Mémoire présenté devant le
Conservatoire National des Arts et Métiers

Pour l'obtention du Master 2 spécialité Actuariat et l'admission à l'Institut des Actuaires

Le 07/07/2021

Par : **Jonathan KHAMPHANH**

Titre : **Application des méthodes de Machine Learning dans le pilotage de portefeuille : Cas en assurance transport**

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

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

Membres présents du jury du Conservatoire National des Arts et Métiers

Sandrine LEMERY

Olivier DESMETTRE

David FAURE

Nathanaël ABECERA

Membres présents du jury de l'Institut des Actuaires

Emmanuel SOTTO

Jérémie DEVUN

Secrétariat

Bibliothèque

Entreprise : **CHUBB**

Directeur de mémoire en entreprise :

Nom : **Léa NEFUSSI**

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

Signature du responsable entreprise

Signature du candidat

APPLICATION DES METHODES DE MACHINE LEARNING DANS LE PILOTAGE DE PORTEFEUILLE : CAS EN ASSURANCE TRANSPORT

Par: Jonathan Khamphanh
Superviseur: Léa Nefussi – IA CERA

CHUBB
Actuariat – Europe Continentale
Tour Carpe Diem
31 Place des Corolles
92400 Courbevoie

SOMMAIRE

REMERCIEMENTS

Avant tout développement, je tiens à remercier tout particulièrement mes anciens collègues d'Allianz Partners : Sisi Yé, Jean-Marc Quéau et Anh Tran, sans qui cette aventure dans l'actuariat n'aurait pas été possible.

Je voudrais également remercier l'ensemble des personnes avec qui j'ai pu travailler au sein d'Allianz au département *Risk Management* et Actuariat pour leur soutien et leurs encouragements dans ce long périple ; le corps professoral du Conservatoire National des Arts et Métiers pour la qualité des cours ainsi que mes camarades de promotion pour les échanges que nous avons pu avoir.

Enfin, je souhaite remercier mes collègues de Chubb et notamment Léa Nefussi, Anne-Laure Boisseau et Guillaume Bertheau pour les conseils, la relecture et le suivi de mes travaux.

Tous ont contribué à la construction de l'actuaire que j'aspire à devenir.

RESUME

Mots-clefs : *Arbre de décision, Assurance Non-Vie, Assurance Transport, Classification Ascendante Hiérarchique, Forêt Aléatoire, Intelligence Artificielle, K-Moyennes, Méthodes d'apprentissage, Modèle Linéaire Généralisé, Pilotage de portefeuille, Rentabilité, Risque d'entreprise, Séparateurs à Vaste Marge, Shortfall, Solvabilité 2 et Tarification.*

A la sortie de la crise des *subprimes* en 2008, le secteur de l'assurance transport fait face à un contexte difficile. La mondialisation et le développement des porte-conteneurs ont créé une accumulation des valeurs assurées. Combinés à l'apparition de conséquences majeures provenant du changement climatique, les sinistres larges et les catastrophes naturelles deviennent plus fréquents. La baisse des primes d'assurances et la diminution des résultats financiers impliquent une recherche de l'optimisation des résultats pour les assureurs sous peine de faire faillite comme un certain nombre d'acteurs du *Lloyd's of London.*

Afin d'utiliser les contraintes règlementaires comme levier pour répondre aux objectifs de rentabilité de l'actionnariat, il est important de mettre en place un pilotage rigoureux du portefeuille. Avec l'essor du *Big Data,* l'utilisation des algorithmes d'apprentissage a été facilitée et popularisée dans la littérature actuarielle ces dernières années. Ils restent en pratique peu utilisés en tarification dans les compagnies d'assurance traditionnelles. Les travaux de ce mémoire cherchent à proposer une utilisation de ces méthodes pour le pilotage de la rentabilité d'un portefeuille d'assurance transport.

La première phase de l'étude consiste à la visualisation des données avec l'emploi d'analyses factorielles et d'arbres de décisions. Dans notre cas précis, le portefeuille étudié présente un nombre faible de comptes non-profitables, il a donc fallu utiliser une méthode de suréchantillonnage synthétique dite *SMOTE* afin d'améliorer les résultats des arbres de décisions. L'extension aux forêts aléatoires a permis de réduire les dimensions et de choisir les facteurs de risques importants pour la suite des algorithmes et ultérieurement dans le recalibrage du Modèle Linéaire Généralisé dans un but de re-tarification.

La deuxième phase de l'étude consiste à segmenter le portefeuille afin d'identifier les zones de sous tarification. L'utilisation des Séparateurs à Vaste Marge, des k-Moyennes et de la Classification Ascendante Hiérarchique permet de mettre en lumière des profils de comptes et leur performance.

Les conclusions de la deuxième phase permettent de prendre des décisions stratégiques sur le portefeuille. Outre la simple résiliation ou la mise en place de franchises et de clauses d'exclusions, nous avons proposé dans une troisième partie, l'étude de l'impact d'un recalibrage du modèle. La validation d'une version simplifiée du modèle de tarification est testée en même temps qu'une correction partielle tenant compte d'un rehaussement des facteurs de risques des profils sous tarifés.

ABSTRACT

Keywords: *Artificial Intelligence, Cargo Insurance, Classification Tree, Commercial Risk, General Insurance, Generalised Linear Models, Hierarchical Clustering, K-Means, Machine Learning, Portfolio Management, Pricing, Profitability, Random Forest, Solvency II, Shortfall and Support Vector Machines.*

After the subprime crisis in 2008, the cargo insurance sector was facing a challenging situation. Globalisation and development of massive containerships have created an accumulation of insured values. Recently, material impacts from climate change entered into the equation, leading to a reduction of man-made and natural catastrophes' return period. The fall of insurance premium combined with the shrink of investment return have forced insurers to search for profit optimisation or otherwise would have to shut up shop like several syndicates in the Lloyd's of London.

In order to leverage regulatory constraints to meet the hurdle rate defined by shareholders, it is of the utmost importance to put in place a strict portfolio management. With Big Data being more common now, the use of Machine Learning algorithms has been facilitated and became all the rage in the actuarial science literature. They are nonetheless, rarely used for pricing in traditional insurance companies. This paper's objective is to propose the use of those algorithms for portfolio management in cargo insurance.

The first stage of the study consists in the data visualisation using factor analysis and decision trees. In our case, the portfolio under analysis only has a few non-performing accounts. To overcome the unbalanced dataset, the results of the decision trees were improved using a synthetic minority over sampling technique called SMOTE. The extension to Random Forest allowed for dimension reduction for the following algorithms and to select important risk factors used at a later stage in our Generalised Linear Model for pricing purposes.

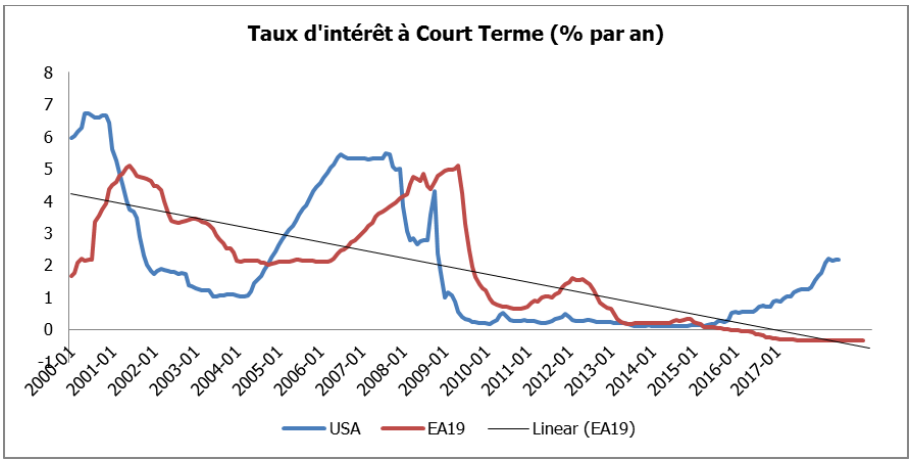
The second stage of the study is to divide the portfolio in order to identify the under-priced areas. Algorithms such as Support Vector Machines, k-Means, and Hierarchical Clustering have been used to detect the profile of profitable and under-profitable accounts.

The outcomes of the segmentation are used to feed the strategic decisions on the portfolio. Besides the cancellation or the implementation of deductibles and exclusion clauses, we have proposed in a third stage to analyse the impacts of recalibrating the models. Therefore, we are back-testing a simplified version of the pricing model, together with another model where we apply a partial uplift of risk factors on the identified under-priced areas.

INTRODUCTION

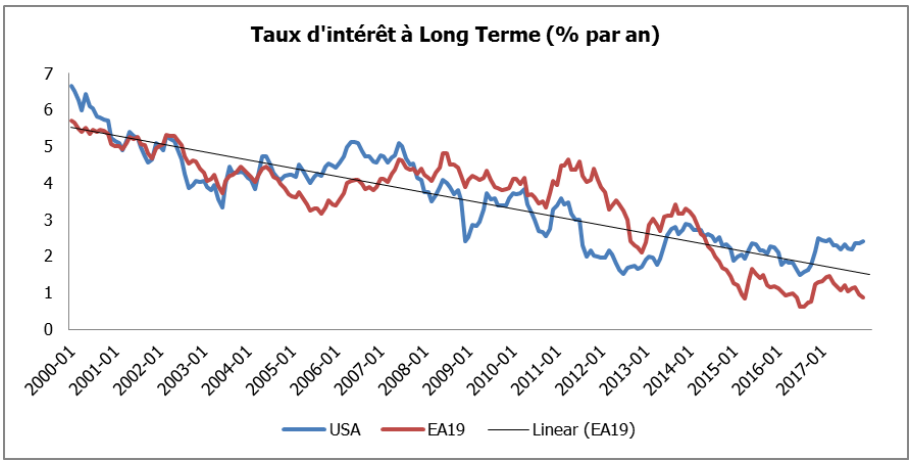
A la suite de la crise des *subprimes* aux Etats-Unis et de la faillite de la banque Lehman Brothers en 2008, la plupart des économies entrèrent en récession économique. Les banques centrales ont ainsi révisé leur politique monétaire afin de faire face à la crise de liquidité qui en découla. En agissant sur les taux directeurs, elles ont permis de refinancer les banques mais ont entrainé une chute des taux d'intérêts de manière durable.

Figure 1 –Taux d'intérêt à Court Terme pour les Etats-Unis et la zone Euro - OCDE



Malgré une légère reprise aux Etats-Unis, les taux sont restés bas dans la zone euro, et sont devenus négatifs pour la première fois dans l'histoire. Les assureurs ont ainsi vu leur résultat financier se réduire avec la baisse durable des taux d'intérêts. Ces chamboulements macroéconomiques ont non seulement impacté la proposition de produits, notamment en assurance vie, mais également les stratégies d'investissements des assureurs. La hausse brusque des taux d'intérêts pouvant réduire fortement la valeur des produits de taux dont sont composés majoritairement les bilans des assureurs.

Figure 2 - Taux d'intérêt à Long Terme pour les Etats-Unis et la zone Euro - OCDE

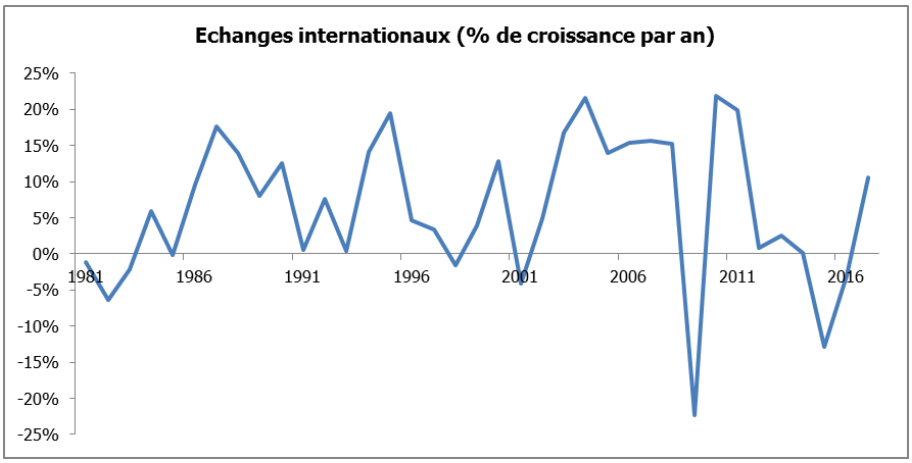


D'un point de vue de l'actionnariat, la problématique de la rentabilité des capitaux propres devient réelle, entrainant une volonté d'optimisation du capital, pouvant avoir des impacts sur la solidité financière en cas de sous-capitalisation.

Une baisse des résultats financiers, conjuguée à une pression sur les prix

Ces dernières années, l'assurance transport a souffert de la baisse du commerce international résultant de la récession mais aussi de l'incertitude provenant de l'annonce du Brexit. La montée en puissance des politiques protectionnistes comme celle des Etats-Unis, provoquera sans nul doute un repli des échanges internationaux, et un impact négatif sur le développement de l'assurance transport.

Figure 3 - Echanges internationaux - OMC

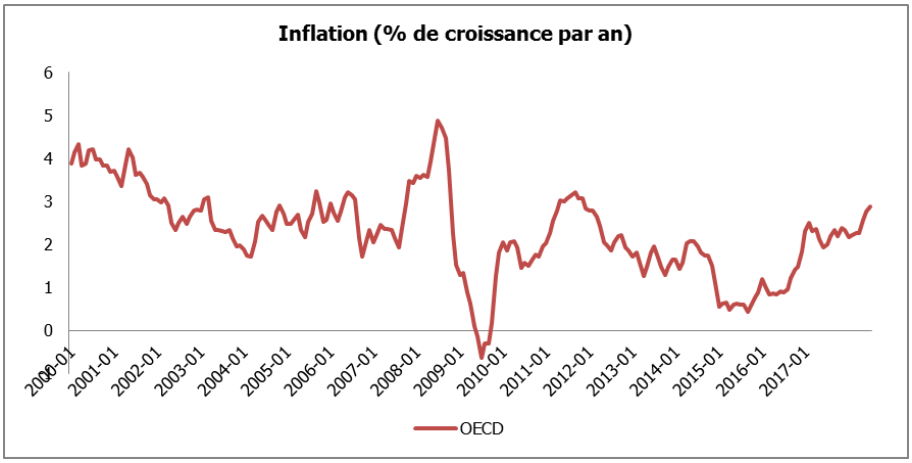


Au même moment, la conjoncture économique a poussé beaucoup d'entreprises industrielles ou financières à rechercher des moyens de réaliser des économies pour faire face à la baisse de leurs revenus. Les assurances étant jugées comme non prioritaires par la plupart des directions financières, font l'objet d'âpres négociations lors des renouvellements. La qualité du service et l'expérience client deviennent secondaires par rapport au prix. Dans ce contexte de marché d'assurance mou, la compétition sur les prix se fait donc de plus en plus dure. Les concurrents n'hésitent pas à réduire leurs marges pour attirer de nouveaux clients.

L'inflation des coûts provoque des difficultés pour garder des primes adéquates

Avec une baisse incontrôlée des marges, les affaires peuvent devenir souvent non profitables. En effet, malgré les nombreuses politiques monétaires mises en place par les banques centrales contre l'inflation, le prix des marchandises a continué à croitre ces dernières années. Une augmentation de l'inflation peut être observée sur l'indice des prix à la consommation, mesurée par l'OCDE, ayant notamment un impact sur le coût de la sinistralité. Le remplacement des cargaisons de marchandises entraine une augmentation des coûts des prestations d'assurances.

Figure 4 - Inflation mesurée par l'Indice des prix à la consommation - OCDE.



Ces dernières années, l'inflation des coûts s'est accompagnée d'une augmentation exponentielle des valeurs assurées. L'assurance transport devient de plus en plus exposée à de larges pertes dues à l'accumulation des marchandises et à leurs hautes valeurs. En effet, compte tenu de la hausse du coût du pétrole, les transporteurs opèrent des navires de plus en plus gros pour améliorer leur rentabilité et les ports doivent s'agrandir afin de les accueillir.

Cette accumulation des valeurs entraine une augmentation de la probabilité d'un évènement catastrophique à la fois d'origines humaines ou d'origines naturelles. Récemment, les assureurs transports ont été touchés par les ouragans Harvey, Irma ou Maria en 2017 mais également par des catastrophes d'origines humaines telles que l'incendie du Maersk Honam en 2018.

Table 1 - Top 10 des évènements catastrophiques en Marine de 2012 à 2016 – Munich Re

Evènements	Année	Pertes estimées
Superstorm Sandy	2012	$ 3 bn
Tianjin	2015	$ 2 bn
Costa Concordia	2012	$ 2 bn
MOL Comfort	2013	$ 500 m
Vermillion	2015	$ 400 m
Space X	2016	$ 285 m
Alpine Eternity	2015	$ 275 m
Hanjin	2016	$ 250 m
MSC Flaminia	2012	$ 150 m
Hurricane Matthew	2016	$ 100 m

Un pilotage actuariel rigoureux, permettra d'assurer la rentabilité et la solvabilité des assureurs face à la baisse des marges et à l'augmentation de la volatilité des résultats. L'adéquation des modèles de tarification est donc un enjeu primordial à la compétitivité mais aussi à la solidité financière des assureurs. Les résultats du *Lloyd's of London*, plus vieux marché d'assurance maritime du monde, montrent cette dégradation des marges et affichent des ratios combinés supérieurs à 100%. Avec la

pression sur les prix précédemment évoquée, ces résultats pourraient être expliqués soit par l'inadéquation des modèles de tarification, soit due à une inflation de la sévérité non prise en compte dans la tarification au fil du temps.

Compte tenu des faibles revenus d'investissement, la profitabilité de la ligne provient essentiellement du résultat technique du fait du niveau bas, voire négatif, des taux d'intérêts à court terme et du développement rapide des risques de l'assurance transport. Il convient par ailleurs de prendre également en compte l'environnement règlementaire dans le pilotage afin de pouvoir satisfaire les attentes des actionnaires en termes de rentabilité contre les capitaux propres engagés.

En effet, l'entrée en vigueur de la directive européenne Solvabilité 2, impose aux compagnies d'assurance des exigences minimales de capitaux règlementaires plus strictes. La seule considération de la profitabilité ne sera probablement pas suffisante pour assurer un retour sur investissement après impôts compétitif par rapport à d'autres lignes d'assurances ou à d'autres produits financiers. Dans un contexte de raréfaction des capitaux propres, l'enjeu sera de démontrer la rentabilité de cette ligne pour justifier d'éventuels investissements supplémentaires.

Une surveillance de la performance et la révision des facteurs de tarification

Le portefeuille de Chubb présente une bonne diversification géographique avec une présence dans plus d'une quinzaine de pays en Europe Continentale. Chubb assure une grande diversité de risques sur des durations différentes avec cependant une concentration du portefeuille en responsabilité civile et en dommages. Il convient cependant de mettre en place un processus de pilotage de portefeuille pour gérer la profitabilité et affiner à postériori le modèle de tarification.

Table 2 - Répartition des volumes de primes par LoB SII – Chubb SFCR au 31.12.2018

Lignes d'activités SII	GWP %
Incendie et autres dommages aux biens	36%
Responsabilité civile générale	31%
Pertes pécuniaires diverses	14%
Maritime, aérienne et transport	7%
Autres	12%

L'utilisation des algorithmes d'apprentissage a souvent été étudiée avec une finalité en tarification, mais ils restent en pratique peu utilisés, car jugés peu transparents par la direction et ne permettant pas de visualiser la relation entre les facteurs de risques. Par ailleurs, la nécessité d'avoir une part de marché importante pour l'obtention d'une quantité de données représentative pour la construction d'un modèle robuste est un frein pour bon nombre d'acteurs.

La proposition de ce mémoire est de trouver une autre utilité à ces méthodes d'apprentissages, avec dans un premier temps leur emploi pour la visualisation des données. Cela permettra d'avoir une idée

de la composition du portefeuille, pour s'assurer d'une bonne diversification et pour identifier les facteurs de risques.

Dans un second temps, l'objectif est de segmenter le portefeuille afin de mettre en exergue les zones de sous-tarification, aux vues des contraintes de rentabilité imposées par l'actionnariat sur le capital économique engagé. Par la suite, ces informations permettront d'améliorer la prise de décision stratégique quant à la résiliation des affaires ou à réajuster le modèle pour obtenir une meilleure estimation du risque. L'impact du recalibrage du modèle sur la rentabilité sera étudié ultérieurement.

Par mesure de confidentialité, les données de l'étude seront regroupées en bandes et une partie des informations de rentabilité sera remplacée par des hypothèses fictives. Les analyses sont réalisées à l'aide du logiciel statistique R.

1. L'ASSURANCE TRANSPORT

1.1. Historique

L'assurance transport fut à la genèse de toute forme d'assurance. Les premiers aspects peuvent être retrouvés en antiquité, avec le prêt à la grosse aventure, permettant à un marchand ayant financé son expédition par un prêt, de ne pas rembourser son emprunt en cas de vol ou en cas de naufrage de son navire.

Des formes de mutualisation du risque peuvent être également retrouvées à cette période et des principes similaires se sont mis en place tout au long de la route de la Soie pour assurer les caravanes et diminuer les pertes des marchands en mutualisant les risques. L'assurance maritime se développa par la suite, avec les marchands italiens de Venise et de Gênes au 13ème siècle.

Le monde change avec les premières grandes découvertes maritimes du 15ème siècle et la montée en puissance des premiers grands empires coloniaux de la péninsule ibérique, mais il a fallu attendre le 17ème siècle pour que l'assurance maritime soit organisée et légiféré en France. En effet, avec la création de la Compagnie Générale des Assurances et Grosses Aventures par Colbert, Louis XIV souhaite faire de la France une grande puissance maritime.

Dans le même temps, le développement de la puissance commerciale de l'empire coloniale britannique, autre grande puissance maritime de l'époque, présente aux caraïbes et dans les Indes orientales permet le développement rapide du premier marché d'assurance maritime à Londres avec le *Lloyd's of London*.

Aujourd'hui, l'assurance transport ne concerne plus seulement l'assurance maritime et le transport des marchandises, mais s'est étendue à couvrir tous les modes de transports et à assurer l'intégrité physique de ces véhicules. A l'heure de la mondialisation et du commerce international, les traités de libres échanges ont permis de favoriser le développement de l'importation et l'exportation des marchandises, au bénéfice de l'assurance transport. Avec la digitalisation et le boom du e-commerce, l'assurance transport conserve une opportunité de croissance considérable dans les années à venir.

1.2. Dispositions relatives à l'assurance transport

En France, les articles 171-1[1] et suivants du Code des Assurances régissent les contrats d'assurances ayant pour objet de garantir les risques relatifs au transport de marchandises et les risques maritimes, aériens ou aéronautiques. Deux sous branches se distinguent : l'assurance corps de navire ou d'aéronef nommée « *Hull* » en anglais et l'assurance transport de marchandises nommée « *Cargo* » en anglais. La suite de ce mémoire s'intéressera uniquement à la partie transport de marchandises.

1.2.1. Description des conditions générales

1.2.1.1. Durée d'assurance

Ces assurances protègent les commerçants lors du transport par voies maritimes, aériennes, fluviales ou terrestres. Sauf convention contraire, les marchandises sont couvertes en transit depuis la sortie des entrepôts de l'expéditeur jusqu'à l'entrée dans les magasins du destinataire. Cela inclut donc les stockages intermédiaires et les risques de chargement ou de déchargement, par exemple lors du transfert des porte-conteneurs dans les ports maritimes.

1.2.1.2. Les risques assurés

En cours de transit, les marchandises sont exposées soit :

- à des risques ordinaires de transport, provenant de causes naturelles ou à des causes humaines, telles que des dommages ou des pertes.
- à des risques exceptionnels ou des risques de guerre trouvant leurs origines dans des conflits politiques ou sociaux.

Ces risques sont qualifiés d'avaries particulières lorsqu'ils résultent de dommages ou de pertes survenus pendant le transport ou durant le stockage. En revanche, on qualifie ces risques d'avaries communes, lorsque, pour échapper à un danger menaçant à la fois le navire et la cargaison, le capitaine est conduit à décider à un sacrifice du navire ou des marchandises dans l'intérêt commun. Le sacrifice financier est ensuite réparti au prorata des valeurs de chacun des propriétaires de marchandises sur le navire.

Les avaries communes ou *general average* en anglais, peuvent exposer les assureurs à d'importants montants, avec l'accumulation des valeurs sur les porte-conteneurs. La branche transport peut donc

[1] **Art 171-1** du CdA : « Est régi par le présent titre tout contrat d'assurance qui a pour objet de garantir :
1° Les risques maritimes ;
2° Les risques aériens ou aéronautiques ;
3° Les risques relatifs à la responsabilité civile au titre d'une opération spatiale ;
4° Les risques relatifs au transport de marchandises par voie maritime, aérienne ou terrestre. […] »

être exposés à des sinistres atypiques. Les dispositions des avaries communes en France sont décrites dans le Code des Transports aux articles L5133-1 et suivants.

1.2.1.3. Les types de polices

En assurance transport, comme dispose l'article L173-17[2] du Code des Assurances, les polices d'assurances peuvent couvrir un unique voyage ou plusieurs sur une période annuelle, ces dernières sont des polices dites d'abonnement ou à aliment.

- Les polices à aliment sont adaptées aux expéditions échelonnées sur une période indéterminée et où l'assuré informe avant chaque envoi de la nature et de la valeur de l'expédition.
- Les polices d'abonnement couvrent à contrario automatiquement toutes les expéditions, quels que soient les marchandises, les destinations et le mode de transport.

Le mémoire traitera les polices annuelles qui présente un intérêt stratégique dans le pilotage d'un portefeuille. En revanche, de par sa nature, il est difficile d'évaluer précisément l'exposition intrinsèque des polices à abonnement, contrairement à la police au voyage et à aliment où le risque de chaque expédition est bien identifié. De fait, l'absence d'information sur chaque voyage rend le calibrage du modèle de tarification moins précis et peut amener de la volatilité dans les résultats.

1.2.2. Les conditions particulières

1.2.2.1. Les couvertures

Deux sortes de conventions peuvent être trouvées en général :

- Les garanties restrictives où la couverture est limitée aux pertes et dommages causés par l'un des événements énumérés dans le contrat. L'assuré doit donc établir la preuve de la survenance de l'un des événements et établir le lien de causalité entre le dommage et l'événement.
- Les garanties Tous Risques où la couverture est étendue à tous les dommages et pertes survenus causés par des événements autres que ceux listés dans les exclusions. La charge de la preuve est ici inversée, c'est à l'assureur de prouver l'application d'une exclusion.

Des pratiques propres à chaque pays peuvent être retrouvées dans les clauses types proposées par le marché. Ainsi en France, les couvertures sont bien souvent plus étendues que sur le reste du marché européen. Dans certains cas, des conditions générales types sont utilisées, telles que les *Institutes Cargo Clauses ICC – A/B/C* développées par la Chambre de Commerce Internationale. Le A est considéré

[2] **Art L172-17** du CdA : « Les marchandises sont assurées, soit par une police n'ayant d'effet que pour un voyage, soit par une police fonctionnant par déclaration d'aliment. »

comme étant « Tout risques » et des restrictions additionnelles s'appliquent sur le B et le C. Une comparaison des clauses réalisée par une entreprise de logistique peut être visualisée à la Table 3.

Table 3 - Comparaison des clauses ICC par TLC Logistics

Causes	ICC-A	ICC-B	ICC-C
Naufrage, échouage, retournement du navire ou tout autre événement similaire	OUI	OUI	OUI
Retournement des véhicules de transport terrestres	OUI	OUI	OUI
Abordage de navires ou d'autres moyens de transport	OUI	OUI	OUI
Collision avec tout autre objet à l'exception du moyen de transport (sauf transport maritime)	OUI	OUI	OUI
Débarquement forcé à l'aéroport	OUI	OUI	OUI
Incendie ou explosion	OUI	OUI	OUI
Tremblement de terre, éruption volcanique ou foudre	OUI	OUI	NON
Actes malveillants de tiers	OUI	NON*	NON*
Vols et cambriolages	OUI	NON	NON
Accidents divers	OUI	OUI	OUI
Déchargement forcé par-dessus bord	OUI	OUI	OUI
Pertes de la cargaison par-dessus bord	OUI	OUI	NON
Risques de guerre, émeutes et grèves	NON*	NON*	NON*
Piraterie	OUI	NON	NON
Pénétration de l'eau dans le navire, dans le lieu de stockage, dans le conteneur ou tout autre unité de transport	OUI	OUI	NON
Perte au chargement/déchargement (seulement en cas de perte totale)	OUI	OUI	NON
Tout autre risque de perte ou de dommage physique à la cargaison non défini ci-dessus	OUI	NON	NON

** Option*

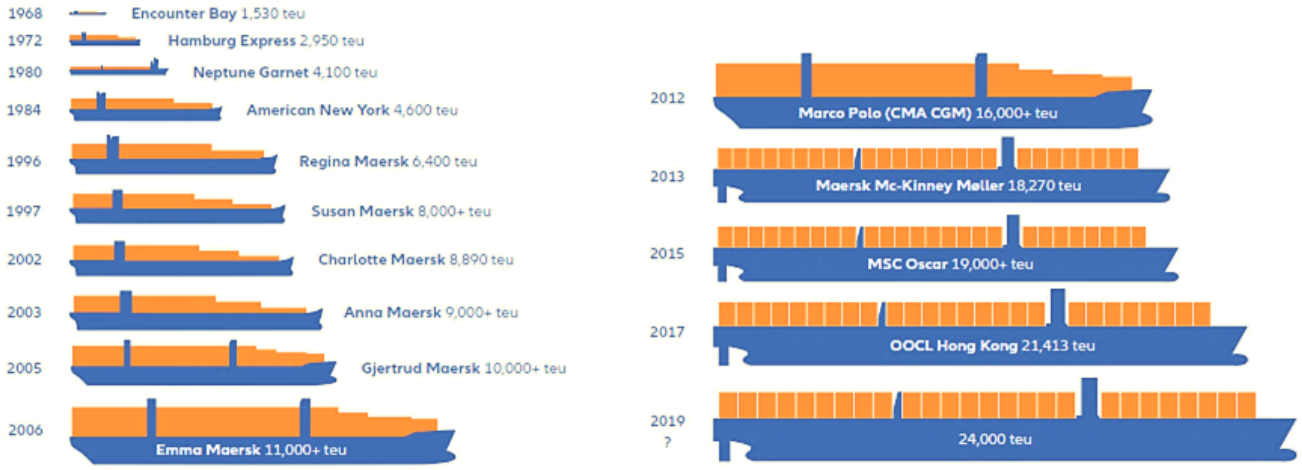
1.2.2.2. Les spécificités de responsabilités

Publiés par la Chambre du Commerce Internationale, les *Incoterms* disposent les responsabilités des vendeurs et des acheteurs durant l'acheminement des biens. Compte tenu des règles complexes de responsabilité, il est souvent possible de former un recours contre un tiers pour la responsabilité du sinistre, ce qui entraine ainsi des développements négatifs de la charge de sinistre pour l'assureur. Par ailleurs, le paiement des indemnités est relativement rapide dans cette branche d'assurance.

1.3. L'assurance transport en chiffres

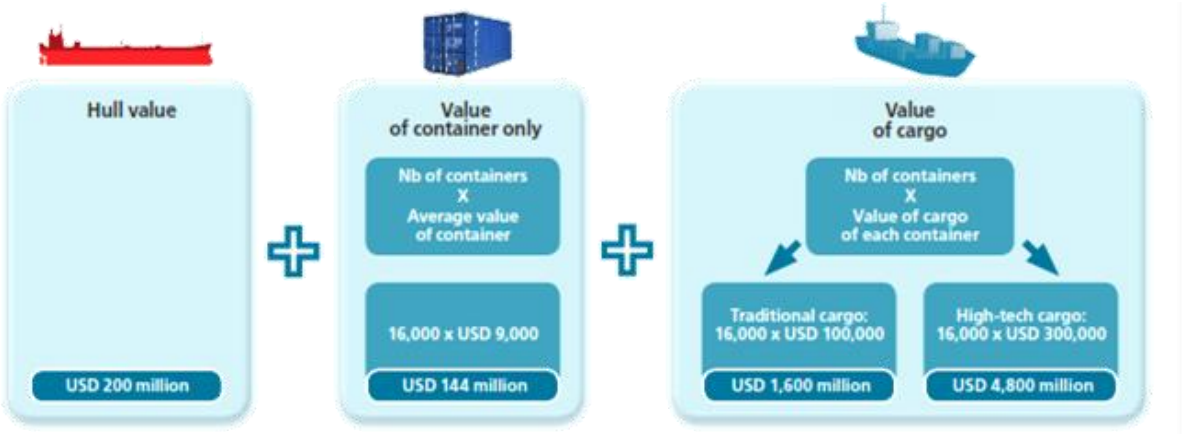
Ces dernières années, les prix des assurances transport ont durablement baissé avec l'excédent de capacité d'assurance sur les marchés. Avec les mauvais résultats chroniques et la pression sur la rentabilité des capitaux propres, de nombreux assureurs se retirent de certains pays d'Europe. La mondialisation et la hausse des prix du pétrole font que les transporteurs opèrent désormais des navires de plus en plus gigantesques, afin de continuer à proposer des coûts de transport de marchandises faibles grâce aux économies d'échelles. En l'espace de 50 ans, la capacité des porte-conteneurs a été multipliée par 14.

Figure 5 - Développement des porte-conteneurs depuis 1968 – AGCS Safety Shipping Review 2019



Cette accumulation de valeurs s'observe également sur le continent, avec l'agrandissement des entrepôts et des ports pour les accueillir, entrainant in-fine une probabilité accrue de sinistres graves. Le sinistre maximum possible pouvant survenir notamment en cas d'avarie commune, de feux à bord ou lors de catastrophes naturelles.

Figure 6 - Scenario d'accumulation sur un porte-conteneurs – SCOR Technical Newsletter 2011



Ainsi, un porte-conteneurs de classe triple E comme le Maersk Mc-Kinney Møller, pouvant transporter plus de 18,000 conteneurs, peut ainsi représenter un sinistre maximum possible compris entre 2 et 5 milliards USD comme le montre le scénario d'accumulation de la SCOR. A titre de comparaison, la prime mondiale en 2017, estimée par *l'International Union of Marine Insurance* est de 16.1 milliards USD en assurance transport et 28.5 milliards USD en assurance corps de navire. Les catastrophes d'origines humaines ou de catastrophes naturelles étant dorénavant de plus en plus fréquentes, elles peuvent rapidement annihiler la capacité disponible.

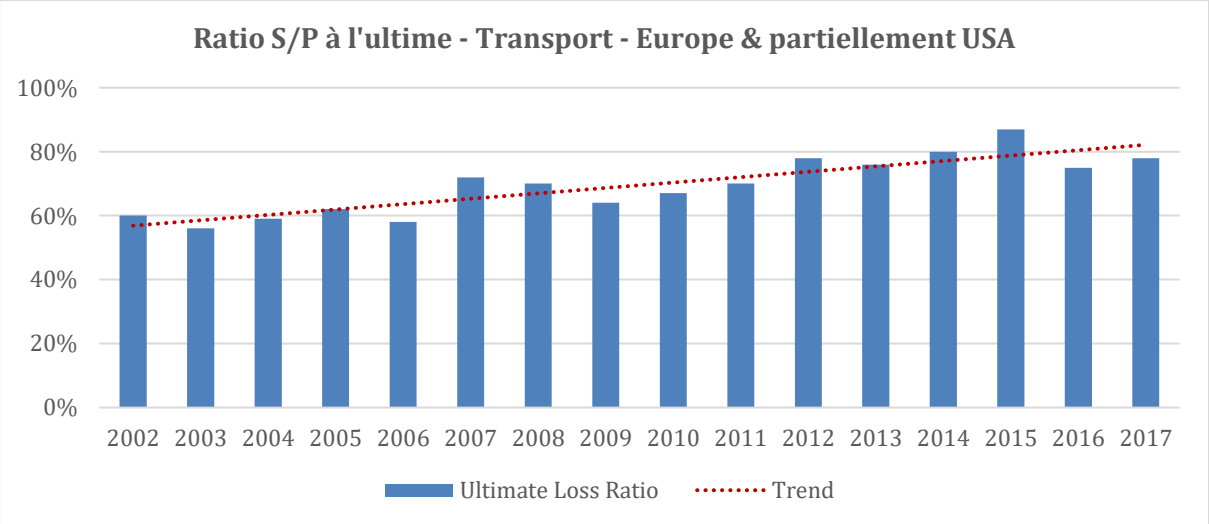
Le marché international du Lloyd's a ainsi accumulé des pertes sur les trois dernières années consécutives, avec respectivement -343 millions de GBP en 2018, -469 millions de GBP dues aux catastrophes naturelles en 2017 et -129 millions de GBP en 2016. Avec 57 syndicats déficitaires parmi les 88 souscrivant des risques en *Marine*, le Lloyd's a été obligé de prendre des mesures drastiques. Ainsi, de nombreux syndicats[3] ont dû arrêter leurs opérations et les restants ont été priés de mettre en place des actions correctives sur leur portefeuille.

Table 4 - Performance annuelle de la ligne "Marine" au Lloyd's of London – Rapport annuel 2018

Année	Primes émises brutes (£m)	Ratio Combiné (%)	Résultat technique (£m)
2013	2,195	95%	84
2014	2,140	95%	84
2015	2,245	94%	108
2016	2,470	106%	-129
2017	2,506	122%	-469
2018	2,603	116%	-343

Cette tendance se vérifie sur les résultats des marchés Européens et d'une partie des Etats-Unis, avec des ratios S/P en hausse depuis 2002.

Figure 7 - Ratio S/P des marchés Européen & USA - IUMI Global Marine Insurance Report 2018



[3] C'est le cas notamment d'A*cappella, AmTrust, Advent*, *Argo, Barbican* et *Sirius.*

2. CADRE REGLEMENTAIRE

2.1. Solvabilité 1

Le secteur de l'assurance a pour principale particularité l'inversion du cycle de production. Contrairement au secteur industriel, le coût de revient n'est pas connu lors de la fixation du prix du produit. En effet, bien que certains coûts soient connus comme les frais de personnel ou les loyers, le coût des sinistres suit une variable aléatoire. Par ailleurs, le décalage temporel entre l'encaissement des primes d'assurance et le paiement effectif des sinistres nécessite un contrôle de la solvabilité des compagnies d'assurance, d'autant plus important dans le cas d'assurance à long-terme comme dans le cas de l'assurance décennale. Il est donc primordial d'assurer la pérennité des compagnies d'assurance, de par leur rôle de protection des assurés lors de la survenance des aléas, mais aussi pour leur rôle de financement de l'économie avec l'investissement des provisions techniques.

2.1.1. Principes

Pour les raisons mentionnées précédemment, l'assurance est un secteur régulé, nécessitant un agrément pour exercer cette activité, comme le dispose l'article L321-1 du Code des Assurances[4]. Outre les différentes règles d'exercices comme le principe de spécialisation, décrites dans le Code des Assurances en France, des lois existent pour garantir la solvabilité des assureurs. Ainsi, les législateurs de certains pays ont promulgué des lois afin de fixer des seuils minimums de capital comme les Etats-Unis d'Amériques avec le *Risk Based Capital* ou le *Swiss Solvency Test* en Suisse, pour ne citer que les plus importants. La France a de son côté, mis en place en 1973, des règles pouvant s'apparenter à un régime dit « Solvabilité 1 ». Elles comprennent notamment trois principes :

- Des provisions techniques suffisantes, c'est-à-dire calculées avec des hypothèses prudentes. Ces dispositions sont décrites à l'Art. R343-1du Code des Assurances[5]
- Des provisions techniques représentées par des Actifs éligibles, d'un montant équivalent et d'une certaine qualité décrit à l'Art. R332-1 du Code des Assurances[6]
- Disposer d'une marge de solvabilité, constituée des éléments décrits à l'Art. R334-3 du Code des Assurances[7]

[4] **Art. L321-1** du CdA : Les entreprises […] ne peuvent commencer leurs opérations qu'après avoir obtenu un agrément administratif délivré par le l'Autorité de contrôle prudentiel […]
[5] **Art. R343-1** du CdA : Les entreprises […] doivent, être en mesure de justifier de l'évaluation des éléments suivants : 1° Les provisions techniques suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés […]
[6] **Art. R332-1** du CdA : Les engagements réglementés mentionnés à l'article R. 331-1 doivent, à toute époque, être représentés par des actifs équivalents. […] Les engagements pris dans une monnaie doivent être couverts par des actifs congruents, c'est-à-dire libellés ou réalisables dans cette monnaie.
[7] **Art. R334-3** du CdA : La marge de solvabilité est constituée de […]

2.1.2. Marge de Solvabilité

La marge de Solvabilité a pour but d'absorber les pertes exceptionnelles. De ce fait, elle est constituée principalement des fonds propres nets d'actifs incorporels, auxquels on ajoute les plus-values latentes et des quasi fonds propres. Ces fonds mis à disposition par les actionnaires ont un rendement attendu.

Le montant de cette marge est déterminé à l'aide de l'article R334-5 du Code des Assurances, qui dispose les différentes méthodes de calcul de l'exigence minimale de la marge de solvabilité en se basant sur un calcul forfaitaire suivant deux méthodes décrites à l'*Annexe 1 – Calcul de la marge de Solvabilité 1.*

La marge la plus élevée résultant des deux méthodes de calcul est gardée, et le montant de marge du dernier exercice est retenu si celui-ci est plus élevé. La marge constituée doit être supérieure à la marge de solvabilité requise, calculée à l'aide de la formule fournie par le Code des Assurances et au-dessus d'un montant minimum défini par branches et formes sociales.

L'assureur doit détenir une marge minimum de solvabilité comprise entre un tiers du minimum et 2 ou 3 millions selon la branche. Les méthodes de calculs ont l'avantage d'être simple à mettre en œuvre et à piloter, cependant les mesures du régime Solvabilité 1 ne permettent pas de prendre en compte l'exhaustivité des risques inhérents à un assureur, notamment les risques financiers et opérationnels.

