

Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuaires

Par : Monsieur/Madame Laura PEREZ

Titre du mémoire :

Partage de la valeur et comparaison multinormes

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

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

*Membres présents du jury de
l'Institut des Actuaires*

signature

Entreprise : NATIXIS ASSURANCES

Nom : BIENAIME Nathalie

Signature :

Directeur de mémoire en
entreprise :

Nom : GEORGES Nicolas

Signature : Georges

Invité :

Nom :

Signature :

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

Signature du responsable
entreprise

Georges

Signature du candidat

Laura Perez

*Membres présents du jury de la
filière*



ISUP - SORBONNE UNIVERSITÉ

MÉMOIRE
ANNÉE 2019 - 2020

Partage de la valeur et comparaison multinormes

Laura PEREZ

Tuteur de mémoire entreprise : Nicolas GEORGES
Tuteur de mémoire académique : Olivier LOPEZ

Avant-propos

Pour des raisons de confidentialité, les données présentées dans ce mémoire ont été modifiées sans perte de généralités.

Remerciements

Tout d'abord, je tiens à remercier mon tuteur en entreprise, M. Nicolas GEORGES pour m'avoir accompagnée tout au long de mon alternance, pour son aide dans la réalisation de mon mémoire et pour les relectures de celui-ci. Par ailleurs, je tiens à remercier toute la Direction Actuariat Produits ainsi que Natixis Assurances pour m'avoir permis d'effectuer mon alternance dans l'équipe de Prévoyance Individuelle, pour m'avoir fait une place au sein de leur équipe et pour avoir répondu à mes questions tout au long de l'année. En particulier, un grand merci à Mme Clotilde ROUBY, pour toutes ses connaissances des produits de prévoyance individuelle, à Mme Nathalie BIENAIME, ma responsable, pour sa bienveillance et ses conseils et à Mme Amandine ROMAND-PIQUANT pour toute son aide et ses propositions de relecture. Je remercie également M. Charles CLIET pour son aide concernant le calcul des éléments du Produit Net Assurance et M. Jimmy BARRIERE pour l'entraide que nous avons pu nous procurer dans la réalisation de nos mémoires respectifs. De même, je souhaite remercier l'équipe de l'assurance des emprunteurs ainsi que M. Adrien Helary pour les conseils qu'ils m'ont donnés lors des présentations de mon mémoire. Je tiens également à remercier la Direction des Risques et la tribu IFRS 17 qui ont su m'apporter leur aide quant à mes interrogations sur la norme IFRS 17.

Ensuite, je souhaiterais remercier l'ISUP et les professeurs pour les cours qu'ils nous ont dispensés ainsi que M. Olivier LOPEZ pour ses qualités de directeur de l'ISUP, de professeur et de tuteur académique.

Par ailleurs, je tiens à remercier mon père, pour sa passion des mathématiques qu'il m'a transmise et pour m'avoir toujours poussée vers la réussite, et ma mère pour son soutien durant mes études. Je remercie également ma soeur pour ses conseils et la relecture de mon mémoire.

Enfin, je remercie mes lecteurs pour l'attention qu'ils porteront à la lecture de ce mémoire.

Résumé

Mots-clés : Rentabilité, Produit Net Assurance, Produit Net Bancaire, French GAAP, Solvabilité 2, IFRS 17, comparaison multinormes, Machine Learning.

Ce mémoire a été écrit afin de répondre à plusieurs objectifs. Le premier était la description et l'explication du partage de la valeur pour un bancassureur. En effet, cette répartition de la valeur permet de représenter la manière dont la rentabilité d'un contrat d'assurance est partagée entre l'apporteur d'affaires, la banque, et l'assureur. Ainsi, il est nécessaire de comprendre ce que sont le Produit Net Assurance, indicateur de la rentabilité de l'assureur, et le Produit Net Bancaire, indicateur de la rentabilité de la banque, chez Natixis Assurances. Afin que le partage entre ces deux indicateurs de rentabilité soit équitable, le ratio PNA/PNB doit être proche de 1. Nos comparaisons ont été réalisées par famille de produits car il est très important pour le département Actuariat Produits Prévoyance de comparer les indicateurs à cette maille afin de classer les familles par rentabilité et d'effectuer d'éventuelles revues tarifaires. Par ailleurs, calculer un PNA et un PNB par tête, comme nous avons décidé de le faire, permet d'agrèger les données à différentes mailles souhaitées et, ainsi, d'observer l'évolution du PNA en fonction de l'ancienneté des contrats ou le ratio PNA/PNB par banque, par exemple. Pour obtenir ces deux indicateurs par tête, il a fallu construire une base de données avec les variables nécessaires à ce calcul. Nous avons donc utilisé les données des garanties Décès Toutes Causes, Décès Accidentel et Arrêt de Travail sur les années 2016 à 2019. En outre, cette base de données aura également permis de créer des modèles de Machine Learning afin de prédire le PNA à un an et de le comparer au PNA réel. Nous avons donc été amenés à faire tourner différents types de modèles, du simple arbre de décision au XGBoost en passant par l'algorithme Random Forest et le Generalized Linear Model.

De nos jours, beaucoup de normes sont présentes pour permettre aux compagnies d'assurance d'établir des comptes de résultat, sous French GAAP et bientôt sous IFRS 17, mais également pour assurer leur pérennité grâce à la norme prudentielle Solvabilité 2. Cependant, qu'en est-il du lien entre ces trois normes ? Afin de mieux comprendre la nouvelle norme internationale IFRS 17, il est tout d'abord nécessaire d'avoir connaissance de ses enjeux. Cette nouvelle norme mêle les normes French GAAP et Solvabilité 2 : le compte de résultat IFRS 17 comprend les Best Estimate de sinistres évalués sous Solvabilité 2 tandis que les primes et les règlements de sinistres sont évalués sous French GAAP. En effet, l'univers normatif sous lequel les entreprises d'assurance sont soumises a beaucoup évolué ces dernières années et les contraintes de production réglementaire forcent les différents acteurs à développer les outils de projection et de reporting nécessaires. Ainsi, l'analyse rétrospective, i.e., la comparaison des projections obtenues avec le réel, est très souvent oubliée. Les entreprises sont plutôt dans une optique "Les outils sont développés ou du moins avancent bien" afin de satisfaire les autorités de contrôle et ne sont pas assez dans l'optique "Les projections réalisées sont proches du réel". Ce mémoire nous a donc permis de comparer des résultats projetés avec des résultats réels et, ainsi, de développer un meilleur esprit critique et une meilleure analyse des résultats.

Dans le cadre de ce mémoire, nous avons calculé le résultat IFRS 17 sur le périmètre Décès Toutes Causes, Décès Accidentel et Arrêt de Travail de la prévoyance individuelle grâce à la méthode Premium Allocation Approach. Puis, nous avons été amenés à comparer les trois normes via différents éléments. Nous avons notamment comparé des résultats obtenus par des prédictions de modèles de Machine Learning. Ceux-ci nous ont permis de comparer le Best Estimate de Solvabilité 2 avec le Produit Net Assurance French GAAP réel et prédit par Machine Learning. Nous avons pu remarquer que, très souvent, une prédiction à un an par Machine Learning est plus fiable que la projection obtenue par Solvabilité 2. De plus, l'utilisation des prédictions d'un modèle de Machine Learning plutôt que des BE S2 pour la comparaison avec le PNA réel a permis un gain de temps. En effet, sous Solvabilité 2, il faut préparer tous les inputs modèle, puis, tout projeter et calculer les résultats de tous les scénarios paramétrés. Par ailleurs, le Machine Learning évite les biais d'hypothèses du modèle de la Direction des Risques qui projette les hypothèses S2. En outre, nous avons comparé les Return On Equity obtenus pour Solvabilité 2 et pour IFRS 17 et nous avons comparé le résultat net obtenu sous French GAAP et sous IFRS 17. Dès lors, nous avons pu comparer ces trois normes deux à deux.

Abstract

Keywords : Profitability, Net Insurance Income, Net Banking Income, French GAAP, Solvency 2, IFRS 17, multi-standards comparison, Machine Learning.

This thesis has been written to meet several objectives. The first was the description and explanation of value sharing for a bankinsurer. Indeed, this distribution of value makes it possible to represent the way in which the profitability of an insurance contract is shared between the business provider, the bank, and the insurer. Thus, it is necessary to understand what the Net Insurance Income, indicator of the insurer's profitability, and the Net Banking Income, indicator of the bank's profitability, are at Natixis Assurances. In order for the sharing between these two profitability indicators to be fair, the NII/NBI ratio must be close to 1. Our comparisons were made by product family because it is very important for the Product Actuarial department to compare the indicators to this granularity in order to classify the families by profitability and to carry out possible pricing reviews. In addition, computing a NII and a NBI per capita, as we have decided to do, allows the data to be aggregated to different desired granularities and, thus, to observe the evolution of the NII according to the contracts'age or the NII/NBI ratio per bank, for example. To obtain these two indicators per capita, it was necessary to build a database with the necessary variables for this computing. We, therefore, used data from All Causes of Deaths, Accidental Deaths and Disability guarantees from 2016 to 2019. In addition, this database also permitted to create Machine Learning models in order to predict the NII at one year and to compare it to the real NII. We, therefore, had to run different types of models, from the simple decision tree to the XGBoost, including the Random Forest algorithm and the Generalized Linear Model.

Nowadays, many standards are present to allow insurance companies to draw up income statements, under French GAAP and soon under IFRS 17, but also to ensure their sustainability thanks to the Solvency 2 prudential standard. However, what is the link between these three standards? In order to understand better the new international standard IFRS 17, it is, firstly, necessary to be aware of its challenges. This new standard combines French GAAP and Solvency 2 standards : the IFRS 17 income statement includes the Best Estimate of claims assessed under Solvency 2 while premiums and claims settlements are assessed under French GAAP. Indeed, the normative universe under which insurance companies are subject has evolved a lot recently and regulatory production constraints force insurers to develop the necessary projection and reporting tools. Thus, the retrospective analysis, i.e., the comparison of the obtained projections with the real, is very often forgotten. The companies are rather in a perspective "The tools are developed or at least are progressing well" in order to satisfy the supervisory authorities and are not enough in the perspective "The projections made are close to reality". This thesis, therefore, allowed us to compare projected results with real results and, thus, to develop a better critical mind and a better analysis of the results.

In the context of this thesis, we computed the IFRS 17 result on the All Causes of Deaths, Accidental Deaths and Disability perimeter thanks to the Premium Allocation Approach. Then, we compared the three standards via different elements. In particular, we compared results obtained by predictions of Machine Learning models. These allowed us to compare the Solvency 2 Best Estimate with the real French GAAP Net Insurance Income and the one predicted by Machine Learning. We have noticed that, very often, a one-year prediction by Machine Learning is more reliable than the projection obtained by Solvency 2. Moreover, using predictions from a Machine Learning model rather than Solvency 2 Best Estimate for the comparison with the real NII saved time. In fact, under Solvency 2, we have to prepare all the model inputs, then project everything and compute the results of all the parameterized scenarios. In addition, Machine Learning avoids the assumptions bias of the risks model which projects the S2 assumptions. In addition, we compared the Return On Equity obtained for Solvency 2 and for IFRS 17 and we compared the net result obtained under French GAAP and under IFRS 17. We were, therefore, able to compare these three standards two by two.

Table des matières

Liste des abréviations	1
Introduction	3
1 Quelques définitions	5
2 Présentation de l'entreprise	7
3 Partage de la valeur	10
3.1 Utilisation d'une base permettant de calculer le PNA	10
3.1.1 Création de la base	11
3.1.2 Calcul du PNA, du PNB et du RBE et préparation de la base pour le Machine Learning	15
3.2 Distribution du chiffre d'affaires	17
3.3 Distribution des règlements par risque	18
3.4 Distribution du ratio PNA/PNB par banque	18
3.5 Distribution du PNA, du PNB et du RBE par famille de produits	23
4 La norme IFRS 17 et ses enjeux	26
4.1 French GAAP	26
4.2 Solvabilité 2	26
4.2.1 Le SCR global	27
4.2.2 Les trois piliers	28
4.2.2.a Pilier I : Les exigences quantitatives	28
4.2.2.b Pilier II : Les exigences qualitatives	30
4.2.2.c Pilier III : Publication de l'information	30
4.2.3 Calcul des Best Estimate de Solvabilité 2	31
4.2.4 Transformation du Best Estimate en Produit Net Assurance	32
4.3 IFRS 17	33
4.3.1 Présentation de la norme	34
4.3.1.a Transition	36
4.3.1.b Les informations à fournir	36
4.3.1.c Les avantages d'IFRS 17	37
4.3.1.d Cadre réglementaire européen	38
4.3.1.e Grand principe de la norme IFRS 17	38
4.3.2 La granularité	39
4.3.3 Le passif sous IFRS 17	41
4.3.3.a Le Best Estimate	41
4.3.3.b Les taux d'actualisation	43
4.3.3.c Le Risk Adjustment	44
4.3.3.d La Contractual Service Margin	46
4.3.4 L'évolution du passif d'assurance durant la période de couverture	48
4.3.5 Approches comptables	49
4.3.5.a Le modèle Building Block Approach	50
4.3.5.b Le modèle Premium Allocation Approach	51
4.3.5.c Le modèle Variable Fee Approach	54
4.3.6 La rentabilité sous IFRS 17	54
4.3.6.a Le chiffre d'affaires	54
4.3.6.b L'option Other Comprehensive Income	55
4.3.6.c Les frais d'acquisition	55
4.3.7 Conclusion IFRS 17	56
4.4 Évaluation de la profitabilité des contrats	56
4.4.1 Première comptabilisation	56
4.4.2 Modèle Building Block Approach : Exemple calculatoire	57
4.5 Calcul du résultat IFRS 17	58
4.5.1 Ajout des postes nécessaires au calcul du résultat brut de réassurance IFRS 17	59
4.5.2 Ajout des postes nécessaires au calcul du résultat de réassurance IFRS 17	62
4.6 Comparaison IFRS 17 et Solvabilité 2	63
4.7 Comparaison IFRS 17 et French GAAP	67

5	Machine Learning	70
5.1	Les méthodes de Machine Learning	71
5.1.1	Les principaux types de données	71
5.1.2	Data pre-processing	71
5.1.3	L'apprentissage non-supervisé	72
5.1.3.a	Méthodes de partitionnement	72
5.1.3.b	Méthodes hiérarchiques	73
5.1.4	L'apprentissage supervisé	73
5.1.4.a	Generalized Linear Model	75
5.1.4.b	Arbre de décision	75
5.1.4.c	Bagging	76
5.1.4.d	Random Forest	77
5.1.4.e	Boosting	78
5.1.4.f	Gradient Boosting	79
5.1.4.g	XGBoost	80
5.1.5	Indicateurs de performance	80
5.1.5.a	La validation croisée	81
5.1.5.b	La Root-Mean-Square Error	82
5.1.5.c	Le critère d'information d'Akaike	82
5.2	Présentation des modèles construits et des résultats obtenus	82
5.2.1	Pre-processing	84
5.2.2	Arbre de décision	86
5.2.3	XGBoost	88
5.2.4	XGBoost avec les nouvelles variables	94
5.2.5	Generalized Linear Model loi gaussienne BPCE	96
5.2.6	Generalized Linear Model loi gaussienne CE	100
5.2.7	Generalized Linear Model loi Gamma CE	103
5.2.8	XGBoost sur les Caisses d'Épargne	104
5.2.9	XGBoost sur les Banques Populaires	106
5.2.10	Random Forest	108
5.2.11	Comparaison des modèles et modèle retenu	109
	Conclusion	114
	Table des figures	116
	Liste des tableaux	117
	Bibliographie	118
	Annexes	119
	A Détail des provisions totales d'ouverture et de clôture	119
	B Prédiction du PNA pour l'année 2020	120

Liste des abréviations

ACPR	Autorité de Contrôle Prudentiel et de Résolution
AdaBoost	ADaptive BOOSTing
ADE	Assurance des Emprunteurs
AIC	Akaike Information Criterion
ANC	Autorité des Normes Comptables
AT	Arrêt de Travail
BBA	Building Block Approach
BE	Best Estimate
BP	Banque Populaire
BPCE	Banque Populaire - Caisse d'Épargne
BSCR	Basic Solvency Capital Requirement
CA	Chiffre d'Affaires
CE	Caisse d'Épargne
CoC	Cost of Capital
CSM	Contractual Service Margin
CSR	Commission Sur Résultat
CTE	Conditional Tail Expectation
DAB	Dommage Aux Biens
DC ACC	Décès Accidentel
DC TC	Décès Toutes Causes
DCAT	Décès Toutes Causes, Décès Accidentel et Arrêt de Travail
DEP	Dépendance
DR	Direction des Risques
EIOPA	European Insurance and Occupational Pensions Authority
EFRAG	European Financial Reporting Advisory Group
FCF	Fulfilment Cash Flows
FG	French GAAP
GAAP	Generally Accepted Accounting Principles
GC	Gestion Centralisée
GCS	Garantie Commerciale Simple
GD	Gestion Déléguée
GE	Garantie Élémentaire
GLM	Generalized Linear Model
HRG	Homogeneous Risk Groups
HT	Hors Taxes
IAD	Invalidité Absolue et Définitive
IARD	Incendie, Accidents et Risques Divers
IAS	International Accounting Standards
IASB	International Accounting Standards Board
IASC	International Accounting Standards Committee
IFRS	International Financial Reporting Standards
LC	Loss Component
LIC	Liability for Incurred Claims
LKOV	Leave-K-Out cross-Validation
LoB	Line of Business
LOOV	Leave-One-Out cross-Validation
LRC	Liability for Remaining Coverage
MAV	Multirisque des Accidents de la Vie
MCR	Minimum Capital Requirement
ML	Machine Learning
MRH	MultiRisque Habitation
NA	Natixis Assurances
OBSQ	Obsèques
OCI	Other Comprehensive Income
ORSA	Own Risk and Solvency Assessment
PAA	Premium Allocation Approach
PC	Produit Commercial
PCG	Plan Comptable Général

PFGS	Provision pour Frais de Gestion de Sinistres
PI	Prévoyance Individuelle
PNA	Produit Net Assurance
PNB	Produit Net Bancaire
PP	Perte Pécuniaire
PPNA	Provision pour Primes Non Acquis
PSAP	Provision pour Sinistres A Payer
PTIA	Perte Totale et Irréversible d'Autonomie
QRT	Quantitative Reporting Templates
RBE	Résultat Brut d'Exploitation
RA	Risk Adjustment
ROE	Return On Equity
RM	Risk Margin
RMSE	Root-Mean-Square Error
RSR	Regular Supervisory Report
TVaR	Tail Value at Risk
S2	Solvabilité 2
SCR	Solvency Capital Requirement
SFCR	Solvency Financial and Condition Report
TTC	Toutes Taxes Comprises
VaR	Value at Risk
VFA	Variable Fee Approach
XGBoost	Extreme Gradient Boosting

Introduction

L'objectif de ce mémoire est multiple. Dans un premier temps, il consiste à expliquer le partage de la valeur pour un bancassureur. Comme nous allons le voir dans la première grande partie de ce mémoire, le partage de la valeur est une notion importante pour un bancassureur car cela représente la manière dont la rentabilité des contrats d'assurance est partagée entre la banque et l'assureur. Dans la suite de ce mémoire, nous parlerons de Produit Net Assurance (PNA) et de Produit Net Bancaire (PNB) pour désigner respectivement la rentabilité de l'assureur et celle de la banque. Afin de pouvoir étudier le PNA et le PNB comme nous le souhaitons, nous avons construit une base de données permettant de calculer ces deux indicateurs de rentabilité par tête. Cette base a été construite à partir des contrats présents en portefeuille entre 2016 et 2019 pour les garanties de Prévoyance Individuelle (PI) suivantes : Décès Toutes Causes, Décès Accidentel et Arrêt de Travail (DCAT). Posséder une base tête par tête nous a permis d'agrèger les données aux mailles décidées. Nous avons donc, par exemple, pu obtenir un ratio PNA/PNB par banque et l'évolution du PNA par rapport à l'ancienneté des contrats. Nous avons effectué nos comparaisons par famille de produits car il est très important pour le département Actuariat Produits Prévoyance de comparer les indicateurs à cette maille afin de classer les familles par rentabilité et d'effectuer d'éventuelles revues tarifaires.

Dans ce mémoire, nous avons également pour objectif de réaliser une comparaison multinormes de la rentabilité ainsi qu'une comparaison entre une rentabilité déterminée par les normes et une rentabilité déterminée par des méthodes de Machine Learning (ML). En effet, de nos jours, beaucoup de normes sont présentes pour permettre aux compagnies d'assurance d'établir des comptes de résultat, sous French Generally Accepted Accounting Principles (GAAP) et bientôt sous International Financial Reporting Standards (IFRS) 17, mais également pour assurer leur pérennité grâce à la norme prudentielle Solvabilité 2 (S2). Par abus de langage, les résultats obtenus sous French GAAP seront appelés "résultats réels". Afin de comparer Solvabilité 2 à French GAAP et d'observer les performances de modèles de Machine Learning, nous avons décidé de comparer les éléments suivants :

- PNA réel (indicateur de la rentabilité French GAAP).
- PNA prédit à un an par un modèle de Machine Learning.
- Best Estimate (BE) de Solvabilité 2.

En effet, l'univers normatif sous lequel les entreprises d'assurance sont soumises a beaucoup évolué ces dernières années et les contraintes de production réglementaire forcent les différents acteurs à développer les outils de projection et de reporting nécessaires. Ainsi, dans le domaine de l'actuariat, nous sommes souvent amenés à projeter des données avec des modèles et à comparer nos résultats d'une année sur l'autre. Cependant, nous ne nous attardons pas sur la comparaison avec les données réelles. Les entreprises sont plutôt dans une optique : "Les outils sont développés ou du moins avancent bien" afin de satisfaire les autorités de contrôle et ne sont pas assez dans l'optique : "Les projections réalisées sont proches du réel". Ce mémoire nous a donc permis de comparer des résultats projetés avec des résultats réels et, ainsi, de développer un meilleur esprit critique et une meilleure analyse des résultats.

Les modèles de Machine Learning que nous avons créés dans ce mémoire ont permis :

- L'obtention d'une rentabilité par tête :
 - Avec la possibilité d'agrèger aux mailles souhaitées. Cela est impossible sous S2 car il faudrait créer une base par tête. En effet, le modèle S2 ressort les résultats de la projection de la base en input. Ainsi, lorsque nous agrégeons à la maille "famille de produits", le modèle ressort un fichier par famille de produits. Il s'ensuit que, pour obtenir les résultats tête par tête, il faut projeter chaque tête une à une ce qui fait autant de paramétrages que de nombre de têtes et c'est impossible. Le paramétrage est très lourd pour chaque base de données donc il n'est pas possible de paramétrer des millions de bases.
 - Cela permet d'avoir les résultats par année d'effet rapidement grâce à une agrégation des résultats par tête et cela peut nous être très utile afin d'évaluer la rentabilité des contrats plus anciens et plus récents. Sous S2, il faudrait faire des bases de projection par année d'effet et par famille de produits, par exemple, soit environ 30×20 scénarios pour avoir une ligne pour chaque année d'effet et chaque famille de produits.

- Un gain de temps. En effet, sous Solvabilité 2, il faut préparer tous les inputs modèle, puis, tout projeter et calculer les résultats de tous les scénarios paramétrés.
- D'être plus proche de la réalité. Le Machine Learning évite les biais d'hypothèses du modèle de la Direction des Risques (DR) qui projette les hypothèses S2.

Pour nos comparaisons, nous avons dû agréger les résultats de Machine Learning par famille de produits et par risque car les résultats obtenus sous Solvabilité 2 sont regroupés de la sorte. Nous pouvons alors nous demander pourquoi calculer des résultats par tête alors qu'ils vont être agrégés in fine. Tout d'abord, pour avoir un bon modèle de Machine Learning, il est nécessaire d'avoir une grande quantité de données. Or, en regroupant les données par famille de produits, nous perdons énormément de données, il ne nous reste qu'une vingtaine de lignes par année comptable, soit environ 80 lignes, nous ne pouvons pas créer un bon modèle avec si peu de données. De plus, en agrégeant par famille de produits, les variables telles que l'âge ou l'ancienneté ne sont qu'une moyenne de toutes les têtes présentes dans chaque famille, nous perdons donc l'information de la variance de ces variables ce qui est nécessaire aux modèles de Machine Learning. Par ailleurs, outre l'objectif de comparaison de la rentabilité, le calcul du Produit Net Assurance et du Produit Net Bancaire répond à un besoin du département Actuariat Produits Prévoyance. En effet, aucun calcul de PNA ni de PNB par tête n'a été effectué jusqu'à présent alors qu'il peut être intéressant de voir à quel point un client peut être rentable ou non pour l'assureur et pour la banque. Ainsi, le travail effectué sur la base de données en amont du Machine Learning sera utile au département. La prédiction de la rentabilité par Machine Learning permettra, de plus, de prévoir la rentabilité tête par tête afin de revoir la tarification, si nécessaire, de certains produits de prévoyance individuelle, à savoir, ceux pour lesquels nous avons effectué nos calculs, i.e., les garanties Décès Toutes Causes, Décès Accidentel et Arrêt de Travail.

Cependant, le modèle de ML n'est pas auditable ("effet boîte noire" du Machine Learning) donc le département ne pourra pas l'utiliser pour remplacer le modèle de la Direction des Risques mais il pourra être utilisé en complément. En effet, comme nous l'avons vu, il permet d'avoir un certain nombre d'informations sur les contrats des clients, c'est donc un bon proxy. Par ailleurs, à partir des traitements et des calculs effectués sur la base de données, il sera possible de l'enrichir avec les données de l'an prochain lorsque nous les aurons afin de lancer la création d'un nouveau modèle permettant de prédire la rentabilité de l'année suivante. Le nouveau modèle pourra être réalisé avec les paramètres retenus cette année afin de tourner plus rapidement. Toutefois, le modèle retenu cette année pourra aussi bien être réutilisé l'an prochain mais il est vrai que l'ajout de données à la base d'apprentissage pourrait permettre de créer un nouveau modèle plus performant.

Dans un premier temps, nous aborderons le partage de la valeur entre la banque et l'assureur. Puis, nous décrirons la norme IFRS 17, ses enjeux et nous la comparerons à French GAAP et à Solvabilité 2. Enfin, dans un troisième temps, nous évoquerons les techniques de Machine Learning ainsi que les résultats obtenus.

