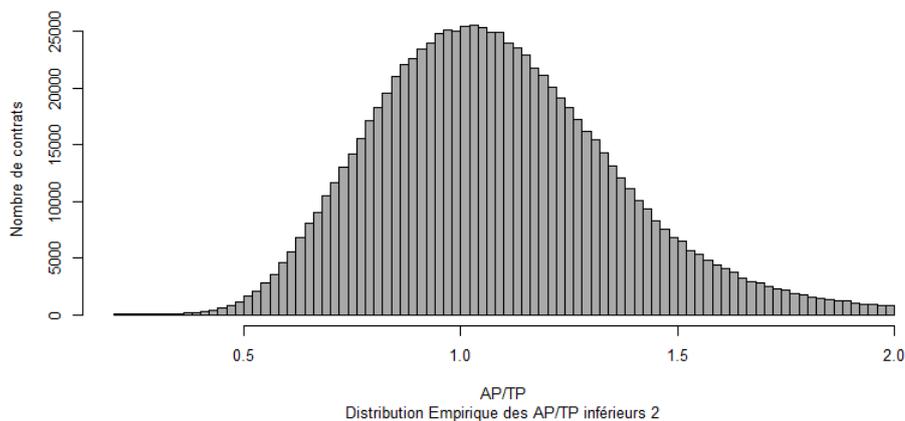
L'adéquation de la distribution de l'AP/TP avec celles de lois classiques sera faite plus tard dans ce mémoire, lors de la modélisation de l'indicateur.

Des zooms sur trois tranches d'AP/TP sont réalisés afin d'obtenir une meilleure appréciation de la distribution de l'AP/TP. On définit les trois segments d'AP/TP de la manière suivante en fixant les seuils aux valeurs d'AP/TP à 2 et à 4:

Classe d'AP/TP	Valeurs d'AP/TP
AP/TP « ordinaires »	Inférieur à 2
AP/TP « extrêmes »	Compris entre 2 et 4
AP/TP « exceptionnels »	Supérieur à 4, assimilé à du bruit statistique et non modélisés par la suite

Figure 19: Récapitulatif des classes d'AP/TP du portefeuille « 4RP Agents »

Ces trois segments représentent respectivement en proportion 98.84%, 1.14% et 0.01% du portefeuille « 4RP Agents ». Leurs distributions respectives sont les suivantes :



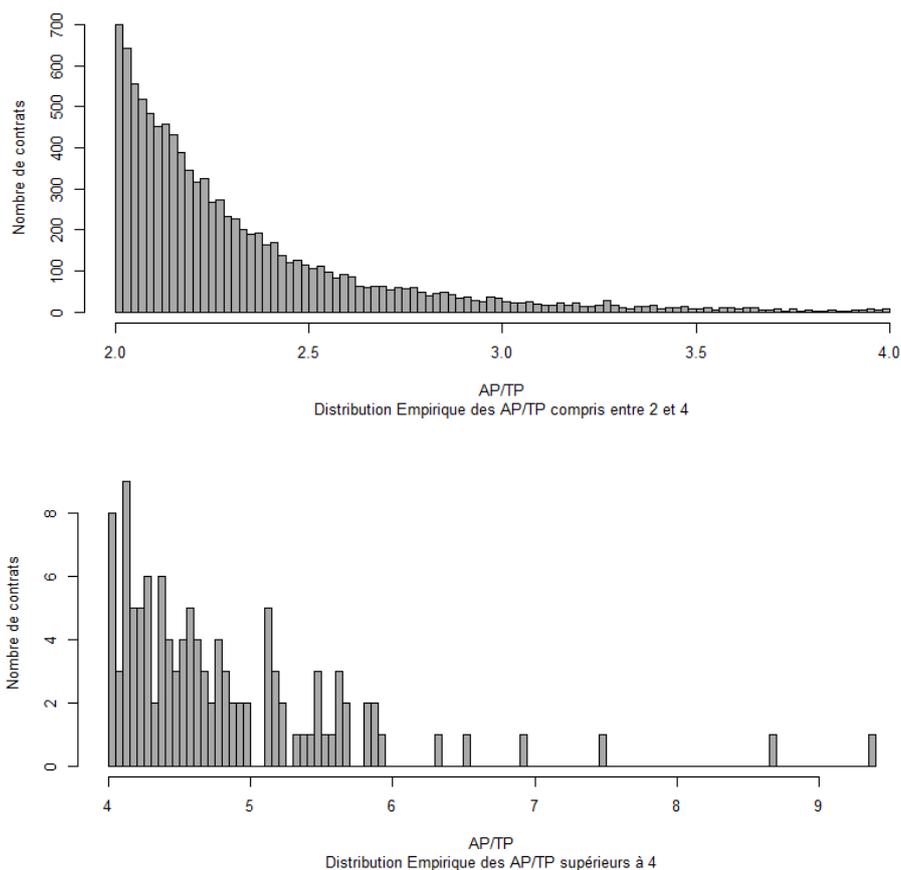


Figure 20: ZOOMS sur la distribution empirique des classes d'AP/TP du portefeuille « 4RP Agents »

L'essentiel des AP/TP se retrouve concentré entre 0.2 et 2. En effet, nous avons relevé que plus de 98% des contrats ont un AP/TP dans cet intervalle. On a précédemment expliqué cela par le fait que la valeur de l'AP/TP dépend des revalorisations au cours des échéances principales et des avenants, qui ne peuvent être excessives afin d'éviter la résiliation de l'assuré. Les AP/TP de la première tranche seront appelés « AP/TP ordinaires ». Ce nom fait référence aux sinistres qualifiés d'attritionnels ou ordinaires en tarification lorsqu'ils sont caractérisés par une fréquence élevée et un coût moyen ordinaire. Ces AP/TP représentent l'essentiel de la distribution de notre indicateur de rentabilité.

La densité empirique de l'AP/TP fait apparaître la présence globale d'une queue à droite d'AP/TP supérieurs à 2 que nous allons vouloir étudier dans la mesure du possible. En effet, bien qu'en effectif moindre, il est essentiel de considérer ces contrats associés à une sur-tarification élevée puisque ce sont les contrats présentant un AP/TP important et dégagent ainsi la rentabilité la plus importante. Ce sont, avec les contrats ayant un AP/TP très faible, les contrats que nous souhaitons capter car ils constituent des poches de sous-rentabilité et sur-rentabilité. La distribution des AP/TP compris entre 2 et 4 nous dévoile un certain nombre de profils intéressants et ayant une distribution lisse. Ces AP/TP dont la valeur se distingue de l'essentiel de la distribution seront appelés « AP/TP extrêmes »

Enfin, on constate que les AP/TP ayant des valeurs supérieures à 4 affichent une distribution moins structurée, étant en effectif moindre. Ils seront qualifiés d'AP/TP « exceptionnels » car ils constituent 209 contrats sur les 1 738 508 contrats, soit 0.01% du portefeuille. Ces AP/TP, dont les valeurs sont difficilement explicables, seront assimilés à du bruit statistique et ne seront pas étudiés dans la suite de ce mémoire.

Voici un bilan de la distribution des AP/TP de notre portefeuille :

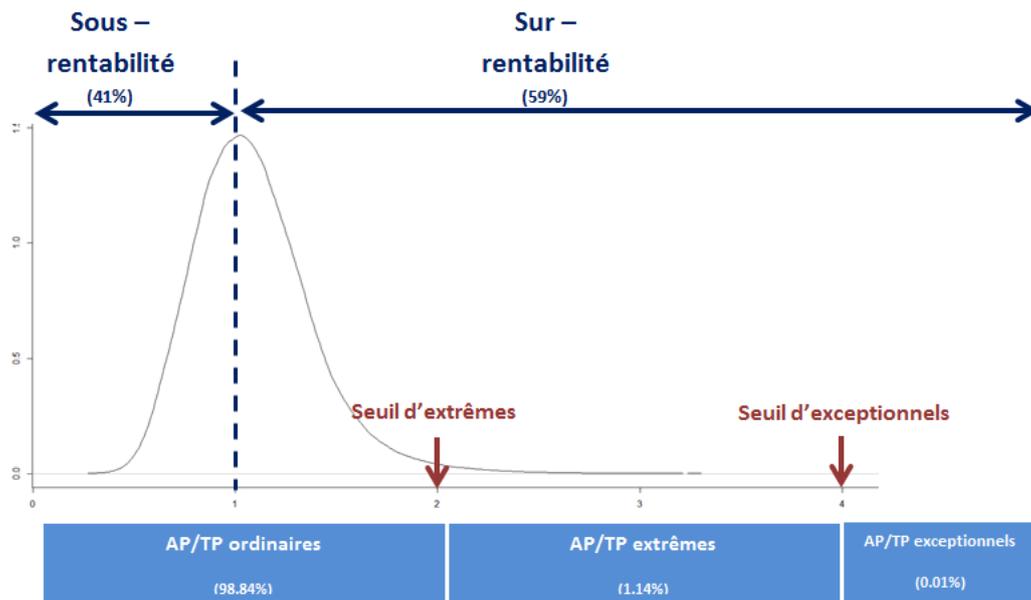


Figure 21: Récapitulatif de la densité de l'AP/TP du portefeuille « 4RP Agents »

La partie suivante va nous permettre de nous familiariser avec l'ensemble des variables disponibles dans notre base et pouvant avoir une influence sur l'AP/TP.

### Partie 3 : Etude de la discrimination des variables sur l'AP/TP

L'objectif de cette partie est de comprendre quelles sont les variables ayant une plus forte influence sur la valeur de l'AP/TP et donc dégageant de la sur-rentabilité ou au contraire de la sous-rentabilité technique. Les analyses uni-variées expliqueront les effets d'un critère sur notre indicateur et détermineront ainsi les écarts de valeurs d'AP/TP selon les modalités de la variable considérée. Il faudra ensuite s'intéresser aux corrélations inter-variables afin de prendre en compte les interactions qui existent entre celles-ci. On étudiera également les corrélations des différentes variables avec l'AP/TP, ce qui permettra de vérifier les constats ayant été réalisés quant à la discrimination d'une variable sur l'AP/TP.

#### a) Analyse uni-variée : influence des variables sur l'AP/TP

La majorité des variables présentes dans notre support sont de nature qualitative (exemple : catégorie socio-professionnelle du conducteur principal). Les variables quantitatives (exemple: ancienneté du contrat) ont été discrétisées en variables qualitatives après les avoir subdivisées en classe pour créer des modalités. La subdivision des variables qualitatives sera guidée par les analyses statistiques uni-variées mais sera également réalisée de manière à avoir un sens d'un point de vue métier. L'ensemble des variables du support se trouvent être désormais toutes de nature qualitative.

Pour chaque variable du support, on s'assure également que chacune des modalités est suffisamment représentée. Suite à un choix personnel, on considérera qu'une modalité est en effectif suffisant dès lors qu'elle contiendra 4 % du portefeuille (plus de 69 000 contrats). Les regroupements sont faits selon un critère de proximité des modalités ou selon un critère de proximité de la valeur des AP/TP moyens des classes initiales.

Avant d'aborder ces analyses il est important d'avoir en tête qu'il est en général difficile d'anticiper, contrairement au Tarif Technique ou au Tarif Réel, dans quel sens les modalités d'une variable de tarification influenceront sur l'AP/TP. En effet, étant donné qu'il s'agit d'un rapport de variables dont nous pouvons en général anticiper la variation en fonction des modalités, il est difficile de prévoir dans quel sens jouera le rapport des deux. Tout dépendra de la manière dont le Tarif Réel se détachera du Tarif Technique par le biais des modulations commerciales et des revalorisations successives. On rappellera cependant que, pour certaines variables, de potentielles discriminations ont été anticipées dans le chapitre précédent lorsque les variables intervenaient dans le processus de revalorisation des différentes primes. C'était notamment le cas de l'ancienneté du contrat, de la formule souscrite et de la sinistralité passée.

Une variable sera considérée comme discriminante sur l'AP/TP si une ou plusieurs de ses modalités présentent des valeurs d'AP/TP particulièrement faibles ou particulièrement élevées. Ceci illustre bien la notion de recherche de niche de sur-tarification ou sous-tarification.

- *Analyse de l'effet de l'âge du conducteur*

Prenons l'exemple de l'âge du conducteur principal pour expliquer ce constat :

On rappelle que le portefeuille exclut les conducteurs novices. Par conséquent, il ne comporte notamment pas d'assurés de moins de 21 ans. Les personnes morales font référence aux assurés non rattachables à une personne physique, de type taxis ou profession libérale. La variable âge du conducteur étant initialement de type quantitatif, des regroupements par continuité ont été réalisés de sorte que chaque modalité contienne au moins 4% de l'effectif. Ceci nous permet de nous assurer de la robustesse des conclusions de nos analyses.

La figure ci-dessous représente l'évolution du Tarif Technique en fonction de l'âge du conducteur :

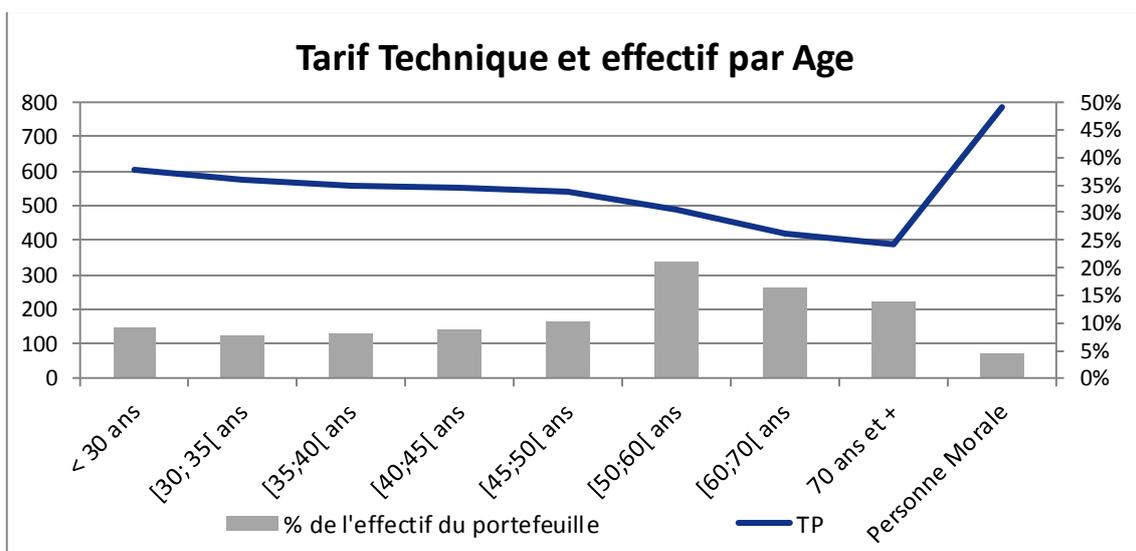
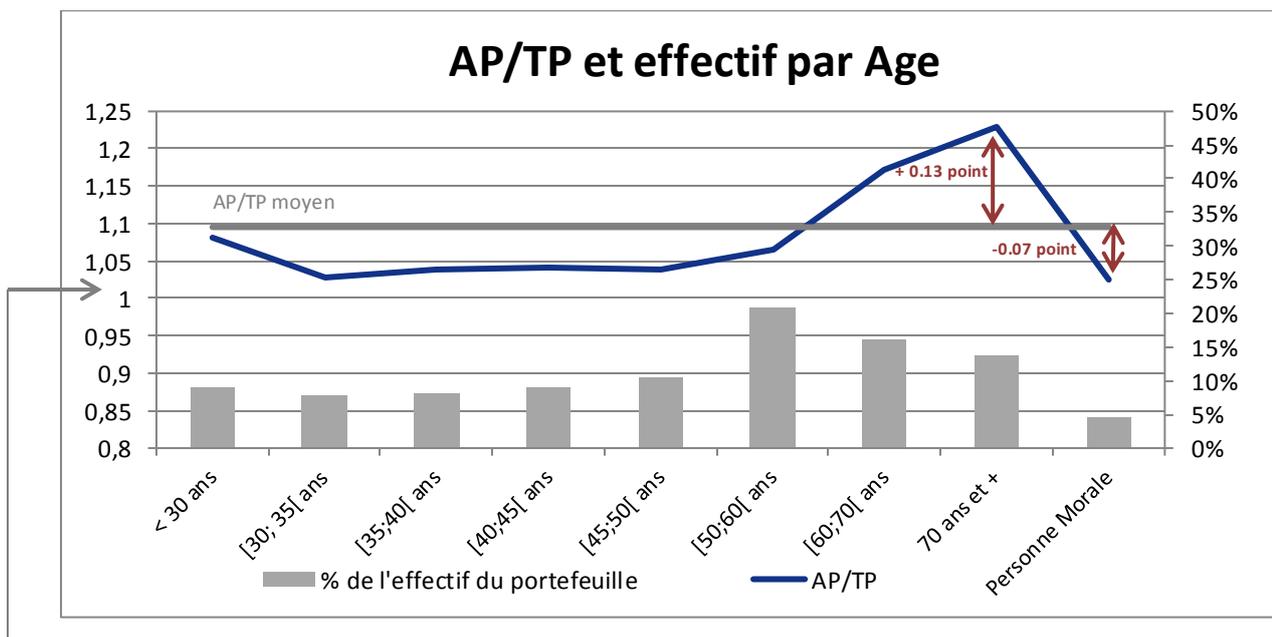


Figure 22: Analyse uni-variée du Tarif Technique et effectif par Age

Le Tarif Technique décroît nettement avec l'âge de l'assuré sur l'ensemble des modalités. Ceci est explicable par le fait que les conducteurs sont de plus en plus expérimentés et prudents. Les conducteurs plus jeunes ont donc en principe une fréquence plus élevée qui pèse sur la valeur de la Prime Pure. Par conséquent les conducteurs plus jeunes ont une prime technique plus élevée que les conducteurs plus âgés. Il faut également noter que les personnes morales ont un Tarif Technique nettement supérieur au reste du portefeuille. Ceci est souvent explicable par le fait que nous avons affaire à des contrats assurant des véhicules sur-équipés.

Dans le cas de l'AP/TP, on ne peut pas anticiper le lien existant entre la modalité d'un critère considéré et le sens de variation de l'AP/TP. En effet, ce n'est pas parce qu'une modalité est associée à un risque moindre qu'elle sera plus ou moins rentable. Tout dépend de la manière dont le Tarif Technique a été corrigé de manière commerciale pour donner le Tarif Réel et ainsi créer un écart.



Rentabilité du contrat  
à l'équilibre

Figure 23: Analyse uni-variée de l'AP/TP et effectif par Age

Le graphique ci-dessus nous montre qu'avec l'AP/TP, on constate une certaine monotonie, sauf exception des moins de 30 ans et des personnes morales, sur la courbe d'AP/TP : on observe globalement une croissance de la rentabilité mesurée par l'AP/TP avec l'âge du conducteur. En effet, l'AP/TP est supérieur pour les moins de 30 ans que pour les 30-35 ans. Il croît ensuite avec l'âge du conducteur pour les assurés de 30 ans et plus, avec un pic de rentabilité pour les assurés de 70 ans et plus et un écart de + 0.13 points par rapport à la rentabilité moyenne. Les 60 ans et plus constituent ainsi une niche de rentabilité. Les personnes morales constituent la modalité la moins rentable du portefeuille. En effet, la valeur de l'AP/TP moyen sur cette modalité se distingue de la moyenne du portefeuille, avec un écart d'AP/TP de 0.07 point en dessous de la moyenne. Ce graphique nous permet de mettre en lumière la détection d'une possible niche de rentabilité auprès des plus de 60 ans.

- Analyse de l'effet de l'ancienneté du contrat

Dans la partie 2, il a été expliqué que le Tarif Réel (AP) subit des revalorisations successives lors des échéances principales et que celles-ci sont globalement minorées par les revalorisations de Tarif Technique (TP) par le mécanisme d'ETP. La valeur du Tarif Réel est par conséquent très fortement liée à l'ancienneté du

contrat. L'ancienneté du contrat est alors une variable qui nous vient à l'esprit lorsque l'on cherche à identifier les critères pouvant être les plus discriminants sur l'AP/TP.

Les valeurs d'AP/TP moyen en fonction de l'ancienneté du contrat sont les suivantes :

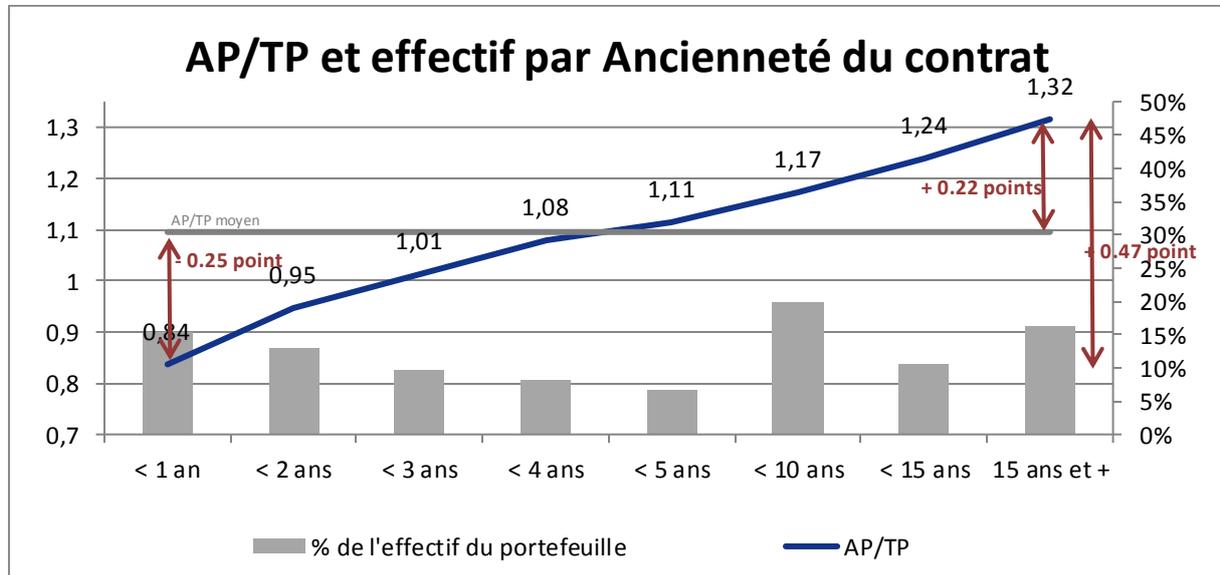


Figure 24: Analyse uni-variée de l'AP/TP et effectif par Ancienneté du contrat

On constate, comme anticipé, une croissance nette de l'AP/TP avec l'ancienneté du contrat. Les contrats d'âge inférieur à 2 ans ne sont pas rentables techniquement en moyenne puisque l'AP/TP moyen sur la modalité est inférieur à 1. Cependant la croissance de l'indicateur avec l'ancienneté du contrat permet de rattraper de la rentabilité sur les contrats plus anciens.

Les écarts d'AP/TP entre les différentes modalités d'âge du contrat sont de grande ampleur. Les contrats en affaire nouvelle (<1 an d'ancienneté) affichent un écart d'AP/TP de - 0.25 points par rapport à l'AP/TP moyen du portefeuille. Au contraire à partir d'une ancienneté de 4 ans, les contrats sont plus rentables que la moyenne. On va jusqu'à constater une rentabilité de + 0.22 points par rapport à la moyenne pour les contrats de 15 ans et plus. On justifie cela par le fait que l'on veut attirer les affaires nouvelles en leur accordant de fortes réductions commerciales à la souscription et lors des premières années du contrat. Elles seront ensuite réduites au cours de la vie du contrat, lors des échéances et avenants par le biais des revalorisations. Les affaires nouvelles et contrats jeunes sont ainsi financés par les contrats plus matures.

Cette croissance nette de l'AP/TP peut être expliquée par différents effets combinés :

- Les revalorisations successives qui ont été appliquées au Tarif Réel lors des échéances et avenants cumulées au cours de la vie d'un contrat, en moyenne largement plus conséquentes que les revalorisations appliquées au Tarif Technique, tendent à faire augmenter le Tarif Réel plus vite que le Tarif Technique. Ceci implique alors une croissance de l'AP/TP avec l'ancienneté du contrat.

- Pour veiller à la rentabilité du portefeuille, une surveillance des contrats est faite en permanence. Celle-ci engendre la résiliation ou la forte majoration des cotisations des contrats de mauvais risques et non rentables. Les contrats non rentables sont alors écartés du portefeuille au cours de leur vie.

- Il existe un grignotage des modulations commerciales au cours de la vie du contrat : Lors de la souscription d'un contrat, on accorde de fortes réductions aux assurés, puis celles-ci sont peu à peu de moins en moins prises en compte dans le Tarif Commercial suite aux échéances portefeuille ou avenant qui s'accumulent avec l'ancienneté du contrat. Par le mécanisme d'ETP, les grignotages viennent rebooster la croissance de l'AP/TP.

L'âge du conducteur augmente en même temps que l'ancienneté de son contrat. Par conséquent, plus un contrat est ancien, plus la moyenne d'âge des conducteurs augmente. Il y a, par conséquent, ici un effet de structure : la rentabilité s'améliorant avec l'âge du contrat, l'AP/TP augmente par conséquent avec l'âge du conducteur. Ceci explique la croissance de la rentabilité avec l'âge du conducteur principal.

L'âge du contrat est ainsi un critère particulièrement discriminant sur l'AP/TP. Ceci nous permet de dégager des profils de sur et sous-rentabilité : Les contrats les plus anciens constituent ainsi une source de sur-rentabilité, en particulier les contrats de plus de 5 ans qui affichent une rentabilité moyenne supérieure à la rentabilité globale du portefeuille. Les contrats de moins de 2 ans provoquent eux des pertes de rentabilité technique. Cette observation souligne le fait qu'il est donc essentiel de s'attacher à l'importance de la fidélisation des contrats, afin d'assurer la pérennité de la compensation des rentabilités entre les contrats d'âge jeune, qui n'ont pas encore permis de dégager du profit, et les contrats plus anciens, qui eux dégagent de la rentabilité.

- *Analyse de l'effet de la formule de garantie souscrite*

On s'intéresse désormais aux écarts de rentabilité selon la formule souscrite par l'assuré. Les cinq formules proposées par Allianz en assurance de quatre roues de particuliers sont définies dans le tableau suivant :

Garantie	Formule				
	CE Formule aux Tiers réduite	C1 « Formule aux Tiers »	CB Formule Intermédiaire réduite	C2 « Formule Intermédiaire »	C3 « Formule Tous Risques »
Responsabilité Civile					
Conducteur					
Assistance réduite					
Assistance essentielle					
Bris de Glace					
Catastrophe Naturelle					
Attentat					
Vol Incendie					
Dommages Tous Accidents					

  
**Formule de plus en plus couvrante**

Figure 25: Présentation des différentes formules de couvertures disponibles

La formule C1 correspond à la formule communément appelée « Formule aux Tiers » en assurance automobile, la formule C2 correspond à la « Formule Intermédiaire » et la formule C3 à la « Formule Tous Risques ». La formule CE est une formule C1 économique ayant une garantie assistance réduite. Elle est donc moins couvrante que la formule C1. La formule CB est une formule C1 complétée par les garanties Bris de Glace, Catastrophe Naturelle, Catastrophe Technologique et Attentat. Elle ne diffère de la formule C2 que par l'absence de garantie Vol et Incendie. Elle se situe donc entre la formule C1 et la formule C2 en termes de couverture.

Par souci d'effectif des contrats ayant souscrit la formule aux Tiers réduite (<4% du portefeuille), les variables CE et C1 ont été regroupées.

Après regroupement, on dispose du graphique suivant :

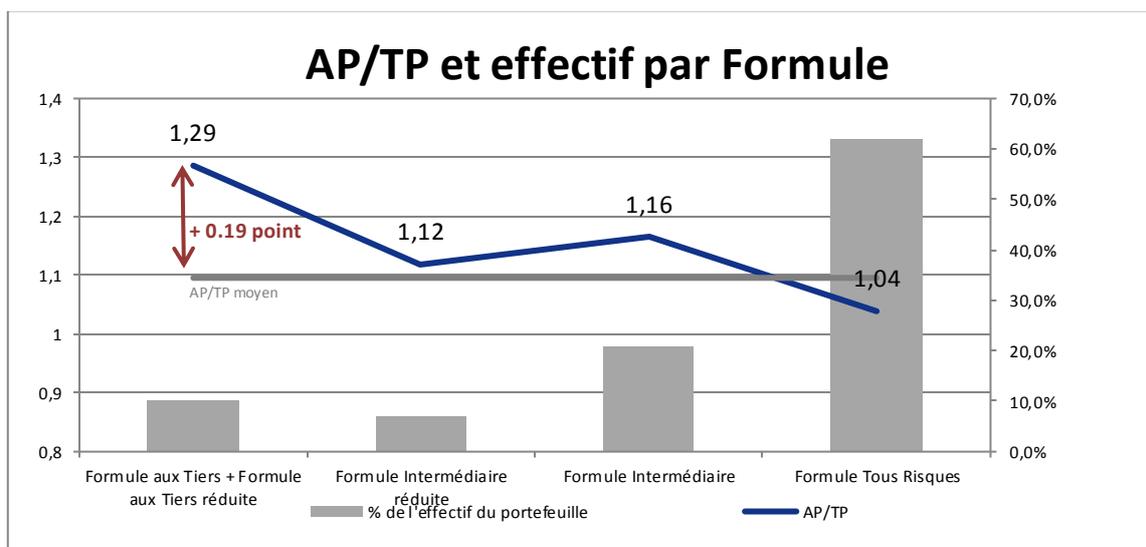


Figure 26: Analyse uni-variée de l'AP/TP et effectif par formule souscrite

Le choix de la formule souscrite paraît discriminant sur l'AP/TP. Les formules aux Tiers et aux Tiers réduite ont un AP/TP moyen relativement bon et apparaissent comme étant une possible source de rentabilité.

Ce constat est souvent observé et expliqué par le contexte commercial actuel présent en assurance automobile : En général les formules aux tiers sont rendues assez rentables relativement aux formules Tous Risques par une sur-tarification des formules aux Tiers et une sous tarification des formules Tous Risques. Le tarif proposé par l'ensemble de la concurrence pour une formule aux tiers est ainsi fixé à un niveau supérieur au coût du risque assurantiel. Au contraire pour les formules Tous Risques la valeur de la cotisation est fixée en-dessous du Tarif Technique. Par conséquent, l'AP/TP moyen est élevé pour les formules aux tiers au détriment d'un AP/TP plus équilibré pour les formules Tous Risques.

- Analyse de l'effet de l'offre double auto

Prenons le cas d'une variable ayant été relevée non discriminante sur l'AP/TP :

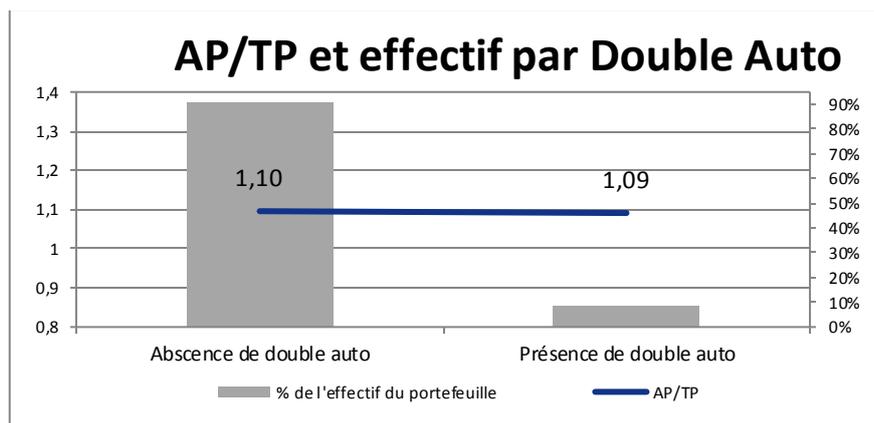


Figure 27: Analyse uni-variée de l'AP/TP et effectif par présence de l'offre double auto

Les clients Allianz assurant au moins deux véhicules quatre roues, dont l'un en forfait kilométrique avec relevé, bénéficient automatiquement de l'offre Double Auto.

L'offre double auto est un exemple de variable non discriminante sur l'AP/TP car les deux modalités de la variable affichent un AP/TP moyen quasiment identique. Cette variable ne sera donc pas conservée pour la suite de nos études.

- *Synthèse de notre analyse*

Une étude à plat du même type a été réalisée sur l'ensemble des variables de notre support, permettant ainsi de rendre compte de la discrimination de chacune des variables sur l'AP/TP et d'écartier un certain nombre de variables ayant été révélées comme non discriminantes ou non exploitables (variable ayant des modalités en effectif réduit et aucun regroupement pertinent ne pouvant être effectué). Les variables correspondant au coefficient bonus-malus, au sexe du conducteur principal, à la marque du véhicule assuré, à sa classe de réparation, à la présence d'un véhicule propre, à sa classe de vol, au garage, à la présence de l'option valeur plus et au taux protocole appliqué n'ont ainsi pas été retenues.

On relève que dans le cas où le sexe du conducteur avait été révélé comme étant discriminant sur l'AP/TP, légalement aucune action visant à favoriser les contrats rentables ou défavoriser les contrats non rentables n'aurait pu être entreprise puisqu'aucune différenciation tarifaire sur le sexe n'est autorisée en assurance automobile depuis 2011. La suppression de la variable sexe n'aura donc aucune conséquence car celle-ci n'aurait pu être utilisée pour piloter notre portefeuille.

Suite au nettoyage préalable de notre support, l'ensemble des variables conservées est le suivant :

Intitulé de la variable
<b>Données sur le conducteur principal assuré</b>
Age du conducteur
Catégorie Socio-Professionnelle
Age d'obtention du permis
Antécédent sinistre
<b>Données sur le véhicule assuré</b>
Ancienneté du véhicule
Classe de prix
Groupe SRA

Année d'acquisition
Energie
Carrosserie
Note de sécurité
<b>Données géographie</b>
Zone Responsabilité Civile
Zone Bris de Glace
Zone Vol/Incendie
Zone dommage
<b>Données sur le contrat</b>
Formule
Age du contrat
Fractionnement
Usage kilométrique
Pack Réparation
Franchise Bris de Glace
Franchise Vol
Franchise dommage
Clause tarifaire
<b>Données sur les modulations tarifaires</b>
Avantage client
Avantage prospect
Taux budget
Taux firme

Figure 28: Ensemble des variables conservées suite aux analyses uni-variées

## b) Analyse quantitative des liaisons entre variables

L'analyse uni-variée n'est pas suffisante pour déterminer l'influence d'une variable sur l'AP/TP : Elle ne permet pas de corriger l'effet des liens entre les variables. En effet, deux variables reflétant la même information ne peuvent être toutes deux intégrées au modèle car cela y ajouterait de la complexité sans pour autant améliorer son pouvoir prédictif. La variation d'un critère selon ses modalités peut être induite par une autre variable à laquelle elle est très corrélée. Il est donc essentiel de quantifier les corrélations existantes entre les différentes variables conservées.