2.2. Solvabilité 2

Afin de tirer les leçons de la crise de liquidité de 2008, un groupe de travail a été constitué afin de plancher sur le projet Solvabilité 2 avec l'objectif d'harmoniser au niveau européen la supervision et d'améliorer l'évaluation et le contrôle des risques pour protéger le consommateur d'une faillite.

Entrée en vigueur le 1er janvier 2016, la directive Solvabilité 2 s'inspire de la règlementation bancaire Bâle 2 en se déclinant sous trois piliers : Exigences quantitatives, exigences qualitatives et de gouvernance et les exigences d'information à destination du public et du superviseur.

Le mémoire se focalisera sur les exigences quantitatives pour déterminer la rentabilité du capital économique. Le reste des conditions d'applications de la directive Solvabilité 2 et les formules de calculs du modèle standard sont présents dans la « Directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice » et dans le « Règlement délégué (UE) 2015/35 de la Commission du 10 octobre 2014 ».

Méthode de calcul – Exigences quantitatives

La directive fournit une liste prédéfinie de scénarios et de chocs forfaitaires avec le modèle standard mais permet une évaluation différente de ces risques sous autorisation de l'autorité de contrôle prudentiel lorsque le profil de risque de l'entité le justifie :

1. Modèle standard
2. Modèle standard avec des paramètres propres à l'entité (*USP*)
3. Modèle interne partiel
4. Modèle interne

Il faut cependant garder à l'esprit que la complexité et les coûts associés à l'implémentation peuvent parfois être beaucoup plus grands que l'économie en capital lorsque l'on veut diverger du modèle standard. Le pilotage du capital peut également s'avérer plus complexe avec un modèle interne. Enfin, bien que le calibrage de la directive ait fait l'objet de multiples consultations publiques, la prise en compte du profil de risque de chaque compagnie n'est pas aisée.

Il est ainsi prévu plusieurs clauses de revoyure afin de réviser les paramètres inhérents à la formule standard. En effet, les différents facteurs peuvent avoir des impacts macroéconomiques puisque les différents assureurs peuvent être amenés à prendre des décisions stratégiques pour réduire la consommation en capital comme revoir leur allocation d'actifs et donc le financement de l'économie ou favoriser la souscription sur certaines lignes d'activités.

Chronologie

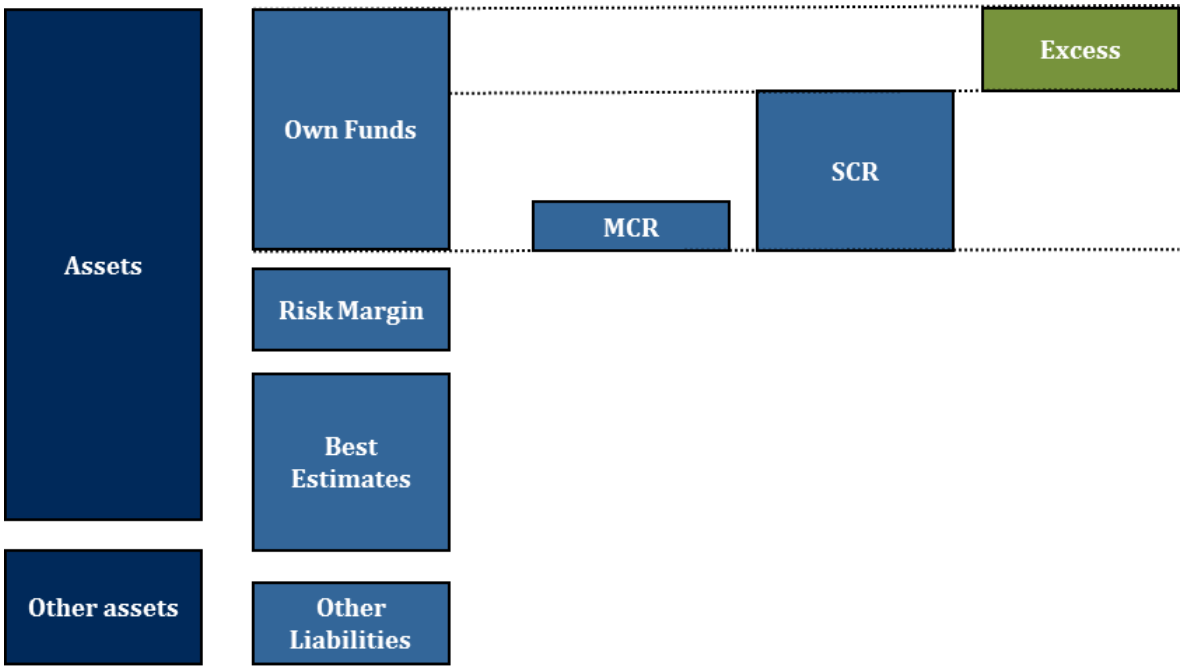
- 2016 : Entrée en vigueur de la directive Solvabilité 2
- 2018 : Consultation publique pour la révision des actes délégués
- 2020 : Implémentation des mesures ayant fait l'objet de la première consultation et mise en place de la deuxième consultation publique.
- 2022 : Implémentation des mesures ayant fait l'objet de la deuxième consultation

2.2.1. Bilan Prudentiel

La nouveauté de la directive Solvabilité 2 repose principalement sur la construction d'un bilan prudentiel avec une évaluation économique des différents postes du bilan. Avec la convergence des normes comptables internationales, ce dernier point a pour objectif de tenir compte de la fluctuation de valeur des éléments du bilan, qui est d'autant plus vraie en période de crise. Les actifs sont donc évalués à leur valeur de marché et le passif avec la meilleure estimation possible en utilisant des hypothèses réalistes, contrairement aux normes locales qui préconisent bien souvent un provisionnement conservateur.

Le bilan prudentiel est réalisé pour les entités légales d'assurances, au niveau solo mais également au niveau groupe et sous-groupe des organismes d'assurances comme le dispose l'article 212 de la Directive Solvabilité 2.

Figure 8 - Bilan Prudentiel Solvabilité 2 - Market Value Balance Sheet



2.2.1.1. Les fonds propres

Les fonds propres correspondent en partie aux capitaux propres amenés par les actionnaires. Ils sont par la suite abondés des excédents de profits qui n'ont pas été utilisés pour payer les dividendes. Avec les autres ressources stables considérées comme quasi-fonds propres, ces ressources sont utilisées comme marge de sécurité en cas de pertes sur les marchés financiers ou sur le portefeuille de l'assureur.

Contrairement à Solvabilité 1, l'actif net est considéré en tenant compte des fonds propres auxiliaires, et non exclusivement des capitaux propres des actionnaires et des quasi-fonds propres. L'objectif est donc de calculer au mieux les ressources restantes à la suite de la survenance d'un évènement rare afin de rembourser les créanciers et les assurés. La valeur des actifs et du passif pouvant évoluer, il faut donc bien évaluer les risques de variation à la baisse des actifs mais aussi leur liquidité et les risques de hausse du passif.

2.2.1.2. L'évaluation des risques

L'évaluation du capital nécessaire se fait en imposant un choc sur l'ensemble des postes du bilan. Ainsi, les différents risques inhérents sont évalués, sur un horizon d'un an en considérant un évènement ayant une période de retour d'un tous les deux cents ans.

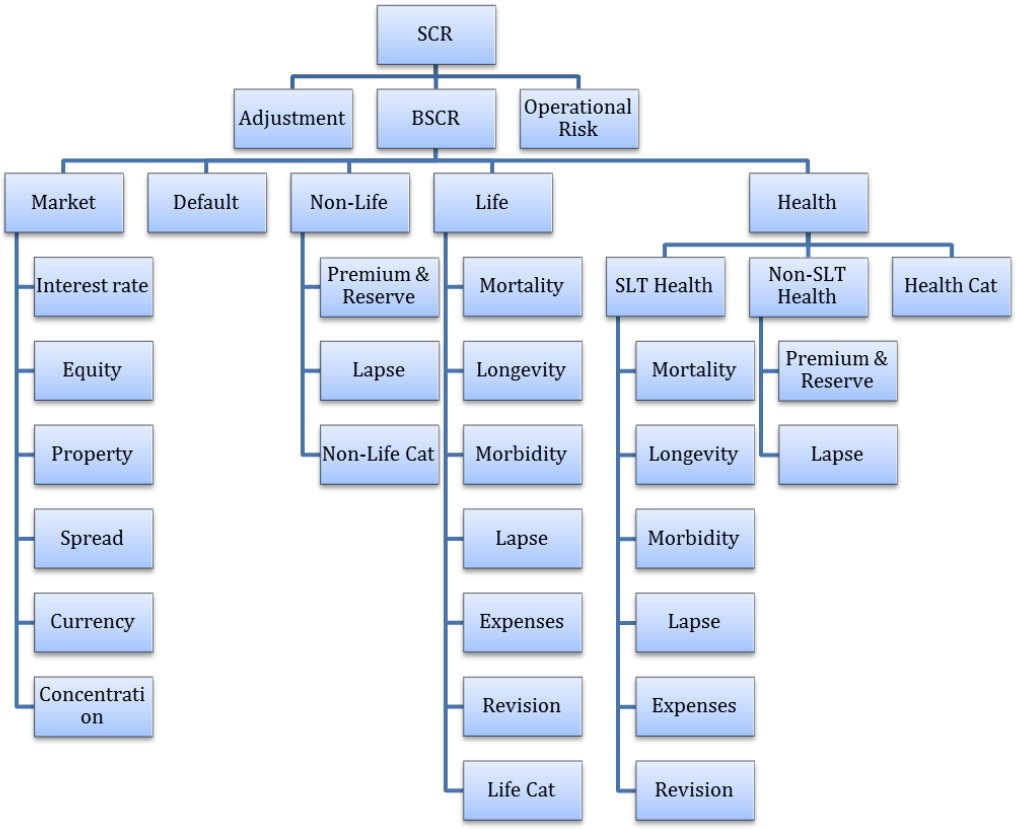
Comparée à Solvabilité 1, la directive fournit une liste plus complète de risques avec une matrice de corrélation permettant l'agrégation des différents risques en prenant en compte une certaine relation dans leur survenance. Les risques non prévus dans le modèle standard doivent également être considérés dans le cadre de l'*Own Risk Solvency Assessment* (ORSA) avec l'évaluation du Besoin Global de Solvabilité (BGS).

2.2.2. Matrice des risques du Modèle Standard

La matrice des risques du modèle standard montre les différents modules et sous modules de risques qui composent le besoin en capital dénommé Solvency Capital Requirement (SCR). Il convient de se référer au Règlement délégué (UE) 2015/35 de la Commission du 10 octobre 2014 pour les modalités et formules de calcul des différents modules de risques.

Le besoin en capital est défini comme la somme des différents sous-modules de risques diversifiés dans le *Basic Solvency Capital Requirement (BSCR)* avec le risque opérationnel qui est réputé non diversifiable. Enfin, les divers ajustements liés aux provisions techniques et aux impôts différés résultant des mouvements sur le bilan prudentiel sont également pris en compte pour réduire le besoin en capital.

Figure 9 - Matrice des risques du Modèle Standard Solvabilité 2



Les étapes de calcul du risque de prime et de réserve et l'agrégation pour le risque de souscription non-vie pour le Modèle Standard sont affichées à l'*Annexe 2 – Calcul du risque de souscription non-vie* et à l'*Annexe 3 – Calcul du risque de prime et de réserve en Modèle Standard*.

2.2.3. Le risque de prime en Modèle Interne

Puisque le pilotage de portefeuille ne porte qu'uniquement sur le résultat technique, la volonté est de ne garder que le capital économique provenant du risque de prime pour calculer l'indicateur de rendement. Ainsi, conformément à la directive, le risque de prime est le risque d'avoir un volume de prime insuffisant pour faire face à un évènement survenant tous les 200 ans. Chubb possède un modèle interne pour la modélisation du capital requis Solvabilité 2. Pour ce module, la charge de sinistre doit être calculée à un quantile de 99.5%. La différence avec l'estimation moyenne de sinistre, aussi utilisé pour l'estimation de la prime, permet de déterminer le capital requis que l'on nomme l'*Expected Shortfall* $\mathbb{ES}_{99.5\%}$.

$$SCR_{Prime} = \mathbb{ES}_{99.5\%} = VaR_{99.5\%} - \mathbb{E}(S)$$

Dans un premier temps, il faut calculer la charge totale de sinistre *S* qui surviendrait pendant la durée de couverture du portefeuille étudié à l'aide du modèle collectif.

Elle est définie comme :

$$S = \sum_{k=1}^{N} X_k$$

La charge totale de sinistre suit une loi composée de la fréquence et de la sévérité Cf. Denuit et Charpentier [2004]. La fréquence est supposée indépendante des coûts. Il convient alors d'estimer à l'aide de lois statistiques appropriées, la fréquence de survenance d'un sinistre N et sa sévérité X. Mathématiquement, cela revient à calculer l'espérance du nombre moyen de sinistres $\mathbb{E}(N)$ et leur coût moyen $\mathbb{E}(X)$.

Pour une variable aléatoire de comptage :

$$\mathbb{E}(N) = \sum_{n \in N} \mathbb{P}(N = n) = \sum_{n \epsilon N} \mathbb{P}(N > n)$$

Pour une variable aléatoire continue positive :

$$\mathbb{E}(X) = \int_{x \in \mathbb{R}+}^{\infty} x f(x)\, dx = \int_{x \in \mathbb{R}+}^{\infty} \mathbb{P}(X > x)\, dx$$

La combinaison des premiers moments des lois utilisées pour modéliser la fréquence et la sévérité permet l'estimation du montant moyen de sinistres attendu.

$$\mathbb{E}(S) = \mathbb{E}(N) \cdot \mathbb{E}(X)$$

$$\mathbb{V}(S) = \mathbb{E}(N) \cdot \mathbb{V}(X) + \mathbb{V}(N) \cdot \mathbb{E}(X)^2$$

Le calcul de la $VaR_{99.5\%}$ correspond à la charge totale de sinistre S définie ci-dessus, mais qui surviendrait à la suite d'un évènement avec une période de retour d'une fois tous les deux cents ans. L'évaluation se fait avec les valeurs aux quantiles de 99.5%. Le capital requis du portefeuille sera utilisé dans le chapitre suivant pour le pilotage du portefeuille.

3. PILOTAGE DE PORTEFEUILLE

3.1. Rentabilité du portefeuille

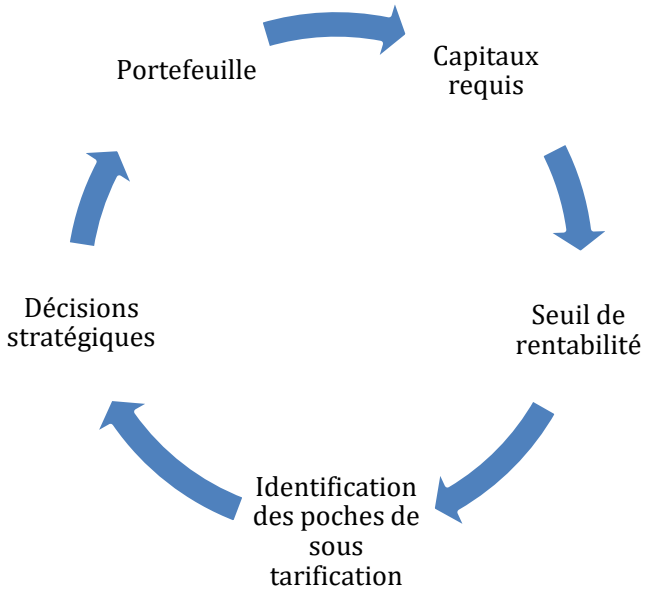
3.1.1. Revue de portefeuille

Compte tenu des changements macroéconomiques, il est important de réaliser des revues approfondies de la performance du portefeuille pour déterminer les segments profitables et non-profitables. Cet exercice permet in fine de vérifier l'adéquation des modèles de tarification et de les réajuster si besoin.

Dans le cadre du cycle de revue du portefeuille, il est important d'identifier les caractéristiques des comptes qui ne permettent pas de répondre aux objectifs de rentabilité fixés par l'actionnariat vis-à-vis des capitaux économiques immobilisés.

Pour ce faire, l'utilisation des méthodes d'apprentissage permettra de segmenter le portefeuille en fonction de la performance et d'en déterminer un profil. Subséquemment à l'identification de ces comptes non-profitables, la direction pourra décider entre autres de la résiliation ou de la re-tarification de ces affaires. Parallèlement à cet exercice, une redéfinition de la stratégie de souscription, permettra de cibler des profils de clients rentables en tenant compte des menaces et opportunités du marché.

Figure 10 - Cycle de revue du portefeuille

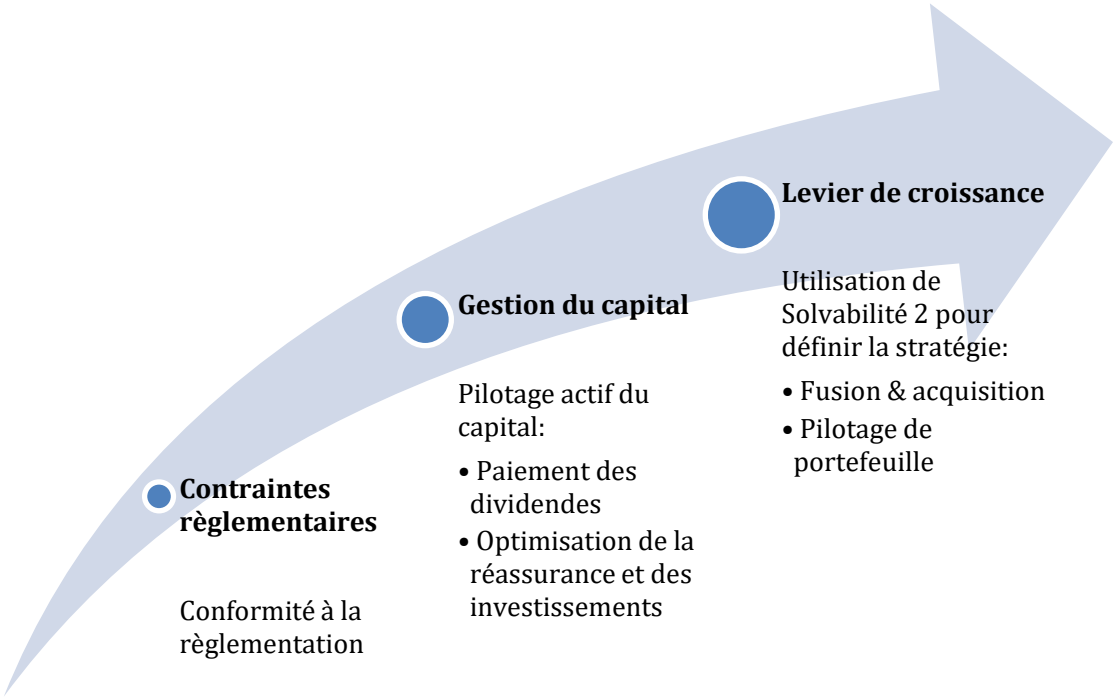


3.1.2. Intégration de la dimension risque dans la stratégie

Contrairement à une grande partie de l'industrie, l'assurance est un secteur très règlementé. Avec l'apparition de la directive Solvabilité 2, les normes ont été harmonisées en Europe, réduisant la disparité des « règles du jeu » entre les différents assureurs. Des différences restent encore présentes, dans une certaine mesure, car la directive ne s'applique qu'aux assureurs européens même si les filiales étrangères des groupes européens sont comprises dans le périmètre de supervision.

Les éléments de la directive peuvent être vus comme des leviers de compétitivité, avec l'accent mis sur une prise de décision avec la dimension risque. L'intégration de la directive dans la stratégie des entreprises peut prendre plusieurs étapes : la simple conformité à la règlementation, ou le choix d'optimiser la prise de risque en utilisant la directive Solvabilité 2 comme un levier de croissance, pour définir une stratégie de maximisation de la rentabilité.

Figure 11 - Stades d'implémentations de la directive Solvabilité 2 dans la stratégie



Dans notre objectif de pilotage, un seuil de rentabilité sera déterminé afin de tenir compte de la volatilité et de la duration du risque. Les différents comptes du portefeuille seront classés comme profitables ou non en fonction du seuil pour des capitaux règlementaires requis en plus du ratio S/P.

3.1.3. Détermination du seuil de rentabilité

La plupart des marchés développés sont soumis à un contexte d'austérité avec des taux de rendements des actifs financiers bas voire négatifs. Les clients font face à des coupes budgétaires entrainant les marchés d'assurances dans un cycle baissier où les primes sont tirées vers le bas.

Avec la baisse des rendements financiers et le risque de subir des pertes techniques, il est important de rassurer les actionnaires sur le potentiel de rentabilité de leurs placements et garantir les investissements pour le développement du portefeuille.

En partant du rendement attendu par les actionnaires, exprimé en coût du capital pour l'entreprise, un seuil de rentabilité minimum est déterminé pour les comptes du portefeuille. Il faudra ainsi que les ratios S/P ne dépassent pas ce seuil afin de rémunérer les actionnaires.

Pour un portefeuille donné, cela revient à inclure le coût du capital immobilisé, noté *Cost of Capital - CoC,* dans le calcul de la prime. La prime pure revient à modéliser l'espérance de sinistre :

$$Prime\ pure = Espérance\ de\ sinistre$$

Elle est ensuite chargée des frais fixe en valeur monétaire et d'un pourcentage de frais variables :

$$Prime\ chargée = \frac{Espérance\ de\ sinistre + Frais\ fixe}{(1 - Frais\ variables\ \%)}$$

Le coût d'immobilisation du capital requis est ensuite inclus dans la prime comme des « frais » additionnels. Puisque le paiement des sinistres sont des flux financiers, ils seront actualisés en utilisant une courbe de taux sans risque provenant de l'EIOPA.

$$Prime\ cible = \frac{(Capital \times CoC\ \%)/(1 - Taxe\ \%) + Espérance\ de\ sinistre\ actualisée + Frais\ fixe}{(1 - Frais\ variables\ \%)}$$

Ainsi le ratio combiné et le ratio S/P cible deviennent :

$$Ratio\ Combiné\ Cible = \frac{Espérance\ de\ sinistre + Frais\ fixe}{Prime\ cible} + Frais\ variables\ \%$$

$$Ratio\ S/P\ Cible = \ Ratio\ Combiné\ Cible - Frais\ variables\ \% - Frais\ fixe\ \%$$

3.2. Base de données

3.2.1. Traitement des données

Une base de données interne portant sur le portefeuille Transport des différents pays d'Europe est utilisée pour la suite de l'étude. Elle contient les caractéristiques d'exposition des clients, à laquelle a été fusionnée la base contenant leurs primes et sinistres.

3.2.1.1. Nettoyage des données

Dans un premier temps, un nettoyage des données est réalisé afin de supprimer les erreurs telles que des comptes ayant des sinistres mais pas de primes. Les observations des variables continues sont ensuite regroupées en classes afin d'éviter une trop grande volatilité avec des données aberrantes et pour obtenir des groupes contenant des observations homogènes. De la même manière, les valeurs manquantes ont été retraité lors du regroupement en classes, avec par exemple, les limites et les valeurs assurées manquantes reclassées en groupe 7.

3.2.1.2. Limites de l'étude

La précision de l'analyse est limitée par certains points pratiques. En effet, les profils de risques d'entreprises ne sont pas toujours comparables dans le portefeuille. Par exemple, les compagnies multinationales ont parfois des captives ou des programmes complexes contenant de la coassurance. Ces différentes structures ne sont pas toujours bien prises en compte par l'outil de tarification, les comptes sont de ce fait tarifés individuellement.

Par ailleurs, les polices annuelles ne permettent pas de suivre avec précision la destination de chaque expédition dans une année. Nous verrons par la suite que la variable destination n'est pas prise en compte dans l'étude, potentiellement due à ce manque de précision statistique. De la même manière seulement la marchandise principale est déclarée dans la base de données. L'impact de l'incertitude est atténué lorsque l'entreprise transporte des marchandises d'un même type, puisque les risques semblables sont rangés dans des classes aux taux identiques.

3.2.1.3. Développements des sinistres et des primes

Dans un second temps, les données de primes et de sinistres sont projetées à leurs valeurs ultimes à l'aide de triangles de liquidation afin de pouvoir comparer la performance finale des différents assurés.

La méthode Chain-Ladder est utilisée pour déterminer les facteurs de développements. Cette méthode est relativement simple à appliquer et permet de prendre en compte le développement négatif provenant des recours.

Les hypothèses sous-jacentes du modèle sont les suivantes :

- les années de survenance sont indépendantes entre elles
- la cadence de règlement dépend uniquement des années de développement

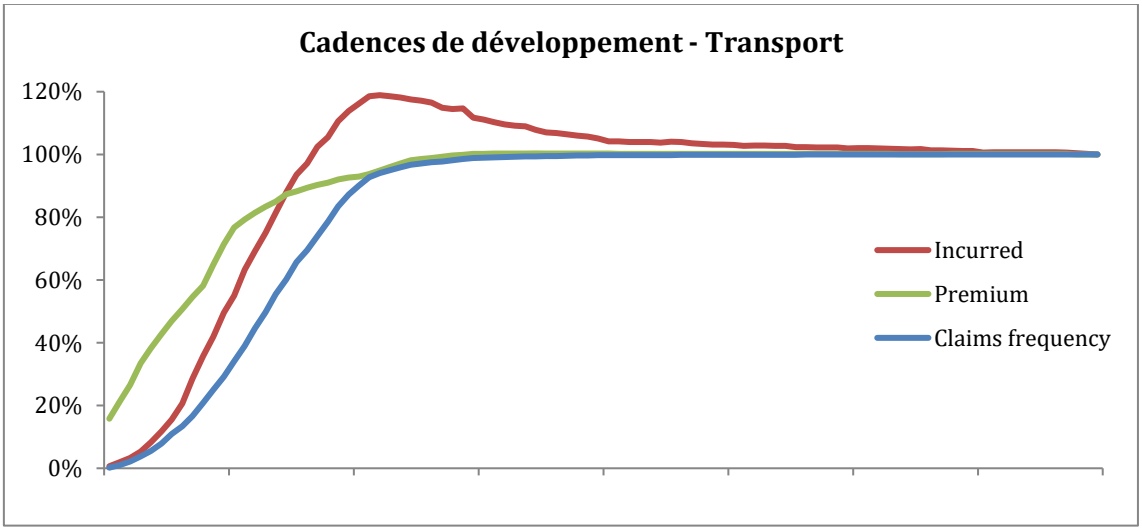
On note :

- Les années de survenance par la lettre i, $i \in \{1, \dots, I\}$
- Les années de développement par la lettre j, $j \in \{1, \dots, J\}$
- $X_{i,j}$ le montant incrémental réglé durant l'année de développement j pour l'année i
- $C_{i,j}$ désigne les règlements cumulés en j années de développement pour l'année i
- f_j désigne le facteur de développement de l'année de développement j

Les facteurs de développements :

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j}}{\sum_{i=1}^{I-j} C_{i,j+1}}$$

Figure 12 - Cadences de développement des sinistres et des primes du portefeuille



Un développement rapide des sinistres peut être observé, ainsi qu'un développement stable du nombre de sinistres au cours du temps, permettant de supposer un nombre faible de sinistres tardifs dits *IBNYR – Incurred But Not Yet Reported.* Par soucis de confidentialité, les triangles et les facteurs de développements du portefeuille Chubb ne seront pas affichés. Les triangles fournies par l'*International Union of Marine Insurance*, incluant la Belgique, la France, l'Allemagne, la Hollande, l'Italie, le Royaume-Uni et les Etats-Unis sont affichés pour comparaison.

Table 5 - Triangle cumulé de primes en Transport $ m - IUMI

Primes	1	2	3	4	5	6	7	8	9	10
2009	2,747	3,691	3,822	3,846	3,850	3,857	3,855	3,855	3,855	3,855
2010	2,486	3,463	3,631	3,666	3,689	3,689	3,686	3,686	3,686	-
2011	2,512	3,553	3,728	3,753	3,758	3,744	3,742	3,743	-	-
2012	2,674	3,643	3,795	3,807	3,817	3,816	3,814	-	-	-
2013	2,730	3,732	3,879	3,900	3,905	3,912	-	-	-	-
2014	2,468	3,389	3,492	3,509	3,520	-	-	-	-	-
2015	2,278	3,249	3,329	3,350	-	-	-	-	-	-
2016	2,227	3,221	3,347	-	-	-	-	-	-	-
2017	2,494	3,433	-	-	-	-	-	-	-	-
2018	2,478	-	-	-	-	-	-	-	-	-
Loss Development Factor	1.387	1.039	1.006	1.003	1.000	0.999	1.000	1.000	1.000	1.000
Factor to Ultimate	1.452	1.047	1.008	1.002	0.999	0.999	1.000	1.000	1.000	1.000
% Development	69%	96%	99%	100%	100%	100%	100%	100%	100%	100%

Table 6 - Triangle cumulé de sinistres en Transport $ m - IUMI

Sinistres	1	2	3	4	5	6	7	8	9	10
2009	1,762	2,500	2,533	2,546	2,485	2,467	2,468	2,471	2,458	2,453
2010	1,663	2,479	2,573	2,533	2,511	2,505	2,506	2,492	2,490	-
2011	1,821	2,616	2,710	2,666	2,661	2,629	2,622	2,619	-	-
2012	1,943	2,851	2,939	2,991	2,912	2,916	2,910	-	-	-
2013	1,978	2,811	2,969	2,879	2,904	2,871	-	-	-	-
2014	1,645	2,536	2,839	2,908	2,859	-	-	-	-	-
2015	1,624	2,731	3,023	2,934	-	-	-	-	-	-
2016	1,347	2,440	2,536	-	-	-	-	-	-	-
2017	1,692	2,490	-	-	-	-	-	-	-	-
2018	1,719	-	-	-	-	-	-	-	-	-
Loss Development Factor	1.515	1.055	0.993	0.988	0.994	0.999	0.998	0.997	0.998	1.000
Factor to Ultimate	1.548	1.021	0.968	0.974	0.985	0.992	0.993	0.995	0.998	1.000
% Development	65%	98%	103%	103%	101%	101%	101%	101%	100%	100%

Table 7 - Développement des Ratio S/P en Transport - IUMI

Ratio S/P	1	2	3	4	5	6	7	8	9	10
2009	64%	68%	66%	66%	65%	64%	64%	64%	64%	64%
2010	67%	72%	71%	69%	68%	68%	68%	68%	68%	
2011	72%	74%	73%	71%	71%	70%	70%	70%		
2012	73%	78%	77%	79%	76%	76%	76%			
2013	72%	75%	77%	74%	74%	73%				
2014	67%	75%	81%	83%	81%					
2015	71%	84%	91%	88%						
2016	60%	76%	76%							
2017	68%	73%								
2018	69%									

3.2.1.4. Catégorisation de la performance

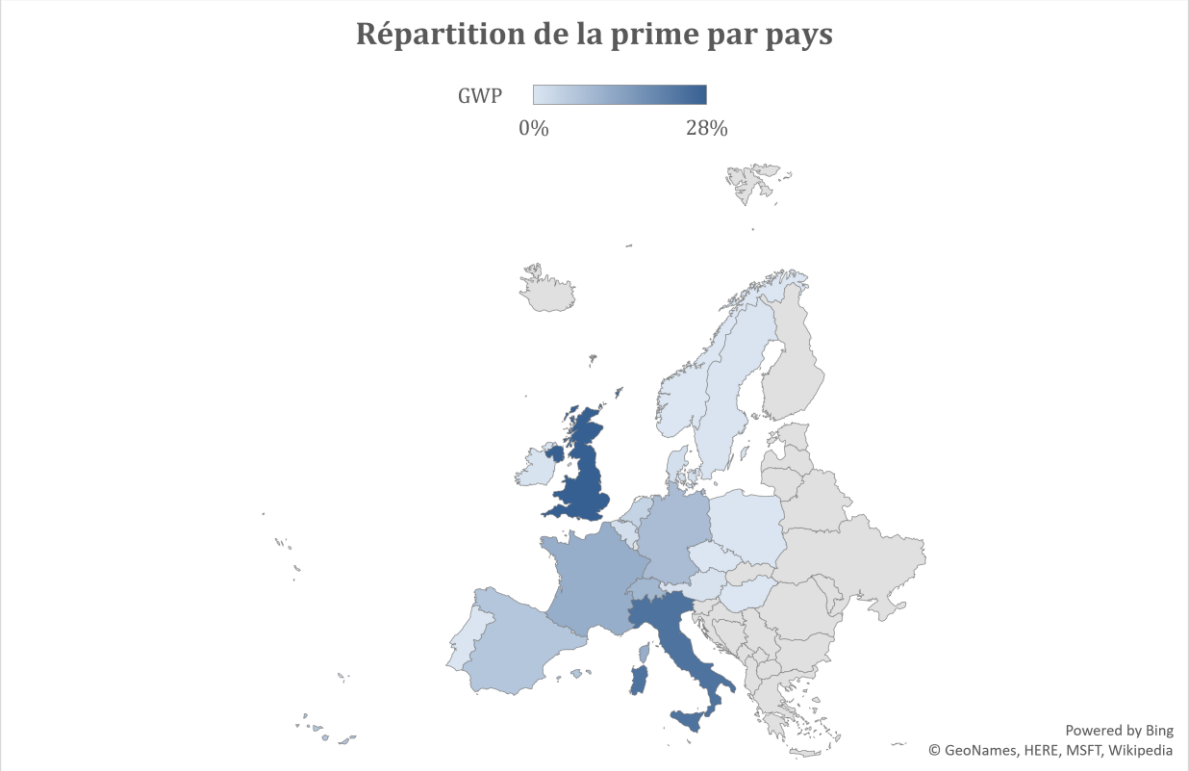
A l'aide du seuil de rentabilité présenté au chapitre précédent et qui sera calculé au chapitre 3.3.3. Profil de risque du portefeuille, les comptes seront segmentés de façon binaire suivant leur contribution à la rémunération des capitaux immobilisés.

3.2.2. Description de la base de données

Avec une base de données couvrant l'Europe, la difficulté opérationnelle réside dans le fait qu'il existe des pratiques et des conditions différentes selon le pays, rendant la comparaison des portefeuilles difficile compte tenu de l'hétérogénéité des données.

Bien que le modèle de tarification soit un outil mondial, prenant en compte un facteur différentiateur pour chaque pays, l'étude est réalisée dans la suite de ce mémoire au niveau d'un pays, pour éviter de rajouter de la dispersion résultant de la divergence des pratiques précédemment évoquées. Le choix de ce pays est fait compte tenu du nombre de risques présents et du volume de primes qu'il représente au sein du portefeuille Européen.

Figure 13 – Répartition géographique des volumes de prime du portefeuille Européen



En sélectionnant le portefeuille du pays, la base de données disponible contient désormais 24,006 lignes. Il faut noter qu'une base de données ne possédant pas assez d'observations, ne permettrait pas d'utiliser de manière convenable l'ensemble des modèles qui seront proposées à l'étude. Les méthodes permettant de générer des données de manière stochastique ne donnent pas toujours de résultats satisfaisants pour les modèles non linéaires. Il y a donc un risque de surapprentissage pouvant conduire à des conclusions biaisées si une ligne d'activité du portefeuille n'a pas une taille critique.

Les variables contenues dans notre base de données peuvent être rangées dans trois groupes : Les variables explicatives, à expliquer et les autres. Les colonnes n'ayant pas d'intérêts statistiques ne

seront pas considérées, telles que l'identifiant ou encore le nom du client etc. Les modalités des variables ci-dessous sont présentées à *l'Annexe 4 – Description de la base de données* avec les ordres de grandeurs de chaque variables.

Variables explicatives :

- **Commodity Class**: Classe de marchandises
- **Limit**: Limite d'assurance
- **Insured Value**: Valeur assurée
- **Coverage**: Type de conditions générales ex. ICC-A, ICC-B etc.
- **Stowage**: Type de conditionnement ex. Full Container, Breakbulk etc.
- **Voyage** : Zone géographique de destination
- **Conveyance**: Moyen de transport
- **Segment** : Le type de client ex. SME, Middle-Market, Global etc.

Variables à expliquer :

- **Premium**: Prime souscrite
- **Technical Premium** : Prime technique du modèle
- **Incurred**: Montant des sinistres
- **Attritional Incurred**: Montant des sinistres attritionnels
- **Large Incurred**: Montant des sinistres graves supérieurs à 500kUSD. Seuil défini par le département d'actuariat provisionnement pour chaque ligne de produit
- **Claims Count** : Nombre de sinistres
- **Ultimate Loss Ratio**: Ratio Sinistre sur Prime à l'ultime
- **Catégorie** : Catégorisation en fonction du seuil de rentabilité ex. profitable ou non

Autres variables :

- **Policy Holder ID** : Identifiant du client
- **Policy Holder Name** : Nom du client
- **Underwriting Year**: Année de souscription
- **Commissions**: Commissions
- **Administration Cost**: Frais d'administration
- **Unallocated Claims Expenses** : Coût de gestion des sinistres

3.2.3. Test d'indépendance χ^2

Afin de considérer uniquement les variables ayant un impact sur la catégorie du compte, soit « Profitable » ou « Non-Profitable », un tri est réalisé au sein des variables explicatives. Pour ce faire, un test d'indépendance du χ^2 avec la performance du compte est effectué.

$$\begin{cases} H_0 : \text{La variable est indépendante de la performance} \\ H_1 : \text{La variable est corrélée avec la performance} \end{cases}$$

Si la p-value ne dépasse pas le seuil de 5%, on rejette l'hypothèse nulle, on peut alors en déduire que la variable est corrélée avec la performance du compte.

Table 8 - Résultats des tests du χ^2

Variable	Statistique	Degré de liberté	p-value
Commodity	X-squared = 30.114	df = 10	p-value = 0.0008205
Limit	X-squared = 56.888	df = 6	p-value = 1.925e-10
Insured Value	X-squared = 48.974	df = 6	p-value = 7.546e-09
Voyage	X-squared = 0.40103	df = 2	p-value = 0.8183
Conveyance	X-squared = 5.4519	df = 2	p-value = 0.06548
Stowage	X-squared = 12.754	df = 6	p-value = 0.04711
Segment	X-squared = 13.131	df = 2	p-value = 0.001408
Coverage	X-squared = 43.143	df = 6	p-value = 1.093e-07

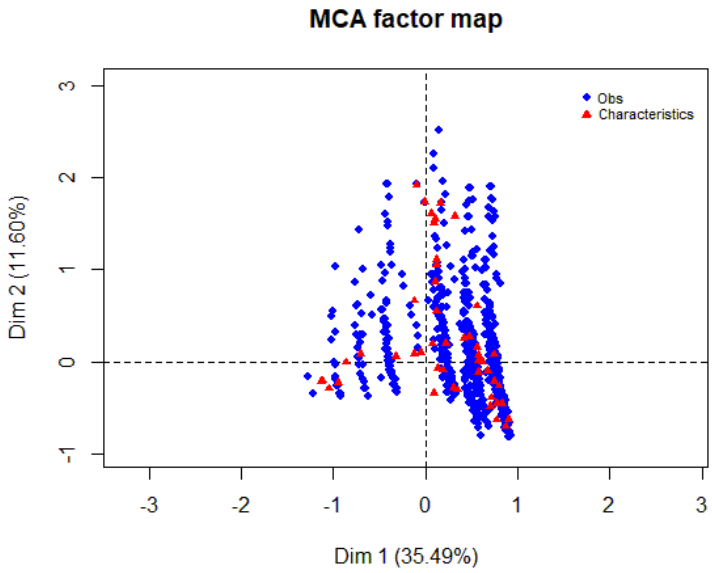
Les p-values sont inférieure au risque de première espèce $\alpha = 5\%$ sauf dans le cas de la variable « *Voyage* » et de la variable « *Conveyance* » qui semblent indépendantes de la performance du compte.

3.2.4. Visualisation des données

Avant toute étude, une analyse des correspondances multiples est réalisée afin de visualiser les caractéristiques du portefeuille. Les données sont regroupées dans un tableau disjonctif complet comportant I lignes pour les individus et J colonnes pour chacune des modalités que peuvent prendre les variables. En analysant les inerties sur une suite d'axes orthogonaux, une représentation factorielle des individus et des modalités est obtenue.

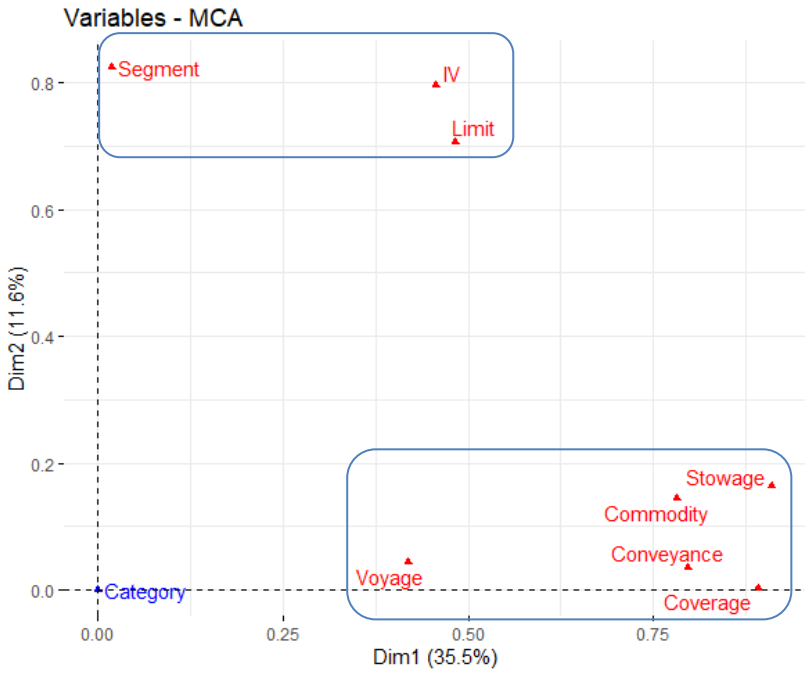
L'affichage du nuage de points permet de constater que la répartition des individus et des barycentres des modalités sur les deux premières dimensions ne présente pas d'anomalies particulières et est plus ou moins concentrée autour de l'origine.