1 Quelques définitions

1. Contrat d'assurance : Un contrat par lequel, une partie (l'émetteur) accepte de la part d'une autre partie (l'assuré), un risque d'assurance significatif en s'engageant à indemniser l'assuré si un événement futur incertain (l'évènement assuré) causait un dommage à l'assuré.
2. Risque d'assurance : Risque, autre que financier, transféré du détenteur d'un contrat à son émetteur (par exemple, risque de décès).
3. Assuré : Une partie qui a le droit à une indemnisation si un événement assuré par un contrat d'assurance se produit.
4. Participation aux bénéfices discrétionnaire : Il s'agit d'une part des produits financiers réalisés pendant l'année que l'assureur peut décider de distribuer à tous ou à une partie des bénéficiaires des contrats d'assurance vie.
5. Contractual Service Margin : C'est une composante de la valeur comptable du passif d'un groupe de contrats d'assurance représentant le bénéfice ou la perte non acquis que l'entité reconnaîtra à mesure qu'elle fournira des services au titre des contrats d'assurance.
6. Risk Adjustment : Il s'agit de la rémunération dont une entité a besoin afin de supporter l'incertitude concernant le montant et les dates de versements des flux de trésorerie résultant du risque non-financier.
7. Dérivé incorporé : C'est une composante d'un instrument financier (action, obligation et swap, par exemple) hybride qui inclut également un contrat hôte non dérivé ce qui a pour effet de faire varier une partie des flux de trésorerie de l'instrument hybride de manière analogue à celle d'un dérivé autonome.
8. Composante d'investissement : Le montant qu'une entité est tenue de payer à l'assuré même en l'absence de survenance d'un événement assuré.
9. Service d'assurance fourni : Il s'agit de l'augmentation du passif en raison des sinistres survenus et des charges engagées au cours de la période en excluant les composantes d'investissement.
10. Dérivé : C'est un instrument financier ou un autre contrat qui possède les trois caractéristiques suivantes :
 - Sa valeur varie en fonction d'une variation d'un taux d'intérêt spécifié, du prix d'un instrument financier, du prix d'une marchandise, d'un cours de change, d'un indice de prix ou de taux, d'une notation de crédit ou d'un indice de crédit ou d'une autre variable parfois appelée le "sous-jacent".
 - Il requiert peu ou pas d'investissement initial.
 - Il est réglé à une date future.
11. Fulfilment Cash Flows : Il s'agit d'une estimation explicite, objective et pondérée (i.e., la valeur attendue) de la valeur actuelle des sorties de trésorerie futures diminuée de la valeur actuelle des entrées de trésorerie futures qui adviendront à mesure que l'entité honorera les contrats d'assurance, y compris un ajustement pour le risque non-financier.
12. Revenu d'assurance : Il s'agit d'un indicateur constitué des prestations attendues, des frais attendus (hors composante d'investissement) et du relâchement de la CSM, du RA et du coût des contrats déficitaires. Il mesure la réduction de l'engagement initial de l'entité au cours de l'exercice. Il permet de reconnaître progressivement la prime initiale en produits, i.e., il est égal au montant des primes encaissées alloué à la période considérée.
13. Loss Component : Une composante de perte est un enregistrement théorique des pertes attribuables à chaque groupe de contrats d'assurance onéreux.
14. Période de couverture : Il s'agit de la période pendant laquelle l'entité fournit une couverture

pour les événements assurés. Cette période comprend la couverture relative à toutes les primes comprises dans le périmètre du contrat d'assurance.

15. Portefeuille de contrats d'assurance : Il s'agit d'un groupe de contrats d'assurance soumis à des risques similaires et gérés ensemble. Nous nous attendons à ce que différentes gammes de produits (par exemple, une assurance en cas de dépendance et une assurance automobile) se retrouvent dans différents portefeuilles.
16. Liability for Remaining Coverage : Il s'agit de l'obligation pour une entité d'enquêter et de payer les sinistres valides en vertu des contrats d'assurance existants pour les événements assurés qui ne se sont pas encore produits (i.e., l'obligation qui se rapporte à la partie non expirée de la période de couverture).
17. Charges ou produits financiers d'assurance : Ils comprennent la variation de la valeur comptable du groupe de contrats d'assurance résultant de :
 - L'effet de la valeur temps de l'argent et des variations dans la valeur temps de l'argent.
 - L'effet du risque financier et de ses variations.
 - En excluant de telles variations pour les groupes de contrats d'assurance avec participation aux bénéfices directe qui ajusteraient la CSM, mais qui ne le font pas parce que les groupes de contrats sont onéreux et, par conséquent, il n'y a pas de CSM. Ceux-ci sont inclus dans les frais de services d'assurance.
18. Liability for Incurred Claims : Il s'agit de l'obligation pour une entité d'enquêter et de payer les sinistres valides pour les événements assurés qui se sont déjà produits, y compris les événements qui se sont produits mais pour lesquels les réclamations n'ont pas été déclarées et les autres dépenses d'assurance survenues.
19. Frais de services d'assurance : Ils comprennent :
 - Les réclamations encourues (à l'exclusion des paiements en vertu des composantes d'investissement) et d'autres frais de service engagés.
 - L'amortissement des flux de trésorerie liés à l'acquisition de contrats d'assurance.
 - Les variations dans les FCF liés aux services passés, i.e., liés au LIC.
 - Les changements dans les FCF liés aux services futurs mais qui n'ajustent pas la CSM, i.e., les pertes sur les groupes de contrats onéreux et les reprises de ces pertes.
20. Éléments sous-jacents : Ce sont les éléments qui déterminent certains des montants payables à l'assuré. Les éléments sous-jacents peuvent comprendre n'importe quels éléments, par exemple, un portefeuille d'actifs de référence, l'actif net de l'entité ou un sous-ensemble spécifié de l'actif net de l'entité.

2 Présentation de l'entreprise

Le groupe Banque Populaire - Caisse d'Épargne (BPCE) est né de la fusion, en 2009, de deux réseaux coopératifs : Banque Populaire (BP) et Caisse d'Épargne (CE), il est le deuxième acteur bancaire en France. Natixis, filiale du groupe BPCE, a été créée en 2006, elle est issue de la fusion des banques d'investissement de Banque Populaire et Caisse d'Épargne. Elle regroupe les pôles de banque internationale de financement, de gestion, d'assurance et de services financiers. Natixis Assurances (NA) se situe au sein du pôle épargne et assurance de Natixis. Le nom "Natixis Assurances" comprend le mot "assurance", pourtant, Natixis Assurances n'est pas un assureur à proprement parler. Cependant, ses filiales, couvrant chacune un domaine particulier, possèdent les accréditations nécessaires auprès de l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) qui est l'organe de supervision français de la banque et de l'assurance.

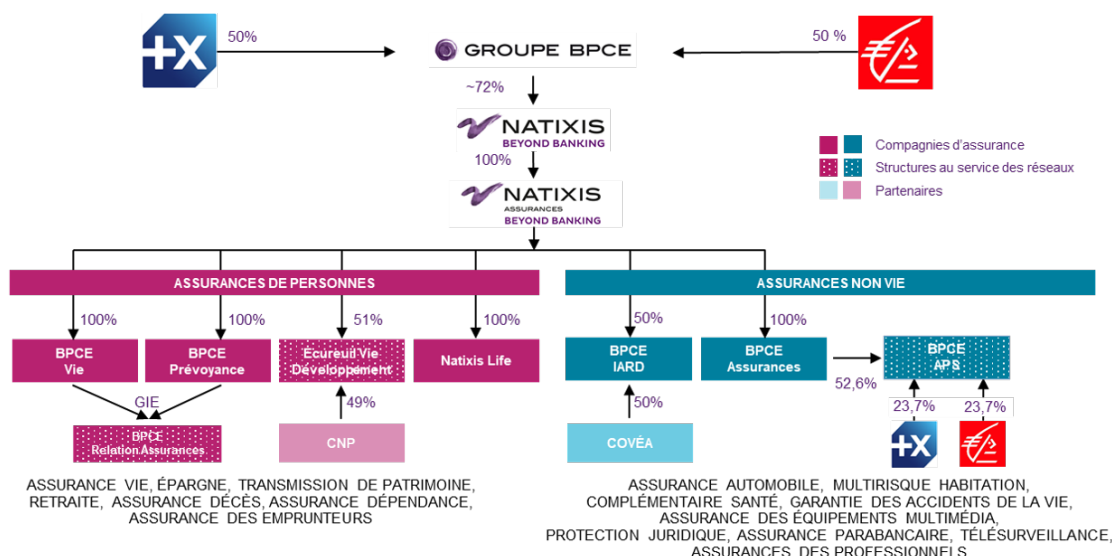


FIGURE 1 – Natixis Assurances : Deux métiers au service des banques du groupe BPCE

Les produits dont nous allons chercher à déterminer la rentabilité sont ceux de la branche "Prévoyance Individuelle" du groupe BPCE et ils sont gérés par Natixis Assurances. En effet, Natixis Assurances couvre les besoins en assurances de personnes pour les particuliers, le grand public et la gestion privée, les professionnels, les entreprises, ou encore, les professions libérales.

Natixis Assurances crée de nouveaux produits qui vont répondre aux nouveaux besoins des clients et des banques du groupe BPCE, par exemple, la signature électronique, la vente à distance et les applications mobiles. Elle suit également les produits existants mais en s'adaptant aux nouvelles évolutions réglementaires qui apparaissent au cours du temps. Natixis Assurances n'a pas de réseau de distribution qui lui est propre. En effet, la commercialisation de tous les produits se fait via le réseau bancaire BPCE.

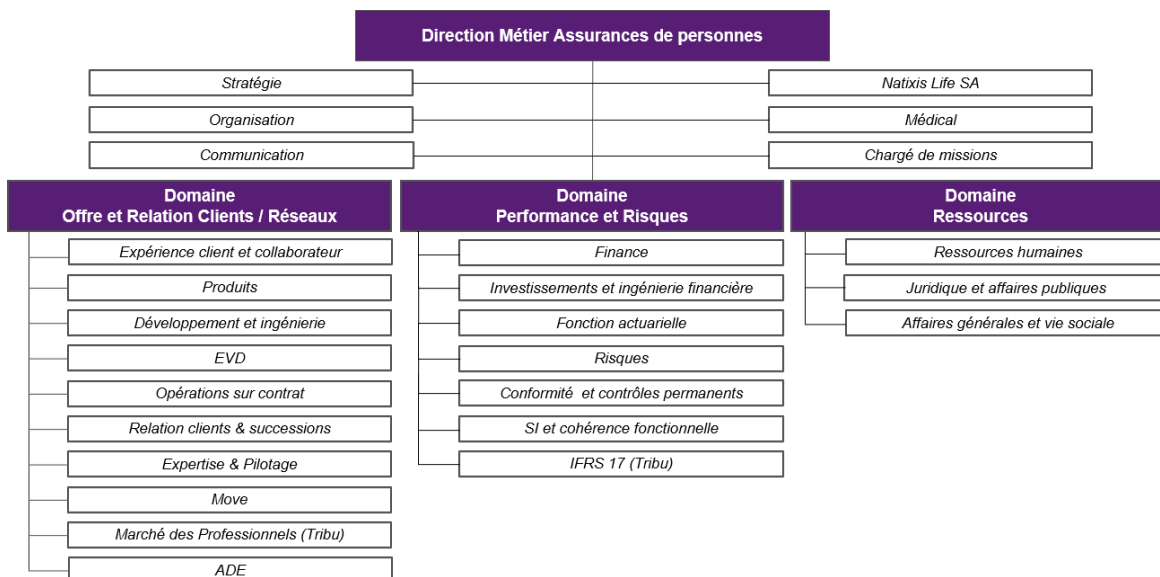


FIGURE 2 – Une organisation et quatre grands domaines

Le pôle principal de la structure de Natixis Assurances est le pôle Assurances de personnes qui comprend notamment l'assurance vie, l'épargne, la retraite, l'assurance des emprunteurs, la transmission de patrimoine et l'assurance dépendance. C'est dans ce pôle que se trouve le département Actuariat Produits Prévoyance au sein duquel j'ai effectué mon alternance. Dans ce département, nous traitons les contrats (cf. définition 1) de Prévoyance Individuelle. Comme son nom l'indique, cette entité couvre les individus (et non les groupes) contre les risques d'assurance (cf. définition 2) suivants :

- **Décès** : Ce risque couvre le Décès Toutes Causes (DC TC), Accidentel (DC ACC) ou provenant d'une maladie.
- **Perte Totale et Irréversible d'Autonomie (PTIA)** : Ce risque est aussi appelé Invalidité Absolue et Définitive (IAD). Il concerne les individus qui sont reconnus définitivement incapables d'effectuer une quelconque activité procurant un gain et qui se trouvent, alors, dans l'obligation d'avoir recours à l'assistance d'une tierce personne afin d'accomplir les actes ordinaires de la vie quotidienne.
- **Arrêt de Travail (AT)** : Ce risque concerne l'incapacité ou l'invalidité, qu'elle soit partielle ou totale, qui empêche un individu de travailler. L'invalidité est un stade supérieur à l'incapacité : si un assuré (cf. définition 3) est encore en incapacité au bout de trois ans, il passe en invalidité (mais il peut y passer plus tôt si son état le justifie). Il existe trois niveaux d'invalidité qui sont définis par l'article L.341-4 du code de la sécurité sociale¹ et le troisième niveau correspond au risque de PTIA.
- **Dépendance (DEP)** : Elle peut être totale ou partielle selon la capacité de la personne à effectuer les actes de la vie quotidienne (s'alimenter, faire sa toilette, s'habiller/se déshabiller, effectuer ses transferts). Elle peut, par exemple, résulter d'un accident, d'un cancer ou d'une maladie neurologique telle que la maladie d'Alzheimer.

Il existe différentes garanties permettant de couvrir les risques décrits ci-dessus. Par exemple, les risques Décès et PTIA peuvent être couverts par des garanties de type maladie/accident ou de type accident seul. La prestation à verser peut être sous la forme :

- D'un capital choisi à l'adhésion qui peut être fixe ou non.
- D'un capital compris dans une tranche choisie à l'adhésion et qui sera déterminé au moment du décès en fonction d'éléments relatifs aux comptes bancaires de l'assuré.
- D'une rente certaine dont le montant a été choisi à l'adhésion.

La garantie Décès maladie/accident offre la possibilité d'être complétée par une garantie maladie grave, sur option, avec versement anticipé d'une partie du capital décès. Il existe d'autres

1. Il est possible de trouver cet article au lien suivant : <http://www.code-securite-sociale.fr/article-L341-4-a6742597.html>.

garanties telles que la garantie Multirisque des Accidents de la Vie (MAV), la garantie Obsèques (OBSQ), la garantie Perte Pécuniaire (PP) ou la garantie Dommages Aux Biens (DAB) qui sera regroupée avec la garantie PP dans le tableau ci-après. Il peut être un peu surprenant de trouver la garantie DAB en Prévoyance Individuelle mais la raison est, en fait, plutôt historique et concerne seulement deux produits. Dans le tableau ci-dessous, nous représentons le Chiffre d'Affaires (CA) Hors Taxes (HT) en 2019 et le nombre de contrats par risque :

Risque	CA BP	Nombre de contrats BP	CA CE	Nombre de contrats CE
AT	7 775 501	17 124	4 339	13
DC ACC	36 870 732	2 266 603	1 214 965	529 980
DC TC	91 030 247	1 466 815	25 033 540	723 366
DEP	6 201 857	16 562	/	/
MAV	33 037 653	279 272	/	/
OBSQ	11 618 757	55 245	43 301 481	102 305
PP	83 096 128	4 291 332	/	/

TABLE 1 – Chiffre d'affaires hors taxes et nombre de contrats par risque

Toutes les garanties mentionnées sont proposées aux clients par l'intermédiaire de différents produits. Certains produits ne comportent qu'une seule garantie comme le produit obsèques, par exemple, et d'autres, regroupent plusieurs garanties. Les produits commercialisés par Natixis Assurances ont différentes cibles (clients particuliers ou professionnels) mais aussi différents types de prestations (forfaitaires, indemnitaires ou rentes).

Pour des raisons historiques, nous retrouvons deux types de gestion :

- La Gestion Centralisée (GC) : Il s'agit d'une gestion nationale des produits par Natixis, celle-ci fixe le tarif des produits et effectue leur suivi.
- La Gestion Déléguée (GD) : Dans ce type de gestion, la banque peut participer à la détermination du prix des produits et elle assurera une partie de leur gestion.

Au moment de la constitution du réseau Banque Populaire en assurance, une partie des produits a été créée en gestion centralisée car certaines BP n'étaient pas équipées pour de la gestion déléguée. Le fait de créer des produits en gestion centralisée a, par ailleurs, permis de proposer des tarifs nationaux harmonisés entre les différentes banques régionales. Le premier distributeur des produits Natixis Assurances était le réseau BP car la partie assurance des Caisses d'Épargne ne s'est pas intégrée tout de suite au réseau. Au fur et à mesure du temps, les BP ont développé la commercialisation des produits en gestion déléguée car cela leur permettait plusieurs choses :

- Elles pouvaient demander des augmentations de tarif indépendamment des autres BP (c'est-à-dire qu'elles n'avaient pas à respecter un tarif national).
- Elles pouvaient augmenter leur part de commissions fixes afin de mieux couvrir les frais de gestion (augmentation du Produit Net Bancaire).
- Cela facilitait la souscription des produits de prévoyance ou d'épargne lors de l'ouverture de comptes bancaires. En effet, le fait de pouvoir gérer directement les contrats permet d'accélérer le processus de vente.

Nous noterons que certains produits essentiellement commercialisés en GD existent encore en GC. C'est, par exemple, le cas d'anciens produits dont Natixis Assurances gère le stock ou de produits dont le montant garanti dépasse le montant maximum accepté en gestion déléguée. En effet, le montant garanti maximum accepté pour une souscription en GD est de 75 000 €, or, certains produits proposent une couverture jusqu'à 100 000 €, cette couverture est donc uniquement proposée en GC.

3 Partage de la valeur

Dans un premier temps, nous nous sommes concentrés sur le partage de la valeur pour Natixis Assurances. L'objectif de cette partie était de présenter deux indicateurs de la rentabilité : le Produit Net Assurance et le Produit Net Bancaire. Après cela, nous avons expliqué comment nous avons construit une base par tête permettant d'observer ces deux indicateurs à différentes mailles et nous avons également regardé le ratio PNA/PNB, notamment à l'aide de cartographies.

Le partage de la valeur est une notion importante pour un bancassureur, cela représente la répartition de la rentabilité des contrats d'assurance entre la banque et l'assurance. En effet, c'est un grand enjeu pour un bancassureur d'assurer le bon équilibre entre le Produit Net Assurance et le Produit Net Bancaire. En tant que bancassureur, les apporteurs d'affaires sont les banques, ainsi, soigner le PNB a toute son importance afin que les banques vendent le maximum de contrats d'assurance possible. Un bancassureur tel que Natixis Assurances, doit respecter un certain partage de la valeur, sinon, les deux réseaux (Banque Populaire et Caisse d'Épargne) auront moins d'intérêt à commercialiser des contrats pour Natixis Assurances. En effet, les contrats d'assurance gérés par Natixis Assurances sont les contrats détenus par les clients de ses deux réseaux, elle doit donc partager la rentabilité de ses contrats avec les différents établissements bancaires des deux réseaux.

Le partage de la valeur représente donc l'équitabilité entre deux rentabilités :

- Le Produit Net Assurance : Indicateur de la rentabilité de l'activité d'assurance pour l'assureur.
- Le Produit Net Bancaire : Indicateur de la rentabilité de l'activité d'assurance pour la banque. L'assureur doit verser des commissions (commissions d'acquisition et Commission Sur Résultat (CSR)) à la banque. Les commissions d'acquisition correspondent aux frais que l'assureur va payer à la banque pour l'acquisition de chaque contrat. Ces frais sont un pourcentage des primes perçues par l'assureur et sont payés périodiquement. La commission sur résultat représente la partie du résultat réalisé par l'assureur qui doit être reversée à la banque.

Dans la suite de ce mémoire, lorsque nous évoquerons le chiffre d'affaires, il s'agira du chiffre d'affaires hors taxes qui correspond aux primes hors taxes perçues par Natixis Assurances.

Après la description des deux indicateurs de rentabilité, il nous a semblé intéressant d'étudier le ratio de ces deux indicateurs. Pour Natixis Assurances, un ratio PNA/PNB est équitable lorsqu'il est égal à 1, c'est-à-dire que le PNA est égal au PNB. Privilégier le PNB des banques au détriment du PNA est intéressant pour les banques apportant le plus d'affaires dans le sens où, même si le ratio PNA/PNB n'est pas en faveur de l'assureur, i.e., le ratio est inférieur à 1, le chiffre d'affaires obtenu est conséquent. Afin de plus se rémunérer, il faut augmenter la prime d'assurance et ainsi contrôler ses marges vis-à-vis de la concurrence. En effet, à vouloir trop se rémunérer et avoir des contrats plus rentables, nous perdons des clients et donc du chiffre d'affaires. L'enjeu est donc triple entre la rémunération des banques, la rémunération de l'activité d'assurance et la concurrence.

Dès lors, le partage de la valeur entre l'assureur et la banque est important mais un bon ratio en faveur de l'assureur ne signifie pas plus de chiffre d'affaires, par exemple, dans le cas où la banque n'apporte pas beaucoup d'affaires.

Afin d'étudier le PNA et le PNB, nous avons construit une base de données permettant d'obtenir ces indicateurs de rentabilité par tête. En effet, le fait de posséder des données par tête nous a permis de les agréger aux mailles souhaitées, puis, de comparer le ratio PNA/PNB par banque et d'observer l'évolution du PNA en fonction de l'ancienneté des contrats, par exemple.

3.1 Utilisation d'une base permettant de calculer le PNA

L'objectif de cette sous-partie est de présenter la manière dont nous avons construit la base de données qui nous a permis de calculer le Produit Net Assurance pour nos contrats sur le périmètre DCAT (Décès Toutes Causes, Décès Accidentel et Arrêt de Travail). Cette base de données a ensuite permis de prédire le PNA à un an à l'aide de méthodes de Machine Learning qui vont

être évoquées dans la partie 5. Par ailleurs, la base de données construite nous a également permis d'obtenir un Produit Net Bancaire par tête ce qui permet de suivre la rentabilité du côté des banques pour chaque client.

3.1.1 Création de la base

Nous avons décidé de déterminer la rentabilité de l'assureur et de la banque en commençant par un calcul du PNA et du PNB par tête. En effet, tous les ans, les calculs de ces deux indicateurs sont réalisés par famille de produits, code société, réseau ou même risque mais aucun calcul par tête n'est effectué. Or, si nous souhaitons agréger nos données de quelque manière que ce soit et si nous voulons être capable d'observer l'évolution du PNA ou du PNB par année d'effet ou la répartition de ces deux indicateurs par banque, avoir des données par tête peut se révéler très utile. En outre, afin de déterminer la rentabilité a posteriori à l'aide de méthodes de Machine Learning, il est important d'avoir le plus de données possibles et donc de ne pas regrouper les contrats. Voici ce que représentent les termes précédemment employés :

- Famille de produits : Il s'agit des noms des produits commercialisés par les réseaux Banque Populaire et Caisse d'Épargne.
- Code société : Ce code est égal à 109 dans le cas de risques concernant les contrats décès (DC TC, une partie des DC ACC et obsèques) et il est égal à 124 pour le reste des garanties proposées par Natixis Assurances.
- Réseau : Il vaut "BP" si le réseau concerné est Banque Populaire et "CE" s'il s'agit des Caisses d'Épargne.
- Risque : Il vaut "DC TC" s'il s'agit du risque Décès Toutes Causes, "DC ACC", s'il s'agit du risque Décès Accidentel et "AT" s'il s'agit du risque Arrêt de Travail.

Afin de réaliser nos calculs, nous avons dû regrouper plusieurs bases de données. Il s'agit uniquement des données de la comptabilité en gestion centralisée et nous avons gardé les risques suivants : Décès Toutes Causes, Décès Accidentel et Arrêt de Travail. Ces bases de données nous ont permis d'obtenir un certain nombre d'informations sur les contrats. Voici les variables présentes dans ces bases et qui nous ont été utiles :

- Année comptable : Il s'agit du 31/12 de l'année à laquelle nous regardons le contrat.
- Client : C'est le numéro de client.
- Contrat : C'est le numéro de contrat associé au client.
- Code société.
- Réseau.
- Homogeneous Risk Groups (HRG) : Il s'agit de numéros associés aux différents risques, selon leur code société et leur réseau. Ainsi, le DC ACC 109 en BP n'aura pas le même HRG que le DC ACC 109 en CE.
- Produit Commercial (PC) : Il s'agit d'un code de produit associé au contrat.
- Produit informatique : C'est un autre code de produit associé au contrat.
- Famille de produits.
- Risque.
- Garantie Élémentaire (GE) : Il s'agit d'un code qui est associé au risque assuré. Il existe plusieurs codes différents pour un même risque.
- Garantie Commerciale Simple (GCS) : Tout comme la GE, il s'agit d'un code qui est associé au risque assuré. Il existe plusieurs codes différents pour un même risque.
- Date d'effet : Il s'agit de la date à laquelle le contrat a pris effet.
- Date de naissance : C'est la date de naissance de l'assuré.
- Sexe : C'est le genre de l'assuré, cette variable vaut 1 si l'assuré est un homme et 2 sinon.
- Périodicité : Il s'agit de la fréquence de règlement des primes. Cette variable peut valoir : 1, 2, 4 ou 12 selon que les primes sont annuelles, semestrielles, trimestrielles ou mensuelles.
- Montant assuré : Il s'agit du montant assuré au client en cas de sinistre.
- Montant de primes hors taxes : Il s'agit des primes périodiques payées par l'assuré.
- Taxes : Montant des taxes associées aux primes.

- Commissions d'acquisition : Il s'agit du montant des commissions d'acquisition payées périodiquement par l'assureur à la banque.
- Chargements de gestion : C'est le montant des chargements de gestion imputés aux primes périodiques. Ils permettent de couvrir les futurs coûts de gestion de la vie du contrat.
- Montant de primes hors taxes annuel : Il s'agit du montant total de primes versées par l'assuré sur l'année.
- Commissions d'acquisition annuelles : Il s'agit du montant des commissions d'acquisition payées annuellement par l'assureur à la banque.
- Chargements de gestion annuels : C'est le montant des chargements de gestion annuels.
- Top affaires nouvelles : Il s'agit d'une variable qui vaut 1 si le contrat est une affaire nouvelle et 0 sinon.
- Nom établissement bancaire : Il s'agit de l'agence bancaire régionale à laquelle est rattachée le client.