- *Principe et éléments de théorie*

Les variables de notre support étant de type qualitatif, nous ne pouvons utiliser le coefficient de corrélation linéaire défini par  $\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$  habituellement utilisé pour mesurer le degré de corrélation entre deux variables quantitatives.

Afin de déterminer les liens quantitatifs existants entre les variables retenues qui sont de type qualitatif, l'outil statistique « V de Cramer » est utilisé. Celui-ci repose sur la distance du  $\chi^2$  qui mesure l'écart entre le tableau de contingence observé pour un couple de variables et celui que l'on observerait sous l'hypothèse d'indépendance entre ces deux variables. Si la distance du  $\chi^2$  est nulle, alors les variables sont indépendantes

et leur degré de liaison est nul. Cependant, le critère du  $\chi^2$  dépendant fortement de la taille de l'échantillon et du nombre de modalités des deux variables, nous utilisons alors le V de Cramer défini par :

$$V = \frac{\sqrt{\chi^2}}{\sqrt{\chi^2_{max}}} = \frac{\sqrt{\chi^2}}{\sqrt{n(\min(p_1, p_2) - 1)}}$$

où  $n$  est le nombre d'observations,  $p_1$  le nombre de modalités possibles de la première variable et  $p_2$  le nombre de modalités de la deuxième variable.

Le V de Cramer a l'avantage de détecter la présence d'un lien entre deux variables qualitatives mais aussi de quantifier l'intensité de ce lien. Plus la valeur du V de Cramer est proche de 0%, plus il y a indépendance entre les deux variables et plus le V de Cramer est proche de 100%, plus la dépendance entre les variables est élevée. Les différents niveaux de corrélation peuvent être hiérarchisés de la manière suivante :

Valeur du V de Cramer	Force du lien statistique entre les variables
Entre 0% et 20%	Peu corrélées, indépendantes
Entre 20% et 40%	Moyennement corrélées
Entre 40% et 80%	Corrélées
Entre 80% et 100%	Très corrélées, colinéarité

Figure 29: Interprétation de la force du lien entre deux variables à partir du V de Cramer

- *Etudes des corrélations inter-variables*

Pour mesurer les corrélations entre les potentielles variables explicatives restantes, nous utilisons la valeur du V de Cramer pour chaque couple de variable. Les résultats sont organisés sous forme d'un triangle de corrélations présenté en annexe.

Lorsque certaines variables sont trop corrélées, on ne peut conserver l'ensemble des variables car cela accorderait un trop gros poids à ces variables lors de la modélisation. Les solutions sont diverses : on peut soit croiser les variables trop corrélées si elles ont un nombre raisonnable de modalité, soit on peut choisir de conserver une de ces deux variables en la sélectionnant avec bon sens. On considérera que deux variables sont trop fortement corrélées pour être toutes deux conservées si elles ont un V de Cramer supérieur à 40%.

Le graphique ci-dessous permet de visualiser les liens les plus forts (V de Cramer supérieur à 40%) existants entre les variables. Plus le trait est épais, plus le lien entre les deux variables est fort.

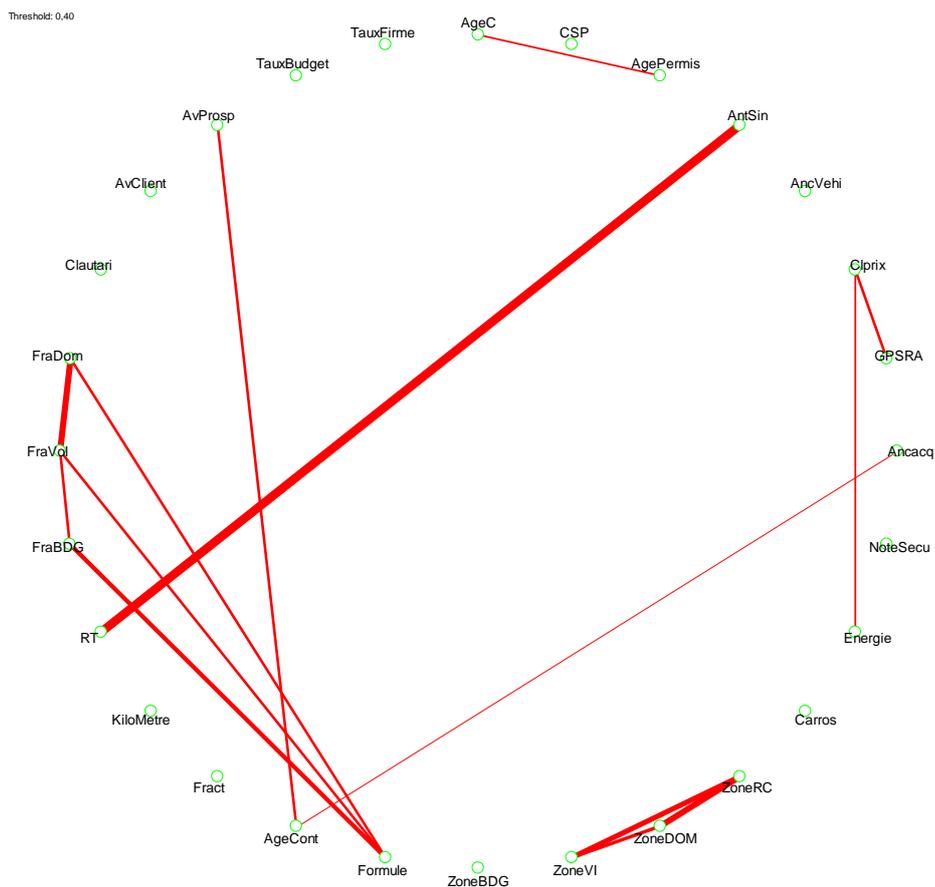


Figure 30: Visualisation des principaux liens entre les variables du support ( $V$  de Cramer  $> 0.4$ )

Le triangle des corrélations nous informe que les valeurs des franchises « FraBDG », « FraVol » et « FraDOM » ont un très fort  $V$  de Cramer entre elles et avec le choix de la « Formule ». Ceci s’explique par le fait que seules les garanties souscrites dans la formule pourront avoir une franchise non nulle. On choisit donc de garder uniquement la variable « Formule » qui est plus générale. De même les différentes zones « ZoneBDG », « ZoneVI » et « ZoneDOM » sont très corrélées. Ceci est explicable par le fait que les zones risquées en bris de glace sont plus ou moins semblables aux zones risquées en vol/incendie ou dommage tous accidents, et vis-versa. La variable « ZoneRC » n’est cependant pas trop corrélée avec les autres zones. On gardera ainsi la variable « ZoneRC » ainsi que la variable « ZoneBDG » qui est la garantie la plus souvent souscrite. L’âge du conducteur et l’âge du permis sont très fortement corrélées. Nous allons choisir de conserver plutôt l’âge du conducteur que l’ancienneté du permis car, comme nous allons le voir dans le paragraphe suivant, celui-ci est plus discriminant sur l’AP/TP.

Ces analyses de corrélations entre variables nous ont ainsi permis de faire un tri des données et de considérer un ensemble de variables peu corrélées entre elles.

Pour conclure ce paragraphe, voici une synthèse des variables conservées suite à cette étude pour la modélisation de l’AP/TP :

<b>Intitulé de la variable</b>
<b>Données sur le conducteur principal assuré</b>
Age du conducteur
Catégorie Socio-Professionnelle
Antécédent sinistre
<b>Données sur le véhicule assuré</b>
Ancienneté du véhicule
Groupe SRA
Année d'acquisition
Energie
Carrosserie
Note de sécurité
<b>Données sur la géographie</b>
Zone Bris de Glace
Zone Vol/Incendie
<b>Données sur le contrat</b>
Formule
Age du contrat
Fractionnement
Usage kilométrique
Pack Réparation
Clause tarifaire
<b>Données sur les modulations tarifaires</b>
Avantage client
Avantage prospect
Taux budget
Taux firme

- *Etudes des corrélations entre variables et AP/TP*

Le V de Cramer est également utilisé afin de quantifier l'intensité de la corrélation des variables avec l'AP/TP. Le V de Cramer considérant le lien entre deux variables qualitatives, nous avons ainsi créé des classes très fines de dixième d'AP/TP.

Variable	V de Cramer avec l'AP/TP
TauxFirme	34,5%
AvProsp	32,2%
Ancacq	29,0%
AgeCont	23,8%
NoteSecu	23,5%
ClauTari	19,5%
AntSin	17,9%
FraBDG	17,5%
TauxBudget	17,2%
Formule	16,6%
FraVol	16,1%
FraDom	16,0%
AncVehi	15,0%
Energie	14,0%
Fract	12,6%
CSP	11,2%
AvClient	10,6%
ZoneVI	10,2%
KiloMetre	10,0%
ZoneDOM	9,4%
AgeC	9,1%
ZoneRC	8,4%
Clprix	8,1%
GPSRA	7,6%
AgePermis	5,9%
Carros	5,6%
ZoneBDG	2,3%

Figure 31: V de Cramer des variables du support avec l'AP/TP

Le V de Cramer permet de vérifier que les variables du support sont bien corrélées avec notre indicateur de rentabilité.

On constate que les modulations commerciales ressortent en tête des corrélations avec l'AP/TP. En effet, le taux firme et l'avantage prospect se trouvent être les variables les plus corrélées à l'AP/TP, soulignant ainsi que les modulations commerciales creusent l'écart au Tarif Technique. On retrouve également, dans les variables fortement corrélées à l'AP/TP, des variables d'ancienneté telle que l'ancienneté d'achat du véhicule et l'ancienneté du contrat. Ceci confirme les observations faites en analyse uni-variée lorsque l'on avait constaté de forts écarts de Tarif Réel au Tarif Technique en fonction de l'ancienneté du contrat. Enfin, la formule et la clause tarifaire souscrite nous font part d'une importante corrélation avec la rentabilité technique individuelle du contrat.

### c) Test de significativité de l'analyse de la variance

L'analyse de la variance (ANOVA) nous permet de juger de la pertinence d'une variable sur la modélisation de l'AP/TP. Le principe général de l'analyse de la variance vise à tester les différences significatives entre les moyennes d'AP/TP sur les différentes modalités d'une variable explicative considérée. Le test est alors le suivant :

$H_0$ : les moyennes d'AP/TP de toutes les modalités sont égales

$H_1$ : au moins deux groupes ont des moyennes d'AP/TP différentes

Le test repose sur le rapport entre la variance totale et la variance à l'intérieur d'une modalité

On définit alors les quantités SCE (Somme des Carrés des Ecart), SCE résiduelle et SCE factorielle par :

*Variance totale = Variance inter-groupe + Variance intra-groupe*

$$SCE = SCE_{factoriel} + SCE_{résiduel}$$

$$\sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 = \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2$$

Où  $p$  est le nombre de modalités prises par une variable et  $n_i$  l'effectif de la modalité  $i$ .

On calcule ensuite  $F = \frac{SCE_{factoriel} / \frac{p-1}{p-1}}{SCE_{résiduel} / \frac{n-p}{n-p}}$ . Sous l'hypothèse nulle  $H_0$ ,  $F$  suit une loi de Fisher de paramètres

$F(p-1, n-p)$ .

Pour une variable donnée, si sa p-value est supérieure à 5% alors on ne peut conclure et si sa p-value est inférieure à 5% alors l'hypothèse  $H_0$  est rejetée en faveur de  $H_1$ : la variable est significative par rapport à l'AP/TP.

Le test a été réalisé sur l'ensemble des variables restantes de notre support à partir de la procédure « Proc ANOVA » de SAS. L'ensemble des variables explicatives à ce jour retenues ont été confirmées comme étant significatives par rapport à l'AP/TP.

### d) Analyse multi-variées sur l'AP/TP et confortation des résultats uni-variés

Afin de confirmer les observations obtenues en uni-varié quant aux influences de variables sur l'AP/TP, une Analyse des Correspondances Multiples est réalisée. Etant donné l'importance du nombre de variables disponibles dans notre support initial, nous allons limiter notre analyse à un certain nombre de variables. En effet, considérer un nombre important de variables pourrait engendrer l'obtention d'un nuage sur-condensé en information et serait ainsi difficilement interprétable. Nous allons ainsi retenir les variables qui ont été révélées discriminantes sur l'AP/TP en partie 1.

- Principe et éléments de théorie de l'ACM

L'Analyse des Correspondances Multiples ou ACM est une technique d'analyse de données permettant de transformer un nuage de variables corrélées en de nouvelles variables indépendantes en synthétisant l'information en un nombre plus limité de dimensions. Les nouvelles variables formées sont appelées

composantes principales ou axes. Cet outil va permettre d'émettre des hypothèses sur les modalités qui seront sous-profitables ou au contraire sur-profitables, permettant de confirmer les résultats relevés en analyse univariée.

Si on note  $n$  le nombre d'individus de notre base,  $p$  le nombre total de variables considérées et  $m_k$  le nombre de modalités de la variable  $k$  pour  $k = \{1, \dots, p\}$ , alors le nombre total de modalités est  $m = \sum_{k=1}^p m_k$ . L'ACM va être réalisée à partir du tableau disjonctif complet de notre base, matrice de taille  $n \times m$  défini par :

$$X = [X_{i,j}] \text{ où } x_{i,j} = \begin{cases} 1 & \text{si l'individu } i \text{ possède la modalité } j \\ 0 & \text{sinon} \end{cases}$$

A partir de cette matrice, nous allons construire une seconde matrice appelée tableau de Burt  $B$  donnée par  $B = X^T X$ . Les axes factoriels de l'ACM sont enfin définis comme étant les vecteurs propres de la matrice  $\frac{1}{p} B D^{-1}$  où  $D$  est la diagonale de  $X$ . L'inertie associée à l'axe  $k$  est définie comme étant la valeur propre associée à ce vecteur propre.

L'AP/TP va être placé en variable supplémentaire, c'est-à-dire que celui-ci ne fera pas partie des variables entrant dans la création des axes de notre ACM, afin d'analyser les corrélations existantes entre les différentes modalités de variables et l'AP/TP. L'AP/TP apparaîtra sous la forme de trois classes d'AP/TP : AP/TP sous-profitables (AP/TP < 0.8), AP/TP sur profitables (AP/TP > 1.2) et AP/TP neutres.

- Application et confrontation avec les résultats uni-variés

L'Analyse des Correspondances Multiples a été effectuée sous SAS à partir de la procédure « Proc Corresp ». L'ensemble des variables ayant précédemment été relevées comme étant discriminantes sur l'AP/TP sont placées en explicatives tandis que l'AP/TP est placé en supplémentaire. L'indicateur n'intervient ainsi pas dans la modélisation puisque c'est la variable que l'on souhaite expliquer.

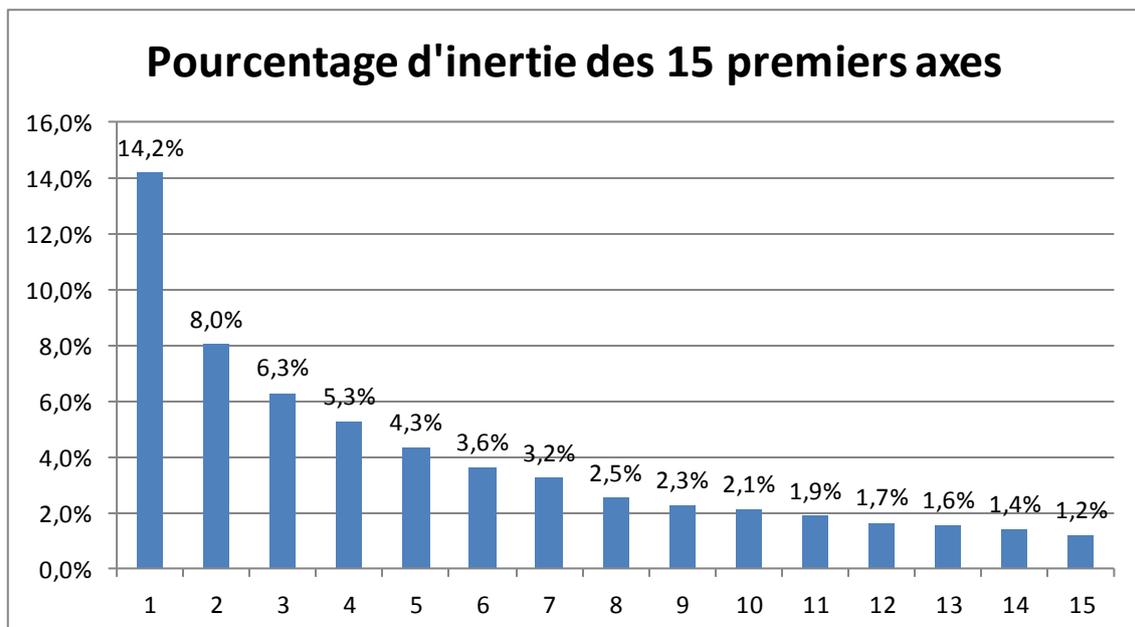


Figure 32: Inertie des 15 premiers axes de l'ACM

Les résultats de l'ACM ne sont pas vraiment satisfaisants étant donné que pour pouvoir conserver 50% de l'information contenue dans nos données initiales il faudrait analyser au moins les 10 premiers axes, alors que dans certains cas on arrive à expliquer 80% de l'information avec 5 axes uniquement environ. Dans notre cas, les variables ne sont pas suffisamment explicatives pour expliquer 80% de l'information.

L'analyse d'une dizaine d'axes serait trop volumineuse, nous nous sommes alors ainsi restreints à l'étude d'un nombre plus réduit d'axes. Il faudra alors considérer nos résultats avec du recul étant donné que les axes étudiés ne constitueront pas de manière cumulée l'essentiel de l'information contenue dans la base.

Le critère du coude permet de déterminer le nombre d'axes à retenir pour notre ACM, en ne conservant que les axes dont l'apport en inertie est important. Au vu du graphique précédent, les quatre premiers axes seront considérés.

On dit qu'une modalité contribue à la création de l'axe si sa contribution est supérieure à la contribution moyenne d'une modalité qui est égale à  $1/\text{nombre total de modalités} = 1/127 \approx 0.008$  ici.

<b>Interprétation de l'axe 1</b>			
<b>Variable concernée</b>	<b>Côté négatif = Rentabilité Positive</b>	<b>Côté positif = Rentabilité Négative</b>	<b>Tendance globale</b>
Ancienneté du véhicule	Véhicule ancien (10 ans et plus)	Véhicules récents (<10 ans)	X
Note de sécurité	Note nulle ou mauvaise sur la sécurité du véhicule (O,P)	Bonne note de sécurité du véhicule (N)	
Formule	Formule moins complète (C1/CE, CB, C2)	Formule plus complète (C3)	X
Franchise Vol	Garantie Vol non souscrite	Garantie Vol souscrite	
Franchise Dommage	Garantie Dommage non souscrite	Garantie Dommage non souscrite	

<b>Interprétation de l'axe 2</b>			
<b>Variable concernée</b>	<b>Côté négatif = Rentabilité Positive</b>	<b>Côté positif = Rentabilité Négative</b>	<b>Tendance globale</b>
Age du conducteur	60 ans et plus	Moins de 60 ans et Personnes morales	X
Ancienneté du contrat	5 ans et plus	Moins de 5 ans	X
Ancienneté acquisition de véhicule	10 ans et plus	Moins de 10 ans	X
Zone RC	Zones moins risquées (1 à 10)	Zones plus risquées (11 à 26)	X
Zone Dommage	Zones moins risquées (1 à 6)	Zones plus risquées (7 à 26)	X
Zone Vol/Incendie	Zones moins risquées (1 à 6)	Zones plus risquées (7 à 26)	X
Franchise Bris de Glace	Franchise nulle	Franchise non souscrite et non nulle	
Franchise Vol	Franchise nulle ou sous-franchise	Garantie non souscrite et sur-franchise	
Franchise Dommage	Franchise nulle ou sous-franchise	Garantie non souscrite et sur-franchise	

Interprétation de l'axe 3			
Variable concernée	Côté négatif = Rentabilité Négative	Côté positif = Rentabilité Positive	Tendance globale
Age du conducteur	Moins de 60 ans et Personnes morales	60 ans et plus	X
Age d'obtention du permis	18 ans	18 ans et plus	X
Energie	Autres dont Gazole	Essence	
Fractionnement	Mensuel	Annuel, Semestriel, Trimestriel	
Clause tarifaire	Conduite couple	Autres	

Interprétation de l'axe 4			
Variable concernée	Côté négatif = Rentabilité Négative	Côté positif = Rentabilité Positive	Tendance globale
Catégorie socio-professionnelle	Fonctionnaire, Salarié	Artisan, Inactif, Autre	
Note de sécurité	Négative (Bonne note)	Positive, Nulle	
Clause tarifaire	Conduite couple	Autre	

La croix en colonne 3 fait référence aux variables continues dont la rentabilité est monotone avec les modalités, c'est-à-dire que la rentabilité est d'autant plus élevée que sa coordonnée sur l'axe considéré est élevée ou au contraire faible. Les variables n'ayant pas de croix sont alors des critères tels que des profils de rentabilité ont été dégagés sur certaines modalités uniquement.

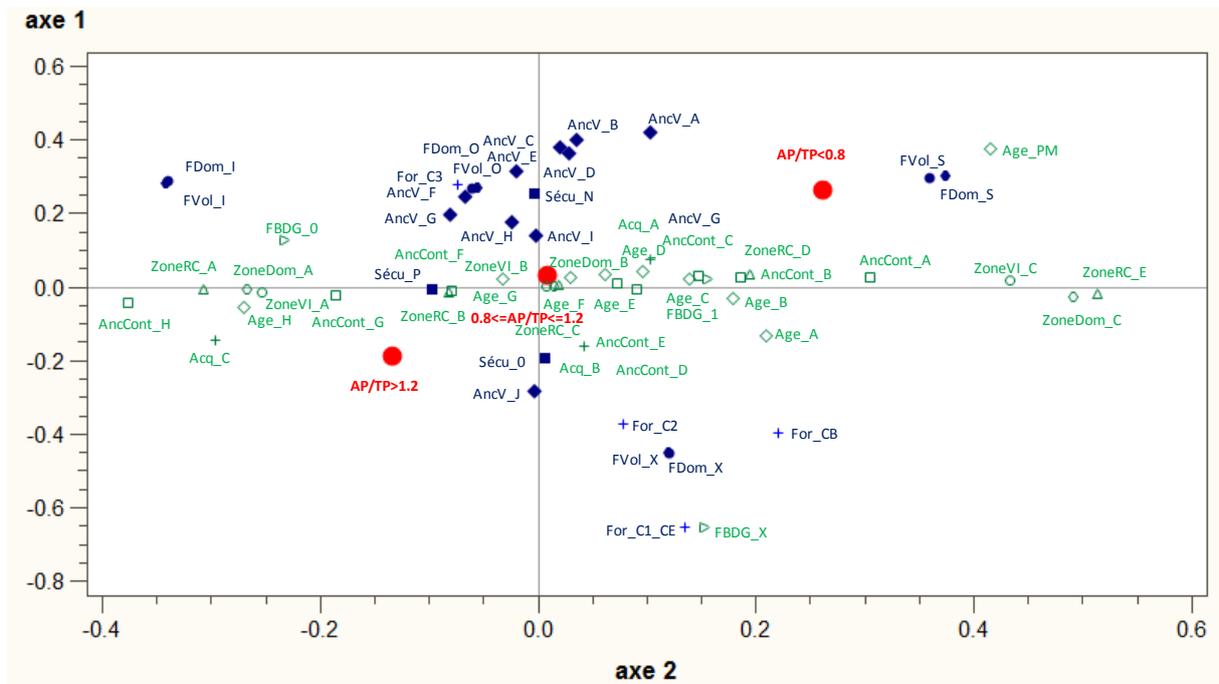


Figure 33: Représentation des modalités sur les axes 1 et 2 de notre ACM

- *Interprétation de l'axe 1*

L'axe 1 est un axe clé puisqu'il constitue à lui seul 14.2% de l'inertie du nuage. C'est lui qui va alors expliquer au mieux l'information contenue dans nos données. Il nous donne des informations sur l'ancienneté du véhicule, sur sa note de sécurité ainsi que sur la formule souscrite. On note que le montant des franchises Vol et Dommage souscrites s'est également vu caractériser l'axe 2. Cependant, ces variables étant très fortement liées à la formule souscrite, leurs coordonnées se retrouvent alors très proches de la formule souscrite correspondante et l'information contenue n'est pas tant sur le montant de la franchise (franchise standard, sur-franchise, sous-franchise) mais plutôt sur la présence ou absence de la garantie souscrite. Ces modalités ne seront alors pas plus analysées.

Sur l'axe 1, le côté négatif est associé à des modalités de contrats qui s'avèrent dégager en moyenne de la rentabilité. Au contraire, le côté positif de l'axe 1 est associé à une fuite de profitabilité des contrats car il s'agit d'AP/TP inférieurs à 1.

En termes de formule souscrite, on retrouve du côté négatif les formules moins complètes, associées à un bon niveau de rentabilité. Le contrat est même d'autant plus rentable que la formule est complète. On retrouve également les modalités des franchises dommage et Vol non souscrites, confirmant alors le fait que les formules moins couvrantes sont plus rentables techniquement. Dans les contrats rentables, on retrouve également les véhicules anciens (de 10 ans et plus). Les véhicules ayant une note de sécurité neutre ou positive, c'est-à-dire avec un équipement de sécurité pauvre, sont associés à une certaine rentabilité.

Du côté positif de l'axe 1, associé à une sous-rentabilité des contrats, on retrouve en particulier les contrats d'ancienneté inférieure à 10 ans. Les modalités ont une coordonnée d'autant plus élevée sur l'axe 1 que le contrat est jeune, ce qui signifie que plus le contrat est jeune plus il est révélé rentable. Les contrats ayant une note de sécurité négative, donc bien équipés, font partie des profils de contrats non rentables sur l'axe 1. La modalité formule C3 a une coordonnée élevée sur l'axe 1, les contrats ayant souscrit pour une formule C3 sont alors détectés comme étant des fuites de profitabilité.

- *Interprétation de l'axe 2*

L'axe 2 fait ressortir des informations quant à l'ancienneté du contrat et à l'acquisition du véhicule, à l'âge du conducteur ainsi qu'au risque des zones des différentes garanties.

Sur l'axe 2, le côté négatif est associé à des modalités de contrats sur-rentables. Au contraire, le côté positif de l'axe 2 est associé à des profils de contrats non rentables.

Du côté négatif, on trouve les contrats anciens, de plus de 5 ans et les véhicules dont l'acquisition date de plus de 10 ans. On y trouve également les contrats dont les zones de risque RC, DOM et VI sont faibles. Ces modalités correspondent aux contrats sur-tarifés et considérés comme étant plus rentables techniquement.

Du côté positif, associé aux contrats moins rentables techniquement, on retrouve les contrats plus récents, dont l'ancienneté est inférieure à 10 ans. La coordonnée sur l'axe 2 des modalités d'âge du contrat est même d'autant plus élevée que le contrat est jeune, un contrat s'avère d'autant moins rentable qu'il est récent.

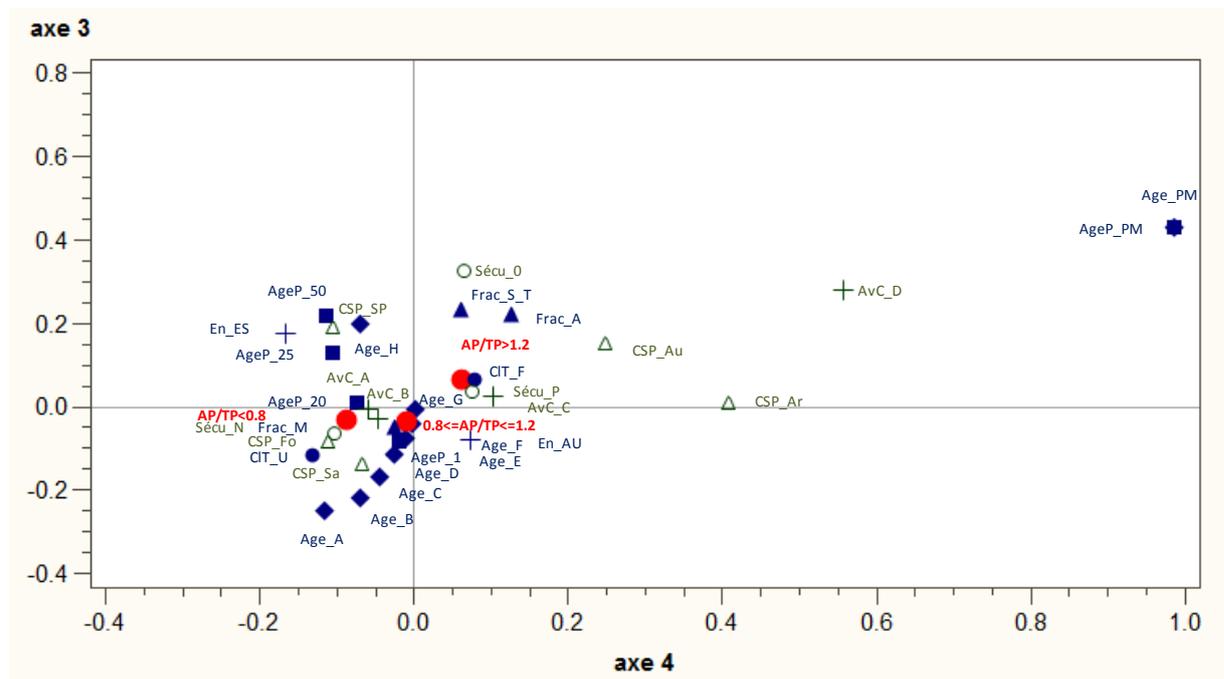


Figure 34: Représentation des modalités sur les axes 3 et 4 de notre ACM

On constate que dès l'axe 3 et 4, les modalités d'AP/TP se retrouvent relativement proches les unes des autres dans le nuage formé par les nouveaux axes. Cela complique alors notre analyse et confirme la particularité de la structure de notre base rendant difficile l'analyse de l'influence des variables sur l'AP/TP.

- *Interprétation de l'axe 3*

L'axe 3 fait ressortir des informations sur l'âge du conducteur, l'âge d'obtention du permis, l'énergie d'alimentation du véhicule, la clause tarifaire souscrite ainsi que sur le mode de fractionnement de la cotisation.

On retrouve du côté positif les modalités s'avérant caractériser les profils dégagant de la rentabilité. Au contraire, du côté négatif, on retrouve les modalités techniquement peu rentables.

Du côté positif, associé à une bonne rentabilité des contrats, on retrouve les conducteurs de plus de 70 ans et ceux ayant obtenu leur permis après 20 ans. Du côté positif, on retrouve également les véhicules alimentés à l'essence, les modalités de fractionnement annuel, trimestriel et semestriel.

Du côté négatif, on retrouve les conducteurs jeunes et ayant obtenu leur permis entre 18 et 20 ans. Le fractionnement mensuel est associé à une rentabilité plutôt mauvaise.

- *Interprétation de l'axe 4*

Sur l'axe 4, on obtient des informations quant à la rentabilité des catégories socio-professionnelles, à la note de sécurité du véhicule ainsi qu'à la clause tarifaire.

Les modalités se trouvant du côté positif de l'axe sont celles qui sont corrélées avec des valeurs d'AP/TP favorables tandis que le côté négatif caractérise les modalités liées à une faible valeur d'AP/TP.

Du côté positif on retrouve les véhicules ayant obtenu une mauvaise note de sécurité et ayant souscrit une clause tarifaire différente de la conduite couple. Ceux-ci sont associés à une bonne rentabilité.

Au contraire, du côté négatif, on retrouve les véhicules bien équipés du point de vue de la sécurité et les contrats associés à une conduite couple.

- *Conclusion sur notre ACM et comparaison avec les analyses uni-variées*

Initialement, l'idée de l'ACM était de construire des profils type d'assurés associés à une sur ou sous-rentabilité. Cependant, les résultats de notre ACM et la composition du nuage formé ne nous ont pas permis d'aboutir à de tels résultats.

Bien que la construction des axes de l'ACM ne soit pas aussi pertinente que souhaitée, l'ACM nous a en général permis de conforter les résultats obtenus en uni-varié quant à la discrimination des variables sur l'AP/TP en tenant compte des corrélations qui existent entre celles-ci. En effet, nous avons retrouvé des résultats similaires aux analyses uni-variées. Les profils qui ressortent comme étant rentables ou au contraire peu rentables concernent la note de sécurité, des critères d'ancienneté tels que l'ancienneté du contrat ou du véhicule ou la formule souscrite. Cependant on observe grâce à l'ACM très peu la discrimination des modulations tarifaires sur l'AP/TP.

Ce chapitre nous a ainsi permis de faire une première analyse de notre portefeuille à travers l'AP/TP. Ceci nous a également permis d'affiner, d'une part, le regroupement des modalités de l'ensemble de nos variables et, d'autre part, d'identifier les variables les plus discriminantes sur l'AP /TP et celles corrélées entre elles, nous permettant ainsi de nous ramener à un support pertinent. Nous pouvons désormais procéder à la modélisation de l'indicateur de rentabilité.

## CHAPITRE 3 : Modélisation de l'indicateur AP/TP par les Modèles Linéaires Généralisés

## Partie 1 : Quelques notions théoriques utiles à la modélisation

Dans cette partie, divers rappels théoriques vont être effectués de manière à poser les bases de la modélisation qui sera faite par la suite. Un certain nombre de résultats obtenus dans cette partie seront par la suite directement utilisés dans la partie 2 de ce chapitre.

### a) Rappels sur certaines lois usuelles

L'AP/TP est une variable strictement positive. Les lois usuelles permettant de modéliser ce type de variable en actuariat sont, entre autre, les lois Exponentielle, Log-normale, Gamma et Weibull. Nous allons ainsi rappeler certaines propriétés sur ces lois classiques sélectionnées. La loi Normale va également être présentée car celle-ci interviendra lors des tests d'adéquation de la loi des AP/TP à la loi Log-normale. Pour chacune des lois, les deux premiers moments seront également calculés car ceux-ci interviendront par la suite dans les calculs d'estimation de paramètres de lois.

#### -Loi exponentielle $\varepsilon(\lambda)$

La loi exponentielle dépend d'un unique paramètre  $\lambda > 0$ .

Sa densité et sa fonction de répartition sont définies par :  $f_\lambda(x) = \lambda e^{-\lambda x}$  et  $F_\lambda(x) = 1 - e^{-\lambda x}$ .

Ses moments sont donnés par :  $m_k = E(X^k) = \frac{k!}{\lambda^k}$  pour tout  $k \in \mathbb{N}$ .