Figure 14 - Répartition des observations en bleu et barycentres des modalités en rouge



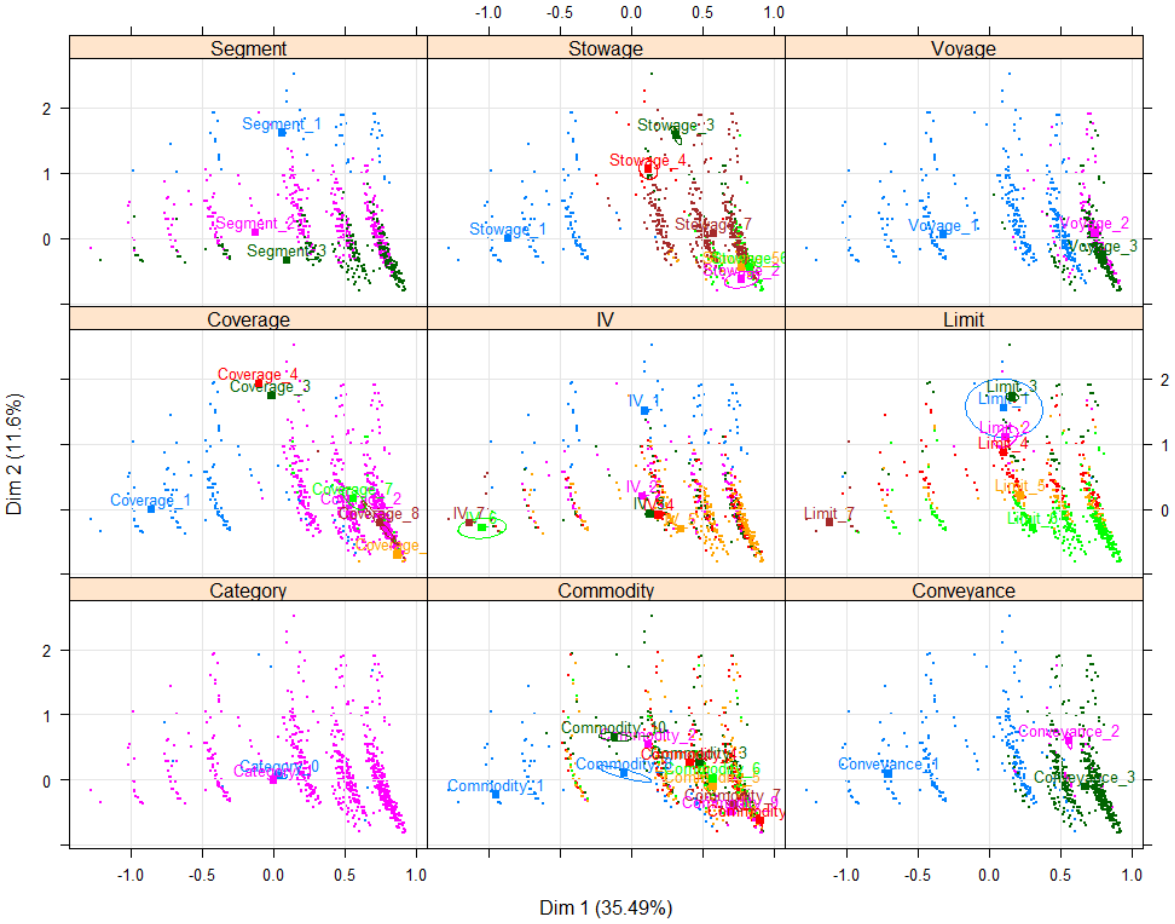
Les corrélations des variables sont analysées par la suite et deux groupes se distinguent pour l'explication des deux axes, composés d'une part, par le segment, la valeur assurée ainsi que la limite, et de l'autre part, des caractéristiques du compte.

Figure 15 - Analyse des corrélations



Les variables à l'intérieur de ces groupes sont corrélées ensemble et permettent d'interpréter les deux premières dimensions comme étant les caractéristiques du compte pour la dimension une et la taille du compte pour la dimension deux. La catégorie de profitabilité ne dépend pas d'un axe plus qu'un autre en particulier.

Figure 16 – Répartition des modalités des variables



Cette sortie permet de donner des informations très utiles sur le portefeuille en affichant les proportions pour chacune des modalités. De ce fait, le portefeuille est composé de très peu de clients multinationaux – *Segment 1* mais majoritairement de petites et moyennes entreprises – *Segment 3 et 2*. Le rangement par ordre croissant est par ailleurs respecté dans l'axe des ordonnées qui représente la dimension taille. Par ailleurs, conformément à l'analyse des corrélations, la profitabilité n'est à priori pas dépendante de ces deux axes et est quasi équirépartie dans le portefeuille en regardant la variable *category*.

Enfin, la visualisation individuelle par variable permet d'obtenir une idée du profil du portefeuille lors de la combinaison visuelle. Ainsi, une grande variabilité de profil des clients peut s'observer dans les petites et moyennes entreprises – respectivement les *segments 3* et 2. Les entreprises multinationales dans le *Segment 1* partagent généralement un profil similaire dans le type de contrat souscrit.

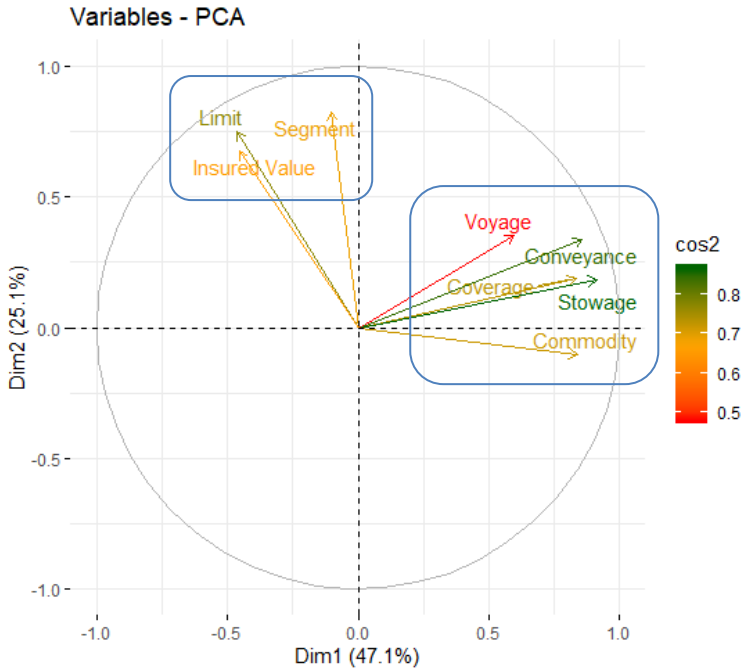
Par exemple, les multinationales – *Segment 1* sont généralement composées :

- de conteneurs pleins – *Stowage 7,*
- des destinations internationales – *Voyage 1,*
- des couvertures tout risques - *Coverage 1* et de couvertures standard Chubb – *Coverage 2*,
- de très grandes valeurs assurées – *IV 1*,
- d'une multitudes de tailles de limite d'assurance,
- de comptes profitables – *Category 1*,
- d'une variété de marchandises,
- de moyens de transport maritime – *Conveyance 1* et terrestre – *Conveyance 3*

3.2.5. Réduction de dimensions

La visualisation des données devient plus compliquée dès lors qu'un grand nombre de dimensions est à l'étude. Dans la suite des tests d'indépendance, la réduction de dimensions va permettre de se focaliser seulement sur les variables qui auront un impact sur la sinistralité des clients et entrainera par ailleurs une amélioration de la vitesse de calcul. Une ACP est réalisée afin de décrire la qualité de représentation des différentes variables. Contrairement à l'ACM qui a été réalisée auparavant, l'objectif n'est pas de savoir la qualité de représentation des modalités mais la qualité de représentation de toute la variable.

Figure 17 - Cercle des corrélations de l'Analyse en Composantes Principales



En observant le cercle des corrélations de l'ACP, deux groupes de variables se démarquent comme pour l'ACM. De par leur orthogonalité, cela indique une corrélation inter-groupe proche de zéro.

Les variables à l'intérieur de ces groupes montrent une corrélation proche de un, comme pour la limite et la valeur assurée. Au regard des variables, on conclut également que la première dimension est relative aux caractéristiques du programme d'assurance et la deuxième dimension est liée à la taille du risque.

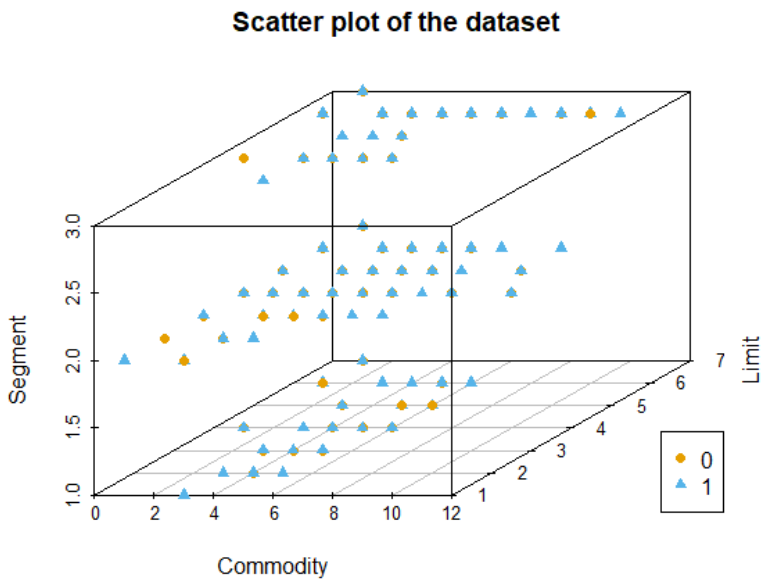
Par ailleurs, la qualité de représentation des variables sur ces deux axes, déterminée par le cosinus carré montre que la destination – *Voyage* est mal représentée ce qui est matérialisée par un faible cosinus carré en rouge et étant à la fois loin du cercle des corrélations. Au vu des résultats obtenues au test d'indépendance, la destination – *Voyage* n'est pas retenue car n'ayant que peu de variabilité dans les observations.

3.2.6. Sur-échantillonnage synthétique

Compte tenu de la bonne santé du portefeuille, il n'y a qu'un faible nombre de comptes en dessous du seuil de rentabilité fixé : 1,149 mauvais comptes sur 24,006 observations, soit 4.7%.

Les données du portefeuille sont ainsi fortement déséquilibrées, ce qui ne permettrait pas d'identifier correctement les caractéristiques des classes minoritaires avec certaines méthodes d'apprentissages. Dans le graphique ci-dessous, les points orange représentent les comptes non-profitables, présents dans une part minoritaire.

Figure 18 - Répartition des comptes selon la limite, le segment et la marchandise

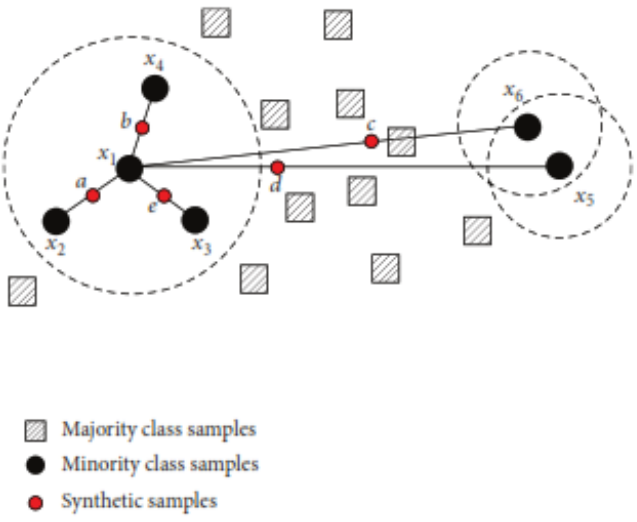


Tremblay [2017] décrit les avantages et inconvénients de différentes méthodes de rééchantillonnage dont celles proposées par Japkowicz [2000], et Chawla et. Al [2002] pour rééquilibrer une base de données.

Le *SMOTE*, *Synthetic Minority Over Sampling Technique*, qui est une méthode de sur-échantillonnage synthétique de la classe minoritaire, est choisi car étant adapté dans les cas multi-dimensionnels et présentant l'avantage d'éviter le surapprentissage lié à la simple réplication des observations. La méthode créée ainsi des observations synthétiques pour les classes minoritaires en faisant appel au *bootstrap* et aux *k-Nearest Neighbors*.

En fonction du nombre de plus proche voisin choisi, l'algorithme va choisir des groupes d'observations et créer des observations « artificielles » en interpolation linéaire entre deux observations en minorité. De cette manière, il n'y a pas de duplication d'observations et donc de surapprentissage, mais un rééquilibrage des poids de ces groupes dans les différents algorithmes de classification et de clustering. Le *SMOTE* va en revanche, aussi réduire le nombre d'observations prépondérante ce qui peut conduire dans certain cas à une perte d'information. Hu & Li [2013] proposent une étude comparant les performances de méthodes alternatives telles que l'*ASMOTE*, le *Borderline-SMOTE*, *SMOTE-RSB* et le *NRSBoundary-SMOTE* sur différents algorithmes comme le *CART*, les *KNN* ou les *SVM*. Bien que ces méthodes présentent de meilleurs résultats en général, elles ne sont pas facilement disponibles sur le logiciel R. Compte tenu de l'amélioration marginale des autres méthodes présentée dans l'étude de Hu & Li et que le profil de la catégorie prépondérante soit homogène, le choix d'utiliser le *SMOTE* est fait pour faciliter l'implémentation.

Figure 19 - Schéma du fonctionnement du SMOTE – Hu & Li [2013]



Afin de fonctionner, l'algorithme a besoin de plusieurs paramètres :

- Perc.over : Le pourcentage de suréchantillonnage synthétique pour la classe minoritaire
- Perc.under : Le pourcentage de sous-échantillonnage pour la classe majoritaire
- k : Le nombre de plus proches voisins pour la création d'observations

Comme le montre Tremblay [2017], pour déterminer ces paramètres, on spécifie la notation de :

- o : Le pourcentage de suréchantillonnage synthétique pour la classe minoritaire
- u : Le pourcentage de sous-échantillonnage pour la classe majoritaire
- n : L'effectif initial de la classe minoritaire
- x : Le pourcentage de la classe minoritaire voulue
- N : L'effectif total après le suréchantillonnage

L'algorithme va générer pour chaque observation de la classe minoritaire, $o + 1$ observations synthétiques et enlève $u \cdot o$ observations de la classe majoritaire.

En résolvant les équations suivantes :

$$x = \frac{o + 1}{o + 1 + u \cdot o}$$

$$N = n(o + 1 + u \cdot o)$$

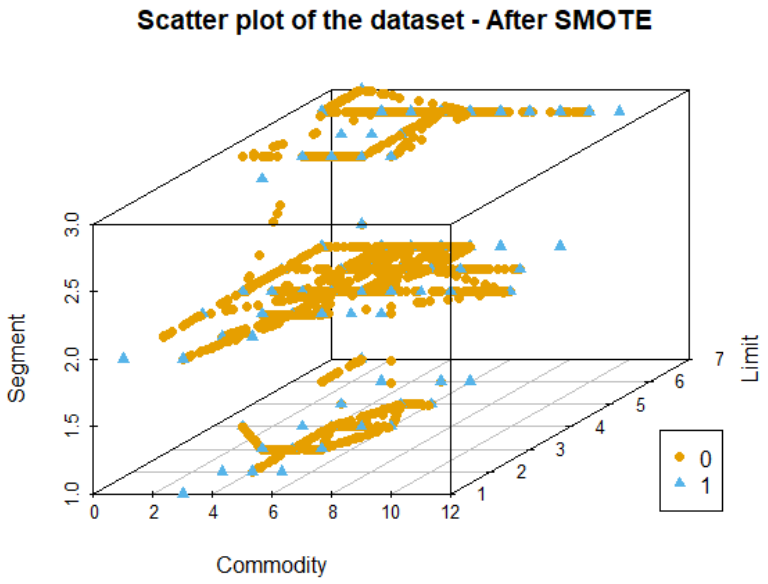
Pour une base équilibrée, on choisit pour paramètres :

$$o = \frac{N_x}{n} = \frac{25000}{1149}$$

et

$$u = \frac{1 - x}{x - \frac{n}{N}} \cdot 100 = \frac{1 - 50\%}{50\% - \frac{1149}{25000}} \cdot 100$$

Figure 20 - Répartition des comptes selon la limite, le segment et la marchandise après SMOTE



Une nouvelle base totalisant 51,849 observations est obtenue, avec désormais une part de 49% pour les comptes non-profitables (0) et 51% pour les comptes profitables (1). Cette première visualisation ne permet pas de tenir compte de la taille des groupes, car les observations se superposent entre elles. Les observations synthétiques générées par le *SMOTE* sont visibles entre les différentes modalités des variables dans le graphique ci-dessus.

3.3. Profil de risque du portefeuille

Dans un but de pilotage, le risque de prime est quantifié pour l'ensemble du portefeuille européen. La considération du portefeuille d'un seul pays peut en effet, entrainer de la volatilité dans la modélisation et ne permet pas de prendre en compte la diversification géographique des risques. Seuls les risques d'assurance sont pris en compte, puisque, suivant la duration de la branche, les risques d'investissements peuvent être fondamentalement différents. La mesure de risque interne à Chubb nommée *Expected Shortfall* $\mathbb{ES}_{99.5\%}$ décrite au chapitre 2.2.3 est retenue.

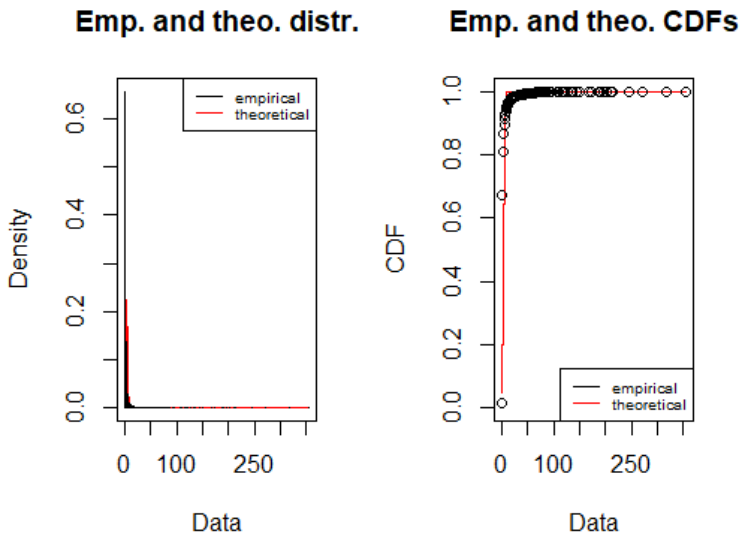
3.3.1. Modélisation de la fréquence

Pour la modélisation de la fréquence des sinistres graves, les lois de Poisson, Binomiale Négative et Géométrique sont choisies parmi une série de lois discrètes. De manière visuelle, les lois Binomiale Négative et Géométrique montrent une fonction de densité de probabilité plus adéquate que la loi de Poisson à *l'Annexe 5 – Graphiques – Fréquence des sinistres graves.*

En consultant les résultats des tests à *l'Annexe 6 – Résultats tests – Fréquence des sinistres graves* la loi Binomiale Négative possède le critère d'information d'Akaike et le critère d'information Bayésien les plus faibles. Ces critères étant des mesures de la qualité d'un modèle statistique, ils permettent de quantifier la quantité d'information perdue lors de la modélisation.

	Estimate	***Std Error***
m	0.90	0.01
p	3.01	0.02

Figure 21 - Loi Binomiale Négative



3.3.2. Modélisation de la sévérité

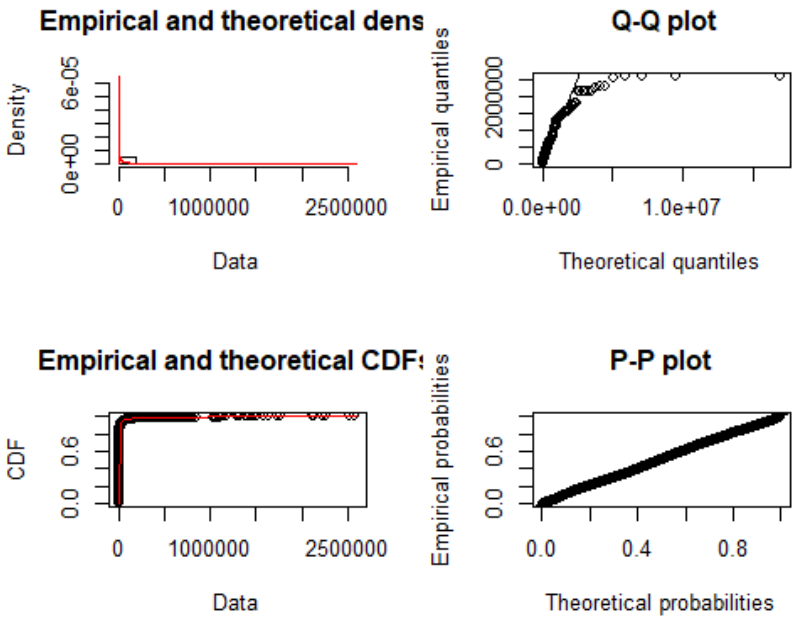
3.3.2.1. Sinistres attritionnels

La sévérité des sinistres attritionnels et la sévérité des sinistres graves supérieurs à un seuil de 500,000 USD seront modélisées séparément afin de tenir compte de la queue de distribution. Ce seuil étant défini par le département provisionnement pour chaque ligne de produit, l'adéquation du seuil ne sera pas étudiée dans ces travaux. Une série de lois continues strictement définies positives est testée : Parmi elles, la loi log-normale, la loi de Pareto, la loi de Weibull et la loi Gamma. La densité de probabilité et le P-P plot sont plus ou moins bien ajustés.

Les différents tests à l'*Annexe 8 – Résultats tests – Sévérité des sinistres attritionnels* montrent que le test de Kolmogorov-Smirnov ne permet pas de départager les quatres lois mais que le test de Cramer-von Mises affiche une distance moins élevée dans l'ajustement des densités de probabilités. La loi log-normale est donc choisie pour la modélisation des sinistres attritionnels avec les paramètres suivants :

	Estimate	***Std Error***
μ	7.20	0.01
σ	2.27	0.01

Figure 22 - Loi Log-Normale



3.3.2.2. Sinistres graves

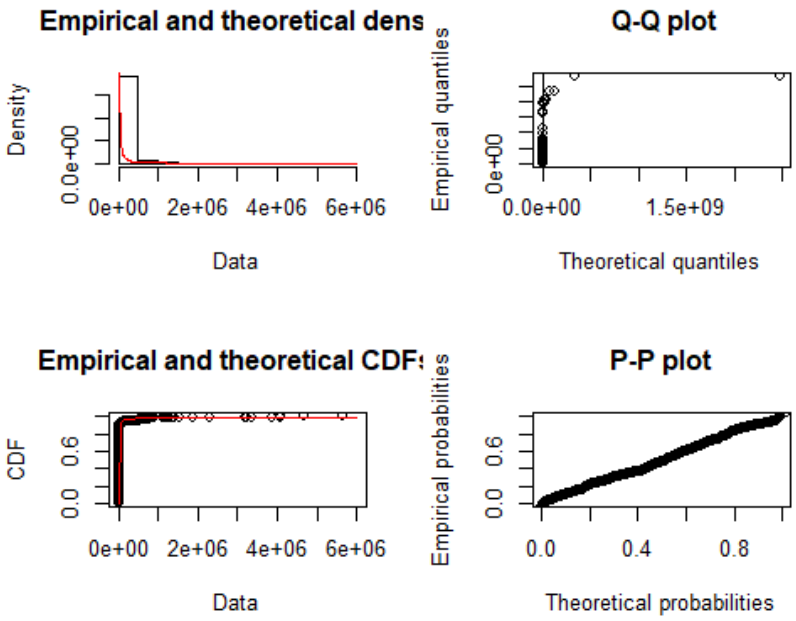
Tout comme pour la sévérité des sinistres attritionnels, les lois précédentes seront essayées mais également quelques lois additionnelles à queue épaisse.

Les tests de Kolmogorov-Smirnov et Cramer-von Mises disponibles à l'*Annexe 10 – Résultats tests – Sévérité des sinistres graves*, montrent que la loi de Burr obtient les meilleurs résultats pour la modélisation des sinistres larges.

Les paramètres de la loi de Burr sont :

	Estimate	***Std Error***
x	0.00	0.00
c	0.59	0.02
k	0.92	0.02

Figure 23 - Loi de Burr



3.3.3. Profil de risque du portefeuille

Table 9 - Profil de risque du portefeuille Européen

Percentile	Return Period (Years)	Premium ($m)	Incurred ($m)	Loss Ratio
Plan		164	78	48%
99.5th	**200**		124	76%
99.8th	**500**		763	466%

Avec un volume de prime planifié de USD 164m, le modèle de fréquence-sévérité donne un SCR_{Marine} de USD 46m pour couvrir le risque d'une mauvaise tarification de notre portefeuille.

Table 10 - Shortfall

Premium ($m)	Loss Ratio	VaR 99.5th	% Shortfall	$m Shortfall
164	48%	76%	28%	46

3.3.4. Seuil de rentabilité du portefeuille

Le seuil de rentabilité du portefeuille est calculé en tenant compte du profil de risque précédemment déterminé et en prenant les hypothèses fictives suivantes :

- Un coût du capital de 15% avant impôt
- Un taux d'imposition moyen de 18.5%
- Des frais variables de 4.0%
- Des frais fixes de USD 56m
- Une courbe des taux sans risques provenant de la directive Solvabilité 2 établie par EIOPA au 31.12.2018 présent à l'*Annexe 11 – Risk* Free Rate EIOPA

Table 11 - Actualisation des Cash-Flows

Years	RFR	Discounted Loss (USD m)
1	0.91%	38
2	1.49%	49
3	2.16%	-2
4	2.60%	-5
5	2.89%	-1
6	3.08%	-1
7	3.22%	-1
8	3.32%	-1
Total		78

En appliquant les formules précédemment exposés dans le chapitre *Détermination du seuil de* rentabilité, les calculs donnent :

$$Prime\ cible = \frac{((46 \times 15\ \%)/(1 - 18.5\%)) + 78 + 56}{(1 - 4\ \%)}$$

$$Prime\ cible = \ 148$$

$$Ratio\ Combiné\ Cible = \frac{78 + 56}{148} + 4\ \%$$

$$Ratio\ Combiné\ Cible = \ 95\ \%$$

Compte tenu du montant de frais fixes exposés précédemment qui s'élèvent à 34%, le seuil de rentabilité requis pour répondre aux attentes des actionnaires est déterminé pour un ratio S/P de 57%. Celui-ci sera utilisé pour classer un compte comme étant bon ou mauvais dans la suite de l'étude de pilotage de portefeuille pour les algorithmes

$$Ratio\ S/P\ Cible = \ 95\% - 4\% - 34\%$$

$$Ratio\ S/P\ Cible = \ 57\%$$

3.4. Segmentation de portefeuille

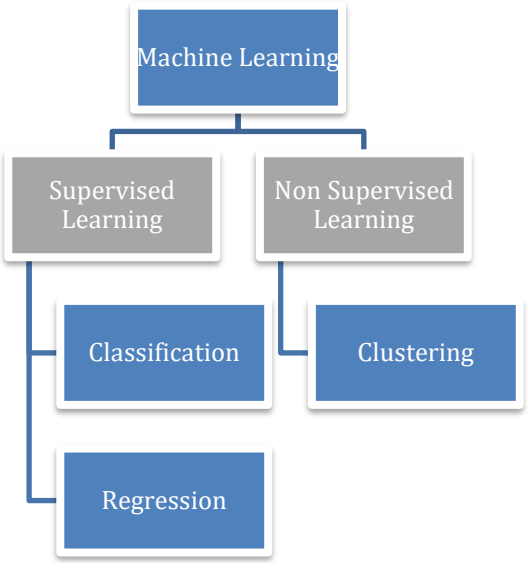
3.4.1. Mise en œuvre

L'étude d'un portefeuille de risques d'entreprise peut s'avérer complexe comparée à des risques de masse où la tarification suit un modèle bien précis. Outre le fait d'avoir moins de données disponibles, le portefeuille sera bien souvent composé de risques hétérogènes en fonction de la taille et de l'industrie du client. Il sera toutefois intéressant de voir les résultats que les méthodes d'apprentissages peuvent apporter, tout en gardant à l'esprit les limitations dans l'élaboration des conclusions de l'étude.

Durant ces récentes années, des études sont apparues dans la littérature technique proposant l'utilisation de méthodes d'apprentissages pour la tarification. Dans la pratique, il est encore rare de voir apparaitre ces méthodes utilisées à des fins de tarification. Cela est principalement dû à la difficulté d'interprétation de certains algorithmes et à l'absence d'explications sur la corrélation des facteurs qui sont à contrario disponibles dans les modèles linéaires. Il existe parmi les différents algorithmes, des méthodes dites d'apprentissages supervisés et des méthodes dites d'apprentissages non supervisés :

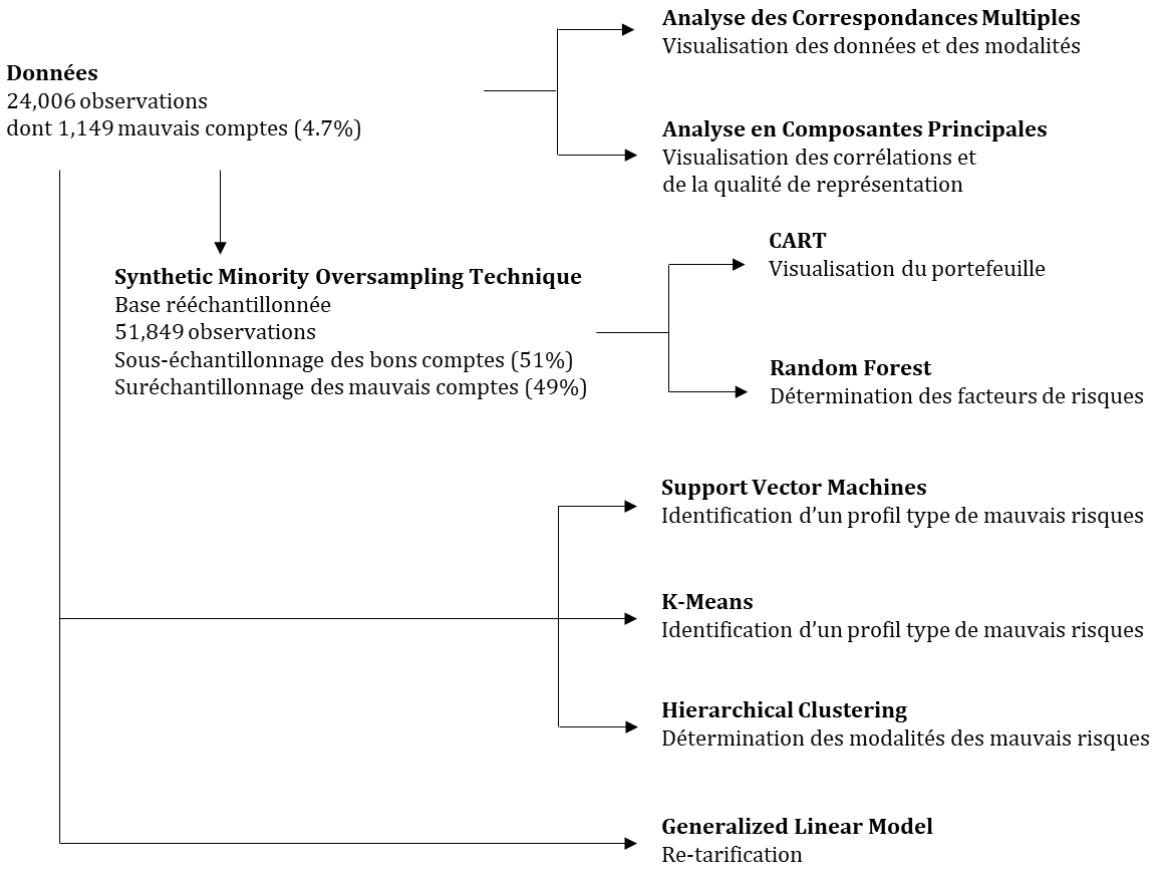
Dans le cas supervisé, une base de données est utilisée pour apprendre un résultat au modèle afin de classer ou d'estimer un résultat en régression. Dans le cas non-supervisé, l'algorithme va essayer de déterminer des groupes en fonction de la ressemblance des données. Les algorithmes permettant une interprétation aisée des résultats seront privilégiés dans la liste disponible.

Figure 24 - Type d'algorithmes



Le travail préliminaire d'analyse et de nettoyage des données a permis de mettre en lumière la nécessité d'appliquer un algorithme de suréchantillonnage synthétique pour rééquilibrer la base. L'identification des caractéristiques des comptes non-profitables par les arbres de décision sera plus aisée comme le décrit Japkowicz [2000]. Les algorithmes suivants seront appliqués pour répondre à un objectif spécifique :

Figure 25 - Détail et objectif des algorithmes utilisés



3.4.2. Arbre de décisions

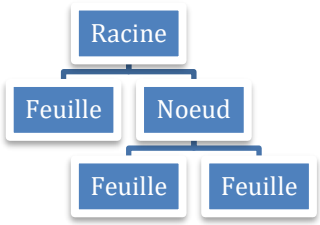
Afin de construire un modèle de tarification par exposition, il est impératif de posséder un grand nombre de données pour obtenir un modèle précis et robuste. Les facteurs peuvent cependant ne pas être bien calibrés lorsque les données ne sont pas représentatives des comptes à tarifer. Il est donc important d'avoir un processus de vérification des taux.

3.4.2.1. Principes des arbres de décisions

L'arbre de classification et de régression - CART, permet de segmenter un jeu de données. Il présente l'avantage d'être facilement lisible.

Un arbre de décision se présente de la manière suivante :

Figure 26 - Fonctionnement de l'algorithme CART



Pour construire un arbre CART, on part de la racine de l'arbre, qui contient toutes les observations de la base, ici les polices. La première étape de CART consiste à découper de façon binaire la racine en deux sous-groupes appelés nœuds fils. En classification, la segmentation est réalisée à l'aide de l'indice de Gini qui définit l'impureté et en régression, on cherchera à minimiser la variance intra-groupe.

L'indice de Gini compris entre 0 et 1, est une mesure de la dispersion d'une distribution. Plus le coefficient est proche de 1 et plus le nœud est hétérogène. L'arbre va ainsi maximiser l'indice et séparer les nœuds dits impurs, en deux nœuds fils qui seront plus homogènes.

L'indice de Gini d'un nœud t est défini par :

$$\Phi(t) = \sum_{c=1}^{L} \hat{p}_t^c (1 - \hat{p}_t^c)$$

où $\hat{p}_t^c$ est la proportion d'observation de la classe c dans le nœud t. Le découpage en deux nœuds fils t_L et t_R est réalisé si la valeur $\hat{\Delta}$ correspondant à la différence entre la pureté du nœud père et la pureté des nœuds fils est maximisée.

$$\hat{\Delta} = \Phi(t) - [\Phi(t_L) + \Phi(t_R)]$$

En régression, on cherche à diminuer la variance dans les nœuds obtenus. La variance d'un nœud t étant définie par :

$$V(t) = \frac{1}{n} \sum_{y_i \in t} (y_i - \bar{y}_t)^2$$

où $\bar{y}_t$ est la moyenne des y_i observations présentes dans le nœud *t*. On cherche à minimiser la somme des variances des deux nœuds fils t_L et t_R : min $V(t_L) + V(t_R)$.

L'arbre est ainsi développé jusqu'à atteindre une condition d'arrêt, lorsqu'il est homogène ou lorsqu'il n'existe plus de partition possible. Il est possible de fixer un seuil pour ne pas découper les nœuds contenant moins d'un certain nombre d'observations. Les arbres de décisions permettent ici d'identifier les variables les plus discriminantes d'un jeu de données, en fonction de leur présence le long des nœuds.

Dans notre cas, le portefeuille sera partitionné afin de mettre en exergue les variables ayant une incidence sur la variable explicative. Attention cependant, l'arbre de décision n'est pas un outil prédictif adapté dans notre cas. En effet, en présence d'un portefeuille non diversifié et donc trop homogène, il n'y aura pas de découpages des nœuds. Il sera avant tout utilisé comme un outil de contrôle et de visualisation des données.

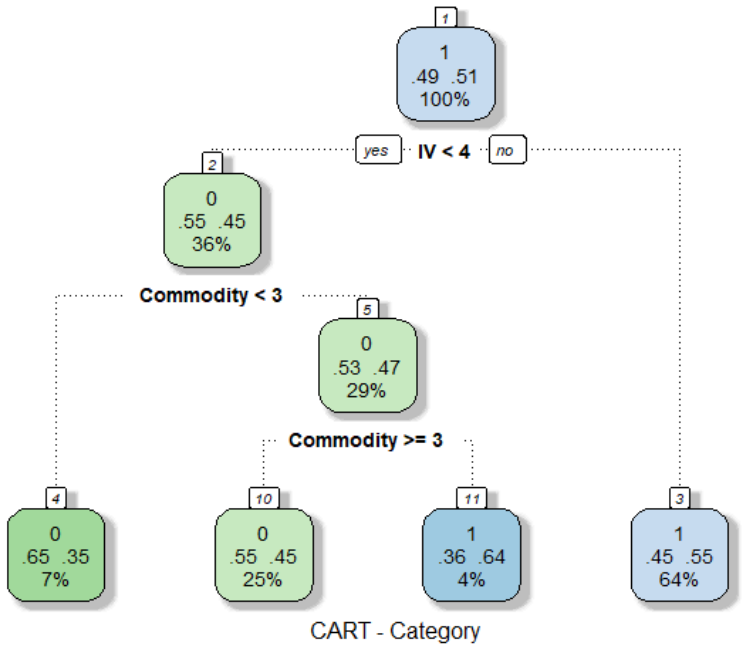
3.4.2.2. Application des arbres de décisions

Afin d'avoir une première idée de la segmentation du portefeuille, différentes variables d'intérêts seront étudiées. Comme il n'existe pas assez d'hétérogénéité des ratios S/P, le découpage n'est pas possible. La catégorie de profitabilité, permet de remplacer implicitement les ratios S/P. Seront également expliquées, la prime technique du modèle de tarification, la prime effectivement souscrite et la sinistralité.

Les informations affichées à l'intérieur des bulles de la sortie graphique varient en fonction du caractère qualitatif ou quantitatif de la variable analysée :

- La première ligne correspond à la modalité la plus représentée dans le nœud ou à une moyenne lorsqu'on est en présence d'une variable quantitative.
- La deuxième ligne correspond aux pourcentages de représentation de chaque modalité dans le nœud dans le cas de variables qualitatives. Cette information n'est pas présente dans le cas contraire.
- Enfin la troisième ligne correspond à la proportion que représente chaque nœud par rapport à la base complète. Cette information est accompagnée du nombre d'observations dans le cas de variables quantitatives.

Figure 27 - CART Profitabilité



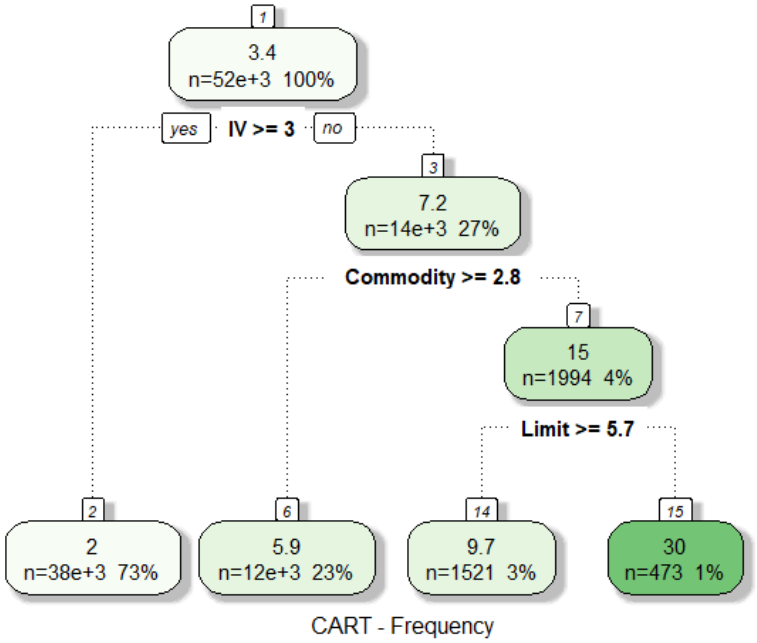

```
1) root 51849 25278 1 (0.4875311 0.5124689)
   2) IV< 3.999367 18698  8394 0 (0.5510750 0.4489250)
     4) Commodity< 2.990334 3684  1298 0 (0.6476656 0.3523344) *
     5) Commodity>=2.990334 15014  7096 0 (0.5273745 0.4726255)
      10) Commodity>=3.001368 13147  5892 0 (0.5518369 0.4481631) *
      11) Commodity< 3.001368 1867   663 1 (0.3551152 0.6448848) *
   3) IV>=3.999367 33151 14974 1 (0.4516907 0.5483093) *
```

Etant en présence d'une variable qualitative prenant les valeurs 0 et 1, la sortie affiche trois lignes. En prenant le premier nœud pour exemple, la première ligne correspond à la valeur majoritaire dans le nœud, à savoir les comptes notés 1 - Profitable. La deuxième ligne est la répartition des modalités, c'est à dire : 49% de 0 - Non-profitable et 51% de 1 - Profitable. Ce quasi-équilibre est une résultante de l'algorithme de sur-échantillonnage SMOTE. La troisième ligne est la proportion que représente le nœud dans la base totale.