Nous noterons que certaines variables ont été utilisées pour en construire de nouvelles, il s'agit, par exemple, du montant de primes HT qui a permis de déterminer le PNA comme nous le verrons plus tard, tandis que, d'autres ont été utilisées comme variables pour aider les prédictions des modèles de Machine Learning, c'est par exemple le cas du montant assuré. En effet, le montant de primes HT a été utilisé pour déterminer d'autres variables et a également été utilisé comme variable prédictive des modèles de Machine Learning alors que le montant assuré a uniquement été utilisé pour la prédiction des modèles de ML.

Dans le but d'avoir un historique à la fois assez long et complet, nous avons sélectionné les données des contrats présents entre 2016 et 2019, cela signifie que nous avons censuré nos données à gauche. En effet, la période que nous avons définie ne comprend pas les contrats sur toute leur durée de vie, nous n'observons donc les contrats que sur une durée déterminée à l'avance. Si un contrat a une date d'effet antérieure à 2016, nous n'allons pas l'observer avant le 01/01/2016, la donnée est donc censurée à gauche.

Nous noterons, par ailleurs, que si un contrat a une date d'effet au 03/04/2014 et une date de clôture au 31/10/2015, il n'apparaîtra pas dans notre base de données. De plus, à la fin de notre observation (au 31/12/2019), tous les contrats n'ont pas été clôturés, ainsi, pour ces contrats, nous dirons que les données ont été censurées à droite. Dès lors, la censure des contrats peut avoir lieu à gauche, à droite, voire même, à gauche et à droite, c'est par exemple le cas d'un contrat ouvert le 07/08/2013 qui n'a pas encore été clôturé au 31/12/2019.

Nous avons donc mentionné les contrats censurés à gauche et/ou à droite, cependant, les contrats censurés à gauche proviennent nécessairement du réseau Banque Populaire car les premiers contrats distribués par le réseau Caisse d'Épargne datent de 2016, ils ne peuvent donc pas être censurés à gauche mais peuvent l'être à droite.

En plus de la censure, il est à noter qu'il peut également y avoir des troncatures. Dans ces cas-là, l'information est entièrement perdue alors que, dans le cas de la censure (à droite ou à gauche), nous perdons une partie de l'information mais nous conservons le reste. Une troncature aura lieu si un client décède ou si le contrat est clôturé et disparaît donc de notre base de données, il est donc nécessaire de contrer au maximum la troncature.

Nous avons utilisé trois bases portefeuilles qui sont les suivantes :

- La base portefeuille GC 2016-2017
- La base portefeuille GC 2017-2018
- La base portefeuille GC 2018-2019

Nous avons commencé par accorder ces trois bases, i.e., nous avons vérifié que l'année 2017 de 2016-2017 était bien égale à l'année 2017 de 2017-2018 et, de même, pour l'année 2018. Au départ, ce n'était pas le cas tête par tête car des agrégations avaient été faites de manière différente selon les années. En effet, dans la base 2018-2019, en plus d'être séparées selon les risques, les lignes des contrats l'étaient selon la Garantie Commerciale Simple. Cependant, dans la base 2017-2018, elles n'étaient séparées que par risque. Dès lors, il y avait nettement plus de lignes dans la base

2018-2019 que dans l'autre. Nous avons donc agrégé ces lignes afin de n'obtenir qu'une ligne par risque comme pour la base 2017-2018 (la base 2016-2017 était elle aussi agrégée de cette manière), cela nous a donc permis d'obtenir des bases homogènes. Nous avons, par ailleurs, pu vérifier que l'agrégation en sommant les montants assurés, les montants réassurés, les capitaux sous risques, les primes, les commissions d'acquisition et les chargements de gestion par les autres colonnes telles que le numéro de client, le numéro de contrat, le risque et l'année comptable donnait bien les mêmes valeurs sur la base 2018-2019 que celles obtenues sur la base 2017-2018 pour l'année comptable 2018. Au niveau de l'année 2017 présente dans les bases 2016-2017 et 2017-2018, nous avons remarqué que les valeurs étaient bien identiques et déjà agrégées de la même manière donc nous n'avons pas eu à modifier ces deux bases.

Une fois que nos bases étaient en accord, nous avons dû supprimer les doublons avant de les réunir, sinon, nous aurions eu les clients de 2017 et de 2018 en double. Nous avons donc gardé l'année 2016 de la base 2016-2017, l'année 2017 de la base 2017-2018 et les années 2018 et 2019 de la base 2018-2019, puis, nous avons joint les trois bases et gardé l'année comptable pour chaque ligne : un contrat présent entre 2016 et 2019 et n'ayant pas été clôturé ni sinistré apparaîtra donc quatre fois pour chacun des risques pour lesquels le client est assuré. En effet, un même numéro de contrat et de client peut être associé à plusieurs risques différents. Cependant, comme nous l'avons vu plus tôt, les clients décédés et les contrats clos l'année N n'apparaissent plus dans la base à l'année N mais ils peuvent apparaître à l'année N-1 s'ils avaient déjà été souscrits à ce moment-là.

Afin de palier à la troncature des clients décédés et des clôtures, nous avons récupéré les données des contrats sinistrés, i.e., les contrats pour lesquels un sinistre est survenu, et également les données des contrats clôturés dans les fichiers de portefeuilles intermédiaires. Dès lors, si un client décède en cours d'année, il ne sera pas présent au 31/12/N dans la base portefeuille mais il pourra être présent au 31/09/N, par exemple, dans un fichier de portefeuille intermédiaire. En effet, des fichiers de portefeuilles intermédiaires sont créés à chaque fin de trimestre. Cela nous a permis de récupérer bon nombre de données sinistrées ou résiliées et ainsi d'éviter la troncature de ces données.

Nous avons donc ajouté, à notre base, les contrats présents dans les fichiers intermédiaires (mars, juin et septembre) des différentes années observées et qui n'étaient pas présents dans le fichier final de l'année. Nous nous sommes donc retrouvés avec une base de données contenant :

- La base portefeuille GC 2016-2017
- La base portefeuille GC 2017-2018
- La base portefeuille GC 2018-2019
- Les bases portefeuilles intermédiaires de 2016
- Les bases portefeuilles intermédiaires de 2017
- Les bases portefeuilles intermédiaires de 2018
- Les bases portefeuilles intermédiaires de 2019

Nous avons ensuite joint cette base de données au fichier recensant les sinistres en cours depuis 2014. La jointure a été effectuée par année comptable, code société, numéro de client, numéro de contrat, réseau, risque, famille de produits et code de produit. Il existe deux codes de produit donnés aux contrats, un code de produit commercial et un code de produit informatique, nous avons joint nos deux bases à l'aide de ces deux codes de produit.

Cette jointure nous a permis d'obtenir plusieurs variables importantes pour la suite :

- La date de survenance du sinistre
- La date de déclaration du sinistre
- Les provisions de sinistres de fin d'année
- Les prestations versées pendant l'année

Après cela, nous avons pu remarquer qu'il manquait également les données sinistrées pour un contrat qui, par exemple, aurait été souscrit en octobre N et aurait été sinistré en novembre N. Le contrat n'ayant pas été dans le portefeuille au 31/03/N, ni au 30/06/N, ni au 30/09/N, ni au 31/12/N, il n'apparaissait ni dans les fichiers intermédiaires, ni dans le fichier de portefeuille final. Pourtant, une prestation a pu être versée pour le sinistre et ce contrat peut donc fortement

impacter le PNA de l'année. Nous avons donc dû aller chercher ce type de contrats dans la base historisée qui recense les contrats en cours, clôturés et sinistrés avec une date de clôture et une date de survenance des sinistres. Seulement, dans cette base de données, il n'y avait pas le montant des commissions d'acquisition et le montant des chargements de gestion et nous avons besoin de ces données pour calculer le PNA et le PNB par tête. Dès lors, en sachant que les montants de commissions d'acquisition et de chargements de gestion sont calculés à partir de la prime HT, nous avons calculé les quotients suivants par Garantie Élémentaire, Année d'effet et Âge comptable (différence entre la date comptable (31/12/N) et la date de naissance de l'assuré) :

$$\frac{\text{Montant de commissions annuel}}{\text{Montant de primes HT annuel}} \text{ et } \frac{\text{Montant de chargements de gestion annuel}}{\text{Montant de primes HT annuel}}.$$

Ces calculs nous ont permis d'ajouter une approximation du montant de commissions d'acquisition et de chargements de gestion sur la base historisée. Par ailleurs, le montant de primes dans la base historisée était le montant TTC (Toutes Taxes Comprises), nous avons donc dû retirer les taxes avant d'ajouter l'approximation des commissions d'acquisition et des chargements de gestion. Or, nous savons que, pour le risque AT, les taxes peuvent être de 9% ou de 14%, pour le risque DC ACC, les taxes sont de 9% et, pour le DC TC, il n'y a pas de taxes, le montant HT était donc égal au montant TTC pour ce risque.

Par ailleurs, de la même manière que nous avons dû joindre nos données de portefeuille GC à la base recensant les sinistres, il a également fallu joindre les données de la base historisée avec la même base afin de récupérer les provisions de sinistres de fin d'année ainsi que les prestations versées pendant l'année. Cependant, dans la base historisée, les dates de survenance et de déclaration étaient déjà renseignées donc nous n'avons pas besoin de les récupérer à partir de la base recensant les sinistres, nous avons, au contraire, pu les utiliser pour la jointure des deux bases. Ainsi, après quelques retraitements, nous avons pu joindre la base historisée à notre base portefeuille GC.

Grâce aux données de clôture et de survenance des sinistres, nous avons pu calculer le temps de présence des clients. Ce temps de présence sur l'année est compris entre 0 et 1 et vaut 1 si le client était présent toute l'année. Il vaut 0 si le client n'était plus présent pendant l'année, c'est-à-dire, s'il est décédé avant le 01/01/N. En effet, un client ayant clôturé son contrat avant le 01/01/N n'apparaîtra pas dans la base de données de l'année N donc il n'aura pas de temps de présence rattaché à l'année N tandis que, s'il est décédé avant le 01/01/N, le sinistre est peut être encore en cours donc il est possible que le client apparaisse dans la base de données de l'année N mais sans pour autant être physiquement présent dans le portefeuille et, dans ce cas-là, la ligne d'observation appartient au portefeuille mais le temps de présence est nul. Nous noterons, par ailleurs, qu'une affaire nouvelle n'aura pas un temps de présence dans l'année égal à 1. En effet, un assuré arrivant dans le portefeuille de contrats le 01/02/2018 et ne sortant pas de ce portefeuille au 31/12/2018, aura un temps de présence d'environ 0,92 car il ne sera présent que 11 mois dans l'année.

Ainsi, grâce à ces temps de présence, nous avons pu calculer les montants de primes, de commissions d'acquisition et de chargements de gestion proratisés, i.e., pour reprendre l'exemple précédent du client qui adhère à un contrat le 01/02/2018, celui-ci ne payera que 92% de la prime annuelle. Il en va de même pour un assuré qui clôture son contrat ou qui décède avant la fin de l'année. Pour les différents calculs permettant d'arriver au PNA, nous avons donc utilisé les données proratisées. Toutefois, pour les modèles de Machine Learning permettant de prédire le PNA au 31/12/N à partir des données au 01/01/N, les variables explicatives des modèles contenaient les données non proratisées car, au 01/01/N, nous n'avons pas connaissance de ce qui va se produire durant l'année. Par exemple, les affaires nouvelles sont retirées du modèle de Machine Learning car nous ne pouvons pas savoir, au 01/01/N, que tel client va adhérer à tel contrat. De plus, si un client clôture son contrat, s'il décède ou s'il a un sinistre dans l'année, nous ne pouvons pas le savoir à l'avance.

Nous avons calculé quelques variables supplémentaires utiles aux modèles de Machine Learning. En effet, nous n'avons pas mis de dates en input des modèles mais plutôt des âges :

- *Ancienneté en mois = Date comptable - Date d'effet*
- *Âge adhésion = Date d'effet - Date de naissance*
- *Âge comptable = Date comptable - Date de naissance*
- *Ancienneté à la survenance en mois = Date de survenance - Date d'effet*
- *Âge à la survenance = Date de survenance - Date de naissance*

- *Ancienneté depuis la survenance en mois = Date comptable - Date de survenance*

Lors de la création de notre base de données, nous avons pu remarquer des erreurs. En effet, parfois un même code de produit commercial était rattaché à différentes familles de produits, or, ceci n'est pas possible. Il peut y avoir plusieurs codes de produit commercial pour une seule et même famille de produits, en revanche, il ne peut pas y avoir plusieurs famille de produits pour un seul code de produit commercial. Ainsi, nous avons dû aller chercher la vraie famille de produits associée à chaque code de produit commercial dans une table de correspondance afin de corriger ces erreurs. Par ailleurs, nous avons vu que la garantie Décès Accidentel peut être rattachée au code société 109 ou au code société 124, ceci peut créer des erreurs. Parfois, nous avons remarqué qu'un client assuré pour le Décès Accidentel avec le code société 109 pour l'année 2016 et 2017 passait au code société 124 en 2018 et revenait au code société 109 en 2019, par exemple. Dans ce type de cas, nous avons corrigé l'année 2018 en remplaçant le code société par 109 afin que le client soit toujours assuré en société 109.

A partir de notre base de données, nous avons pu calculer le PNA, le PNB et le Résultat Brut d'Exploitation (RBE) par tête.

3.1.2 Calcul du PNA, du PNB et du RBE et préparation de la base pour le Machine Learning

Nous avons commencé par calculer la variation de provisions de sinistres, pour cela, nous avons utilisé la colonne "Provisions" de notre base de données, correspondant aux provisions de fin d'année. Les provisions de début d'année N correspondent aux provisions de fin d'année N-1 donc nous avons pu récupérer les provisions de début d'année grâce aux provisions de fin d'année que nous avons. Pour les provisions de début d'année 2016, nous n'avons pas les provisions de fin d'année 2015 dans notre base car nous n'avons pas l'année comptable 2015 mais il a tout de même été possible de récupérer ces provisions grâce à une jointure (avec décalage d'un an) avec la base de sinistres précédemment utilisée. En effet, cette base de sinistres recense tous les sinistres survenus depuis 2014, elle contenait donc les provisions de fin d'année 2015 et nous avons alors pu joindre ces provisions avec l'année 2016 de notre base de données afin d'obtenir les provisions de début d'année 2016.

Voici les formules que nous avons utilisées pour calculer un proxy du PNA, du PNB et du RBE à partir de notre base de données :

- *Variation de provisions de sinistres = Provisions de sinistres de fin - Provisions de sinistres de début*
- *Primes pures acquises = Primes HT annuelles - Commissions d'acquisition - Chargements de gestion*
- *Charge de sinistres comptable = Prestations + Variation de provisions de sinistres*

Afin de réaliser la suite de ces calculs, nous avons dû joindre notre base de données à un fichier permettant d'obtenir certaines données en fonction du code société, du réseau et du risque. Grâce à cela, nous avons pu obtenir les taux suivants :

- Taux de cession de primes
- Taux de cession de prestations
- Taux de frais forfaitaires
- Taux de commission sur résultat
- Taux de commission sur résultat cédée
- Taux de Provision pour Égalisation
- Taux de frais généraux de primes
- Taux de frais généraux de prestations

Nous avons calculé un top (variable qui vaut 0 ou 1) pouvant être utile en input du modèle de ML, il s'agit d'un top qui va illustrer si un sinistre était en cours l'année N. Il vaut 1 s'il y avait un sinistre en cours l'année N et 0 si ce n'était pas le cas. Nous considérons qu'un sinistre était en cours l'année N s'il y a une provision de fin d'année N ou que des prestations ont été versées

pendant l'année N. Nous avons également pu calculer un deuxième top qui vaut 1 si le sinistre était en cours l'année N-1 et 0 si ce n'était pas le cas. Ce deuxième top a été créé afin d'être utilisé en remplacement du premier pour la prédiction du PNA au 31/12/N car nous ne pouvons utiliser que les données présentes dans la base au 01/01/N. Ainsi, il n'est pas possible d'utiliser un top qui dépend de l'année N pour prédire le PNA à la fin de l'année N, nous avons donc utilisé un top relatif à l'année précédente.

Nous avons ensuite utilisé les formules suivantes :

- $Primes\ TTC\ annuelles = Primes\ HT\ annuelles + Taxes$
- $Provision\ pour\ Égalisation = \min((1 - Taux\ de\ cession\ de\ primes) \times (Primes\ pures\ acquises - Charge\ de\ sinistres\ comptable) \times 0.75, (1 - Taux\ de\ cession\ de\ primes) \times (Primes\ TTC\ annuelles \times Taux\ de\ Provision\ pour\ Égalisation))$

Tout comme pour le calcul de la variation de provisions de sinistres, nous avons utilisé la colonne "Provision pour Égalisation" de notre base de données, correspondant à la provision pour égalisation de fin d'année, pour calculer la variation de provision pour égalisation. La provision pour égalisation de début d'année N correspond à la provision pour égalisation de fin d'année N-1 donc nous avons pu récupérer la provision pour égalisation de début d'année grâce à la provision pour égalisation de fin d'année que nous avons. Il en a été de même pour le calcul de la variation de provision de primes pures et celui de la variation du fonds de revalorisation.

Toutefois, pour ces variations, contrairement à ce qui a pu être fait pour la variation de provisions de sinistres, nous n'avions pas la provision (respectivement, le fonds de revalorisation) de début d'année 2016 et nous ne pouvions pas l'obtenir par tête. Nous allons donc décrire la manière dont nous avons obtenu une approximation de la provision pour égalisation de fin d'année 2015, et donc de début d'année 2016, par tête. Nous avons procédé de manière analogue pour ce qui était de la provision de primes pures de fin d'année 2015 et pour le fonds de revalorisation de fin d'année 2015.

Pour obtenir la provision pour égalisation de fin d'année 2015 par tête, nous avons utilisé la provision pour égalisation de fin d'année 2015 qui était présente dans les fichiers comptables agrégés par produit commercial, famille de produits, risque et année comptable. En effet, nous avons ces données pour l'année comptable 2015 et nous avons donc pu les ventiler par rapport au montant de provisions de fin d'année 2016 afin d'obtenir la provision de début d'année 2016 ainsi que la variation de provision sur l'année 2016. Il semble cohérent de penser que la provision de fin d'année N-1 est corrélée à celle de fin d'année N. Dès lors, si le total des provisions pour égalisation de fin d'année 2015 vaut 1 000 pour un certain produit commercial et qu'il vaut 1 200 en fin d'année 2016 pour un même produit commercial et si nous savons que la provision pour égalisation de notre client était de 120 à fin 2016, il est cohérent de lui associer une provision pour égalisation de 100 à fin 2015. C'est donc de cette manière que nous avons obtenu l'estimation de la provision pour égalisation à fin 2015.

Nous avons donc pu utiliser les formules suivantes :

- $Variation\ de\ provision\ pour\ égalisation = Provision\ pour\ égalisation\ de\ fin - Provision\ pour\ égalisation\ de\ début$
- $Primes\ pures\ cédées = Primes\ pures\ acquises \times Taux\ de\ cession\ de\ primes$
- $Prestations\ cédées = Charge\ de\ sinistres\ comptable \times Taux\ de\ cession\ de\ prestations$
- $Commission\ sur\ résultat\ cédée = (Primes\ pures\ cédées - Prestations\ cédées) \times Taux\ de\ commission\ sur\ résultat\ cédée$
- $Frais\ forfaitaires = Primes\ pures\ acquises \times Taux\ de\ cession\ de\ primes \times Taux\ de\ frais\ forfaitaires$
- $Solde\ de\ réassurance = Primes\ pures\ cédées - Prestations\ cédées - Commission\ sur\ résultat\ cédée + Frais\ forfaitaires$
- $Commission\ sur\ résultat = (Primes\ pures\ acquises - Charge\ de\ sinistres\ comptable - Solde\ de\ réassurance) \times Taux\ de\ commission\ sur\ résultat$
- $Variation\ de\ provision\ de\ primes\ pures = Provision\ de\ primes\ pures\ de\ fin - Provision\ de\ primes\ pures\ de\ début$

- $Variation\ du\ fonds\ de\ revalorisation = Fonds\ de\ revalorisation\ de\ fin - Fonds\ de\ revalorisation\ de\ début$

Grâce aux formules précédentes, nous avons pu obtenir les données nécessaires au calcul du PNA, du PNB et du RBE :

- $PNA = Primes\ HT\ annuelles - Variation\ de\ provision\ de\ primes\ pures - Commissions\ d'acquisition - Charge\ de\ sinistres\ comptable - Commission\ sur\ résultat - Variation\ de\ provision\ pour\ égalisation - Variation\ du\ fonds\ de\ revalorisation - Solde\ de\ réassurance$
- $PNB = Commissions\ d'acquisition + Commission\ sur\ résultat$
- $Frais\ généraux\ de\ primes = Primes\ HT\ annuelles \times Taux\ de\ frais\ généraux\ de\ primes$
- $Frais\ généraux\ de\ prestations = Prestations \times Taux\ de\ frais\ généraux\ de\ prestations$
- $RBE = PNA - Frais\ généraux\ de\ primes - Frais\ généraux\ de\ prestations$

Étant donné la formule du PNA, nous savons que le PNA sera négatif dans le cas de contrats sinistrés (avec des provisions et/ou des prestations positives) et qu'il sera positif dans le cas de contrats non encore sinistrés. En effet, la charge de sinistres est plus élevée que les primes de l'année dans le cas de contrats sinistrés, c'est pour cela que le PNA sera négatif.

Une fois la base de données prête, nous avons voulu vérifier sa cohérence par rapport à la base comptable agrégée par produit commercial, une maille très fine car il y a plus d'une centaine de codes différents. Ainsi, nous avons tout d'abord commencé par comparer les résultats par risque, code société et par année comptable que nous avons obtenus avec le portefeuille en gestion centralisée avec ceux obtenus par la base comptable. Nous avons, pour cela, comparé, le montant de primes HT, les commissions d'acquisition, la charge de sinistres comptable, la commission sur résultat et le solde de réassurance, i.e., les principaux éléments composant le PNA. Nous avons pu remarquer quelques légers écarts probablement dus au fait que nous avons dû assembler plusieurs bases afin de récupérer des données tronquées ou tout simplement dus à des erreurs d'association de code société, par exemple, comme nous l'avions mentionné plus tôt. Après cela, nous avons pu comparer le PNA, le PNB et le RBE en constatant toujours quelques écarts provoqués par les écarts présents dans les principaux composants de ces trois éléments de rentabilité. Nous avons décidé de ne pas corriger ces écarts car il s'agissait d'écarts très faibles ne pouvant pas impacter les prédictions des modèles de Machine Learning donc cela n'a pas été un problème de les garder. En effet, il aurait été très coûteux en temps de corriger ces écarts pour un gain minime voire même inexistant sur les performances des modèles de Machine Learning.

Par la suite, nous avons observé nos données afin d'avoir une idée de la distribution des primes et des règlements de sinistres selon les risques ou les familles de produits. Il est important de suivre le PNA et le PNB par famille de produits afin d'avoir une vue détaillée de la rentabilité de nos offres, de classer les familles par rentabilité et d'effectuer d'éventuelles revues tarifaires. Ainsi, nous avons observé le PNA, le PNB et également le RBE par famille de produits.

3.2 Distribution du chiffre d'affaires

Pour évaluer la distribution du chiffre d'affaires, nous avons commencé par observer sa répartition en fonction des différents risques étudiés, i.e., le Décès Toutes Causes, le Décès Accidentel et l'Arrêt de Travail. Nous avons obtenu le tableau suivant :

Risque	CA 2016 (k€)	CA 2017 (k€)	CA 2018 (k€)	CA 2019 (k€)
DC TC	54 165	64 996	77 217	89 358
DC ACC	15 237	16 541	17 850	18 858
AT	8 232	8 110	8 032	7 780

TABLE 2 – Distribution du chiffre d'affaires par risque

Pour les quatre années comptables étudiées, c'est le DC TC qui apporte le plus de chiffre d'affaires et l'AT qui en apporte le moins. Par ailleurs, le CA rapporté par le DC TC et le DC ACC a augmenté au fil des années alors que celui rapporté par l'AT a légèrement diminué. Cela signifie que, soit les primes ont diminué, soit il y a moins d'affaires nouvelles que de clôtures.

Nous avons ensuite étudié le poids des différentes familles de produits vis-à-vis du chiffre d'affaires rapporté en 2019 :

Famille de produits	Chiffre d'Affaires (k€)	Poids
FAMILLE FRUCTI-FAMILLE	47 524	41,0%
SECUR FAMILLE CE DC PTIA	23 055	19,9%
FRUCTI-HOMME CLE	21 309	18,4%
FRUCTI-PROFESSIONNEL	11 736	10,1%
FRUCTI-PATRIMOINE	4 407	3,8%
SECUR URGENCE CE	2 878	2,5%
FAMILLE ASSURANCE FAMILLE	1 071	0,9%
DEDICACE	1 017	0,9%
FAMILLE FRUCTI-BUDGET	1 015	0,9%
D.S.EPARGNE SALARIALE	687	0,6%
ASSURANCE HOMME-CLE C.E.	319	0,3%
FRUCTI-FACILITES AGRI	308	0,3%
CREDIT MARITIME VIE	267	0,2%
FAMILLE FRUCTI-FACILITES	123	0,1%
CERTI CAPITAL	85	0,1%
FRUCTI-FACILITES PRO	78	0,1%
SECURI-CREDIT BPROP	72	0,1%
FAMILLE DOUBLE SECURITE	40	0,0%
FAMILLE FRUCTI-SECURITE	5	0,0%

TABLE 3 – Poids des différentes familles de produits en 2019

Ici, nous pouvons remarquer que les deux familles de produits qui rapportent le plus de chiffre d'affaires sont Famille Fructi-Famille, qui appartient aux Banques Populaires, et Secur Famille CE DC PTIA, qui appartient aux Caisses d'Épargne. Ainsi, le CA rapporté par ces deux familles est supérieur à 60% du CA total. En observant uniquement les quatre premières familles de produits, nous pouvons comprendre la provenance de la quasi-totalité du chiffre d'affaires, les autres familles de produits sont relativement faibles à côté de celles-ci.