On a donc de plus :  $E(X) = m_1 = \frac{1}{\lambda}$  et  $V(X) = m_2 - m_1^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ .

#### -Loi Normale $N(\mu, \sigma^2)$

La loi est caractérisée par les paramètres  $\mu \in \mathbb{R}$  et  $\sigma > 0$ , qui représentent respectivement l'espérance et l'écart type de la loi.

La densité et la fonction de répartition de la loi normale sont données par :  $f_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ .

et  $F_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy = \Phi\left(\frac{x-\mu}{\sigma}\right)$  où  $\phi$  est la fonction de répartition de la loi normale centrée réduite définie par  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$ .

Les moments centrés d'ordre  $k$  de la loi Normale  $N(\mu, \sigma^2)$  sont donnés par :

$$\begin{cases} \mu_{2k} = E((X - \mu)^{2k}) = \sigma^{2k} \frac{(2k)!}{2^k k!} \\ \mu_{2k+1} = E((X - \mu)^{2k+1}) = 0 \end{cases}$$

Ces moments centrés peuvent être décomposés en une fonction des moments d'ordres inférieurs ou égaux à  $k$ , donc le moment d'ordre  $k$  peut s'obtenir à partir des moments centrés d'ordres inférieurs ou égaux à  $k$ . Les premiers moments de la loi Normale  $N(\mu, \sigma^2)$  sont ainsi donnés par :

$$m_1 = E(X) = E(X - \mu) + \mu = \mu$$

$$m_2 = E(X^2) = E((X - \mu)^2) + 2\mu E(X) - \mu^2 = \sigma^2 + 2\mu^2 - \mu^2 = \sigma^2 + \mu^2$$

De la même manière, on obtient :  $m_3 = 3\mu\sigma^2 + \mu^3$  et  $m_4 = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$ .

On peut donc vérifier que  $E(X) = \mu$  et  $V(X) = m_2 - m_1^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$ .

### -Loi Log-normale $\text{LogN}(\mu, \sigma^2)$

La loi log-normale dépend des paramètres  $\mu \in \mathbb{R}$  et  $\sigma > 0$ .

La variable aléatoire  $X$  suit une loi log-normale  $\text{LogN}(\mu, \sigma^2)$  si et seulement si la variable aléatoire  $\ln X$  suit la loi normale  $N(\mu, \sigma^2)$ .

Sa densité et sa fonction de répartition sont définies par les expressions suivantes :

$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}$  et  $F_{\mu, \sigma^2}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$  où  $\Phi$  est la fonction de répartition de la loi normale centrée réduite.

Ses moments sont donnés par :  $m_k = e^{k\mu + \frac{k^2}{2}\sigma^2}$  donc  $E(X) = e^{\mu + \frac{\sigma^2}{2}}$  et  $V(X) = e^{2\mu} \cdot e^{\sigma^2} \cdot (e^{\sigma^2} - 1)$

### -Loi Gamma $\Gamma(\alpha, \lambda)$

Le paramètre  $\alpha$  de la loi gamma est appelé « paramètre de forme » tandis que le paramètre  $\lambda$  est appelé « paramètre d'échelle ». Ils sont tous deux strictement positifs.

Sa densité et sa fonction de répartition sont définis pour  $x > 0$  par :  $f_{\alpha, \lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \mathbb{1}_{x>0}$  et  $F_{\alpha, \lambda}(x) = \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda t} t^{\alpha-1} dt$  avec  $\Gamma$  la fonction gamma définie par  $\Gamma(y) = \int_0^{+\infty} e^{-t} t^{y-1} dt$  pour  $y > 0$ .

Ses moments sont donnés par :  $m_k = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\lambda^k}$ .

On a ainsi :  $E(X) = \frac{\alpha}{\lambda}$  et  $V(X) = \frac{\alpha}{\lambda^2}$

### -Loi de Weibull $W(\alpha, \lambda)$

Le paramètre  $\alpha$  de la loi de Weibull est appelé « paramètre de forme » tandis que  $\lambda$  est appelé « paramètre d'échelle ». Ils sont tous deux strictement positifs.

Sa densité est définie par  $f_{\alpha, \lambda}(x) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} e^{-\left(\frac{x}{\lambda}\right)^\alpha}$  et sa fonction de répartition par

$$F_{\alpha, \lambda}(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^\alpha}$$

Ses moments sont donnés par  $m_k = \lambda^k \Gamma\left(\frac{k}{\alpha} + 1\right)$  avec  $\Gamma$  la fonction gamma, donc  $E(X) = \lambda \Gamma\left(\frac{1}{\alpha} + 1\right)$  et  $V(X) = \lambda^2 \left[ \Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right) \right]$

## **b) Méthodes d'estimation des paramètres**

Afin de tester l'adéquation de la loi suivie par notre échantillon d'AP/TP à une des lois classiques présentées dans le paragraphe précédent, nous devons au préalable estimer les paramètres de ces lois classiques à tester à partir de l'échantillon dont nous disposons.

Soit  $x_1, x_2, \dots, x_n$  un échantillon de  $n$  réalisations indépendantes et identiquement distribuées selon une loi paramétrique de paramètre  $\theta$ .

Il existe deux méthodes d'estimation des paramètres des lois principalement utilisées en statistique non paramétrique : la méthode des moments et la méthode du maximum de vraisemblance.

La méthode des moments :

La méthode des moments consiste à égaliser les moments théoriques aux moments empiriques. Cette procédure se justifie par la loi des grands nombres qui implique que les moyennes empiriques convergent vers les moments théoriques lorsque  $n$  est grand.

Si on note  $m_k = E(X^k)$  le moment d'ordre  $k$  de la loi  $X$ , la méthode des moments nous donne ainsi l'estimateur  $\widehat{m}_k = \sum_{i=1}^n x_i^k$  pour tout  $k$ . Dans le paragraphe précédent, nous avons vu que les différents moments peuvent en général être exprimés à partir des paramètres de chacune des lois.

La détermination des valeurs des paramètres est alors obtenue en résolvant un système d'équations.

La méthode du maximum de vraisemblance :

Soit  $X$  une loi paramétrique de paramètre  $\theta \in \mathbb{R}^p$ . Soit  $\widetilde{X}_n = (X_1, X_2, \dots, X_n)$  un  $n$ -échantillon de loi  $X$ .

La fonction de vraisemblance associée à l'échantillon  $\widetilde{X}_n$  est alors définie comme :

$$L: (x_n, \theta) \mapsto L(x_n, \theta) = \prod_{i=1}^n f_{\theta}(x_i)$$

La méthode du maximum de vraisemblance permet d'estimer le paramètre  $\theta \in \mathbb{R}^p$  tel que son estimation  $\widehat{\theta}$  maximise la vraisemblance de l'observation  $\widetilde{x}_n$ .

On a donc que  $\widehat{\theta}$  est solution des équations suivantes :

$$\frac{\partial L(x_n, \theta)}{\partial \theta_k} = 0 \text{ pour } k \in [1, \dots, p]$$

Application aux lois usuelles retenues:

-Loi exponentielle  $\varepsilon(\lambda)$

Soit  $x_1, x_2, \dots, x_n$  un  $n$ -échantillon d'observations de loi  $\varepsilon(\lambda)$ . Les méthodes des moments et du maximum de vraisemblance donnent le même estimateur du paramètre  $\lambda$  égal à  $\widehat{\lambda} = \frac{1}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}_n}$  où  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$  est la moyenne empirique de l'échantillon.

-Loi Normale  $N(\mu, \sigma^2)$

Soit  $x_1, x_2, \dots, x_n$  un  $n$ -échantillon d'observations de loi  $N(\mu, \sigma^2)$ . Les méthodes des moments et du maximum de vraisemblance donnent les mêmes estimateurs des paramètres  $\mu$  et  $\sigma^2$  :

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n \text{ et } \widehat{\sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \text{ où } \bar{x}_n \text{ est la moyenne empirique de l'échantillon.}$$

### -Loi Log-normale $LogN(\mu, \sigma^2)$

Soit  $x_1, x_2, \dots, x_n$  un  $n$ -échantillon d'observations de loi  $LogN(\mu, \sigma^2)$ . Les méthodes des moments et du maximum de vraisemblance donnent les mêmes estimateurs des paramètres  $\mu$  et  $\sigma$  :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i \text{ et } \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu})^2$$

### -Loi Gamma $\Gamma(\alpha, \lambda)$

Soit  $x_1, x_2, \dots, x_n$  un  $n$ -échantillon d'observations de loi  $\Gamma(\alpha, \lambda)$ . La méthode des moments permet d'obtenir directement une estimation des paramètres  $\alpha$  et  $\lambda$  :

$$\hat{\alpha} = \frac{\overline{x_n^2}}{\overline{\sigma_n^2}} \text{ et } \hat{\lambda} = \frac{\overline{x_n}}{\overline{\sigma_n^2}} \text{ avec } \overline{x_n} \text{ et } \overline{\sigma_n^2} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x_n})^2 \text{ les moyennes et variances empiriques de l'échantillon.}$$

Cependant l'obtention de l'estimation par la méthode du maximum de vraisemblance est plus complexe. En définissant la loi Digamma  $\Psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ , l'estimateur du maximum de vraisemblance de  $\alpha$  est défini comme la solution de l'équation suivante :  $\ln(\hat{\alpha}) - \Psi(\hat{\alpha}) = \ln\left(\frac{\overline{x_n}}{(\prod_{i=1}^n x_i)^{1/n}}\right)$ .

On en déduit alors l'estimateur du maximum de vraisemblance de  $\beta$ , défini par :  $\hat{\beta} = \frac{\hat{\alpha}}{\overline{x_n}}$

### -Loi de Weibull $W(\alpha, \lambda)$

Soit  $x_1, x_2, \dots, x_n$  un  $n$ -échantillon d'observation de loi  $W(\alpha, \lambda)$ .

L'estimateur de la méthode des moments de  $\alpha$  est solution de l'équation suivante :

$$\ln \Gamma\left(\frac{2}{\alpha} + 1\right) - 2 \ln \Gamma\left(\frac{1}{\alpha} + 1\right) = \ln\left(1 + \frac{\overline{\sigma_n^2}}{\overline{x_n^2}}\right).$$

L'estimateur de la méthode des moments de  $\lambda$  vérifie alors :  $\hat{\lambda} = \frac{\overline{x_n}}{\Gamma\left(\frac{1}{\hat{\alpha}} + 1\right)}$ .

La méthode du maximum de vraisemblance donne les résultats suivants :

$$\alpha \text{ est solution de l'équation } \frac{1}{\alpha} = \frac{\sum_{i=1}^n x_i^\alpha \ln x_i}{\sum_{i=1}^n x_i^\alpha} - \frac{1}{n} \sum_{i=1}^n \ln x_i, \text{ qui permet de définir } \hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\hat{\alpha}}\right)^{\frac{1}{\hat{\alpha}}}$$

## **c) Tests d'adéquations aux lois usuelles**

Soit un échantillon de  $n$  données  $x_1, x_2, \dots, x_n$  indépendantes et identiquement distribuées selon une loi que l'on cherche identifier. On cherche ainsi à déterminer une loi classique permettant de modéliser la distribution empirique de cet échantillon. Nous disposons d'un certain nombre de méthodes permettant de juger de l'adéquation à une loi.

### Comparaisons graphiques

On peut simplement comparer graphiquement les allures des densités et des fonctions de répartition empiriques de notre échantillon avec celles des lois théoriques avec lesquelles on souhaite tester l'adéquation. La comparaison graphique est ainsi une indication simple et visuelle de la pertinence de l'ajustement d'une distribution à une distribution théorique.

### Le QQ plot et Le PP plot

- Le « QQ-plot » ou « Graph Quantile-Quantile » en français, permet de tester l'adéquation d'une loi empirique à une loi classique théorique en utilisant la notion de quantiles. Il compare en fait les quantiles de la loi empirique à ceux de la loi théorique testée. Le QQ-Plot porte une attention particulière à l'adéquation de la loi au niveau des extrêmes.

Soit  $x \in \mathbb{R}^+$ , on définit les fonctions  $x \mapsto y = F_{obs}(x)$

$$\text{et } y \mapsto z = F_{th}^{-1}(y)$$

où  $F_{obs}$  est la fonction de répartition empirique des observations et  $F_{th}$  est la fonction de répartition de la loi théorique.

Si l'adéquation entre les deux lois est bonne, on a alors  $F_{obs} \approx F_{th}$  donc  $z = F_{th}^{-1} \circ F_{obs}(x) \approx x$ . Donc plus les points sont proches de la première diagonale, plus l'ajustement semble pertinent.

Cependant, si la courbe est au-dessus de la diagonale, on surestime la réalité tandis que si la courbe est en-dessous de la diagonale, on sous-estime la réalité.

- Le « PP-plot » ou « Graph Pourcentage-Pourcentage » en français, permet de tester l'adéquation d'une loi empirique à une loi classique théorique en se concentrant sur l'essentiel de la distribution.

Soit  $x \in [0,1]$ , on définit les fonctions  $x \mapsto y = F_{obs}^{-1}(x)$

$$\text{et } y \mapsto z = F_{th}(y)$$

où  $F_{obs}^{-1}$  est l'inverse de la fonction de répartition empirique des observations et  $F_{th}$  est la fonction de répartition de la loi théorique.

Si l'adéquation entre les deux lois est bonne, on a alors  $F_{obs} \approx F_{th}$  donc  $z = F_{th} \circ F_{obs}^{-1}(x) \approx x$

Contrairement aux QQ-Plot, si la courbe est au-dessus de la diagonale, on sous-estime la réalité tandis que si la courbe est en-dessous de la diagonale, on surestime la réalité.

#### **d) Théorie sur les Modèles Linéaires Généralisés**

Le modèle linéaire généralisé ou GLM (Generalized Linear Model) est un modèle statistique classique qui consiste à étudier la liaison entre une variable  $Y$  dite « variable réponse » et un ensemble de variables explicatives  $X_1, X_2, \dots, X_p$ . Les variables  $X_i$  doivent être choisies de manière à être les plus prédictives possibles tout en étant les plus décorréelées possibles entre elles.

Cette méthode nécessitera alors de faire des hypothèses sur la loi suivie par les observations de la variable  $Y$  à expliquer. Celle-ci devra appartenir à la famille exponentielle, regroupant un grand nombre de lois classiques.

Contrairement au modèle linéaire standard où l'espérance de la variable observée est décomposable en une fonction linéaire des variables explicatives, dans le modèle linéaire généralisé c'est une fonction des observations, appelé paramètre naturel, qui est défini comme une fonction linéaire des variables explicatives. Le modèle capture ainsi les effets non linéaires des variables explicatives sur la variable réponse. Le modèle s'écrit donc comme suit :

$$g(E(Y)) = \beta \cdot X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

où  $Y$  est la variable réponse,  $X$  est le vecteur des variables explicatives,  $g$  la fonction de lien strictement monotone sur  $\mathbb{R}$  et  $\beta$  le vecteur des coefficients à estimer.

Concrètement, les GLM permettront d'expliquer de manière individuelle la valeur de l'AP/TP d'un contrat à partir de l'ensemble des critères disponibles le décrivant.

### **CHOIX DU MODELE :**

D'après la relation définissant les GLM, trois éléments le décrivent et sont ainsi à définir lorsque l'on a recours à cette méthode : la loi de  $Y$ , la fonction de lien  $g$  ainsi que les variables explicatives  $X_1, X_2, \dots, X_p$ .

- *La variable réponse, une composante aléatoire*

La variable aléatoire réponse est la variable dont on souhaite expliquer l'espérance de manière individuelle. On dispose de  $n$  réalisations  $Y_1, Y_2, \dots, Y_n$  de la variable  $Y$ . La loi de  $Y$  doit appartenir à la famille exponentielle. Celle-ci est une famille de lois de distribution à laquelle appartient l'ensemble des lois dont la densité peut s'écrire sous la forme suivante :

$$f(y) = e^{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)}$$

où  $\theta$  et  $\phi$  sont des paramètres respectivement appelés « paramètre naturel » et « paramètre de dispersion ».  $a$ ,  $b$  et  $c$  sont des fonctions.

On peut montrer que :

$$E(Y) = b'(\theta) = \mu$$

$$V(Y) = b''(\theta) \cdot \phi = \sigma^2$$

L'ensemble des lois étudiées précédemment, à l'exception de la loi Log-normale, appartient à la famille exponentielle. Si l'adéquation de la loi des AP/TP avec la loi Log-normale s'avère être bonne, on pourra cependant se ramener à un modèle généralisé en considérant le logarithme de la variable modélisée.

- *Les variables explicatives, composantes déterministes*

La prédictivité du modèle repose sur le choix stratégique des variables explicatives. En effet, celui-ci doit se faire de manière à ce que le nombre de variables retenues soit raisonnable afin que le modèle soit utilisable en pratique, mais suffisant pour qu'il soit aussi pertinent. Il faut ainsi choisir les meilleures variables au sens prédictif du terme. Il faut ainsi étudier l'influence des variables sur la variable à expliquer, en retirant tout effet de corrélation qui peut exister entre celles-ci.

### Critères de sélection de variables :

Il existe divers critères pour le choix des variables explicatives : AIC, AIC corrigé, BIC,... Seul le critère AIC sera exposé car c'est celui qui sera utilisé par la suite.

Le Critère d'Information d'Akaike (AIC) est défini par  $AIC = -2\ln L(\theta) + 2p$  où  $L$  est la vraisemblance du modèle et  $p$  est le nombre total de modalités explicatives retenues.

Celui-ci permet de sélectionner le modèle ayant un critère AIC minimum parmi un ensemble de modèles candidats. Ce critère est ainsi un compromis entre la qualité d'ajustement, augmentant avec le nombre de paramètres retenus et la parcimonie du modèle, volonté de décrire les données avec le plus petit nombre de paramètres possibles. En pénalisant les modèles intégrant un trop gros nombre de modalités, on limite les effets de sur-ajustement d'un modèle.

### Procédure de sélection de variables :

Afin de trouver un compromis entre l'exhaustivité et le coût d'obtention de l'information, diverses procédures de sélection des variables explicatives existent : la sélection exhaustive, la sélection forward, la sélection backward, la sélection stepwise. Seule la procédure forward sera exposée car c'est celle qui sera utilisée par la suite dans nos modélisations.

Dans la sélection forward, on effectue dans un premier temps des régressions avec chacune des  $d$  variables explicatives du support. Pour chacune, on réalise un test de significativité (Wald) de la variable et on retient pour notre modèle celle qui est la plus explicative. Ensuite, on effectue les  $d-1$  régressions possibles en introduisant une seconde variable. On réalise à nouveau un test de significativité pour chacune des variables introduites et on retient la variable la plus significative. Si l'ajout d'aucune variable n'est significatif, alors on arrête le processus.

### Outils de validation des variables :

Le test de Wald permet de tester la significativité des variables explicatives et de l'ensemble des modalités composant le modèle choisi.

Pour le test de la significativité d'une modalité  $j$  de la variable  $k$ , nous testons l'hypothèse nulle  $H_0: \beta_{j,k} = 0$  contre l'hypothèse alternative  $H_1: \beta_{j,k} \neq 0$

Pour le test de la significativité de la variable  $k$ , nous testons l'hypothèse nulle  $H_0: \beta_{1,k} = \beta_{2,k} = \dots = \beta_{n_k,k} = 0$  où  $n_k$  est le nombre de modalités de la variable  $k$ , contre l'hypothèse alternative  $H_1: \exists j, \beta_{j,k} \neq 0$ .

On calcule ensuite la statistique  $W$  correspondant au rapport de log vraisemblance du modèle non contraint et du modèle contraint par l'hypothèse  $H_0: W = 2(\log L(\hat{\beta}) - \log L(\hat{\beta}_c))$  où  $\hat{\beta}$  et  $\hat{\beta}_c$  sont respectivement les estimateurs des paramètres du modèle contraint et du modèle non contraint.

Sous l'hypothèse  $H_0$ , la statistique  $W$  suit une loi de Khi-Deux de degré de liberté égal à 1 si on teste la significativité d'une modalité et égale à  $n_k-1$  si on teste la significativité d'une variable à  $n_k$  modalités. Les variables dont la  $p$ -valeur excède 5% ne peuvent être retenues.

- *La fonction de lien : généralisation du modèle linéaire*

La fonction de lien contient le lien entre l'espérance de la variable à expliquer et la combinaison linéaire des variables explicatives.

$$g(E(Y)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

où  $g$  est supposée monotone et différentiable.

Le choix de la fonction de lien est souvent donné par le lien canonique qui est le lien  $g$  associant la moyenne au paramètre naturel  $\theta : g = b'(\theta)^{-1}$ .

### **CALCUL DES PARAMETRES DU MODELE :**

Une fois le choix du modèle effectué, au travers de la détermination de la loi suivie par la variable réponse, le choix des variables explicatives et la sélection de la fonction de lien, il faut désormais estimer les paramètres  $\beta_0, \beta_1, \dots, \beta_p$  intervenant dans la formule du GLM à partir des données  $y_1, y_2, \dots, y_n$ . On en déduira ainsi une espérance de la variable réponse des observations en calculant :

$$\hat{y}_i = g^{-1}(\beta \cdot x) = g^{-1}(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p)$$

La méthode utilisée est la technique du maximum de vraisemblance. On dispose de  $n$  observations  $y_1, y_2, \dots, y_n$  et de l'ensemble de leurs  $p$  variables explicatives  $X_{1,1}, X_{1,2}, \dots, X_{1,p}, X_{2,1}, X_{2,2}, \dots, X_{2,p}, \dots, X_{n,1}, X_{n,2}, \dots, X_{n,p}$ . On regroupe les données explicatives dans une matrice de design  $X$  à  $n$  lignes et  $k$  colonnes.

La solution du maximum de vraisemblance se calcule ensuite par la procédure itérative de descente de gradient de Fisher ou alternativement par celle de Newton-Raphson.

### **VALIDATION ET QUALITE DU MODELE :**

En général, on procède par validation croisée : La base de données initiale est systématiquement séparée en deux échantillons contenant 80% et 20% de nos observations. Le premier échantillon, appelée échantillon d'apprentissage, permet d'engendrer le modèle tandis que le deuxième, appelé échantillon test, permet de valider le modèle. Ainsi, nous vérifierons que nos résultats ne sont pas le fruit du hasard et ne dépendent pas des données utilisées pour engendrer le modèle. En effet, on peut construire un modèle qui présente d'excellentes qualités de prédiction sur l'échantillon d'apprentissage mais qui ne s'adapte alors pas bien sur l'échantillon test. On parle dans ce cas de sur-apprentissage du modèle.

Il existe diverses méthodes permettant de juger de la qualité du modèle linéaire généralisé obtenu. On retiendra parmi celles-ci l'étude des résidus de Pearson. Cet outil est étudié à la fois sur la base d'échantillonnage pour vérifier la cohérence du modèle élaboré ainsi que sur la base de test pour vérifier le sur-apprentissage.

On définit les résidus bruts du modèle comme l'écart entre la valeur prédite et la variable observée :

$$\text{Résidu}_B^i = Y_i - \hat{Y}_i \text{ pour toute observation } i$$

Les résidus bruts doivent être en principe concentrés autour de 0 si le modèle est satisfaisant. Cependant, ces derniers n'ayant pas la même variance selon le modèle étudié, on doit alors se référer aux résidus de Pearson qui sont définis par :

$$\text{Résidu } p^i = \frac{AP/TP_{\text{observé}}^i - AP/TP_{\text{estimé}}^i}{\sigma(AP/TP_{\text{estimé}})} \text{ pour toute observation } i.$$

En théorie, lorsque les résidus de Pearson suivent une loi normale centrée réduite, l'erreur du modèle est assimilable à du bruit statistique.

## Partie 2 : Application des Modèles Linéaires Généralisés à la modélisation de l'AP/TP

Le but de cette partie est de modéliser l'AP/TP des contrats en portefeuille. Les modèles généralisés vont être utilisés, permettant d'évaluer de manière prédictive la valeur de l'écart au Tarif Technique contrat par contrat à partir de l'ensemble des caractéristiques le définissant. A chaque catégorie de risque sera associé un coefficient, permettant ainsi de juger de l'influence d'une modalité de variable sur la rentabilité technique.

Etant donné l'allure de la distribution de l'AP/TP présentée antérieurement, nous allons tout d'abord chercher à identifier la loi globale suivie par l'échantillon d'AP/TP. Ensuite, nous sélectionnerons les facteurs permettant la meilleure segmentation de notre modèle et la prédiction optimale de l'indicateur à partir du support d'information sur le risque de l'assuré et sur les modulations commerciales qui lui sont accordées. Cette sélection intègrera les analyses qui ont été faites dans le chapitre précédent. Enfin, nous nous intéresserons à la validation du modèle obtenu, en se référant aux indicateurs jugeant de la qualité du modèle.

### a) Sélection du modèle d'AP/TP

Pour chaque case tarifaire, l'AP/TP va être estimé par un modèle linéaire généralisé. Il est alors essentiel de choisir méthodiquement les trois éléments définissant un GLM, de manière à obtenir le modèle le plus pertinent.

Les contrats ayant des AP/TP supérieurs à 4, qualifiés d'exceptionnels, ont été retirés de la modélisation, ceux-ci étant considérés comme du bruit statistique. Ils sont au nombre de 209, ce qui représente 0.01% de la base initiale pour 0.02% du chiffre d'affaire réalisé sur le portefeuille « 4RP Agents ». Ainsi, en les retirant de la base modélisée, nous ne perdons pas vraiment d'information sur le portefeuille étudié.

En partant de notre base initiale et après retrait des AP/TP supérieurs à 4, deux échantillons ont été construits contenant respectivement 80% et 20% de celle-ci et correspondant aux échantillons d'apprentissage et de test. Cette démarche nous permet de nous assurer de la robustesse du modèle entraîné sur la base d'apprentissage et que les résultats ne dépendent pas des données utilisées. Afin de garantir une certaine qualité de notre modèle, nous avons construit les échantillons d'apprentissage et de test de manière à ce que les données soient équi-réparties sur l'AP/TP.

- *Identification de la loi suivie par la variable réponse*

Dans un modèle linéaire généralisé, la variable à expliquer suit une loi qui appartient à la famille exponentielle. Différentes lois de la famille vont ainsi être testées pour identifier la loi théorique suivie par l'échantillon d'AP/TP dont nous disposons. Etant donné les caractéristiques relevées sur la distribution de l'AP/TP dans le chapitre 2, nous allons nous intéresser aux lois Exponentielle, Normale, Gamma et Weibull. Nous testerons également la loi log-normale, qui semble être pertinente, en passant par la modélisation du logarithme de l'AP/TP permettant de se ramener à une loi appartenant à la famille exponentielle.

Pour tester l'adéquation aux lois, des méthodes de comparaisons graphiques de densité et fonction de répartition ainsi que des tests d'adéquation des quantiles théoriques et réels ont été réalisés. Les paramètres de chacune des lois testées ont été estimés au préalable par la méthode du maximum de vraisemblance. Les résultats des tests d'adéquation aux différentes lois sont disponibles en annexe.

Les résultats obtenus nous ont permis de retenir la loi Gamma comme étant la loi la plus proche de la distribution de notre échantillon d'AP/TP sur portefeuille « 4RP Agents ». C'est une distribution peu restrictive qui permet de modéliser des données positives continues à queue intermédiaire. Elle est notamment très utilisée en actuariat pour la modélisation des coûts moyens de sinistres.

Les résultats des tests d'adéquation de l'indicateur à la loi gamma sont les suivants :

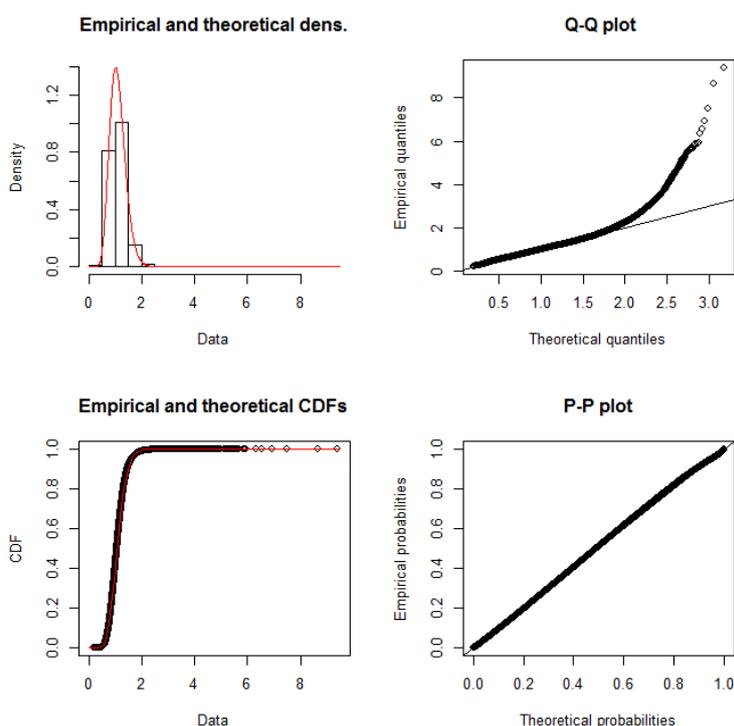


Figure 35: Tests d'adéquation de la densité empirique de l'AP/TP avec la loi Gamma

Le QQ-Plot fait ressortir un certain écart sur les valeurs d'AP/TP supérieures à 2. En effet, la loi gamma correspondante semble sous-estimer la proportion d'AP/TP extrêmes. Cependant, l'ensemble des contrats en portefeuille ayant un AP/TP supérieur à 2 représente 1.14% du portefeuille pour 1.66% de son chiffre d'affaire. La loi gamma est donc acceptable pour la modélisation de l'AP/TP de l'ensemble du portefeuille.

La loi retenue pour la modélisation des AP/TP du portefeuille quatre roues considéré est la loi gamma, dont la densité est la suivante :

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \text{ avec } x \in \mathbb{R}_*^+, \alpha \in \mathbb{R}_*^+ \text{ et } \lambda \in \mathbb{R}_*^+$$

Où  $\alpha$  et  $\lambda$  ayant été estimés par le maximum de vraisemblance sont données par :

$$\hat{\alpha} = 13.68 \text{ et } \hat{\lambda} = 12.49$$

- *Choix de la fonction de lien*

Le lien canonique de la loi gamma est la fonction inverse. Cependant, cette fonction pose deux problèmes dans son utilisation pour les modèles linéaires généralisés :

- La positivité des valeurs prédites n'est pas garantie. Or la loi gamma ne prend que des valeurs strictement positives
- Les coefficients sont à interpréter à l'inverse de la logique habituelle : plus il est élevé en valeur absolue, moins son influence est grande sur la valeur prédite.

Afin de palier simultanément à ces deux problèmes, la fonction logarithmique a été utilisée comme fonction de lien. Il a été constaté que l'utilisation de la fonction logarithmique plutôt que du lien canonique ne baisse pas pour autant la qualité du modèle. De plus, le modèle est ainsi rendu multiplicatif, ce qui simplifie l'utilisation des coefficients. Le lien logarithmique est ainsi fréquemment utilisé en actuariat.

- *Choix des variables explicatives*

Les variables disponibles pour la modélisation de l'AP/TP sont les variables que nous avons retenus en fin d'analyse dans le chapitre 2. Il s'agit des variables ayant été révélées discriminantes sur l'AP/TP et ayant été corrigées des trop fortes corrélations qui existent entre les variables restantes et qui pourraient ainsi biaiser le modèle. Notre support est le suivant :

<b>Intitulé de la variable</b>
<b>Données sur le conducteur principal assuré</b>
Age du conducteur
Catégorie Socio-Professionnelle
Antécédent sinistre
<b>Données sur le véhicule assuré</b>
Ancienneté du véhicule
Groupe SRA
Année d'acquisition
Energie
Carrosserie
Note de sécurité

Données sur la géographie
Zone Bris de Glace
Zone Vol/Incendie
Données sur le contrat
Formule
Age du contrat
Fractionnement
Usage kilométrique
Pack Réparation
Clause tarifaire
Données sur les modulations tarifaires
Avantage client
Avantage prospect
Taux budget
Taux firme

Figure rappel : Support des critères disponibles à la modélisation de l'AP/TP

Pour choisir les variables explicatives composant notre modèle GLM, nous allons utiliser le Critère d'Information d'Akaike (AIC), celui-ci permet de sélectionner le modèle ayant un critère AIC minimum parmi un ensemble de modèles candidats.

Parmi les différentes procédures de sélection de variables, la procédure de sélection forward a été retenue. Le choix des variables utilisées pour la modélisation de l'AP/TP à partir du support de variables a été réalisé à partir de la commande « Stepwise Regression » du logiciel Emblem.

Le graphique ci-dessous nous donne l'AIC du modèle à chaque ajout de variable ainsi que le pourcentage de réduction d'AIC associé :

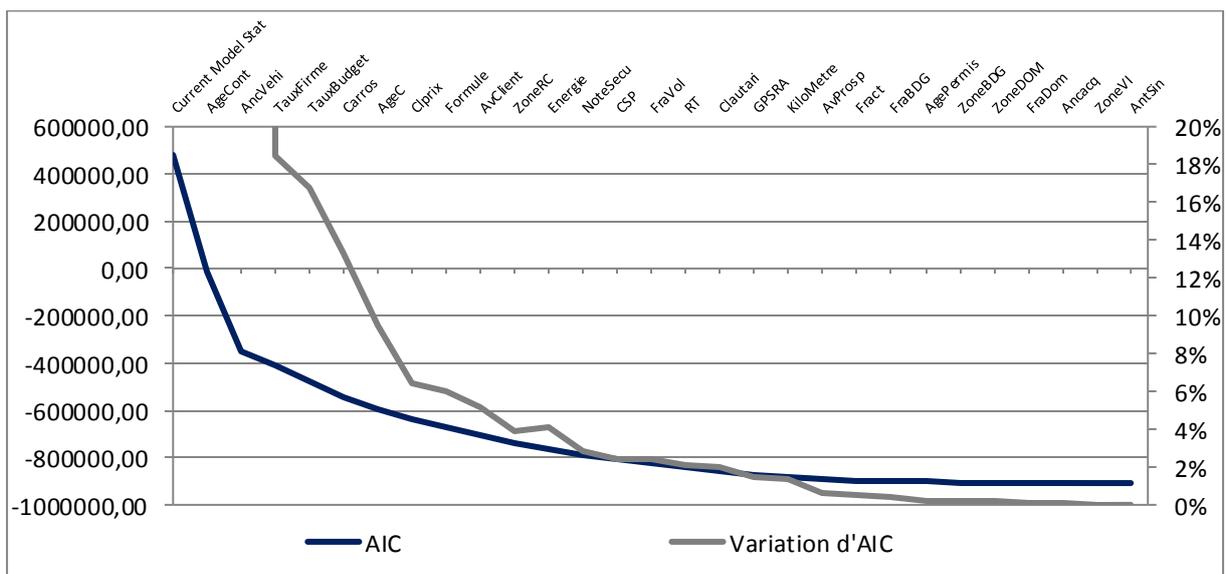


Figure 36: Résultat de la sélection forward des variables explicatives du modèle d'AP/TP sous Emblem

Nous allons conserver l'ensemble des critères jusqu'à la variable « AvClient », le reste des variables ne réduisant que de manière non notable le critère AIC et leur ajout nécessitant un grand nombre de regroupements étant donné l'apparition de modalités non significatives. L'AIC qui était de -13 555 en début de modélisation lors de l'ajout de la première variable diminue à -70 8237 sur le modèle final. On note également que les variables présentant une trop forte corrélation entre elles n'ont ainsi pas été ajoutées toutes deux. Par exemple, la variable « Anacq », ayant un V de Cramer égal à 43% avec la variable « AgeCont » déjà intégrée au modèle, n'a pas été ajoutée. De même, le modèle n'a pas ajouté simultanément la variable « Classe de Prix » et l'« Energie » ou le « Groupe SRA », ces deux dernières variables étant fortement corrélées avec la variable « Classe de Prix ».