En analysant les résultats de l'arbre sur la profitabilité, les comptes avec de grandes valeurs assurées, inférieures à 4, sont en majorité déficitaires. En descendant l'arbre, parmi ces clients, ceux transportant des marchandises risquées, supérieur à 3, sont profitables. La bonne performance de ces grands comptes est probablement due à une attention accrue dans la tarification de ces risques et à de meilleurs dispositifs de gestion des risques.

Du fait du *SMOTE*, les modalités peuvent être des nombres non entiers, ce qui perturbe la lecture de l'arbre. Une incohérence sur la marchandise peut être lue, avec un double découpage au niveau du groupe de marchandise 3, ce qui est dû en fait à un problème d'arrondi. En regardant la sortie brute, le découpage se fait en réalité au niveau 2.99 et 3.00.

Figure 28 - CART Fréquence de sinistres



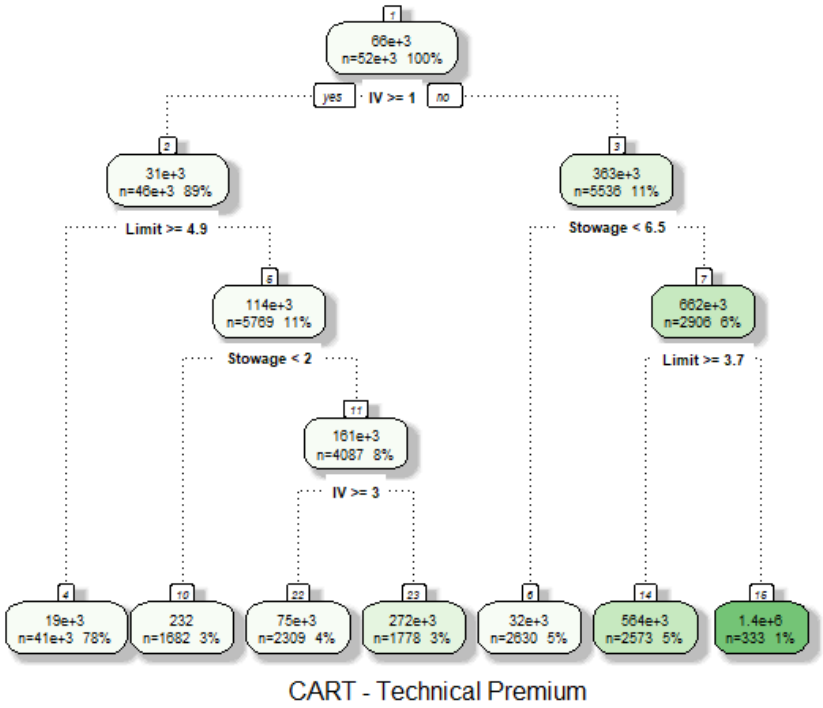

```
1) root 51849 6707398  3.407280
   2) IV>=2.999953 37952 1299227  2.024706 *
   3) IV< 2.999953 13897 5137507  7.183018
     6) Commodity>=2.833049 11903 2452930  5.948325 *
     7) Commodity< 2.833049 1994 2558112 14.553400
      14) Limit>=5.746375 1521 1323999  9.656512 *
      15) Limit< 5.746375 473 1080357 30.300060 *
```

Etant en présence d'une variable quantitative, les bulles contiennent seulement deux lignes. En prenant le premier nœud comme exemple, la première ligne indique la valeur moyenne du nombre de sinistres. La deuxième donne le nombre d'observation et la taille du nœud dans la base totale.

En analysant le nombre de sinistres, ce sont principalement les clients avec de grandes valeurs assurées et des limites d'assurances hautes qui possèdent une fréquence élevée de sinistres. Les clients ayant des profils similaires sont souvent des grands comptes multinationaux avec de nombreuses filiales incluses dans le programme d'assurance.

Afin de comprendre si la performance des comptes non rentables est due à un mauvais calibrage de l'outil de tarification ou à une non-adéquation de la prime au risque, il est nécessaire d'obtenir plus de détails. Pour ce faire, seront analysés la prime effectivement souscrite, la prime technique provenant du modèle existant et la sinistralité.

Figure 29 - CART Prime technique



CART - Technical Premium

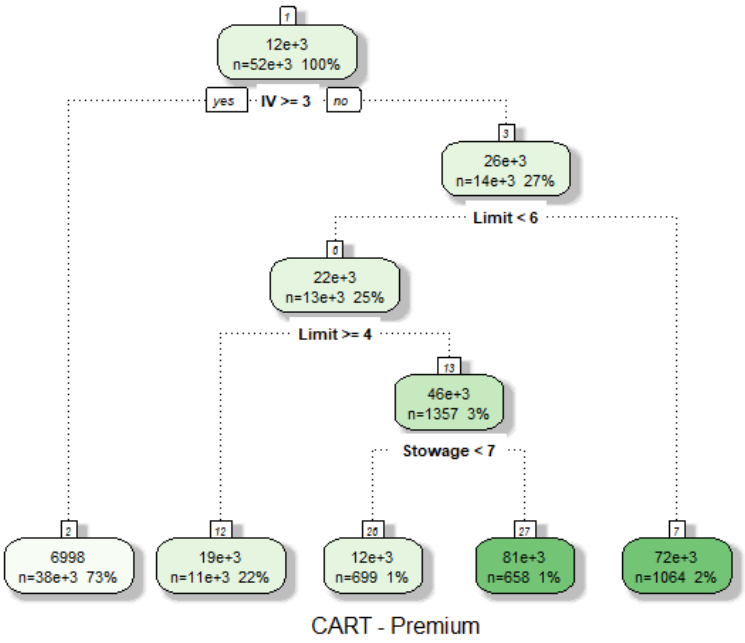
```
1) root 51849 2.775396e+15   66016.920
   2) IV>=1.001085 46313 3.431845e+14   30565.510
     4) Limit>=4.934418 40544 1.002327e+14   18675.750 *
     5) Limit< 4.934418 5769 1.969393e+14  114125.700
      10) Stowage< 2.036229 1682 2.276044e+10     232.165 *
      11) Stowage>=2.036229 4087 1.661188e+14  160998.400
        22) IV>=2.999634 2309 2.156839e+13   75241.650 *
        23) IV< 2.999634 1778 1.055173e+14  272366.400 *
   3) IV< 1.001085 5536 1.887064e+15  362595.900
     6) Stowage< 6.5 2630 3.295354e+13   31780.940 *
     7) Stowage>=6.5 2906 1.305800e+15  661991.400
      14) Limit>=3.699442 2573 2.305901e+14  563938.700 *
      15) Limit< 3.699442 333 8.593312e+14 1419618.000 *
```

L'arbre montre ici que la prime technique est issue d'un modèle plus complexe faisant appel à plus de variables :

- La valeur assurée
- Le mode de conditionnement
- La limite d'assurance

Le modèle discrimine d'un côté les très grandes valeurs assurées et les autres. Selon le modèle, la limite permet également de jouer un rôle dans la limitation de la sinistralité. De même, le mode de conditionnement des marchandises semble influencer la sinistralité des deux côtés.

Figure 30 - CART Prime souscrite

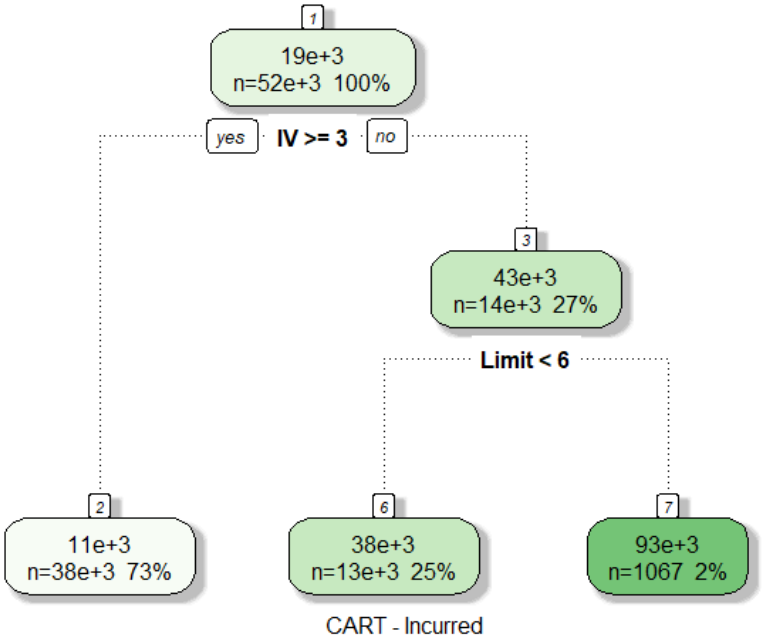



```
1) root 51849 8.193334e+13 12081.310
   2) IV>=2.998347 37956 7.341824e+12  6998.223 *
   3) IV< 2.998347 13893 7.093153e+13 25968.430
     6) Limit< 6.020139 12829 3.762600e+13 22179.400
      12) Limit>=3.971082 11472 1.793380e+13 19398.020 *
      13) Limit< 3.971082 1357 1.885317e+13 45693.070
        26) Stowage< 6.989044 699 1.998056e+11 12244.330 *
        27) Stowage>=6.989044 658 1.704053e+13 81225.990 *
     7) Limit>=6.020139 1064 3.090060e+13 71653.970 *
```

En utilisant l'arbre de classification sur la prime souscrite, les facteurs influant sur la prime sont similaires au modèle de tarification mais la différentiation de la prime est nettement moins complexe. Quelques explications pouvant mener à ce constat :

- Il n'y a pas de dispersion significative dans les risques du portefeuille pouvant mener à un découpage plus fin par l'arbre.
- Le souscripteur ne suit peut-être pas le modèle de tarification pour la détermination de la prime. Il y a donc des décisions de souscription pour réduire ou simplifier le mode de calcul.
- Il peut y avoir une inadéquation du modèle par rapport au marché, puisque la prime effectivement souscrite n'est pas en accord avec le modèle de tarification. En effet, la prime moyenne au niveau des différents nœuds est beaucoup plus faible que le précédent CART sur la prime technique. La logique est cependant respectée ici puisque les risques possédant une grande valeur assurée et une grande limite se voient attribuer une prime élevée.

Figure 31 - CART Sinistres




```
1) root 51849 2.372765e+14 19300.43
  2) IV>=2.998347 37956 3.216097e+13 10748.90 *
  3) IV< 2.998347 13893 1.947567e+14 42663.43
    6) Limit< 6.006564 12826 1.414323e+14 38448.42 *
    7) Limit>=6.006564 1067 5.035733e+13 93330.47 *
```

En utilisant l'arbre sur les sinistres, le découpage diffère du modèle de tarification car le type de conditionnement n'a pas d'influence sur le niveau de sinistralité. En comparant avec la prime souscrite, les deux arbres sont relativement similaires et suivent la même logique, puisque les grandes valeurs assurées et les grandes limites possèdent une sinistralité plus grande. Il semble cependant que le mode de conditionnement est pris en compte par les souscripteurs lorsqu'ils déterminent la prime mais que cette caractéristique n'a pas d'impact sur la sinistralité.

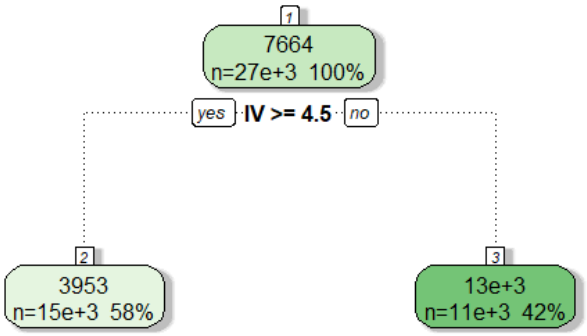
3.4.2.3. Conclusion intermédiaire

Le modèle de tarification est complexe et apparait inadapté. Il peut parfois donner des primes supérieures au marché et apporte une différentiation non nécessaire pour des risques similaires. Cela est probablement dû à la volonté lors de la construction du modèle, que chaque changement de caractéristique implique un changement de prix. Le modèle de tarification cherche également à discriminer un nombre de profils différents pour chaque pays ce qui peut entrainer de la volatilité contrairement à ce portefeuille homogène qui nécessite peu de facteurs explicatifs.

Il est intéressant de regarder par la suite si les comptes des sous-groupes « profitables » et « non-profitables » qui ont été déterminés à l'aide du seuil de rentabilité, partagent des similarités avec le portefeuille global. Autrement dit, ont-ils des caractéristiques propres expliquant leur performance.

3.4.2.4. CART sous-groupe « Profitables »

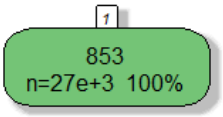
Figure 32 - CART Groupe « Profitables » – Prime Souscrite



CART - Premium - Good

```
1) root 26571 1.428337e+13  7785.612
  2) IV>=4.5 15247 8.155262e+11  3854.688 *
  3) IV< 4.5 11324 1.291502e+13 13078.340 *
```

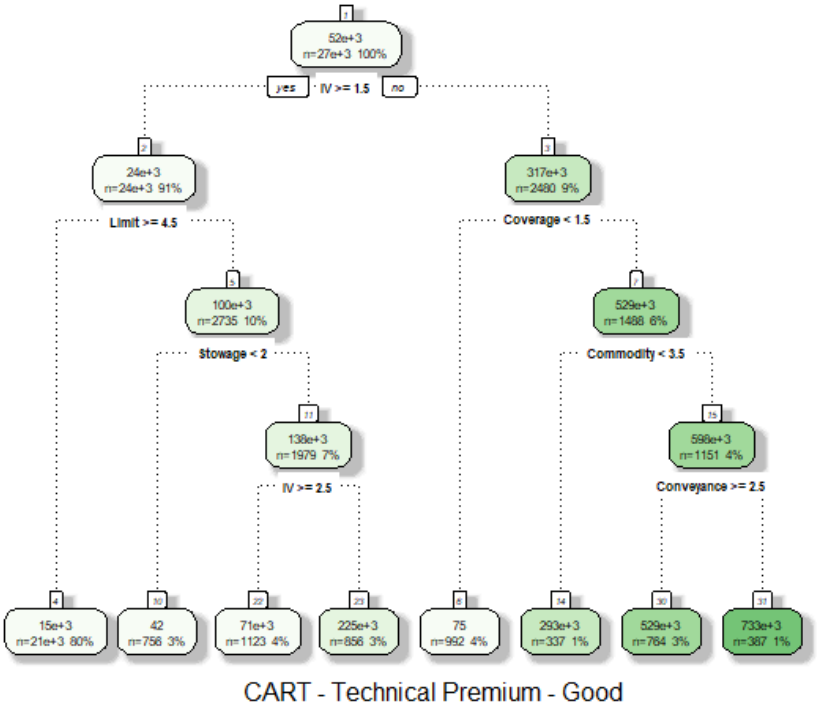
Figure 33 - CART Groupe « Profitables » - Sinistres



CART - Incurred - Good

```
1) root 26571 1.257346e+12 852.8205 *
```

Figure 34 - CART Groupe « Profitables » - Prime Technique



CART - Technical Premium - Good

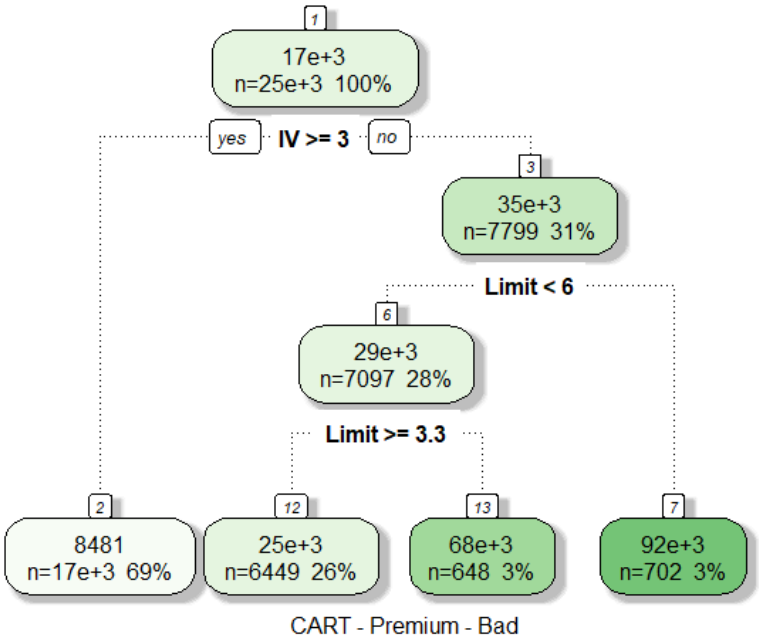
```
1) root 26571 7.256324e+14  51773.42000
   2) IV>=1.5 24091 1.261550e+14  24446.62000
     4) Limit>=4.5 21356 3.600197e+13  14818.04000 *
     5) Limit< 4.5 2735 7.271320e+13  99630.52000
      10) Stowage< 2 756 2.503981e+08     41.97354 *
      11) Stowage>=2 1979 6.235074e+13 137674.50000
        22) IV>=2.5 1123 9.774953e+12  71113.55000 *
        23) IV< 2.5 856 4.107334e+13 224996.80000 *
   3) IV< 1.5 2480 4.067301e+14 317229.00000
     6) Coverage< 1.5 992 5.470923e+09     74.59980 *
     7) Coverage>=1.5 1488 2.404209e+14 528665.30000
      14) Commodity< 3.5 337 1.204190e+13 293147.10000 *
      15) Commodity>=3.5 1151 2.042129e+14 597622.40000
        30) Conveyance>=2.5 764 1.115193e+14 528922.30000 *
        31) Conveyance< 2.5 387 8.196921e+13 733247.40000 *
```

3.4.2.5. Conclusion intermédiaire sur les comptes profitables

Au sein des comptes profitables, l'arbre des sinistres ne peut être découpé et reste à la racine. Cela est dû au nombre faible de comptes sinistrés, et qu'il n'y a pas d'hétérogénéité dans les données permettant un découpage des facteurs. La valeur assurée permet de déterminer la prime en appliquant un taux et est ainsi la principale variable qui se retrouve ici. La prime technique du modèle suit les mêmes facteurs que le portefeuille global sur la partie gauche mais une légère différence peut être observée sur la partie droite de l'arbre correspondant aux grandes primes avec l'apparition de nouvelles variables comme les conditions de couvertures, les marchandises et le mode de transport.

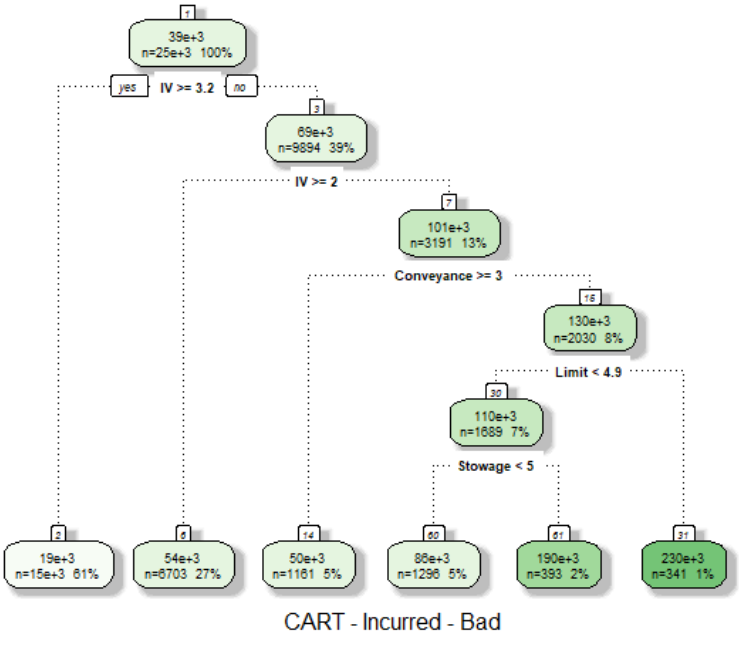
3.4.2.6. CART sous-groupe « Non-profitables »

Figure 35 - CART Groupe « Non-Profitable » – Prime Souscrite



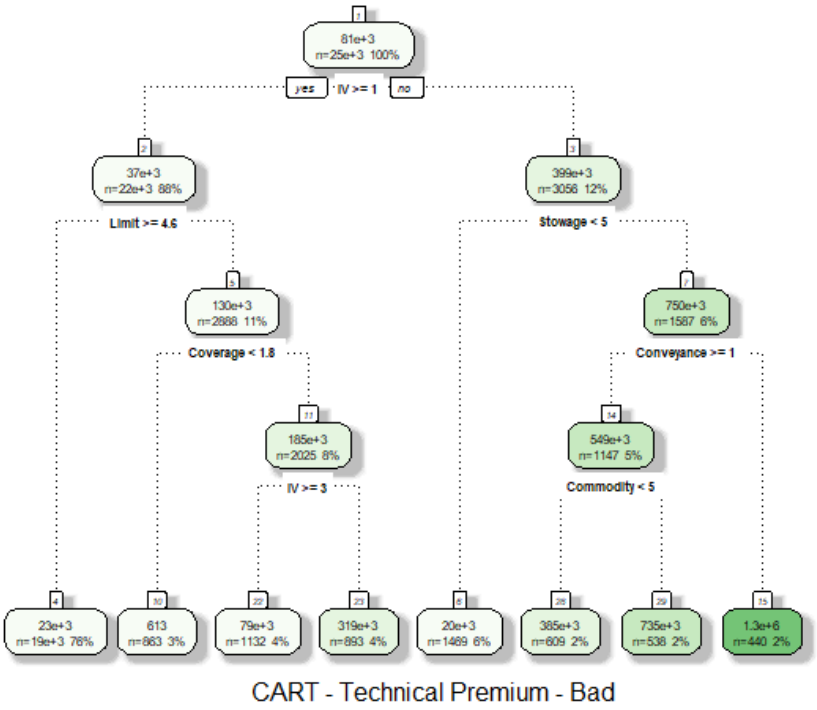

```
1) root 25278 6.664427e+13 16596.740
   2) IV>=2.956332 17479 4.536204e+12  8480.795 *
   3) IV< 2.956332 7799 5.837643e+13 34786.080
     6) Limit< 6.020139 7097 3.038599e+13 29098.890
      12) Limit>=3.286363 6449 1.376916e+13 25216.560 *
      13) Limit< 3.286363 648 1.555225e+13 67736.460 *
     7) Limit>=6.020139 702 2.544025e+13 92281.770 *
```

Figure 36 - CART Groupe « Non-Profitable » – Sinistres



```
1) root 25278 2.174717e+14  38691.66
   2) IV>=3.223855 15384 1.874624e+13  19304.46 *
   3) IV< 3.223855 9894 1.839524e+14  68836.47
     6) IV>=1.977664 6703 6.142961e+13  53585.58 *
     7) IV< 1.977664 3191 1.176888e+14 100872.40
      14) Conveyance>=2.956701 1161 3.324919e+12  49532.63 *
      15) Conveyance< 2.956701 2030 1.095536e+14 130234.70
        30) Limit< 4.916795 1689 6.764918e+13 110084.80
          60) Stowage< 5 1296 2.936188e+13  85945.52 *
          61) Stowage>=5 393 3.504172e+13 189689.10 *
        31) Limit>=4.916795 341 3.782197e+13 230038.90 *
```

Figure 37 - CART Groupe « Non-Profitable » – Prime Technique



CART - Technical Premium - Bad

```
1) root 25278 2.038707e+15   80988.9900
   2) IV>=1.001085 22222 2.151497e+14   37199.0300
     4) Limit>=4.628723 19334 6.575226e+13   23338.3500 *
     5) Limit< 4.628723 2888 1.208165e+14  129990.7000
      10) Coverage< 1.792587 863 3.674089e+10     612.6066 *
      11) Coverage>=1.792587 2025 1.001780e+14  185128.1000
        22) IV>=2.999634 1132 1.121319e+13   79142.8600 *
        23) IV< 2.999634 893 6.013043e+13  319479.0000 *
   3) IV< 1.001085 3056 1.471088e+15  399411.9000
     6) Stowage< 5 1469 9.628882e+12   20437.2100 *
     7) Stowage>=5 1587 1.055185e+15  750208.3000
      14) Conveyance>=1.014646 1147 1.482347e+14  548899.4000
        28) Commodity< 4.999577 609 7.231378e+13  384568.4000 *
        29) Commodity>=4.999577 538 4.085878e+13  734917.3000 *
      15) Conveyance< 1.014646 440 7.392967e+14 1274984.0000 *
```

3.4.2.7. Conclusion intermédiaire sur les comptes non-profitables

En observant la catégorie « Non-profitable », une inadéquation de la prime souscrite avec la sinistralité est constatée. La prime souscrite ne semble pas prendre en compte la complexité des comptes puisque seulement basée sur la valeur assurée et la limite d'assurance. La prime technique diffère cette fois ci du portefeuille global et du groupe « Profitable ». Cela montre de ce fait, qu'il existe une probabilité forte que le modèle de tarification soit mal calibré.

3.4.2.8. Conclusion

L'algorithme CART présente l'avantage de classer les données en ayant une représentation graphique simple. Il est donc intuitif et très utile pour la prise de décision lorsqu'il est appliqué sur un portefeuille. La volatilité des résultats suivant le nombre minimum d'observations et le choix du sous-ensemble rend la généralisation difficile car très sensible au surapprentissage.

La conclusion qui peut être prise avec relativement de précaution en analysant les différents arbres, est que la méthode de calcul des souscripteurs est en réalité en lien avec le profil de risque réel. Le modèle de tarification présente également une complexité inadéquate et des incohérences de différentiation du risque pouvant résulter à une prime souvent beaucoup plus élevée par rapport au marché.

Enfin, le CART seul ne permet pas de répondre à la problématique initiale à savoir quelles sont les caractéristiques communes aux comptes avec une mauvaise performance. L'arbre ne fonctionne pas lorsque le portefeuille n'est pas assez hétérogène pour permettre un découpage grâce à l'indice de Gini. Par ailleurs, si le portefeuille n'a pas une taille critique, les algorithmes sont exposés à un risque de surapprentissage fort, puisqu'ils basent leurs décisions sur un nombre limité d'observations.

3.4.3. Forêt aléatoire

Les forêts aléatoires permettent à la fois de faire de la classification mais aussi de la régression. Elles sont donc théoriquement utilisables pour prédire un montant de sinistralité et être utilisées pour la tarification. Contrairement au CART, les forêts aléatoires permettent de pallier un sur-apprentissage et de réduire considérablement la variance de l'estimation afin d'obtenir des résultats plus robustes.

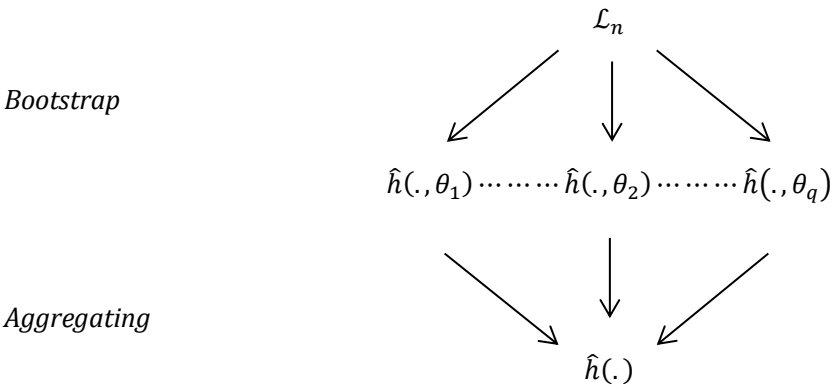
3.4.3.1. Principes des forêts aléatoires

Dans la famille des arbres de décisions, il existe des méthodes d'ensemble pour converger vers des résultats plus stables en agrégeant une série de prédicteurs. Le principe consiste à construire une série de prédicteurs et à les agréger.

En régression, les prédictions reviennent à faire une moyenne des estimations $\hat{y}_t$ et en classification, cela revient à calculer les labels de classes qui reviennent le plus souvent.

Ainsi, en construisant des prédicteurs individuels optimisés avec des sous-ensembles de l'univers global, la probabilité d'erreur est réduite, puisqu'il est nécessaire d'avoir la même erreur provenant de différents prédicteurs pour que cela soit pris en compte. De plus, si en utilisant des sous-ensembles différents, on converge vers un même résultat, il sera jugé fiable.

Les forêts aléatoires appartiennent aux méthodes dites de *Bagging,* introduites par Breiman [1996] et qui consiste à tirer un grand nombre d'échantillons indépendants des uns et des autres et à construire des prédicteurs qui seront ensuite agrégés entre eux avec une moyenne ou un vote majoritaire des labels de classes.



D'autres méthodes existent comme le Boosting, une méthode introduite par Freund et al. [1996] et qui diffère par rapport au Bagging, principalement dans la sélection des échantillons qui ne sont plus indépendants entre eux.

Un premier échantillon commence par être tiré, sur lequel un prédicteur $\hat{h}_t$ est déterminé. Les échantillons successifs sont ensuite choisis en fonction de l'erreur dans l'estimation du prédicteur $\hat{h}_t$

précédant, en cherchant à tirer une observation mal prédite. On cherche ainsi à se concentrer sur les mauvaises prédictions afin de diminuer l'erreur globale. L'ensemble des prédicteurs est ensuite agrégé avec une moyenne pondérée par des poids exponentiels.

3.4.3.2. Calibrage des forêts aléatoires

Afin d'optimiser l'algorithme, deux paramètres peuvent être modifiés dans la construction des forêts aléatoires : Le nombre de variables utilisées pour la construction de chaque nœud et le nombre d'arbres dans l'échantillon.

Le taux d'erreur décroit généralement avec le nombre d'arbres construits jusqu'à converger vers un seuil. Concernant le nombre de variables, il convient habituellement de multiplier ou de diviser par deux le nombre de variables utilisées à chaque découpage (mtry) afin de tester la sensibilité et réduire le taux d'erreur du modèle.

Avec la construction de multiples arbres, les forêts aléatoires permettent d'identifier les principaux facteurs influant sur la prédiction. En effet, le *Mean Decrease Gini* mesure la réduction d'impureté de chaque facteur en calculant, la moyenne de réduction pondérée par le nombre d'occurrence du facteur dans les arbres de la forêt. Plus le *Mean Decrease Gini* est haut, plus l'importance du facteur est haute dans la prédiction.

3.4.3.3. Application des forêts aléatoires

L'utilisation des forêts aléatoires va permettre dans notre cas d'identifier les variables qui reviennent le plus souvent dans le découpage des arbres et ainsi de réduire les dimensions utilisées sans perdre trop d'information dans la suite de l'étude. L'algorithme sera appliqué pour expliquer ces variables :

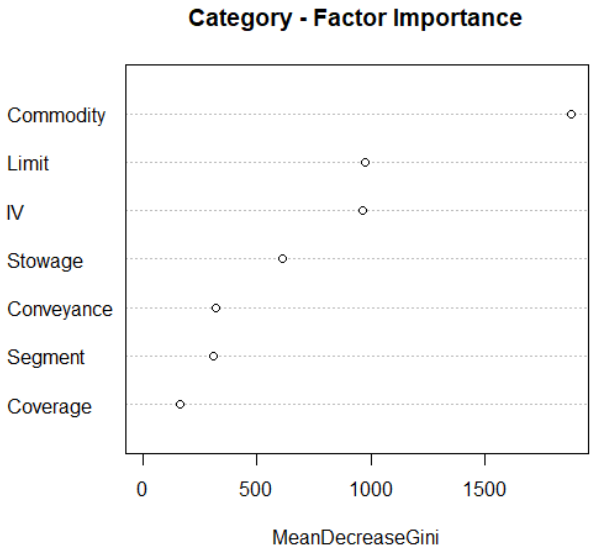
- La catégorie représentant l'adéquation de la prime avec la sinistralité.
- La fréquence de sinistres
- La sévérité des sinistres

Catégorie

```
randomForest(formula = Category ~ ., data = data_rf_class, mtry = 4, ntree = 500, na.action
= na.omit)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 4

        OOB estimate of error rate: 33.62%
Confusion matrix:
      0     1 class.error
0 14146 11132   0.4403829
1  6300 20271   0.2371006
```

En appliquant les forêts aléatoires sur la catégorie de profitabilité du compte, c'est-à-dire implicitement le ratio S/P, les facteurs revenant les plus souvent dans le découpage des nœuds sont la marchandise, la limite, les valeurs assurées et dans une moindre mesure le conditionnement. Malgré l'optimisation des paramètres, l'erreur de classification reste relativement grande. En effet, le ratio S/P a une faible variabilité dans notre portefeuille, et l'adéquation de la prime sur la sinistralité n'est pas toujours liée. Le modèle sera appliqué par la suite, sur la fréquence et la sévérité des sinistres pour obtenir une idée plus précise.

Figure 38 - Importance des facteurs pour l'explication de la profitabilité

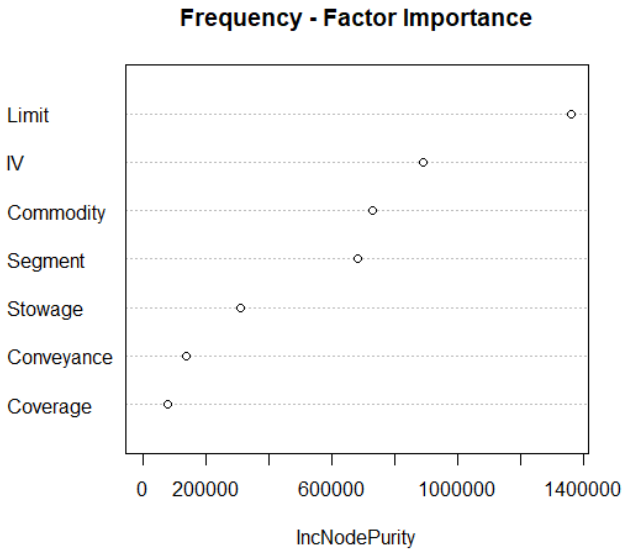


Fréquence

```
randomForest(formula = ClaimsCount ~ ., data = data_rf_class, mtry = 4, ntree = 500, na.act
ion = na.omit)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 4

          Mean of squared residuals: 54.93581
                    % Var explained: 58.21
```

L'application des forêts aléatoires sur la fréquence de sinistres montre que le segment du client influence également sur le nombre de sinistres après les précédents facteurs, mais n'est pas le plus important. En effet, la probabilité d'avoir un nombre élevé de sinistres semble dépendre de la taille du client mais le même nombre de sinistres d'une petite-moyenne entreprise peut être retrouvé chez une multinationale qui possède des franchises généralement plus hautes.

Figure 39 - Importance des facteurs pour l'explication de la fréquence

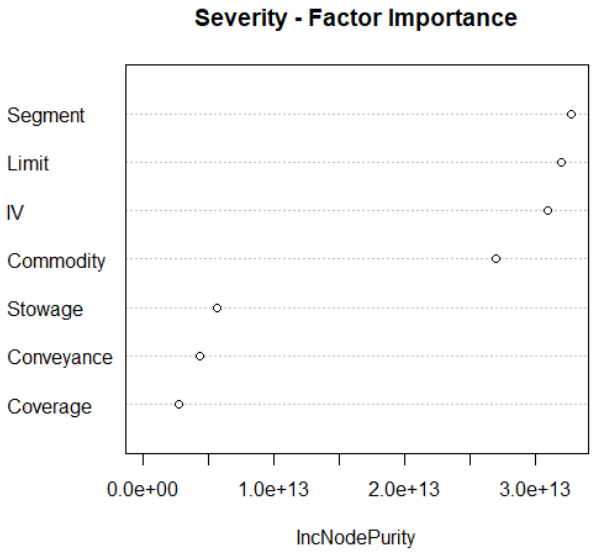


Sévérité

```
randomForest(formula = Incurred ~ ., data = data_rf_class, mtry = 4, ntree = 500, na.action
= na.omit)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 4

          Mean of squared residuals: 1898444646
                    % Var explained: 55.83
```

Les mêmes facteurs que pour la fréquence reviennent ici mais dans un ordre différent. En effet, le segment du client va principalement influencer sur la sévérité des sinistres, puisqu'un programme multinational possède souvent des valeurs assurées plus importantes qu'une petite-moyenne entreprise et cette accumulation de valeurs peut entrainer des sinistres plus larges. Bien que l'étude serait beaucoup plus précise avec les sous-limites, l'influence de la limite annuelle d'assurance va également jouer un rôle important dans la limitation des pertes.

Figure 40 - Importance des facteurs pour l'explication de la sévérité



L'utilisation des forêts aléatoires a permis une réduction de dimensions pour ne garder que les seules variables ayant un impact sans perte d'information significatif. Ces facteurs identifiés seront utilisés par la suite pour le reste des algorithmes et pour la construction d'un modèle linéaire généralisé simplifié. Il faut néanmoins noter qu'il n'a pas été possible de réduire plus encore le taux d'erreur de classification et que les pourcentages de variance expliquée pour la fréquence et la sévérité sont optimaux. Etant donné que l'utilisation des résultats se limite à l'identification des variables importantes, l'optimisation de l'erreur ne sera pas recherchée avec un autre algorithme tel que le *Boosting*.

3.4.4. Séparateurs à Vaste Marge

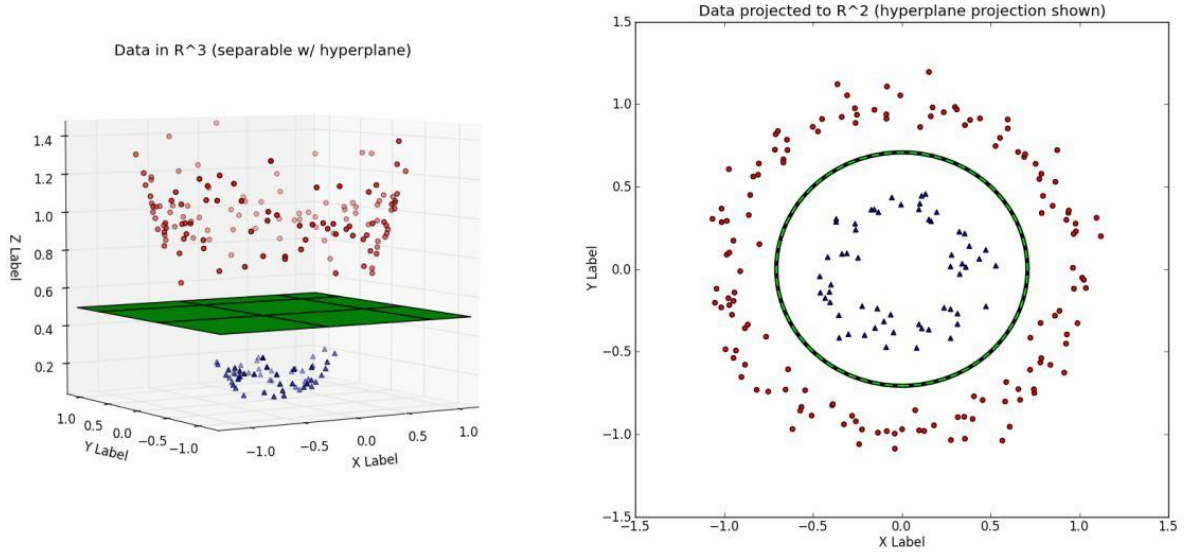
Le Séparateur à Vaste Marge ou *SVM* est utilisé pour essayer de déterminer un profil type des comptes du portefeuille dont la profitabilité n'est pas bonne. Cette information permet d'identifier les zones stratégiques de croissance et les zones où des actions doivent être prises pour redresser le portefeuille. Bien que le SVM soit une méthode supervisée, l'utilisation sans base d'apprentissage ou de base de test est choisie car l'utilisation de ses capacités prédictives n'est pas requise. En effet, la généralisation du comportement d'une sous-partie du portefeuille pour un autre profil n'est pas souhaitée car le profil du portefeuille d'un pays à un autre n'est pas toujours le même.

3.4.4.1. Principes des Séparateurs à Vaste Marge

Introduits par Vapnik [1998], les SVM consistent à classer des données à l'aide de séparateurs linéaires. Ces séparateurs peuvent prendre la forme de vecteurs, mais également d'hyperplans dans un espace de plus grande dimension.

Dans la pratique, il est rare d'avoir des données où la séparation se fait de manière linéaire. Une méthode pour surmonter cette contrainte est de faire appel à une fonction Φ transformant l'espace d'entrée $\mathbb{R}^d$ en un espace de Hilbert $\mathcal{H}$ de plus grande dimension où les projections seront linéairement séparables comme le montre l'exemple fournie par Kim [2017] ci-dessous.

Figure 41 - Exemple de séparation par hyperplan d'observations projetées - Kim [2017]



La méthode des SVM utilise pour ce faire, les produits scalaires donnés par un noyau : $K(x, x') = \big(\Phi(x), \Phi(x')\big)$. Parmi les différents types de noyaux existants, les noyaux polynomiaux de la forme $K(x, x') = \big(\alpha + \beta(x + x')\big)^{\delta}$ sont choisis dans le but de réaliser des séparations non linéaires.