3.3 Distribution des règlements par risque

Au vu de la répartition du CA par risque que nous avons obtenue sur le tableau 2, il peut être intéressant de regarder la distribution des règlements à cette même maille. En effet, nous avons pu constater que c'est le risque DC TC qui rapporte le plus de CA, nous allons donc regarder ce qu'il en est des règlements pour chacun des trois risques étudiés :

Risque	Règlements 2016 (k€)	Règlements 2017 (k€)	Règlements 2018 (k€)	Règlements 2019 (k€)
DC TC	14 283	14 856	15 251	17 700
AT	2 599	2 623	2 749	2 209
DC ACC	1 353	1 867	2 721	1 645

TABLE 4 – Évolution des règlements par risque

C'est sur la garantie DC TC que l'on retrouve le plus de règlements, cela est cohérent car il s'agit du risque qui rapporte le plus de CA et beaucoup de clients possèdent cette garantie. Or, plus il y a de clients, plus le risque de survenance d'un sinistre est élevé donc cela augmente les règlements. Toutefois, le risque arrêt de travail est celui qui rapporte le moins de CA, pourtant, il y a plus de règlements pour cette garantie que pour le décès accidentel, et ce, peu importe l'année comptable que nous observons. Cela peut être expliqué par le fait que les sinistres sont rares en DC ACC car il s'agit de causes de décès particulières, il y a un taux de refus très élevé pour cette garantie. En effet, les bénéficiaires des contrats vont souvent déclarer un décès accidentel pour la personne assurée alors qu'il s'agissait en fait d'un décès causé par la vieillesse, par exemple. La garantie de décès accidentel inclut le décès soudain, violent et brutal et dont la cause n'est pas médicale.

3.4 Distribution du ratio PNA/PNB par banque

Afin de mieux visualiser le partage de la valeur sur nos réseaux au sein de toute la France, nous avons réalisé une application R Shiny permettant d'analyser des résultats à une maille très détaillée. Cette application nous a permis de réaliser des cartographies représentant le partage de la valeur. En effet, ces cartographies permettent de voir comment se répartissent le PNA et le PNB sur la France mais également comment se répartit le ratio PNA/PNB.

Shiny est un package R qui permet la création de pages web interactives. Normalement, une application Shiny est composée de deux fichiers : un fichier "ui.R" ainsi qu'un fichier "server.R".

La notre était constituée de plusieurs parties "ui" et "server" imbriquées afin de réaliser une application Shiny évoluée.

La partie "ui" (User Interface script) permet de contrôler la mise en page et l'apparence de l'application et la partie "server" (Server script) est la partie contenant les instructions dont l'ordinateur a besoin pour construire l'application.

Tout d'abord, nous avons observé le ratio PNA/PNB pour toutes les familles de produits et tous les risques sur quatre années d'observation (2016 à 2019) :

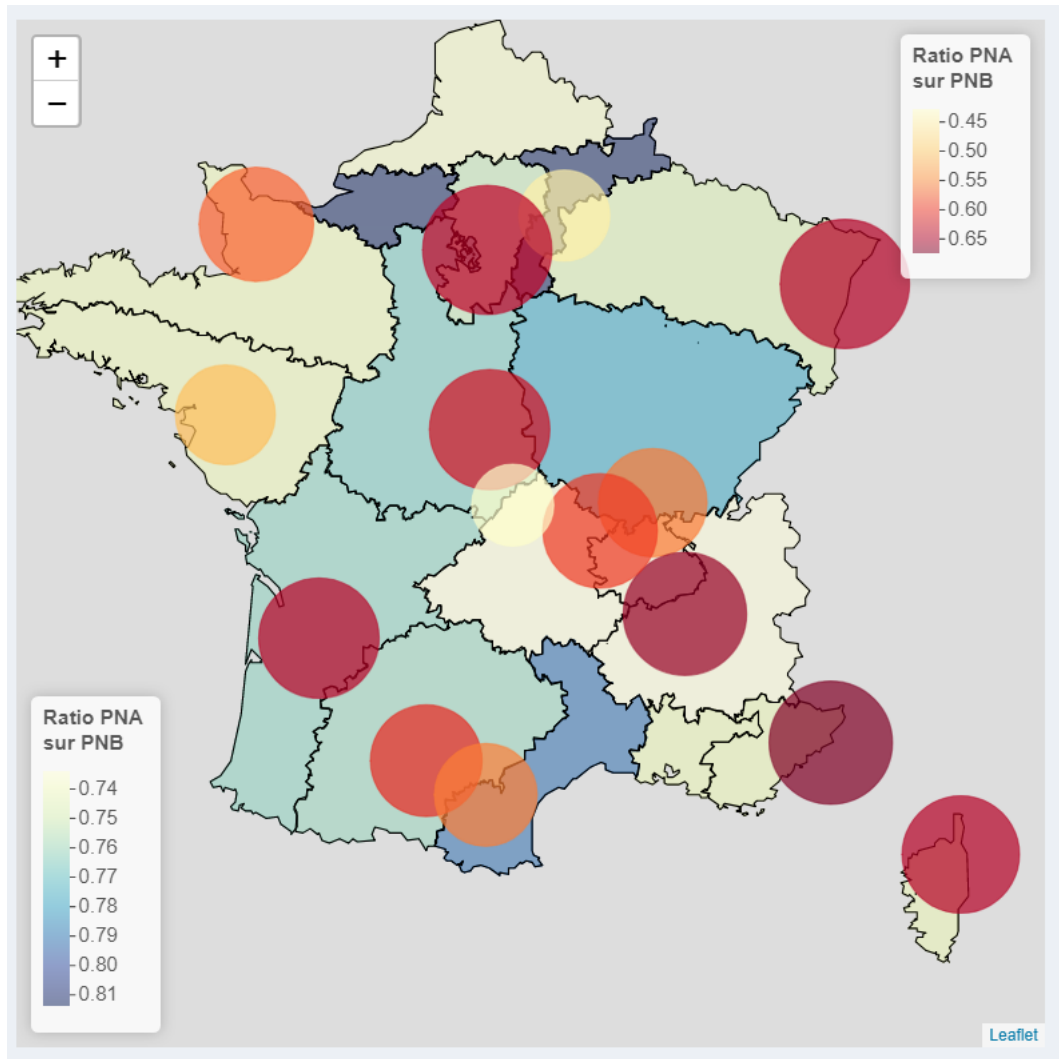


FIGURE 3 – Ratio PNA/PNB pour toutes les familles de produits et tous les risques 2016-2019

Sur cette carte de France, nous pouvons regarder, d'une part, le partage de la valeur entre le PNA et le PNB pour les Banques Populaires (régions en bleu) et, d'autre part, pour les Caisses d'Épargne (cercles rouges). Plus le bleu (respectivement le rouge) est foncé et plus le ratio est proche de 1 ce qui signifie que le partage entre l'assureur et la banque est équitable. À l'inverse, plus les couleurs sont claires et plus le ratio est déséquilibré, le PNB est alors plus élevé que le PNA car $\frac{PNA}{PNB} < 1$ et la rentabilité de la banque est privilégiée par rapport à celle de l'assureur. Ainsi, pour les Banques Populaires, c'est dans le nord et dans le Languedoc-Roussillon que le partage de la valeur est le plus équitable, au contraire, il ne l'est pas dans l'ouest et dans le sud-est. Quant aux Caisses d'Épargne, c'est dans le sud-est que le ratio est le plus équitable.

Nous pouvons maintenant nous intéresser au partage de la valeur sur la plus grande famille BP (Famille Fructi-Famille) et CE (Secur Famille CE DC PTIA) pour l'année 2019 :

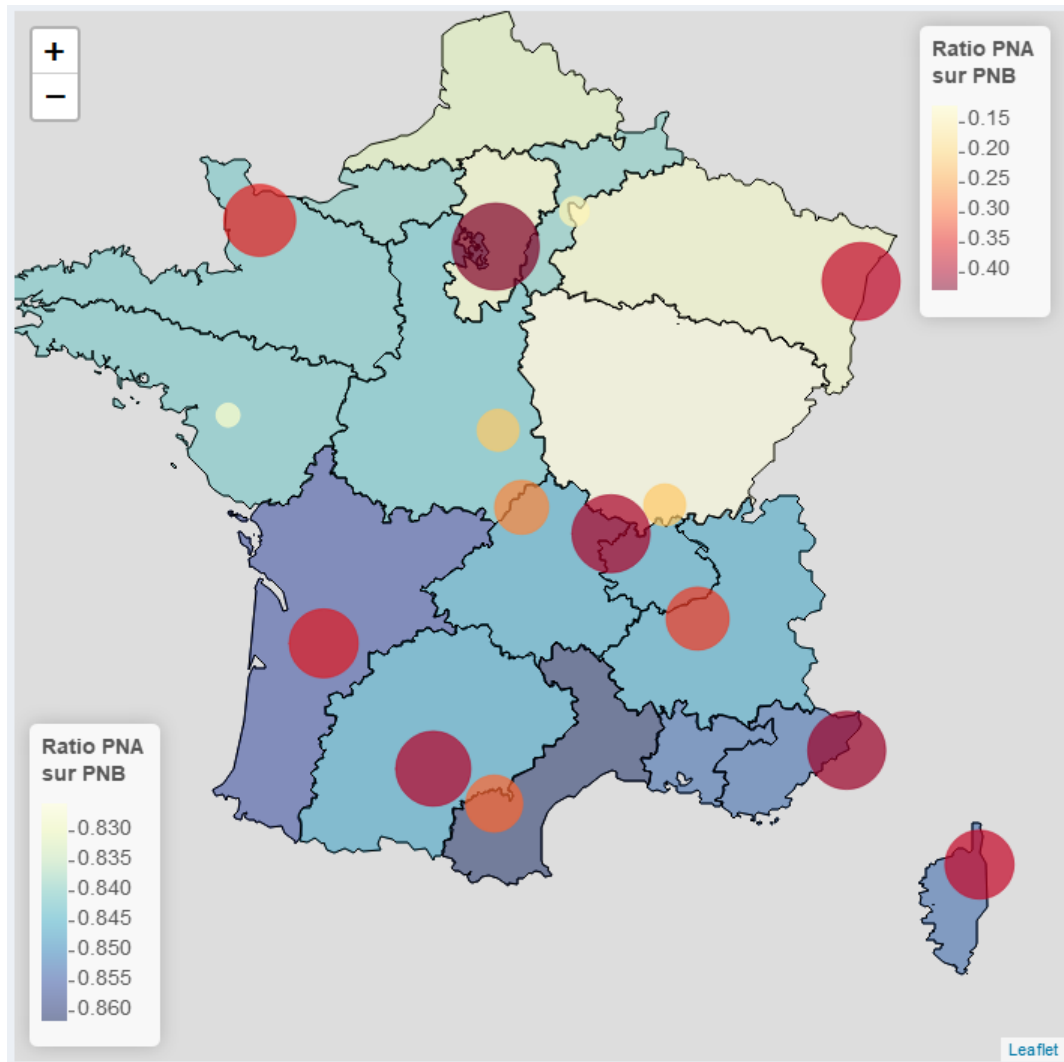


FIGURE 4 – Ratio PNA/PNB des plus grandes familles BP et CE 2019

Pour les BP, tous les ratios sont assez proches car ils sont compris entre 0,83 et 0,86 mais c'est dans le sud que nous trouvons les ratios les plus élevés. Pour les CE, les ratios sont très faibles, ils sont toujours inférieurs à 0,4, ce qui signifie que le PNA est très petit devant le PNB, le partage de la valeur n'est pas équitable dans ce cas. Cependant, comme nous l'avons précisé plus haut, ce n'est pas parce que le ratio n'est pas équitable pour l'assureur qu'il va récupérer très peu de chiffre d'affaires. En effet, les CE rapportent beaucoup d'affaires à l'assureur pour la garantie Secur Famille CE DC PTIA.

Nous allons ensuite nous intéresser à l'évolution du partage de la valeur pour le risque Décès Toutes Causes entre 2016 et 2019 :

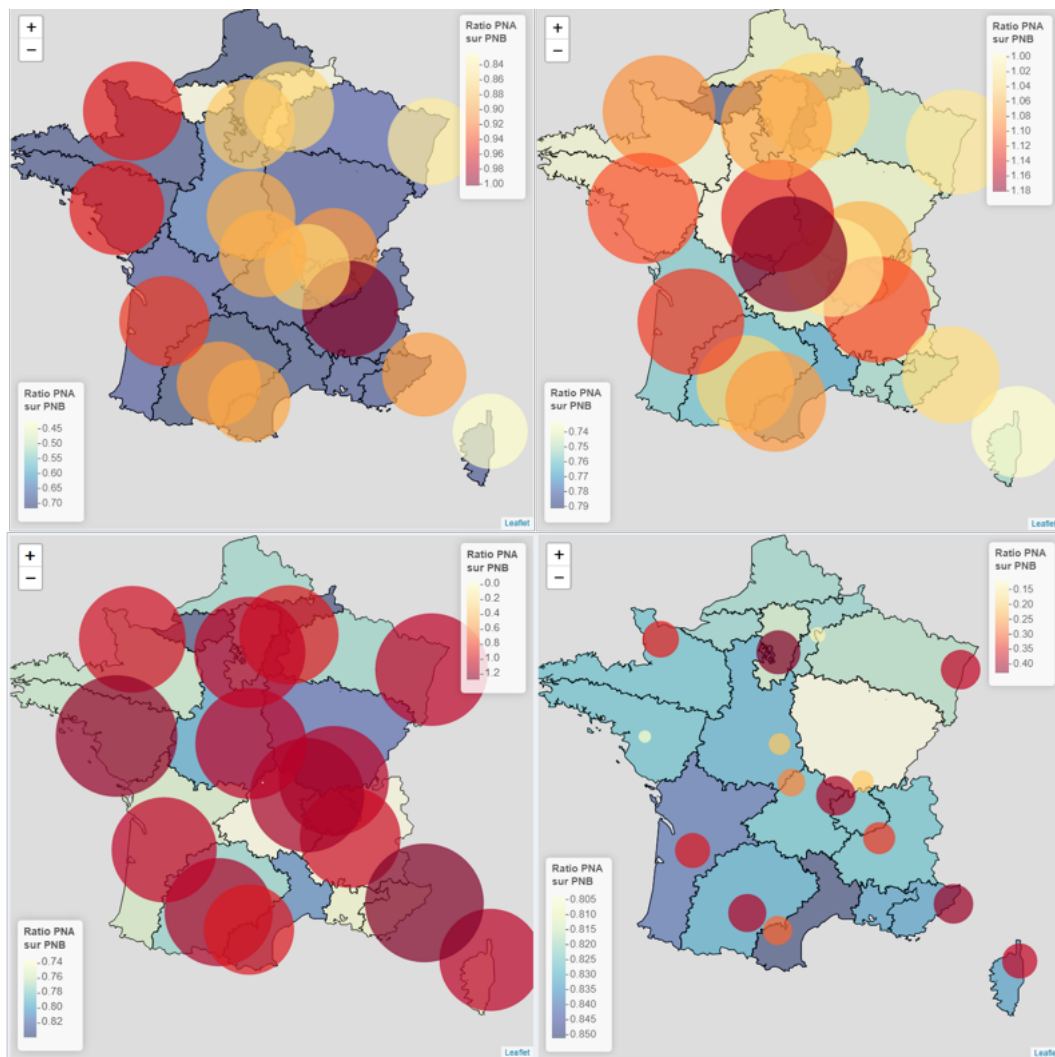


FIGURE 5 – Évolution du ratio PNA/PNB DC TC de 2016 à 2019

Pour les BP, nous pouvons voir que le partage de la valeur est devenu plus équitable au fil des années. En effet, le ratio PNA/PNB a augmenté de plus en plus entre 2016 et 2019.

Pour les CE, le ratio PNA/PNB a bien évolué tout au long des années. Nous pouvons voir qu'il a augmenté entre 2016 et 2017 pour ensuite rester assez stable en 2018 et diminuer en 2019. Le ratio PNA/PNB a donc été très équitable pendant les trois premières années sur toutes les régions de France. Cependant, en 2019, le PNA est devenu inférieur à 40% du PNB, nous pouvons donc en déduire que, soit le PNA a beaucoup diminué, soit le PNB a beaucoup augmenté. En réalité, c'est le PNA qui a diminué à cause de la montée en charge du portefeuille CE. En effet, comme nous l'avons déjà évoqué, les contrats d'assurance pour les CE ne sont commercialisés que depuis 2016, la sinistralité n'est donc pas encore en régime de croisière et cela rend le provisionnement peu précis ce qui impacte fortement le ratio PNA/PNB.

En plus de la vision du ratio PNA/PNB sur les cartographies, nous avons réalisé un tableau permettant de comparer le ratio au fur et à mesure des années par banque. Les différents noms d'établissements bancaires BP et CE ont été renommés afin de simplifier la compréhension du tableau :

Banque	PNA/PNB 2016 (en %)	PNA/PNB 2017 (en %)	PNA/PNB 2018 (en %)	PNA/PNB 2019 (en %)
Banque Populaire 1	68,37	74,75	77,43	77,06
Banque Populaire 2	72,34	75,29	75,97	79,84
Banque Populaire 3	69,94	73,02	73,56	77,21
Banque Populaire 4	72,95	75,59	82,20	79,60
Banque Populaire 5	68,95	72,16	77,40	76,71
Banque Populaire 6	73,39	77,46	81,18	82,70
Banque Populaire 7	69,96	72,99	76,49	77,62
Banque Populaire 8	71,14	72,68	74,54	79,30
Banque Populaire 9	71,60	75,79	77,70	78,98
Banque Populaire 10	70,38	73,46	78,47	78,81
Banque Populaire 11	68,34	74,88	80,06	81,60
Banque Populaire 12	82,53	80,50	83,88	79,36
Caisse d'Épargne 1	99,35	99,87	104,54	21,45
Caisse d'Épargne 2	93,53	102,72	110,43	25,27
Caisse d'Épargne 3	95,35	98,47	122,31	7,04
Caisse d'Épargne 4	90,33	96,14	127,20	33,32
Caisse d'Épargne 5	89,34	111,18	1,11	19,55
Caisse d'Épargne 6	90,72	95,57	112,45	1,57
Caisse d'Épargne 7	90,36	95,14	116,00	29,53
Caisse d'Épargne 8	82,65	92,34	111,44	30,43
Caisse d'Épargne 9	87,03	92,45	96,35	1,79
Caisse d'Épargne 10	84,89	93,30	109,98	30,62
Caisse d'Épargne 11	87,89	96,94	108,50	35,69
Caisse d'Épargne 12	89,85	95,26	97,30	9,54
Caisse d'Épargne 13	87,50	92,41	117,25	35,55
Caisse d'Épargne 14	89,23	102,73	111,49	6,14
Caisse d'Épargne 15	95,61	96,21	102,78	20,19

TABLE 5 – Évolution du ratio PNA/PNB par banque au fil du temps

Sur ce tableau, nous pouvons remarquer que le ratio PNA/PNB est légèrement inférieur à 1 pour la plupart des banques. Ainsi, c'est la rentabilité de la banque qui est majoritairement favorisée. Toutefois, certains ratios sont extrêmes, pour la CE 5 en 2018, l'assureur est lésé par rapport à la banque. En effet, le PNA ne représente qu'1 % du PNB pour cette Caisse d'Épargne.

Ensuite, nous nous sommes intéressés à l'évolution du ratio PNB/CA par banque au fil des années :

Banque	PNB/CA 2016 (en %)	PNB/CA 2017 (en %)	PNB/CA 2018 (en %)	PNB/CA 2019 (en %)
Banque Populaire 1	33,42	40,60	42,21	38,43
Banque Populaire 2	33,72	44,53	37,35	39,66
Banque Populaire 3	39,31	40,67	36,06	33,50
Banque Populaire 4	39,34	40,01	43,01	40,59
Banque Populaire 5	41,99	42,67	35,82	40,62
Banque Populaire 6	30,57	39,18	38,59	42,76
Banque Populaire 7	39,93	37,37	39,03	32,55
Banque Populaire 8	46,34	40,78	43,19	35,25
Banque Populaire 9	42,76	52,37	45,03	38,71
Banque Populaire 10	41,46	45,17	46,71	41,29
Banque Populaire 11	33,53	44,38	33,28	39,18
Banque Populaire 12	18,56	57,35	57,63	57,65
Caisse d'Épargne 1	36,80	33,96	29,10	32,43
Caisse d'Épargne 2	42,43	29,71	30,56	33,53
Caisse d'Épargne 3	42,40	31,69	15,83	20,48
Caisse d'Épargne 4	42,46	43,34	21,53	35,67
Caisse d'Épargne 5	42,57	23,27	5,28	30,01
Caisse d'Épargne 6	42,58	33,94	18,72	23,56
Caisse d'Épargne 7	42,55	43,18	23,04	37,04
Caisse d'Épargne 8	42,79	42,53	30,73	35,82
Caisse d'Épargne 9	42,64	38,63	28,57	23,39
Caisse d'Épargne 10	42,66	41,17	27,37	37,09
Caisse d'Épargne 11	42,62	41,36	25,51	42,79
Caisse d'Épargne 12	42,45	39,91	27,08	27,37
Caisse d'Épargne 13	42,42	44,71	16,27	43,02
Caisse d'Épargne 14	40,39	26,66	26,62	25,53
Caisse d'Épargne 15	42,45	38,98	24,09	32,80

TABLE 6 – Évolution du ratio PNB/CA par banque au fil du temps

Nous pouvons remarquer que le ratio est souvent compris entre 30 et 40%. Ainsi, le PNB représente environ un tiers du montant des primes hors taxes. Les banques qui rapportent beaucoup

de chiffre d'affaires à l'assureur ont donc un PNB plus élevé, en effet, le PNB évolue avec le CA. Dès lors, plus les banques vendent des contrats pour Natixis Assurances et plus elles sont récompensées par un PNB plus élevé. Par ailleurs, certains ratios ont augmenté au fil des années tandis que d'autres ont diminué.

3.5 Distribution du PNA, du PNB et du RBE par famille de produits

Après l'observation de plusieurs cartes, nous nous sommes intéressés à la répartition du PNA et du PNB par famille de produits.

Pour cela, nous avons commencé par regarder l'évolution du PNA en fonction de l'ancienneté des contrats. Les données présentées ci-dessous ont été comptabilisées sur les quatre années comptables étudiées. Ainsi, pour un contrat ouvert en 2014, le PNA rapporté en 2016, 2017, 2018 et 2019 sera comptabilisé avec les anciennetés 2, 3, 4 et 5 ans si le contrat est toujours en cours en 2019. Le PNA rapporté pour les contrats d'ancienneté égale à 0 comprend alors le PNA rapporté en 2016 par les contrats ouverts en 2016 et ainsi de suite jusqu'à 2019. Le tableau présenté ci-après est donc un tableau toutes générations de contrats confondues :

Ancienneté (année)	PNA (k€)	Ancienneté (année)	PNA (k€)	Ancienneté (année)	PNA (k€)
0	17 619	10	3 603	20	1 172
1	13 274	11	3 309	21	1 355
2	9 110	12	2 728	22	724
3	8 485	13	3 084	23	659
4	6 754	14	1 964	24	446
5	6 120	15	2 420	25	167
6	5 758	16	2 215	26	132
7	5 634	17	2 101	27	53
8	4 685	18	1 956	28	1
9	4 273	19	1 685		

TABLE 7 – Évolution du PNA en fonction de l'ancienneté des contrats

Sur ces tableaux, nous pouvons remarquer que les contrats très anciens rapportent très peu de PNA désormais, cela est dû au fait qu'ils ne sont pas très nombreux. Par ailleurs, étant donné que les Caisses d'Épargne commercialisent les contrats de Natixis Assurances uniquement depuis 2016, le PNA rapporté par ces contrats n'a que trois ans d'ancienneté au plus.

Nous nous sommes ensuite attardés sur l'évolution du PNA et du PNB par famille de produits au fil des quatre années comptables étudiées. Pour le PNA, nous avons obtenu le tableau suivant :

Famille de produits	PNA 2016 (k€)	PNA 2017 (k€)	PNA 2018 (k€)	PNA 2019 (k€)
FAMILLE FRUCTI-FAMILLE	10 607	12 754	11 854	12 690
FRUCTI-HOMME CLE	5 537	4 842	6 496	5 944
FRUCTI-PROFESSIONNEL	4 436	3 515	3 584	3 447
FRUCTI-PATRIMOINE	1 105	1 663	612	1 610
FAMILLE FRUCTI-BUDGET	892	257	695	501
DEDICACE	663	431	473	466
SECUR FAMILLE CE DC PTIA	333	2 356	3 005	6 682
FRUCTI-FACILITES AGRI	98	94	126	-74
FAMILLE DOUBLE SECURITE	92	34	23	7
CERTI CAPITAL	83	26	23	36
SECUR URGENCE CE	40	569	570	842
CREDIT MARITIME VIE	26	9	140	73
ASSURANCE HOMME-CLE C.E.	0	8	31	162
FAMILLE ASSURANCE FAMILLE	0	0	0	408
FRUCTI-FACILITES PRO	-3	-15	20	1
SECURI-CREDIT BPROP	-19	45	5	24
FAMILLE FRUCTI-FACILITES	-30	29	37	29
FAMILLE FRUCTI-SECURITE	-30	12	-9	-12
D.S.EPARGNE SALARIALE	-422	-23	-105	-274

TABLE 8 – Évolution du PNA au fil du temps

Sur ce tableau, nous avons ordonné les PNA du plus élevé au plus faible pour l'année 2016, cela permet de voir que l'importance des familles de produits a changé au fil du temps. En effet, comme nous pouvions nous en douter, les familles de produits des Caisses d'Épargne (Secur Famille CE DC PTIA, Secur Urgence CE et Assurance Homme-Clé CE) ont grossi au fil des années. Ainsi, Secur Famille CE DC PTIA est la famille du réseau CE qui rapporte le plus de PNA en 2019 et c'est aussi la deuxième famille au global qui rapporte le plus de PNA. Par ailleurs, comme nous l'avions vu sur le tableau 3, c'est également la deuxième famille à rapporter le plus de CA en 2019.