Cette sélection nous permet de retenir les variables les plus pertinentes pour la modélisation de notre indicateur et en nombre raisonnable, permettant ainsi d'identifier les variables les plus influentes sur la valeur de l'AP/TP. Ainsi, l'ensemble des variables participant à la modélisation de l'AP/TP est le suivant :

Variables retenues pour la modélisation de l'AP/TP								
AgeCont	AncVehi	TauxFirme	TauxBudget	Carros	AgeC	Clprix	Formule	AvClient

Figure 37: Variables explicatives retenues pour la modélisation de l'AP/TP

On constate que la première variable sélectionnée pour la modélisation de l'AP/TP et donc la plus discriminante sur l'AP/TP est l'ancienneté du contrat. Nous avons relevé antérieurement, dans les analyses uni-variées, que l'influence de cette variable sur l'AP/TP pouvait s'expliquer par différents points. Parmi ceux-là on rappelle qu'interviennent au cours de la vie d'un contrat :

- un processus de surveillance du portefeuille, revalorisant fortement les cotisations voir résiliant les contrats sous-profitables. Les contrats non rentables sont alors écartés au cours de leur vie
- un processus complexe de revalorisation des primes qui tend à faire croître l'écart existant entre le Tarif Technique et le Tarif Réel
- un système de grignotage des modulations commerciales au cours de la vie du contrat, résultant alors en l'augmentation de son AP/TP

Pour ces différentes explications, la rentabilité est alors d'autant meilleure que le contrat est ancien.

Un certain nombre de modalités relatives aux modulations tarifaires ressortent également lors de la modélisation de l'indicateur de rentabilité. Cela s'explique par le fait que la valeur de l'écart au Tarif Technique est la conséquence d'un processus complexe de revalorisation lors des avenants et des échéances de portefeuille dépendant fortement des modulations tarifaires accordées au contrat.

On note enfin l'importance de la formule souscrite sur le niveau de rentabilité, puisque ce facteur est pris en compte dans les grilles de revalorisation du Tarif Réel.

## b) Implémentation du modèle

On rappelle que le modèle linéaire généralisé utilisé est un modèle de fonction de lien logarithme et que la densité de la loi réponse est la loi gamma. Les variables explicatives sélectionnées pour la prédiction de l'AP/TP sont les suivantes :

Variables retenues pour la modélisation de l'AP/TP								
AgeCont	AncVehi	TauxFirme	TauxBudget	Carros	AgeC	Clprix	Formule	AvClient

Figure rappel : Variables explicatives retenues pour la modélisation de l'AP/TP en loi unique

Si on note  $m_k$  le nombre de modalités de la variable  $k$  pour  $k = \{1, \dots, 9\}$ , alors le nombre total de cases tarifaires est  $\prod_{k=1}^9 m_k$ . On note également  $m = \sum_{k=1}^9 m_k$  le nombre total de modalités du modèle.

Si on note  $AP/TP_i^{pred}$  la valeur prédite de l'AP/TP pour la case  $i$ , la vraisemblance du modèle s'écrit sous la forme :

$$\log(AP/TP_i^{pred}) = \beta_0 + \sum_{j=1}^m \beta_j \mathbb{1}_{\text{case } i \text{ vérifie la modalité } j}$$

D'où

$$AP/TP_i^{pred} = e^{\beta_0 + \sum_{j=1}^m \beta_j \mathbb{1}_{\text{case } i \text{ vérifie la modalité } j}}$$

- Estimation des paramètres à partir de la méthode du maximum de vraisemblance

L'ensemble des  $m + 1$  coefficients  $\beta$  sont déterminés par la méthode du maximum de vraisemblance sur l'échantillon de données compris dans la base d'apprentissage.

L'estimation des paramètres de chacune des modalités a été réalisée sous le logiciel Emblem. Les résultats se trouvent en annexe.

- Significativité des paramètres du modèle

Les résultats des tests de significativité des variables par le test de Wald sur le modèle d'AP/TP retenu sont les suivants :

Factor	Wald p Value (Chi-Sq)
AgeC	0,0000%
AncVehi	0,0000%
Clprix	0,0000%
Carros	0,0000%
Formule	0,0000%
AgeCont	0,0000%
AvClient	0,0000%
TauxBudget	0,0000%
TauxFirme	0,0000%

Figure 38: Résultats des tests de Wald des variables explicatives sur la base d'apprentissage

L'ensemble des variables explicatives retenues sont significatives dans la modélisation des AP/TP.

Les erreurs standards données dans l'onglet « Parameters » du logiciel Emblem et exprimées en pourcentage nous permettent de vérifier également que les coefficients de chacune des modalités sont bien

différents de 0. Ici, les résultats nous indiquent que l'ensemble des modalités de chaque variable sont significatives.

Les résultats de significativité des modalités ainsi que les coefficients du modèle obtenu sont présentés en annexe.

Prenons en exemple les coefficients obtenus pour la variable ancienneté du contrat. Pour obtenir les coefficients multiplicatifs, on doit passer les coefficients obtenus à l'exponentielle. En effet, le modèle s'écrit :

$$AP/TP_i^{pred} = e^{\beta_0 + \sum_{j=1}^m \beta_j \mathbb{1}_{\substack{\text{case } i \text{ vérifie} \\ \text{la modalité } j}}}$$

$$\text{d'où } AP/TP_i^{pred} = e^{\beta_0} \cdot \prod_{j=1}^m e^{\beta_j \cdot \mathbb{1}_{\substack{\text{case } i \text{ vérifie} \\ \text{la modalité } j}}}$$

$$AP/TP_i^{pred} = \beta'_0 \cdot \prod_{j=1}^m \beta'_j \cdot \mathbb{1}_{\substack{\text{case } i \text{ vérifie} \\ \text{la modalité } j}}$$

Les coefficients obtenus pour l'ancienneté du contrat sont les suivants :

	Coefficient	Exponentiel du coefficient
AgeCont='ACont_A'	-0,2731	0,76
AgeCont='ACont_B'	-0,1685	0,84
AgeCont='ACont_C'	-0,1093	0,90
AgeCont='ACont_D'	-0,075	0,93
AgeCont='ACont_E'	-0,0424	0,96
AgeCont='ACont_F'	0	1,00
AgeCont='ACont_G'	0,0334	1,03
AgeCont='ACont_H'	0,095	1,10

Figure 39: Exemple de relativité du modèle avec l'ancienneté du contrat

Les contrats d'ancienneté inférieure à 1 an ('ACont\_A') auront une sous-tarification de 24% par rapport à la moyenne du portefeuille tandis que les contrats d'ancienneté supérieure à 10 ans auront une surtarification de 10% par rapport à la modalité de référence.

### c) Validation du modèle

- *Qualité d'ajustement du modèle et analyse des résidus*

Il est ensuite essentiel d'évaluer la qualité du modèle obtenu. Pour ce faire, on étudie usuellement la part de la variable réponse non expliquée ligne à ligne par le modèle, c'est-à-dire non explicable linéairement par les variables explicatives sélectionnées. Il s'agit des résidus bruts du modèle, qui sont définis par :  $\text{Résidu}_B^i = AP/TP_{\text{observé}}^i - AP/TP_{\text{estimé}}^i$  pour tout contrat  $i$ . Un modèle est doté d'une bonne prédictivité si ses résidus bruts sont proches de zéro.

Nous disposons des statistiques suivantes sur les résidus bruts :

Moments			
<b>N</b>	1390640	<b>Sum Weights</b>	1390640
<b>Mean</b>	-5.8723E-6	<b>Sum Observations</b>	-8.1662179
<b>Std Deviation</b>	0.21133971	<b>Variance</b>	0.04466447
<b>Skewness</b>	1.7249623	<b>Kurtosis</b>	8.9140474
<b>Uncorrected SS</b>	62112.1587	<b>Corrected SS</b>	62112.1586
<b>Coeff Variation</b>	-3598942.1	<b>Std Error Mean</b>	0.00017921

Figure 40: Statistiques sur les résidus bruts du modèle d'AP/TP sur l'échantillon d'apprentissage

Les résidus bruts du modèle ont une moyenne proche de zéro. Cependant, les résidus bruts n'ont pas la même variance selon le modèle étudié. On a alors tendance à étudier les résidus réduits.

Les résidus de Pearson, définis comme les résidus bruts réduits permettent de juger de la qualité du modèle. Ils sont définis par : 
$$\text{Résidu}_p^i = \frac{AP/TP_{\text{observé}}^i - AP/TP_{\text{estimé}}^i}{\sigma(AP/TP_{\text{estimé}})}$$

La représentation des résidus de Pearson permet de vérifier la qualité du modèle. Le graphique doit être symétrique en 0, ayant des valeurs entre -3 et +3 et le plus regroupé possible en 0. Nous disposons des résultats suivants sur notre modèle :

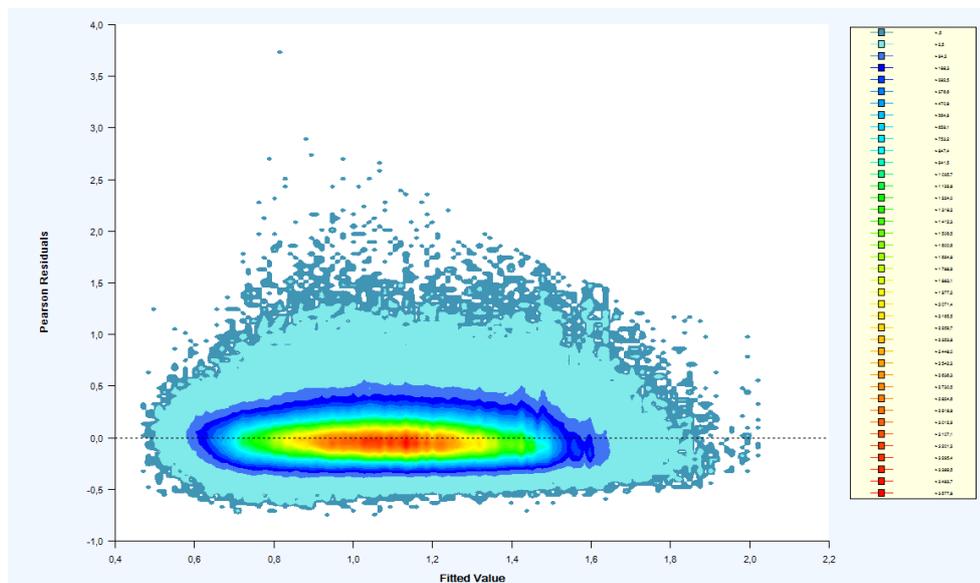


Figure 41: Résidus de Pearson du modèle d'AP/TP sur l'échantillon d'apprentissage

Le graphique ci-dessus présente les résidus de Pearson en fonction de la valeur des AP/TP prédits par le modèle. La zone est d'autant plus rouge qu'elle est concentrée en points et d'autant plus bleue qu'elle est peu concentrée. On constate la présence d'un unique îlot, faisant preuve qu'aucune variable explicative pertinente n'a été négligée. De plus, les résidus sont centrés en zéro et d'une forme ellipsoïde. Ils sont également bien concentrés entre -3 et +3, mis à part un point qui sort de cet intervalle. Le modèle obtenu est ainsi plutôt satisfaisant sur la base d'apprentissage.

On note, cependant, une certaine asymétrie des résidus de Pearson, ceux-ci ayant plus souvent des valeurs positives que négatives. Le modèle sous-estime la valeur des AP/TP des contrats en portefeuille en moyenne.

Nous souhaitons ensuite comparer la distribution des résidus de Pearson avec celle de la loi normale centrée réduite. En effet, le modèle est correct si l'histogramme des résidus de Pearson suit une telle loi. La distribution des résidus de Pearson obtenus sur notre modèle est la suivante :

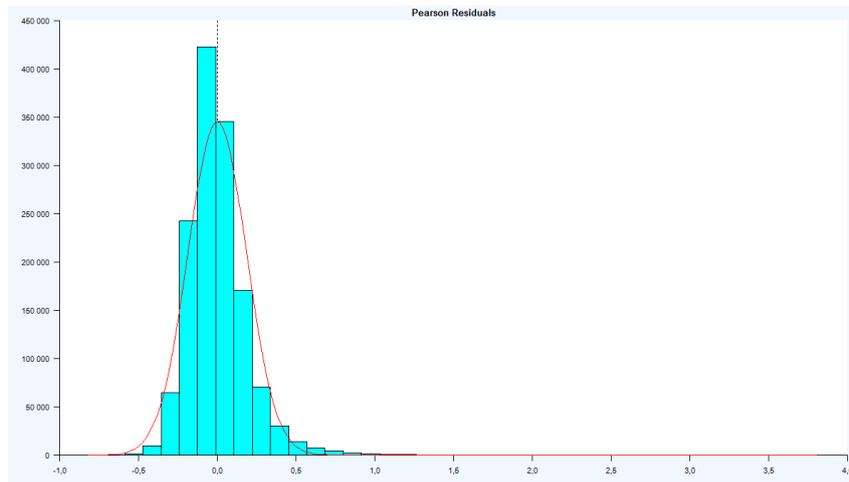


Figure 42: Histogramme des résidus de Pearson du modèle d'AP/TP sur l'échantillon d'apprentissage

L'histogramme des résidus standardisés ci-dessus indique que ceux-ci s'ajustent plutôt bien avec une loi normale centrée réduite, jugeant de la bonne qualité de notre modèle.

On note tout de même que l'histogramme relève à nouveau une asymétrie de la distribution des résidus de Pearson du modèle par la présence d'une queue de distribution à droite légèrement plus fournie que la queue de distribution à gauche.

Un test d'adéquation des résidus de Pearson à la loi normale a été réalisé :

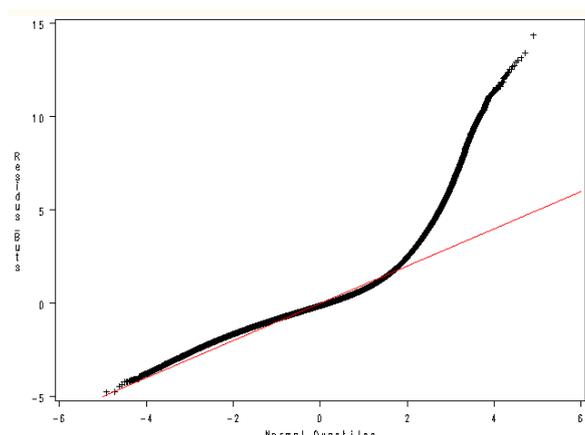


Figure 43: Test d'adéquation des résidus bruts à la loi normale sur l'échantillon d'apprentissage

Le test de QQ-Plot, qui porte un intérêt particulier aux valeurs extrêmes, nous indique que l'adéquation des résidus de Pearson à la loi normale est plutôt bonne, à l'exception de la queue de distribution positive qui ne s'y ajuste pas correctement. Nous souhaitons naturellement comprendre quels sont les contrats qui sont associés à un tel écart des résidus.

Pour rappel, lors du choix de la loi suivie par les AP/TP à modéliser, nous avons sélectionné la loi gamma comme étant la loi la plus adaptée. Nous avons cependant noté qu'un certain décrochement entre la loi théorique et la loi empirique se faisait remarquer pour les AP/TP supérieures à 2 environ. En effet, la loi gamma sous-estimait la queue de distribution de droite car la loi gamma utilisée est caractérisée par une queue de distribution intermédiaire, alors que la distribution des AP/TP admet une lourde queue à droite. On peut donc en supposer que l'importance des erreurs positives s'expliquerait par le fait que notre modèle sous-estime les valeurs extrêmes d'AP/TP. En effet, le modèle prédit 0.0006% de cas extrêmes contre 1.14% en réalité pour un seuil d'AP/TP extrême fixé à 2.

- *Validation du modèle sur la base de test*

Il est ensuite essentiel de tester le modèle qui a été élaboré à partir de la base d'apprentissage sur la base de test, permettant ainsi de nous assurer de la robustesse du modèle et de vérifier qu'il n'est pas sur-ajusté aux données d'apprentissage.

Le graphique représentant les estimations des AP/TP en fonction des résidus de Pearson sur la base de test est le suivant :

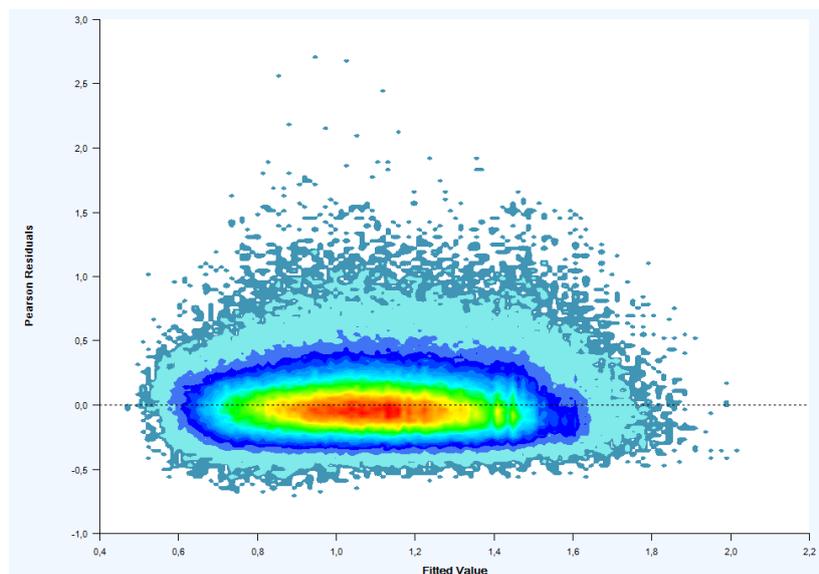


Figure 44: Résidus de Pearson du modèle sur l'échantillon test

Des résultats similaires à ceux obtenus sur la base d'apprentissage sont observés : les résidus de Pearson sont centrés en 0 avec une forte concentration autour de cette valeur, faisant gage d'un modèle de bonne qualité et qui s'avère être robuste car il est également ajusté à notre base de test.

L'histogramme des résidus de Pearson sur la base de test est le suivant :

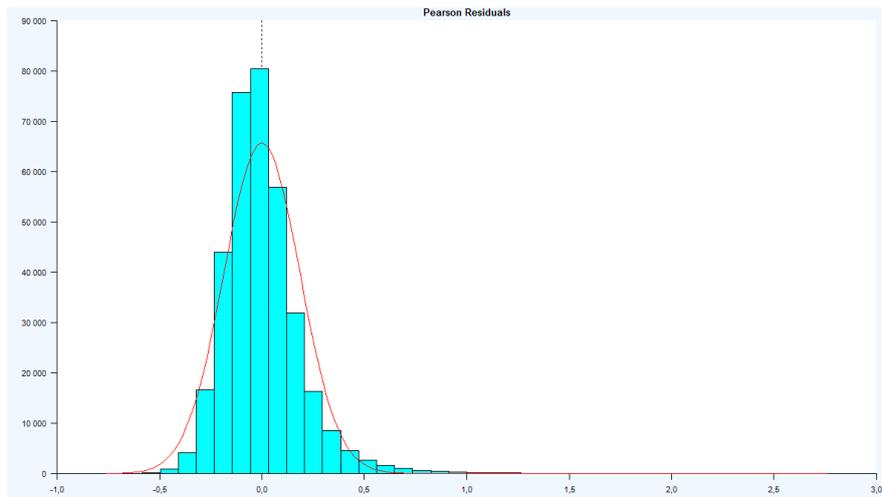


Figure 45: Histogramme des résidus de Pearson du modèle d'AP/TP sur l'échantillon test

On obtient, avec l'échantillon test, des résultats assez proches de ceux obtenus avec l'échantillon d'apprentissage : les résidus bruts du modèle sont globalement concentrés autour de 0 et la loi suivie par les résidus semble être proche de la loi normale, à l'exception d'une queue de distribution plus importante à droite qu'à gauche.

Cette observation est confirmée par le test d'adéquation à la loi normale par un QQ-plot, qui montre en effet la présence d'une plus forte queue de distribution du côté positif des résidus et donc une sous-estimation en moyenne des AP/TP.

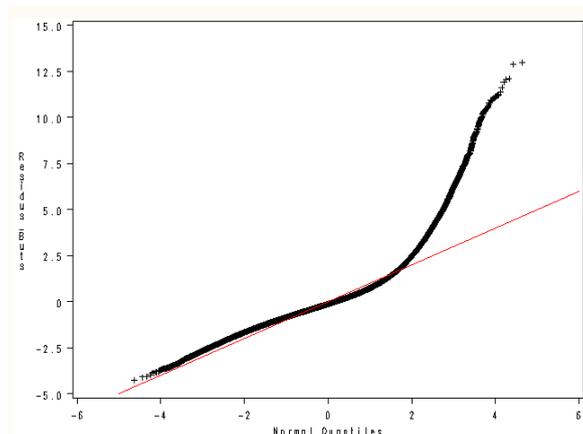


Figure 46: Test d'adéquation des résidus de Pearson du modèle d'AP/TP à la loi normale sur l'échantillon test

Une certaine erreur de modélisation est ainsi commise au niveau des valeurs extrêmes d'AP/TP, soulignant ainsi le fait que la modélisation n'est pas ajustée au niveau de ces cas.

- *Stabilité du modèle dans le temps*

Pour qu'un modèle soit fiable, il faut que sa capacité prédictive soit stable dans le temps. Nous souhaitons ici tester l'hypothèse de stabilité du modèle dans le temps. Ainsi nous avons appliqué le modèle d'AP/TP généré sur la base de 2016 sur des échantillons des bases d'AP/TP de 2014 et 2015. Les résidus de Pearson sont ensuite analysés. On obtient les histogrammes de résidus suivants :

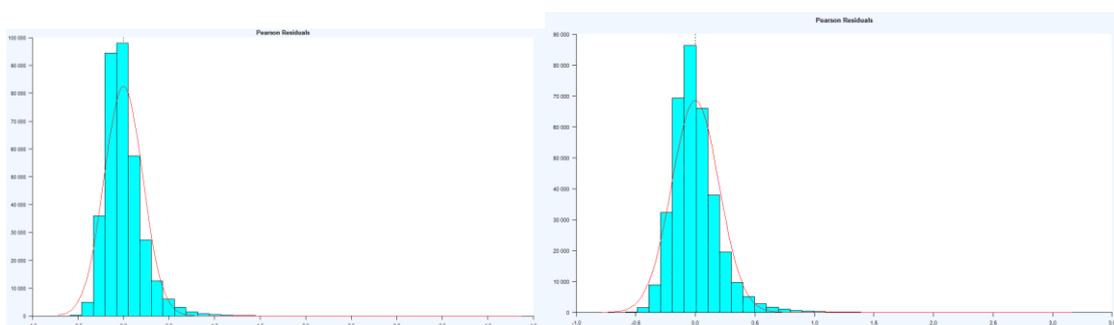


Figure 47: Histogrammes des résidus de Pearson du modèle d'AP/TP sur les bases de 2014 (à gauche) et 2015 (à droite)

Les résidus sont centrés et assimilables à des lois normales. Le modèle est ainsi stable dans le temps.

Dans cette partie, nous avons mis en place un modèle robuste de segmentation et d'estimation de notre indicateur de rentabilité technique qu'est l'AP/TP sur le portefeuille « 4RP Agents ». L'importance de l'ancienneté du contrat, de la formule souscrite et des modulations tarifaires dans la valeur de l'AP/TP y a été soulignée.

Nous avons néanmoins relevé que le modèle trouvait ses limites dans la modélisation des AP/TP ayant des valeurs élevées. Ces contrats, bien qu'en effectif réduit puisqu'ils représentent 1.14% du portefeuille pour un seuil de 2, semblent faire apparaître une certaine hétérogénéité dans leurs caractéristiques par rapport au reste du portefeuille. Nous souhaitons naturellement mieux comprendre ces contrats caractérisés par une rentabilité notable sur le portefeuille et ayant une faible occurrence. Pour procéder, nous allons réaliser une analyse supplémentaire sur ces extrêmes et qui fera l'objet du chapitre 4 de ce mémoire.

## CHAPITRE 4 : Etude portée sur les AP/TP extrêmes

Dans le chapitre précédent, nous avons relevé la présence de contrats en portefeuille faisant apparaître une rentabilité particulièrement importante, composant la queue de distribution droite de la distribution des AP/TP, et dont la modélisation des AP/TP du portefeuille « 4RP Agents » ne permettait pas la bonne appréciation de ces profils particuliers étant donné leur faible effectif relativement à l'ensemble du portefeuille. Cette partie va donc nous permettre de caractériser au mieux ces profils associés à une certaine rentabilité des contrats et que l'on souhaite ainsi particulièrement capter.

## Partie 1 : Détermination du seuil des extrêmes par la théorie des valeurs extrêmes

Dans les chapitres précédents, les contrats avaient été définis comme étant extrêmes dès lors que leur AP/TP avait une valeur supérieure à 2. Ce seuil avait été déterminé par un raisonnement économique, ces contrats étant ceux qui payent une cotisation d'un montant égal à deux fois celui du risque qu'ils assurent.

Nous souhaitons désormais fixer ce seuil de manière plus rationnelle, c'est à dire que ce nouveau seuil nous permettrait d'isoler les AP/TP, relevés précédemment, faisant apparaître une certaine hétérogénéité dans leur composition et dans leur distribution par rapport à l'essentiel du portefeuille.

Nous allons tout d'abord réaliser quelques rappels sur la théorie des valeurs extrêmes afin d'avoir en main les outils principaux de cette théorie, nous permettant par la suite de définir le seuil des extrêmes.

### a) La théorie des valeurs extrêmes

La théorie des valeurs extrêmes s'intéresse au comportement des extrêmes et est alors souvent utilisée en actuariat lors de la modélisation des sinistres extrêmes. Celle-ci propose un certain nombre d'outils permettant de procéder à la définition du seuil des extrêmes.

- *Les distributions des extrêmes généralisés*

Soit  $X$  une variable aléatoire et  $M_n = \max(X_i, i \in \llbracket 1, \dots, n \rrbracket)$ . Le théorème clé de la Théorie des valeurs extrêmes est le théorème de Fisher-Tippett. Celui-ci indique que s'il existe deux suites réelles  $(a_n)_{n \geq 1}$  et  $(b_n)_{n \geq 1}$  telles que  $a_n > 0$  pour tout  $n$  et que  $P \left[ \frac{M_n - a_n}{b_n} \leq x \right] \rightarrow G(x)$  lorsque  $n \rightarrow +\infty$ , alors  $G$  ne peut suivre que trois lois :

- 1) La loi de Gumbel  $G(x) = e^{-e^{-x}}$
- 2) La loi de Weibull  $G_\alpha(x) = \begin{cases} e^{-x^{-\alpha}}, & x \geq 0 \text{ et } \alpha > 0 \\ 0, & x < 0 \end{cases}$
- 3) La loi de Fréchet  $G_\alpha(x) = \begin{cases} 1, & x \geq 0 \\ e^{-(-x)^\alpha}, & x < 0 \text{ et } \alpha > 0 \end{cases}$

Ces trois lois peuvent être regroupées sous le nom de lois des extrêmes généralisée (GEV), dont la fonction de répartition générale est définie par :

$$G_{\mu, \sigma, \xi}(x) = e^{\left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right)_+^{-\frac{1}{\xi}} \right] \right\}}$$

, où  $\mu \in \mathbb{R}$  est un paramètre de « position »,  $\sigma > 0$  est un paramètre de « dispersion » et  $\xi \in \mathbb{R}$  est un paramètre de « forme », appelé également « indice des valeurs extrêmes ». Ce dernier caractérise l'épaisseur de la queue de la distribution de la variable  $X$  et le type de loi de la distribution  $G$  des extrêmes généralisés :

- si  $\xi \rightarrow 0$  alors on retrouve les distributions de type Gumbel à queue fine ou moyenne
- si  $\xi < 0$  alors on retrouve les distributions de type Weibull à queue finie
- si  $\xi > 0$  alors on retrouve les distributions de type Fréchet à queue épaisse

Voici un tableau donnant des exemples de lois associées à chacun des domaines des extrêmes généralisés :

Weibull ( $\xi < 0$ )	Gumbel ( $\xi \rightarrow 0$ )	Fréchet ( $\xi > 0$ )
Lois à queue finie à droite	Lois à queue fine	Lois à queue épaisse
Uniforme, beta	Gamma, Normale, Lof-Normale, Weibull, Exponentielle	Cauchy, Pareto, Student, Log-gamma

Figure 48: Distribution des extrêmes généralisés et lois classiques associées

La recherche d'un seuil revient à étudier la distribution de la loi dépassant ce seuil  $u$ . Nous disposons, par la théorie des valeurs extrêmes, de plusieurs outils permettant de le définir :

- *Le Mean Excess Loss Plot*

Parmi les outils proposés par la théorie des valeurs extrêmes, la fonction moyenne des excès est un moyen simple et visuel de déterminer la valeur du seuil des extrêmes.

La fonction moyenne des excès par rapport au seuil  $u$  est définie par :  $e(u) = E(X - u | X > u)$ . On calcule alors pour chaque seuil  $u$  la fonction moyenne des excès supérieurs à ce seuil. La fonction moyenne des excès empirique est alors définie par :  $\hat{e}(u) = \frac{\sum_{i=1}^n (X_i - u) \mathbb{1}_{X_i > u}}{\sum_{i=1}^n \mathbb{1}_{X_i > u}}$ .

Dans le cas des lois extrêmes généralisées, la fonction moyenne des excès a la particularité que l'allure de sa courbe nous donne des informations quant à la loi suivie par la variable initiale  $X$ . Voici un tableau donnant des exemples d'association d'allure de courbe de la fonction moyenne des excès et de loi classique affiliée :

$e(u)$	Loi associée
Tend vers 0	Normale
Constante égale à $\lambda$	Exponentielle de paramètre la constante
Tend en décroissant vers $\frac{1}{\lambda}$	Gamma ( $n, \lambda$ )
Croît approximativement comme $\frac{u}{\ln(u)}$	Normale
Croît approximativement comme $\frac{u}{\lambda - 1}$ si $\lambda > 1$ et l'infini sinon	Log-gamma ( $n, \lambda$ )
Croît linéairement avec coefficient directeur égal à $\frac{\alpha}{\alpha - 1}$ si $\alpha > 1$ et l'infini sinon	Pareto ( $\alpha$ )

Figure 49: Allure du Mean Excess Loss et distributions associées

- *Le Hill Plot*

L'estimateur de Hill est un outil permettant de déterminer un seuil des extrêmes dans le cas des distributions de Fréchet uniquement.

Il est défini par  $H(u) = E(\ln(X) - \ln(u) | X > u)$  et est donc calculé de manière empirique par  $\hat{H}(u) = \frac{\sum_{i=1}^n (\ln(X_i) - \ln(u)) \cdot \mathbb{1}_{X_i > u}}{\sum_{i=1}^n \mathbb{1}_{X_i > u}}$ .

Celui-ci s'avère tendre vers  $\xi$  lorsque  $X$  suit une loi de Fréchet.

### b) Application à la sélection du seuil des AP/TP extrêmes

Pour définir la valeur des AP/TP extrêmes, nous allons ainsi utiliser les deux outils précédemment cités. Le Mean Excess Plot sur les AP/TP de notre base nous donne la courbe suivante :

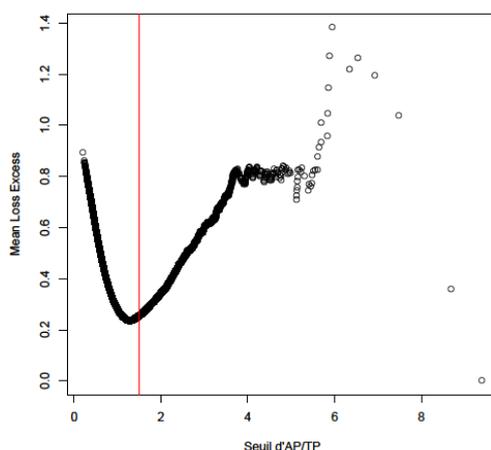


Figure 50: Mean Excess Loss des AP/TP

Le graphique ci-dessus nous permet de retenir 1.5 comme étant un seuil des extrêmes. En effet, sur le graphique du Mean Excess Loss, on observe trois tendances de la courbe associées à trois segments distincts d'AP/TP :

- Un premier segment ( $AP/TP < 1.5$ ) faisant apparaître une décroissance de la courbe. Cette tendance de la courbe est associée à l'adéquation des AP/TP de ce segment à une loi gamma. Cette observation vient confirmer l'hypothèse de sélection de la loi gamma lors de la modélisation des AP/TP par un GLM étant donné que les contrats caractérisés par un AP/TP ordinaire représentent l'essentiel de notre portefeuille en termes d'effectif.

- Un deuxième segment ( $1.5 < AP/TP < 4$ ) sur lequel la courbe du Mean Excess Loss est linéaire avec un coefficient directeur positif. On pourrait ainsi proposer l'appartenance à une distribution de type Fréchet à queue lourde.

- Un troisième segment ( $AP/TP > 4$ ) sur lequel la courbe du Mean Excess Loss est constante. Les observations dépassant ce seuil avaient, lors de la modélisation GLM, été assimilées à du bruit statistique et avaient été retirées de la base car considérées comme des valeurs exceptionnelles. Ce deuxième seuil sera ainsi conservé et ne rentrera pas en compte dans la suite de nos modélisations.

Le Mean Excess Loss nous permet ainsi de retenir 1.5 comme seuil des extrêmes.