3.4.4.2. Calibrage des Séparateurs à Vaste Marge

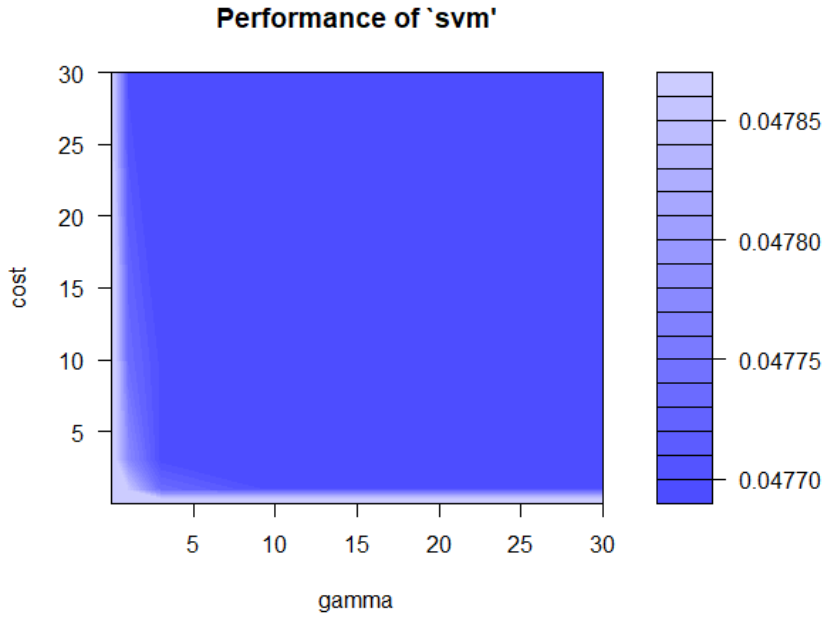
Le calibrage des *SVMs* se fait en choisissant deux paramètres : Le *gamma* et le *cost*

- Le *gamma* représente la zone d'influence qu'une observation va avoir sur la frontière. La relation est inverse, ainsi, une valeur faible va étendre cette zone et une valeur haute va la réduire. Une zone faible d'influence va augmenter le risque de surapprentissage car le modèle va dépendre d'un nombre d'observations plus faible.
- Le *cost* représente quant à lui la sensibilité de la frontière aux observations aberrantes. Ainsi une valeur haute va induire une maximisation de l'ajustement de la frontière. Tous les points vont donc être pris en compte. Une valeur faible va permettre de lisser la frontière et éviter un surapprentissage.

L'optimisation du couple de paramètres est réalisée automatiquement sur la base d'apprentissage par une fonction. Le couple optimal retenu étant respectivement 10 et 1.

```
Parameter tuning of 'svm':
- sampling method: fixed training/validation set
- best parameters:
 gamma cost
    10    1
- best performance: 0.04769641
```

Figure 42 - Performance de l'algorithme SVM en fonction du coût et du gamma



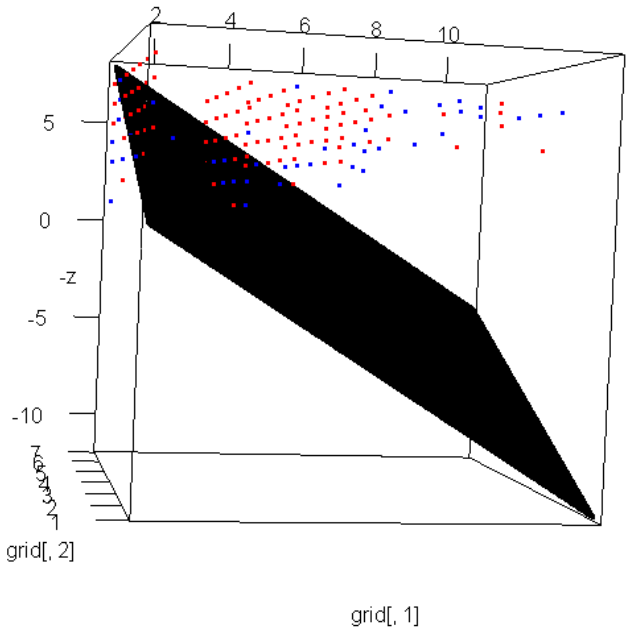
3.4.4.3. Application des Séparateurs à Vaste Marge

Après un premier essai avec la base de données issue du sur-échantillonnage, les observations synthétiques entrainent une lenteur de calcul pouvant aller à plusieurs jours et donnent un nombre de vecteurs de support important. Malgré la projection par noyau, les observations synthétiques bruitent la détermination des frontières car les comptes non-profitables sont équirépartis dans le portefeuille. Hu & Li [2013] montrent également que le rééchantillonnage a un bénéfice relatif avec les SVMs. Le choix est donc fait d'utiliser les *SVMs* sur la base initiale.

```
svm(formula = Category ~ ., data = data_svm, type = "C-classification", kernel = "polynomia
l",
    cost = best_cost, gamma = best_gamma)

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  polynomial
       cost:  1
     degree:  3
     coef.0:  0

Number of Support Vectors:  1672
 ( 797 875 )
Number of Classes:  2
Levels:
 0 1
```

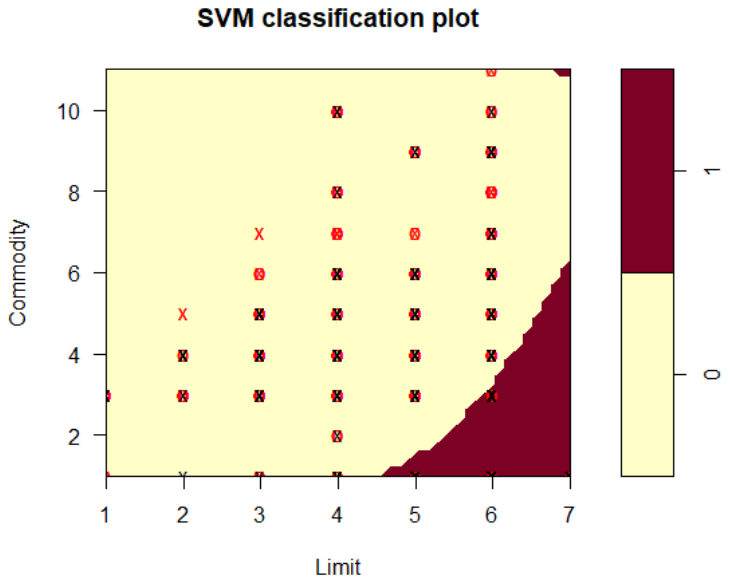
Pour faciliter la visualisation du portefeuille sur trois dimensions, le segment est mis de côté. Nous souhaitons représenter en 3D l'hyperplan et les observations mais nous rencontrons des difficultés pour afficher les projections des observations par le noyau. En l'état actuel, il n'existe pas de séparation linéaire des observations sans projection.

Figure 43 - Représentation 3D de l'hyperplan et des observations



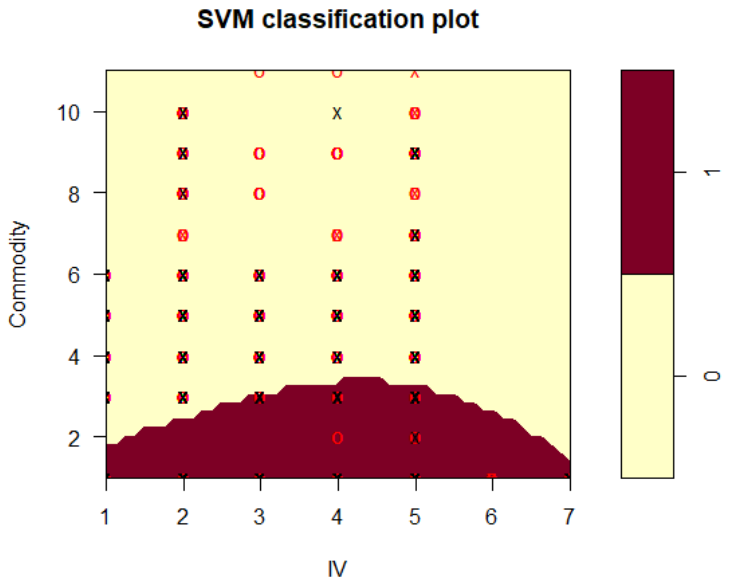
Nous choisissons d'afficher la représentation 2D de la frontière, pour chacun des couples de variables mais il est nécessaire de garder à l'esprit qu'une partie de l'information est tronquée sur le troisième axe manquant. Dans les représentations qui suivent, la zone en beige correspond aux comptes non-profitables et la zone en pourpre correspond aux comptes profitables. Les x et o représentent respectivement les vecteurs de supports et les observations.

Figure 44 - Représentation 2D de la frontière Marchandise et Limite



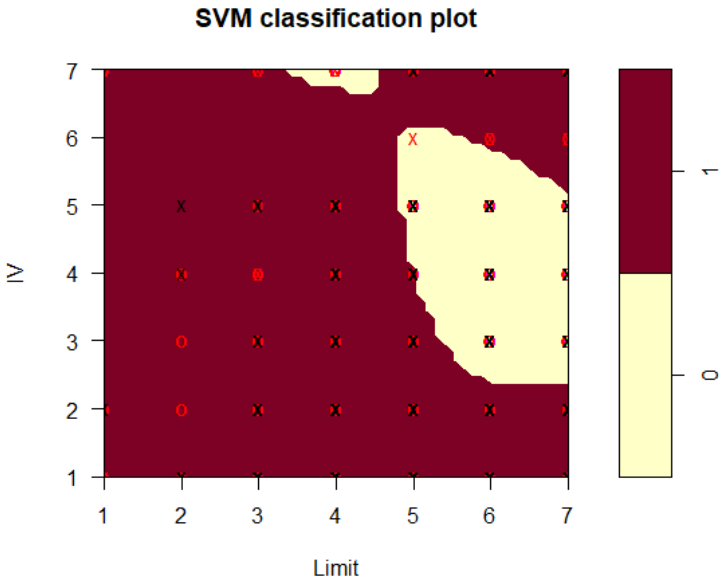
Les entreprises assurant des marchandises peu risquées et ayant une limite d'assurance faible présentent de bonnes performances.

Figure 45 - Représentation 2D de la frontière Marchandise et Valeur assurées



La représentation montre encore une fois que les entreprises ayant des marchandises peu risquées, possèdent de meilleurs résultats quel que soit le volume de biens assurées.

Figure 46 - Représentation 2D de la frontière Limite et Valeur assurées



Dans la dernière dimension, les comptes non-profitables ont principalement de petites limites et cela est le cas pour toutes les entreprises sauf celles ayant de très grandes valeurs assurées.

3.4.4.4. Conclusion

Ces trois représentations graphiques donnent une idée globale des *clusters* :

- Petites limites et marchandises peu risquées
- Marchandises peu risquées pour toutes les valeurs assurées
- Tout sauf les petites limites avec des petites et moyennes valeurs assurées

Les résultats du SVM montrent qu'une partie de l'information est perdue sans la combinaison de toutes les variables importantes car les conclusions sont parfois incohérentes entre elles. En fin de compte, cette méthode n'est pas adaptée en tant que telle, dans notre objectif de déterminer les caractéristiques des comptes non-profitables. En effet, 4 variables ayant un intérêt statistique avaient été identifiées et seulement une visualisation en 2D qui ne comporte pas toute l'information a pu être affichée. Enfin si une représentation 3D avait pu être affichée, elle n'aurait probablement pas suffi sur des portefeuilles ou des produits ayant plus de 3 variables d'intérêts. D'un point de vue pratique, le SVM requiert beaucoup de puissance de calcul, et les machines ont été poussé à leurs limites avec des calculs pouvant prendre jusqu'à plusieurs journées.

3.4.5. k-Moyennes

Dans la continuité de l'algorithme précédent, le souhait est de vérifier s'il existe un profil type de comptes dans le portefeuille dont la profitabilité n'est pas bonne. Pour cela, le recours aux k-Moyennes ou « k-Means » appartenant aux algorithmes d'apprentissage non supervisés, va permettre d'identifier des groupes de k individus ayant des caractéristiques similaires sans l'utilisation de base d'apprentissage au préalable.

3.4.5.1. Principes des k-Moyennes

Le principe de l'algorithme est de considérer les populations dans un environnement multidimensionnel représentant les différentes caractéristiques. L'algorithme s'appuie ensuite sur des éléments de mécanique du solide provenant du théorème de Huygens qui montre une relation entre les moments d'inerties des différents groupes.

Pour la construction de l'algorithme, il est nécessaire de déterminer à priori un nombre de *clusters*. Dans la phase d'initialisation, le barycentre des *clusters* sera généré aléatoirement sur le plan et les plus proches individus seront affectés à ces groupes. Les barycentres sont ensuite recalculés et l'affectation des individus change jusqu'à ce que le processus converge vers l'optimalité.

La recherche de l'optimalité vise à minimiser la distance des individus au sein du même groupe et à maximiser la distance des individus appartenant à des groupes différents. L'optimisation de ces deux distances est importante, car le risque de surapprentissage est grand. En effet, un individu peut en soi être considéré comme un groupe à lui tout seul.

Inertie totale = Inertie interclasses + Inertie intra-classes

$$\sum_{i=1}^{n} d^2(i,G) = \sum_{k=1}^{K} n_k d^2(G_k,G) + \sum_{k=1}^{K}\sum_{i=1}^{n_k} d^2(i,G_k)$$

1 2

1. Dispersion des barycentres de chaque groupe autour du barycentre global – Between_ss
2. Dispersion à l'intérieur de chaque groupe – Within_ss

En se basant sur ces deux formules, un procédé itératif est construit afin de trouver le nombre optimal de groupes en maximisant le rapport suivant, que nous appellerons dans la suite de l'étude, ratio d'inertie par simplification :

$$\arg max \frac{\sum_{k=1}^{K} n_k d^2(G_k,G)}{\sum_{i=1}^{n} d^2(i,G)}$$

3.4.5.2. Application des k-Moyennes

Après un premier essai avec la base de données issue du sur-échantillonnage, nous nous rendons compte que l'algorithme identifie des clusters censés n'être composés que de mauvais comptes, mais ne sont en réalité que des observations synthétiques. Nous décidons de réaliser cette étude sur la base originale qui semble plus adéquate. A l'étape du paramétrage, l'évolution de la dispersion intragroupe et intergroupe peut être observée dans les deux graphiques suivant en fonction du nombre de *clusters*. Le ratio d'inertie est maximisé pour un nombre de *clusters* de 86 sur la figure 48. Cette valeur est retenue pour la construction de l'algorithme de k-Means.

Figure 47 - Représentation des inerties en fonction du nombre de clusters

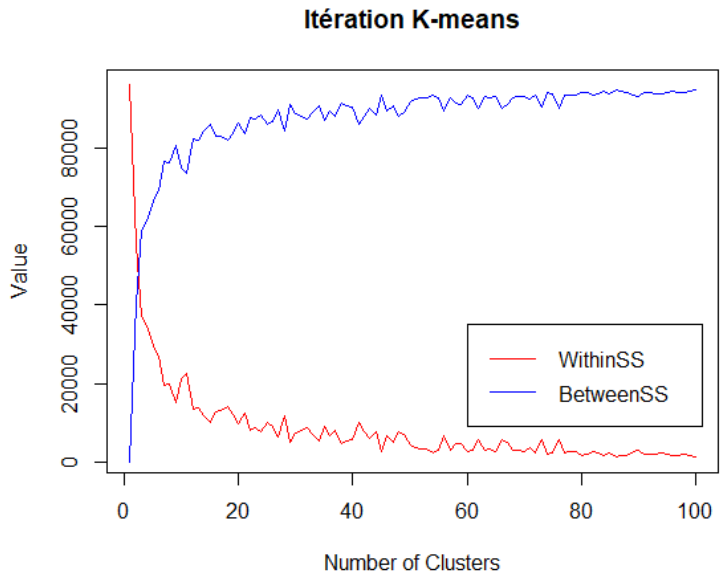
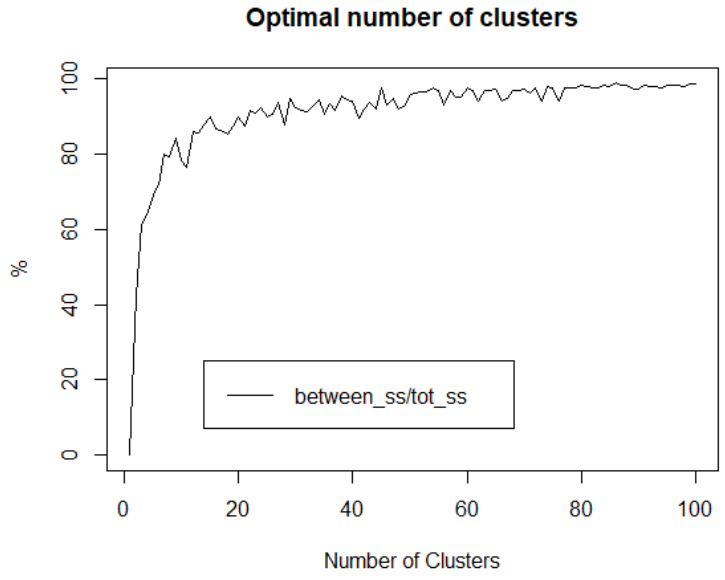


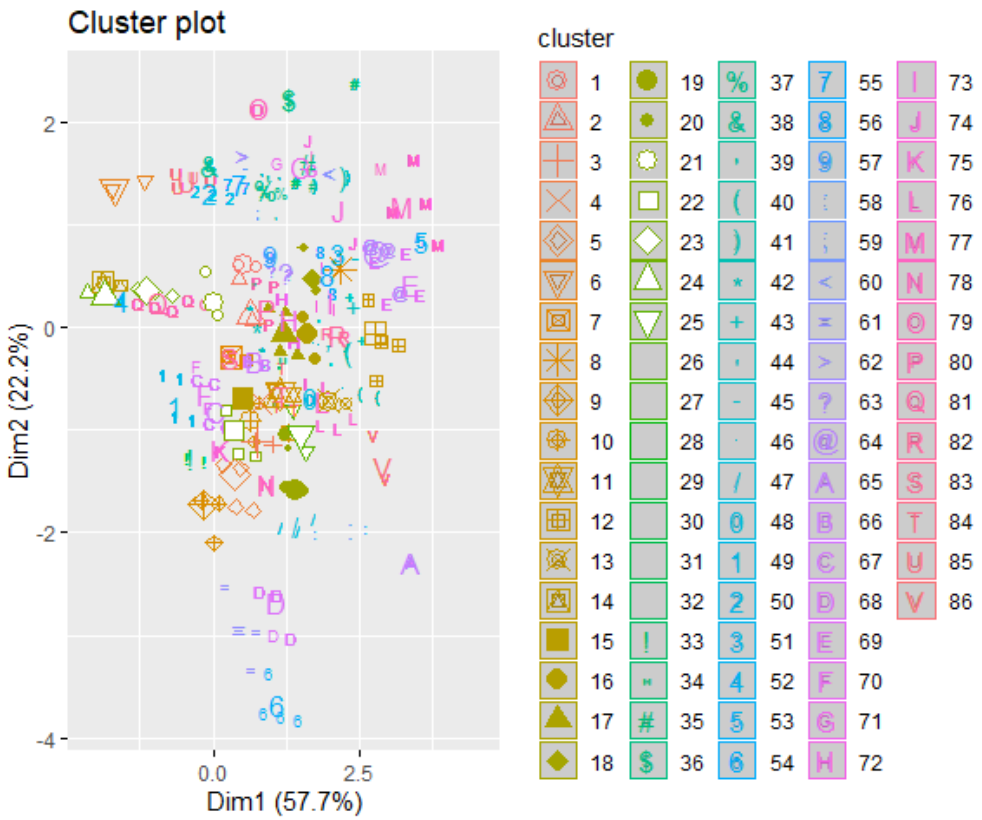
Figure 48 – Ratio d'inertie affichant le nombre optimal de clusters



Le périmètre de l'algorithme est restreint aux facteurs identifiés par les forêts aléatoires comme pour l'algorithme précédent. Due à la quantité de dimensions choisie et au grand nombre de *clusters*, la visualisation du résultat n'est pas aisée. Avec 86 *clusters*, le ratio d'inertie est de 97.8% ce qui évoque une bonne qualité de représentation. Un nombre k beaucoup plus faible aurait pu être choisi, avec un ratio d'inertie convenable mais l'objectif est d'avoir une segmentation assez granulaire pour ne pas impacter un nombre de comptes profitables trop important lors de l'augmentation tarifaire dans les *clusters* retenus.

```
K-means clustering with 86 clusters of sizes 208, 118, 13, 30, 397, 2455, 13, 18, 652, 22,
48, 26, 83, 1419, 90, 363, 219, 4, 3, 13, 567, 15, 1168, 1190, 7, 110, 13, 202, 90, 236, 14
, 7, 2156, 16, 37, 4, 18, 27, 40, 257, 93, 401, 37, 97, 38, 472, 12, 54, 3407, 200, 2, 323,
7, 4, 64, 9, 75, 20, 31, 15, 125, 3, 224, 22, 83, 324, 599, 31, 247, 3, 3, 189, 712, 12, 13
, 21, 1489, 6, 49, 338, 730, 5, 348, 327, 348, 26

[…]

(between_SS / total_SS =  97.8 %)
```

Figure 49 - Représentation des clusters



Afin de reprendre les informations fournies par l'algorithme d'une façon plus lisible, le tableau 9 affiche le profil de tous les *clusters*, avec la répartition du nombre de comptes et du volume de prime et de sinistres par catégorie de performance. Une équi-distribution de mauvais comptes peut être observée au sein de chaque *clusters*, et est représentative du portefeuille global et du constat initial. Il n'y a donc à priori, pas un profil type de mauvais compte qui se démarquerait.

Table 12 - Tableau des clusters, des ratio S/P et volumes de primes

Category	0	Bad performers		1	Good performers						
Cluster	**Count**	**Incurred**	**Premium**	**Count**	**Incurred**	**Premium**	**LR - Bad**	**LR - Good**	**LR Total**	**Bad - Premium Volume**	**Good - Premium Volume**
1	9	138,875	41,986	199	101,498	1,325,486	331%	8%	18%	3%	97%
2	5	42,007	22,619	113	166,958	1,361,908	186%	12%	15%	2%	98%
3	0	-	-	13	2,739	68,191		4%	4%	0%	100%
4	2	152,550	35,472	28	46,345	274,118	430%	17%	64%	11%	89%
5	19	3,277,820	1,216,668	378	591,462	6,039,678	269%	10%	53%	17%	83%
6	63	1,139,291	668,698	2392	470,615	4,618,203	170%	10%	30%	13%	87%
7	1	42,011	8,895	12	1,607	134,829	472%	1%	30%	6%	94%
8	0	-	-	18	134,208	262,655		51%	51%	0%	100%
9	30	561,670	289,312	622	289,862	3,245,794	194%	9%	24%	8%	92%
10	0	-	-	22	2,709	137,179		2%	2%	0%	100%
11	2	431,269	26,841	46	52,225	543,104	1607%	10%	85%	5%	95%
12	3	98,164	16,683	23	174,555	1,447,400	588%	12%	19%	1%	99%
13	9	388,658	200,910	74	135,880	980,915	193%	14%	44%	17%	83%
14	62	1,281,571	776,710	1357	448,368	4,465,636	165%	10%	33%	15%	85%
15	5	61,186	38,247	85	49,864	501,076	160%	10%	21%	7%	93%
16	27	1,084,846	461,768	336	212,058	3,936,309	235%	5%	29%	10%	90%
17	9	1,086,392	133,962	210	333,170	3,108,345	811%	11%	44%	4%	96%
18	3	32,405	34,579	1	-	545	94%	0%	92%	98%	2%
19	1	1,860	592	2	-	42,643	314%	0%	4%	1%	99%
20	2	331,017	28,863	11	-	186,809	1147%	0%	153%	13%	87%
21	26	616,891	85,907	541	218,600	2,771,447	718%	8%	29%	3%	97%
22	0	-	-	15	9,244	83,824		11%	11%	0%	100%
23	75	2,084,422	1,070,942	1093	836,003	7,244,654	195%	12%	35%	13%	87%
24	37	693,669	194,954	1153	210,477	2,500,206	356%	8%	34%	7%	93%
25	1	48,524	7,789	6	6,878	294,937	623%	2%	18%	3%	97%
26	9	734,302	251,576	101	104,670	1,619,889	292%	6%	45%	13%	87%
27	0	-	-	13	103,633	479,118		22%	22%	0%	100%
28	14	1,682,582	259,510	188	359,075	3,052,674	648%	12%	62%	8%	92%
29	2	26,560	20,621	88	88,704	783,152	129%	11%	14%	3%	97%
30	13	322,946	116,512	223	138,636	1,752,857	277%	8%	25%	6%	94%
31	5	2,178,742	1,634,230	9	26,774	505,706	133%	5%	103%	76%	24%
32	0	-	-	7	-	117,332		0%	0%	0%	100%
33	109	1,688,481	570,578	2047	864,723	10,288,608	296%	8%	24%	5%	95%
34	1	29,369	31,530	15	1,042	132,959	93%	1%	18%	19%	81%
35	8	904,500	588,970	29	259,313	999,702	154%	26%	73%	37%	63%
36	3	178,199	107,707	1	-	178	165%	0%	165%	100%	0%
37	6	244,025	123,741	12	38,215	215,551	197%	18%	83%	36%	64%
38	0	-	-	27	13,896	407,892		3%	3%	0%	100%
39	2	31,039	47,681	38	70,470	564,470	65%	12%	17%	8%	92%
40	14	1,589,462	670,819	243	902,618	4,964,631	237%	18%	44%	12%	88%
41	4	335,956	31,516	89	94,141	734,347	1066%	13%	56%	4%	96%
42	26	679,162	188,230	375	156,562	2,134,600	361%	7%	36%	8%	92%
43	4	159,322	109,531	33	53,665	1,018,364	145%	5%	19%	10%	90%
44	7	183,652	97,564	90	96,712	924,206	188%	10%	27%	10%	90%
45	3	82,789	25,871	35	5,376	508,421	320%	1%	17%	5%	95%
46	20	679,722	520,773	452	404,996	4,353,357	131%	9%	22%	11%	89%
47	0	-	-	12	8,963	55,923		16%	16%	0%	100%
48	2	54,877	3,205	52	68,147	714,526	1712%	10%	17%	0%	100%
49	115	1,473,532	504,050	3292	1,289,686	15,537,625	292%	8%	17%	3%	97%
50	15	390,775	329,646	185	313,137	3,305,582	119%	9%	19%	9%	91%
51	0	-	-	2	14,402	28,422		51%	51%	0%	100%
52	9	381,405	85,048	314	68,407	1,277,786	448%	5%	33%	6%	94%
53	2	14,039	10,238	5	-	54,356	137%	0%	22%	16%	84%
54	1	14,410	7,906	3	4,867	44,829	182%	11%	37%	15%	85%
55	13	577,402	333,833	51	112,164	969,276	173%	12%	53%	26%	74%
56	0	-	-	9	71,388	189,105		38%	38%	0%	100%
57	1	63,208	58,817	74	121,218	515,665	107%	24%	32%	10%	90%
58	2	60,010	11,962	18	5,104	120,314	502%	4%	49%	9%	91%
59	4	87,572	14,414	27	7,642	287,394	608%	3%	32%	5%	95%
60	1	157,766	33,045	14	140,825	518,435	477%	27%	54%	6%	94%
61	14	178,097	64,224	111	53,785	449,598	277%	12%	45%	12%	88%
62	0	-	-	3	-	57,049		0%	0%	0%	100%
63	14	631,770	213,484	210	186,858	2,124,636	296%	9%	35%	9%	91%
64	0	-	-	22	291,866	663,501		44%	44%	0%	100%
65	2	7,156	4,783	81	47,366	442,098	150%	11%	12%	1%	99%
66	17	709,251	277,922	307	401,299	3,899,487	255%	10%	27%	7%	93%
67	29	1,059,359	566,631	570	770,932	6,881,082	187%	11%	25%	8%	92%
68	1	98,453	62,622	30	42,534	334,578	157%	13%	35%	16%	84%
69	18	2,856,007	955,448	229	1,202,472	6,330,007	299%	19%	56%	13%	87%
70	0	-	-	3	-	7,341		0%	0%	0%	100%
71	0	-	-	3	-	80,004		0%	0%	0%	100%
72	13	779,556	157,856	176	122,394	1,359,024	494%	9%	59%	10%	90%
73	29	2,049,646	508,724	683	977,906	7,692,723	403%	13%	37%	6%	94%
74	2	93,163	78,750	10	-	365,107	118%	0%	21%	18%	82%
75	1	7,303	12,332	12	-	61,255	59%	0%	10%	17%	83%
76	1	14,120	10,150	20	23,066	257,973	139%	9%	14%	4%	96%
77	84	6,903,261	2,562,906	1405	1,831,751	14,698,563	269%	12%	51%	15%	85%
78	0	-	-	6	-	517,707		0%	0%	0%	100%
79	5	833,072	269,670	44	216,475	1,725,573	309%	13%	53%	14%	86%
80	16	1,059,839	440,716	322	241,455	3,339,473	240%	7%	34%	12%	88%
81	48	4,705,983	2,815,033	682	808,743	8,981,133	167%	9%	47%	24%	76%
82	0	-	-	5	159	254,190		0%	0%	0%	100%
83	17	215,292	146,805	331	142,227	1,741,325	147%	8%	19%	8%	92%
84	14	747,324	241,104	313	369,286	3,673,555	310%	10%	29%	6%	94%
85	17	408,167	176,757	331	210,698	2,514,343	231%	8%	23%	7%	93%
86	1	323,665	51,556	25	46,560	379,059	628%	12%	86%	12%	88%

Plusieurs types de profil se démarquent :

1. Les *clusters* composés seulement de bons comptes ex 3
2. Les *clusters* composés en grande majorité de bons comptes ex 16
3. Les *clusters* composés en majorité de mauvais comptes ex 18
4. Les *clusters* avec de bons résultats mais avec un volume significatif de mauvais comptes ex 77
5. Les autres *clusters*

Un échantillon des mauvais profils et des bons profils de *clusters* est affiché ci-après, en tenant compte du barycentre du *cluster*, qui représente l'individu moyen.

Table 13 - Profil des clusters contenant une majorité de mauvais comptes

Clusters	Commodity	Limit	IV	Segment	Coverage	Stowage	Voyage	Conveyance
18	4	3	5	2	2	6	1	1
31	4	2	1	1	2	7	1	1
36	1	6	1	1	1	1	1	1

Compte tenu du faible nombre de comptes non-profitables dans le portefeuille, on identifie seulement trois profils contenant majoritairement des comptes non-profitables. Il y a, à la fois des multinationales avec de grandes valeurs assurées et des petites limites transportant des marchandises peu risquées et des moyennes entreprises avec des valeurs assurées modestes et une grande limite.

Table 14 - Profil des clusters contenant un gros volume de comptes non-profitables

Clusters	Commodity	Limit	IV	Segment	Coverage	Stowage	Voyage	Conveyance
77	4	4	1	1	1	3	1	1
81	1	7	2	3	1	1	1	1

Les *clusters* ci-dessus contiennent un volume important de comptes non-profitables. Une revue du modèle de tarification pour ces profils n'est cependant pas à l'étude puisque la Table 12 montre que la mutualisation au sein des groupes permet d'obtenir des résultats en dessous du ratio S/P cible. La hausse tarifaire impacterait également les clients profitables dont la rétention serait plus difficile. L'information sur ces profils permet néanmoins aux souscripteurs de revoir au cas par cas les primes de ces comptes.

Table 15 - Profil des clusters contenant une majorité de bons comptes

Clusters	Commodity	Limit	IV	Segment	Coverage	Stowage	Voyage	Conveyance
2	4	6	3	2	2	7	2	3
3	7	6	5	2	2	7	2	3
8	3	4	1	2	2	7	2	3
10	5	4	5	3	2	7	1	3
12	5	3	2	2	2	6	1	2
19	6	4	3	3	2	7	3	3
22	4	5	4	3	2	7	2	3
27	3	1	1	1	2	4	1	1
32	4	5	3	3	2	7	2	3
38	1	4	7	2	1	1	1	1
47	9	6	5	2	2	6	2	2
48	6	5	3	2	2	6	1	3
51	3	3	3	2	2	7	1	3
56	4	3	4	2	2	7	1	2
62	1	3	7	2	1	1	1	1
64	3	2	3	2	2	5	1	1
65	10	4	2	2	1	4	1	1
70	4	5	5	3	2	7	2	3
71	1	3	4	2	1	1	1	1
78	6	4	5	3	2	7	3	3
82	5	3	4	2	4	5	3	3

De la même manière, les *clusters* des comptes profitables peuvent être isolés pour déterminer les profils des clients cibles pour la stratégie de croissance. Ce sont principalement des entreprises de taille moyenne transportant beaucoup de marchandises peu risquées et avec de hautes limites d'assurances. Nous pouvons imaginer que ces profils cherchent principalement à s'assurer contre des risques d'accumulation, en choisissant de grandes limites pour leurs marchandises peu couteuses.

L'affichage du profil des 5 plus gros et 5 plus petits clusters en termes de nombre de polices d'assurance, permet d'avoir une idée des principaux profils qui composent le portefeuille. Ces *clusters* sont majoritairement des comptes profitables :

Table 16 - Profil des 5 plus gros clusters

Clusters	Commodity	Limit	IV	Segment	Coverage	Stowage	Voyage	Conveyance
6	1	7	7	2	1	1	1	1
14	1	6	7	3	1	1	1	1
33	5	6	5	3	2	7	2	3
49	4	6	5	3	2	7	2	3
77	4	4	1	1	1	3	1	1

Table 17 - Profil des 5 plus petits clusters

Clusters	Commodity	Limit	IV	Segment	Coverage	Stowage	Voyage	Conveyance
19	6	4	3	3	2	7	3	3
51	3	3	3	2	2	7	1	3
62	1	3	7	2	1	1	1	1
70	4	5	5	3	2	7	2	3
71	1	3	4	2	1	1	1	1

Conclusion

Le k-Means est un algorithme qui est difficile à mettre en œuvre. En effet, il est nécessaire de définir un nombre de *clusters* à priori, et cela peut être une tâche ardue car le risque de surapprentissage est grand puisqu'il est théoriquement possible d'avoir autant de groupes que d'individus. De même, l'algorithme peut donner des résultats légèrement différents, d'une fois à l'autre, car la convergence dépend de la localisation initiale des barycentres des groupes.

Il a cependant donné des résultats intéressants en montrant que les clients non-profitables sont présents dans des proportions plus ou moins grande dans tous les *clusters*, De plus, il a également permis d'identifier les profils composées quasi-exclusivement de clients profitables et non-profitables.

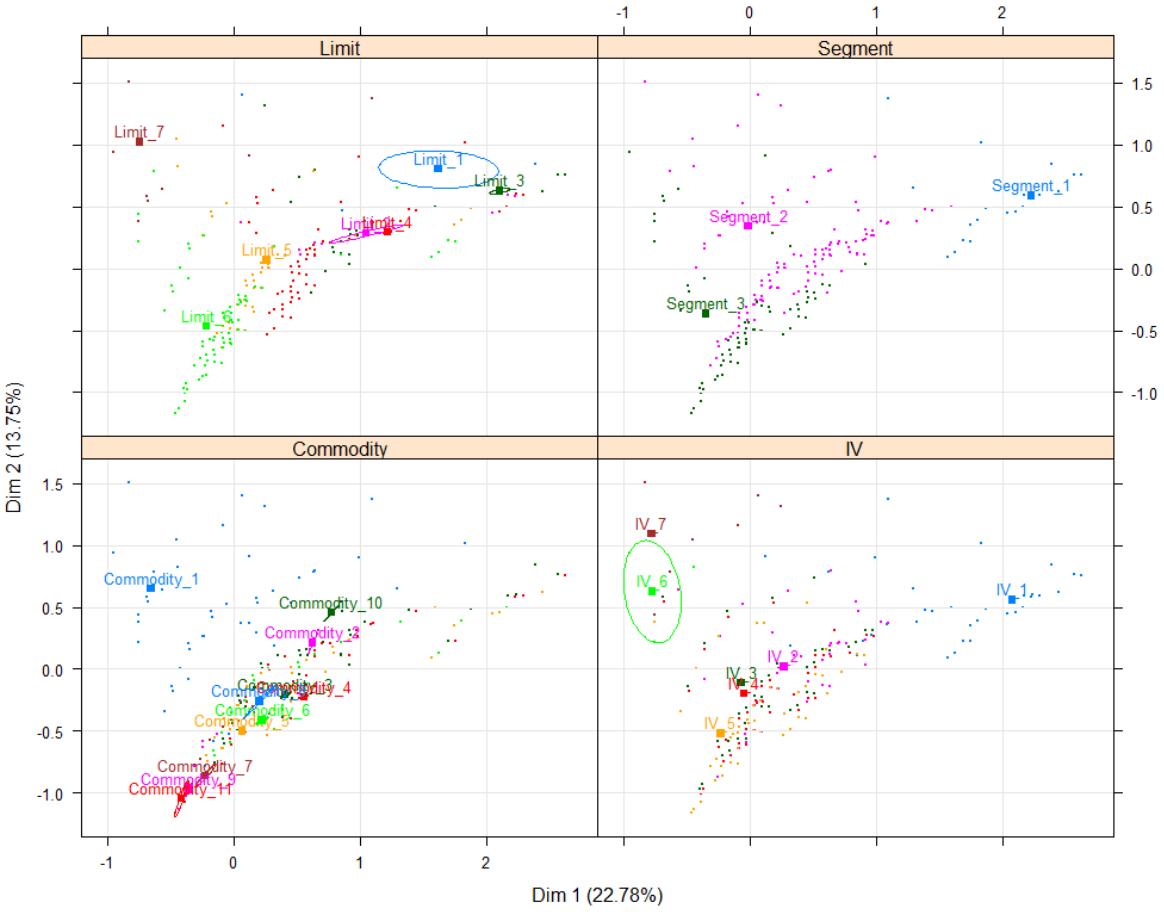
3.4.6. Classification Ascendante Hiérarchique

L'identification d'un profil s'est révélée difficile avec les précédents algorithmes, même si les k-Means ont apporté de premiers éléments de réponses. La Classification Ascendante Hiérarchique ou CAH est une autre méthode non supervisée pour la classification. C'est une méthode « ascendante » comme son nom l'indique, contrairement à l'arbre de décision qui est une méthode « descendante », car elle va agréger des groupes de plus en plus grands. Les sorties de la CAH permettent d'identifier les modalités prédominantes pour un échantillon de données, ce qui sera utilisé pour déterminer le profil des comptes dans un groupe.

3.4.6.1. Analyse des Correspondances Multiples

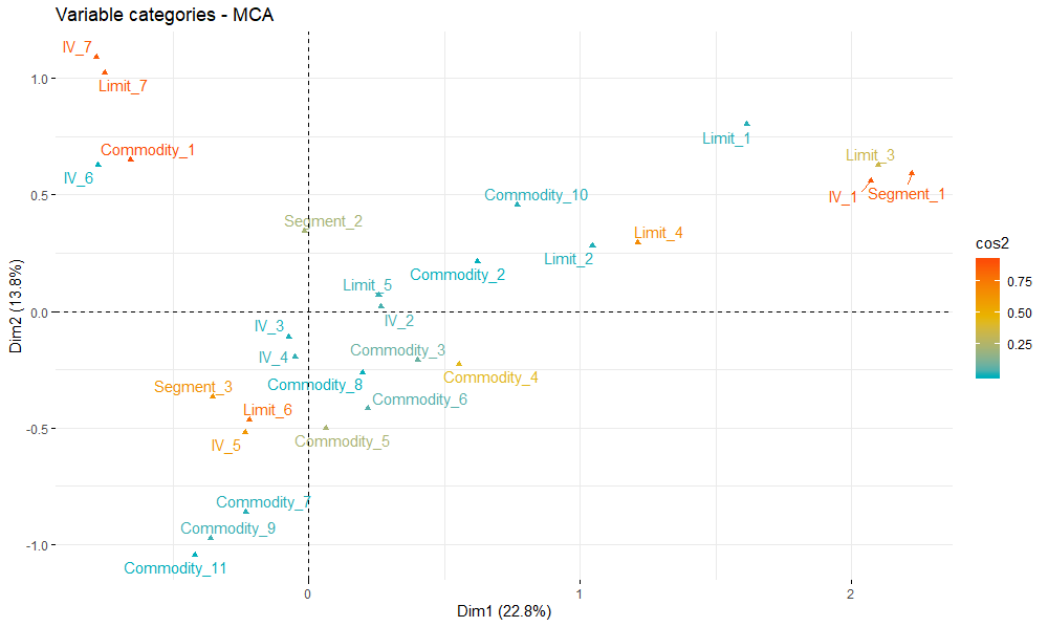
La CAH est un algorithme nécessitant des variables quantitatives en entrée, il est possible de l'utiliser sur des variables qualitatives après avoir réalisé une analyse factorielle au préalable. L'exploitation de l'analyse des correspondances multiples réalisée en amont, mais seulement sur les variables identifiés par les forêts aléatoires cette fois ci, va permettre d'avoir une première idée visuelle des différents groupes. Le résultat sera par ailleurs plus stable car tronqué du bruit non conservé lors du choix du nombre de dimensions.

Figure 50 – Répartition des modalités des variables



En reprenant l'analyse des correspondances multiples initiale, ciblée cette fois-ci sur les quatre principaux facteurs, la forme du nuage de points a légèrement changé. Les axes restent les mêmes, à savoir la taille du risque pour la dimension 2 et la variation du profil de risque pour la dimension 1.

Figure 51 – Nuage des modalités et qualité de représentation



La représentation des modalités des variables permet de donner de précieuses informations sur le portefeuille. La construction de cette figure est réalisée à l'aide des distances euclidiennes, et les modalités représentent le barycentre des observations qui les possèdent. L'ordonnée à l'origine étant le barycentre du nuage contenant toutes les observations, les individus se trouvant loin de l'origine possèdent des caractéristiques rares dans la base de données. Pour l'étape suivante de classification, deux groupes se détachent visuellement, un, en haut à gauche et le deuxième en diagonale.