Après cette vision de l'évolution du PNA au fil des années, nous allons pouvoir comparer nos observations avec l'évolution du PNB. Nous avons construit le tableau suivant :

Famille de produits	PNB 2016 (k€)	PNB 2017 (k€)	PNB 2018 (k€)	PNB 2019 (k€)
FAMILLE FRUCTI-FAMILLE	12 405	16 944	16 187	16 653
FRUCTI-HOMME CLE	8 658	8 774	10 225	9 690
FRUCTI-PROFESSIONNEL	4 465	4 434	4 359	4 312
FRUCTI-PATRIMOINE	1 919	2 746	1 894	1 945
DEDICACE	775	676	639	590
FAMILLE FRUCTI-BUDGET	699	647	785	661
SECUR FAMILLE CE DC PTIA	375	2 669	3 504	7 670
FRUCTI-FACILITES AGRI	150	142	138	89
SECUR URGENCE CE	69	228	688	1 090
FAMILLE FRUCTI-FACILITES	47	57	58	54
CERTI CAPITAL	46	42	35	34
CREDIT MARITIME VIE	39	36	35	33
D.S.EPARGNE SALARIALE	39	50	19	27
FRUCTI-FACILITES PRO	29	31	22	19
FAMILLE DOUBLE SECURITE	19	24	19	17
SECURI-CREDIT BPROP	12	17	31	22
FAMILLE FRUCTI-SECURITE	3	2	2	2
ASSURANCE HOMME-CLE C.E.	1	15	34	82
FAMILLE ASSURANCE FAMILLE	0	0	0	407

TABLE 9 – Évolution du PNB au fil du temps

L'ordre d'importance des familles de produits est assez ressemblant entre le PNA et le PNB et c'est ce que l'on souhaite. En effet, dans l'idéal, il faut avoir un ratio PNA/PNB proche de 1 pour que le partage de la valeur soit équitable et, pour cela, il faut que les valeurs du PNA et du PNB soient proches donc que les familles de produits aient la même importance pour ces deux indicateurs de rentabilité.

Maintenant que nous avons observé l'évolution du PNA et du PNB au fil des années comptables étudiées, nous allons pouvoir regarder l'évolution du ratio PNA/PNB. Voici le tableau que nous avons obtenu :

Famille de produits	PNA/PNB 2016	PNA/PNB 2017	PNA/PNB 2018	PNA/PNB 2019	Poids du CA (2019)
FAMILLE FRUCTI-FAMILLE	86 %	75 %	73 %	76 %	41,0%
SECUR FAMILLE CE DC PTIA	89 %	88 %	86 %	87 %	19,9%
FRUCTI-HOMME CLE	64 %	55 %	64 %	61 %	18,4%
FRUCTI-PROFESSIONNEL	99 %	79 %	82 %	80 %	10,1%
FRUCTI-PATRIMOINE	58 %	61 %	32 %	83 %	3,8%
SECUR URGENCE CE	59 %	250 %	83 %	77 %	2,5%
FAMILLE ASSURANCE FAMILLE				100 %	0,9%
DEDICACE	86 %	64 %	74 %	79 %	0,9%
FAMILLE FRUCTI-BUDGET	128 %	40 %	89 %	76 %	0,9%
D.S.EPARGNE SALARIALE	-1092 %	-47 %	-559 %	-1014 %	0,6%
ASSURANCE HOMME-CLE C.E.	11 %	49 %	89 %	198 %	0,3%
FRUCTI-FACILITES AGRI	65 %	66 %	91 %	-83 %	0,3%
CREDIT MARITIME VIE	67 %	24 %	399 %	222 %	0,2%
FAMILLE FRUCTI-FACILITES	-64 %	50 %	64 %	54 %	0,1%
CERTI CAPITAL	181 %	61 %	65 %	105 %	0,1%
FRUCTI-FACILITES PRO	-9 %	-50 %	90 %	8 %	0,1%
SECURI-CREDIT BPROP	-149 %	256 %	18 %	106 %	0,1%
FAMILLE DOUBLE SECURITE	482 %	143 %	120 %	44 %	0,0%
FAMILLE FRUCTI-SECURITE	-1041 %	480 %	-432 %	-645 %	0,0%

TABLE 10 – Évolution du ratio PNA/PNB au fil du temps

Sur ce tableau, ce qui est préférable pour une famille de produits, c'est d'obtenir un ratio PNA/PNB le plus proche de 100% possible. Nous pouvons voir que le ratio est rarement très proche de cette valeur. En effet, sur les différentes années, il y a même quelques ratios extrêmes, par exemple, pour D.S. Épargne Salariale en 2016, le ratio est de -1 092%. Les ratios négatifs comme celui-ci sont une preuve du mauvais partage de la valeur car ils signifient que le PNA est négatif alors que le PNB est positif, l'assureur est donc lésé dans ce type de situations. Toutefois, pour les deux familles de produits les plus importantes (Famille Fructi-Famille et Secur Famille CE DC PTIA), le ratio est resté assez stable au fil des années et il s'agit d'un bon ratio qui favorise légèrement la banque car le PNA est environ égal à 80% du PNB pour ces deux familles.

À partir de ce tableau et du tableau du chiffre d'affaires par famille de produits (tableau 3), nous avons pu comparer le ratio PNA/PNB et le chiffre d'affaires de deux familles de produits en 2019 :

- Famille Assurance Famille :
 - Contrats commercialisés depuis 2019
 - Ratio PNA/PNB = 100% VS chiffre d'affaires = 1 071 k€ (moins d'1% du CA total)
- Fructi-Homme Clé :

— Ratio PNA/PNB = 61% VS chiffre d'affaires = 21 309 k€ (près de 20% du CA total)

Cela montre donc bien qu'un ratio qui favorise le PNB au détriment du PNA ne va pas pour autant signifier moins de chiffre d'affaires pour l'assureur. En effet, le ratio de Famille Assurance Famille est très équitable mais cette famille rapporte à peine 1% du CA total tandis que le ratio de Fructi-Homme Clé privilégie la rentabilité de la banque mais cette famille représente presque 20% du CA total.

Enfin, nous avons décidé d'observer l'évolution du RBE au fil du temps et nous avons obtenu le tableau suivant :

Famille de produits	RBE 2016 (k€)	RBE 2017 (k€)	RBE 2018 (k€)	RBE 2019 (k€)
FRUCTI-PROFESSIONNEL	1 658	1 349	1 605	2 823
FRUCTI-HOMME CLE	1 034	674	2 237	3 023
FAMILLE FRUCTI-FAMILLE	666	3 675	1 905	5 950
FAMILLE FRUCTI-BUDGET	583	-18	550	341
DEDICACE	325	163	291	303
FAMILLE DOUBLE SECURITE	82	25	18	1
CERTI CAPITAL	55	1	12	23
FRUCTI-PATRIMOINE	36	690	-550	984
FRUCTI-FACILITES AGRI	6	26	57	-119
FAMILLE ASSURANCE FAMILLE	0	0	0	284
FAMILLE FRUCTI-SECURITE	-31	11	-18	-13
FRUCTI-FACILITES PRO	-37	-44	-5	-14
SECURI-CREDIT BPROP	-49	19	-14	11
CREDIT MARITIME VIE	-50	-74	67	30
ASSURANCE HOMME-CLE C.E.	-51	-98	-495	112
FAMILLE FRUCTI-FACILITES	-81	-8	9	11
SECUR URGENCE CE	-219	128	202	243
D.S.EPARGNE SALARIALE	-613	-136	-294	-449
SECUR FAMILLE CE DC PTIA	-819	-249	442	1 869

TABLE 11 – Évolution du RBE au fil du temps

Nous pouvons voir que l'importance des familles de produits au sens du RBE n'est pas la même qu'au sens du PNA ou du PNB. Cependant, Famille Fructi-Famille reste la famille de produits la plus importante au sens du RBE en 2019. Comme nous l'avons vu plus tôt,

$$RBE = PNA - \text{Frais généraux totaux.}$$

Ce sont donc les frais généraux qui bouleversent l'ordre d'importance des familles de produits. Ainsi, la ventilation de ces frais est en train d'être revue en interne.

Dans cette partie, nous avons donc pu comprendre l'importance du partage de la valeur pour un bancassureur. Par ailleurs, nous avons vu l'utilité de la construction d'une base par tête afin d'agréger nos données aux mailles souhaitées et de pouvoir observer le PNA et le PNB de diverses manières. Dans la partie suivante, nous nous concentrerons sur les enjeux d'une nouvelle norme, IFRS 17, et sur sa comparaison à deux normes plus anciennes, French GAAP et Solvabilité 2.

4 La norme IFRS 17 et ses enjeux

Avant d'aborder la nouvelle norme IFRS 17 ainsi que ses enjeux, il nous a semblé important de donner les principaux éléments de la norme comptable French GAAP ainsi que de la norme prudentielle Solvabilité 2. En effet, l'objectif de cette partie étant une comparaison multinormes, il faut tout d'abord avoir connaissance du fonctionnement de ces deux normes.

Quand Solvabilité 2 est arrivée, nous n'avons pas fait assez de rétrospectif pour voir si ce qui avait été implémenté collait au réel ou non. Par la suite, la norme IFRS 17 a été publiée donc il a fallu implémenter de nouvelles méthodes. Nous sommes donc toujours dans le développement mais sans prendre le temps d'évaluer la justesse des résultats et nous ne pensons pas assez aux paramètres utilisés (par exemple, aux hypothèses de mortalité). Ainsi, comparer ces trois normes, comme nous l'avons fait dans ce mémoire, nous a permis de regarder les différences d'évaluation entre ces normes et de comparer des résultats projetés avec des résultats réels.

4.1 French GAAP

French GAAP, également appelée "Plan Comptable Général (PCG)" est une norme comptable édictée par l'Autorité des Normes Comptables (ANC). Elle définit les règles de comptabilité applicables aux entreprises domiciliées en France mais, également, la façon dont celles-ci doivent présenter leurs bilans et comptes de résultat, par exemple.

French GAAP rappelle les principes de la comptabilité, elle précise les règles de tenue des comptes et elle donne la nomenclature complète des comptes.

Sous French GAAP, il existe plusieurs classes de comptes permettant de bien répartir les différentes opérations :

- Classe 1 : Comptes de capitaux.
- Classe 2 : Comptes d'immobilisations.
- Classe 3 : Comptes de stocks et encours.
- Classe 4 : Comptes de tiers.
- Classe 5 : Comptes financiers.
- Classe 6 : Comptes de charges.
- Classe 7 : Comptes de produits.

Les cinq premières classes font partie du bilan tandis que les deux dernières font partie du compte de résultat.

4.2 Solvabilité 2

Solvabilité 2 est une norme prudentielle possédant plusieurs objectifs :

- Elle demande aux entreprises d'avoir des capitaux suffisants en vue de leurs propres risques.
- Elle encourage une gestion des risques in fine.
- Elle vise à harmoniser les normes ainsi que les pratiques prudentielles en Europe.
- Une évaluation économique et non historique.

La norme Solvabilité 2 est constituée de trois piliers qui définissent la quantification des risques, la gestion des risques et la communication que l'assureur doit apporter au sujet de ses risques en termes d'exigences, par exemple, à travers le Solvency Capital Requirement (SCR) et le Minimum Capital Requirement (MCR). Le SCR fait partie du pilier I de la norme et permet de capturer la volatilité de l'ensemble des risques découlant du portefeuille d'un assureur.

Sous Solvabilité 2, les calculs de SCR sont réalisés par projection des flux futurs du portefeuille. Dès lors, la tarification de l'assureur va agir sur le calcul des primes futures et sur le provisionnement tandis que les chocs de la formule standard de S2 permettent de se placer dans un scénario dont l'occurrence est d'une fois tous les 200 ans.

4.2.1 Le SCR global

Sous Solvabilité 2, il existe plusieurs manières de calculer le SCR d'un assureur :

- La formule standard : C'est l'approche qui est utilisée par la majorité des organismes assureurs, elle a été proposée par la Commission Européenne.
- Le modèle interne partiel : Si un assureur estime que la formule standard ne représente pas ses risques de la meilleure manière, il peut utiliser un modèle partiel, i.e., il utilise ses propres hypothèses pour certains risques, cependant, il garde la formule standard pour les autres risques.
- Le modèle interne : Ce modèle est propre à l'entreprise, néanmoins, sa mise en oeuvre est coûteuse et compliquée pour certains risques. Si un assureur veut utiliser son propre modèle, il a besoin de l'approbation de l'Autorité de Contrôle Prudentiel et de Résolution (ACPR).

Une fois qu'un assureur a choisi une approche pour calculer son SCR, il ne peut plus changer de méthode quand bon lui semble.

Le SCR global se décompose de la façon suivante (schéma provenant du document référencé [EIOPA-2014] en bibliographie) :

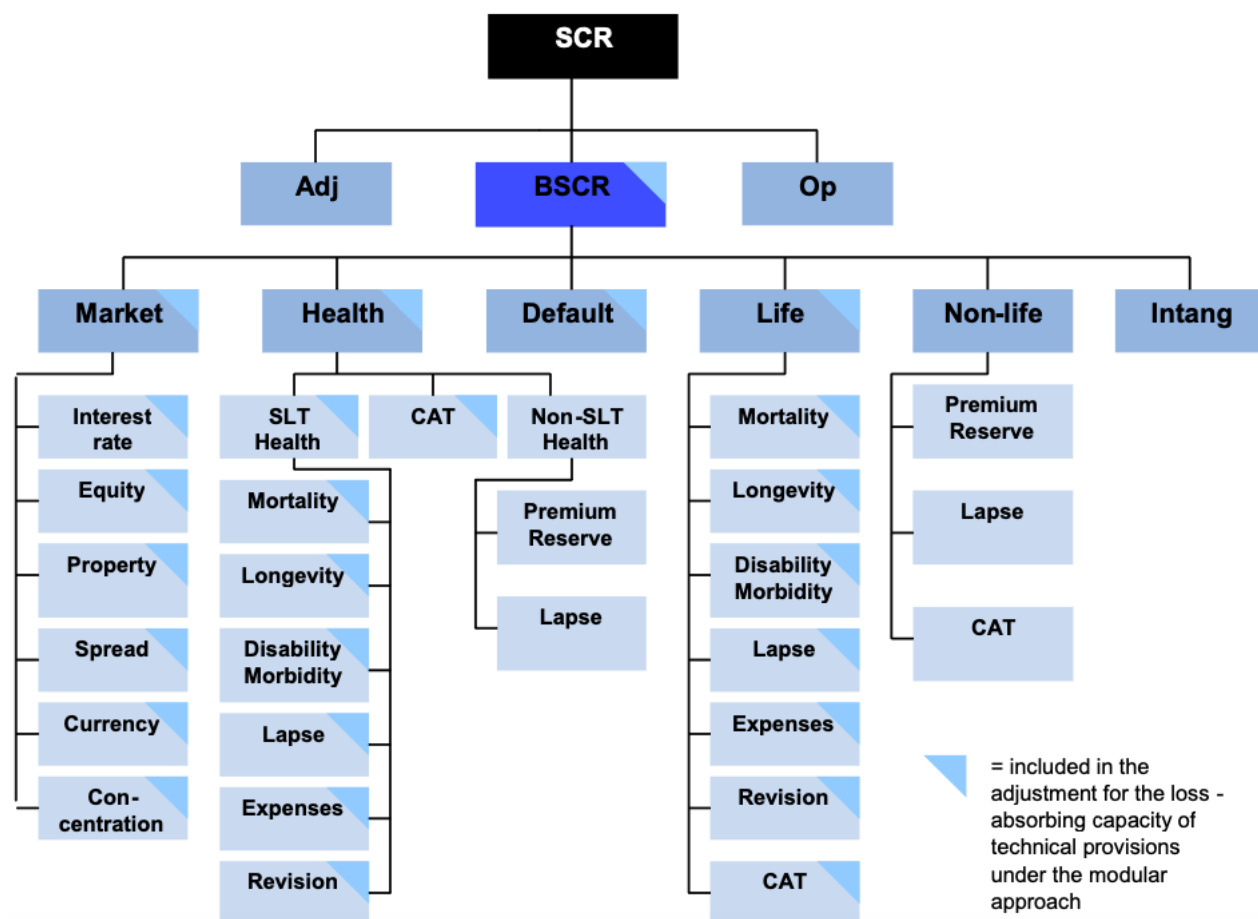


FIGURE 6 – Pieuvre S2

Ainsi, la formule du SCR global obtenue grâce à la formule standard peut s'écrire de la manière suivante : $SCR = BSCR + Adj + SCR_{Op}$, où :

- Basic Solvency Capital Requirement (BSCR) correspond au capital requis de base. Il s'obtient en calculant, puis, en agrégeant les SCR suivants à l'aide d'une matrice d'agrégation :
 - SCR_{market} correspond au capital requis pour le risque de marché.
 - SCR_{health} correspond au capital requis pour le risque de souscription santé.
 - $SCR_{default}$ correspond au capital requis pour le risque de contrepartie.

- SCR_{life} correspond au capital requis pour le risque de souscription vie.
- $SCR_{non-life}$ correspond au capital requis pour le risque de souscription non-vie.
- $SCR_{intangible}$ correspond au capital requis au titre des actifs incorporels.
- Adj correspond à l'ajustement lié à la capacité d'absorption des pertes des provisions techniques et des impôts différés.
- SCR_{Op} correspond au capital requis pour le risque opérationnel.

Le BSCR est défini de la façon suivante :

$$BSCR = \sqrt{\sum_{i,j} Corr(BSCR)_{i,j} \times SCR_i \times SCR_j} + SCR_{intangible},$$

où :

- $Corr(BSCR)_{i,j}$ représente le facteur de corrélation des risques i et j . La matrice de corrélation, quant à elle, est notée $Corr(BSCR)$.
- SCR_i et SCR_j représentent respectivement le SCR des risques i et j relativement à la ligne i et à la colonne j de la matrice de corrélation $Corr(BSCR)$.

Nous pouvons définir la matrice de corrélation $Corr(BSCR)$ de la façon suivante (schéma provenant du document référencé [EIOPA-2014] en bibliographie) :

i \ j	Market	Default	Life	Health	Non-life
Market	1				
Default	0.25	1			
Life	0.25	0.25	1		
Health	0.25	0.25	0.25	1	
Non-life	0.25	0.5	0	0	1

FIGURE 7 – Matrice de corrélation BSCR

4.2.2 Les trois piliers

Comme nous l'avons vu, Solvabilité 2 s'organise autour de trois piliers :

4.2.2.a Pilier I : Les exigences quantitatives

Sous Solvabilité 2, de nouvelles normes calculatoires ont été définies :

- La valeur de marché pour l'évaluation des actifs.
- Le Best Estimate pour le passif.

Le Best Estimate est la valeur actuelle attendue des flux de trésorerie futurs (notamment les primes, les prestations et les frais), où :

- Les flux sont estimés grâce à des hypothèses et des paramètres probabilistes réalistes à propos des différents facteurs de risque pouvant les affecter.
- Actuelle : Il s'agit de l'actualisation des flux futurs avec la courbe des taux sans risque fournie par l'European Insurance and Occupational Pensions Authority (EIOPA) à la date d'évaluation.
- Attendue : Il s'agit de l'espérance des flux de trésorerie.

Le BE correspond donc à une évaluation économique des engagements d'une compagnie d'assurance. Par ailleurs, nous noterons que le périmètre de calcul utilisé pour le calcul du BE inclut uniquement les contrats en stock, i.e., les contrats en cours à la date d'évaluation, les affaires nouvelles ne sont donc pas prises en compte dans le calcul du BE.

Il est possible d'observer les différences entre Solvabilité 1 et Solvabilité 2 à l'aide de leurs bilans prudentiels respectifs (schéma provenant du document référencé [GEORGES-2016] en bibliographie) :

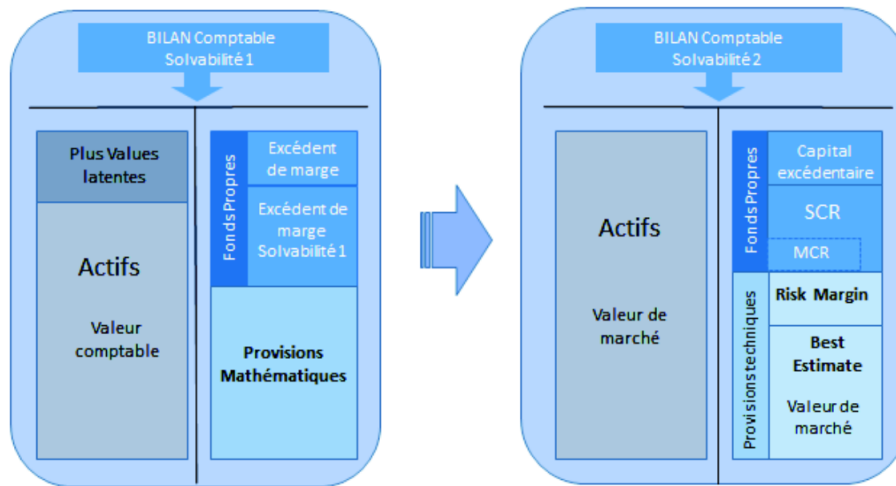


FIGURE 8 – Passage de Solvabilité 1 à Solvabilité 2

Au niveau de l'évaluation des actifs sous Solvabilité 2, nous pouvons remarquer qu'ils sont valorisés à leur valeur de marché à la date de clôture du bilan, il s'agit d'une approche "mark-to-market". Néanmoins, si la valeur de marché n'existe pas, ou bien, si elle est incohérente, les actifs sont valorisés selon des modèles de valorisation. Dans ce cas là, nous n'utilisons pas la méthode "mark-to-market" mais "mark-to-model" à la place, celle-ci valorise une position sur la base d'un modèle financier pour lequel les hypothèses ont été formulées par l'évaluateur. Ainsi, cette méthode est soumise au risque de modèle.

En ce qui concerne l'évaluation des passifs, sous S2, nous étendons aux provisions techniques, la notion de "cohérence avec le marché" de façon à ce que le bilan lui-même soit cohérent. A contrario, sous S1, nous déterminons les provisions techniques en intégrant une marge de prudence dans chacun des paramètres que nous utilisons lors du calcul des provisions. Sous S2, les provisions techniques sont définies de la manière suivante : *Provisions techniques S2 = Best Estimate de la valeur des engagements d'assurance + Risk Margin (RM)*.

Il existe deux niveaux de seuils pour les fonds propres, ceux-ci ont été définis par Solvabilité 2 pour caractériser l'exigence de capital réglementaire :

- Le SCR : Ce seuil représente le capital cible qu'il serait nécessaire d'avoir afin d'être en mesure de régler les frais engendrés par des événements imprévus. Le SCR est supposé couvrir, à un horizon d'un an, les contrats en cours de l'assureur. Les paramètres utilisés pour le calcul du SCR sont calibrés afin de représenter la Value at Risk (VaR) à un horizon d'un an avec un niveau de confiance de 99,5%. Ainsi, le SCR doit couvrir, à un horizon d'un an, la probabilité de ruine qui vaut 0,5%, soit une occurrence tous les 200 ans ($\frac{1}{0,005} = 200$).
- Le MCR : Ce seuil correspond au niveau minimum de fonds propres que l'assureur doit déterminer, en-dessous duquel, il ne serait plus en mesure de tenir ses engagements et l'ACPR serait dans l'obligation d'intervenir.

Par ailleurs, au niveau du passif du bilan Solvabilité 2, il existe une Risk Margin. Celle-ci est calculée et ajoutée au BE afin que le montant total des provisions soit en accord avec celui exigé par S2 de manière à ce que l'assureur honore ses engagements. La RM est calculée sur l'activité entière de l'assureur, elle tient ainsi compte des effets de la diversification des branches. Ce calcul est ensuite redistribué entre les différentes branches selon la contribution marginale de chacune au SCR. La RM est évaluée par l'actualisation du Cost of Capital (CoC) annuel généré par l'immobilisation du SCR, estimé à 6% par an sur la durée résiduelle des engagements utilisée pour le calcul

du BE. La RM peut s'écrire de la manière suivante :

$$RM = 6\% \sum_{t>0} \frac{SCR_t}{(1+r_t)^t},$$

où :

- SCR_t est le SCR à la date t .
- r_t est le taux sans risque à la date t .

4.2.2.b Pilier II : Les exigences qualitatives

Le deuxième pilier de Solvabilité 2 concerne les exigences qualitatives, il implique la création d'un dispositif interne de maîtrise des risques qu'il s'agisse de risques opérationnels, techniques ou financiers auxquels une compagnie d'assurance peut être confrontée. Dès lors, ce pilier permet de donner une vision précise de l'état de la solvabilité de l'assureur. Ainsi, sous le deuxième pilier de S2, il est nécessaire d'avoir une gouvernance saine, prudente et effective et il est nécessaire de s'assurer de la qualité des données, celles-ci doivent être exactes, exhaustives et pertinentes.

Par ailleurs, quatre fonctions clé ont été créées, il s'agit des suivantes : Actuariat, Audit interne, Conformité et Gestion des risques. La fonction clé "Actuariat" consiste, en partie, à étudier le ratio Sinistre/Primes dans le but de provisionner les sommes qui serviront à payer les sinistres. La fonction "Gestion des risques" définit l'appétit pour le risque des compagnies d'assurance. La "Conformité" veille à ce que les décisions de l'entreprise soient en accord avec la réglementation mais aussi avec les objectifs que l'entreprise se fixe en matière de bonnes pratiques notamment. Enfin, la fonction "Audit interne" a pour objectif de vérifier toutes les actions menées.

De plus, l'Own Risk and Solvency Assessment (ORSA) a été mis en place, c'est un dispositif interne de maîtrise des risques qui a pour but d'évaluer le besoin global de solvabilité. Il illustre la capacité de l'assureur à identifier, mesurer et gérer les éléments pouvant altérer sa solvabilité ou sa situation financière. Il sert d'outil de pilotage de l'activité en fonction des risques auxquels l'assureur fait face.

4.2.2.c Pilier III : Publication de l'information

Dans le troisième pilier de Solvabilité 2, le principe de discipline de marché a été introduit. Ainsi, l'organisme assureur est dans l'obligation de publier des informations détaillées à propos de son activité. Cela a permis d'améliorer la transparence du marché et a facilité la comparaison entre les différents assureurs européens.