L'estimateur de Hill des AP/TP du portefeuille admet l'allure suivante :

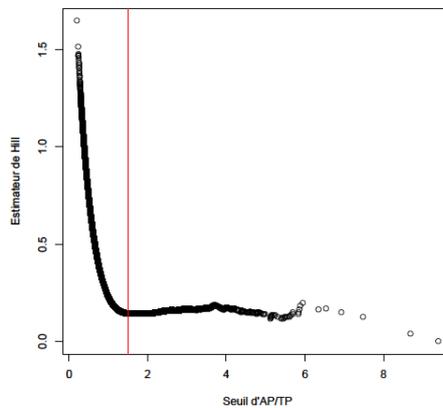


Figure 51: Estimateur de Hill des AP/TP

Le graphique de l'estimateur de Hill fait apparaître une certaine constance à partir du seuil de 1.5 environ. Cette observation nous permet ainsi de retenir 1.5 comme seuil et d'assimiler la loi des AP/TP extrêmes à une loi de type Fréchet.

Suite à l'utilisation de ces deux outils de la théorie des valeurs extrêmes, le seuil retenu est alors le suivant :

Seuil proposé	% contrats en portefeuille	% cotisations du portefeuille
1.5	8,59%	9,93%

## Partie 2 : Caractérisation des extrêmes

Dans cette partie, nous souhaitons comprendre quels sont les profils caractérisant cette rentabilité technique très favorable. La meilleure compréhension de ces contrats sur-rentables nous permettra ainsi de les identifier à la souscription ainsi que de piloter le portefeuille de manière à favoriser ces profils au détriment de ceux caractérisés par un AP/TP faible.

### a) Caractérisation en termes de primes

Nous nous sommes tout d'abord intéressés aux écarts de Tarif Technique et de Tarif Réel moyens qui apparaissent entre ces contrats sur-rentables et l'ensemble du portefeuille, pouvant expliquer les AP/TP élevés des contrats extrêmes :

	AP moyen	TP moyen
Total portefeuille	519.23	510.95
Ordinaires	511.62	526.63
Extrêmes	599.96	344.54

Figure 52: Tarif Réel et Tarif Technique moyen par classe d'AP/TP

Il s'avère finalement que les AP/TP extrêmes sont caractérisés par une combinaison de Tarif Réel élevé comparé à l'ensemble du portefeuille et de Tarif Technique au contraire moindre. Il s'agit de petits risques et payant une prime élevée en moyenne, comparée à leur risque mais aussi par rapport à l'ensemble du portefeuille.

#### b) Caractérisation en termes de variables descriptives

Nous souhaitons désormais relever les caractéristiques particulières communes aux profils d'AP/TP extrêmes. Pour ce faire, nous allons considérer le seuil des extrêmes défini précédemment à 1.5 et un deuxième seuil d'extrême permettant de réunir les contrats les 5% les plus rentables du portefeuille correspondant à un seuil d'AP/TP de 1.63. Pour ces études, nous testons les divers critères descriptifs disponibles dans notre base, représentant le risque observé sur le contrat ainsi que l'ensemble des modulations tarifaires qui lui sont accordées. Nous regardons ensuite pour chacun des critères disponibles quelles sont les modalités surreprésentées par rapport à l'ensemble du portefeuille et pour ces deux seuils. Parmi ces critères, certains semblent être particulièrement caractérisant des AP/TP extrêmes et sont présentés dans le tableau suivant :

		Total portefeuille	Seuil à 1.5	5% plus rentables
<b>Coefficient Bonus-Malus</b>	50	70,6%	68,0%	66,6%
	]50;60]	10,5%	10,6%	10,7%
	]60;70]	6,5%	6,7%	7,0%
	]70;80]	6,9%	7,3%	7,6%
	]80;90]	3,2%	3,3%	3,4%
	]90;100]	1,5%	<b>2,5%</b>	<b>2,9%</b>
	plus de 100	0,7%	<b>1,6%</b>	<b>1,8%</b>
<b>Classe prix</b>	Classe Prix A	9,6%	10,8%	10,7%
	Classe Prix F	12,1%	11,2%	10,8%
	Classe Prix H	18,2%	14,6%	14,3%
	Classe Prix J	19,3%	1,2%	17,8%
	Classe Prix L	10,1%	<b>18,3%</b>	8,2%
	Classe Prix M	9,9%	8,5%	9,0%
	Classe Prix N	7,0%	9,0%	7,6%
	Classe Prix O	4,5%	<b>7,6%</b>	<b>5,9%</b>
	Classe Prix P	4,5%	5,7%	<b>6,8%</b>
	Classe Prix RST	3,8%	<b>6,4%</b>	<b>7,7%</b>

	Classe Prix HC	1,1%	6,8%	1,4%
Energie	ES	31,1%	43,8%	45,8%
	GO	68,6%	56,2%	54,1%
	Autres	0,4%	0,2%	0,1%
Ancienneté du contrat	< 1 an	15,4%	1,9%	1,8%
	< 2 ans	12,9%	3,7%	3,4%
	< 3 ans	9,8%	4,5%	4,3%
	< 4 ans	8,3%	5,7%	5,3%
	< 5 ans	6,7%	5,9%	5,7%
	< 10 ans	20,0%	23,5%	22,9%
	< 15 ans	10,7%	17,6%	17,6%
	15 ans et +	16,3%	37,2%	39,1%
Formule	Formule aux Tiers + Formule aux Tiers réduite	10,0%	23,2%	24,3%
	Formule Intermédiaire réduite	7,0%	5,6%	5,0%
	Formule Intermédiaire	20,9%	28,2%	28,8%
	Formule Tous Risques	62,1%	43,1%	41,9%

Figure 53: Caractérisation des extrêmes à partir de variables explicatives

On note en particulier la sur-représentation, par rapport à l'ensemble du portefeuille, des coefficients bonus-malus supérieur à 100 sur les deux échantillons d'extrêmes formés par les deux seuils. Cette variable est particulièrement intéressante car bien que celle-ci avait été révélée non discriminante sur l'AP/TP lors de l'étude de l'ensemble du portefeuille, celle-ci s'avère être une variable clé de la caractérisation des contrats sur-rentables. Cette analyse relève à nouveau la forte rentabilité des contrats anciens, des formules moins couvrantes et des véhicules de classe de prix élevée ou véhicules classés « Hors Classe ». Enfin, les véhicules alimentés par essence sont associés à une rentabilité élevée.

Une fois avoir mieux compris de quoi étaient composés les contrats associés à une rentabilité extrême, nous souhaitons mettre en place des outils permettant d'accompagner l'aide à la décision dans des problématiques de détection de contrats sur-profitables. Pour ce faire nous allons faire appel à deux modèles de prévision statistique binaire : les régressions logistiques et les arbres de classification binaires.

### Partie 3 : Détection de profils associés à une rentabilité extrême

L'objectif de cette partie est de construire un modèle permettant la détection de profils caractérisés par une rentabilité extrême, c'est-à-dire ayant un AP/TP supérieur au seuil de 1.5. Ainsi, deux méthodes statistiques répandues permettant de réaliser des classifications binaires vont être utilisées : la régression logistique et l'arbre de classification CART binaire.

## a) Utilisation d'une régression logistique pour la détection de contrats extrêmes

- *Théorie sur les régressions logistiques*

Les régressions logistiques sont des cas particuliers des modèles linéaires généralisés. On s'intéresse ici à la régression logistique binaire, c'est-à-dire que l'on considère une variable cible prenant deux valeurs,  $Y=0$  ou  $Y=1$ , que l'on souhaite prédire à partir de  $p$  variables explicatives. La variable  $Y$  suit alors une loi de Bernoulli de paramètre  $p(x) = P(Y = 1|X = x)$ .

Notre variable  $Y$  sera, dans le cas de la détection des extrêmes, le fait qu'un contrat est caractérisé par un AP/TP extrême ( $Y=1$ ) ou caractérisé par un AP/TP ordinaire ( $Y=0$ ).

Le but de la régression logistique est, comme tout modèle linéaire généralisé, de prédire  $E(Y|X = x)$  à partir des variables explicatives disponibles. Cependant, étant donné que la variable  $Y$  est binaire, on a :  $E(Y|X = x) = 1 \cdot P(Y = 1|X = x) + 0 \cdot P(Y = 0|X = x) = P(Y = 1|X = x)$

La fonction de lien utilisée dans les régressions logistiques est la fonction de lien canonique associée à une variable à expliquer de loi Bernoulli. Il s'agit de la fonction logit, définie par  $g(y) = \ln\left(\frac{y}{1-y}\right)$ . Et puisque  $P(Y = 1|X = x) + P(Y = 0|X = x) = 1$ , le modèle logistique s'écrit sous la forme :

$$\ln\left(\frac{P(Y = 1|X = x)}{P(Y = 0|X = x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p.$$

Par conséquent, on a également  $P(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}$  et  $P(Y = 0|X = x) = \frac{1}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}$ .

La sélection des variables explicatives du modèle est réalisée à partir des mêmes outils que les modèles linéaires généralisés : Critère AIC, test de significativité de Wald des variables... L'estimation des coefficients est également réalisée par une méthode de maximisation de la vraisemblance. Le modèle ainsi construit se trouve en annexe.

On considèrera ici que le seuil standard de la règle de décision est égal à 0.5, c'est à dire qu'une observation sera classée positive dès lors que la valeur prédite de  $P(Y = 1|X = x)$  sera supérieure au seuil de 0.5 et classée négative dans l'autre cas.

Pour attester de la qualité du modèle linéaire construit, les mêmes outils que ceux présentés sur les modèles linéaires généralisés sont utilisés. On dispose cependant d'outils supplémentaires et spécifiques aux méthodes de régressions logistiques : les matrices de confusion ainsi que la courbe ROC (Receiver Operating Characteristic).

Pour un seuil de décision donné, la matrice de confusion confronte les valeurs observées de la variable cible  $Y$  avec celles qui sont prédites et comptabilise les bonnes et les mauvaises prédictions :

		Décision		
		Négatif: $\hat{Y}=0$	Positif: $\hat{Y}=1$	Total
Appartenance	Négatif: $Y=0$	Vrais négatifs (VN)	Faux positifs (FP)	Total des négatifs observés (VN + FP)
	Positif: $Y=1$	Faux négatifs (FN)	Vrais positifs (VP)	Total des positifs observés (FN + VP)
	Total	Total des négatifs prédits (VN + FN)	Total des positifs prédits (FP + VP)	Taille totale de l'échantillon (VN + FP + FN + VP = N)

Figure 54: Matrice de confusion

En plus de nous indiquer le taux d'erreur, la matrice de confusion nous permet de nous informer du type d'erreur réalisé en faisant appel à différents indicateurs :

- Le pourcentage de bien classés :  $\frac{VN + VP}{N}$
- Le pourcentage de mal classés :  $\frac{FN + FP}{N}$
- La sensibilité, qui correspond au taux de positifs détectés :  $\frac{VP}{VP + FN} = \frac{VP}{\#Positifs}$
- La spécificité, qui correspond au taux de négatifs détectés :  $\frac{VN}{VN + FP} = \frac{VN}{\#Négatifs}$
- La précision :  $\frac{\#Positifs}{N} \cdot Sensibilité + \frac{\#Négatifs}{N} \cdot Spécificité$

La courbe ROC représente la sensibilité en fonction du complémentaire de la spécificité, avec un seuil de décision qui varie entre 0 et 1. Cette courbe permet de visualiser le pouvoir discriminant d'un modèle. En effet, si la courbe ROC coïncide avec la diagonale le modèle ne sera pas plus performant qu'un modèle aléatoire, par contre, plus la courbe coïncide avec le coin carré supérieur gauche et plus le modèle sera performant étant donné qu'il capturera d'avantage de vrais positifs avec le moins possible de faux positifs.

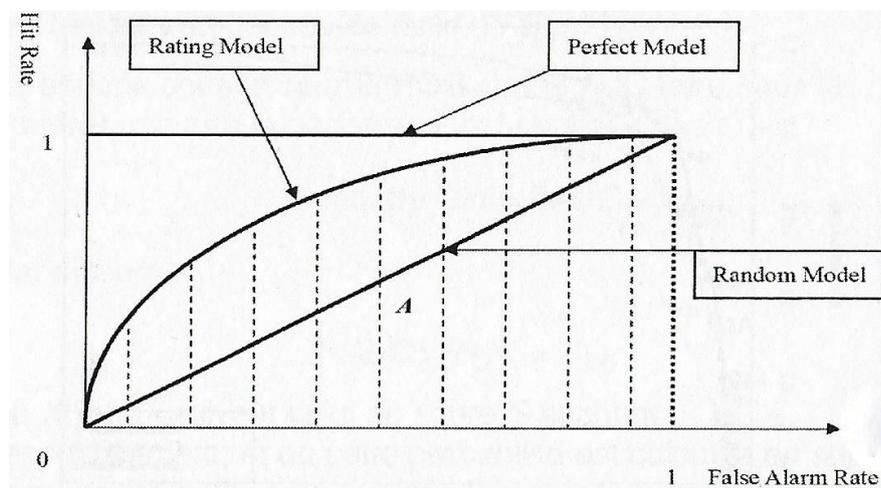


Figure 55: Courbes ROC et discrimination du modèle (Source: Bertrand P., Université Paris Dauphine)

C'est donc l'air sous la courbe, appelée AUC (Area Under the Curve) et hachurée sur le graphique ci-dessus, qui sera utilisée pour juger de la qualité du modèle :

AUC	Discrimination du modèle
=0.5	Aucune discrimination
]0.5 ;0.7[	Discrimination faible
[0.7 ;0.8[	Discrimination acceptable
[0.8 ;0.9[	Discrimination excellente
[0.9 ;1]	Discrimination parfaite

Figure 56: Air sous la courbe (AUC) et discrimination du modèle

- Application à la détection des extrêmes

Nous allons ici utiliser les régressions logistiques précédemment introduites afin de détecter les contrats qui s'avèrent être caractérisés par un AP/TP extrême. Un contrat sera alors classé positif lorsque l'on estime que celui-ci est associé à une rentabilité particulièrement élevée ( $AP/TP > 1.5$ ).

Le modèle a tout d'abord été élaboré sur l'échantillon d'apprentissage composé de 80% de la base de données.

L'ensemble des variables explicatives sont avant tout rendues binaires, c'est à dire qu'une variable à K modalités est remplacée par K variables binaires, chacune correspondant à une modalité de la variable.

Les variables explicatives suivantes ont été retenues suite à une procédure de sélection forward :

Variables explicatives retenues pour la régression logistique								
AncacqAcq_C	AncVehiAncV_J	AgeContACont_H	FormuleFor_CI	GPSRAGPSRA_G_H	FraBDGFBDG_X	ClauTariClTari_F	TauxBudgetTxbudg_A	TauxFirmeTxFirme_A

Figure 57: Variables explicatives retenues pour le scoring des AP/TP extrêmes par régression logistique

C'est sans surprise et conformément aux résultats obtenus dans la partie sur la caractérisation des extrêmes que l'on retrouve des variables concernant les modulations tarifaires, l'ancienneté du contrat et les formules aux tiers parmi les variables segmentant les contrats ayant un AP/TP ordinaire des contrats ayant un AP/TP extrême.

Lorsque nous appliquons notre modèle à l'échantillon test, nous obtenons un AUC de 0.839 et la courbe ROC suivante :

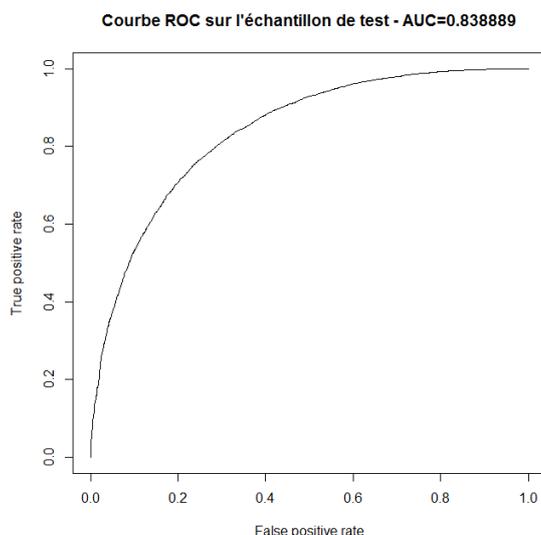


Figure 58: Courbe ROC du scoring par régression logistique sur l'échantillon test

La courbe ROC et l'AUC sont alors associés à un modèle ayant une discrimination excellente. Il faut cependant nous intéresser de manière plus précise au type d'erreur qui est réalisé en utilisant les différents indicateurs de qualité mis à notre disposition.

La matrice de confusion pour un seuil de décision égal à 0.5 est la suivante :

		Décision			
		Négatif	Positif	Total	
Appartenance	Négatif	315 435	2 418	317 853	% Bien classé: 91,72%
	Positif	26 380	3 468	29 848	% Mal classé: 8,28%
	Total	341 815	5 886	347 701	Sensibilité: 11,62%
					Spécificité: 99,24%
					Précision: 91,72%

Figure 59: Matrice de confusion du scoring par régression logistique sur la base de test

Le taux de bien classés du modèle est de 91.72% sur la base de test, ce qui semble à première vue être un résultat très satisfaisant et conforme aux bons résultats ayant été révélés précédemment par la courbe ROC.

Lorsque l'on s'intéresse aux autres indicateurs de qualité du modèle, on constate que la sensibilité est de 11.62%, qui correspond au taux de contrats ayant un AP/TP extrême affectés à la bonne classe, tandis que la spécificité est de 99.24%, qui correspond au taux d'AP/TP ordinaires bien classés. Le modèle décèle alors particulièrement bien les AP/TP ordinaires au détriment des contrats caractérisés par des AP/TP extrêmes qui sont très difficilement captés. De tels résultats sont souvent observables lorsque les deux classes n'ont pas les mêmes proportions en observé, le modèle tendant plus facilement à classer une observation dans la modalité dominante. Cependant, d'un point de vue métier, il faut garder en mémoire que nous souhaitons que le modèle identifie les contrats ayant un AP/TP extrême, mais avant tout qu'il ne prédise pas trop de contrats en extrêmes lorsqu'il s'agit de contrats ayant une valeur ordinaire d'AP/TP. Le modèle est ainsi satisfaisant.

## b) Utilisation d'un arbre de décision pour la détection de contrats extrêmes

- *Théorie sur les arbres de classification*

Le principe des arbres de décision consiste à segmenter successivement l'espace des variables explicatives disponibles en différentes classes homogènes du point de vue de la variable à prédire.

L'arbre de décision est un algorithme répandu en apprentissage statistique : son but est d'apprendre sur le lien entre la valeur de la variable réponse et les variables explicatives mises à disposition à partir de la base d'apprentissage. Sa robustesse est ensuite testée sur l'échantillon test. Cette méthode permet de détecter et relever facilement la structure sous-jacente responsable de la valeur de la variable à expliquer.

L'algorithme utilisé est l'algorithme CART (Classification And Regression Trees) développé en 1984 par L.Breiman. Cette méthode est utilisée pour des problèmes de classification lorsque la variable à expliquer est de type qualitatif, ou pour des problèmes de régression lorsque la variable à expliquer est continue.

Dans le cadre de la détection des AP/TP extrêmes, il s'agira d'un arbre de classification : un contrat sera classé 1 ou « positif » s'il est associé à un AP/TP extrême et 0 ou « négatif » s'il est associé à une rentabilité ordinaire. L'ensemble des variables explicatives a d'abord été discrétisé, c'est à dire qu'une variable à K modalités est remplacée par K variables binaires, chacune correspondant à une modalité.

Pour initier le partitionnement, l'algorithme cherche la variable qui permet le meilleur fractionnement de la base initiale en créant deux sous-groupes homogènes.

Pour un nœud  $i$ , on note :

-  $N_i$  le nombre d'éléments associés au nœud  $i$

-  $N_i^0$  le nombre d'éléments tels que  $Y=0$  au sein du nœud  $i$  et  $N_i^1$  le nombre d'individus tel que  $Y=1$  au sein du nœud  $i$

-  $P_i^0 = \frac{N_i^0}{N_i}$  la proportion d'éléments tels que  $Y=0$  au sein du nœud  $i$  et  $P_i^1 = \frac{N_i^1}{N_i}$  la proportion d'éléments tels que  $Y=1$  au sein du nœud  $i$

On évalue la séparation pour chacune des variables explicatives grâce à un indicateur qui mesure la pureté des nœuds. L'algorithme CART se base sur l'indice de Gini des nœuds, défini par :  $G(i) = \sum_{i \in \text{Noeud}} 1 - P_i^{0^2} - P_i^{1^2}$ . On sélectionnera ainsi la variable permettant d'obtenir des indices de Gini les plus faibles possibles pour les nœuds fils créés. Chacun des deux nœuds fils subira alors à leur tour une étape de fractionnement similaire et indépendante du nœud fils de droite au nœud fils de gauche.

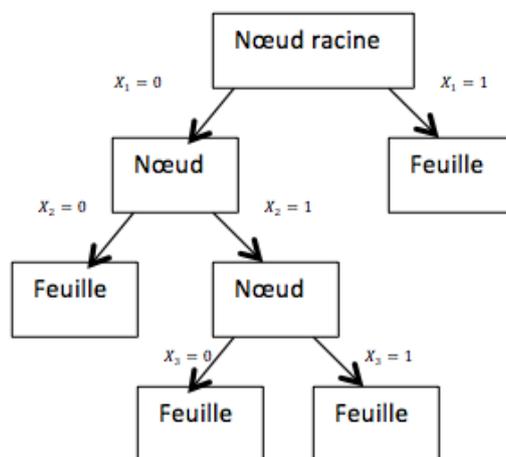


Figure 60 : Représentation d'un arbre de décision

Plus l'indice de Gini est faible, plus les éléments contenus dans le nœud sont homogènes et plus le nœud est pur. La pureté est alors maximale lorsque tous les individus au sein d'un même nœud possèdent la même modalité de la variable Y. Dans le cas d'un nœud pur, le fractionnement sera achevé au niveau de ce nœud qui formera désormais un nœud terminal appelé « feuille ».

Il existe divers critères d'arrêt : le nombre minimal d'individus dans une feuille, le nombre minimum d'individus dans un nœud pour réaliser une nouvelle étape de partitionnement, la profondeur maximale de l'arbre, le critère de complexité qui limite le sur-apprentissage induit lorsque l'arbre élaboré contient trop de feuilles. Le sur-apprentissage peut également être évité en réalisant dans un second temps, après l'établissement de l'arbre principal, une phase d'élagage de l'arbre.

Par ailleurs, l'arbre maximal construit suite à l'application de cet algorithme peut présenter d'excellentes qualités de prédiction sur l'échantillon d'apprentissage, puisque c'est celui-ci qui aura engendré la segmentation, mais peut être très mauvais prédicteur sur l'échantillon test. Cela se produit par exemple lorsque la profondeur de l'arbre maximum retenu est trop élevée. Il aura alors intégré des informations qui ont en réalité un faible pouvoir de séparation et qui sont finalement spécifiques aux données d'apprentissage. L'arbre maximal ainsi retenu aura du mal à généraliser son pouvoir de prédiction sur les nouvelles données constituant la base de test. L'intérêt de l'élagage est alors de remédier à l'overfitting de façon à ce que l'arbre admette une plus petite profondeur et que chacune des feuilles contienne plus d'information.

L'arbre de décision est une méthode qui a l'avantage d'être très visuelle puisqu'elle permet une lecture à plusieurs niveaux lorsque la profondeur n'est pas trop élevée. Elle permet de hiérarchiser l'importance d'une variable quant à sa discrimination sur la variable à expliquer. En effet, plus la variable intervient en début d'arbre et plus elle explique la variable à prédire. L'arbre retenu minimisera alors un compromis entre l'erreur réalisé par l'arbre et le nombre de feuilles terminales.

La matrice de confusion et la courbe ROC permettent également de juger de la qualité du modèle.

- *Application à la détection d'AP/TP extrêmes*

L'utilisation d'arbre de décision pour la détection d'AP/TP extrême a été réalisée à partir du package « RPart » (Recursive Partitioning) de R de T.Therneau qui permet de construire des arbres de décisions d'algorithme CART. Ce package inclut la mise en pratique des techniques de validation croisée et d'élagage d'arbre afin d'obtenir un arbre optimal en termes de qualité et de complexité de prédiction.

Le modèle de construction de l'arbre maximal s'appuie essentiellement sur la commande suivante permettant de définir la variable à prédire, les variables explicatives entrant dans la prédiction et les critères d'arrêt par la fonction `rpart.control` :

```
TAppren<- rpart (Score ~ AgeContACont_A + AgeContACont_B +..., data= appren2, control=rpart.control
(xval=10, minbucket=20, cp=0, maxdepth=6))
```

- *Xval* est le nombre de validations croisées réalisées permettant de générer l'arbre. La validation croisée consiste à subdiviser l'échantillon d'apprentissage en *k* sous-échantillons indépendants et de taille égale. Le modèle est construit sur l'ensemble formé par les *k-1* premiers échantillons, puis l'évaluation de la performance est réalisée sur le dernier échantillon. On réalise ensuite ces processus d'apprentissage et validation en excluant tour à tour chacun des *k-1* sous-échantillons restant. Enfin, le modèle retenu est sélectionné en minimisant l'erreur de prédiction des *k* configurations. *xval=10* indique ici que l'on va séparer la base d'apprentissage en 10 parties de manière aléatoire.
- *Minbucket* désigne le nombre minimal d'individus contenus dans un nœud terminal
- *Cp* est le paramètre de complexité qui est un critère d'arrêt calculé à chaque nouvelle création de nœud et est fonction de l'erreur d'apprentissage ainsi que du nombre de nœuds. Il permet ainsi de pénaliser les arbres dont la complexité est élevée. Nous n'imposons tout d'abord pas de contraintes sur ce paramètre, celui-ci intervenant plus tard lors de l'élagage de l'arbre.
- *Maxdepth* désigne la profondeur maximum de l'arbre. Nous avons ainsi choisi de réaliser un premier arbre de faible profondeur égale à 6 afin que sa représentation graphique soit ainsi facilement interprétable.

L'arbre maximal de profondeur 6 ainsi obtenu est ensuite élagué à partir de la commande suivante qui permet de retenir le nombre de feuille tel que l'erreur d'apprentissage est la plus faible:

```
TAppren_elague <- prune (TAppren, TAppren$cptable [which.min(TAppren$cptable[,4]),1])
```

Les variables qui interviennent le plus tôt dans le partitionnement de la base par l'arbre de classification et donc qui s'avèrent être les plus discriminantes sur l'indicateur à expliquer sont des modalités relatives à l'ancienneté du contrat, à l'âge du véhicule, à la formule souscrite ainsi qu'à la classe SRA du véhicule. Ces variables ont déjà été révélées comme étant pertinentes pour la segmentation des contrats ayant un AP/TP extrême des contrats ayant un AP/TP ordinaire.

Lorsque nous appliquons notre modèle à l'échantillon test, nous obtenons un AUC de 0.847 et la courbe ROC suivante :

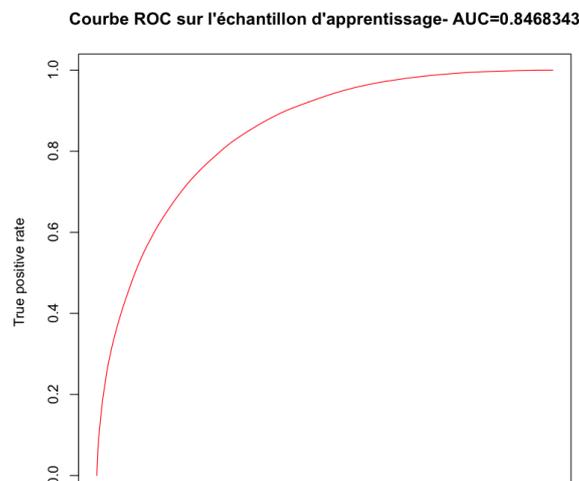


Figure 61: Courbe ROC du scoring par l'arbre de classification sur l'échantillon test

La courbe ROC et l'AUC sont alors associés à un modèle ayant une discrimination excellente. Il faut cependant nous intéresser de manière plus précise au type d'erreur qui est réalisé en utilisant les différents indicateurs de qualité du modèle mis à notre disposition.

La matrice de confusion obtenue grâce à l'arbre de classification, pour un seuil de décision égal à 0.5, est la suivante :

		Décision			% Bien classé: 91,86%	% Mal classé: 8,14%
		Négatif	Positif	Total		
Appartenance	Négatif	314 604	3 249	317 853	Sensibilité: 16,06%	Spécificité: 98,98%
	Positif	25 055	4 793	29 848		
	Total	339 659	8 042	347 701		

Figure 62: Matrice de confusion du scoring par arbre de classification sur la base de test

Les pourcentages de bien classés et de mal classés sont respectivement égaux à 91.86% et 8.14%, qui sont des résultats très satisfaisants. Lorsque l'on s'intéresse aux indicateurs complémentaires, nous constatons que la sensibilité est de 16.06%, qui correspond au taux de contrats caractérisés par un AP/TP extrême effectivement détectés, et que la spécificité est de 98.98%, qui correspond au taux de contrats associés à une rentabilité ordinaire détectés. Les résultats sont ici d'autant plus performants pour la détection des AP/TP effectivement extrêmes qu'avec la régression logistique. Cependant ce modèle a tendance à plus facilement classer les AP/TP en extrêmes pour un seuil de décision de 0.5 et donc notamment à classer en AP/TP extrêmes les contrats qui ont en réalité un AP/TP ordinaire, ce qui pourrait avoir une mauvaise incidence en cas d'accord de faveurs commerciales envers ces contrats qui ne sont finalement pas sur-rentables.

Pour mieux comparer les modèles, nous superposons leurs courbes ROC :

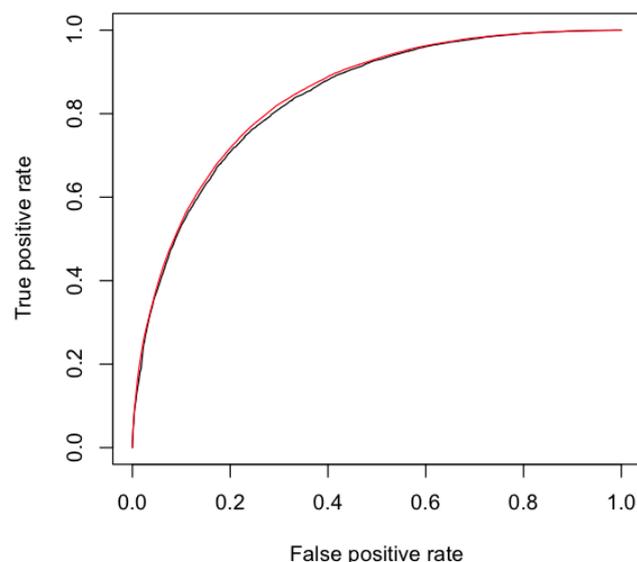


Figure 63: Comparaison des courbes ROC des deux modèles de score

La courbe noire correspond à la courbe ROC du modèle construit par la régression logistique et la courbe rouge correspond à la courbe ROC du modèle construit par l'arbre de décision. On constate que la courbe de l'arbre de classification est toujours légèrement au-dessus de la courbe de la régression logistique. L'arbre est ainsi plus performant quel que soit le seuil de décision.

Les deux modèles mis en place constituent une aide à la décision dans la mesure où ils nous permettent de pouvoir capter un maximum de contrats caractérisés par un AP/TP extrême. Des actions commerciales ciblées pourraient ainsi être réalisées en faveur de ces contrats, détectés comme apportant une certaine rentabilité à l'entreprise, au moment de leur souscription et permettant ainsi de capter ces bons assurés au sein du marché. Il faudrait cependant au préalable tester ces modèles sur d'autres années afin de s'assurer de leur fiabilité dans le temps. Une piste d'amélioration aurait été de travailler avec des arbres prenant en compte plus de variables ou en utilisant des méthodes plus complexes de Machine Learning telles que les forêts aléatoires ou gradient boosting, pour lesquels le modèle construit est plus indépendant de la structure des données de la base d'apprentissage puisqu'ils est basé sur la construction d'un grand nombre d'arbres et évite ainsi le sur-apprentissage. Ces modèles s'avèrent en général être plus robustes que les arbres de décision.

## CONCLUSION

L'objectif de cette étude était de fournir une meilleure compréhension de l'AP/TP, indicateur de rentabilité police à police nouvellement développé au sein de la Direction Technique IARD d'Allianz France dans un contexte de marché où la pression sur la rentabilité technique est de plus en plus forte (concurrence vive, mise en place de la loi Hamon, taux financier bas). Ce mémoire s'est tout particulièrement appuyé sur le portefeuille automobile de particuliers distribué par le biais d'agents généraux, qui est le portefeuille le plus important dont la Direction Technique IARD est à charge (environ 1.7 millions de contrats à fin 2016) et sur lequel l'utilisation de l'AP/TP a été la plus étendue.

La grande particularité de l'AP/TP est qu'il permet de mesurer une rentabilité assurantielle sur la maille très fine qu'est le contrat, lui accordant ainsi une réelle valeur ajoutée par rapport aux indicateurs classiques utilisés en assurance IARD (S/P, CoR). Par ailleurs, sa pertinence dans le pilotage de portefeuille IARD lui a permis de rapidement se voir accorder une place très importante dans la gestion de ces derniers. L'AP/TP intervient désormais lors des revalorisations annuelles des cotisations ainsi que lors de la surveillance de portefeuille.

Diverses méthodes statistiques descriptives et de modélisation de type modèles linéaire généralisés ont été mises en place afin de repérer certains profils particuliers présentant une sur ou sous-rentabilité. Nous avons notamment souligné l'importance de l'ancienneté du contrat, de la formule souscrite ou encore des modulations tarifaires. De plus, la stabilité de ces modèles dans le temps nous conforte dans la significativité des résultats obtenus, permettant ainsi de les implémenter lors de la souscription d'un nouveau contrat ainsi que d'intégrer au mieux l'AP/TP ou les critères discriminant sur cette rentabilité technique dans la gestion des portefeuilles d'assurance IARD.

La modélisation par les GLM de l'ensemble des AP/TP du portefeuille trouve ses limites dans la modélisation des extrêmes par la présence d'une queue de distribution lourde à droite. Cela nous a montré qu'il existait des profils se distinguant du reste du portefeuille et associés à des AP/TP qualifiés d'« extrêmes » qui nécessitaient une analyse plus approfondie. Ils représentent 8.5% des contrats en portefeuille pour 10% des cotisations. Nous avons ainsi pu constater que les contrats associés à des AP/TP extrêmes sont caractérisés par une ancienneté des contrats, par des formules peu complètes et par des classes de prix élevés notamment. Deux modèles de scoring ont également été mis en place afin de détecter ces contrats apportant une rentabilité extrême. Sous réserve de stabilité dans le temps, ces modèles prometteurs pourraient être intégrés dans le processus de sélection des affaires nouvelles et dans le pilotage du fond de portefeuille.