3.4.6.2. Principes de la Classification Ascendante Hiérarchique

La classification ascendante hiérarchique tient son nom de la façon dont la classification est réalisée. A la différence des arbres de décisions, ce n'est pas un découpage de la base mais un regroupement de plusieurs classes. Tout comme pour les k-Means, le regroupement utilisant la méthode de Ward fonctionne sur la base du théorème de Huygens. En partant d'un individu représentant une classe, on a une inertie intra classe de 1. L'agrégation avec une autre classe se fera en cherchant à diminuer la perte d'inertie. Cette opération revient à minimiser l'inertie interclasse définie dans la formule ci-dessous cf. Husson, Lê, Pages [2009]

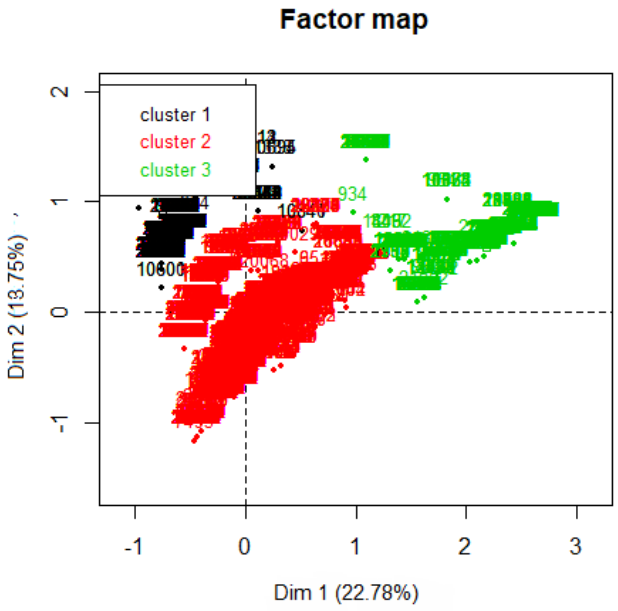
$$Inertie(a) + Inertie(b) = Inertie(a \cup b) - \frac{m_a m_b}{m_a + m_b} d^2(a,b)$$

- Avec m = le nombre d'observations de la classe
- d^2 = la distance entre les centres de gravité des classes

3.4.6.3. Application de la Classification Ascendante Hiérarchique

Le dendrogramme reprend d'une manière visuelle les regroupements réalisés par l'algorithme. Un choix doit ensuite être fait pour déterminer à quel niveau se fera le découpage. La meilleure coupe possible est obtenue, au regard des gains d'inerties en fonction du nombre de *clusters*, qui est une mesure de la qualité du partitionnement. Le découpage optimal donne trois *clusters :*

Figure 52 – Découpage optimal des facteurs de l'ACM



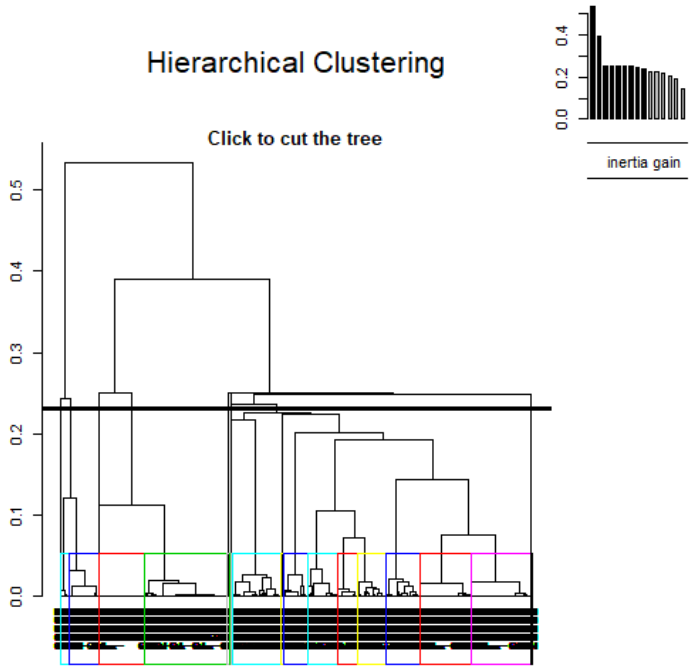
La description des principales modalités de ces trois *clusters* peut être observée à *l'Annexe 12 – Description des 3 clusters suite à l'agrégation optimale du CAH.* En ayant en tête le nuage des modalités de l'ACM, on remarque sur la carte factorielle ci-dessus que l'agrégation optimale donnerait trois *clusters* correspondant plus ou moins aux segments des entreprises :

- Le *cluster* 1 aux risques atypiques
- Le *cluster* 2 aux petites et moyennes entreprises qui partagent des profils similaires
- Le *cluster* 3 aux grandes entreprises multinationales

L'objectif ici n'est pas de regrouper les observations en groupes homogènes, mais en revanche, d'identifier à un niveau relativement granulaire, les groupes possédant une grande majorité de mauvais comptes. L'algorithme des k-Means a montré précédemment, que ces comptes étaient plus ou moins équiréparties au sein des différents profils de clients. L'analyse du découpage optimal ne permettrait donc pas une précision nécessaire pour identifier les zones nécessitant la mise en place d'actions correctives du portefeuille.

En choisissant un découpage au-dessus de la partie volatile composée d'un grand nombre de petits groupes, l'algorithme identifie 21 *clusters*, visibles sur le dendrogramme ci-dessous :

Figure 53 - Dendrogramme



La description des principales modalités de ces *clusters*, permettra l'identification des principaux profils des clients qui possèdent de mauvais résultats. La description complète est disponible à l'*Annexe 13 – Description des 21 clusters.* Ne sont gardés ci-dessous que les *clusters* contenant une majorité de comptes non-profitables par soucis de lisibilité. C'est-à-dire les *clusters 5, 8, 13 et 18.*

Table 18 - Description des principales modalités des clusters identifiés

```
Description of each cluster by the categories
============================================
$`5`
                          Cla/Mod   Mod/Cla    Global      p.value     v.test
Commodity=Commodity_9 100.0000000 100.00000  0.633175 0.000000e+00        Inf
Limit=Limit_6           1.0277101  88.81579 54.719653 9.305047e-20   9.096779
IV=IV_5                 1.1780510  75.00000 40.310756 3.851594e-18   8.683090
Segment=Segment_3       0.9741050  78.94737 51.316338 1.902244e-12   7.041468
Category=Category_0     1.4795474  11.18421  4.786303 1.363721e-03   3.202221
Limit=Limit_3           0.0000000   0.00000  2.203616 3.344778e-02  -2.126667
Category=Category_1     0.5906287  88.81579 95.213697 1.363721e-03  -3.202221
Commodity=Commodity_6   0.0000000   0.00000  5.977672 8.276654e-05  -3.936245
Segment=Segment_1       0.0000000   0.00000  8.439557 1.446696e-06  -4.818461
Segment=Segment_2       0.3312287  21.05263 40.244106 5.085903e-07  -5.023043
IV=IV_1                 0.0000000   0.00000  9.710072 1.716909e-07  -5.227636
Commodity=Commodity_3   0.0000000   0.00000  9.930851 1.181866e-07  -5.296278
Limit=Limit_4           0.0000000   0.00000 15.287845 1.023554e-11  -6.803151
IV=IV_7                 0.0000000   0.00000 16.791635 6.657176e-13  -7.186298
Limit=Limit_7           0.0000000   0.00000 18.441223 3.136048e-14  -7.592577
Commodity=Commodity_5   0.0000000   0.00000 20.390736 7.816962e-16  -8.057031
Commodity=Commodity_4   0.0000000   0.00000 28.226277 1.058501e-22  -9.806229
Commodity=Commodity_1   0.0000000   0.00000 33.837374 4.230504e-28 -10.990823
```

```
$`8`
                        Cla/Mod    Mod/Cla     Global       p.value     v.test
IV=IV_3                80.675147 100.000000  8.5145380  0.000000e+00        Inf
Commodity=Commodity_5  11.235955  33.353548 20.3907356  2.077625e-37  12.781596
Segment=Segment_3       7.614254  56.882959 51.3163376  2.689160e-06   4.693237
Category=Category_0     8.790252   6.124924  4.7863034  1.069611e-02   2.552473
Segment=Segment_2       7.359487  43.117041 40.2441056  1.398851e-02   2.457558
Limit=Limit_4           7.820163  17.404488 15.2878447  1.470584e-02   2.439543
Limit=Limit_2           0.000000   0.000000  0.1874531  4.054104e-02  -2.048193
Commodity=Commodity_8   0.000000   0.000000  0.2041156  3.048043e-02  -2.163790
Commodity=Commodity_4   6.242621  25.651910 28.2262768  1.536260e-02  -2.423716
Category=Category_1     6.772542  93.875076 95.2136966  1.069611e-02  -2.552473
Commodity=Commodity_10  0.000000   0.000000  0.3790719  1.520724e-03  -3.170699
Commodity=Commodity_7   0.000000   0.000000  0.3832375  1.415867e-03  -3.191397
Commodity=Commodity_9   0.000000   0.000000  0.6331750  1.935825e-05  -4.272167
Limit=Limit_5           4.549632   6.003639  9.0644006  2.381652e-06  -4.718008
Commodity=Commodity_3   3.859060   5.579139  9.9308506  5.110873e-11  -6.567670
Commodity=Commodity_6   0.000000   0.000000  5.9776722  1.648538e-46 -14.319691
Segment=Segment_1       0.000000   0.000000  8.4395568  2.990702e-66 -17.193106
IV=IV_1                 0.000000   0.000000  9.7100725  1.176866e-76 -18.530274
IV=IV_4                 0.000000   0.000000 10.9555944  5.265135e-87 -19.771254
IV=IV_2                 0.000000   0.000000 13.6965759 2.540082e-110 -22.316767
IV=IV_7                 0.000000   0.000000 16.7916354 1.347596e-137 -24.968413
IV=IV_5                 0.000000   0.000000 40.3107556  0.000000e+00       -Inf

$`13`
                          Cla/Mod     Mod/Cla     Global       p.value     v.test
Segment=Segment_2      16.97546838  99.8174072 40.2441056  0.000000e+00        Inf
Limit=Limit_5          75.50551471 100.0000000  9.0644006  0.000000e+00        Inf
IV=IV_2                16.78832117  33.5970785 13.6965759 4.279896e-102  21.453015
Commodity=Commodity_3  18.24664430  26.4759586  9.9308506  3.322265e-89  20.025192
Commodity=Commodity_4   8.14639906  33.5970785 28.2262768  8.123451e-07   4.932378
Commodity=Commodity_5   8.41675179  25.0760803 20.3907356  1.836073e-06   4.770677
Category=Category_0     9.05134900   6.3298844  4.7863034  3.498107e-03   2.920197
Limit=Limit_2           0.00000000   0.0000000  0.1874531  4.103404e-02  -2.043186
IV=IV_3                 5.72407045   7.1211199  8.5145380  3.278275e-02  -2.134734
Commodity=Commodity_8   0.00000000   0.0000000  0.2041156  3.088429e-02  -2.158560
Category=Category_1     6.73316708  93.6701156 95.2136966  3.498107e-03  -2.920197
Commodity=Commodity_10  0.00000000   0.0000000  0.3790719  1.558392e-03  -3.163584
Commodity=Commodity_7   0.00000000   0.0000000  0.3832375  1.451328e-03  -3.184244
Commodity=Commodity_9   0.00000000   0.0000000  0.6331750  2.016693e-05  -4.263035
IV=IV_1                 4.20420420   5.9646987  9.7100725  1.715381e-08  -5.638500
IV=IV_4                 4.03041825   6.4516129 10.9555944  1.071015e-10  -6.456572
Limit=Limit_3           0.00000000   0.0000000  2.2036158  3.339889e-17  -8.434020
IV=IV_7                 3.12577524   7.6688984 16.7916354  5.429965e-29 -11.174591
Commodity=Commodity_6   0.00000000   0.0000000  5.9776722  2.454331e-46 -14.292007
Segment=Segment_1       0.00000000   0.0000000  8.4395568  5.287632e-66 -17.160041
Commodity=Commodity_1   3.00381632  14.8508825 33.8373740  3.373813e-72 -17.969560
Limit=Limit_4           0.00000000   0.0000000 15.2878447 9.523105e-124 -23.659058
Limit=Limit_7           0.00000000   0.0000000 18.4412230 5.386928e-152 -26.260756
Segment=Segment_3       0.02435263   0.1825928 51.3163376  0.000000e+00       -Inf
Limit=Limit_6           0.00000000   0.0000000 54.7196534  0.000000e+00       -Inf
```

```
$`18`
                          Cla/Mod     Mod/Cla      Global       p.value    v.test
Limit=Limit_2          100.00000000 100.000000  0.1874531 9.579482e-142 25.347483
Commodity=Commodity_3    1.17449664  62.222222  9.9308506  1.539349e-17  8.524151
IV=IV_3                  0.83170254  37.777778  8.5145380  7.561728e-08  5.377285
Segment=Segment_1        0.69101678  31.111111  8.4395568  1.432438e-05  4.338829
IV=IV_1                  0.60060060  31.111111  9.7100725  7.086319e-05  3.973369
Segment=Segment_2        0.32087776  68.888889 40.2441056  1.214951e-04  3.843090
Category=Category_0      0.60922541  15.555556  4.7863034  6.405004e-03  2.726293
Limit=Limit_5            0.00000000   0.000000  9.0644006  1.384329e-02 -2.461304
Category=Category_1      0.16625104  84.444444 95.2136966  6.405004e-03 -2.726293
Commodity=Commodity_5    0.04085802   4.444444 20.3907356  3.123480e-03 -2.955317
Limit=Limit_4            0.00000000   0.000000 15.2878447  5.679647e-04 -3.446469
IV=IV_7                  0.00000000   0.000000 16.7916354  2.534517e-04 -3.658747
Limit=Limit_7            0.00000000   0.000000 18.4412230  1.028316e-04 -3.883812
Commodity=Commodity_1    0.01231072   2.222222 33.8373740  2.078235e-07 -5.192200
IV=IV_5                  0.01033378   2.222222 40.3107556  2.599433e-09 -5.955081
Segment=Segment_3        0.00000000   0.000000 51.3163376  8.190898e-15 -7.764591
Limit=Limit_6            0.00000000   0.000000 54.7196534  3.120697e-16 -8.168569
```

Table 19 – Nombre d'observations par modalités et par clusters ex : Cluster 5

	Cluster_5	Autres	Total
IV_5	114	9,563	9,677
Autres	38	14,291	14,329
Total	152	23,854	24,006

La description des *clusters* donne des informations sur les proportions et les modalités des observations appartenant à chaque *cluster*. En prenant l'exemple du *cluster* 5 et de l'IV_5, la colonne Cla/Mod, représente le pourcentage de la modalité IV_5 présente dans le *cluster_5*. Ici seulement 1.17% d'observations possédant la modalité IV_5 appartiennent au *cluster_5*.

$$Cla/Mod = \frac{114}{9677} = 1.17\%$$

La colonne Mod/Cla représente le pourcentage d'individus du *cluster_5* possédant la modalité IV_5. Ici 75% d'observations du *cluster_5* sont composées de l'IV_5.

$$Mod/Cla = \frac{114}{152} = 75.00\%$$

Enfin, la colonne Global représente la proportion de la modalité IV_5 dans la totalité du portefeuille. L'IV_5 est une modalité très présente dans le portefeuille puisque 40.31% d'observations possèdent cette modalité.

$$Global = \frac{9677}{24006} = 40.31\%$$

Un test est réalisé et la p-value est là pour indiquer si la proportion de l'échantillon incluse dans le *cluster* est normale ou pas.

$$Valeur - test = \frac{\overline{X_q} - \bar{x}}{\sqrt{\frac{s^2}{n_q}\left(\frac{N - n_q}{N - 1}\right)}} \sim \mathcal{N}(0,1)$$

La valeur-test traduit le résultat du test pour indiquer si cet échantillon est sur-représenté avec une valeur positive ou sous-représenté avec une valeur négative. La modalité caractérise d'autant mieux le cluster que sa valeur-test est grande. Seule les variables caractéristiques dont la valeur absolue de la valeur-test est supérieure à 1.96 sont gardées en sortie.

Dans les *clusters 5, 8, 13 et 18,* la quantité de *category_0*, indique que les comptes non-profitables, sont sur-représentés dans ces clusters. Encore une fois, ils ne représentent qu'une faible proportion d'observations au sein de leur *cluster*, avec respectivement 11%, 6%, 6% et 15%. Cela est dû au faible nombre de comptes non-profitables dans le portefeuille global à savoir 4.7%.

En reprenant le ratio S/P des profils identifiés, il s'avère que les *clusters* 13 et 18 ne correspondent pas à des profils sinistrés.

Table 20 - Ratio S/P des clusters contenant une grande proportion de comptes non-profitables

	Commodity	Limit	IV	Segment	Initial LR
Cluster 5	9	6	5	3	122.1%
Cluster 8	5	4	3	3	197.1%
Cluster 13	3	5	2	2	21.1%
Cluster 13 bis	4	5	2	2	26.2%
Cluster 13 ter	5	5	2	2	23.8%
Cluster 18	3	2	3	1	0.0%

3.4.6.4. Conclusion

La classification ascendante hiérarchique est un algorithme polyvalent, qui permet de choisir le degré d'homogénéité des groupes obtenus en fonction de l'utilisation voulue. Contrairement aux k-Means, la difficulté et le risque de paramétrisation sont moindres.

L'étude des groupes permet de faire ressortir la sur-représentation ou sous-représentation d'une modalité au sein de l'échantillonnage. Cela a permis d'identifier quatre *clusters*, où la part des comptes non-profitables était prépondérante ainsi que les modalités qu'ils possèdent. Il faut cependant être vigilant sur les conclusions, puisqu'en reprenant le ratio S/P des *clusters*, il s'avère que la mutualisation avec le reste des comptes permet d'avoir une profitabilité convenable.

Des actions correctives seront mises en place sur les deux premiers profils, en partant du principe que la sur-représentation de ces comptes ne permet plus une mutualisation suffisante avec les autres comptes d'un même groupe.

4. PRISE DE DECISIONS

Au-delà de la simple résiliation des comptes appartenant aux poches de sous tarification, plusieurs possibilités s'offrent aux gestionnaires de portefeuille. En effet, l'étude des causes de sinistres pourrait également permettre d'agir directement sur la sinistralité en mettant en place des actions préventives, telles que le changement des franchises ou la mise en place de clauses d'exclusions. Une hausse de la prime peut être réalisée au cas par cas mais un recalibrage du modèle de tarification peut permettre d'éviter que les résultats continuent de se dégrader dans le futur. C'est ce dernier cas qui sera étudié par la suite afin de valider les résultats de la segmentation du portefeuille.

4.1. Modèle de tarification

La mauvaise performance d'un compte peut être due à un niveau de prime trop faible provenant d'une concurrence forte sur le marché, à de mauvaises pratiques des souscripteurs ou encore due à une mauvaise adéquation du modèle. C'est ce dernier cas qui a été identifié par le CART, puisque le modèle se révèle trop complexe. Les bénéfices d'une recalibration du modèle de tarification seront démontrés dans la suite de cette partie.

4.1.1. Théorie

Du fait de l'inversion du cycle de production en assurance, la tarification consiste à déterminer la charge future des sinistres sur une période de temps donnée. Dénommée prime pure, l'assureur rajoutera à cette somme des chargements pour couvrir ses frais fixes et variables ainsi que sa marge technique pour déterminer la prime commerciale. Le calcul de la prime pure peut se faire de manière plus ou moins sophistiqué selon la ligne d'activité et des données disponibles.

Dans la pratique, la prime souscrite ne provient pas toujours de modèles stochastiques. Cela est d'autant plus vrai dans les marchés où la capacité d'assurance est excédentaire à la demande et où le prix est tiré par la concurrence. En assurance transport, deux méthodes de tarifications actuarielles coexistent et seront détaillées par la suite.

4.1.1.1. Tarification par expérience

La technique la plus simple est la tarification par expérience, aussi connue sous le nom de *Burning Cost*. Cette technique consiste à utiliser l'expérience de sinistralité du propre compte pour déterminer le prix. Ainsi une hypothèse forte est prise quant à la représentativité des résultats passés pour prédire le futur. La prime pure est dans ce cas une moyenne des pertes du passée, ajustée pour tenir compte des changements de l'exposition et de l'inflation.

$$Burning\ Cost = \frac{1}{n}\sum_{j=1}^{n}\frac{\sum_{i=1}^{mj} Rev(X_{i,j})}{Ej}$$

Avec :

- $X_{i,j}$ le montant de l'$i^{ème}$ sinistre de la $j^{ème}$ année
- $Rev(X_{i,j}) = X_{i,j} \cdot (1+t)^{N-n}$ le montant réévalué pour tenir compte d'une inflation constante t.
- m_j Le nombre de sinistres pour l'année j,
- E_j L'exposition relative à l'année j et n le nombre d'années statistiques utilisé.

Le *Burning Cost* est relativement simple à calculer, mais présente des limitations du fait qu'il se base exclusivement sur les sinistres du passé. Il permet de prendre en compte, seulement dans une certaine mesure, les changements dans les conditions générales du contrat. Par exemple, seulement les changements de limites ou de franchises des périls étant déjà survenus dans le passé peuvent être quantifiés. Il est par ailleurs nécessaire de faire appel à des méthodes de modélisation statistique afin de quantifier le prix des tranches du contrat n'ayant jamais été consommées par le passé, en extrapolant la courbe de sévérité et de fréquence des sinistres.

4.1.1.2. Tarification par exposition

Pour construire une tarification par exposition, il est nécessaire de recourir à une base de données conséquente. A l'aide de cette base, un modèle benchmark est construit pour la tarification des autres comptes. Il est donc plus fréquent de voir ces modèles en assurance pour particuliers qu'en assurance pour entreprise. C'est notamment le cas en assurance santé ou en assurance automobile.

En assurance transport ou en assurance corps de navire, des modèles de tarification par exposition peuvent être trouvés. En transport, ces modèles se basent souvent sur un taux de base par type de marchandises multiplié par la valeur transportée. Ce taux de base est augmenté ou réduit par des facteurs de risques tels que le mode de transport, la destination, le mode de stockage, le mode conditionnement etc.

La liste des facteurs peut être très longue, néanmoins, il est dans la pratique difficile d'évaluer précisément le risque dans une police transport annuelle sachant la multiplicité des marchandises transportées, des destinations, ou des modes de transport. Il est donc nécessaire de ne retenir dans la construction du modèle que les facteurs de risques significatifs. Bien souvent, les modèles de tarification par exposition sont construits à l'aide de modèles linéaires généralisés qui présentent l'avantage de montrer les relativités des facteurs de risques sur la prime.

4.1.2. Modèle Linéaire Généralisé

4.1.2.1. Principes du modèle linéaire généralisé

Le modèle linéaire généralisé ou GLM permet d'estimer une variable aléatoire telle que le montant de sinistres. Il étudie la liaison entre cette variable quantitative avec une ou plusieurs variables. Il est donc particulièrement utile en assurance pour expliquer la sinistralité en fonction des facteurs de risques de l'assuré pour déterminer la prime pure. La mise en œuvre cherche donc à montrer un lien entre une variable réponse avec une ou plusieurs variables explicatives. Cela revient à ajuster une équation de la forme suivante cf. Denuit et Charpentier (2005) :

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

Les modèles linéaires généralisés sont formés de trois composantes :

Composante aléatoire

La variable à expliquer Y à laquelle on associe une loi de probabilité, qui est nécessairement de la famille exponentielle.

Soit $(Y_1, \dots, Y_n)$ un échantillon de taille n de la variable Y, alors les variables aléatoires $(Y_1, \dots, Y_n)$ sont supposées indépendantes. On cherche donc une loi pouvant s'écrire sous la forme :

$$f(y|\theta, \varphi) = exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi)\right\}$$

Où $a(.), b(.)$et $c(.)$ sont des fonctions, θ le paramètre d'intérêt et φ le paramètre de nuisance.

Composante déterministe

Les variables explicatives $X_1, \dots, X_n$ sont utilisées comme prédicteur et définissent sous forme d'une combinaison linéaire $\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$ la composante déterministe Y.

Fonction de lien

La fonction lien permet de linéariser le phénomène modélisé et décrit la relation fonctionnelle entre les variables explicatives et la variable réponse. Il spécifie comment l'espérance mathématique de Y notée μ est liée au prédicteur linéaire construit à partir des variables explicatives. Finalement le modèle à ajuster est de la forme suivante :

$$f(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

Où f représente notre fonction lien, et μ l'espérance mathématique de Y.

4.1.2.2. Application du modèle linéaire généralisé

Pour la modélisation du montant de sinistres attendu, deux modèles linéaires généralisés sont construits. Un modèle estimera la fréquence de sinistres et un autre la sévérité moyenne des sinistres. Nous faisons le choix de ne retenir que les facteurs de risques identifiés préalablement par les forêts aléatoires. Il se peut ainsi que ces facteurs obtenus par une méthode non linéaire puissent être identifiés comme non significatifs par le modèle linéaire généralisé.

Les lois qui seront retenues sont celles préexistantes dans le modèle de tarification initial, afin d'avoir une base de comparaison similaire. Ainsi l'impact de la sélection des variables par un modèle non paramétrique est isolé et la variance induite par un changement de loi n'est pas prise en compte. Bien sur le choix optimal des lois doit faire l'objet d'une revue régulière lors de la révision du modèle de tarification mais cela ne sera pas abordée dans cette étude.

Fréquence

Parmi les lois de la famille exponentielle, la loi de poisson est choisie afin de modéliser la fréquence.

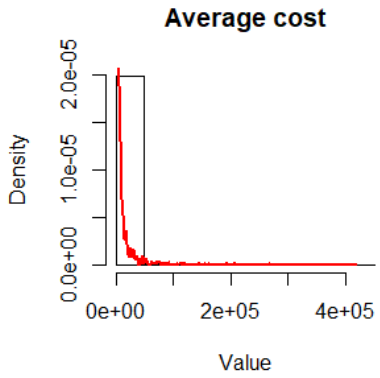
```
Call:
glm(formula = ClaimsCount ~ Commodity + Limit + IV + Segment,
    family = poisson(link = "log"), data = dt)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.877  -1.420  -1.023   0.073  43.612

             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.594787   0.034789  45.842  < 2e-16 ***
Commodity   -0.040653   0.003451 -11.778  < 2e-16 ***
Limit       -0.025802   0.007190  -3.589 0.000333 ***
IV          -0.332796   0.003974 -83.734  < 2e-16 ***
Segment      0.117838   0.010912  10.799  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 123901  on 24005  degrees of freedom
Residual deviance: 113126  on 24001  degrees of freedom

AIC: 141247
Number of Fisher Scoring iterations: 7
```

Pour la modélisation de la fréquence, tous les variables précédemment identifiés par les forêts aléatoires sont reconnues comme très significatifs avec un risque de première espèce égal à 0%.

Sévérité

Sachant qu'il existe peu de volatilité dans les montants de sinistres au vu de la Figure 58, le *GLM* va être utilisé pour modéliser un coût moyen. Les sinistres atypiques ne seront pas modélisés et pourront expliquer une certaine déviation la sinistralité sur la prime mais la mutualisation des risques devrait permettre de limiter son impact sur le résultat.

Figure 54 - Densité du coût moyen des sinistres



Parmi les lois de la famille exponentielle, la loi gamma est choisie pour modéliser le coût moyen.

```
Call:
glm(formula = avgcost ~ Commodity + Limit + IV + Segment, family = Gamma(link = "log"),
    data = dataglm)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8816  -1.7320  -1.0388  -0.2868  13.5125

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.56204    0.33534  31.496  < 2e-16 ***
Commodity   -0.01195    0.02924  -0.409    0.683
Limit       -0.28740    0.06561  -4.380 1.20e-05 ***
IV          -0.21670    0.03373  -6.424 1.41e-10 ***
Segment      0.04455    0.08997   0.495    0.621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 15.40692)
    Null deviance: 20623  on 7015  degrees of freedom
Residual deviance: 18552  on 7011  degrees of freedom

AIC: 124047
Number of Fisher Scoring iterations: 12
```

Pour la sévérité, contrairement aux forêts aléatoires, la marchandise et le *segment* de l'entreprise ne sont pas jugés comme des variables significatives par le GLM. Ces variables seront gardées dans le modèle et un second modèle sera construit avec les variables proposées par le GLM afin de mesurer le bénéfice d'une sélection préalable par les forêts aléatoires.

```
Call:
glm(formula = avgcost ~ Limit + IV, family = Gamma(link = "log"),
    data = dataglm)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8793  -1.7355  -1.0364  -0.2824  13.5485

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.49956    0.26817  39.152  < 2e-16 ***
Limit       -0.26857    0.05711  -4.703 2.61e-06 ***
IV          -0.21138    0.03266  -6.472 1.03e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 15.40304)

    Null deviance: 20623  on 7015  degrees of freedom
Residual deviance: 18553  on 7013  degrees of freedom
AIC: 124044
Number of Fisher Scoring iterations: 10
```

Un deuxième modèle de sévérité avec seulement deux variables est obtenu. Avec la combinaison de ces trois modèles linéaires généralisés, deux modèles sont construits :

- Modèle 1 appelé « GLM » : Fréquence avec 4 variables et une Sévérité avec 4 variables
- Modèle 2 appelé « GLM 2 » : Fréquence avec 4 variables et une Sévérité avec 2 variables

4.2. Test de validation

4.2.1. Description de l'approche

Afin de valider l'impact du recalibrage du modèle, la prime sera recalculée pour l'ensemble des comptes et comparée avec la sinistralité historique. Ceci n'est qu'une simulation historique puisque la rétention client n'est pas prise en compte. Il faut cependant s'assurer qu'il n'y ait pas trop de dispersion dans les différents *segments* de clients. Le *Mean Square Error* ou *MSE* sera choisi pour mesurer la distance entre la prime modélisée avec la sinistralité historique. Cet indicateur est défini comme :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Plusieurs modèles seront analysés et comparés avec le ratio S/P actuel :

- *Original Model* : Modèle actuel donnant la prime technique
- *GLM* : Modèle proposé avec des variables sélectionnées par les forêts aléatoires
- *GLM 2* : Modèle proposé avec les variables sélectionnées par le GLM
- *Random Forest* : Forêts aléatoires qui avaient été utilisées pour la sélection des variables
- *Partial Model KM* : Modèle actuel avec la correction des facteurs identifiés par les k-Means
- *Partial Model CAH* : Modèle actuel avec la correction des facteurs identifiés par le CAH

Bien que l'utilisation des méthodes d'apprentissage dans un but de tarification, ne soit pas l'objet principal du mémoire, les forêts aléatoires seront quand même utilisées pour vérifier leurs capacités prédictives. Il faut cependant noter, que l'algorithme n'a pas été spécialement optimisé et que d'autres types de méthodes d'apprentissage peuvent probablement donner de meilleurs résultats tels que l'*Extreme Gradient Boosting* ou encore les *Neural Networks*. Ces méthodes non linéaires permettent entre autres de mieux estimer les sinistres atypiques.

Enfin, les comptes appartenant aux zones de sous-tarifications identifiées par les k-Means et la CAH vont faire l'objet d'une hausse de prime dont le pourcentage est déterminé afin d'obtenir le ratio S/P cible sur ces profils. Seulement les zones contenant une majorité de comptes non-profitables sont sélectionnées afin de ne pas impacter négativement les comptes profitables.

Table 21 - Taux de correction appliqués

	Commodity	Limit	IV	Segment	Initial LR	Revised LR	Revised Factor
Cluster 5	9	6	5	3	122%	57%	+114%
Cluster 8	5	4	3	3	197%	57%	+246%
kM 18	4	3	5	2	176%	57%	+210%
kM 31	4	2	1	1	306%	57%	+437%
kM 36	1	6	1	1	198%	57%	+247%

4.2.2. Résultats de la re-tarification

L'analyse des résultats montre clairement une meilleure performance des forêts aléatoires, avec à la fois un ratio S/P plus bas, mais également une volatilité plus faible. Le modèle proposé « GLM », avec seulement 4 facteurs, donne un ratio S/P plus élevé que le modèle actuel « Original Model » mais avec une volatilité plus faible. A l'inverse les modèles bénéficiant d'un rehaussement des primes des profils trop sinistrés, montrent une légère amélioration du résultat. Cela s'explique par la part marginale des comptes non-profitables dans le portefeuille global. Enfin, le modèle proposé « GLM 2 » dont les variables ont été sélectionné avec le GLM, affiche une plus grande volatilité que le modèle dont les facteurs ont été déterminé par les forêts aléatoires. Cette présélection par modèle non paramétrique permet une meilleure adéquation.

Table 22 – Résultats global des différents modèles de tarification

	Actual LR	Original Model	GLM	GLM 2	Random Forest	Partial Model KM	Partial Model CAH
LR	36.2%	39.3%	39.9%	39.9%	23.4%	38.5%	39.2%
MSE		935,764,161	926,686,560	930,243,038	539,439,479	937,746,226	935,829,433

Bien que le *MSE* donne une idée de la qualité de l'ajustement global, il ne permet pas de savoir dans quel cas l'estimation est moins précise. Dans ce cas, la décomposition des résultats selon les trois *segments* permettra de mieux visualiser si la prédiction est proche de la sinistralité réelle selon la taille des clients. En effet, l'analyse en correspondance multiples avait montré que les profils de clients pouvaient différer fortement selon le *segment*.

Table 23 - Résultats des différents modèles de tarification pour les entreprises multinationales

Premium Volume		31,977,809					
Segment 1	**Actual LR**	**Original Model**	**GLM**	**GLM 2**	**Random Forest**	**Partial Model KM**	**Partial Model CAH**
LR	55%	26%	29%	27%	22%	25%	26%
MSE		346,953,366	335,543,746	341,104,833	236,990,026	348,930,134	346,953,366

Table 24 - Résultats des différents modèles de tarification pour les moyennes entreprises

Premium Volume		86,052,011					
Segment 2	**Actual LR**	**Original Model**	**GLM**	**GLM 2**	**Random Forest**	**Partial Model KM**	**Partial Model CAH**
LR	34%	41%	41%	42%	27%	41%	41%
MSE		169,693,156	171,660,117	169,848,205	93,677,267	169,698,452	169,693,156

Table 25 - Résultats des différents modèles de tarification pour les petites entreprises

Premium Volume		77,322,701					
Segment 3	**Actual LR**	**Original Model**	**GLM**	**GLM 2**	**Random Forest**	**Partial Model KM**	**Partial Model CAH**
LR	31%	57%	53%	57%	21%	57%	56%
MSE		419,117,640	419,482,697	419,289,999	208,772,186	419,117,640	419,182,911

L'analyse par *segment* de la prédiction des modèles est intéressante car elle montre que le ratio S/P est relativement élevé pour les entreprises multinationales. Ce *segment* étant très exposés à des sinistres atypiques, cela n'est en revanche pas la raison des mauvais résultats puisqu'on observe que le modèle original et les autres apportent une bonne qualité de prédiction avec des ratio S/P plus faibles partout. Ce ratio S/P est donc expliqué soit par des marges plus faibles car les clients multinationaux ont une force de négociation plus puissante, ou parce que le souscripteur donne trop de rabais dû à la concurrence accrue sur ces comptes qui ont un intérêt marketing avec le prestige de les assurer et ouvrent des opportunités de ventes croisées.

Le *segment* des moyennes entreprises est celui où la volatilité des modèles est la plus faible. Le modèle simplifié à 4 facteurs n'apporte pas un grand intérêt avec en sus, une plus grande volatilité. Avec le *segment* des petites entreprises, le modèle simplifié donne une meilleure estimation ce qui permet de réduire le ratio S/P sans trop perdre en stabilité avec un *MSE* relativement similaire.

L'analyse des résultats montre que le portefeuille historique surperforme les modèles de tarification pour les petites et moyennes entreprises car le ratio S/P est plus faible que celui prédit. En revanche, la situation s'inverse avec les entreprises multinationales, où les résultats sont plus élevés probablement dues à la concurrence. L'utilisation des méthodes non-linaires pour la tarification est prometteuse car les forêts aléatoires donnent de bonnes estimations avec une volatilité plus faible et ce, quel que soit le *segment*. A mon sens, son emploi sera plus fréquent lorsque nous changerons du paradigme de la mutualisation des risques vers une tarification individuelle au plus juste avec la collecte de données externes, pour enrichir le modèle, avec par exemple le *web scrapping* ou l'accès à *l'open data*.

Enfin il est intéressant de voir que le modèle appelé « GLM » dont les variables ont été sélectionné par les forêts aléatoires offre de meilleurs résultats que le modèle « GLM 2 » dont la sélection des variables est issue d'un méthode paramétrique. En effet, bien que le résultat est plus ou moins similaire au global, le ratio S/P du modèle « GLM 2 » devient beaucoup plus élevé pour les petites et moyennes entreprises où le type de marchandises transporté peut différer significativement. Pour les entreprises multinationales dont la sinistralité dépend beaucoup plus de la limite et de la valeur assurée, ce modèle arrive à tirer son épingle du jeu.

Notre proposition d'un modèle de tarification simplifié donne des résultats en amélioration avec une meilleure appréhension des grands risques des entreprises multinationales. Pour ce qui est de la correction partielle du modèle avec les zones de sous tarification identifiées par les *k-Means* et le CAH, il serait intéressant de voir l'impact sur un portefeuille plus sinistré. En effet, la correction de ces petits *clusters* ne permet pas d'observer une amélioration matérielle du modèle de tarification.

CONCLUSION

Avec l'essor du *Big Data,* l'utilisation des algorithmes d'apprentissage a été facilitée et popularisée dans la littérature actuarielle ces dernières années. Ils restent en pratique peu utilisés en tarification dans les compagnies d'assurance traditionnelles. Les travaux de ce mémoire proposent une utilisation de ces méthodes pour le pilotage de la rentabilité d'un portefeuille de risques spéciaux en assurance transport. Cela se révèle d'autant plus important que le marché international souffre de mauvais résultats due à la baisse des primes d'assurances, conjuguée à une inflation des sinistres et à une accumulation des marchandises avec la mondialisation.

Le département actuariel de Chubb étant une fonction centralisée à Paris pour la région Europe continentale, l'étude se concentre sur un seul pays disposant d'un portefeuille suffisamment grand pour être exploité seul afin de ne pas créer de dispersion additionnelle avec l'hétérogénéité des pratiques de tarification. L'intégration d'une dimension de retour sur capitaux économiques dans le pilotage, a cependant montré que le pays possédait un portefeuille en bonne santé avec une part faible de comptes non-profitables.

La première phase de l'étude a consisté à permettre une visualisation des données améliorée. L'utilisation de méthodes d'analyses factorielles a permis d'offrir une bonne synthétisation de l'information et de la corrélation de certaines variables entre elles. La variable « destination » a notamment pu être supprimée à la suite de ces différentes analyses. Par la suite, l'application d'un algorithme de suréchantillonnage synthétique, dit *SMOTE,* a permis de palier au déséquilibre de la base de données pour améliorer la prédiction des arbres de décisions. Le *CART* a permis de mettre en lumière un modèle de tarification trop complexe par rapport à la sinistralité réelle. En revanche, l'algorithme montre également que les bons résultats sont expliqués par une adéquation de la prime souscrite avec la sinistralité, et donc un effort rigoureux de souscription. L'extension vers les forêts aléatoires, a été utile pour la sélection des variables significatives dans le reste du mémoire. A la fois pour réduire les dimensions des algorithmes d'apprentissages, mais également pour le choix des variables d'un modèle *GLM* alternatif qui s'est avéré meilleur qu'un modèle *GLM* avec une sélection par une méthode paramétrique dans la phase de re-tarification.

Après avoir calculé le profil de risque pour intégrer la rentabilité du capital économique dans le pilotage, la deuxième phase de l'étude a consisté à étudier la segmentation du portefeuille. L'utilisation des *SVM* s'est révélée non concluante. En plus d'être un algorithme supervisé, difficile à calibrer, la visualisation du résultat n'a pas été aisée, que ce soit en 3D (représentation non aboutie) ou en 2D (représentation tronquée). Dans les cas de risques comportant plus de dimensions, la qualité de la visualisation serait encore plus aggravée. L'exploration de l'algorithme des *k-Means* par la suite a

permis de montrer que les comptes non profitables étaient équirépartis dans le portefeuille. Des profils contenant une part prépondérante de comptes profitables et non-profitables ont également pu être déterminé pour une prise de décision ultérieure. En fin de compte, les *k-Means* ont donné des résultats intéressants, mais le risque de sur-apprentissage est grand, car il est difficile de calibrer cet algorithme qui nécessite de fixer à priori un nombre de *clusters*. Les résultats peuvent également diverger d'une fois sur l'autre, car la convergence vers l'optimalité du découpage dépend de la localisation initiale des barycentres des groupes. Enfin, l'utilisation de la CAH a également donné des profils de comptes profitables et non-profitables, mais différents de l'algorithme des *k-Means.* Il s'est montré relativement facile à calibrer et polyvalent, puisqu'il permet de choisir le degré d'homogénéité des groupes obtenus. La valeur ajoutée est de donner de façon automatique les modalités des différents *clusters*, et d'indiquer les différentes proportions pour déterminer si la sur-représentation ou la sous-représentation d'une modalité dans le *cluster* est gaussienne ou non.

Dans la dernière phase de l'étude qui consistait à décomposer l'impact d'une re-tarification des comptes, nous avons analysé l'amélioration du résultat avec le ratio S/P et de la volatilité du modèle avec le *MSE.* Nous avons choisi d'inclure les forêts aléatoires bien que l'objectif ne fût pas de l'utiliser en tarification, mais il s'est montré plus profitable et moins volatile que les autres dans tous les *segments* d'entreprises. Notre proposition d'un modèle simplifié à 4 facteurs a donné des résultats en amélioration, avec une meilleure estimation des grands risques, qui n'était pas possible dans le modèle original. Enfin, la correction partielle du modèle avec les zones de sous-tarification identifiées par les *k-Means* et la Classification Ascendante Hiérarchique, n'apporte pas de résultats substantiels mais cela est due au nombre faible de comptes sous tarifés initialement présents dans le portefeuille. Ces solutions permettraient ainsi de réduire la volatilité du modèle et de gagner en rentabilité sur certains *segments*.

En conclusion, la mise en place d'un processus de pilotage de portefeuille, tenant compte des contraintes de capital économique a permis de mettre en lumière les zones de sous-tarification du portefeuille, et la mise en place d'actions correctives. Nous avons présenté entre autres les bénéfices d'un recalibrage du modèle de tarification. Ces mêmes méthodes d'apprentissage peuvent également être appliquées avec des objectifs connexes dans le pilotage de portefeuille comme l'analyse de la résiliation étudié par Herboch [2016]. L'utilisation sur le portefeuille des autres pays ou sur d'autres lignes d'activités au sein de Chubb permettrait de confirmer les conclusions des différents algorithmes. L'extension à d'autres méthodes d'apprentissage non supervisées permettrait également d'enrichir la proposition de ce mémoire.