Les informations mentionnées dans le pilier III de S2 sont transmises par l'assureur périodiquement grâce à deux types de documents :

- Les rapports dits narratifs : Il s'agit de documents qui décrivent la politique prudentielle mise en place par l'organisme assureur, il en existe deux différents :
 - Le Solvency Financial and Condition Report (SFCR) qui est destiné au public, il est remis annuellement.
 - Le Regular Supervisory Report (RSR) qui est destiné au superviseur de l'ACPR, il s'agit d'un document plus complet que le SFCR, il doit être remis au moins une fois tous les trois ans.
- Les reportings quantitatifs dit Quantitative Reporting Templates (QRT) : Il s'agit de documents qui prennent la forme d'un tableau de bord et dont le format a été décidé par l'EIOPA. Ils permettent de couvrir les éléments essentiels de l'activité d'une compagnie d'assurance tels que les fonds propres, les provisions techniques, le bilan, les placements, le SCR et le MCR. La plupart des QRT sont remis annuellement, cependant, environ 25% de ces rapports sont remis tous les trimestres.

Ces trois piliers peuvent être résumés de la manière suivante (schéma provenant du document référencé [KAMEGA-2014] en bibliographie) :

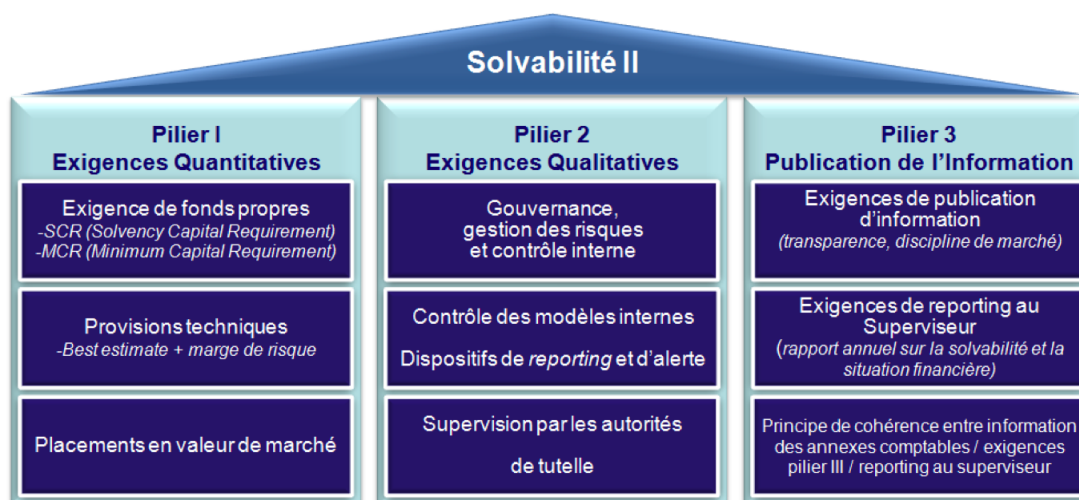


FIGURE 9 – Les trois piliers de Solvabilité 2

4.2.3 Calcul des Best Estimate de Solvabilité 2

Afin de pouvoir comparer la norme Solvabilité 2 à French GAAP et à IFRS 17, il a fallu calculer les Best Estimate de Solvabilité 2 pour l'année 2019. Pour cela, nous avons projeté les différents éléments nécessaires au calcul du BE avec le modèle de projection de la Direction des Risques de Natixis Assurances. Le modèle de projection de la DR permet de faire vivre le stock de nos contrats en cours à une certaine date t de manière prospective. Ainsi, il tient compte de beaucoup d'hypothèses modélisées permettant de projeter les flux futurs : la loi de mortalité par âge et par sexe, la loi de résiliation par âge, la loi d'incidence en arrêt de travail ainsi que la loi de maintien en arrêt de travail par âge et par ancienneté dans l'arrêt de travail. À partir de ces différentes lois Best Estimate réalisées par l'Actuariat Produits, des chaînes de Markov sont construites. Grâce à celles-ci, nous avons les différents états de passage et les probabilités associées à chaque état. Ainsi, les contrats sont projetés sous forme de chaînes de Markov, i.e., les primes et les prestations, par exemple, vont être projetées selon les différents états du contrat dans la chaîne. Le modèle considéré est donc un modèle multi-états et les assurés sont susceptibles de changer d'état selon le schéma suivant :

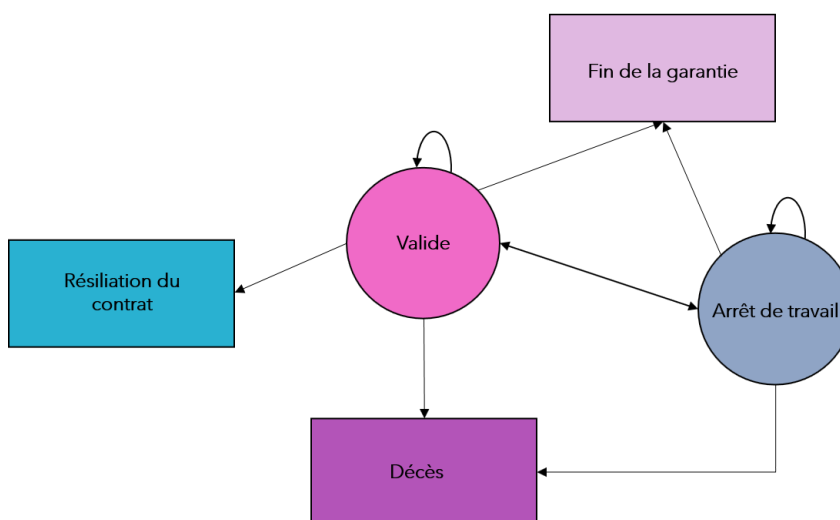


FIGURE 10 – Chaîne de Markov : Espace des états du modèle de projection

Le modèle de la DR est déterministe et il projette les flux mois par mois. Par ailleurs, comme nous l'avons indiqué, il n'utilise que les contrats en cours à la date de projection. Le calcul du Best Estimate est séparé en deux parties :

- Le BE de sinistres : Cette provision a pour but de couvrir les sinistres déjà survenus qu'ils soient connus ou inconnus. Elle contient essentiellement la Provision pour Sinistres A Payer (PSAP).
- Le BE de primes : Cette provision a pour objectif de couvrir le risque sur une période de garantie qui n'a pas encore commencé. Ainsi, elle vise à couvrir le risque d'insuffisance des primes non acquises à la date d'évaluation. En effet, si les primes sont insuffisantes, elles ne permettront pas d'absorber les prestations pour les sinistres futurs, c'est pour couvrir ce risque que le BE de primes a été créé.

Tout d'abord, il faut expliquer ce que sont les sinistres connus et les sinistres inconnus, voici un schéma donnant des exemples :

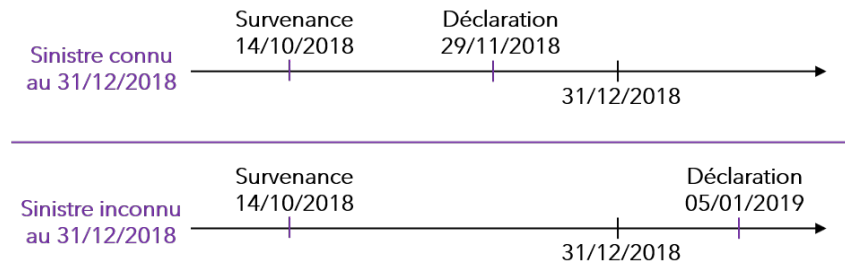


FIGURE 11 – Schéma sinistres connus et inconnus

Le premier schéma représente un sinistre connu au 31/12/2018, i.e., un sinistre qui est survenu et a été déclaré avant le 31/12/2018. Sur le deuxième schéma, un sinistre inconnu au 31/12/2018 est représenté, il s'agit d'un sinistre qui est survenu avant le 31/12/2018 mais qui a été déclaré après cette date.

Nous avons les formules suivantes pour les calculs de BE :

- « BE de primes brut » = - Primes TTC + Prestations futures + Commissions d'acquisition + Taxes + Frais sur primes + Frais prestations futures + Commission sur résultat de primes.
- « BE de sinistres brut » = Prestations connues + Prestations inconnues + Frais prestations connues + Frais prestations inconnues + Commission sur résultat de sinistres.
- « BE de primes cédé » = - Primes HT cédées + Prestations futures cédées + Commissions de cession + Participation aux bénéfices de réassurance de primes.
- « BE de sinistres cédé » = Prestations connues cédées + Prestations inconnues cédées + Participation aux bénéfices de réassurance de sinistres.

Nous noterons, par ailleurs, que la participation aux bénéfices de réassurance correspond à la commission sur résultat cédée. Avec ces formules-ci, nous avons pu calculer les BE totaux :

- « BE total brut » = « BE de primes brut » + « BE de sinistres brut ».
- « BE total cédé » = « BE de primes cédé » + « BE de sinistres cédé ».
- « BE total » = « BE total brut » - « BE total cédé ».

La dernière formule s'applique également au BE de sinistres seul de la manière suivante : « *BE de sinistres* » = « *BE de sinistres brut* » - « *BE de sinistres cédé* ». Le BE de sinistres seul sera, en effet, utile au calcul du résultat IFRS 17 dont nous parlerons dans la sous-partie 4.5.1.

4.2.4 Transformation du Best Estimate en Produit Net Assurance

Les modèles de Machine Learning que nous avons créés sont des modèles permettant de prédire le Produit Net Assurance à un an. Ces modèles nous ont permis de comparer le Produit Net Assurance prédit à un an et le PNA réel mais pas uniquement. En effet, l'objectif de ces modèles était de comparer les résultats obtenus par Machine Learning avec ceux obtenus par les projections S2. Nous avons donc désiré comparer le Best Estimate total de S2 avec le PNA prédit par Machine Learning sur la base Solvency qui sert de base de calcul au BE. Cependant, le PNA et le BE total S2 ne sont pas tout à fait égaux et les modèles de ML réalisés permettent de prédire un PNA et non

un BE. Toutefois, il est possible de passer d'un BE à un PNA en procédant à quelques ajustements.

Ainsi, avant de pouvoir comparer le BE total S2 et le PNA, il a fallu transformer le BE. En effet, le BE total S2 est calculé avec des provisions de sinistres de fin d'année car il fait intervenir les prestations pour sinistres connus et pour sinistres inconnus alors que le PNA est calculé avec une variation de la provision de sinistres, c'est-à-dire que nous avons calculé la différence entre la provision de sinistres de fin d'année et la provision de sinistres de début d'année. Il a donc fallu retirer la provision de sinistres de début d'année du PNA afin de compenser la différence de provisions :

$$- (Provision\ de\ sinistres\ de\ fin\ d'année - Provision\ de\ sinistres\ de\ début\ d'année) - Provision\ de\ sinistres\ de\ début\ d'année = - Provision\ de\ sinistres\ de\ fin\ d'année.$$

En effet, dans le calcul du PNA, nous soustrayons la charge de sinistres comptable donc nous retirons la variation de provisions de sinistres, ainsi, nous soustrayons la provision de sinistres de fin d'année et nous ajoutons la provision de sinistres de début d'année. Par ailleurs, sous S2, les frais généraux sont comptabilisés dans le BE, ce n'est pas le cas dans le PNA, il a donc fallu imputer les frais généraux au PNA modifié pour arriver au BE de S2. A noter que, comme nous l'avons vu plus tôt,

$$PNA - Frais\ généraux\ de\ primes - Frais\ généraux\ de\ prestations = RBE.$$

Nous en sommes donc arrivés à la formule suivante :

$$Best\ Estimate\ total = PNA - Provision\ de\ sinistres\ de\ début\ d'année - Frais\ généraux\ de\ primes - Frais\ généraux\ de\ prestations.$$

Cependant, les BE qui sortent du modèle de la DR sont des BE calculés en retirant les primes et en ajoutant les sinistres, contrairement à ce qui a été fait dans le calcul du PNA, la formule exacte est donc :

$$- Best\ Estimate\ total = PNA - Provision\ de\ sinistres\ de\ début\ d'année - Frais\ généraux\ de\ primes - Frais\ généraux\ de\ prestations.$$

Nous pouvons donc en déduire la formule suivante :

$$PNA = - Best\ Estimate\ total + Provision\ de\ sinistres\ de\ début\ d'année + Frais\ généraux\ de\ primes + Frais\ généraux\ de\ prestations.$$

Cette formule nous a donc permis de "transformer" un BE en PNA afin de faciliter les comparaisons.

Par ailleurs, en reprenant la formule de la transformation du BE en PNA, nous pouvons remarquer qu'il est également possible de transformer le BE en Résultat Brut d'Exploitation car le RBE est égal au PNA moins les frais généraux totaux. Ainsi,

$$RBE = - Best\ Estimate\ total + Provision\ de\ sinistres\ de\ début\ d'année.$$

4.3 IFRS 17

Avec la nouvelle norme IFRS 17, la profitabilité des contrats a été redéfinie. Le compte de résultat IFRS 17 intègre les Best Estimate de sinistres évalués sous la norme Solvabilité 2. Cependant, les primes, commissions sur résultat et règlements de sinistres, par exemple, sont évalués sous la norme French GAAP. Ainsi, les deux normes s'entremêlent et IFRS 17 est dépendante des normes French GAAP et Solvabilité 2.

Toujours en conservant la vision produits importante pour classer les familles, nous avons comparé la rentabilité des trois normes par le biais du Résultat Brut d'Exploitation. Pour la norme Solvabilité 2, nous avons considéré les flux projetés au 31/12/N-1 pour la rentabilité au 31/12/N. Les flux générés par les affaires nouvelles de l'année N n'étant pas comptabilisés, nous avons comparé les flux des têtes présentes à la fois au 31/12/N-1 et au 31/12/N.

4.3.1 Présentation de la norme

De nos jours, il est nécessaire d'avoir un référentiel comptable commun à un maximum de pays afin de favoriser des échanges internationaux et de permettre aux investisseurs des groupes multinationaux d'avoir une vision harmonisée des éléments comptables et financiers.

Dès lors, l'International Accounting Standards Board (IASB) a été créé en 2001. Il est venu remplacer l'International Accounting Standards Committee (IASC) créé en 1973. C'est l'organisme international qui est chargé d'élaborer les normes comptables International Accounting Standards (IAS) et IFRS. L'IASB a de nombreux objectifs :

- La réconciliation des normes comptables propres à chaque pays vers une même norme internationale dans le but de fournir aux investisseurs des résultats transparents et universels qui répondent aux standards internationaux.
- La promotion des nouvelles normes au niveau mondial et l'encouragement à les utiliser.
- La publication de documents permettant d'aider les entreprises à bien comprendre les normes et à les appliquer.

La lecture des comptes ne nous permet pas toujours d'avoir les informations que l'on désire, en particulier, au niveau de la rentabilité des opérations, c'est pour cela que la nouvelle norme IFRS 17 va être très utile.

IFRS 17 a été publiée le 18 mai 2017 par l'IASB et devrait entrer en vigueur le 1er janvier 2023. IFRS 17 répond à des enjeux d'harmonisation des pratiques, de comparabilité et de transparence des états financiers de 450 groupes assureurs et réassureurs cotés (ou émettant de la dette cotée). IFRS 17 sera appliquée aux contrats d'assurance et traités de réassurance que les sociétés d'assurance émettent, aux traités de réassurance qu'elles détiennent ainsi qu'aux contrats d'investissement émis avec un élément de participation discrétionnaire (cf. définition 4). IFRS 17 est la première norme comptable internationale complète pour les contrats d'assurance.

Voici le périmètre sur lequel la norme est appliquée :

	Comptes sociaux	Comptes consolidés
Société cotée	Norme française obligatoire et norme IFRS exclue	Norme IFRS obligatoire
Société non cotée		Norme française ou IFRS

FIGURE 12 – Périmètre IFRS 17

Cette norme spécifique aux sociétés d'assurance est entrée en jeu à cause du fonctionnement particulier de celles-ci. En effet, on dit que leur cycle de production est inversé : les primes (prix de vente) sont connues et encaissées avant le paiement des prestations (prix de revient).

IFRS 17 a été créée dans le cadre de l'homogénéisation internationale des normes comptables afin de pouvoir comparer les sociétés cotées. Elle concerne la manière dont les contrats d'assurance seront classés et comptabilisés dans les comptes consolidés. Elle va permettre d'harmoniser l'estimation des passifs d'assurance et de favoriser la cohérence avec les autres normes IFRS.

Les premières discussions sur cette norme de comptabilisation des contrats d'assurance ont débuté en 1997. Ainsi, pendant 20 ans d'élaboration, plusieurs projets de normes ont été publiés en fonction des différentes parties prenantes (compagnies d'assurance et régulateurs nationaux, par exemple). Au fur et à mesure de ces années, les normes ont commencé à évaluer les contrats en valeur de marché et ont défini la reconnaissance des résultats avec la Contractual Service Margin (CSM, cf. définition 5).

IFRS 17 a été introduite pour remplacer IFRS 4. Cette norme, introduite en 2004, permettait aux assureurs de comptabiliser les contrats d'assurance sur la base de normes comptables nationales en utilisant différentes approches. Cependant, le fait d'utiliser différentes approches compliquait la comparaison des performances financières des contrats d'assurance. En effet, au sein d'une même

zone géographique, les compagnies d'assurance vie, non-vie et de réassurance n'ont pas la même vision de leurs engagements ce qui rend difficile une éventuelle comparaison entre ces compagnies. De plus, IFRS 4 n'imposait aucun modèle de valorisation des passifs alors qu'IFRS 17 en propose trois que nous allons détailler dans la sous-partie 4.3.5 : le modèle Building Block Approach (BBA), le modèle Premium Allocation Approach (PAA) et le modèle Variable Fee Approach (VFA) :

- Le modèle BBA est le modèle général. Il est basé sur l'évaluation de trois blocs pour valoriser le passif d'assurance : le BE, le Risk Adjustment (RA, cf. définition 6) et la CSM.
- Le modèle PAA est une simplification du modèle BBA pouvant être appliquée uniquement si le contrat d'assurance a une durée de couverture de moins d'un an ou que l'approximation de ce modèle est raisonnable vis-à-vis du modèle BBA.
- Le modèle VFA a pour objectif d'adapter le modèle BBA aux contrats d'assurance dits "directement participatifs" (ces contrats seront définis dans la sous-partie 4.3.5).

Sous IFRS 17, la notion de contrat d'assurance est redéfinie et il est obligatoire de comptabiliser les contrats de façon homogène, cela réduit ainsi les différences d'évaluation et de comptabilisation des contrats d'assurance entre pays. Cette norme vise à corriger l'inadéquation entre l'évaluation de l'actif et du passif qu'il y avait sous IFRS 4 car les actifs étaient valorisés selon le principe de juste valeur mais les passifs pouvaient encore être évalués selon le coût historique. Pour corriger cela, elle propose une évaluation économique des provisions techniques ce qui assure une meilleure cohérence avec l'évaluation de l'actif.

IFRS 17 va nous permettre d'utiliser des méthodes de valorisation du passif d'assurance qui ressemblent aux méthodes utilisées pour le reporting prudentiel de Solvabilité 2. Cependant, sous IFRS 17, la granularité est beaucoup plus fine que sous S2. De plus, les délais de production des états financiers sont bien plus courts que ceux qu'il y a pour le reporting S2.

En dehors du cadre d'uniformisation du processus comptable, IFRS 17 vise à imposer une nouvelle vision concernant la valorisation des passifs d'une compagnie d'assurance. Dans l'ensemble, les principes comptables locaux conduisent à une valorisation au coût historique alors qu'IFRS 17 tend à valoriser les engagements d'assurance à leur juste valeur.

Cette nouvelle norme a pour objectif de fournir une information pertinente pour les utilisateurs des états financiers. Elle vise à évaluer l'impact des contrats d'assurance sur la situation, la performance financière et les flux de trésorerie des entités. En outre, IFRS 17 va permettre de valoriser les options et les garanties des contrats d'assurance.

L'actif du bilan de l'assureur est, quant à lui, soumis à IFRS 9. En effet, IFRS 9 est une norme relative à l'évaluation et à la comptabilisation des instruments financiers. Elle s'applique aux actifs de placement (par exemple, les actions et les obligations) mais elle ne s'applique pas aux biens immobiliers de placement.

Par ailleurs, les pays peuvent, s'ils le souhaitent, utiliser les normes IFRS comme référentiels locaux pour la présentation de leurs états individuels. IFRS 17 dispose d'une particularité selon laquelle, il faut séparer les composants de chaque contrat d'assurance. En effet, il faut isoler les dérivés incorporés (cf. définition 7), les composantes d'investissement (cf. définition 8) et les fournitures de biens et services (cf. définition 9) car seuls les autres composants sont soumis à IFRS 17. Il faut donc analyser chaque contrat d'assurance afin de classer ses composants s'ils sont séparables et de déterminer les valeurs des paramètres indispensables à leur comptabilisation. Dans la suite de ce mémoire, la notion de "dérivés incorporés" se référera à des dérivés (cf. définition 10) qui n'ont pas été séparés du contrat d'assurance hôte et la notion de "composantes d'investissement" fera référence à des composantes d'investissement qui n'ont pas été séparées du contrat d'assurance hôte.

Il faut appliquer IFRS 9 afin de déterminer si les dérivés incorporés doivent être séparés du contrat d'assurance hôte et, si c'est le cas, ils seront comptabilisés sous IFRS 9. Les composantes d'investissement doivent être séparées du contrat d'assurance hôte si elles sont distinctes de celui-ci. C'est le cas, si et seulement si, les conditions suivantes sont vérifiées :

- La composante d'investissement et la composante d'assurance ne sont pas étroitement liées. C'est-à-dire que l'entité est capable de mesurer une composante sans considérer l'autre et l'assuré est en mesure de bénéficier d'une composante même si l'autre n'est pas présente.

- Un contrat aux conditions équivalentes est vendu ou pourrait être vendu séparément de la composante d'investissement sur le même marché ou la même juridiction. L'entité n'est pas tenue d'effectuer une recherche exhaustive afin de déterminer si la composante d'investissement est vendue séparément ou non.

Comme pour les dérivés incorporés, les composantes d'investissement devant être séparées du contrat d'assurance hôte seront comptabilisées sous IFRS 9.

Voici la représentation du bilan avec l'utilisation d'IFRS 9 à l'actif et d'IFRS 17 au passif :

Actif IFRS 9	Passif IFRS 17
Actifs en juste valeur ou en coût amorti	Fonds propres
	CSM
	RA
	BE

FIGURE 13 – Bilan IFRS 9/IFRS 17

4.3.1.a Transition

L'entrée en vigueur d'IFRS 17 nécessite une année de transition pendant laquelle, les assureurs devront publier leurs comptes IFRS 17 tout en assurant une année comparative ; cette année de transition marquera le passage d'un bilan IFRS 4 à un bilan IFRS 17.

La période de transition est l'année qui précède la date d'implémentation d'IFRS 17 et la date de transition correspond au début de la période de transition, elle est donc située un an avant la date d'implémentation. A l'origine, la date de transition était le 1er janvier 2020 car la norme devait entrer en vigueur le 1er janvier 2021, cependant, l'entrée en vigueur de la norme a été reportée au 1er janvier 2023 donc la date de transition est fixée au 1er janvier 2022. C'est à cette date que la norme sera appliquée pour la première fois pour la préparation du bilan d'ouverture 2022.

Établir un bilan IFRS 17 à la date de transition va revenir à évaluer un BE, un RA et une CSM à cette date. Le BE et le RA seront évalués selon une approche prospective ce qui ne sera pas le cas de la CSM. En effet, elle pourra être évaluée selon trois approches que nous allons détailler plus tard. Nous utiliserons ici le terme de BE en raison de sa ressemblance avec la notion de BE sous S2 mais le terme réellement utilisé dans le texte d'IFRS 17 est : "estimates of future cash flows".

Dès lors, les sociétés d'assurance vont devoir modifier leurs activités et affronter les répercussions sur leur stratégie de communication financière mais aussi sur leurs processus comptables, actuariels et organisationnels. Par ailleurs, mener une transition au plus proche de la norme va être très coûteux en termes de temps. Ainsi, réaliser son premier bilan de transition pour l'entrée en vigueur de la norme semble risqué parce que les résultats qui seront obtenus nécessiteront du recul et une analyse approfondie.

4.3.1.b Les informations à fournir

La communication a pour but de permettre aux utilisateurs des états financiers d'évaluer l'effet des contrats relevant du champ d'application d'IFRS 17 sur différents aspects : la situation financière, la performance financière et les flux de trésorerie de l'entité. Pour cela, l'entreprise doit fournir des informations qualitatives et quantitatives sur (paragraphe 93 d'IFRS 17 Insurance Contracts) :

- Les montants comptabilisés dans ses états financiers pour les contrats relevant du champ d'application d'IFRS 17.

- Les jugements d'expert importants et les changements intervenus dans ces jugements lors de l'application d'IFRS 17.
- La nature et l'ampleur des risques liés aux contrats relevant du champ d'application d'IFRS 17.

Cette norme favorise la pertinence de la communication avec les utilisateurs plutôt que la conformité à des exigences de communication détaillées et prescriptives. Toutefois, si ces publications ne sont pas assez détaillées pour assurer les objectifs de communication de la norme, il est possible d'exiger des compléments d'informations.

Sous IFRS 17, les entreprises doivent publier les réconciliations entre le solde d'ouverture et le solde de clôture pour certains éléments :

- Les provisions au titre de la couverture restante (hors éléments de pertes du passif).
- Les éléments de pertes du passif.
- Les provisions pour les sinistres survenus.
- La valeur actuelle des flux de trésorerie futurs.
- Le RA.
- La CSM.

Pour chacun de ces éléments, une communication séparée doit être effectuée si possible pour ce qui suit :

- Les sinistres survenus.
- L'amortissement des frais d'acquisition.
- La variation des Fulfilment Cash Flows (FCF, cf. définition 11).
- Les changements relatifs aux services futurs.
- Les composantes d'investissement (hors revenus et dépenses liés au service d'assurance).
- Le revenu d'assurance (cf. définition 12).

IFRS 17 exige une présentation séparée des montants relatifs aux contrats d'assurance émis et aux contrats de réassurance détenus dans les états principaux.

4.3.1.c Les avantages d'IFRS 17

IFRS 17 propose un cadre complet et uniformisé pour la comptabilité d'assurance et réduit le recours aux normes locales pour les calculs. Ceci augmente la comparabilité des contrats d'assurance :

- Comptabilisation homogène pour tous les contrats d'assurance.
- Comparaison simplifiée des produits d'assurance selon les zones géographiques de part l'harmonisation des mesures pour les contrats d'un même groupe.