Ce mémoire nous a ainsi confortés dans l'idée que l'AP/TP était un indicateur pertinent pour le pilotage de portefeuille d'assurance IARD et nous encourage d'avantage à développer son utilisation sur d'autres périmètres tels que l'assurance habitation.

## Références bibliographiques

- FFA [2016] *Les Assurances de Biens et de Responsabilité, données clés 2015*
- FFA [juin 2017] *Le marché de l'assurance automobile en 2016*
- ACPR [juillet 2013] *La situation des principaux organismes d'assurance en 2012*
- Suru A. [2016-2017] *Notes de cours de l'Université Paris Dauphine, Principes de base de l'assurance dommage*
- Charpentier A. [2010] *Notes de cours de l'Université de Rennes 1 et Université de Montréal, Statistiques de l'assurance*
- Bertrand P. [2016-2017] *Notes de cours de l'Université Paris Dauphine, Analyse de données et scoring*
- Biernat E., Lutz M. [2015] *Data science : fondamentaux et études de cas*
- Denuit M., Charpentier A. [2005] *Mathématiques de l'assurance Non-Vie – Tome II : Tarification et provisionnement*
- T. Therneau [Mars 2017] *An introduction to recursive partitioning using the RPART routines*
- <https://support.sas.com/documentation/> consultation de manière quotidienne pour les procédures du logiciel SAS

## Table des illustrations

Figure 1: Evolution du ratio combiné en assurance de biens et de responsabilité (Source: FFA).....	17
Figure 2: Répartition des cotisations des particuliers par branches en 2016 (Source : FFA) .....	17
Figure 3: Evolution du ratio combiné en assurance automobile (Source : ACPR, FFA) .....	18
Figure 4: Evolution des produits financiers nets de charges sur le total des placements (Source: ACPR) .....	18
Figure 5: Evolutions du taux de résiliations et du taux d'affaires nouvelles sur le marché de l'assurance automobile (Source : FFA) .....	19
Figure 6: Présentation des différentes primes assurantielles.....	22
Figure 7: Explication théorique du principe de calibration du Tarif Technique.....	29
Figure 8: Grille de revalorisation du Tarif Réel à partir de l'ETP .....	33
Figure 9: Récapitulatif du fonctionnement de l'ETP en cas d'avenant.....	35
Figure 10: Exemple de mécanisme d'ETP avec un contrat sous-tarifé et un risque qui augmente .....	36
Figure 11: Exemple de mécanisme d'ETP avec un contrat sous-tarifé et un risque qui diminue.....	36
Figure 12: Comparaison des indicateurs de rentabilité classique avec l'AP/TP .....	38
Figure 13: Fonctionnement de la surveillance de portefeuilles faisant intervenir l'AP/TP .....	39
Figure 14: AP/TP et suivi de la rentabilité technique au cours de la vie d'un contrat.....	40
Figure 15: Fonctionnement des revalorisations additionnelles faisant intervenir l'AP/TP .....	40
Figure 16: Ensemble de variables de la base initiale sélectionnées pour notre étude .....	45
Figure 17: Distribution empirique de l'AP/TP du portefeuille « 4RP Agents ».....	46
Figure 18: Statistiques sur l'AP/TP du portefeuille « 4RP Agents » .....	46
Figure 19: Récapitulatif des classes d'AP/TP du portefeuille « 4RP Agents » .....	47
Figure 20: Zooms sur la distribution empirique des classes d'AP/TP du portefeuille « 4RP Agents ».....	48
Figure 21: Récapitulatif de la densité de l'AP/TP du portefeuille « 4RP Agents ».....	49
Figure 22: Analyse uni-variée du Tarif Technique et effectif par Age.....	50
Figure 23: Analyse uni-variée de l'AP/TP et effectif par Age.....	51
Figure 24: Analyse uni-variée de l'AP/TP et effectif par Ancienneté du contrat.....	52
Figure 25: Présentation des différentes formules de couvertures disponibles .....	53
Figure 26: Analyse uni-variée de l'AP/TP et effectif par formule souscrite.....	54
Figure 27: Analyse uni-variée de l'AP/TP et effectif par présence de l'offre double auto.....	55
Figure 28: Ensemble des variables conservées suite aux analyses uni-variées .....	56
Figure 29: Interprétation de la force du lien entre deux variables à partir du V de Cramer .....	57
Figure 30: Visualisation des principaux liens entre les variables du support (V de Cramer > 0 .4) .....	58
Figure 31: V de Cramer des variables du support avec l'AP/TP .....	60
Figure 32: Inertie des 15 premiers axes de l'ACM.....	62
Figure 33: Représentation des modalités sur les axes 1 et 2 de notre ACM .....	64
Figure 34: Représentation des modalités sur les axes 3 et 4 de notre ACM .....	66
Figure 35: Tests d'adéquation de la densité empirique de l'AP/TP avec la loi Gamma .....	78
Figure 36: Résultat de la sélection forward des variables explicatives du modèle d'AP/TP sous Emblem .....	80
Figure 37: Variables explicatives retenues pour la modélisation de l'AP/TP .....	81
Figure 38: Résultats des tests de Wald des variables explicatives sur la base d'apprentissage .....	82
Figure 39: Exemple de relativité du modèle avec l'ancienneté du contrat .....	83
Figure 40: Statistiques sur les résidus bruts du modèle d'AP/TP sur l'échantillon d'apprentissage .....	84
Figure 41: Résidus de Pearson du modèle d'AP/TP sur l'échantillon d'apprentissage.....	84
Figure 42: Histogramme des résidus de Pearson du modèle d'AP/TP sur l'échantillon d'apprentissage.....	85
Figure 43: Test d'adéquation des résidus bruts à la loi normale sur l'échantillon d'apprentissage.....	85
Figure 44: Résidus de Pearson du modèle sur l'échantillon test.....	86

Figure 45: Histogramme des résidus de Pearson du modèle d'AP/TP sur l'échantillon test .....	87
Figure 46: Test d'adéquation des résidus de Pearson du modèle d'AP/TP à la loi normale sur l'échantillon test	87
Figure 47: Histogrammes des résidus de Pearson du modèle d'AP/TP sur les bases de 2014 (à gauche) et 2015 (à droite) .....	88
Figure 48: Distribution des extrêmes généralisés et lois classiques associées .....	91
Figure 49: Allure du Mean Excess Loss et distributions associées .....	91
Figure 50: Mean Excess Loss des AP/TP .....	92
Figure 51: Estimateur de Hill des AP/TP.....	93
Figure 52: Tarif Réel et Tarif Technique moyen par classe d'AP/TP .....	94
Figure 53: Caractérisation des extrêmes à partir de variables explicatives.....	95
Figure 54: Matrice de confusion.....	97
Figure 55: Courbes ROC et discrimination du modèle (Source: Bertrand P., Université Paris Dauphine) .....	98
Figure 56: Air sous la courbe (AUC) et discrimination du modèle.....	98
Figure 57: Variables explicatives retenues pour le scoring des AP/TP extrêmes par régression logistique.....	98
Figure 58: Courbe ROC du scoring par régression logistique sur l'échantillon test .....	99
Figure 59: Matrice de confusion du scoring par régression logistique sur la base de test .....	99
Figure 60 : Représentation d'un arbre de décision.....	101
Figure 61: Courbe ROC du scoring par l'arbre de classification sur l'échantillon test .....	102
Figure 62: Matrice de confusion du scoring par arbre de classification sur la base de test .....	103
Figure 63: Comparaison des courbes ROC des deux modèles de score.....	103

# Annexes

## Annexe 1: Analyses uni-variées de l'AP/TP selon les variables

- Variables sur le conducteur assuré:

- Age du conducteur principal

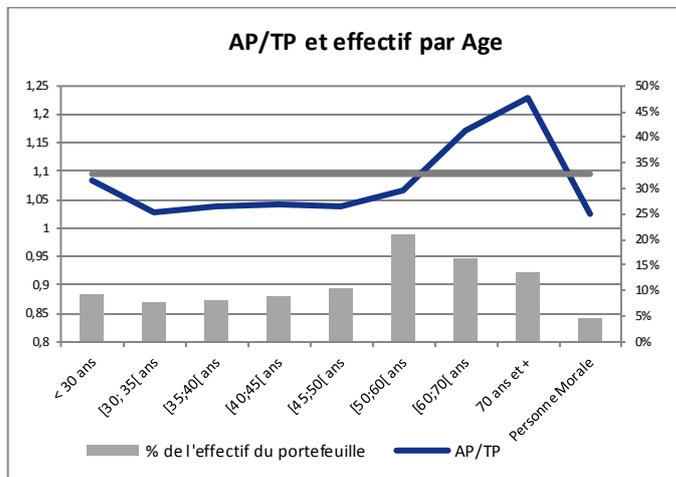
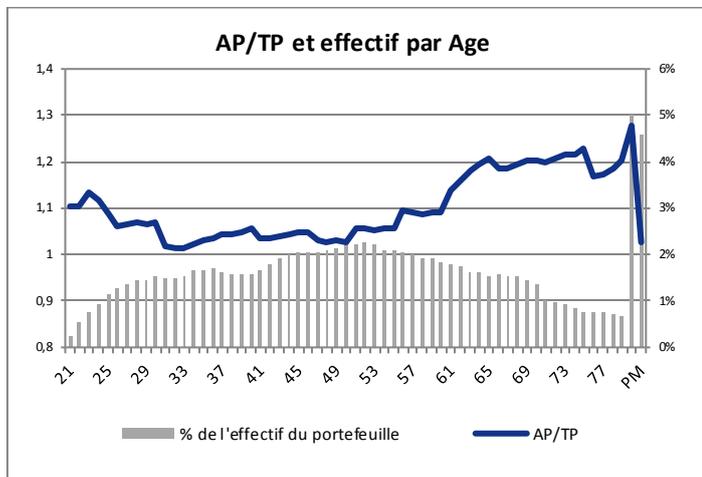


Figure : AP/TP et effectif par âge du conducteur

On observe une certaine tendance de la courbe d'AP/TP, avec une décroissance de l'AP/TP avec l'âge du conducteur, jusqu'à 50 ans puis une augmentation progressive. Cependant, l'âge du conducteur étant une variable continue, des regroupements vont être effectués afin d'obtenir des modalités présentant chacune un effectif suffisant du portefeuille. Après regroupement par tranche d'âge de l'assuré, l'âge du conducteur paraît discriminant sur l'AP/TP. On relève une certaine rentabilité des 60 ans et plus. Les Personnes Morales, qui sont les personnes non rattachables à une personne physique (taxis, profession libérale, ...), sont associés à une rentabilité faible par rapport à l'ensemble du portefeuille.

- Catégorie Socio-Professionnelle

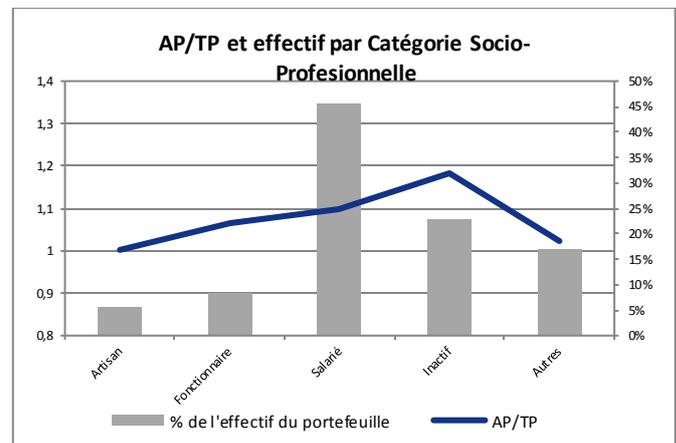
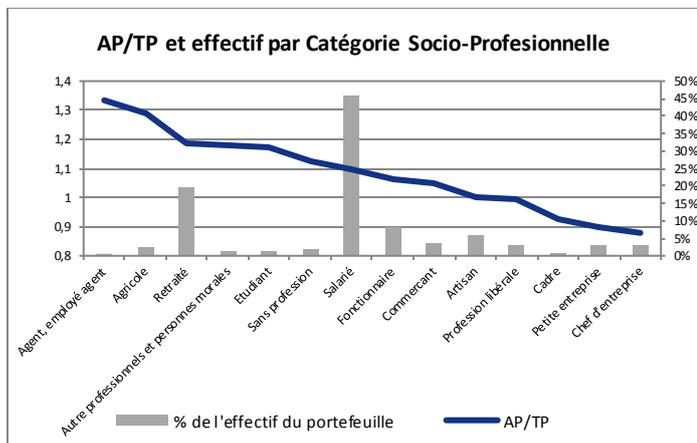


Figure : AP/TP et effectif par Catégorie Socio-Professionnelle

Des niveaux d'AP/TP différents sont obtenus d'un segment à l'autre de Catégorie Socio-Professionnelle, cette dernière semble donc avoir une influence sur l'AP/TP. Cependant, un certain nombre de modalités de la variable CSP ne sont pas suffisamment représentées. Des regroupements vont être effectués par proximité des valeurs de l'AP/TP et du type de catégorie.

Les modalités après regroupement sont alors les suivantes :

- Artisan
- Fonctionnaire
- Salarié
- Inactif : étudiant, retraité et sans profession
- Autres

La variable Catégorie Socio-Professionnelle paraît discriminante : Les inactifs ressortent comme étant sur-tarifés, contrairement aux

artisans et à la modalité « autres ».

- Coefficient Bonus-Malus

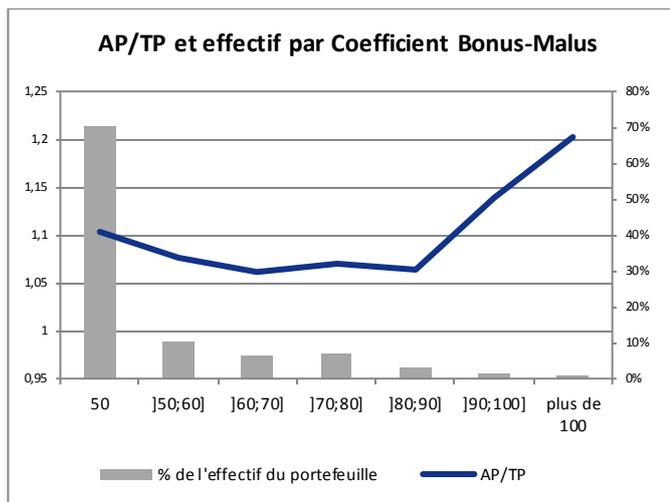


Figure : AP/TP et effectif par Coefficient Bonus-Malus

Les polices faisant référence à des Coefficients Bonus-Malus élevés apparaissent comme étant des profils dégageant de la rentabilité. Cependant ces modalités ne sont pas suffisamment représentées et aucun regroupement ne permettrait de rendre la variable discriminante sur l'AP/TP. La variable Coefficient Bonus-Malus ne sera ainsi pas retenue.

- Age d'obtention du permis

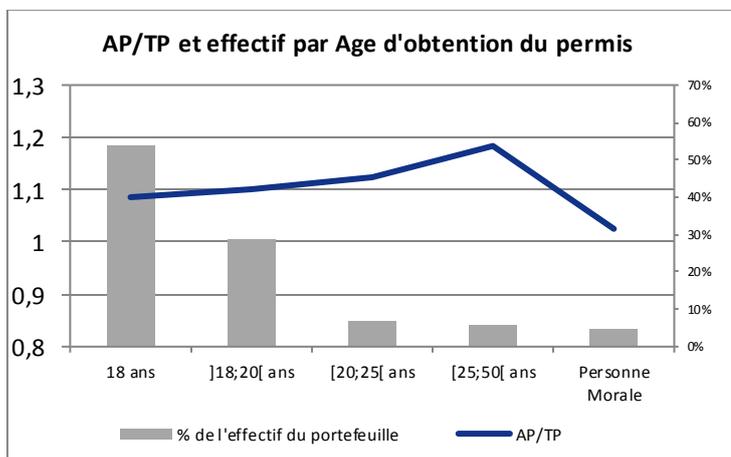


Figure : AP/TP et effectif par Ancienneté du permis

Après regroupement en classe d'âge à l'obtention du permis, cette variable paraît discriminante sur l'AP/TP. On relève une certaine rentabilité des conducteurs ayant obtenu leur permis à 25 ans et plus.

- Antécédent Sinistre

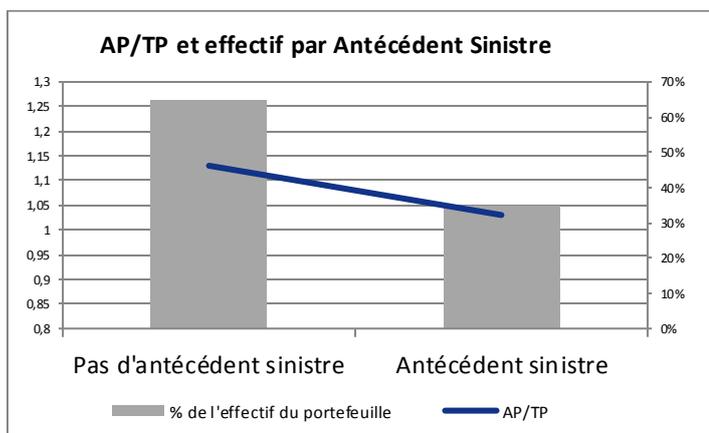


Figure : AP/TP et effectif par Antécédent Sinistre

L'AP/TP moyen varie de manière significative d'une modalité à l'autre de l'antécédent sinistre: les polices n'ayant pas d'antécédent sinistre sur les 36 derniers mois dégagent une meilleure rentabilité que ceux ayant eu des sinistres sur la période. La variable antécédent sinistre sera conservée.

- Sexe

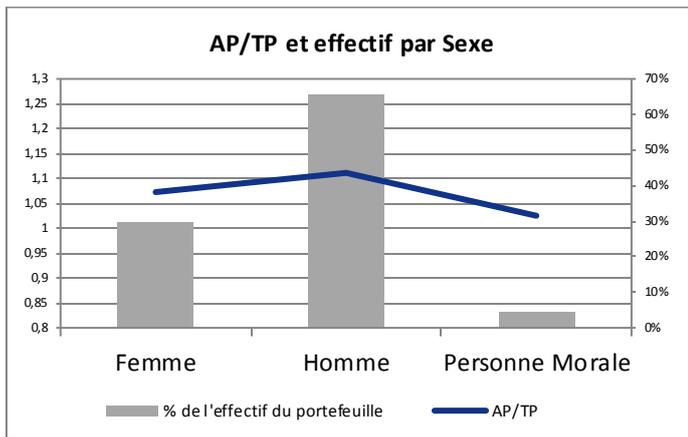


Figure : AP/TP et effectif par Sexe

Le niveau d'AP/TP moyen des Femmes se distingue peu de celui des Hommes. La modalité « Personne Morale », qui affiche une rentabilité faible comparé aux autres modalités de la variable « sexe » est déjà retenue dans l'âge du conducteur. La variable sexe n'apparaît pas comme étant discriminante et ne sera ainsi pas retenue.

On ne souhaiterait de plus pas conserver cette variable puisqu'aucune discrimination ne pourrait être autorisée sur la variable « sexe » lors du pilotage des tarifs du portefeuille.

- Variables sur le véhicule assuré:

- Marque

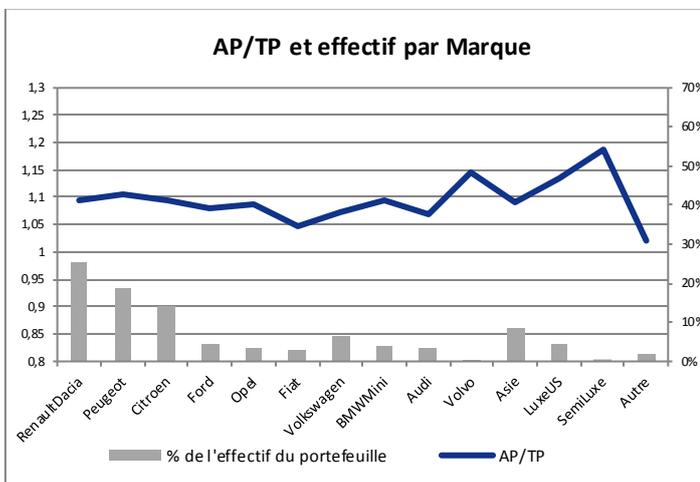


Figure : AP/TP et effectif par Marque

On observe des pics de valeurs d'AP/TP moyen sur certaines marques, notamment sur les marques de « Semi Luxe » et les « Volvo ». Cependant, celles-ci sont peu représentées et aucun regroupement ne peut être réalisé sur cette variable. La variable « Marque » ne sera pas prise en compte par la suite.

- Classe de réparation

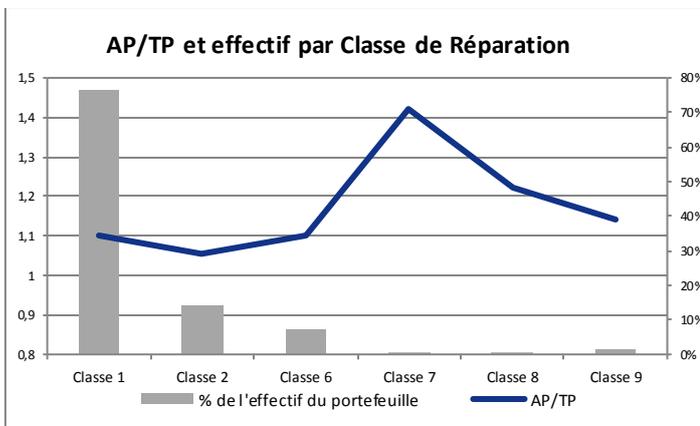


Figure : AP/TP et effectif par Classe de réparation

On observe des pics de valeurs d'AP/TP moyen sur certaines classes de réparation. Cependant, celles-ci sont sous-représentées et aucun regroupement ne peut être réalisé. La variable « Classe de réparation » ne sera pas prise en compte par la suite.

- Ancienneté du véhicule

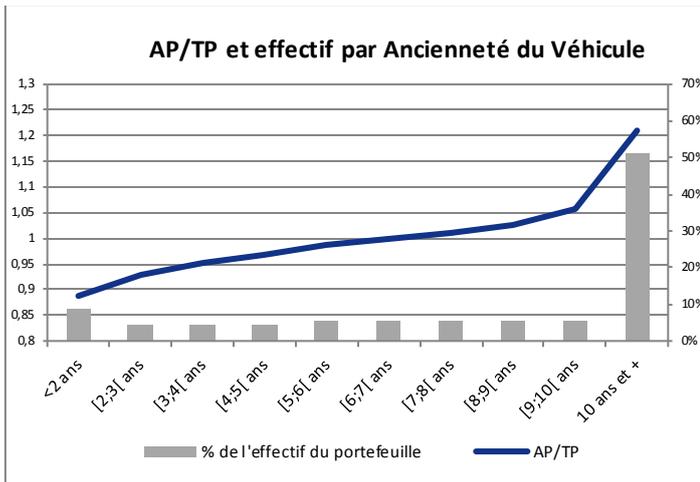


Figure : AP/TP et effectif par ancienneté du véhicule

Après regroupement en classe d'ancienneté du véhicule, cette variable paraît discriminante sur l'AP/TP. On relève une croissance nette de la rentabilité avec l'ancienneté du véhicule, avec une sous-rentabilité pour les véhicules récents et une sur-rentabilité pour les véhicules anciens. Cette variable sera ainsi par la suite retenue.

Remarque : On aurait souhaité avoir plus de tranches d'ancienneté du véhicule pour les véhicules datant de 10 ans afin de perdre moins d'information sur cette variable qui paraît fortement discriminante sur l'AP/TP.

- Propreté du véhicule

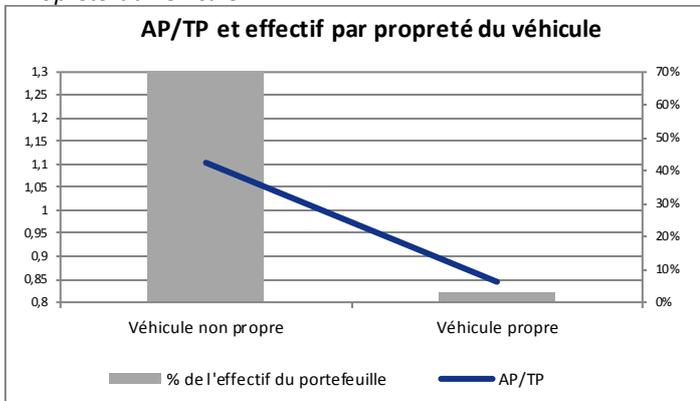


Figure : AP/TP et effectif par propreté du véhicule

Les véhicules propres font référence aux véhicules de moins de 2 ans et bénéficiant du bonus écologique à la date de souscription (à ce jour émission inférieure ou égale à 60 g/km de CO<sub>2</sub>). Bien que des niveaux d'AP/TP assez distincts en fonction de la modalité sont observés, les véhicules propres ne sont pas en effectif suffisant pour que les analyses soient significatives. La variable ne pourra être retenue.

- Classe de prix

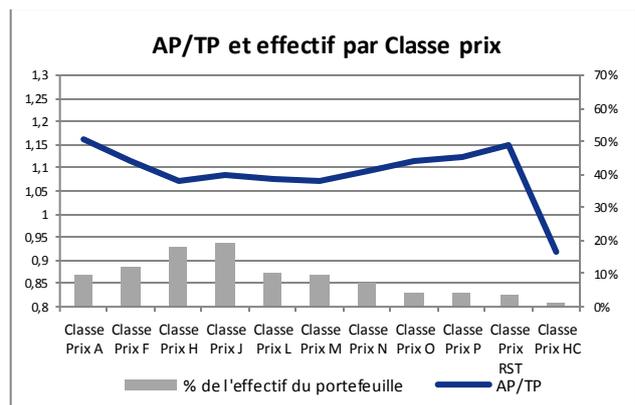
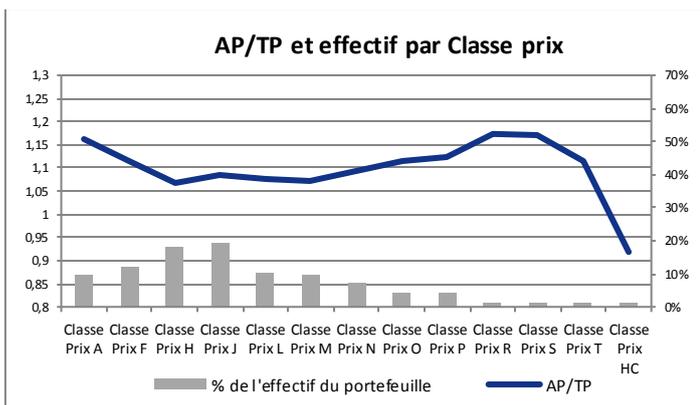


Figure : AP/TP et effectif par Classe de Prix

Les classes de prix R, S et T sont sous représentées et sont alors regroupées afin que les discriminations de la variable soient rendus significatives.

Après regroupement, la variable Classe de Prix paraît discriminante sur l'AP/TP et sera ainsi conservée. On note une forte rentabilité commerciale des Classes A et R/S/T. Les véhicules « Hors Classe » quant à eux paraissent très peu rentables.

- Taux d'émission de CO2

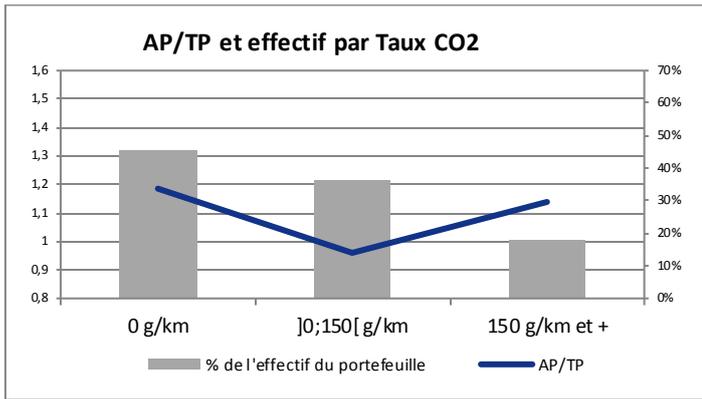


Figure : AP/TP et effectif par Taux d'émission de CO2

Plus de 45% du portefeuille apparaît en modalité 0 pour taux d'émission de CO2. Cela correspond probablement à une mauvaise alimentation de la base SRA sur la variable « Taux d'émission de CO2 » qui donne la modalité 0 par défaut. La variable « taux d'émission de CO2 » ne pourra ainsi pas être exploitée.

- Groupe SRA

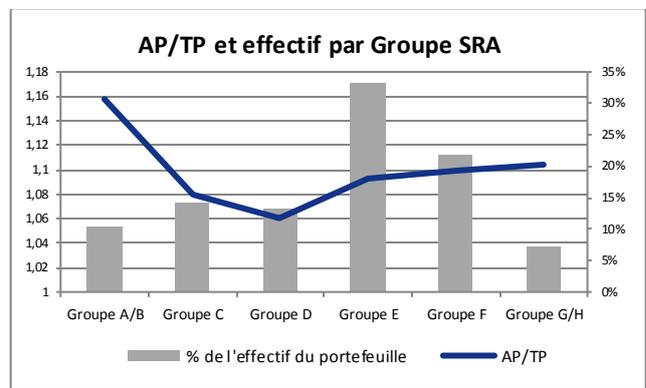
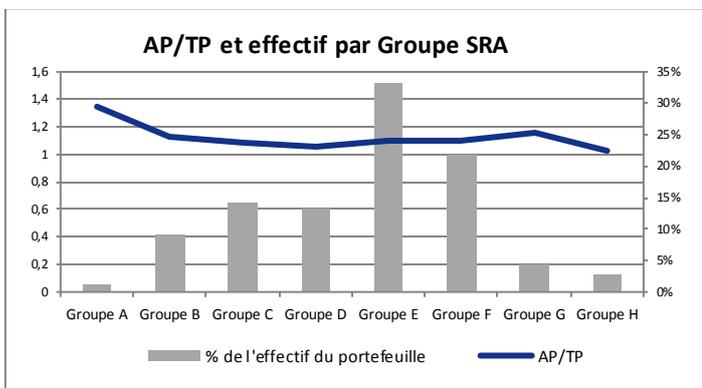


Figure : AP/TP et effectif par Groupe SRA

Un certain nombre de modalités de groupe SRA sont sous-représentées. Les groupes A et B ainsi que les groupes G et H vont être regroupés. Les Classes A/B apparaissent comme étant particulièrement rentables techniquement. La variable « Groupe SRA » va être conservée.

- Ancienneté d'acquisition du véhicule

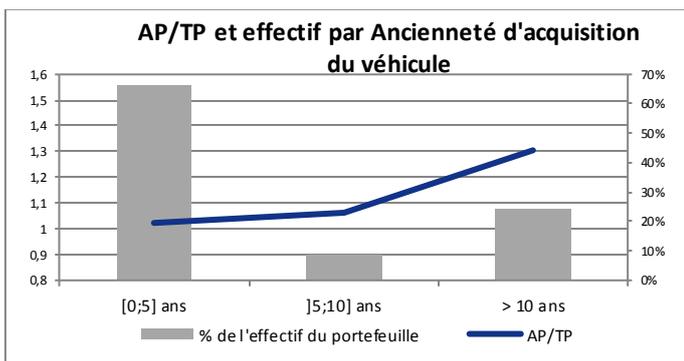


Figure : AP/TP et effectif par ancienneté d'acquisition du véhicule

Les polices dont l'acquisition du véhicule date de plus de 10 ans apparaissent comme étant particulièrement rentables en terme d'AP/TP. La variable « Ancienneté d'Acquisition du Véhicule » paraît discriminante sur l'AP/TP.

- Classe de vol

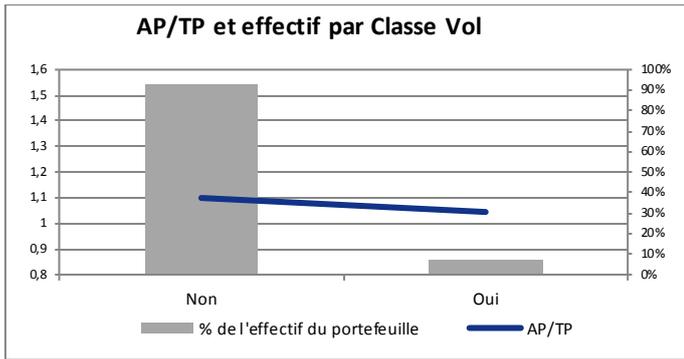


Figure : AP/TP et effectif par Classe de Vol

Un certain nombre de véhicules sont répertoriés par Allianz comme étant fréquemment volés. La présence d'un véhicule sur cette liste fait l'objet d'une variable dans notre base. On ne constate pas d'écart important d'AP/TP sur les deux modalités de la variable « Classe Vol ». La variable Classe Vol ne sera pas retenue par la suite.

- Energie

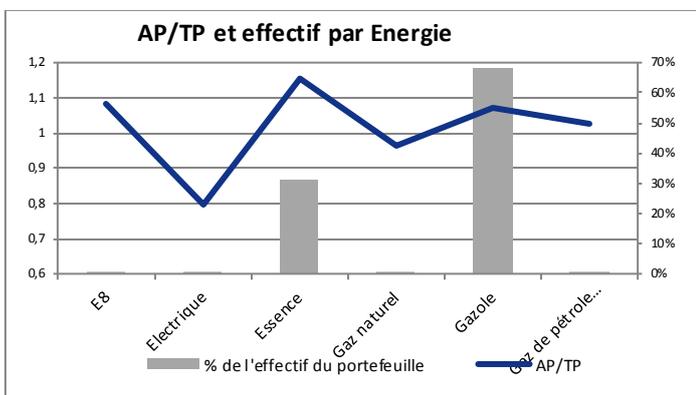
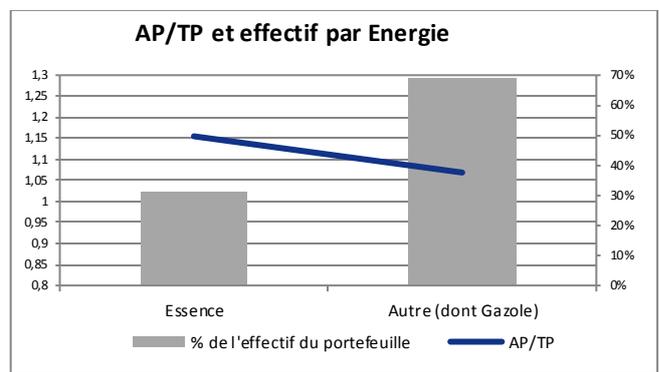


Figure : AP/TP et effectif par Energie



Un certain nombre de modalités d'énergie du véhicule étant sous-représentées, des regroupements doivent être effectués. Etant donnée la plus forte proximité des valeurs d'AP/TP, nous allons considérer deux modalités d'énergies :

- Essence
- Autres (dont Gazole)

On note une plus forte rentabilité en terme d'AP/TP des véhicules alimentés par essence par rapport aux autres modes d'alimentation comportant essentiellement les véhicules à gazole.