Comme tout travail statistique, le temps passé au nettoyage des données a été grand. La mise en pratique de ces méthodes en assurance de risques d'entreprises est donc conditionnée à l'obtention de données propres et conséquentes dès le départ. Le *Lloyd's of London* l'a bien compris, en mettant en place dès aujourd'hui des initiatives grâce aux objets connectés, qui permettront de traquer très

précisément les conteneurs. L'objectif final est d'arriver à une surveillance en continue du portefeuille contrairement à un processus ponctuel aujourd'hui. Cette base de données à grandes dimensions sera également utile pour la tarification, offrant une donnée précise du risque. En revanche, la nécessité d'avoir une puissance de calcul accrue sera peut-être un frein, car avec 50,000 lignes, certains algorithmes ont dû prendre plusieurs heures à converger. L'utilisation de ce processus en assurance de personnes où le portefeuille est composé de plus d'observations et où la tarification est réalisée de façon homogène avec un même outil, et sans structure complexe, (c'est-à-dire sans captive, sans rétention agrégée etc.), devrait permettre d'obtenir de meilleurs résultats.

Enfin, bien que le sujet de ce mémoire ne concernait pas l'utilisation de méthodes d'apprentissages pour la tarification, l'emploi des forêts aléatoires a permis de montrer la précision des modèles non linéaires. L'optimisation de cet algorithme ou l'extension à d'autres techniques telles que l'*Extreme Gradient Boosting*, ou des réseaux de neurones permettrait de comparer leurs résultats et de déterminer le modèle le plus approprié. L'utilisation de méthodes d'apprentissages dans la tarification pourrait survenir dans le futur en cas de changement de paradigme de la mutualisation des risques vers une tarification du risque individuel au plus juste.

OUTCOMES

The development of Big Data has helped popularised Machine Learning algorithms in the actuarial literature in the last few years. However, they are still rarely used in a pricing fashion in traditional insurance companies. This paper suggests the application of those models in optimising the profitability of a specialty risks portfolio in Marine Cargo insurance. This has become even more important as the international market is suffering from bad results. Indeed, the decline of insurance premium combined with the cost inflation and the accumulation of insured values with globalisation has led to negative rate changes.

Chubb's actuarial department is a regionalised function located in Paris for Continental Europe. The study is focusing on a single country to avoid adding any volatility due to heterogeneous pricing methods. The selected country has a sufficiently large portfolio to be analysed on a stand-alone basis. However, the implementation of a return on economic capital dimension, has shown that the portfolio only had a small share of non-profitable accounts.

The first phase of the study has consisted in enhancing the data visualisation. The use of factor analysis provided a summary of the database and the correlation of variables. Subsequently the "voyage" variable has been put aside for the rest of the analysis. Afterwards, the database has been rebalanced using a synthetic oversampling algorithm called SMOTE to improve the outcome of decision trees. The CART showed that the pricing model was too complex, compared to the actual loss experience. On the other hand, it also confirmed a good underwriting discipline. The extension to Random Forest, has proven to be effective in selecting impactful variables, in order to reduce the dimension for other algorithms and to select the variables of an alternative GLM that has proved to be better than a GLM with variables selected by a parametric method for the repricing phase.

Following the calculation of the risk profile for the economic capital, the second phase consisted in clustering the portfolio. The use of SVMs has proven to be unsuccessful. Besides being a supervised algorithm, as well as hard to calibrate, the visualisation of the outputs were not easy, either in 3D (uncompleted representation) or in 2D (truncated representation). For risks composed of more dimensions, the visualisation quality would have been even worse. The exploration of k-Means showed that unprofitable accounts were equally distributed in the portfolio. Cluster profiles holding either a significant amount of profitable or unprofitable accounts were revealed for strategic decision making at a later stage. Overall, k-Means offered interesting results, but the risk of overfitting is high as it requires to set a number of clusters beforehand. The outputs can also differ from time to time as the convergence depends on the starting point of clusters' centroids. Finally, Hierarchical Clustering also provided profitable and unprofitable cluster profiles but were different from k-Means. The model was

easier to calibrate and versatile, as it allows to choose clusters' purity. The added value brought by this algorithm is to give variables' categories and the ratio to determine if the cluster's sample is from a Normal distribution.

In the last phase, we have repriced accounts, and analysed the impacts on the loss ratio and on the accuracy of the model with the MSE. Although the main objective wasn't to use it for a pricing purpose, we have chosen to include the Random Forest as it showed precise estimates on all company segments. Our proposed simplified model with 4 variables has given better results than the original model on major risks. Last of all, the partial uplift of under-priced areas identified by k-Means and Hierarchical Clustering has not provided substantial results, but this is caused by the portfolio only having few under-priced accounts. Those actions would help reduce the model's volatility and gain profitability in some segments.

To conclude, this portfolio review process factoring return on economic capital helped to reveal under-priced areas and determine corrective action plans. Among others, we have presented the benefits of recalibrating the pricing model. Those same algorithms could be used in related portfolio management objectives such as cancellation studied by Herboch [2016]. Replicating this analysis on the portfolio of a different country or on a different line of business would help validate the calibration of the proposed algorithms. Additionally, the extension to other non-supervised algorithms would supplement the proposition of this paper.

Like every statistical analysis, the amount of time spent for data cleaning was tremendous. The implementation and usage of these models in corporate specialty risks rely on a clean and big enough database from the beginning. Lloyd's of London is therefore starting to implement "Internet of Things" initiatives to track containers for instance. The final objective will be to have a continuous and live monitoring instead of an ad-hoc process. This enhanced database will be useful for pricing, offering a precise tracking of the risk. However, the need of computational power will probably put a curb on the usage as some algorithms took several hours to complete with only 50,000 observations. Using this process in personal lines will certainly provide better results as the portfolio is composed of homogenous risks priced using the same pricing model and without any complex structure i.e. captives, annual retention etc.

Although the subject of this paper wasn't on applying machine learning techniques for pricing purposes, the use of Random Forest has shown the better fit of non-linear models. The optimisation of this algorithm or the extension to other techniques such as Extreme Gradient Boosting or to Neural Network would help compare and find the most appropriate model. The application of machine learning techniques in pricing would probably happen in the future in case we change from a mutualisation paradigm into pricing of individual risks as best as possible.

ANNEXES

Annexe 1 – Calcul de la marge de Solvabilité 1

Méthode des primes

Le volume de base est calculé à partir du montant le plus élevé des primes brutes émises ou acquises auquel sont ajoutées les primes acceptées en réassurance. Les primes pour les branches 11, 12 et 13 énumérées à l'article R. 321-1 sont majorées de 50 %.

Le volume de prime est ensuite multiplié par un facteur de 18% pour la part des primes jusqu'à 61.5 millions d'euros et 16% au-dessus. Un ratio de réassurance, limité à 50%, est ensuite appliqué. Il est calculé sur la base de la cession moyenne de sinistres des trois dernières années.

Méthode des sinistres

Le volume de base est calculé à partir de la somme des sinistres payés et des provisions pour sinistres à payer, constituées pour les affaires directes et en acceptation au cours des trois derniers exercices, brute de la réassurance mais avec déduction des recours. Pour les branches 11, 12 et 13 énumérées à l'article R. 321-1, les sinistres, provisions et recours sont majorés de 50 %.

Le tiers du volume est ensuite multiplié par 26 % pour la part jusqu'à 42.9 millions d'euros et 23% au-dessus. Un ratio de réassurance, limité à 50%, est ensuite appliqué. Il est calculé sur la base de la cession moyenne de sinistres des trois dernières années.

Annexe 2 – Calcul du risque de souscription non-vie

Le risque de souscription non-vie représente l'ensemble des risques provenant de contrats d'assurances non-vie.

$$SCR_{non-life} = \sqrt{\sum_{i,j} CorrNL(i,j) \cdot SCR_i \cdot SCR_j}$$

Le risque de souscription non-vie est l'agrégation des SCR_i et SCR_j pour les sous-modules i et j avec la matrice de corrélation $CorrNL(i,j)$ présente en ci-après.

Matrice de corrélation pour le risque de souscription non-vie

	Prime et réserve	Catastrophe	Cessation
Prime et réserve	1	0,25	0
Catastrophe	0,25	1	0
Cessation	0	0	1

Il est donc défini comme l'agrégation de trois sous-modules de risques :

1. Le risque de prime et de réserve
2. Le risque catastrophe non-vie
3. Le risque cessation en non-vie

Annexe 3 – Calcul du risque de prime et de réserve en Modèle Standard

Le risque de prime et de réserve

La directive Solvabilité 2 prévoit une classification des différentes lignes d'activités. Pour estimer le risque, il convient donc de classer les différentes lignes d'activités de l'assureur à l'aide de la nomenclature présente dans les actes délégués et de considérer leurs risques associés, à savoir, le risque d'une mauvaise tarification et le risque d'une mauvaise estimation du niveau de réserves nécessaire pour payer les sinistres.

Risque total du portefeuille

Le risque du portefeuille est composé de l'agrégation des risques de primes et de réserves de chaque ligne d'activités avec l'aide d'une matrice de corrélation. D'après la théorie du portefeuille, l'assureur bénéficie d'une diversification des risques lorsqu'il est présent sur plusieurs lignes d'activités. Cette diversification est matérialisée avec une matrice de corrélation des différents risques.

L'agrégation des risques est donnée par cette formule :

$$SCR_{nl\ prem\ res} = 3 \cdot \sigma_{nl} \cdot V_{nl}$$

σ_{nl} Représente le risque de prime et de réserve en non-vie du portefeuille de l'assureur et V_{nl} le volume consolidé de primes et de réserves.

$$\sigma_{nl} = \frac{1}{V_{nl}} \cdot \sqrt{\sum_{s,t} CorrS_{(s,t)} \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t}$$

Le produit des risques des différentes lignes avec la matrice de corrélation $CorrS_{(s,t)}$ présente ci-dessous et la pondération de leur volume de primes et de réserves permet l'agrégation globale.

Matrice de corrélation pour le risque de prime et de réserve non-vie

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0,5	0,5	0,25	0,5	0,25	0,5	0,25	0,5	0,25	0,25	0,25
2	0,5	1	0,25	0,25	0,25	0,25	0,5	0,5	0,5	0,25	0,25	0,25
3	0,5	0,25	1	0,25	0,25	0,25	0,25	0,5	0,5	0,25	0,5	0,25
4	0,25	0,25	0,25	1	0,25	0,25	0,25	0,5	0,5	0,25	0,5	0,5
5	0,5	0,25	0,25	0,25	1	0,5	0,5	0,25	0,5	0,5	0,25	0,25
6	0,25	0,25	0,25	0,25	0,5	1	0,5	0,25	0,5	0,5	0,25	0,25
7	0,5	0,5	0,25	0,25	0,5	0,5	1	0,25	0,5	0,5	0,25	0,25
8	0,25	0,5	0,5	0,5	0,25	0,25	0,25	1	0,5	0,25	0,25	0,5
9	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	1	0,25	0,5	0,25
10	0,25	0,25	0,25	0,25	0,5	0,5	0,5	0,25	0,25	1	0,25	0,25
11	0,25	0,25	0,5	0,5	0,25	0,25	0,25	0,25	0,5	0,25	1	0,25
12	0,25	0,25	0,25	0,5	0,25	0,25	0,25	0,5	0,25	0,25	0,25	1

Risque d'une ligne d'activité

L'assurance transport est classée par la nomenclature dans la ligne d'activité 6 « Assurance maritime, aérienne et transport ». Cela comprend les engagements d'assurance couvrant les dommages subis par les véhicules fluviaux, lacustres ou maritimes et les aéronefs, et tout dommage subi par les marchandises transportées ou les bagages quel que soit le moyen de transport. De même, cela inclut les engagements d'assurance couvrant tous les passifs découlant de l'utilisation d'aéronefs ou de navires, bateaux ou embarcations naviguant sur la mer, les lacs, rivières ou canaux, y compris la responsabilité du transporteur.

Le risque de la ligne d'activité « s » est défini par la volatilité dans le montant et la date de règlement des sinistres affectant les réserves et la volatilité du résultat provenant de l'incertitude concernant la fréquence et la gravité des événements assurés prise en compte dans le calcul de la prime.

$$\sigma_s = \frac{\sqrt{\sigma^2_{(prem,s)} \cdot V^2_{(prem,s)} + \sigma_{(prem,s)} \cdot V_{(prem,s)} \cdot \sigma_{(res,s)} \cdot V_{(res,s)} + \sigma^2_{(res,s)} \cdot V^2_{(res,s)}}}{V_{(prem,s)} + V_{(res,s)}}$$

Les écarts-types du risque de prime ou du risque de réserve des différents segments sont définis par la directive sont disponibles ci-après **:**

Ecarts types pour le sous-module risque de prime et de réserve non-vie

Segment	Ecart-type risque de prime	Ecart-type risque de réserve
1	10%	9%
2	8%	8%
3	15%	11%
4	8%	10%
5	14%	11%
6	12%	19%
7	7%	12%
8	9%	20%
9	13%	20%
10	17%	20%
11	17%	20%
12	17%	20%

Calcul des volumes de primes et de réserves

La différence entre les risques à développement long et les risques à développement court se fait à la fois dans la volatilité considérée et dans les volumes de primes et de réserves considérés. Pour chaque ligne d'activité, le volume V_s est calculé à l'aide de la formule suivante :

$$V_s = \left(V_{(prem,s)} + V_{(res,s)}\right) \cdot (0{,}75 + 0{,}25 \cdot DIV_s)$$

$V_{(prem,s)}$ Représente la mesure de volume pour risque de prime de la ligne d'activité « s »

$$V_{(prem,s)} = \left[P_s; P_{(last,s)}\right] + FP_{(existing,s)} + FP_{(future,s)}$$

Evaluée comme le maximum des primes acquises durant les 12 prochains mois et les 12 derniers mois et en ajoutant les primes futures à acquérir.

$V_{(res,s)}$ Représente la mesure de volume pour le risque de réserve du segment s

Calcul de la diversification inter ligne d'activité

DIV_s Représente le facteur de diversification géographique de la ligne d'activité « s » défini par :

$$DIV_{(s)} = \frac{\sum_r \left(V_{(prem,r,s)} + V_{(res,r,s}\right)^2}{\left(\sum_r \left(V_{(prem,r,s)} + V_{(res,r,s)}\right)\right)^2}$$

Où $V_{(prem,r,s)}$ et $V_{(res,r,s)}$ représentent respectivement le volume pour le risque de prime et du risque de réserve du segment s et de la région r.

Annexe 4 – Description de la base de données

Insured Value Band	Value
1	High
2	
3	
4	
5	
6	
7	Low

Limit Band	Value
1	High
2	
3	
4	
5	
6	
7	Low

Basis of Coverage	Restriction
1	Extensive
2	
3	
4	
5	
6	
7	
8	Restrictive

Conveyance Band	Type
1	Sea
2	Air
3	Land

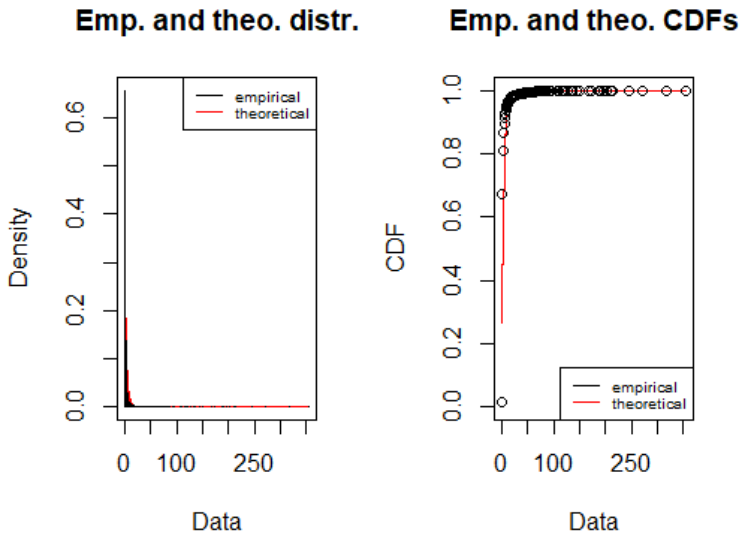
Voyage	Type
1	Worldwide
2	Continental
3	National

Stowage	Risk
1	High
2	
3	
4	
5	
6	
7	Low

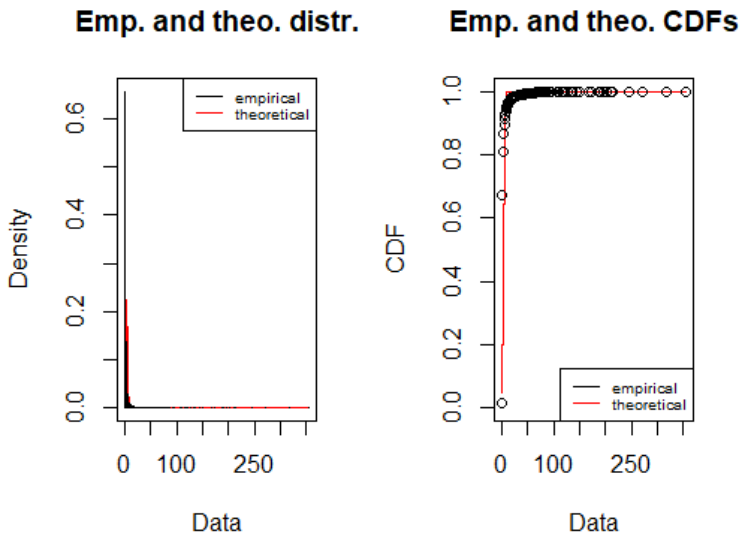
Segment	Type
1	Multinational
2	Middle Market
3	SME

Annexe 5 – Graphiques – Fréquence des sinistres graves

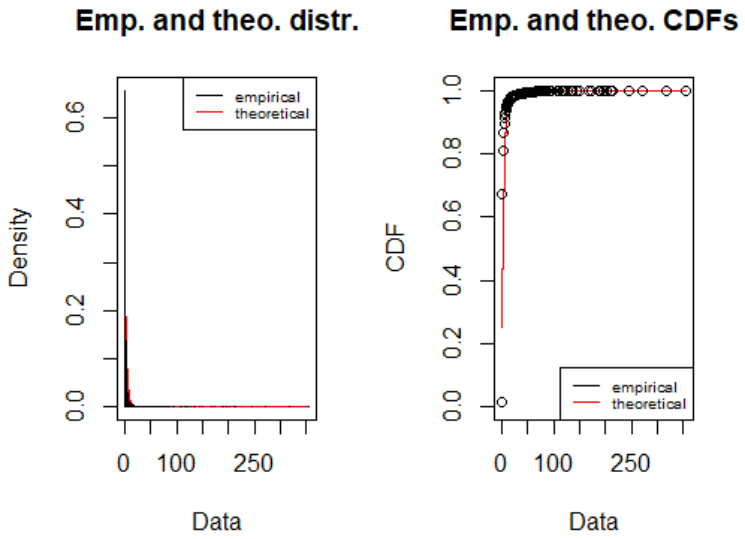
Loi de Poisson



Loi Binomiale Négative



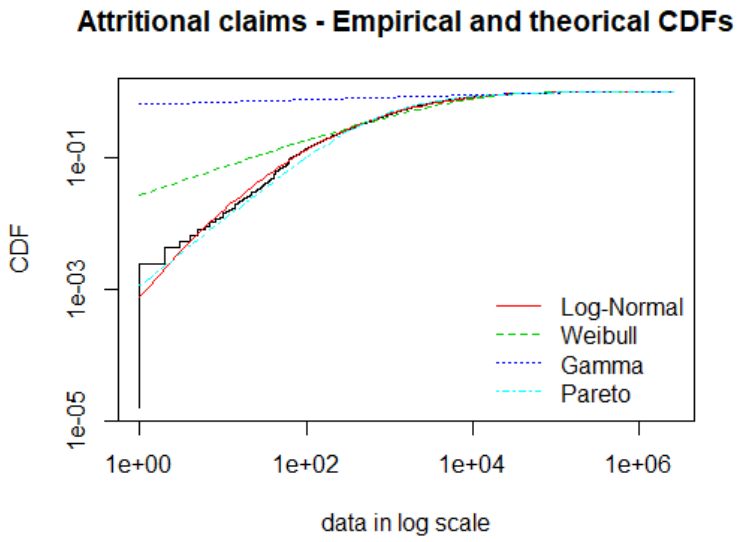
Loi Géométrique

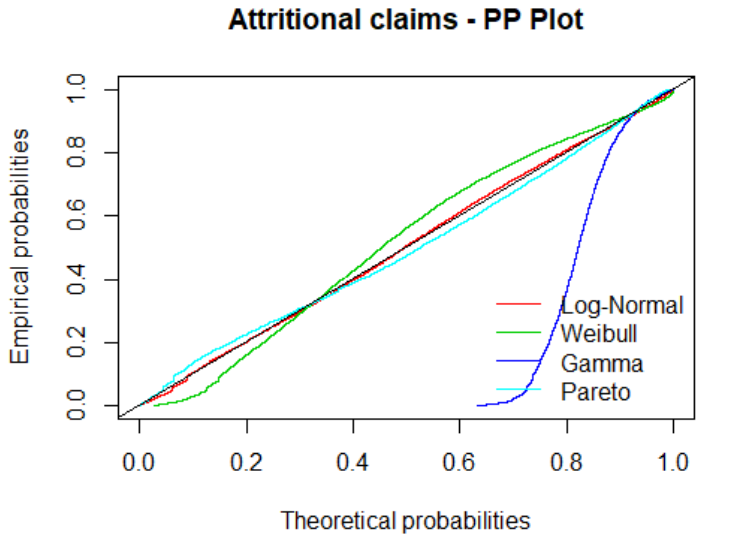


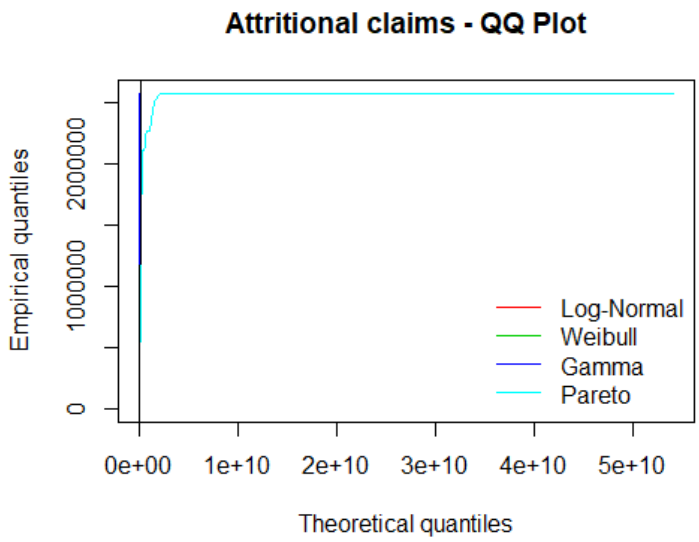
Annexe 6 – Résultats tests – Fréquence des sinistres graves

Chi-squared statistic:			218,202.00	Inf	391,908.60
DF of Khi-squared distribution			7	8	11.00
Chi-squared p-value:			-	-	-
The p-value may be wrong with some theoretical counts < 5					
Chi-squared table		**Obs counts**	**Th. Negative Binomial**	**Th. Poisson**	**Th. Geometric**
<=	-	473.00	8,579.19	1,583.89	8,038.48
<=	1	21,199.00	5,958.72	4,774.14	6,035.96
<=	2	4,386.00	4,362.16	7,195.10	4,532.31
<=	3	1,844.00	3,247.92	7,229.14	3,403.24
<=	4	902.00	2,438.60	5,447.51	2,555.44
<=	5	681.00	1,840.09	3,283.97	1,918.84
<=	6	432.00	1,393.08	1,649.75	1,440.82
<=	7	304.00	1,057.14	710.38	1,081.89
<=	9	488.00	1,415.39	357.29	1,422.37
<=	11	309.00	821.80	34.42	801.97
<=	15	352.00	758.33	2.40	707.12
<=	22	291.00	334.23	0.00	285.21
<=	40	312.00	60.83	0.00	44.10
> 40	40	295.00	0.52	-	0.26
Goodness-of-fit criteria			**Negative Binomial**	**Poisson**	**Geometric**
Akaike's Information Criterion			145,293.70	288,273.80	145,428.40
Bayesian Information Criterion			145,310.40	288,282.20	145,436.80

Annexe 7 – Graphiques – Sévérité des sinistres attritionnels



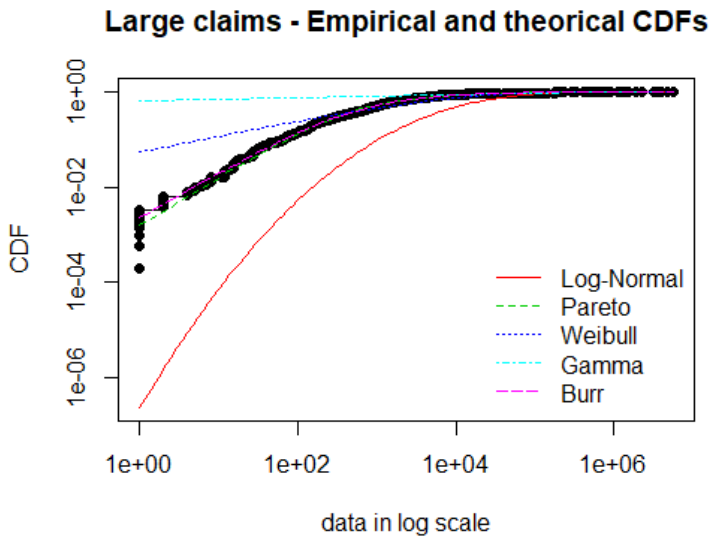


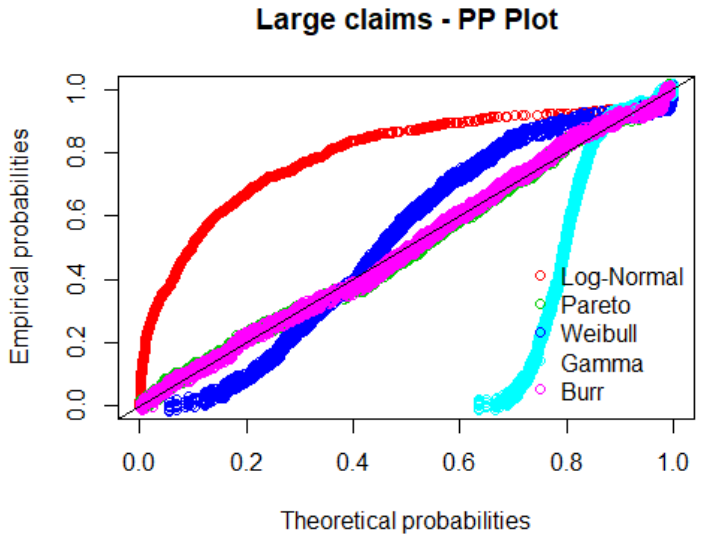


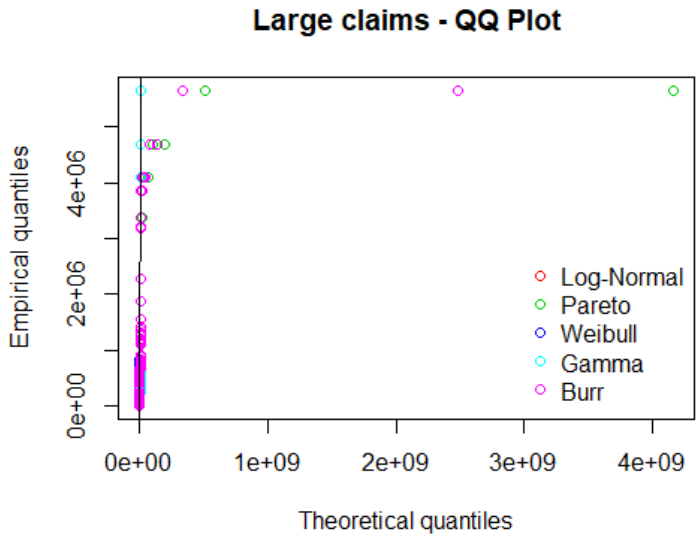
Annexe 8 – Résultats tests – Sévérité des sinistres attritionnels

Attritional				
Goodness-of-fit statistics	Log-Normal	Pareto	Weibull	Gamma
Kolmogorov-Smirnov statistic	0.02	0.04	0.08	0.68
Cramer-von Mises statistic	1.80	13.99	75.10	5,010.08
Anderson-Darling statistic	13.53	97.14	501.37	23,568.66
Goodness-of-fit criteria	Log-Normal	Pareto	Weibull	Gamma
Akaike's Information Criterion	608,197.00	609,234.10	614,284.80	692,254.50
Bayesian Information Criterion	608,213.70	609,250.90	614,301.50	692,271.20

Annexe 9 – Graphiques – Sévérité des sinistres graves



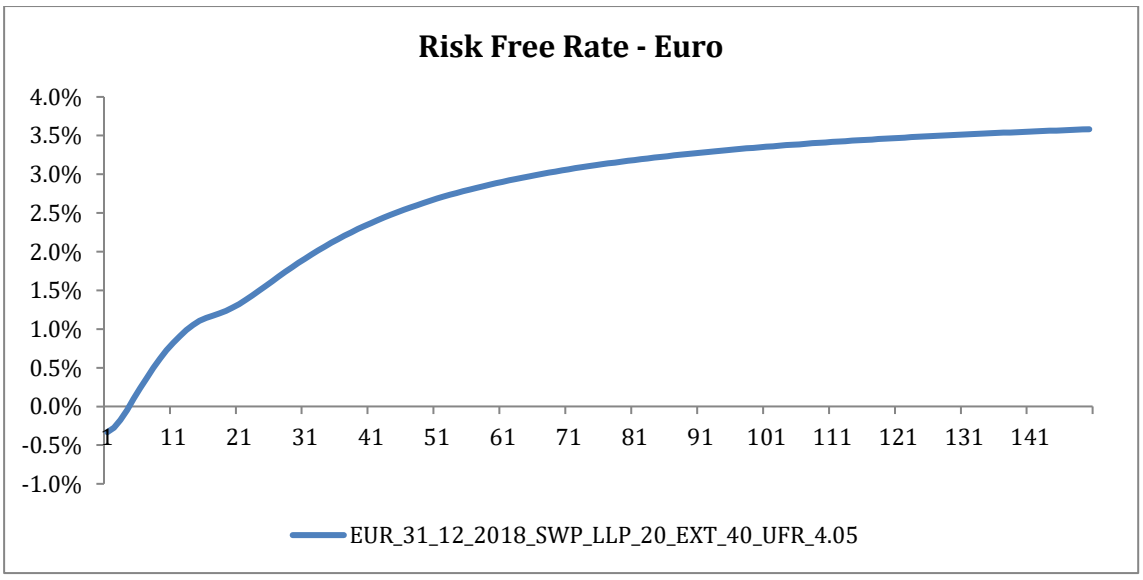




Annexe 10 – Résultats tests – Sévérité des sinistres graves

Large					
Goodness-of-fit statistics	Log-Normal	Pareto	Weibull	Gamma	Burr
Kolmogorov-Smirnov statistic	0.48	0.04	0.15	0.67	0.04
Cramer-von Mises statistic	287.19	0.74	19.54	371.04	0.62
Anderson-Darling statistic	2,022.51	5.49	121.03	1,702.01	4.81
Goodness-of-fit criteria	Log-Normal	Pareto	Weibull	Gamma	Burr
Akaike's Information Criterion	53,310.95	48,551.61	49,707.64	55,714.03	48,546.91
Bayesian Information Criterion	53,322.67	48,563.33	49,719.36	55,725.75	48,564.49

Annexe 11 – Risk Free Rate EIOPA



Annexe 12 – Description des 3 *clusters* suite à l'agrégation optimale du CAH

```
Link between the cluster variable and the categorical variables (chi-square test)
================================================================================
               p.value df
Commodity 0.000000e+00 20
Limit     0.000000e+00 12
IV        0.000000e+00 12
Segment   0.000000e+00  4
Category  6.476556e-08  2

Description of each cluster by the categories
=============================================
$`1`
                          Cla/Mod      Mod/Cla      Global       p.value     v.test
IV=IV_7                100.0000000  66.02784603 16.79163542  0.000000e+00        Inf
Limit=Limit_7           98.8931556  71.71171171 18.44122303  0.000000e+00        Inf
Commodity=Commodity_1   75.1569617 100.00000000 33.83737399  0.000000e+00        Inf
Segment=Segment_3       28.3139865  57.13349713 51.31633758  5.312959e-26  10.545781
Category=Category_1     25.7645360  96.46191646 95.21369658  5.077230e-08   5.448584
Segment=Segment_2       27.0882931  42.86650287 40.24410564  1.377688e-06   4.828206
IV=IV_6                100.0000000   0.08190008  0.02082813  1.062426e-03   3.273452
Limit=Limit_2            0.0000000   0.00000000  0.18745314  1.815391e-06  -4.772959
Commodity=Commodity_8    0.0000000   0.00000000  0.20411564  5.598219e-07  -5.004586
Category=Category_0     18.7989556   3.53808354  4.78630342  5.077230e-08  -5.448584
Commodity=Commodity_10   0.0000000   0.00000000  0.37907190  2.384469e-12  -7.009924
Commodity=Commodity_7    0.0000000   0.00000000  0.38323752  1.775764e-12  -7.051047
Commodity=Commodity_9    0.0000000   0.00000000  0.63317504  3.610630e-20  -9.199069
IV=IV_3                 14.0410959   4.70106470  8.51453803  4.077094e-39 -13.083769
Limit=Limit_3            0.9451796   0.08190008  2.20361576  9.717511e-60 -16.300947
IV=IV_4                 12.2813688   5.29074529 10.95559443  6.283654e-69 -17.546882
Limit=Limit_5            5.8363971   2.08026208  9.06440057 7.916461e-138 -24.989676
IV=IV_2                  7.7250608   4.16052416 13.69657586 1.062417e-169 -27.767801
Commodity=Commodity_6    0.0000000   0.00000000  5.97767225 2.879388e-190 -29.421927
Segment=Segment_1        0.0000000   0.00000000  8.43955678 1.754875e-272 -35.268999
IV=IV_1                  0.7293007   0.27846028  9.71007248 3.541264e-280 -35.767464
IV=IV_5                 12.2765320  19.45945946 40.31075564  0.000000e+00       -Inf
Limit=Limit_6           11.9138246  25.63472563 54.71965342  0.000000e+00       -Inf
Limit=Limit_4            0.7356948   0.44226044 15.28784471  0.000000e+00       -Inf
Commodity=Commodity_5    0.0000000   0.00000000 20.39073565  0.000000e+00       -Inf
Commodity=Commodity_4    0.0000000   0.00000000 28.22627676  0.000000e+00       -Inf
Commodity=Commodity_3    0.0000000   0.00000000  9.93085062  0.000000e+00       -Inf

$`2`
                           Cla/Mod    Mod/Cla      Global       p.value     v.test
IV=IV_5                 87.723468 53.8915693 40.31075564  0.000000e+00        Inf
Limit=Limit_6           87.088916 72.6256983 54.71965342  0.000000e+00        Inf
Commodity=Commodity_5   94.688458 29.4248349 20.39073565  0.000000e+00        Inf
IV=IV_2                 92.274939 19.2610462 13.69657586 6.620480e-322  38.358019
Commodity=Commodity_4   80.121015 34.4654647 28.22627676 2.703643e-205  30.574118
IV=IV_4                 87.718631 14.6457593 10.95559443 1.329265e-163  27.258183
Commodity=Commodity_3   88.003356 13.3189436  9.93085062 2.533712e-151  26.201816
Limit=Limit_5           88.878676 12.2778060  9.06440057 1.647183e-149  26.042240
IV=IV_3                 85.958904 11.1541392  8.51453803 5.446272e-104  21.655058
Segment=Segment_3       71.686013 56.0627222 51.31633758  4.639837e-92  20.350031
Commodity=Commodity_6   86.411150  7.8720163  5.97767225  2.209692e-75  18.371798
Segment=Segment_2       71.638547 43.9372778 40.24410564  3.882368e-59  16.216073
Commodity=Commodity_9  100.000000  0.9649568  0.63317504  1.194118e-28  11.104404
Commodity=Commodity_7  100.000000  0.5840528  0.38323752  1.335529e-17   8.540577
Commodity=Commodity_10 100.000000  0.5777044  0.37907190  2.039408e-17   8.491526
Commodity=Commodity_8  100.000000  0.3110716  0.20411564  1.053446e-09   6.101095
```

```
Category=Category_0     69.625762  5.0787202  4.78630342  3.124123e-03   2.955253
IV=IV_6                  0.000000  0.0000000  0.02082813  4.801512e-03  -2.820057
Category=Category_1     65.415409 94.9212798 95.21369658  3.124123e-03  -2.955253
Limit=Limit_1           30.434783  0.0444388  0.09580938  7.393939e-04  -3.374539
Limit=Limit_4           59.918256 13.9601320 15.28784471  5.198535e-15  -7.822013
Limit=Limit_3           26.654064  0.8951244  2.20361576  3.269815e-76 -18.475205
Segment=Segment_1        0.000000  0.0000000  8.43955678  0.000000e+00       -Inf
IV=IV_7                  0.000000  0.0000000 16.79163542  0.000000e+00       -Inf
IV=IV_1                  7.078507  1.0474860  9.71007248  0.000000e+00       -Inf
Limit=Limit_7            0.000000  0.0000000 18.44122303  0.000000e+00       -Inf
Commodity=Commodity_1   24.092084 12.4238192 33.83737399  0.000000e+00       -Inf

$`3`
                          Cla/Mod     Mod/Cla      Global       p.value     v.test
Segment=Segment_1      100.0000000  94.2764076  8.43955678  0.000000e+00        Inf
IV=IV_1                 92.1921922 100.0000000  9.71007248  0.000000e+00        Inf
Limit=Limit_4           39.3460490  67.1940437 15.28784471  0.000000e+00        Inf
Limit=Limit_3           72.4007561  17.8222429  2.20361576 1.537085e-287  36.237974
Commodity=Commodity_4   19.8789847  62.6803164 28.22627676 9.047609e-269  35.025996
Commodity=Commodity_6   13.5888502   9.0739879  5.97767225  2.376486e-09   5.969729
Limit=Limit_1           56.5217391   0.6049325  0.09580938  1.185795e-08   5.701756
Commodity=Commodity_3   11.9966443  13.3085156  9.93085062  1.337312e-07   5.273658
Limit=Limit_2           31.1111111   0.6514658  0.18745314  2.824220e-05   4.187196
Category=Category_0     11.5752829   6.1889251  4.78630342  2.057030e-03   3.081872
Commodity=Commodity_8    0.0000000   0.0000000  0.20411564  1.005014e-02  -2.574099
Category=Category_1      8.8200551  93.8110749 95.21369658  2.057030e-03  -3.081872
Commodity=Commodity_10   0.0000000   0.0000000  0.37907190  1.933460e-04  -3.727556
Commodity=Commodity_7    0.0000000   0.0000000  0.38323752  1.759719e-04  -3.751226
Commodity=Commodity_9    0.0000000   0.0000000  0.63317504  6.146824e-07  -4.986546
Limit=Limit_5            5.2849265   5.3513262  9.06440057  2.117856e-11  -6.697658
Commodity=Commodity_5    5.3115424  12.0986505 20.39073565  4.483128e-26 -10.561732
IV=IV_3                  0.0000000   0.0000000  8.51453803  6.286760e-88 -19.878185
IV=IV_4                  0.0000000   0.0000000 10.95559443 1.647649e-114 -22.743937
Limit=Limit_7            1.1068444   2.2801303 18.44122303 4.547028e-130 -24.265410
IV=IV_2                  0.0000000   0.0000000 13.69657586 2.737239e-145 -25.666920
IV=IV_7                  0.0000000   0.0000000 16.79163542 2.597893e-181 -28.713371
Commodity=Commodity_1    0.7509541   2.8385295 33.83737399 3.563471e-304 -37.279222
Segment=Segment_2        1.2731601   5.7235924 40.24410564 9.387247e-322 -38.348920
Segment=Segment_3        0.0000000   0.0000000 51.31633758  0.000000e+00       -Inf
IV=IV_5                  0.0000000   0.0000000 40.31075564  0.000000e+00       -Inf
Limit=Limit_6            0.9972594   6.0958585 54.71965342  0.000000e+00       -Inf
```