Cependant, IFRS 17 impose de nouvelles contraintes, en particulier, dans la définition de la maille de calcul. En effet, IFRS 17 impose aux compagnies d'assurance de classer l'ensemble de leurs engagements d'assurance par groupes de contrats partageant les mêmes caractéristiques puis de scinder ces groupes en fonction de leur rentabilité attendue, i.e., si ces contrats vont être profitables ou onéreux. Il s'agit de comptabiliser une CSM représentant les revenus attendus et non encore réalisés sur toute la durée de couverture d'un groupe de contrats. Ce montant doit être ajusté en fonction de la sinistralité observée au cours de la durée de vie du contrat. La CSM est toujours positive ou nulle, elle ne peut pas être négative. Ainsi, pour les contrats onéreux, il n'y a pas de CSM mais une Loss Component (LC, cf. définition 13) est constituée.

4.3.1.d Cadre réglementaire européen

L'European Financial Reporting Advisory Group (EFRAG) est un organisme privé qui a été créé en 2001 par les acteurs européens (des professions comptables, les utilisateurs et les normalisateurs nationaux) avec les encouragements de la Commission Européenne. Il a pour but de promouvoir les normes au niveau européen, de conseiller la Commission Européenne, de servir les intérêts européens auprès de l'IASB et de donner un avis sur les normes publiées par cet organisme.

Le règlement européen n°1606/2002 entré en vigueur le 1er janvier 2005 impose aux sociétés régies par le droit national d'un État membre de préparer leurs comptes consolidés conformément aux normes comptables internationales adoptées, si, à la date de clôture de leur bilan, leurs titres sont admis à la négociation sur le marché réglementé d'un État membre au sens de l'article 1er, point 13, de la directive 93/22/CEE du Conseil du 10 mai 1993² concernant les services d'investissement dans le domaine des valeurs mobilières.

4.3.1.e Grand principe de la norme IFRS 17

Plusieurs phases constituent le grand principe de la norme IFRS 17 :

- La phase de classification permet de définir la nature des contrats. Ainsi, les entreprises d'assurance vont pouvoir regrouper leurs contrats par portefeuille de risques similaires. Cette phase de classification est très importante car les contrats seront évalués différemment s'ils sont considérés comme des contrats d'assurance ou non et s'ils entrent dans le champ d'application d'IFRS 17 ou non.
- La première évaluation permet de déterminer les montants que l'entreprise doit comptabiliser quant à ses contrats détenus ou émis. Pour réaliser cette évaluation, la compagnie d'assurance va devoir estimer la valeur de ses différents portefeuilles sur la base des FCF et de la CSM.
- La phase d'évaluation ultérieure a lieu à chaque date de clôture pour ajuster les montants comptabilisés au départ. Cette phase prend en compte la valeur des contrats sur le reste de la période de couverture (cf. définition 14) et la dette sur les sinistres survenus.
- La dernière phase concerne la présentation des états financiers. Pour cela, il faut enregistrer les variations d'estimation sur les flux de trésorerie à chaque date de clôture en résultat ou en ajustement de la CSM selon la nature des réévaluations. Enfin, ces changements seront présentés au compte de résultat ou en Other Comprehensive Income (OCI) dans les capitaux propres.

Voici les principes clés d'IFRS 17 :

- L'assureur doit identifier les contrats d'assurance comme ceux en vertu desquels l'entité accepte un risque d'assurance significatif d'une autre partie (l'assuré) en s'engageant à indemniser l'assuré si un événement futur incertain (l'événement assuré) causait un dommage à l'assuré.
- Il doit séparer les dérivés incorporés spécifiés, les composantes d'investissement distinctes et les biens ou services distincts, i.e., non liés à l'activité d'assurance, des contrats d'assurance.
- Il doit diviser les contrats en groupes qu'il comptabilisera et évaluera.
- Il doit reconnaître et mesurer les groupes de contrats d'assurance à une valeur actualisée des flux de trésorerie futurs ajustée au risque (les FCF) plus un montant représentant le bénéfice non encore gagné dans le groupe de contrats (la CSM).
- Il doit reconnaître le bénéfice d'un groupe de contrats d'assurance sur la période pendant laquelle il fournit une couverture d'assurance et lorsqu'il est libéré du risque. S'il s'attend à ce qu'un contrat soit onéreux sur la période de couverture restante, il doit comptabiliser la perte immédiatement.
- Il doit fournir des informations afin de permettre aux utilisateurs des états financiers d'évaluer l'effet que les contrats entrant dans le champ d'application d'IFRS 17 ont sur la situation financière, la performance financière et les flux de trésorerie de l'entité.

2. Il est possible de trouver cet article au lien suivant : <https://www.senat.fr/rap/a03-202/a03-20231.html>.

4.3.2 La granularité

Il est important de définir le niveau de granularité à utiliser pour le calcul des provisions car la rentabilité a un impact direct sur le mode de reconnaissance des résultats futurs (les pertes sont reconnues immédiatement alors que les profits sont différés). Nous allons, ici, présenter trois niveaux de définition de la granularité :

- Le portefeuille de contrats d'assurance (cf. définition 15).
- L'année de souscription (génération, cohorte) : Les contrats appartenant à un portefeuille d'assurance peuvent être segmentés en groupes d'amplitude d'une année de souscription au maximum.
- La qualification de la profitabilité : Les contrats peuvent être identifiés comme onéreux, potentiellement onéreux ou profitables.

Les actuaires vont donc passer d'une maille S2 (Line of Business (LoB) × HRG) à une maille IFRS 17 (portefeuilles × génération × profitabilité). Cela va complexifier les modèles, les process et la gestion des données.

Les contrats d'assurance sont agrégés par groupe. Une fois que l'agrégation a été faite, elle ne peut pas être modifiée. Sous IFRS 17, les calculs sont effectués par groupe de contrats. Par ailleurs, les critères de segmentation des groupes doivent respecter la réglementation.

Pour créer un groupe de contrats qui vont être gérés ensemble, nous sélectionnons des contrats exposés à des risques similaires et souscrits la même année. En effet, il est obligatoire que la souscription de tous les contrats d'un même groupe ait eu lieu la même année afin de ne pas agréger des générations profitables avec des générations onéreuses. Cependant, si nous avons des contraintes légales ou réglementaires, il est possible d'intégrer des contrats à profitabilité différente dans un même groupe de contrats.

Après cette agrégation de contrats, nous segmentons les groupes de contrats selon la profitabilité des contrats lors de la première comptabilisation. Cette segmentation donne lieu à trois sous-groupes de contrats :

- Les contrats onéreux à l'origine : L'organisme d'assurance s'attend à des pertes financières sur ces contrats, $CSM = 0$. Par exemple, certains contrats automobile jeunes conducteurs peuvent être onéreux.
- Les contrats potentiellement onéreux : Il s'agit de contrats non-onéreux à l'origine mais qui ont un risque significatif de devenir onéreux, $CSM > 0$.
- Les contrats profitables : Ils ne présentent pas de risque significatif de devenir onéreux, $CSM \gg 0$.

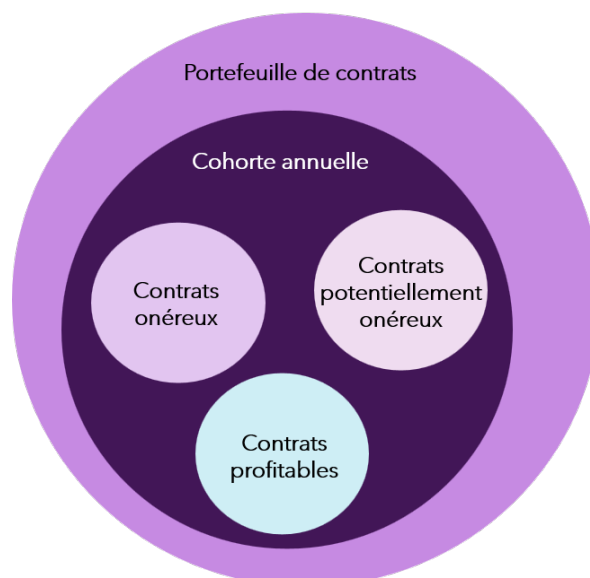


FIGURE 14 – Une nouvelle granularité

Un assureur n'a pas besoin de déterminer au cas par cas le groupe de profitabilité auquel appartiennent les contrats. Si l'entité dispose d'informations raisonnables et justifiables pour conclure qu'un ensemble de contrats appartiendra au même groupe, elle peut les mesurer comme un ensemble afin de déterminer à quel groupe ils appartiennent. Si l'entité ne dispose pas de telles informations, elle doit prendre la décision en évaluant les contrats individuellement.

Un contrat d'assurance est onéreux à la date de la comptabilisation initiale si le total des FCF alloués au contrat (y compris les flux de trésorerie d'acquisition précédemment comptabilisés et les flux de trésorerie résultant du contrat à la date de la comptabilisation initiale) constitue une sortie nette, c'est-à-dire qu'il y a plus d'argent qui sort de l'entité que d'argent qui rentre.

Pour les contrats qui ne sont pas onéreux initialement, l'entité devra distinguer les contrats profitables sans risque significatif de devenir onéreux par la suite et les autres contrats. L'évaluation du fait qu'un contrat ou qu'un ensemble de contrats ne présente pas de risque significatif de devenir onéreux ultérieurement devrait :

- Être basée sur la probabilité de changements d'hypothèses qui rendraient le contrat ou l'ensemble de contrats onéreux s'ils avaient lieu.
- Utiliser des informations sur les estimations fournies par le reporting interne de l'entité.

Un groupe de contrats d'assurance auquel une entité applique le modèle BBA devient onéreux (ou encore plus onéreux) lors d'une évaluation ultérieure si des variations défavorables des FCF alloués au groupe de contrats pour les services futurs dépassent la valeur comptable de la CSM (c'est-à-dire que la CSM va devenir négative, ou plutôt, elle va devenir nulle et une Loss Component va entrer en jeu, alors qu'elle était nulle dans le cas d'un contrat onéreux ou positive dans le cas d'un contrat profitable).

Le niveau d'agrégation des portefeuilles doit se baser de façon à ce que la population du portefeuille soit homogène. En ce qui concerne l'assurance vie, il faut évaluer le caractère onéreux en prenant en compte la mutualisation des produits financiers entre contrats. Ces regroupements de contrats sont très contraignants et supposent que l'on a déjà une connaissance approfondie des contrats commercialisés et que l'on a déjà réalisé des calculs de CSM.

A la date de première valorisation, on évalue la CSM une première fois et cette première estimation va servir tout au long de la vie du contrat d'assurance. Trois cas sont possibles à la date initiale de valorisation : deux cas de contrats déficitaires (ils présentent des pertes dès la première date de valorisation) et un cas de contrats profitables (ils présentent des bénéfices futurs attendus sous forme de CSM).

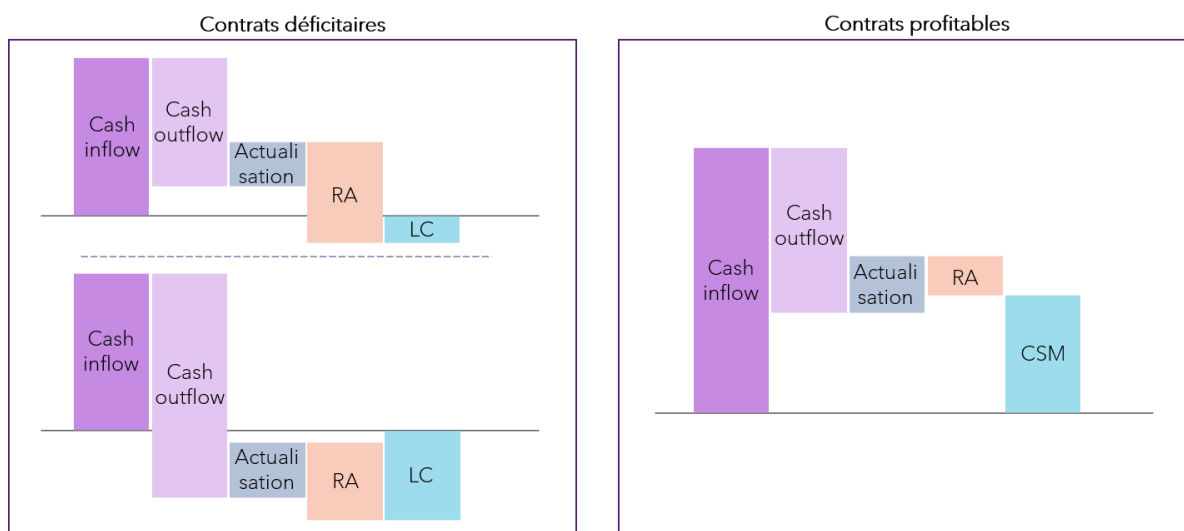


FIGURE 15 – Contrats déficitaires et profitables

Sur le schéma ci-dessus, le mot "cash inflow" fait référence aux revenus de l'entreprise et "cash outflow" fait référence aux dépenses de l'entreprise. Par ailleurs, nous remarquons qu'une LC est constituée à gauche car nous sommes dans le cas de contrats déficitaires et une CSM est constituée à droite car il s'agit de contrats profitables.

Dans le cas des contrats profitables, à la fin de la période, la CSM représente le bénéfice du groupe de contrats d'assurance. Ce bénéfice n'a pas encore été comptabilisé en résultat. Ensuite, à chaque date de revalorisation future, les ajustements de la CSM seront recalculés en prenant en compte les variations des taux d'intérêts, la valeur des nouveaux services futurs et la valeur des services rendus pendant la période de référence.

Pour les contrats onéreux, IFRS 17 impose de comptabiliser immédiatement, en compte de résultat, les pertes futures rattachées à ces contrats et également que ces pertes ne soient pas compensées avec des contrats profitables. Une fois que l'entité a reconnu une perte sur un groupe de contrats onéreux, elle doit la reconnaître dans une composante de perte. Ensuite, l'entité doit systématiquement répartir les variations ultérieures du Liability for Remaining Coverage (LRC, cf. définition 16) entre la composante de perte (qui est contenue dans le LRC) et le reste du LRC.

Les augmentations ou les diminutions ultérieures des FCF alloués au groupe de contrats pour les services futurs doivent être affectées uniquement à la LC jusqu'à ce que cette composante soit réduite à zéro. Les diminutions des FCF pour les services futurs qui réduisent à zéro la composante de perte du LRC rétablissent une CSM.

Pour les contrats non-onéreux, le profit sera amorti au fur et à mesure des montants versés aux assurés. Le profit peut être distribué annuellement au rythme des prestations attendues, par exemple.

Beaucoup de processus actuels d'assureurs ne seront pas adaptés à ces regroupements et les données pourraient ne pas être disponibles à ce niveau de granularité. Par ailleurs, un des principaux objectifs est de déterminer comment identifier les contrats onéreux ou potentiellement onéreux.

IFRS 17 impose de considérer comme date de comptabilisation des contrats, la première des trois dates suivantes :

- La date de commencement de la période de couverture du contrat.
- La date à laquelle la première prime doit être payée par l'assuré ou la date de réception de la première prime s'il n'y a pas de date fixée pour le paiement.
- Pour les contrats onéreux, la date à laquelle l'organisme d'assurance estime que le contrat devient déficitaire.

Dès lors, la maille de calcul différente et la granularité plus fine des données sous IFRS 17 multiplient le nombre de scénarios et le volume des données à traiter. Néanmoins, les problématiques issues de la valorisation de la CSM existent uniquement pendant la période de couverture. Effectivement, comme nous l'avons vu, la CSM est amortie au fur et à mesure de la période de couverture, elle est donc totalement amortie à la fin de cette dernière.

Après l'obtention de la granularité, la modélisation actuarielle va nous permettre de déterminer le passif sur trois blocs principaux :

- Les provisions en Best Estimate.
- Le Risk Adjustment.
- La perte ou le profit futur : Contractual Service Margin.

4.3.3 Le passif sous IFRS 17

4.3.3.a Le Best Estimate

Comme nous l'avons vu plus tôt, le BE correspond à la valeur actuelle attendue des flux de trésorerie futurs.

Dans le contexte IFRS 17, il est important de définir la notion de frontière d'un contrat et la courbe d'actualisation utilisée. En effet, l'estimation des flux de trésorerie futurs se fait en les comptabilisant dans la limite de la frontière des contrats. Ces flux sont actualisés et pondérés par leur probabilité de survenance.

La frontière des contrats est définie ainsi (paragraphe 34 d'IFRS 17) : Les flux de trésorerie sont compris dans la frontière d'un contrat d'assurance s'ils découlent de droits et d'obligations substantiels qui existent au cours de la période de présentation de l'information financière durant laquelle l'entité peut contraindre le titulaire de contrat d'assurance à payer les primes ou, durant laquelle, l'entité a une obligation substantielle de fournir des services au titulaire du contrat d'assurance. Une obligation substantielle de fournir des services cesse dans l'une ou l'autre des situations suivantes :

- L'entité a la capacité pratique de réévaluer les risques posés spécifiquement par le titulaire de contrat d'assurance et peut, par conséquent, fixer un prix ou un niveau de prestations qui reflète intégralement ces risques.
- Les deux critères ci-dessous sont remplis :
 - L'entité a la capacité pratique de réévaluer les risques posés par le portefeuille de contrats d'assurance dont fait partie le contrat en cause et peut, par conséquent, fixer un prix ou un niveau de prestation qui reflète intégralement le risque posé par le portefeuille.
 - La tarification des primes pour la couverture qui s'étend jusqu'à la date de réévaluation des risques ne tient pas compte des risques liés aux périodes postérieures à la réévaluation.

Il faut, cependant, faire attention à la notion de frontière de contrat. En effet, S2 aborde la réévaluation des risques à la maille "portefeuille" alors qu'IFRS 17 prend en compte une vision "contrat".

Pour la valorisation des contrats d'assurance, les données requises sont, en général, regroupées selon les catégories suivantes :

- Les engagements de l'assureur : Il s'agit de tous les montants que l'assureur doit régler à l'assuré si un sinistre ou un événement déclencheur (par exemple, la retraite) a lieu durant la période de couverture du contrat.
- Les engagements de l'assuré : Il s'agit de toutes les primes que l'assuré doit payer pendant la vie du contrat.
- Les frais : Coût d'acquisition, coût de gestion des contrats, coût de gestion des sinistres et frais généraux.
- Les taxes : Notamment, les prélèvements sociaux.
- Les rachats : Rachats conjoncturels et rachats structurels.
- Les flux potentiels de contre-assurance.
- Les flux de réassurance.

L'entité doit estimer ces flux hypothétiques en montant, en probabilité et en date de survenance pour chaque regroupement de contrats.

Dès lors, pour l'évaluation de chaque groupe de contrats, l'assureur doit inclure tous les flux de trésorerie futurs compris dans le périmètre de chacun des contrats du groupe. Les estimations de flux de trésorerie futurs doivent (paragraphe 33 d'IFRS 17 Insurance Contracts) :

- Intégrer, de manière objective, l'ensemble des informations raisonnables et justifiables qu'il est possible d'obtenir, sans coût ou effort excessif, sur le montant, les dates de versements et l'incertitude relative aux flux de trésorerie futurs. Pour ce faire, l'entité doit estimer l'espérance mathématique de l'ensemble des résultats possibles.
- Refléter le point de vue de l'entité à condition que les estimations des variables de marché pertinentes soient cohérentes avec les prix de marché observables pour ces variables.
- Être à jour à la date d'évaluation : Les estimations doivent refléter les conditions existant à la date d'évaluation y compris les hypothèses sur le futur à cette date.

- Être explicites : L'entité doit séparer l'estimation du RA des autres estimations. L'entité doit également séparer l'estimation des flux de trésorerie de l'estimation de l'ajustement au titre de la valeur temps de l'argent et du risque financier sauf si la technique d'évaluation la plus appropriée combine ces estimations.

Une entité doit revoir et mettre à jour ses estimations depuis la clôture de la période de reporting précédente. Ainsi, une entité doit déterminer si les estimations qui ont été mises à jour représentent fidèlement les conditions à la fin de la période de reporting et les changements durant la période ou non.

Pour le calcul du BE, on parle de courbe des taux sans risque pertinente car, celle-ci, doit présenter des caractéristiques de liquidité et de cohérence avec les prix de marché d'instruments financiers adéquats.

4.3.3.b Les taux d'actualisation

La détermination de la courbe des taux d'actualisation est très importante car le montant du BE en dépend. Ainsi, la CSM à l'origine en dépend aussi et c'est donc l'ensemble de la reconnaissance du résultat sous IFRS 17 qui va en dépendre. Le taux d'actualisation est d'autant plus important que les engagements sont de longue durée. Les taux d'actualisation utilisés doivent respecter certains principes en application du paragraphe 36 d'IFRS 17 Insurance Contracts :

- Les taux d'actualisation reflètent la valeur temps de l'argent, les caractéristiques des flux de trésorerie des contrats d'assurance et les caractéristiques de liquidité des contrats d'assurance auxquels ils sont appliqués.
- Les taux d'actualisation doivent être conformes aux prix actuellement observables du marché (si de tels prix existent) pour des actifs avec des flux de trésorerie dont les caractéristiques (telles que le temps, la devise et la liquidité) correspondent à celles des contrats d'assurance.
- Les taux d'actualisation excluent l'impact de quelconques facteurs sur les prix observables du marché mais qui n'affectent pas les flux de trésorerie futurs des contrats d'assurance.

Il est donc demandé à ce que les entreprises utilisent plusieurs courbes de taux différentes selon la nature des flux à actualiser. Il n'y a pas de prise en compte du risque propre à l'entité. Les taux d'actualisation devraient uniquement inclure des facteurs pertinents pour les flux de trésorerie des passifs d'assurance auxquels les taux seront appliqués et ne devraient donc pas, par défaut, être égaux aux rendements attendus des actifs qui sont en fait détenus pour soutenir les passifs d'assurance. De plus, les taux d'actualisation doivent refléter la valeur temps de l'argent car la valeur de l'argent de demain ne sera pas la même que celle d'aujourd'hui.

Pour déterminer le taux d'actualisation, deux approches s'offrent aux compagnies d'assurance :

- L'approche Bottom-Up : Nous partons de la courbe des taux sans risque et nous l'ajustons pour y inclure d'autres éléments tels que la prime de liquidité. Cette approche ressemble à celle utilisée sous S2 (utilisation d'une courbe des taux sans risque ajustée), cependant, IFRS 17 ne donne pas d'indication sur le calcul de la prime de liquidité. Cette prime a pour but de corriger les écarts de liquidité qui existent entre les contrats d'assurance et les instruments financiers qui ont servi à construire la courbe des taux sans risque. Les courbes de taux sans risque sont souvent construites à partir d'obligations très liquides que nous pouvons vendre quand nous le souhaitons alors que les passifs d'assurance sont moins liquides.
- L'approche Top-Down : Nous partons de la courbe de rendement de marché d'un portefeuille d'actifs financiers de référence (qui correspond aux passifs des produits d'assurance que nous étudions). Ensuite, nous lui retirons la rémunération des risques non rattachables aux contrats d'assurance tels que le risque de défaut de l'émetteur des obligations. Cette approche semble difficile à appliquer car il faut trouver un portefeuille d'actifs financiers de référence adapté.

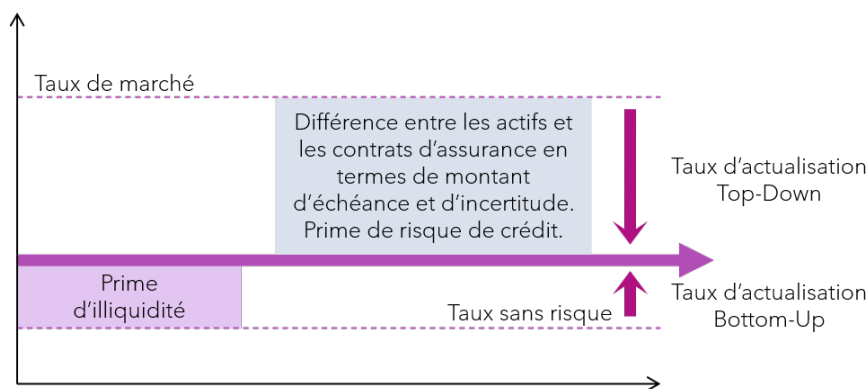


FIGURE 16 – Bottom-Up VS Top-Down

Sur le schéma, nous pouvons constater que, théoriquement, peu importe la méthode utilisée, le taux d'actualisation obtenu est le même. Par ailleurs, selon IFRS 17, il n'est pas nécessaire de vérifier que les deux méthodes arrivent au même résultat.

Pour l'IASB, il est très important de prendre en compte la valeur temps de l'argent. Par principe, tous les flux de trésorerie doivent être actualisés même en assurance non-vie. Toutefois, il faut considérer plusieurs taux :

- Pour les flux de trésorerie qui ne dépendent pas du rendement des actifs, il est possible d'utiliser une courbe des taux sans risque ajustée de la liquidité. Il s'agit d'une méthode Bottom-Up.
- Pour les flux de trésorerie qui dépendent du rendement des actifs et, dans la mesure où ils en dépendent, il est possible de construire une courbe des taux de marché corrigée des caractéristiques du contrat d'assurance (liquidité, échéances et risque, par exemple). Il s'agit d'une méthode Top-Down.

Il est également possible de différencier des taux nominaux et des taux réels lorsque les flux de trésorerie dépendent de l'inflation.

Il faut ensuite ajouter une dimension temporelle à ces différentes courbes de taux. En effet, chaque courbe utilisée pour calculer les données d'un contrat en valeur initiale doit être mémorisée afin d'être réutilisée par la suite. L'assureur est tenu de publier ses courbes de taux, il y aura ainsi plusieurs courbes de taux car ce ne sera pas la même pour chaque génération de contrats, pour chaque nature de flux et pour chaque date initiale de valorisation. En effet, le taux d'actualisation est une hypothèse économique, il doit donc être remis à jour à chaque date de reporting. Dans le modèle général, une variation du taux d'actualisation d'une date d'arrêt à l'autre génère des charges ou des produits financiers d'assurance (cf. définition 17). Avec le modèle PAA, plusieurs simplifications sont possibles quant à l'actualisation :

- Il n'est pas nécessaire d'actualiser le LRC à moins qu'il n'y ait une composante financière importante dans le groupe de contrats. Quand le temps écoulé entre le moment où les primes sont dues et le début de la couverture est inférieur ou égal à un an, le groupe est réputé comme n'ayant pas une composante financière importante.
- Pour le Liability for Incurred Claims (LIC, cf. définition 18), les cash flows n'ont pas à être actualisés si nous nous attendons à les payer dans l'année. Pour les cash flows à durée plus longue, il faudrait utiliser le taux actualisé à la date où le sinistre a été déclaré plutôt que le taux à la date de reconnaissance initiale du contrat.