La variable energie paraît discriminante sur l'AP/TP.

- Carrosserie

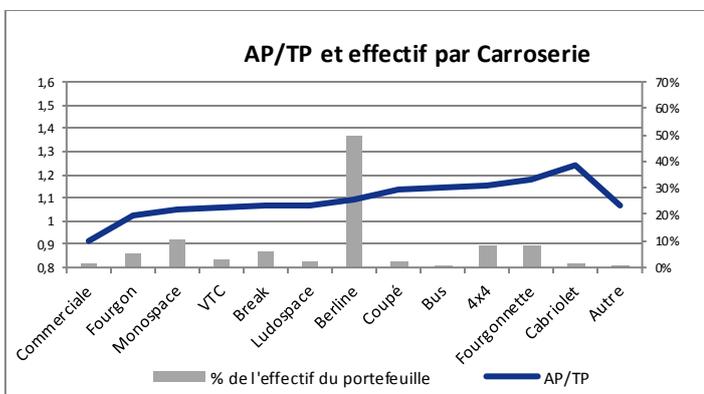
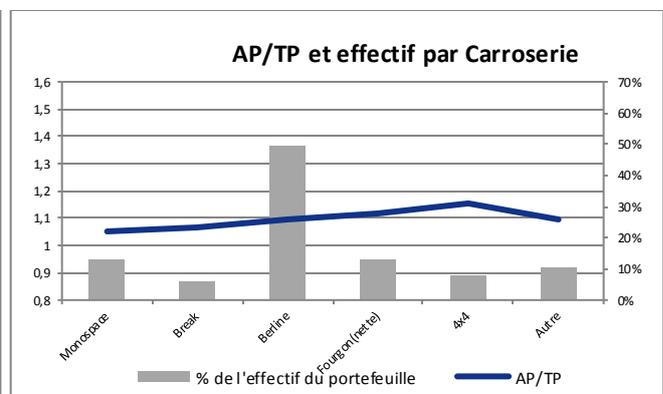


Figure: AP/TP et effectif par Carrosserie



La modalité « Commerciale » de la variable « Carrosserie » apparaît comme étant une modalité sous-tarifée et donc sous-rentable techniquement. Au contraire, la rentabilité de la modalité « Cabriolet » semble être très profitable. Cependant, certaines modalités ne sont pas suffisamment représentées et vont nécessiter des regroupements.

Après avoir effectué les regroupement, la variable « Carrosserie » paraît discriminante sur l'AP/TP.

- Note de sécurité

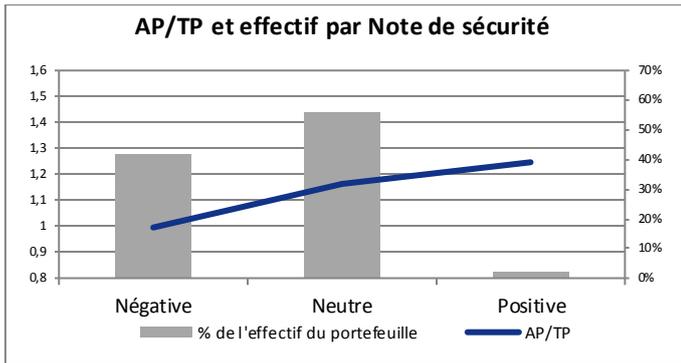


Figure: AP/TP et effectif par Note de Sécurité

La note de sécurité est une donnée fournie par la Société de Sécurité et de Réparation Automobile, prenant en compte la présence de matériel de prévention et de sécurité tels que des airbags dans le véhicule. Plus elle est faible, plus le véhicule est bien équipé.

On constate que plus la note de sécurité est élevée (donc plus l'équipement de sécurité du véhicule est faible), plus la rentabilité est élevée. La variable Note de Sécurité semble être discriminante sur l'AP/TP.

- Variables sur la géographie

Les zones de risques des différentes garanties sont des cartographies de risque à l'échelle de la commune qui ont été établies par un service de la Direction Technique IARD en charge du géocoding.

- Zone de la garantie Responsabilité Civile

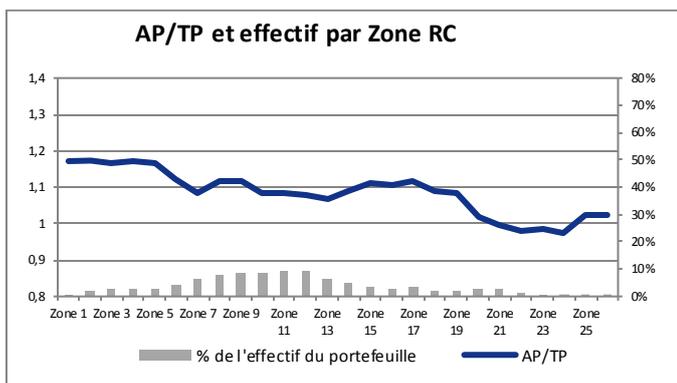
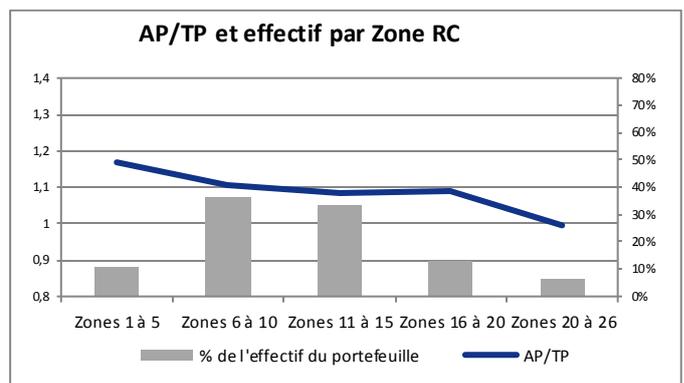


Figure: AP/TP et effectif par Zone RC



Certaines modalités de la variable Zone RC n'étant pas en nombre suffisant, nous devons procéder à des regroupements afin que les résultats soient significatifs.

Une fois les regroupements de zones réalisés, on observe une décroissance de l'AP/TP avec le risque de la Zone RC. La variable Zone RC semble discriminante et sera ainsi conservée.

- Zone de la garantie Bris de Glace

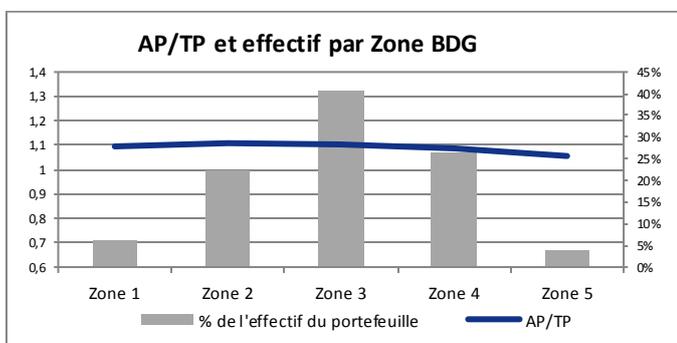


Figure: AP/TP et effectif par Zone BDG

On constate que l'écart d'AP/TP entre les Zones Bris de Glace est très faible. La variable est peu discriminante sur l'AP/TP et ne sera donc pas conservée.

- Zone des garanties Vol et Incendie

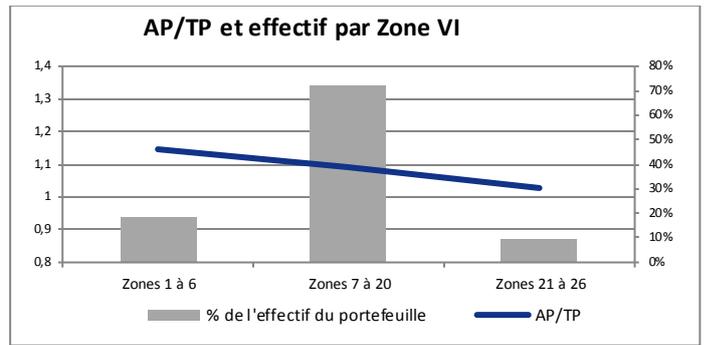
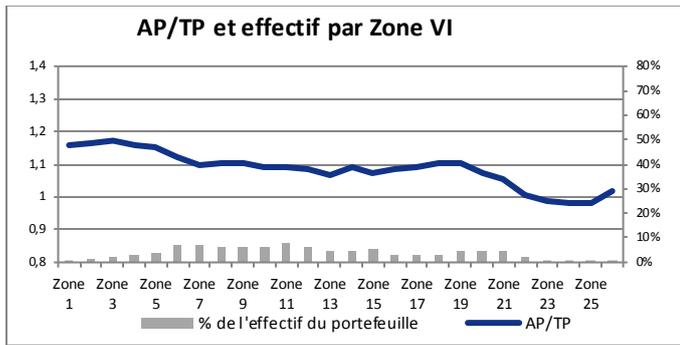


Figure: AP/TP et effectif par Zone VI

La zone de garantie Vol/Incendie semble être discriminante sur notre indicateur de rentabilité technique. Les modalités sont cependant sous représentées et demandent ainsi la réalisation de regroupements de zones.

Après regroupement, on constate que plus la zone Vol/Incendie est risquée, plus l'AP/TP est faible. La variable Zone VI est discriminante sur l'AP/TP.

- Zone de la garantie Dommage

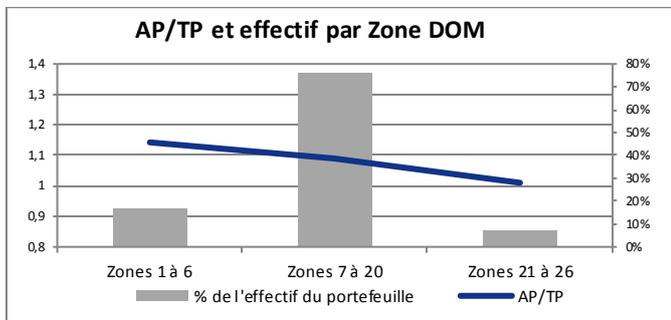


Figure : AP/TP et effectif par Zone DOM

De même, après regroupement des zones Dommages, on constate que plus la zone est risquée et plus l'AP/TP moyen est faible. La variable « Zone DOM », jugée discriminante sur notre indicateur, sera par la suite conservée.

- Garage

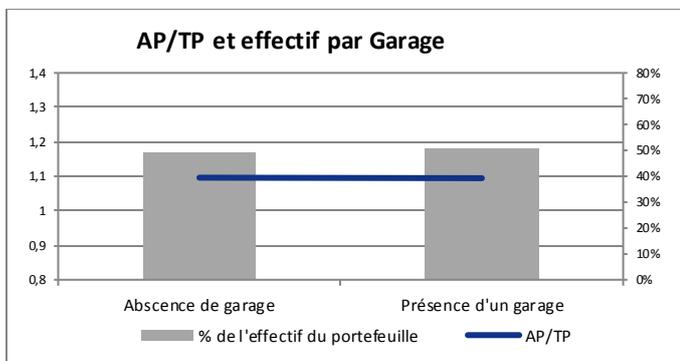


Figure: AP/TP et effectif par Garage

On constate que la rentabilité technique moyenne est équivalente que l'assuré présente un garage pour son véhicule ou non. La variable « garage », n'apportant pas d'information sur la valeur de l'AP/TP, ne sera pas retenue.

- Variables sur le contrat:

- Formule

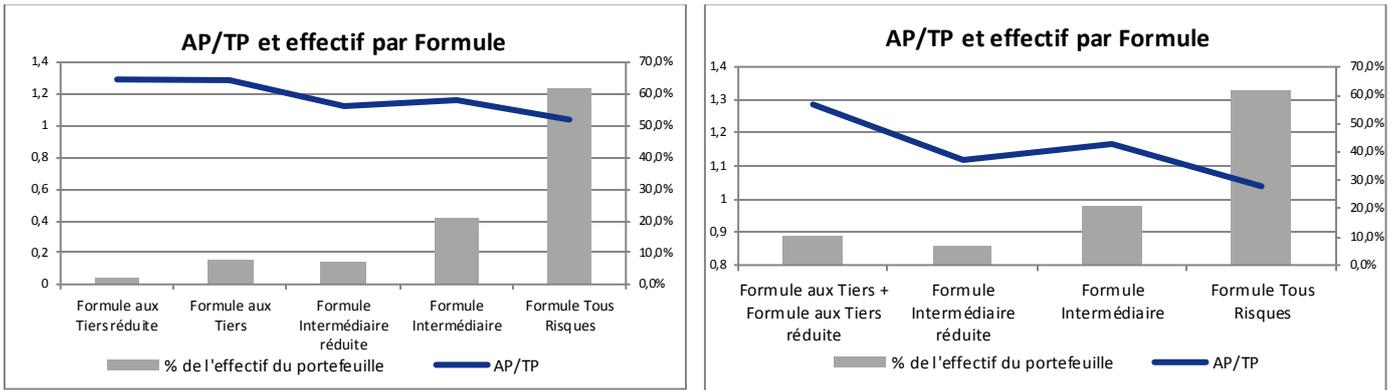


Figure: AP/TP et effectif par formule de souscription

La formule souscrite semble être discriminante sur l'AP/TP. La formule aux tiers réduite et la formule aux tiers, dont la différence réside en la réduction de la garantie assistance pour la formule aux tiers réduite, seront regroupées afin d'avoir un effectif d'au moins 4% du portefeuille pour chacune des modalités de la variable.

On constate après regroupement que la formule souscrite est assez discriminante sur l'AP/TP : en général la rentabilité est d'autant moins bonne que la formule souscrite est complète. Les formules aux tiers et formules aux tiers réduite apparaissent comme étant particulièrement rentable techniquement, contrairement à la formule tous risques qui affiche une rentabilité équilibrée.

- Ancienneté du contrat

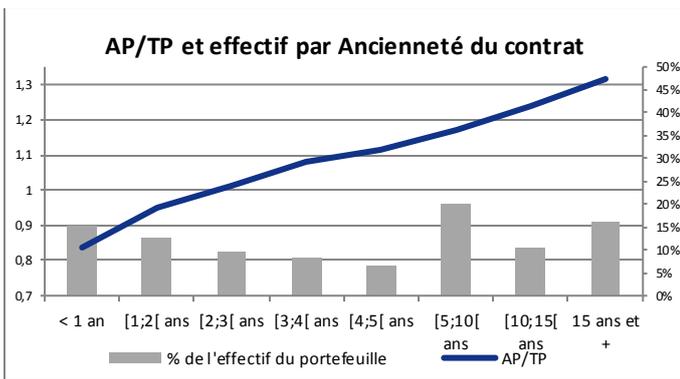


Figure : AP/TP et effectif par ancienneté du contrat

Comme relevé dans la partie 3 du chapitre 2 du mémoire, l'ancienneté du contrat est une variable particulièrement discriminante sur l'AP/TP : on note une nette croissance de l'AP/TP avec l'ancienneté du contrat. Différentes explications de cette discrimination ont été relevées dans le corps du mémoire.

- Fractionnement de la prime

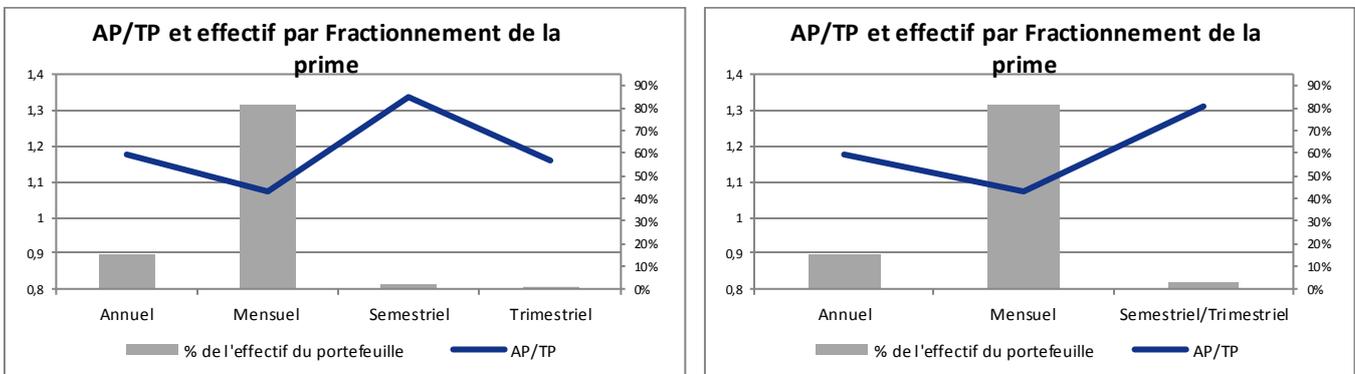


Figure: AP/TP et effectif par fractionnement de la prime

Après regroupement du fractionnement semestriel et trimestriel ayant tous les deux des AP/TP assez élevés comparé aux autres modalités de la variable « fractionnement », on constate une discrimination de la variable sur l'AP/TP : la rentabilité est meilleure pour les contrats dont le fractionnement de la prime se fait de manière annuelle, semestrielle ou trimestrielle. La

variable fractionnement sera ainsi conservée.

- Usage Kilometrique

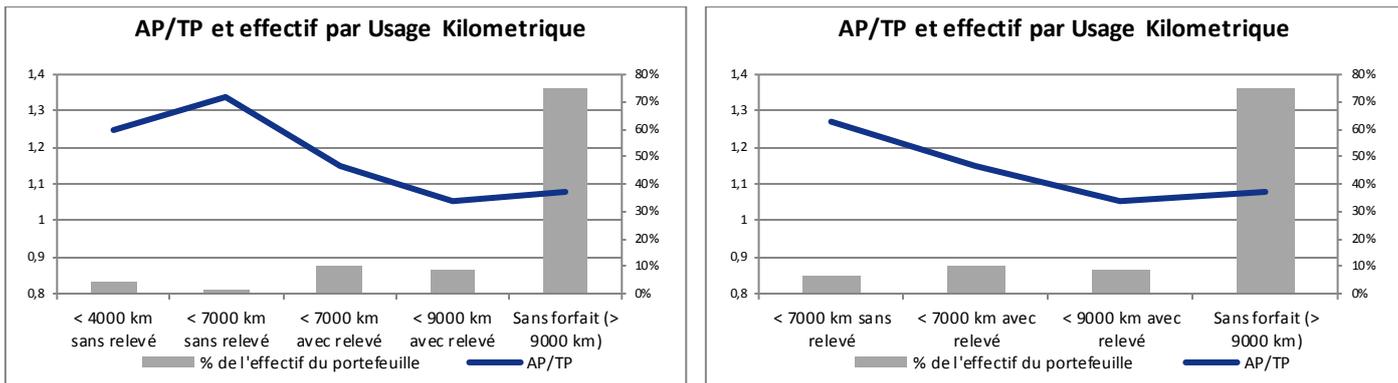


Figure: AP/TP et effectif par Usage Kilometrique

Les forfaits kilométriques sont proposés aux « petits rouleurs » et permettent ainsi de réduire la valeur de leur cotisation en se limitant à un certain nombre de kilomètres parcourables.

Les forfaits avec relevé demandent aux assurés ayant souscrit de se rendre auprès de la Société Française des Compteurs Automobiles (SOFCA) afin de réaliser un relevé de leur kilométrage parcouru aux dates qui leur sont indiquées.

Dans le cas d'un forfait kilométrique avec absence de relevé, tout dépassement de compteur doit être signalé car en cas d'accident le sinistre ne sera pas dédommagé si un dépassement est observé.

Les modalités « < 4000 km sans relevé » et « < 7000 km avec relevé » vont être regroupées car celles-ci sont proportionnellement peu représentées.

Après regroupement, on constate des écarts de rentabilité notables selon la modalité d'usage kilométrique souscrite. Les modalités « < 7000 km sans relevé » et « < 7000 km avec relevé » semblent ainsi être assez rentables techniquement. La variable est révélée discriminante sur l'AP/TP.

- Option Double Auto

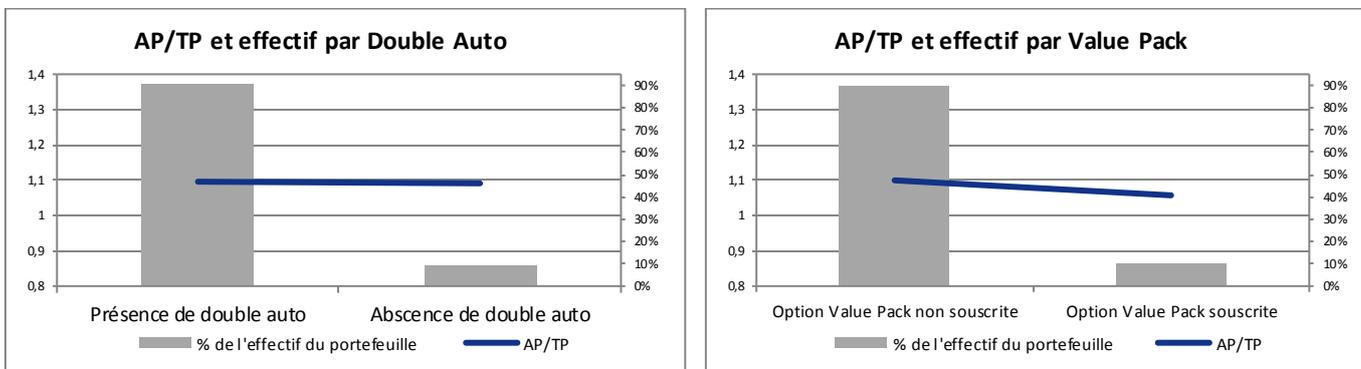


Figure: AP/TP et effectif par Double Auto

Figure: AP/TP et effectif par Value Pack

On constate que la souscription des options « Double Auto » et « Value Pack » n'influent pas ou que très peu sur la valeur de l'AP/TP. Les variables ne seront pas retenues par la suite.

- Option Pack Réparation

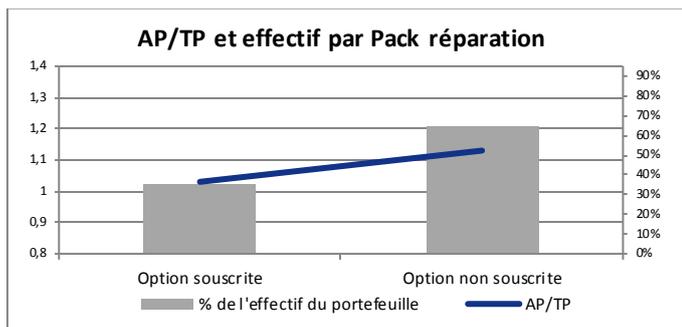


Figure : AP/TP et effectif par Pack de Réparation

Les polices ayant souscrit l'option « Pack Réparation » affichent une rentabilité moyenne inférieure. La variable semble être discriminante sur l'AP/TP.

- Franchise Bris De Glace

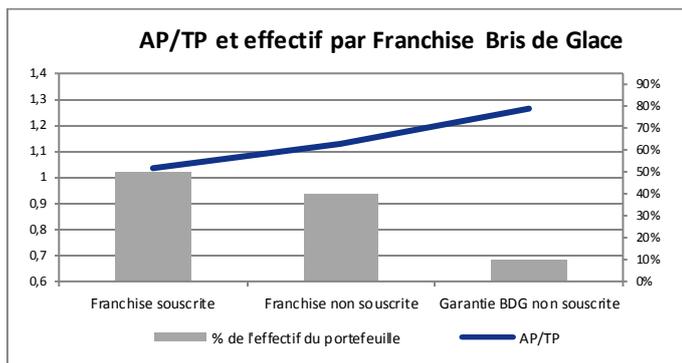


Figure: AP/TP et effectif par Franchise Bris de Glace

Les contrats ayant la modalité « Garantie BDG non souscrite » font référence aux formules aux Tiers (C1) et formules aux tiers réduite (CE) qui n'incluent pas la garantie Bris De Glace. Cette modalité n'est pas à prendre en compte lors du jugement de la discrimination de la franchise Bris De Glace. Cet effet est déjà pris en compte lors de l'étude univariée de la variable Formule. On constate finalement que les contrats ayant souscrit une franchise pour la garantie Bris De Glace sont en général moins rentables commercialement que les contrats n'en ayant pas souscrit. La variable Franchise Bris de Glace est ainsi discriminante sur l'AP/TP.

- Franchise Vol

Pour la variable « Franchise Vol » nous disposons de la valeur de la franchise pour chacun des contrats. A partir d'un tableau à double entrée disponible dans les conditions de souscription du produit automobile de particuliers, nous pouvons ainsi juger selon la classe de prix du véhicule et de la valeur de la franchise si celle-ci est supérieure, égale ou inférieure à la franchise standard pour la classe de véhicule considérée et donc s'il s'agit d'une « sur-franchise », d'une « franchise standard », d'une « sous-franchise » ou d'une « franchise nulle ». C'est cette nouvelle variable que l'on étudie.

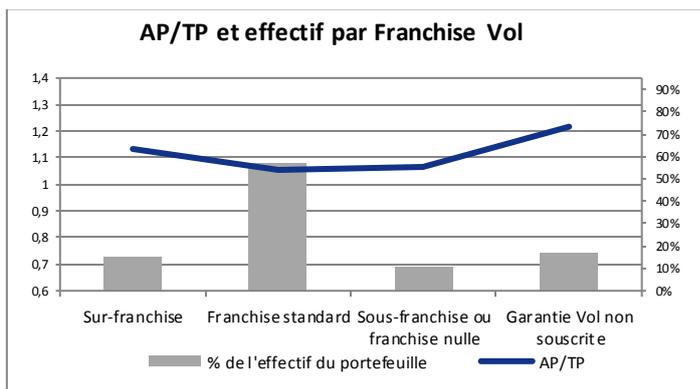
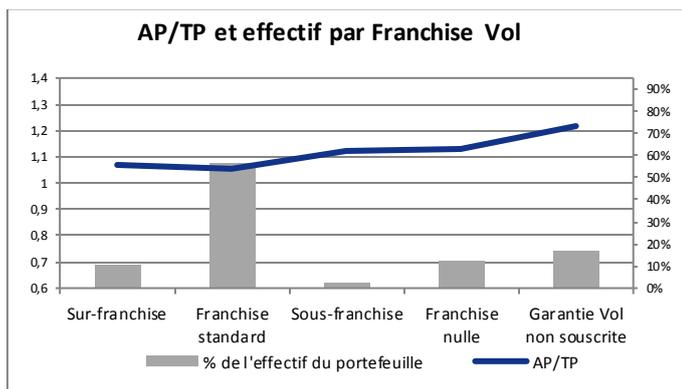


Figure: AP/TP et effectif par Franchise Vol

A nouveau, la modalité « Garantie Vol non souscrite » n'est pas à prendre en compte lors du jugement de la discrimination de la franchise vol sur l'AP/TP.

Les contrats ayant choisi une sous-franchise sont en sous-effectif et sont ainsi regroupés avec les contrats n'ayant pas souscrit de franchise par proximité de la valeur de l'AP/TP de ces modalités.

Après regroupement des classes, on constate que la modalité « sur-franchise » s'avère être légèrement plus rentable que les autres modalités. La variable « Franchise Vol » semble ainsi être discriminante sur l'AP/TP.

- Franchise Dommage

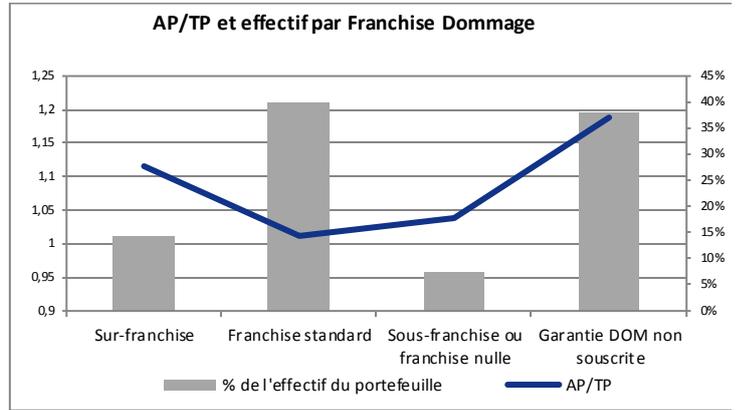
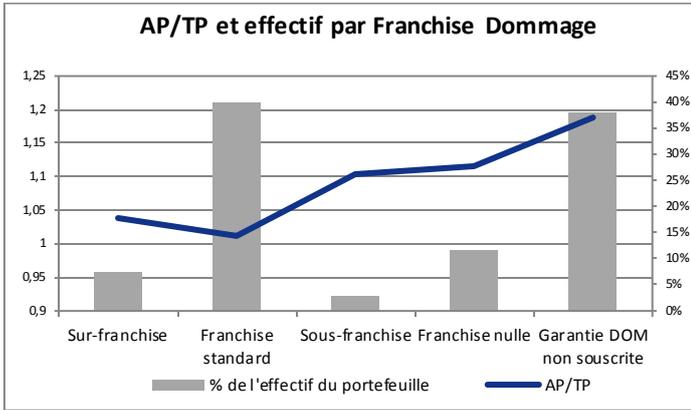
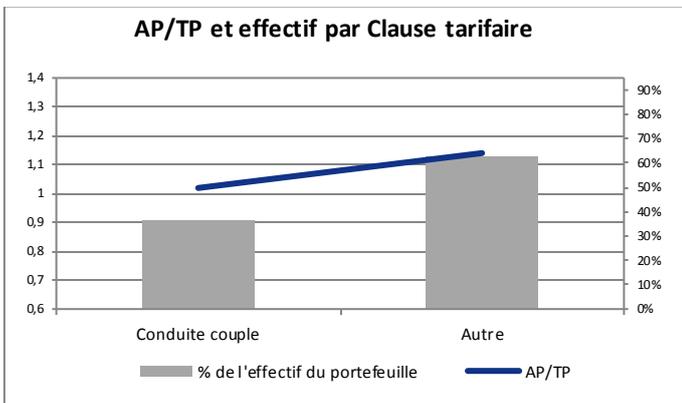


Figure: AP/TP et effectif par Franchise Dommage Tout Accident

De même, les contrats ayant souscrit une « sous-franchise » sont en sous-effectif et seront ainsi regroupés avec la modalité « franchise nulle ».

On constate des écarts d'AP/TP selon la modalité de Franchise Dommage sélectionnée. La modalité « sur-franchise » est notamment plus rentable que les autres modalités. La variable « Franchise Dommage » semble ainsi être discriminante sur l'AP/TP.

- Clause Tarifaire



Les polices ayant souscrit la clause « conduite couple » ont un AP/TP plus faible que les autres clauses tarifaires. La variable « Clause Tarifaire » est discriminante sur l'AP/TP et sera conservée.

Figure: AP/TP et effectif par Clause Tarifaire

- Variables sur les modulations tarifaires :

- Avantage client

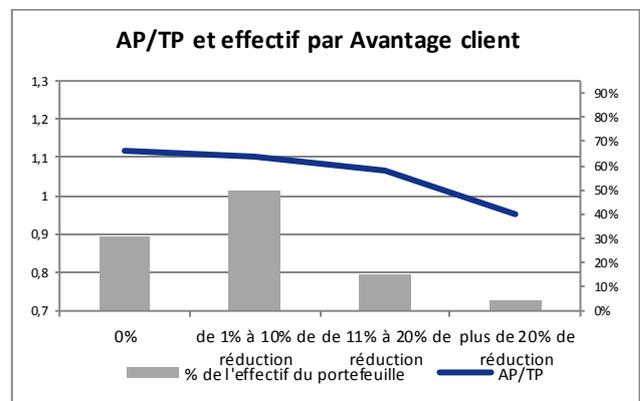
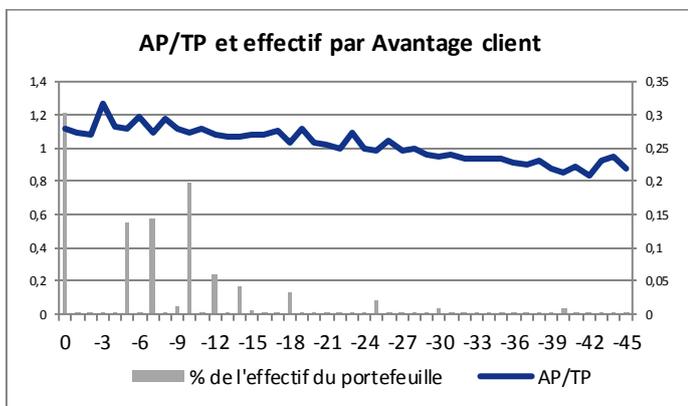


Figure: AP/TP et effectif par Avantage Client

On observe une tendance à la décroissance de l'AP/TP moyen des contrats en fonction de la réduction accordée par l'avantage client lorsque l'on observe la variable « avantage client » en continu. Le même résultat est observé après mise en classe qui permet de garantir une certaine robustesse des conclusions réalisées sur les observations. La variable « avantage client » est

ainsi discriminante sur l'AP/TP.

- *Avantage Prospect*

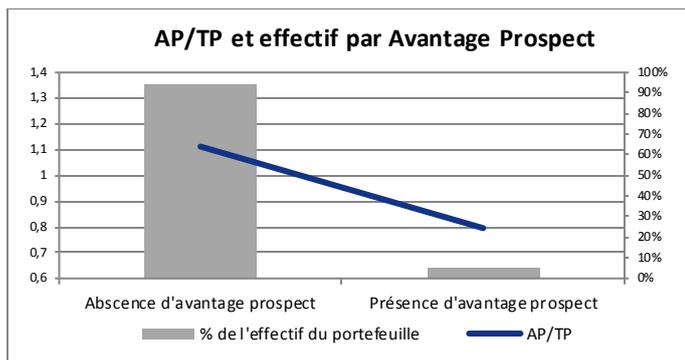


Figure: AP/TP et effectif par Avantage Prospect

On constate un fort écart de la valeur de l'AP/TP en fonction de la présence d'avantage prospect sur la police. En effet, les contrats bénéficiant de l'avantage prospect s'avèrent être beaucoup moins rentables techniquement que les polices n'en bénéficiant pas. L'Avantage Prospect semble être discriminant sur l'AP/TP.