Annexe 13 – Description des 21 *clusters*

```
Link between the cluster variable and the categorical variables (chi-square test)
================================================================================
               p.value  df
Commodity 0.000000e+00 200
Limit     0.000000e+00 120
IV        0.000000e+00 120
Segment   0.000000e+00  40
Category  3.399137e-12  20

Description of each cluster by the categories
=============================================
$`1`
                          Cla/Mod     Mod/Cla     Global       p.value     v.test
Segment=Segment_2      24.2107442 100.0000000 40.2441056  0.000000e+00        Inf
IV=IV_7                58.0253039 100.0000000 16.7916354  0.000000e+00        Inf
Limit=Limit_7          52.1572171  98.7174006 18.4412230  0.000000e+00        Inf
Commodity=Commodity_1  28.7947803 100.0000000 33.8373740  0.000000e+00        Inf
Category=Category_1     9.9925625  97.6485678 95.2136966  2.457816e-10   6.329610
Limit=Limit_2           0.0000000   0.0000000  0.1874531  9.876844e-03  -2.580111
Commodity=Commodity_8   0.0000000   0.0000000  0.2041156  6.548927e-03  -2.718951
Commodity=Commodity_10  0.0000000   0.0000000  0.3790719  8.721048e-05  -3.923669
Commodity=Commodity_7   0.0000000   0.0000000  0.3832375  7.868088e-05  -3.948383
Commodity=Commodity_9   0.0000000   0.0000000  0.6331750  1.622905e-07  -5.238040
Category=Category_0     4.7867711   2.3514322  4.7863034  2.457816e-10  -6.329610
Limit=Limit_3           0.5671078   0.1282599  2.2036158  5.448709e-20  -9.154744
Commodity=Commodity_6   0.0000000   0.0000000  5.9776722  1.033629e-66 -17.254584
Segment=Segment_1       0.0000000   0.0000000  8.4395568  3.464582e-95 -20.699988
IV=IV_3                 0.0000000   0.0000000  8.5145380  4.568115e-96 -20.797411
Limit=Limit_5           0.0000000   0.0000000  9.0644006 1.526213e-102 -21.500923
IV=IV_1                 0.0000000   0.0000000  9.7100725 3.356896e-110 -22.304294
Commodity=Commodity_3   0.0000000   0.0000000  9.9308506 7.837749e-113 -22.573818
IV=IV_4                 0.0000000   0.0000000 10.9555944 3.879234e-125 -23.793719
Limit=Limit_4           0.7356948   1.1543395 15.2878447 1.166983e-134 -24.696473
IV=IV_2                 0.0000000   0.0000000 13.6965759 8.343608e-159 -26.850403
Commodity=Commodity_5   0.0000000   0.0000000 20.3907356 4.691322e-246 -33.500805
Segment=Segment_3       0.0000000   0.0000000 51.3163376  0.000000e+00       -Inf
IV=IV_5                 0.0000000   0.0000000 40.3107556  0.000000e+00       -Inf
Limit=Limit_6           0.0000000   0.0000000 54.7196534  0.000000e+00       -Inf
Commodity=Commodity_4   0.0000000   0.0000000 28.2262768  0.000000e+00       -Inf

$`2`
                          Cla/Mod Mod/Cla      Global      p.value   v.test
IV=IV_6               100.00000000     100  0.02082813 1.505786e-20 9.292590
Commodity=Commodity_1   0.06155361     100 33.83737399 4.432303e-03 2.845635

$`3`
                          Cla/Mod    Mod/Cla      Global       p.value     v.test
Segment=Segment_3      31.0171280  93.468689 51.31633758  0.000000e+00        Inf
Commodity=Commodity_1  50.3262341 100.000000 33.83737399  0.000000e+00        Inf
IV=IV_7                38.7744976  38.233855 16.79163542 6.576441e-299  36.952761
Limit=Limit_7          32.9568557  35.689824 18.44122303 2.286247e-187  29.194368
IV=IV_5                21.4942648  50.880626 40.31075564  6.804169e-51  15.005048
Limit=Limit_6          20.0137028  64.310176 54.71965342  2.845805e-42  13.624956
Limit=Limit_1           0.0000000   0.000000  0.09580938  1.362562e-02  -2.466985
Limit=Limit_2           0.0000000   0.000000  0.18745314  2.228362e-04  -3.691613
Commodity=Commodity_8   0.0000000   0.000000  0.20411564  1.054380e-04  -3.877724
Commodity=Commodity_10  0.0000000   0.000000  0.37907190  4.045872e-08  -5.488838
Commodity=Commodity_7   0.0000000   0.000000  0.38323752  3.354276e-08  -5.521861
IV=IV_2                12.9562044  10.420744 13.69657586  5.842390e-12  -6.883441
```

```
Commodity=Commodity_9   0.0000000   0.000000  0.63317504  4.304584e-13  -7.245622
Limit=Limit_3           0.0000000   0.000000  2.20361576  3.843427e-44 -13.935713
Commodity=Commodity_6   0.0000000   0.000000  5.97767225 4.589685e-121 -23.396956
IV=IV_1                 0.8151008   0.464775  9.71007248 9.647989e-166 -27.438051
Segment=Segment_1       0.0000000   0.000000  8.43955678 4.160355e-173 -28.048552
IV=IV_3                 0.0000000   0.000000  8.51453803 1.022824e-174 -28.180191
Limit=Limit_5           0.0000000   0.000000  9.06440057 1.455810e-186 -29.130963
Commodity=Commodity_3   0.0000000   0.000000  9.93085062 2.153165e-205 -30.581555
IV=IV_4                 0.0000000   0.000000 10.95559443 6.325532e-228 -32.231820
Segment=Segment_2       2.7636891   6.531311 40.24410564  0.000000e+00       -Inf
Limit=Limit_4           0.0000000   0.000000 15.28784471  0.000000e+00       -Inf
Commodity=Commodity_5   0.0000000   0.000000 20.39073565  0.000000e+00       -Inf
Commodity=Commodity_4   0.0000000   0.000000 28.22627676  0.000000e+00       -Inf

$`4`
                     Cla/Mod Mod/Cla     Global      p.value   v.test
Commodity=Commodity_11     100     100 0.01249688 4.337566e-13 7.244588

$`5`
                          Cla/Mod   Mod/Cla    Global      p.value     v.test
Commodity=Commodity_9 100.0000000 100.00000  0.633175 0.000000e+00        Inf
Limit=Limit_6           1.0277101  88.81579 54.719653 9.305047e-20   9.096779
IV=IV_5                 1.1780510  75.00000 40.310756 3.851594e-18   8.683090
Segment=Segment_3       0.9741050  78.94737 51.316338 1.902244e-12   7.041468
Category=Category_0     1.4795474  11.18421  4.786303 1.363721e-03   3.202221
Limit=Limit_3           0.0000000   0.00000  2.203616 3.344778e-02  -2.126667
Category=Category_1     0.5906287  88.81579 95.213697 1.363721e-03  -3.202221
Commodity=Commodity_6   0.0000000   0.00000  5.977672 8.276654e-05  -3.936245
Segment=Segment_1       0.0000000   0.00000  8.439557 1.446696e-06  -4.818461
Segment=Segment_2       0.3312287  21.05263 40.244106 5.085903e-07  -5.023043
IV=IV_1                 0.0000000   0.00000  9.710072 1.716909e-07  -5.227636
Commodity=Commodity_3   0.0000000   0.00000  9.930851 1.181866e-07  -5.296278
Limit=Limit_4           0.0000000   0.00000 15.287845 1.023554e-11  -6.803151
IV=IV_7                 0.0000000   0.00000 16.791635 6.657176e-13  -7.186298
Limit=Limit_7           0.0000000   0.00000 18.441223 3.136048e-14  -7.592577
Commodity=Commodity_5   0.0000000   0.00000 20.390736 7.816962e-16  -8.057031
Commodity=Commodity_4   0.0000000   0.00000 28.226277 1.058501e-22  -9.806229
Commodity=Commodity_1   0.0000000   0.00000 33.837374 4.230504e-28 -10.990823

$`6`
                          Cla/Mod     Mod/Cla     Global       p.value     v.test
IV=IV_5                23.6333574  93.7295082 40.3107556  0.000000e+00        Inf
Limit=Limit_6          18.3922046  99.0163934 54.7196534  0.000000e+00        Inf
Commodity=Commodity_5  49.8467824 100.0000000 20.3907356  0.000000e+00        Inf
Segment=Segment_3      16.9818979  85.7377049 51.3163376 2.036300e-311  37.723628
Limit=Limit_2           0.0000000   0.0000000  0.1874531  8.002247e-03  -2.651975
Commodity=Commodity_8   0.0000000   0.0000000  0.2041156  5.207504e-03  -2.793910
Commodity=Commodity_10  0.0000000   0.0000000  0.3790719  5.695444e-05  -4.025086
Commodity=Commodity_7   0.0000000   0.0000000  0.3832375  5.114349e-05  -4.050339
Commodity=Commodity_9   0.0000000   0.0000000  0.6331750  7.957469e-08  -5.368090
Limit=Limit_3           0.0000000   0.0000000  2.2036158  1.215005e-25 -10.467747
IV=IV_2                 4.6532847   6.2704918 13.6965759  1.183341e-34 -12.278387
Commodity=Commodity_6   0.0000000   0.0000000  5.9776722  1.003241e-69 -17.650800
Limit=Limit_5           0.5974265   0.5327869  9.0644006  2.708408e-85 -19.571454
Segment=Segment_1       0.0000000   0.0000000  8.4395568  1.671174e-99 -21.173645
IV=IV_3                 0.0000000   0.0000000  8.5145380 2.008110e-100 -21.273264
IV=IV_1                 0.0000000   0.0000000  9.7100725 3.319314e-115 -22.814138
Commodity=Commodity_3   0.0000000   0.0000000  9.9308506 5.869396e-118 -23.089748
IV=IV_4                 0.0000000   0.0000000 10.9555944 7.915674e-131 -24.337228
Limit=Limit_4           0.2997275   0.4508197 15.2878447 6.053132e-165 -27.371129
Segment=Segment_2       3.6021116  14.2622951 40.2441056 4.381556e-190 -29.407671
IV=IV_7                 0.0000000   0.0000000 16.7916354 2.877239e-207 -30.722189
```

```
Limit=Limit_7           0.0000000   0.0000000 18.4412230 6.499536e-230 -32.373408
Commodity=Commodity_4   0.0000000   0.0000000 28.2262768  0.000000e+00       -Inf
Commodity=Commodity_1   0.0000000   0.0000000 33.8373740  0.000000e+00       -Inf

$`7`
                          Cla/Mod    Mod/Cla     Global       p.value    v.test
Commodity=Commodity_7 100.0000000 100.000000  0.3832375 1.517641e-261 34.548186
IV=IV_5                 0.8370363  88.043478 40.3107556  2.042759e-21  9.502823
Limit=Limit_6           0.5633374  80.434783 54.7196534  2.687550e-07  5.144136
Segment=Segment_3       0.4951701  66.304348 51.3163376  3.828744e-03  2.891937
Limit=Limit_5           0.1378676   3.260870  9.0644006  3.611610e-02 -2.095618
IV=IV_2                 0.1824818   6.521739 13.6965759  3.366145e-02 -2.124104
Commodity=Commodity_6   0.0000000   0.000000  5.9776722  3.407412e-03 -2.928373
Segment=Segment_1       0.0000000   0.000000  8.4395568  2.952070e-04 -3.619470
IV=IV_3                 0.0000000   0.000000  8.5145380  2.737316e-04 -3.638969
IV=IV_1                 0.0000000   0.000000  9.7100725  8.140479e-05 -3.940226
Commodity=Commodity_3   0.0000000   0.000000  9.9308506  6.495692e-05 -3.994040
IV=IV_7                 0.0000000   0.000000 16.7916354  4.365535e-08 -5.475388
Limit=Limit_7           0.0000000   0.000000 18.4412230  6.888634e-09 -5.793617
Commodity=Commodity_5   0.0000000   0.000000 20.3907356  7.399451e-10 -6.157302
Commodity=Commodity_4   0.0000000   0.000000 28.2262768  5.235414e-14 -7.525911
Commodity=Commodity_1   0.0000000   0.000000 33.8373740  2.867593e-17 -8.451835

$`8`
                         Cla/Mod    Mod/Cla     Global       p.value     v.test
IV=IV_3                80.675147 100.000000  8.5145380  0.000000e+00        Inf
Commodity=Commodity_5  11.235955  33.353548 20.3907356  2.077625e-37  12.781596
Segment=Segment_3       7.614254  56.882959 51.3163376  2.689160e-06   4.693237
Category=Category_0     8.790252   6.124924  4.7863034  1.069611e-02   2.552473
Segment=Segment_2       7.359487  43.117041 40.2441056  1.398851e-02   2.457558
Limit=Limit_4           7.820163  17.404488 15.2878447  1.470584e-02   2.439543
Limit=Limit_2           0.000000   0.000000  0.1874531  4.054104e-02  -2.048193
Commodity=Commodity_8   0.000000   0.000000  0.2041156  3.048043e-02  -2.163790
Commodity=Commodity_4   6.242621  25.651910 28.2262768  1.536260e-02  -2.423716
Category=Category_1     6.772542  93.875076 95.2136966  1.069611e-02  -2.552473
Commodity=Commodity_10  0.000000   0.000000  0.3790719  1.520724e-03  -3.170699
Commodity=Commodity_7   0.000000   0.000000  0.3832375  1.415867e-03  -3.191397
Commodity=Commodity_9   0.000000   0.000000  0.6331750  1.935825e-05  -4.272167
Limit=Limit_5           4.549632   6.003639  9.0644006  2.381652e-06  -4.718008
Commodity=Commodity_3   3.859060   5.579139  9.9308506  5.110873e-11  -6.567670
Commodity=Commodity_6   0.000000   0.000000  5.9776722  1.648538e-46 -14.319691
Segment=Segment_1       0.000000   0.000000  8.4395568  2.990702e-66 -17.193106
IV=IV_1                 0.000000   0.000000  9.7100725  1.176866e-76 -18.530274
IV=IV_4                 0.000000   0.000000 10.9555944  5.265135e-87 -19.771254
IV=IV_2                 0.000000   0.000000 13.6965759 2.540082e-110 -22.316767
IV=IV_7                 0.000000   0.000000 16.7916354 1.347596e-137 -24.968413
IV=IV_5                 0.000000   0.000000 40.3107556  0.000000e+00       -Inf

$`9`
                         Cla/Mod    Mod/Cla     Global       p.value     v.test
IV=IV_4                90.8745247 100.000000 10.9555944  0.000000e+00        Inf
Commodity=Commodity_5  16.2410623  33.263598 20.3907356  9.180536e-55  15.585179
Limit=Limit_4          13.4059946  20.585774 15.2878447  2.651377e-13   7.311012
Limit=Limit_6          11.2058465  61.589958 54.7196534  8.881244e-13   7.146818
Segment=Segment_2      11.2928268  45.648536 40.2441056  1.640449e-08   5.646188
Segment=Segment_3      10.5446871  54.351464 51.3163376  1.749066e-03   3.129832
Limit=Limit_2           0.0000000   0.000000  0.1874531  8.882107e-03  -2.616559
Commodity=Commodity_8   0.0000000   0.000000  0.2041156  5.833985e-03  -2.756969
Commodity=Commodity_10  0.0000000   0.000000  0.3790719  7.034615e-05  -3.975112
Commodity=Commodity_7   0.0000000   0.000000  0.3832375  6.331595e-05  -4.000099
Commodity=Commodity_9   0.0000000   0.000000  0.6331750  1.132878e-07  -5.304007
Commodity=Commodity_1   8.2112520  27.907950 33.8373740  5.918659e-11  -6.545778
```

```
Limit=Limit_7            7.2961373  13.514644 18.4412230  1.309703e-11  -6.767562
Limit=Limit_3            0.5671078   0.125523  2.2036158  1.653519e-20  -9.282626
Limit=Limit_5            4.5955882   4.184100  9.0644006  8.917729e-22  -9.588728
Commodity=Commodity_6    0.0000000   0.000000  5.9776722  3.124573e-68 -17.455531
Segment=Segment_1        0.0000000   0.000000  8.4395568  2.304708e-97 -20.940204
IV=IV_3                  0.0000000   0.000000  8.5145380  2.899801e-98 -21.038740
IV=IV_1                  0.0000000   0.000000  9.7100725 1.004182e-112 -22.562859
IV=IV_2                  0.0000000   0.000000 13.6965759 1.868323e-162 -27.161179
IV=IV_7                  0.0000000   0.000000 16.7916354 8.836398e-203 -30.384374
IV=IV_5                  0.0000000   0.000000 40.3107556  0.000000e+00       -Inf

$`10`
                          Cla/Mod      Mod/Cla     Global       p.value     v.test
Segment=Segment_3      19.69315691  85.27240773 51.3163376  0.000000e+00        Inf
IV=IV_5                27.52919293  93.63796134 40.3107556  0.000000e+00        Inf
Limit=Limit_6          21.57429963  99.61335677 54.7196534  0.000000e+00        Inf
Commodity=Commodity_4  41.98642267 100.00000000 28.2262768  0.000000e+00        Inf
Category=Category_1    11.97444984  96.20386643 95.2136966  6.874815e-03   2.702848
Category=Category_0     9.39947781   3.79613357  4.7863034  6.874815e-03  -2.702848
Limit=Limit_2           0.00000000   0.00000000  0.1874531  3.406640e-03  -2.928443
Commodity=Commodity_8   0.00000000   0.00000000  0.2041156  2.054653e-03  -3.082217
Commodity=Commodity_10  0.00000000   0.00000000  0.3790719  1.010891e-05  -4.414831
Commodity=Commodity_7   0.00000000   0.00000000  0.3832375  8.906320e-06  -4.442153
Commodity=Commodity_9   0.00000000   0.00000000  0.6331750  4.414200e-09  -5.867873
Limit=Limit_3           0.00000000   0.00000000  2.2036158  4.724670e-30 -11.389356
IV=IV_2                 5.50486618   6.36203866 13.6965759  1.152862e-39 -13.179412
Commodity=Commodity_6   0.00000000   0.00000000  5.9776722  5.904761e-82 -19.175718
Segment=Segment_1       0.00000000   0.00000000  8.4395568 4.985174e-117 -22.997084
IV=IV_3                 0.00000000   0.00000000  8.5145380 4.108134e-118 -23.105167
Limit=Limit_5           0.09191176   0.07029877  9.0644006 2.319175e-121 -23.426059
IV=IV_1                 0.00000000   0.00000000  9.7100725 1.582554e-135 -24.777111
Commodity=Commodity_3   0.00000000   0.00000000  9.9308506 9.045111e-139 -25.076196
IV=IV_4                 0.00000000   0.00000000 10.9555944 6.188919e-154 -26.430044
Limit=Limit_4           0.24523161   0.31634446 15.2878447 9.169667e-201 -30.231372
Segment=Segment_2       4.33702515  14.72759227 40.2441056 3.349512e-216 -31.383960
IV=IV_7                 0.00000000   0.00000000 16.7916354 4.786646e-244 -33.362578
Limit=Limit_7           0.00000000   0.00000000 18.4412230 9.303843e-271 -35.156329
Commodity=Commodity_5   0.00000000   0.00000000 20.3907356 4.325436e-303 -37.212246
Commodity=Commodity_1   0.00000000   0.00000000 33.8373740  0.000000e+00       -Inf

$`11`
                         Cla/Mod     Mod/Cla     Global       p.value     v.test
Commodity=Commodity_6  89.407666 100.0000000  5.9776722  0.000000e+00        Inf
Limit=Limit_6           7.795371  79.8129384 54.7196534  5.280496e-83  19.300877
IV=IV_3                11.937378  19.0179267  8.5145380  2.545888e-34  12.216240
IV=IV_5                 7.016637  52.9228371 40.3107556  7.685792e-21   9.363884
Limit=Limit_5           9.696691  16.4458301  9.0644006  5.221001e-18   8.648436
IV=IV_2                 6.690998  17.1473110 13.6965759  3.306675e-04   3.590008
Segment=Segment_2       5.951765  44.8168355 40.2441056  6.363786e-04   3.415616
Commodity=Commodity_10  0.000000   0.0000000  0.3790719  6.684672e-03  -2.712158
Commodity=Commodity_7   0.000000   0.0000000  0.3832375  6.326051e-03  -2.730384
Commodity=Commodity_9   0.000000   0.0000000  0.6331750  2.303554e-04  -3.683166
IV=IV_4                 3.650190   7.4824630 10.9555944  1.816046e-05  -4.286384
Limit=Limit_3           1.134216   0.4676539  2.2036158  2.767329e-07  -5.138641
IV=IV_1                 1.887602   3.4294622  9.7100725  1.188374e-18  -8.815796
Segment=Segment_1       1.480750   2.3382697  8.4395568  9.067225e-21  -9.346411
Limit=Limit_4           1.144414   3.2735776 15.2878447  7.137194e-47 -14.377755
Commodity=Commodity_3   0.000000   0.0000000  9.9308506  1.038174e-60 -16.437071
IV=IV_7                 0.000000   0.0000000 16.7916354 2.809346e-106 -21.896432
Limit=Limit_7           0.000000   0.0000000 18.4412230 8.188203e-118 -23.075351
Commodity=Commodity_5   0.000000   0.0000000 20.3907356 9.227354e-132 -24.425232
Commodity=Commodity_4   0.000000   0.0000000 28.2262768 1.231712e-191 -29.528733
```

```
Commodity=Commodity_1   0.000000   0.0000000 33.8373740 7.557176e-239 -33.002184

$`12`
                          Cla/Mod    Mod/Cla     Global       p.value     v.test
Commodity=Commodity_3  54.7818792 96.6691340  9.9308506  0.000000e+00        Inf
Limit=Limit_6           8.2749695 80.4589193 54.7196534  2.382106e-92  20.382687
IV=IV_5                 9.2487341 66.2472243 40.3107556  3.628195e-87  19.790031
Limit=Limit_3          17.2022684  6.7357513  2.2036158  9.661594e-22   9.580458
IV=IV_2                 9.3369830 22.7239082 13.6965759  1.982981e-20   9.263252
Segment=Segment_3       6.7294423 61.3619541 51.3163376  2.203891e-14   7.638136
Category=Category_1     5.7006606 96.4470762 95.2136966  2.404495e-02   2.256410
Category=Category_0     4.1775457  3.5529238  4.7863034  2.404495e-02  -2.256410
Commodity=Commodity_10  0.0000000  0.0000000  0.3790719  5.086263e-03  -2.801519
Commodity=Commodity_7   0.0000000  0.0000000  0.3832375  4.798931e-03  -2.820230
Commodity=Commodity_9   0.0000000  0.0000000  0.6331750  1.458457e-04  -3.798038
Segment=Segment_1       3.7512340  5.6254626  8.4395568  5.702089e-05  -4.024812
Limit=Limit_5           3.5386029  5.6994819  9.0644006  2.671824e-06  -4.694559
Segment=Segment_2       4.6164993 33.0125833 40.2441056  1.720710e-08  -5.637965
Limit=Limit_4           2.6158038  7.1058475 15.2878447  6.271608e-21  -9.385336
Commodity=Commodity_6   0.0000000  0.0000000  5.9776722  5.543197e-38 -12.883940
IV=IV_3                 0.0000000  0.0000000  8.5145380  1.542296e-54 -15.551993
IV=IV_4                 0.3422053  0.6661732 10.9555944  3.246854e-56 -15.797309
IV=IV_7                 0.0000000  0.0000000 16.7916354 4.700266e-112 -22.494483
Commodity=Commodity_5   0.2042901  0.7401925 20.3907356 8.748561e-120 -23.270861
Limit=Limit_7           0.0000000  0.0000000 18.4412230 3.185541e-124 -23.705218
Commodity=Commodity_4   0.2804014  1.4063657 28.2262768 8.757447e-167 -27.525246
Commodity=Commodity_1   0.1969716  1.1843079 33.8373740 3.075823e-219 -31.605773

$`13`
                          Cla/Mod     Mod/Cla     Global       p.value     v.test
Segment=Segment_2      16.97546838  99.8174072 40.2441056  0.000000e+00        Inf
Limit=Limit_5          75.50551471 100.0000000  9.0644006  0.000000e+00        Inf
IV=IV_2                16.78832117  33.5970785 13.6965759 4.279896e-102  21.453015
Commodity=Commodity_3  18.24664430  26.4759586  9.9308506  3.322265e-89  20.025192
Commodity=Commodity_4   8.14639906  33.5970785 28.2262768  8.123451e-07   4.932378
Commodity=Commodity_5   8.41675179  25.0760803 20.3907356  1.836073e-06   4.770677
Category=Category_0     9.05134900   6.3298844  4.7863034  3.498107e-03   2.920197
Limit=Limit_2           0.00000000   0.0000000  0.1874531  4.103404e-02  -2.043186
IV=IV_3                 5.72407045   7.1211199  8.5145380  3.278275e-02  -2.134734
Commodity=Commodity_8   0.00000000   0.0000000  0.2041156  3.088429e-02  -2.158560
Category=Category_1     6.73316708  93.6701156 95.2136966  3.498107e-03  -2.920197
Commodity=Commodity_10  0.00000000   0.0000000  0.3790719  1.558392e-03  -3.163584
Commodity=Commodity_7   0.00000000   0.0000000  0.3832375  1.451328e-03  -3.184244
Commodity=Commodity_9   0.00000000   0.0000000  0.6331750  2.016693e-05  -4.263035
IV=IV_1                 4.20420420   5.9646987  9.7100725  1.715381e-08  -5.638500
IV=IV_4                 4.03041825   6.4516129 10.9555944  1.071015e-10  -6.456572
Limit=Limit_3           0.00000000   0.0000000  2.2036158  3.339889e-17  -8.434020
IV=IV_7                 3.12577524   7.6688984 16.7916354  5.429965e-29 -11.174591
Commodity=Commodity_6   0.00000000   0.0000000  5.9776722  2.454331e-46 -14.292007
Segment=Segment_1       0.00000000   0.0000000  8.4395568  5.287632e-66 -17.160041
Commodity=Commodity_1   3.00381632  14.8508825 33.8373740  3.373813e-72 -17.969560
Limit=Limit_4           0.00000000   0.0000000 15.2878447 9.523105e-124 -23.659058
Limit=Limit_7           0.00000000   0.0000000 18.4412230 5.386928e-152 -26.260756
Segment=Segment_3       0.02435263   0.1825928 51.3163376  0.000000e+00       -Inf
Limit=Limit_6           0.00000000   0.0000000 54.7196534  0.000000e+00       -Inf

$`14`
                          Cla/Mod    Mod/Cla     Global       p.value    v.test
Commodity=Commodity_8 100.00000000 100.000000  0.2041156 1.478169e-152 26.309882
IV=IV_2                 1.15571776  77.551020 13.6965759  8.264045e-24 10.060425
Limit=Limit_4           0.54495913  40.816327 15.2878447  1.793071e-05  4.289213
Commodity=Commodity_6   0.00000000   0.000000  5.9776722  4.863674e-02 -1.971762
```

```
Segment=Segment_1       0.00000000   0.000000  8.4395568  1.323496e-02 -2.477384
Limit=Limit_5           0.00000000   0.000000  9.0644006  9.458866e-03 -2.595008
IV=IV_1                 0.00000000   0.000000  9.7100725  6.668600e-03 -2.712956
Commodity=Commodity_3   0.00000000   0.000000  9.9308506  5.913954e-03 -2.752514
IV=IV_4                 0.00000000   0.000000 10.9555944  3.373681e-03 -2.931464
IV=IV_7                 0.00000000   0.000000 16.7916354  1.213061e-04 -3.843472
Limit=Limit_7           0.00000000   0.000000 18.4412230  4.541997e-05 -4.078025
Commodity=Commodity_5   0.00000000   0.000000 20.3907356  1.385988e-05 -4.346069
IV=IV_5                 0.04133512   8.163265 40.3107556  5.782775e-07 -4.998334
Commodity=Commodity_4   0.00000000   0.000000 28.2262768  8.588711e-08 -5.354302
Commodity=Commodity_1   0.00000000   0.000000 33.8373740  1.581703e-09 -6.035807

$`15`
                          Cla/Mod    Mod/Cla     Global       p.value     v.test
Segment=Segment_2      15.70230825 99.4754098 40.2441056  0.000000e+00        Inf
IV=IV_2                39.26399027 84.6557377 13.6965759  0.000000e+00        Inf
Limit=Limit_4          31.90735695 76.7868852 15.2878447  0.000000e+00        Inf
Commodity=Commodity_4  12.89846517 57.3114754 28.2262768 1.332965e-133  24.597818
Commodity=Commodity_5   8.74361593 28.0655738 20.3907356  1.304619e-13   7.405699
Commodity=Commodity_8   0.00000000  0.0000000  0.2041156  3.998068e-02  -2.053948
Commodity=Commodity_10  0.00000000  0.0000000  0.3790719  2.518149e-03  -3.021152
Commodity=Commodity_7   0.00000000  0.0000000  0.3832375  2.357573e-03  -3.041047
Commodity=Commodity_3   4.53020134  7.0819672  9.9308506  6.237583e-05  -4.003638
Commodity=Commodity_9   0.00000000  0.0000000  0.6331750  4.500080e-05  -4.080180
Limit=Limit_3           0.00000000  0.0000000  2.2036158  5.588016e-16  -8.097978
Commodity=Commodity_6   0.00000000  0.0000000  5.9776722  6.015803e-43 -13.737946
IV=IV_1                 0.77220077  1.1803279  9.7100725  4.249030e-46 -14.253740
Segment=Segment_1       0.00000000  0.0000000  8.4395568  3.770706e-61 -16.498348
IV=IV_3                 0.00000000  0.0000000  8.5145380  1.032974e-61 -16.576361
Limit=Limit_5           0.00000000  0.0000000  9.0644006  7.510218e-66 -17.139649
IV=IV_4                 0.00000000  0.0000000 10.9555944  2.761376e-80 -18.974690
IV=IV_5                 2.23209672 14.1639344 40.3107556 8.202299e-117 -22.975462
IV=IV_7                 0.00000000  0.0000000 16.7916354 6.360128e-127 -23.965563
Commodity=Commodity_1   1.41573310  7.5409836 33.8373740 7.863564e-139 -25.081769
Limit=Limit_7           0.00000000  0.0000000 18.4412230 1.006346e-140 -25.254672
Limit=Limit_6           2.69488429 23.2131148 54.7196534 6.423860e-148 -25.901390
Segment=Segment_3       0.06494034  0.5245902 51.3163376  0.000000e+00       -Inf

$`16`
                          Cla/Mod Mod/Cla      Global      p.value    v.test
Commodity=Commodity_2 100.00000000     100  0.02499375 3.764307e-24 10.137548
Limit=Limit_4           0.16348774     100 15.28784471 1.272252e-05  4.364824
Segment=Segment_2       0.06210537     100 40.24410564 4.244345e-03  2.859407
Segment=Segment_3       0.00000000       0 51.31633758 1.330495e-02 -2.475501
Limit=Limit_6           0.00000000       0 54.71965342 8.612524e-03 -2.627064

$`17`
                           Cla/Mod    Mod/Cla     Global       p.value    v.test
Commodity=Commodity_10 100.00000000 100.000000  0.3790719 3.945041e-259 34.386997
IV=IV_2                  2.52433090  91.208791 13.6965759  2.451584e-62 16.662594
Limit=Limit_4            2.26158038  91.208791 15.2878447  2.167950e-58 16.110065
Segment=Segment_2        0.85912431  91.208791 40.2441056  2.020397e-24 10.198172
Commodity=Commodity_6    0.00000000   0.000000  5.9776722  3.624922e-03 -2.909082
IV=IV_4                  0.03802281   1.098901 10.9555944  3.365651e-04 -3.585398
Segment=Segment_1        0.00000000   0.000000  8.4395568  3.225307e-04 -3.596497
IV=IV_3                  0.00000000   0.000000  8.5145380  2.993138e-04 -3.615893
Limit=Limit_5            0.00000000   0.000000  9.0644006  1.727364e-04 -3.755876
IV=IV_1                  0.00000000   0.000000  9.7100725  9.019624e-05 -3.915556
Commodity=Commodity_3    0.00000000   0.000000  9.9308506  7.214921e-05 -3.969085
IV=IV_7                  0.00000000   0.000000 16.7916354  5.250542e-08 -5.442610
Limit=Limit_7            0.00000000   0.000000 18.4412230  8.453494e-09 -5.759157
Commodity=Commodity_5    0.00000000   0.000000 20.3907356  9.303778e-10 -6.120918
```

```
IV=IV_5                  0.07233647   7.692308 40.3107556  2.468785e-12 -7.005061
Commodity=Commodity_4    0.00000000   0.000000 28.2262768  7.305264e-14 -7.482265
Commodity=Commodity_1    0.00000000   0.000000 33.8373740  4.342609e-17 -8.403258
Segment=Segment_3        0.06494034   8.791209 51.3163376  4.873011e-18 -8.656306
Limit=Limit_6            0.06090134   8.791209 54.7196534  1.902349e-20 -9.267682

$`18`
                          Cla/Mod    Mod/Cla     Global       p.value    v.test
Limit=Limit_2         100.00000000 100.000000  0.1874531 9.579482e-142 25.347483
Commodity=Commodity_3   1.17449664  62.222222  9.9308506  1.539349e-17  8.524151
IV=IV_3                 0.83170254  37.777778  8.5145380  7.561728e-08  5.377285
Segment=Segment_1       0.69101678  31.111111  8.4395568  1.432438e-05  4.338829
IV=IV_1                 0.60060060  31.111111  9.7100725  7.086319e-05  3.973369
Segment=Segment_2       0.32087776  68.888889 40.2441056  1.214951e-04  3.843090
Category=Category_0     0.60922541  15.555556  4.7863034  6.405004e-03  2.726293
Limit=Limit_5           0.00000000   0.000000  9.0644006  1.384329e-02 -2.461304
Category=Category_1     0.16625104  84.444444 95.2136966  6.405004e-03 -2.726293
Commodity=Commodity_5   0.04085802   4.444444 20.3907356  3.123480e-03 -2.955317
Limit=Limit_4           0.00000000   0.000000 15.2878447  5.679647e-04 -3.446469
IV=IV_7                 0.00000000   0.000000 16.7916354  2.534517e-04 -3.658747
Limit=Limit_7           0.00000000   0.000000 18.4412230  1.028316e-04 -3.883812
Commodity=Commodity_1   0.01231072   2.222222 33.8373740  2.078235e-07 -5.192200
IV=IV_5                 0.01033378   2.222222 40.3107556  2.599433e-09 -5.955081
Segment=Segment_3       0.00000000   0.000000 51.3163376  8.190898e-15 -7.764591
Limit=Limit_6           0.00000000   0.000000 54.7196534  3.120697e-16 -8.168569

$`19`
                          Cla/Mod   Mod/Cla      Global      p.value    v.test
Limit=Limit_1         100.00000000 100.00000  0.09580938 4.674309e-79 18.825426
Commodity=Commodity_3   0.83892617  86.95652  9.93085062 1.083873e-17  8.564668
Segment=Segment_1       0.64165844  56.52174  8.43955678 5.766633e-09  5.823388
IV=IV_1                 0.55770056  56.52174  9.71007248 3.187916e-08  5.530790
IV=IV_2                 0.21289538  30.43478 13.69657586 3.871149e-02  2.067241
Commodity=Commodity_1   0.03693217  13.04348 33.83737399 2.947491e-02 -2.177076
Limit=Limit_4           0.00000000   0.00000 15.28784471 2.197482e-02 -2.290803
Limit=Limit_7           0.00000000   0.00000 18.44122303 9.178899e-03 -2.605318
Commodity=Commodity_5   0.00000000   0.00000 20.39073565 5.260050e-03 -2.790661
Commodity=Commodity_4   0.00000000   0.00000 28.22627676 4.846236e-04 -3.489114
IV=IV_5                 0.00000000   0.00000 40.31075564 6.958509e-06 -4.494954
Segment=Segment_3       0.00000000   0.00000 51.31633758 6.382420e-08 -5.407742
Limit=Limit_6           0.00000000   0.00000 54.71965342 1.203337e-08 -5.699253

$`20`
                         Cla/Mod     Mod/Cla     Global       p.value     v.test
Segment=Segment_1      74.5804541  94.3196005  8.4395568  0.000000e+00        Inf
IV=IV_1                68.7258687 100.0000000  9.7100725  0.000000e+00        Inf
Limit=Limit_4          38.5013624  88.2022472 15.2878447  0.000000e+00        Inf
Commodity=Commodity_4  17.6652893  74.7191011 28.2262768  0.000000e+00        Inf
Commodity=Commodity_6  10.5923345   9.4881398  5.9776722  1.066412e-08   5.719813
Category=Category_0     8.6161880   6.1797753  4.7863034  8.999666e-03   2.612067
Limit=Limit_2           0.0000000   0.0000000  0.1874531  4.456326e-02  -2.008754
Commodity=Commodity_8   0.0000000   0.0000000  0.2041156  3.378772e-02  -2.122596
Category=Category_1     6.5756661  93.8202247 95.2136966  8.999666e-03  -2.612067
Commodity=Commodity_10  0.0000000   0.0000000  0.3790719  1.841676e-03  -3.114643
Commodity=Commodity_7   0.0000000   0.0000000  0.3832375  1.718307e-03  -3.135040
Commodity=Commodity_9   0.0000000   0.0000000  0.6331750  2.666646e-05  -4.200212
Limit=Limit_3           0.0000000   0.0000000  2.2036158  8.904190e-17  -8.318555
Commodity=Commodity_5   3.5546476  10.8614232 20.3907356  1.093953e-25 -10.477679
Commodity=Commodity_3   0.7550336   1.1235955  9.9308506  6.292562e-51 -15.010234
Limit=Limit_5           0.4595588   0.6242197  9.0644006  1.983327e-53 -15.387573
IV=IV_3                 0.0000000   0.0000000  8.5145380  6.618196e-65 -17.012644
Limit=Limit_7           1.1068444   3.0586767 18.4412230  5.288969e-84 -19.419414
```

```
IV=IV_4                   0.0000000   0.0000000 10.9555944  1.875250e-84 -19.472595
IV=IV_2                   0.0000000   0.0000000 13.6965759 4.447326e-107 -21.980278
IV=IV_7                   0.0000000   0.0000000 16.7916354 1.524061e-133 -24.592379
Commodity=Commodity_1     0.7509541   3.8077403 33.8373740 1.918623e-204 -30.510026
Segment=Segment_2         0.9419315   5.6803995 40.2441056 2.004479e-236 -32.832805
Segment=Segment_3         0.0000000   0.0000000 51.3163376  0.000000e+00       -Inf
IV=IV_5                   0.0000000   0.0000000 40.3107556  0.000000e+00       -Inf
Limit=Limit_6             0.9896468   8.1148564 54.7196534  0.000000e+00       -Inf

$`21`
                          Cla/Mod     Mod/Cla    Global       p.value     v.test
Segment=Segment_1     18.85488648  99.2207792  8.439557  0.000000e+00        Inf
IV=IV_1               16.43071643  99.4805195  9.710072  0.000000e+00        Inf
Limit=Limit_3         72.77882798 100.0000000  2.203616  0.000000e+00        Inf
Commodity=Commodity_3  6.79530201  42.0779221  9.930851  2.108787e-62  16.671599
Commodity=Commodity_4  2.05135773  36.1038961 28.226277  7.204447e-04   3.381678
Commodity=Commodity_6  0.00000000   0.0000000  5.977672  4.054757e-11  -6.602063
IV=IV_3                0.09784736   0.5194805  8.514538  7.288916e-13  -7.173903
Limit=Limit_5          0.00000000   0.0000000  9.064401  9.500593e-17  -8.310866
IV=IV_4                0.00000000   0.0000000 10.955594  2.704786e-20  -9.230061
IV=IV_2                0.00000000   0.0000000 13.696576  1.432453e-25 -10.452147
Limit=Limit_4          0.00000000   0.0000000 15.287845  1.034954e-28 -11.117177
IV=IV_7                0.00000000   0.0000000 16.791635  9.799546e-32 -11.722283
Limit=Limit_7          0.00000000   0.0000000 18.441223  4.075716e-35 -12.364335
Commodity=Commodity_1  0.00000000   0.0000000 33.837374  1.748484e-70 -17.749193
Segment=Segment_2      0.03105269   0.7792208 40.244106  3.144938e-81 -19.088530
IV=IV_5                0.00000000   0.0000000 40.310756  6.364196e-88 -19.877570
Segment=Segment_3      0.00000000   0.0000000 51.316338 1.619968e-122 -23.539189
Limit=Limit_6          0.00000000   0.0000000 54.719653 7.601979e-135 -24.713794
```

ARTICLES

[2011] Technical Newsletter September 2011 – SCOR Global P&C

[2017] NAT CAT Losses and Impact on Marine Insurance Market – Munich Re

[2019] Safety Shipping Review - Allianz Global Corporate Specialty

[2019] Solvency and Financial Condition Report – Chubb European Group SE

[2019] Annual Report 2018 – Lloyds of London

[2019] Global Marine Insurance Report 2018 – International Union of Marine Insurance

[2019] Economic Outlook - OCDE

BIBLIOGRAPHIE

[1] **Baccini** [2010]
Statistiques descriptives multidimensionnelles – Institut de mathématiques de Toulouse

[2] **Bellina** [2014]
Méthodes d'apprentissage appliquées à la tarification non-vie

[3] **Breiman** [1996]
Bagging predictors - Berkeley Statistics

[4] **Breiman** [2001]
Random forests. Machine Learning - Berkeley Statistics

[5] **Breiman et al** [1984]
Classification and regression trees – Chapman and Hall/CRC

[6] **Chawla et al** [2002]
SMOTE: Synthetic Minority Over-sampling Technique - Journal of Artificial Intelligence Research

[7] **Denuit, Charpentier** [2005]
Mathématique de l'assurance non-vie volume II - Economica

[8] **Dutang, Charpentier** [2012]
L'actuariat avec R

[9] **Genuer, Poggi** [2017]
Arbres CART et Forêts aléatoires, Importance et sélection de variables.

[10] **Guillot** [2015]
Apprentissage statistique en tarification non-vie : quel avantage opérationnel ?

[11] **Farr et al** [2014]
Marine and Energy Pricing – Institute and Faculty of Actuaries

[12] **Freund, Shapire** [1996]
Experiments with a new Boosting algorithm - AT&T Research Laboratories

[13] **Hu, Li** [2013]
A novel boundary oversampling algorithm based on neighbourhood rough set model: NRSBoundary-SMOTE - Chongqing Key Laboratory of Computational Intelligence.

[14] **Husson et al** [2009]
Analyse des données avec R - Broché

[15] **Herboch** [2016]
Predictive Analytics en actuariat : application à la modélisation de la résiliation non-vie

[16] **Japkowicz** [2000]
Learning from Imbalanced Data Sets: A comparison of Various Strategies

[17] **Kim** [2017]
Everything you wanted to know about kernel trick

[18] **Parodi** [2016]
Basic concepts and techniques of the pricing process in general insurance.

[19] **Therond** [2004]
Le modèle collectif

[20] **Tremblay** [2017]
Prédire les sinistres graves en assurance : les apports de l'apprentissage statistique aux modèles linéaires.

[21] **Vapnik** [1998]
Statistical Learning Theory - AT&T Research Laboratories

TABLE DES FIGURES

TABLE DES TABLEAUX