- *Taux Protocole*

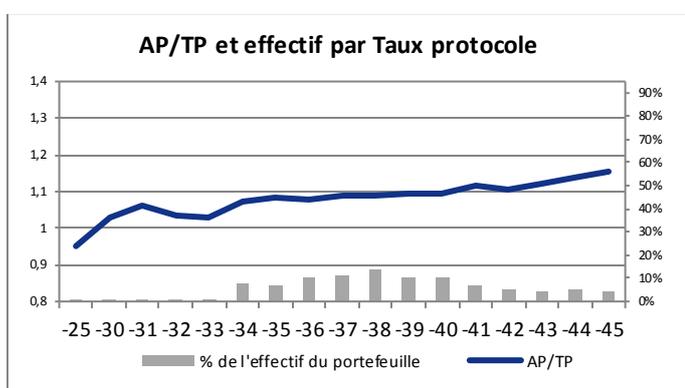
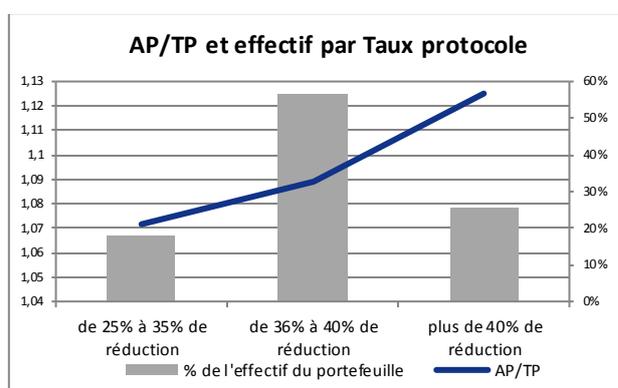


Figure: AP/TP et effectif par Taux Protocole



Lorsqu'il est observé de manière continue, le Taux Protocole semble être discriminant sur l'AP/TP. On constate que plus la réduction est élevée, plus l'AP/TP moyen est élevé.

Des résultats identiques sont observés après mise en classe. Le protocole est une variable discriminante sur l'AP/TP.

- *Taux budget*

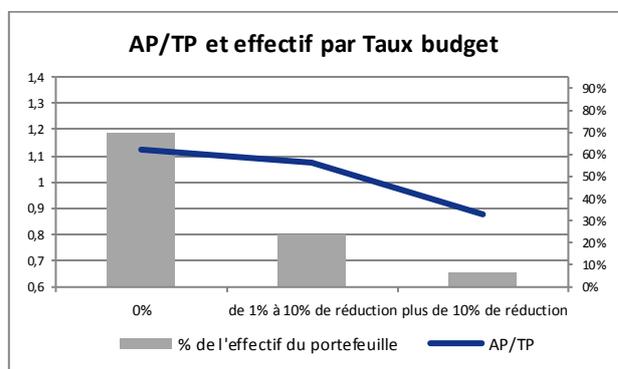
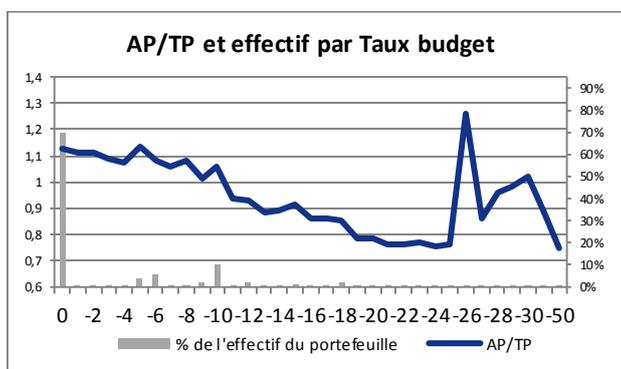


Figure: AP/TP et effectif par Taux budget

Lorsque l'on observe le Taux Budget de manière continue, on constate que, mis à part quelques exceptions, plus la réduction est élevée, plus l'AP/TP est faible.

Les mêmes résultats sont observés après mise en classe. Le réduction du taux budget est une variable discriminante sur l'AP/TP.

- Taux Firme

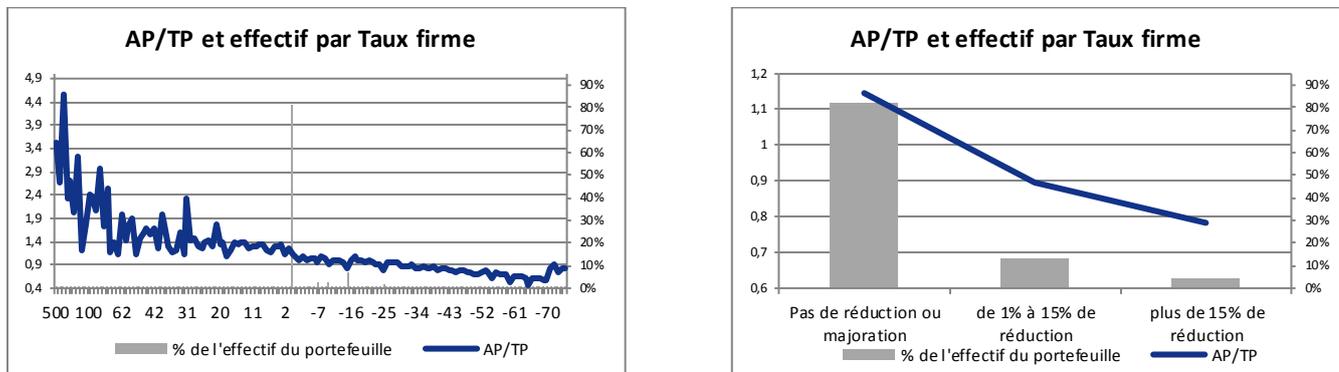


Figure: AP/TP et effectif par Taux Firme

Le Taux Firme fait référence au taux de réduction accordé par les codes firmes. Il peut s'agir d'une majoration comme d'une minoration. On constate qu'en continu, le taux firme semble être nettement discriminant sur l'AP/TP. Les contrats subissant des majorations dues aux codes firmes ne sont pas en effectif suffisant pour former une classe, ils seront alors regroupés avec les contrats qui ne présentent pas de réduction accordées par les codes firmes.

Après mise en classe, le constat reste le même: plus l'assuré bénéficie d'une réduction élevée, plus l'AP/TP est faible.



## Annexe 3: Coordonnées et contribution des modalités de l'ACM sur les 10 premiers axes

Row Coordinates				
	Dim1	Dim2	Dim3	Dim4
Age-A	-0.1302	0.2088	-0.2489	-0.1155
Age-B	-0.0322	0.1783	-0.2198	-0.0708
Age-C	0.0239	0.1385	-0.1675	-0.0442
Age-D	0.0415	0.0955	-0.1155	-0.0251
Age-E	0.0360	0.0620	-0.0769	-0.0115
Age-F	0.0258	0.0303	-0.0421	-0.0030
Age-G	0.0223	-0.0324	-0.0043	0.0010
Age-H	-0.0551	-0.2705	0.1985	-0.0703
Age-PM	0.3771	0.4150	0.4320	0.9851
CSP-Ar	0.1264	0.1278	0.0098	0.4072
CSP-Au	0.2147	0.1866	0.1521	0.2482
CSP-Fo	0.0521	-0.0251	-0.0839	-0.1122
CSP-SP	-0.0850	-0.2780	0.1905	-0.1062
CSP-Sa	-0.0619	0.0590	-0.1369	-0.0686
AgeP-18	0.0367	0.0234	-0.0842	-0.0186
AgeP-20	-0.0476	-0.0487	0.0109	-0.0738
AgeP-25	-0.1430	-0.1069	0.1313	-0.1052
AgeP-50	-0.2255	-0.1707	0.2194	-0.1139
AgeP-PM	0.3771	0.4150	0.4320	0.9851
AntSin-0	-0.0683	-0.0202	0.0166	0.0011
AntSin-1	0.1276	0.0377	-0.0310	-0.0021
AncV-A	0.4222	0.1032	0.1049	-0.0183
AncV-B	0.3996	0.0348	0.0731	-0.0126
AncV-C	0.3788	0.0205	0.0439	-0.0360
AncV-D	0.3629	0.0277	0.0087	-0.0356
AncV-E	0.3159	-0.0194	-0.0033	-0.0396
AncV-F	0.2461	-0.0673	0.0103	-0.0504
AncV-G	0.1978	-0.0807	0.0057	-0.0634
AncV-H	0.1786	-0.0249	-0.0501	-0.0233
AncV-I	0.1395	-0.0018	-0.0631	-0.0243
AncV-J	-0.2830	-0.0040	-0.0175	0.0316
Prix-A	-0.4215	-0.1596	0.2532	-0.1409
Prix-F	-0.2517	-0.0839	0.1276	-0.0752
Prix-H	-0.0416	-0.0424	0.0325	-0.0276
Prix-HC	0.2104	0.2795	0.3648	-0.2328
Prix-J	0.0301	0.0028	-0.0832	-0.0017
Prix-L	0.0963	0.0265	-0.1562	0.0183
Prix-M	0.1825	0.0420	-0.1263	0.0888
Prix-N	0.1992	0.0778	-0.1044	0.1518
Prix-O	0.1843	0.1129	-0.0616	0.1906
Prix-P	0.2346	0.1358	0.0112	0.0757
Prix-R	0.1849	0.2733	0.0778	-0.0984
Prix-S	0.3485	0.1616	0.1416	-0.1089
Prix-T	0.3418	0.1992	0.2395	-0.1205
GPSRA-A/	-0.4338	-0.1292	0.2356	-0.0651
GPSRA-C	-0.1447	-0.0516	0.0695	-0.0158
GPSRA-D	-0.0574	-0.0636	0.0812	-0.0460
GPSRA-E	0.0831	-0.0062	-0.1030	-0.0011
GPSRA-F	0.1348	0.0847	-0.0910	0.1089
GPSRA-G/	0.2432	0.1944	0.1409	-0.1286
Acq-A	0.0744	0.1033	-0.0583	-0.0195
Acq-B	-0.1585	0.0414	-0.0337	-0.0273
Acq-C	-0.1440	-0.2971	0.1715	0.0633
Secu-0	-0.1916	0.0060	0.0366	0.0752
Secu-N	0.2560	-0.0030	-0.0656	-0.1037
Secu-P	-0.0047	-0.0980	0.3258	0.0654
En-AU	0.0667	0.0369	-0.0779	0.0738
En-ES	-0.1497	-0.0829	0.1750	-0.1658
Car-4x4	0.1708	0.1279	0.0446	0.0310
Car-Au	0.2050	0.1177	0.0639	0.0123
Car-BER	-0.0930	-0.0553	0.0268	-0.1204
Car-BRK	0.1293	-0.0085	-0.1523	0.0037
Car-FGN	-0.0878	0.1093	0.0611	0.4360
Car-MSP	0.1192	-0.0640	-0.1682	-0.0088
KM-A	-0.3346	-0.1704	0.2290	0.0396
KM-B	-0.1609	-0.1130	0.1832	-0.1266
KM-C	-0.0027	0.0071	0.0251	-0.1464
KM-D	0.0581	0.0247	-0.0440	0.0324
KM-H	-0.4513	-0.0965	0.1699	0.0124
ZoneRC-A	-0.0073	-0.3081	-0.2479	0.2656
ZoneRC-B	-0.0122	-0.0818	-0.0871	0.0034
ZoneRC-C	0.0052	0.0181	-0.0147	-0.0473
ZoneRC-D	0.0358	0.1941	0.1910	-0.0651
ZoneRC-E	-0.0191	0.5132	0.6178	-0.1027
ZoneDOM-A	-0.0039	-0.2675	-0.2237	0.2131
ZoneDOM-B	0.0034	0.0141	-0.0053	-0.0371
ZoneDOM-C	-0.0276	0.4908	0.5979	-0.1073
ZoneVI-A	-0.0136	-0.2536	-0.2068	0.1973
ZoneVI-B	0.0013	0.0071	-0.0138	-0.0382
ZoneVI-C	0.0166	0.4331	0.5012	-0.0907
ZoneBDG-1	-0.0400	0.1959	0.3071	-0.0911
ZoneBDG-2	-0.0095	0.0105	0.0469	0.0180
ZoneBDG-3	0.0127	-0.0108	-0.0060	0.0316

Partial Contributions to Inertia for the Row Points				
	Dim1	Dim2	Dim3	Dim4
Age-A	0,0021	0,0093	0,0169	0,0043
Age-B	0,0001	0,0069	0,0135	0,0017
Age-C	0	0,0029	0,0054	0,0004
Age-D	0,0002	0,0023	0,0043	0,0002
Age-E	0,0001	0,0007	0,0015	0
Age-F	0,0001	0,0003	0,0007	0
Age-G	0,0001	0,0002	0	0
Age-H	0,0012	0,051	0,035	0,0052
Age-PM	0,0086	0,0184	0,0254	0,1578
CSP-Ar	0,0012	0,0022	0	0,0338
CSP-Au	0,0102	0,0136	0,0116	0,0367
CSP-Fo	0,0003	0,0001	0,0018	0,0038
CSP-SP	0,0022	0,0411	0,0247	0,0091
CSP-Sa	0,0023	0,0037	0,0255	0,0076
AgeP-18	0,0009	0,0007	0,0112	0,0007
AgeP-20	0,0009	0,0016	0,0001	0,0056
AgeP-25	0,0018	0,0018	0,0034	0,0026
AgeP-50	0,004	0,0041	0,0086	0,0028
AgeP-PM	0,0086	0,0184	0,0254	0,1578
AntSin-0	0,004	0,0006	0,0005	0
AntSin-1	0,0075	0,0012	0,001	0
AncV-A	0,02	0,0021	0,0028	0,0001
AncV-B	0,0091	0,0001	0,0007	0
AncV-C	0,0081	0	0,0002	0,0002
AncV-D	0,0081	0,0001	0	0,0002
AncV-E	0,0071	0	0	0,0003
AncV-F	0,0043	0,0006	0	0,0005
AncV-G	0,0027	0,0008	0	0,0008
AncV-H	0,0022	0,0001	0,0004	0,0001
AncV-I	0,0014	0	0,0007	0,0001
AncV-J	0,0539	0	0,0005	0,0018
Prix-A	0,0225	0,0057	0,0183	0,0068
Prix-F	0,0102	0,002	0,0059	0,0024
Prix-H	0,0004	0,0008	0,0006	0,0005
Prix-HC	0,0004	0,0011	0,0025	0,0012
Prix-J	0,0002	0	0,004	0
Prix-L	0,0012	0,0002	0,0073	0,0001
Prix-M	0,0044	0,0004	0,0047	0,0028
Prix-N	0,0037	0,001	0,0023	0,0058
Prix-O	0,002	0,0013	0,0005	0,0058
Prix-P	0,0033	0,002	0	0,0009
Prix-R	0,0004	0,0016	0,0002	0,0003
Prix-S	0,0016	0,0006	0,0006	0,0004
Prix-T	0,0024	0,0014	0,0027	0,0008
GPSRA-A/	0,0257	0,004	0,0171	0,0016
GPSRA-C	0,0039	0,0009	0,002	0,0001
GPSRA-D	0,0006	0,0013	0,0026	0,001
GPSRA-E	0,003	0	0,0105	0
GPSRA-F	0,0053	0,0037	0,0054	0,0092
GPSRA-G/	0,0051	0,0057	0,0038	0,0038
Acq-A	0,0048	0,0165	0,0067	0,0009
Acq-B	0,003	0,0004	0,0003	0,0002
Acq-C	0,0066	0,05	0,0213	0,0035
Secu-0	0,027	0	0,0022	0,0112
Secu-N	0,0361	0	0,0053	0,0159
Secu-P	0	0,0005	0,0068	0,0003
En-AU	0,004	0,0022	0,0124	0,0133
En-ES	0,0091	0,0049	0,0279	0,0299
Car-4x4	0,003	0,003	0,0005	0,0003
Car-Au	0,0056	0,0032	0,0012	0,0001
Car-BER	0,0057	0,0035	0,0011	0,0255
Car-BRK	0,0014	0	0,0043	0
Car-FGN	0,0013	0,0036	0,0014	0,0879
Car-MSP	0,0024	0,0012	0,0109	0
KM-A	0,0025	0,0011	0,0026	0,0001
KM-B	0,0035	0,0031	0,0103	0,0059
KM-C	0	0	0,0002	0,0064
KM-D	0,0033	0,0011	0,0043	0,0028
KM-H	0,0124	0,001	0,004	0
ZoneRC-A	0	0,0245	0,0203	0,0277
ZoneRC-B	0,0001	0,0056	0,0081	0
ZoneRC-C	0	0,0003	0,0002	0,0026
ZoneRC-D	0,0002	0,0115	0,0142	0,002
ZoneRC-E	0	0,0384	0,0711	0,0023
ZoneDOM-A	0	0,0278	0,0248	0,0269
ZoneDOM-B	0	0,0004	0,0001	0,0037
ZoneDOM-C	0,0001	0,0388	0,0736	0,0028
ZoneVI-A	0	0,0274	0,0232	0,0252
ZoneVI-B	0	0,0001	0,0004	0,0037
ZoneVI-C	0	0,0416	0,0711	0,0028
ZoneBDG-1	0,0001	0,0057	0,0178	0,0019
ZoneBDG-2	0	0,0001	0,0015	0,0003
ZoneBDG-3	0,0001	0,0001	0	0,0014

ZoneBDG-4/5	-0.0016	-0.0344	-0.0910	-0.0365
For-C1	-0.6527	0.1345	0.0046	0.1519
For-C2	-0.3706	0.0782	-0.1077	0.0308
For-C3	0.2783	-0.0739	0.0455	-0.0375
For-CB	-0.3981	0.2205	-0.0839	0.0193
ACont-A	0.0281	0.3050	-0.1069	-0.1181
ACont-B	0.0273	0.1864	-0.0795	-0.0215
ACont-C	0.0323	0.1475	-0.0707	0.0050
ACont-D	-0.0060	0.0902	-0.0610	0.0206
ACont-E	0.0094	0.0720	-0.0604	0.0305
ACont-F	-0.0086	-0.0798	0.0027	0.0605
ACont-G	-0.0202	-0.1866	0.0904	0.0555
ACont-H	-0.0441	-0.3759	0.1980	-0.0085
Frac-A	-0.0508	-0.1822	0.2203	0.1248
Frac-M	0.0163	0.0406	-0.0502	-0.0260
Frac-ST	-0.1870	-0.1650	0.2319	0.0611
FBDG-0	0.1286	-0.2330	0.0565	0.0280
FBDG-1	0.0238	0.1557	-0.0428	-0.0516
FBDG-X	-0.6532	0.1519	-0.0109	0.1521
FVol-0	0.2715	-0.0557	-0.0105	-0.0298
FVol-I	0.2845	-0.3416	0.0109	0.0265
FVol-S	0.3043	0.3744	0.4372	-0.2132
FVol-X	-0.4502	0.1195	-0.0736	0.0607
FDom-0	0.2704	-0.0612	-0.0091	-0.0290
FDom-I	0.2908	-0.3411	0.0025	0.0285
FDom-S	0.2994	0.3590	0.4371	-0.2107
FDom-X	-0.4502	0.1195	-0.0736	0.0607
CITari-F	-0.0045	-0.0517	0.0658	0.0780
CITari-U	0.0078	0.0889	-0.1132	-0.1341
AvC-A	-0.0905	0.0744	-0.0053	-0.0586
AvC-B	-0.0300	-0.0593	-0.0294	-0.0466
AvC-C	0.1602	0.0019	0.0235	0.1027
AvC-D	0.3946	0.1485	0.2792	0.5564
AvProsp-0	-0.0021	-0.0223	0.0094	0.0148
AvProsp-1	0.0370	0.3970	-0.1667	-0.2628
Txbudg-A	-0.0048	0.0026	0.0334	-0.0052
Txbudg-B	0.0008	-0.0464	-0.0575	0.0400
Txbudg-C	0.0480	0.1401	-0.1466	-0.0893
TxFirme-A	-0.0406	-0.0486	0.0158	0.0264
TxFirme-B	0.1629	0.2230	-0.0987	-0.1116
TxFirme-C	0.2648	0.2261	0.0117	-0.1545

ZoneBDG-4/5	0	0,0008	0,0075	0,0014
For-C1	0,0565	0,0042	0	0,0082
For-C2	0,038	0,003	0,0072	0,0007
For-C3	0,0629	0,0078	0,0038	0,0031
For-CB	0,0148	0,008	0,0015	0,0001
ACont-A	0,0002	0,0332	0,0052	0,0076
ACont-B	0,0001	0,0104	0,0024	0,0002
ACont-C	0,0001	0,0049	0,0014	0
ACont-D	0	0,0016	0,0009	0,0001
ACont-E	0	0,0008	0,0007	0,0002
ACont-F	0	0,003	0	0,0026
ACont-G	0,0001	0,0087	0,0026	0,0012
ACont-H	0,0004	0,0536	0,019	0
Frac-A	0,0005	0,012	0,0224	0,0086
Frac-M	0,0003	0,0031	0,0061	0,0019
Frac-ST	0,0013	0,0018	0,0046	0,0004
FBDG-0	0,0087	0,0503	0,0038	0,0011
FBDG-1	0,0004	0,0283	0,0027	0,0047
FBDG-X	0,0544	0,0052	0	0,0079
FVol-0	0,039	0,0029	0,0001	0,0013
FVol-I	0,0154	0,0393	0,0001	0,0004
FVol-S	0,0086	0,0229	0,0399	0,0113
FVol-X	0,1018	0,0127	0,0061	0,005
FDom-0	0,0394	0,0036	0,0001	0,0012
FDom-I	0,0151	0,0368	0	0,0004
FDom-S	0,0085	0,0216	0,0408	0,0113
FDom-X	0,1018	0,0127	0,0061	0,005
CITari-F	0	0,0039	0,0081	0,0136
CITari-U	0	0,0067	0,014	0,0234
AvC-A	0,0033	0,0039	0	0,0037
AvC-B	0,0006	0,0041	0,0013	0,0038
AvC-C	0,0051	0	0,0002	0,0057
AvC-D	0,0093	0,0023	0,0105	0,0498
AvProsp-0	0	0,0011	0,0002	0,0007
AvProsp-1	0,0001	0,0194	0,0044	0,013
Txbudg-A	0	0	0,0023	0,0001
Txbudg-B	0	0,0012	0,0023	0,0013
Txbudg-C	0,0002	0,003	0,0042	0,0018
TxFirme-A	0,0018	0,0045	0,0006	0,002
TxFirme-B	0,0048	0,0157	0,0039	0,006
TxFirme-C	0,0039	0,005	0	0,0035
TxFirme-A	0,0018	0,0043	0,0003	0,0022
TxFirme-B	0,0049	0,0155	0,0021	0,0069
TxFirme-C	0,0039	0,0046	0,0002	0,0036

## Annexe 4: Tests d'adéquation de la loi empirique des AP/TP

### Tests d'adéquation de la loi des AP/TP à la loi normale

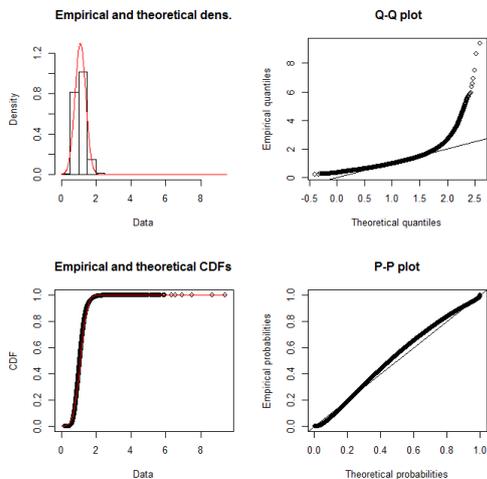


Figure : Résultats des tests d'adéquation des AP/TP avec la loi normale

### Tests d'adéquation de la loi des AP/TP à la loi Log-normale

Une variable aléatoire  $X$  suit une loi log-normale  $\text{LogN}(\mu, \sigma^2)$  si et seulement si la variable aléatoire  $\ln X$  suit la loi normale  $N(\mu, \sigma^2)$ .

On a ainsi testé l'adéquation de la loi suivie par le log des AP/TP de l'échantillon avec la loi Normale.

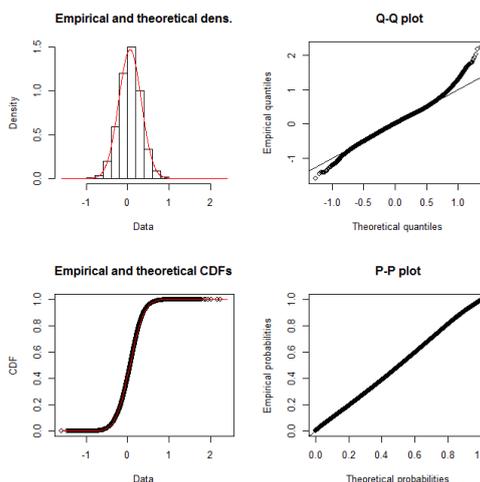


Figure : Résultats des tests d'adéquation du log des AP/TP avec la loi normale

### Tests d'adéquation de la loi des AP/TP à la loi Gamma

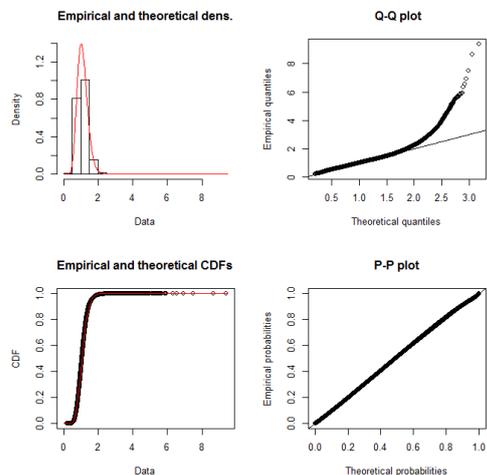


Figure : Résultats des tests d'adéquation des AP/TP avec la loi Gamma

### Tests d'adéquation de la loi des AP/TP à la loi de Weibull

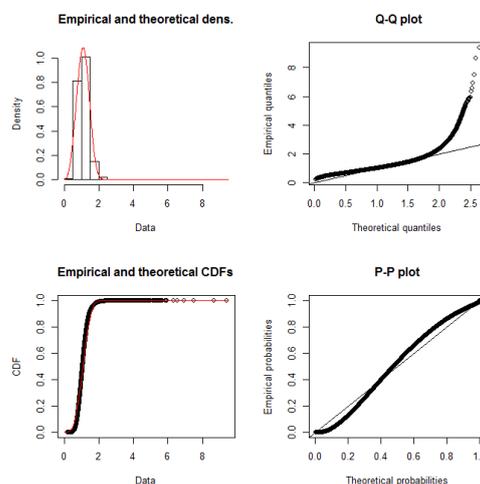


Figure : Résultats des tests d'adéquation des AP/TP avec la loi de Weibull

L'adéquation avec la loi log-normale semble être plutôt satisfaisante mais cette distribution n'est pas adéquate au niveau des deux extrêmes.

Parmi les lois testées, c'est la gamma qui est la loi qui paraît alors la plus adaptée pour qualifier la loi suivie par notre échantillon d'AP/TP. On note cependant que la loi théorique retenue sous-estime les valeurs d'AP/TP supérieures à 2.

Les coefficients de la loi gamma sont donnés par :

```
> summary(fit.gamma)
Fitting of the distribution ' gamma ' by maximum likelihood
Parameters :
      estimate Std. Error
shape 13.67983 0.02050218
rate  12.48983 0.01906589
Loglikelihood: -154035.4  AIC: 308074.8  BIC: 308098.2
Correlation matrix:
      shape      rate
shape 1.0000000 0.9817899
rate  0.9817899 1.0000000
```

## Annexe 5: Résultats de la modélisation des AP/TP en loi unique

Parameter Number	Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
1	Mean	0,2753	0,00065	0,2		1 390 640	100,0	1,3170
2	AgeC (Age_A)	0,0275	0,00061	2,2		127 247	9,2	1,0279
3	AgeC (Age_B)	-0,0525	0,00059	1,1		130 354	9,4	0,9488
4	AgeC (Age_C)	-0,0532	0,00067	1,3		90 565	6,5	0,9482
5	AgeC (Age_D)	-0,0650	0,00054	0,8		153 186	11,0	0,9370
6	AgeC (Age_E)	-0,0834	0,00060	0,7		115 976	8,3	0,9200
7	AgeC (Age_F)	-0,0795	0,00051	0,6		182 414	13,1	0,9236
8	AgeC (Age_G)	-0,0533	0,00061	1,1		109 349	7,9	0,9481
-	AgeC (Age_H)					417 750	30,0	
9	AgeC (Age_PM)	0,0044	0,00083	19,0		63 799	4,6	1,0044
10	AncVehi (AncV_A)	-0,2192	0,00063	0,3		118 990	8,6	0,8032
11	AncVehi (AncV_B)	-0,2176	0,00080	0,4		60 361	4,3	0,8044
12	AncVehi (AncV_C)	-0,1935	0,00080	0,4		59 396	4,3	0,8241
13	AncVehi (AncV_D)	-0,1755	0,00077	0,4		65 127	4,7	0,8390
14	AncVehi (AncV_E)	-0,1607	0,00072	0,4		74 846	5,4	0,8515
15	AncVehi (AncV_F)	-0,1507	0,00071	0,5		75 718	5,4	0,8601
16	AncVehi (AncV_G)	-0,1418	0,00072	0,5		74 186	5,3	0,8678
17	AncVehi (AncV_H)	-0,1248	0,00071	0,6		74 319	5,3	0,8827
18	AncVehi (AncV_I)	-0,1035	0,00069	0,7		77 279	5,6	0,9016
-	AncVehi (AncV_J)					710 418	51,1	
19	Clprix (Prix_A)	-0,0203	0,00062	3,0		133 132	9,6	0,9799
20	Clprix (Prix_F)	-0,0309	0,00056	1,8		168 347	12,1	0,9695
21	Clprix (Prix_H)	-0,0291	0,00050	1,7		252 556	18,2	0,9714
22	Clprix (Prix_HC)	-0,0573	0,00158	2,8		14 828	1,1	0,9443
-	Clprix (Prix_J)					267 605	19,2	
23	Clprix (Prix_L)	0,0063	0,00059	9,3		139 931	10,1	1,0064
24	Clprix (Prix_M)	0,0205	0,00060	2,9		137 173	9,9	1,0207
25	Clprix (Prix_N)	0,0440	0,00067	1,5		97 754	7,0	1,0450
26	Clprix (Prix_O)	0,0622	0,00080	1,3		62 934	4,5	1,0641
27	Clprix (Prix_P)	0,0745	0,00081	1,1		63 111	4,5	1,0773
28	Clprix (Prix_R)	0,1128	0,00143	1,3		17 300	1,2	1,1194
29	Clprix (Prix_S)	0,1342	0,00158	1,2		14 266	1,0	1,1436
30	Clprix (Prix_T)	0,1170	0,00132	1,1		21 703	1,6	1,1241
31	Carros (Car_4x4)	0,0786	0,00065	0,8		113 556	8,2	1,0817
32	Carros (Car_Au)	0,0810	0,00055	0,7		143 511	10,3	1,0843
-	Carros (Car_BER)					687 081	49,4	
33	Carros (Car_BRK)	-0,0079	0,00066	8,4		86 168	6,2	0,9922
34	Carros (Car_FGN)	0,0208	0,00050	2,4		180 699	13,0	1,0210
35	Carros (Car_MSP)	-0,0241	0,00050	2,1		179 625	12,9	0,9762
36	Formule (For_C1)	0,1046	0,00059	0,6		139 223	10,0	1,1103
37	Formule (For_C2)	0,0332	0,00045	1,4		289 732	20,8	1,0337
-	Formule (For_C3)					863 869	62,1	
38	Formule (For_CB)	-0,0029	0,00066	22,8		97 816	7,0	0,9971
39	AgeCont (ACont_A)	-0,2731	0,00058	0,2		214 493	15,4	0,7610
40	AgeCont (ACont_B)	-0,1686	0,00058	0,3		178 847	12,9	0,8449
41	AgeCont (ACont_C)	-0,1093	0,00062	0,6		135 923	9,8	0,8965
42	AgeCont (ACont_D)	-0,0750	0,00063	0,8		114 840	8,3	0,9277
43	AgeCont (ACont_E)	-0,0424	0,00068	1,6		93 622	6,7	0,9585
-	AgeCont (ACont_F)					277 850	20,0	
44	AgeCont (ACont_G)	0,0335	0,00058	1,7		148 951	10,7	1,0340
45	AgeCont (ACont_H)	0,0950	0,00053	0,6		226 114	16,3	1,0996
46	AvClient (AvC_A)	0,0161	0,00035	2,2		422 531	30,4	1,0163
-	AvClient (AvC_B)					690 868	49,7	
47	AvClient (AvC_C)	-0,0264	0,00045	1,7		212 793	15,3	0,9740
48	AvClient (AvC_D)	-0,1309	0,00079	0,6		64 448	4,6	0,8773
-	TauxBudget (Txbudg_A)					970 729	69,8	
49	TauxBudget (Txbudg_B)	-0,0775	0,00039	0,5		328 964	23,7	0,9254
50	TauxBudget (Txbudg_C)	-0,1532	0,00069	0,4		90 947	6,5	0,8579
-	TauxFirme (TxFirme_A)					1 138 173	81,8	
51	TauxFirme (TxFirme_B)	-0,1142	0,00054	0,5		189 490	13,6	0,8921
52	TauxFirme (TxFirme_C)	-0,2687	0,00081	0,3		62 977	4,5	0,7643

## Annexe 6: Résultat de la régression logistique pour la détection des contrats associés à un AP/TP extrême

```
> summary(model)
```

```
Call:
glm(formula = formula, family = binomial(link = lien), data = data)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-2.3974	-0.4039	-0.2391	-0.1266	3.7863

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-7.367482	0.023773	-309.91	<2e-16	***
AncacqAcq_C	0.751573	0.007567	99.33	<2e-16	***
AncVehiAncV_J	1.152194	0.009047	127.36	<2e-16	***
AgeContACont_H	0.833739	0.007975	104.54	<2e-16	***
FormuleFor_C1	1.323313	0.030852	42.89	<2e-16	***
GPSRAGPSRA_G_H	1.953319	0.012705	153.74	<2e-16	***
FraBDGFBDG_X	-0.805293	0.031545	-25.53	<2e-16	***
ClauTariClTari_F	0.674490	0.008241	81.85	<2e-16	***
TauxBudgetTxbudg_A	0.981157	0.009075	108.11	<2e-16	***
TauxFirmeTxFirme_A	2.017492	0.018855	107.00	<2e-16	***
GPSRAGPSRA_F	1.015855	0.009234	110.01	<2e-16	***
GPSRAGPSRA_E	0.776440	0.008480	91.56	<2e-16	***
AvClientAVC_A	0.495747	0.006975	71.07	<2e-16	***
CarrosCar_MSP	-0.832479	0.013175	-63.19	<2e-16	***
FraDomFDom_STD	-0.622882	0.008673	-71.82	<2e-16	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 815968 on 1390806 degrees of freedom
Residual deviance: 627116 on 1390792 degrees of freedom
AIC: 627146
```

```
Number of Fisher Scoring iterations: 7
```

# Annexe 7: Résultat graphique de l'arbre de classification pour la détection des contrats associés à un AP/TP extrême