4.3.3.c Le Risk Adjustment

Prendre seulement en compte une vision BE dans la détermination des provisions ne suffit pas, il est nécessaire d'ajouter un RA au BE. Le RA mesure le montant que l'entité demanderait en plus du BE pour supporter les risques non-financiers. IFRS 17 n'impose pas de méthodologie pour calculer le RA, en revanche, il est demandé de traduire le montant calculé en niveau de quantile équivalent pour permettre de comparer les RA entre eux. Ainsi, même si l'assureur n'utilise pas

la VaR pour évaluer le RA, il devra publier un équivalent du niveau de confiance donné par la VaR.

En fait, selon le paragraphe B87 d'IFRS 17 Insurance Contracts, le RA des contrats d'assurance mesure la rémunération dont l'entité aurait besoin afin d'être indifférente entre :

- Avoir un passif ayant un éventail de résultats possibles découlant d'un risque non-financier.
- Avoir un passif qui générera des flux de trésorerie fixes avec la même valeur actuelle attendue que les contrats d'assurance.

Par ailleurs, le RA reflète le degré du bénéfice de diversification que l'entité inclut lors de la détermination de la compensation dont elle a besoin pour supporter ce risque. Il reflète également les résultats à la fois favorables et défavorables d'une manière qui reflète le degré d'aversion au risque de l'entité.

Le RA valorise la compensation permettant à l'assureur de supporter l'incertitude quant à ses flux de trésorerie futurs pris en compte dans le BE. Au niveau consolidé, le RA bénéficie de la diversification entre les risques mais également de la diversification entre l'ensemble des portefeuilles et des groupes de contrats. Dès lors, plus les risques souscrits sont diversifiés, plus le RA peut être faible. L'importance de cette compensation est différente pour chaque assureur, en particulier, nous noterons que le niveau de diversification à prendre en compte n'est pas régi par IFRS 17.

Le RA peut être comparé à la Risk Margin de S2. Cependant, les définitions diffèrent quelque peu et, parmi ces différences, nous pouvons noter que la RM est définie dans le contexte d'une valeur transactionnelle alors que le RA représente la vue interne du risque non-financier inhérent aux flux de trésorerie du passif et exclut le risque opérationnel général. De plus, IFRS 17 autorise les entreprises à choisir librement la méthode à utiliser pour calculer le RA tandis que, sous S2, la RM est calculée en utilisant une méthode coût du capital définie.

Thème	Risk Margin	Risk Adjustment
Général	Mesure de la marge de prudence requise dans l'hypothèse d'un transfert du portefeuille à une autre société	Mesure de la compensation attendue par l'assureur pour porter l'incertitude sur les flux futurs générée par les risques non-financiers
Méthode générale	Coût du capital	Pas de méthode obligatoire mais il faut donner le quantile correspondant au RA obtenu
Calibration	Le taux de coût du capital est prescrit (6%) et seuil de confiance de 99,5% implicite au calcul des SCR	Possibilité de calibrer l'ensemble des paramètres de la méthode retenue en fonction de l'aversion au risque de l'entité d'assurance
Horizon de risque	Vision à un an	Doit recouvrir la durée des engagements
Périmètre des risques	Risques de souscription, de contrepartie et opérationnel	Risques non-financiers relatifs aux contrats d'assurance
Bénéfices de diversification	Reflète le niveau de diversification de l'assureur au niveau de l'entité	Possibilité de prendre en compte la diversification au niveau groupe
Réassurance	Calculée nette des effets d'atténuation par les contrats de réassurance	Calcul d'un RA séparé pour les contrats de réassurance avec une définition différente de la définition générale

FIGURE 17 – Risk Margin VS Risk Adjustment

Si IFRS 17 permet aux assureurs d'adapter les calculs aux risques individuels, cette norme introduit également une complexité supplémentaire autour de la sélection des méthodes. De plus, le RA est propre à l'entité car il reflète sa propre tolérance au risque.

Il existe plusieurs méthodes d'évaluation du RA mais aucune n'est obligatoire :

- La VaR est une méthode connue et utilisée sous S2 et simple à mettre en oeuvre. Toutefois, elle est moins adaptée pour les distributions de pertes asymétriques. En effet, elle estime la perte maximale que l'assureur risque de subir sur un horizon de temps donné et avec un niveau de confiance donné. Elle possède donc l'avantage d'une communication financière simplifiée mais elle n'est pas applicable pour les risques ayant des pertes potentielles extrêmes. Elle

est utilisée sous S2 pour calculer le SCR de façon à limiter la probabilité de ruine à 0,5% à horizon d'un an (i.e. VaR à 99,5%). $Var_\alpha = \inf\{x | \mathbb{P}(X \leq x) \geq \alpha\}$.

- La Conditional Tail Expectation (CTE) ou Tail Value at Risk (TVaR) est une méthode similaire à la VaR mais qui offre une meilleure estimation. Elle s'interprète comme la moyenne probabilisée de la VaR sur l'intervalle $[VaR_\alpha; +\infty[$. Néanmoins, elle nécessite plus de données extrêmes et est donc plus difficile à mettre en place. Elle possède donc l'avantage d'une communication financière simplifiée comme la VaR et c'est aussi une méthode applicable à l'ensemble des risques contrairement à la VaR mais elle suppose de connaître toute la distribution de probabilité des risques. $TVaR_\alpha = \mathbb{E}[X | X \geq VaR_\alpha(X)]$.
- Le Cost of Capital est une méthode utilisée sous S2 pour calculer la RM (voir sous-partie 4.2.2.a). Cependant, elle nécessite des variables supplémentaires (les hypothèses et les sensibilités) et, contrairement à S2, l'horizon pour le capital est la durée de vie des FCF. En effet, la méthode est fondée sur le coût de la mobilisation des fonds propres nécessaires afin de respecter les engagements d'assurance. Ce coût est déterminé en appliquant un taux de coût du capital qui est fonction de la rentabilité des capitaux investis que les actionnaires attendent. Cette méthode possède l'avantage d'être déjà utilisée par S2 pour calculer la Risk Margin avec un taux de coût du capital fixé à 6%, cependant, elle a l'inconvénient d'une communication financière complexe contrairement aux deux autres méthodes.

En ajoutant le RA au BE calculé, l'assureur sera en mesure de couvrir les quantiles correspondant à des risques plus élevés.

Nous noterons, par ailleurs, que l'analyse des variations nécessitera de calculer le RA plusieurs fois pour les différentes étapes des variations de la CSM. En effet, l'estimation du RA doit se faire à chaque pas de la projection et il doit être libéré tout au long de la période de couverture comme la CSM.

4.3.3.d La Contractual Service Margin

Après considération du BE et du RA, il reste à considérer la CSM. Celle-ci correspond aux profits/pertes non encore reconnus car ils sont associés aux couvertures/services futurs qui seront fournis à l'assuré. La CSM correspond aux primes perçues à la souscription diminuées du BE et du RA. Lorsqu'un contrat contient plusieurs garanties, nous devons mesurer la CSM en additionnant toutes les garanties du contrat. IFRS 17 ne permet pas de reconnaître tout de suite ces profits futurs en fonds propres parce que la CSM est amortie sur la période de couverture du contrat. En effet, la CSM est amortie en suivant la diminution des unités de couverture ("coverage units") au cours du temps. Elle est donc amortie au prorata de la quantité de services fournis sur la période par rapport à la quantité de services restant à fournir à l'échelle du groupe de contrats et elle est comptabilisée dans le compte de résultat au fur et à mesure que ces services sont fournis. Cette quantité de services est exprimée en nombre d'unités de couverture. La part de la CSM qui sera libérée lors des futures clôtures de comptes dépendra donc des unités de couverture et de la durée de couverture du contrat. Sous S2, ces profits futurs étaient reconnus directement en fonds propres.

Un montant de CSM pour un groupe de contrats d'assurance est comptabilisé en résultat pour chaque période afin de refléter les services fournis, dans le cadre du groupe de contrats d'assurance, au cours de cette période. Le montant est déterminé par (paragraphe B119 d'IFRS 17 Insurance Contracts) :

- L'identification des unités de couverture dans le groupe : Le nombre d'unités de couverture dans un groupe est la quantité de couverture fournie par les contrats du groupe. Ce nombre d'unités de couverture attribué à chaque contrat doit refléter la quantité de services fournis et la durée de la couverture.
- La répartition de la CSM à la fin de la période, à parts égales, dans chaque unité de couverture fournie pendant la période en cours et dans chaque unité de couverture qui devrait être fournie pendant les périodes futures (avant de comptabiliser tout montant en résultat afin de refléter les services fournis au cours de la période).
- La comptabilisation, en résultat, du montant alloué aux unités de couverture fournies au cours de la période.

Une entité doit évaluer la CSM lors de la comptabilisation initiale d'un groupe de contrats d'assurance à un montant qui, à moins que le groupe de contrats ne soit onéreux, n'entraîne aucun produit ou charge résultant de (paragraphe 38 d'IFRS 17 Insurance Contracts) :

- La comptabilisation initiale d'un montant pour les FCF.
- La décomptabilisation à la date de la comptabilisation initiale de tout actif ou passif comptabilisé pour les flux de trésorerie liés à l'acquisition de contrats d'assurance.
- Les flux de trésorerie éventuels résultant des contrats du groupe à cette date.

IFRS 17 a introduit trois approches permettant de reconstituer la CSM à la date de transition afin de déterminer les résultats comptables futurs :

- Approche rétrospective complète : Il faut appliquer la norme à chaque groupe de contrats de la même manière que nous l'aurions fait si la norme existait depuis toujours. Cela s'applique aux contrats encore en cours et à certains contrats échus qui auraient pu impacter les montants au bilan d'ouverture. Il est donc nécessaire d'identifier, de reconnaître et de valoriser tous les groupes de contrats comme si IFRS 17 avait toujours existé. Toutefois, il faut aussi décomptabiliser les éléments qui ont été comptabilisés mais qui ne l'auraient pas été sous IFRS 17 et, ainsi, reconnaître en résultat les différences de montants qui résultent de cette approche. Il faut déterminer le bilan à date de première comptabilisation et amortir la CSM jusqu'à la date de transition. Cette approche est très difficile à mettre en oeuvre en pratique car elle requiert beaucoup d'informations techniques et financières concernant chaque groupe de contrats à la date de première comptabilisation. En effet, elle a besoin de toutes ces données afin de pouvoir reconstituer les groupes de contrats, les flux de trésorerie passés et le Risk Adjustment.
- Approche rétrospective modifiée : Si nous ne pouvons pas appliquer la première approche sans engager beaucoup de coûts et d'efforts, il faut effectuer un calcul basé sur les FCF à la date de transition et les données comptabilisées pour le passé. Cette approche permet de regrouper plusieurs cohortes au sein d'un même groupe de contrats. Ainsi, elle permet d'équilibrer la balance entre simplicité de mise en oeuvre et mesure de profits futurs. En effet, cette approche permet d'utiliser, si nécessaire, une série de simplifications qui consistent, par exemple, à remplacer la date de souscription par des estimations en date du bilan d'ouverture, à regrouper des contrats sur des cohortes de plus d'un an ou à négliger les écarts d'expérience des flux sur les périodes passées, i.e., les différences entre les flux historiques réellement survenus et les estimations prospectives que nous aurions pu faire. Si nous possédons les données historiques à une granularité très fine, l'approche permet de comptabiliser un bilan et une CSM de transition. Cette méthode est donc très coûteuse à mettre en oeuvre, malgré les simplifications apportées, car elle nécessite de disposer de nombreuses données passées à une maille très fine.
- Approche juste valeur : Elle ne prend en compte que des flux futurs et la CSM est calculée en faisant la différence entre les provisions techniques (BE et RA) et la juste valeur des actifs en représentation des contrats. Cette approche a l'avantage d'être seulement basée sur les flux de trésorerie futurs. Les normes IFRS ont défini la notion de juste valeur principalement en référence à un prix de transaction, celui-ci est difficile à cerner lorsqu'il s'agit de portefeuilles de contrats d'assurance. En effet, IFRS 13 a énoncé deux approches de détermination de la juste valeur : une approche "marché" dans laquelle l'évaluation est réalisée à partir de prix observés sur le marché des transactions et une approche "modèle" dans laquelle l'évaluation se fait à l'aide d'un modèle en cas d'impossibilité d'utilisation de la première méthode.

Afin de choisir une des trois méthodes, il convient notamment d'évaluer la faisabilité de la méthode, il est possible de faire cela en mesurant la combinaison coût \times effort qui est nécessaire à la réalisation de la méthode. Par ailleurs, il est probable que les trois méthodes aboutissent à un niveau de CSM différent ce qui aura un impact sur les fonds propres. En effet, plus la CSM de transition sera élevée et moins les fonds propres le seront. Cependant, si la date de souscription de certains contrats est proche de la date de transition, il sera nécessaire d'utiliser la méthode rétrospective complète pour ces contrats car cela ne demandera pas autant d'efforts que pour les autres contrats. En effet, il faut utiliser les deux dernières méthodes uniquement dans le cas où la première approche n'est pas réalisable.

La complexité de la modélisation sous IFRS 17 résulte du fait que le RA et la CSM sont des nouvelles notions qui n'étaient pas calculées auparavant. Ainsi, il faudra déterminer la méthodologie et la maille de calcul utilisées pour leur valorisation.

Une CSM n'est pas spécifiquement identifiée pour les contrats soumis à l'approche PAA, bien que le même principe de constatation des bénéfices s'applique : aucun profit n'est reconnu lors de la comptabilisation initiale.

La CSM de clôture est déterminée sur la base de la CSM d'ouverture et d'une série d'ajustements successifs. IFRS 17 impose l'ordre dans lequel il faut procéder à ces ajustements. La CSM à la fin de la période de reporting représente le bénéfice du groupe de contrats d'assurance qui n'a pas encore été comptabilisé en résultat, car il se rapporte aux services futurs à fournir dans le cadre des contrats du groupe.

4.3.4 L'évolution du passif d'assurance durant la période de couverture

À la date où le contrat est souscrit, la période de couverture n'a pas encore débuté. Ainsi, les blocs du passif d'assurance (BE, RA et CSM) correspondent entièrement à un LRC. Durant l'écoulement de la période de couverture, le LRC diminue mais un LIC peut apparaître si des sinistres sont survenus. Il est très important de distinguer le LRC et le LIC car, sous IFRS 17, il est considéré que la diminution du LRC au cours d'une période reflète le niveau d'activité de la période : c'est sur cela que repose la méthode de reconnaissance du chiffre d'affaires sous IFRS 17. Cette distinction permet d'enregistrer certains effets différemment selon qu'ils s'appliquent au LRC ou au LIC. En effet, si nous affectons comptablement l'effet des changements d'estimation (par exemple, le changement de la table de mortalité) sur le LRC, l'effet sera imputé à la CSM tandis que, si nous affectons comptablement l'effet des changements d'estimation (par exemple, la réévaluation du coût d'un sinistre) sur le LIC, l'effet sera directement enregistré en résultat technique. Le compte de résultat est alimenté en analysant les variations du passif d'assurance d'un exercice sur l'autre.

À chaque date de clôture, la valeur comptable d'un groupe de contrats d'assurance doit être égale à la somme de deux montants :

- Le LRC qui est constitué des FCF se rapportant aux services futurs affectés au groupe et de la CSM du groupe à cette date.
- Le LIC qui est constitué des FCF se rapportant aux services passés affectés au groupe à cette date.

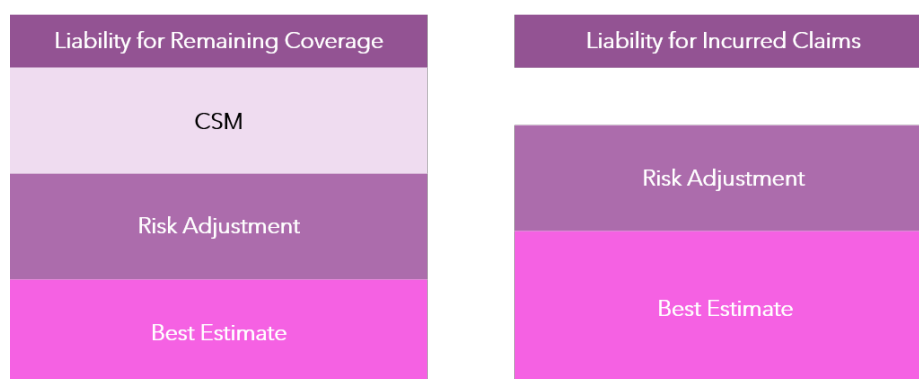


FIGURE 18 – Liability for Remaining Coverage et Liability for Incurred Claims

Sous IFRS 17, l'analyse détaillée des variations du passif d'assurance est indispensable :

- La variation du LRC génère le revenu d'assurance ainsi que du résultat financier d'assurance.
- Les variations d'hypothèses techniques entraînent des reclassifications entre les différentes composantes du LRC (BE, RA et CSM).
- La variation du LIC alimente à la fois les charges d'assurance et le résultat financier d'assurance.

- L'impact de la variation de la courbe des taux d'actualisation doit être identifié lorsque l'option OCI est activée.

Selon IFRS 17, une entité doit présenter, dans le résultat, le revenu d'assurance provenant des groupes de contrats d'assurance émis. Celui-ci doit représenter la fourniture de couvertures et d'autres services résultant du groupe de contrats d'assurance à un montant qui reflète la contrepartie que l'entité s'attend à recevoir en échange de ces services, c'est-à-dire, le montant des primes versées ajusté de l'effet financier (la valeur temps de l'argent) et excluant toute composante d'investissement. En effet, les composantes d'investissement sont comptabilisées séparément et ne font pas partie du résultat des services d'assurance.

La contrepartie totale pour un groupe de contrats couvre les éléments suivants (paragraphe B121 d'IFRS 17 Insurance Contracts) :

- Les montants liés à la fourniture de services comprenant :
 - Les frais de services d'assurance (cf. définition 19) à l'exclusion de tout montant alloué à la LC du LRC.
 - Le RA à l'exclusion de tout montant alloué à la LC du LRC.
 - La CSM.
- Les montants découlant des flux de trésorerie liés à l'acquisition de contrats d'assurance.

Lorsqu'une entité fournit des services au cours d'une période, elle réduit le LRC pour les services fournis et comptabilise un revenu d'assurance. La réduction du LRC qui donne lieu à un revenu d'assurance exclut les variations du passif ne concernant pas les services qui devraient être couverts par la contrepartie reçue par l'entité.

Les pratiques comptables existantes pour l'assurance-vie dans de nombreuses juridictions reconnaissent les primes dues durant une période comme équivalentes au revenu d'assurance. Sous IFRS 17, le revenu d'assurance exclut les composantes d'investissement et comptabilise le revenu au fur et à mesure que les services sont fournis plutôt qu'au moment où les primes doivent être reçues.

4.3.5 Approches comptables

Il existe trois différentes approches pour l'évaluation des passifs des contrats d'assurance. Ces trois modèles s'appliquent en supposant que la CSM a été évaluée à la date de première commercialisation du produit. Or, cette condition n'est pas réalisable pour la période transitoire car les contrats existent déjà avant l'entrée en vigueur d'IFRS 17. Nous évoquerons les trois approches mais nous nous concentrerons sur l'approche Premium Allocation Approach que nous avons utilisée dans la réalisation de nos calculs (voir sous-partie 4.5).

Les contrats d'assurance avec participation aux bénéfices directe sont essentiellement des contrats d'investissement pour lesquels l'entité émettrice promet un rendement basé sur des éléments sous-jacents (cf. définition 20). Ainsi, ils sont définis comme des contrats d'assurance dont (paragraphe B101 d'IFRS 17 Insurance Contracts) :

- Les conditions contractuelles prévoient que l'assuré participe au rendement d'un pool d'éléments sous-jacents clairement identifié.
- L'entité émettrice s'attend à payer à l'assuré une part substantielle du rendement évalué en juste valeur des éléments sous-jacents.
- L'entité émettrice prévoit qu'une part substantielle des flux de trésorerie qu'elle s'attend à payer à l'assuré varie avec les flux de trésorerie des éléments sous-jacents.

Il n'existe pas de pool d'éléments sous-jacents clairement identifié lorsque (paragraphe B106 d'IFRS 17 Insurance Contracts) :

- Une entité peut modifier les éléments sous-jacents qui déterminent le montant de l'obligation de l'entité avec effet rétroactif.

- Aucun élément sous-jacent n'a été identifié, même si l'assuré pourrait obtenir un rendement qui reflète généralement la performance et les attentes globales de l'entité ou celles d'un sous-ensemble d'actifs qu'elle détient.

Il existe aussi les contrats non participatifs et les contrats participatifs indirects :

- Les contrats non participatifs sont des contrats dont les flux ne varient pas avec les éléments sous-jacents tels que les contrats "Assurance Des Emprunteurs (ADE)" et les contrats "Incendie, Accidents et Risques Divers (IARD)". Le modèle adapté à ce type de contrats est le BBA ou le PAA pour les contrats remplissant les critères de ce modèle.
- Les contrats participatifs indirects sont des contrats dont les flux varient avec les éléments sous-jacents mais qui ne sont pas éligibles au modèle VFA contrairement aux contrats participatifs directs. Pour ce type de contrats il faut appliquer le modèle BBA.

Une entité évalue si un contrat est directement participatif en regardant ce qu'elle attend de ce contrat au début de celui-ci et ne réévalue pas les conditions sauf si le contrat est modifié. L'éligibilité à la méthode VFA dépend de l'attente des paiements d'une part substantielle des rendements de la juste valeur à l'assuré plutôt que de l'évaluation comptable des éléments sous-jacents.

Nous pouvons résumer le choix du modèle avec le tableau suivant :

Produit	BBA	PAA	VFA
Dépendance	X		
Épargne-retraite avec PB (euro et UC)			X
Assurance des Emprunteurs	X		
IARD	X	X	
Prévoyance collective	X	X	
Prévoyance individuelle	X	X	

TABLE 12 – Modèles disponibles par activité

Voici les différents modèles disponibles en prévoyance individuelle selon les risques :

BBA	PAA	VFA
Dépendance	DCAT	Obsèques
	MAV	
	PP	

TABLE 13 – Modèles disponibles en prévoyance individuelle

Les garanties que nous avons associées au modèle PAA peuvent également être évaluées sous le modèle BBA.

4.3.5.a Le modèle Building Block Approach

Le modèle BBA ou "General Model" est l'approche générale applicable aux contrats d'assurance. Cette méthode concerne les contrats sans participation directe aux bénéficiaires, en effet, comme nous l'avons évoqué, c'est l'approche VFA qui s'applique à ce type de contrats. Comme son nom l'indique, le modèle BBA est basé sur l'évaluation de trois blocs pour valoriser le passif d'assurance : le BE, le RA et la CSM qui ont été définis précédemment.

Pour le modèle BBA, le BE et le RA sont évalués indépendamment et la CSM est déterminée initialement par déduction des deux autres. Étant donné que la CSM n'existait pas sous S2, les outils de valorisation du passif d'assurance vont évoluer afin de prendre en compte le calcul de cette marge et son amortissement. De plus, sous IFRS 17, les valorisations reposent sur une estimation des flux de trésorerie futurs, i.e., une estimation des encaissements et des décaissements. Or, actuellement, les systèmes se basent sur la reconnaissance des produits et des charges et non

pas sur les flux de trésorerie, cela risque donc de poser des problèmes opérationnels.

Voici le fonctionnement du modèle BBA (schéma provenant du document référencé [ACTUARIS-2017] en bibliographie) :

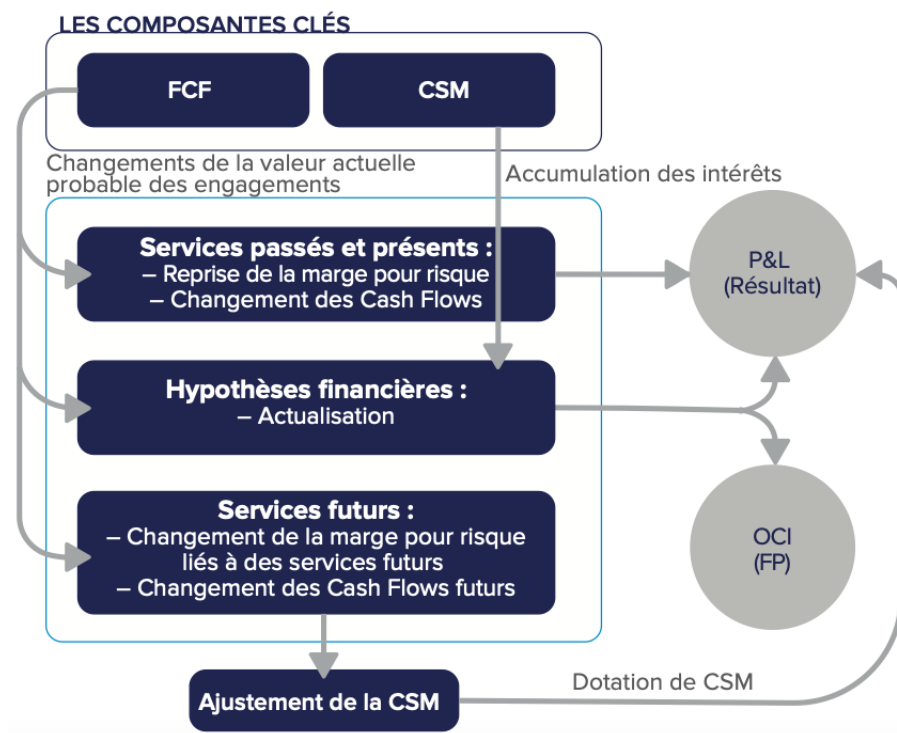


FIGURE 19 – Fonctionnement du modèle BBA

Si nous considérons une période donnée, les gains et les pertes peuvent être :

- Reconnus en résultat : Les gains et les pertes sont calculés avec un taux d'intérêt figé à l'origine.
- Reconnus en fonds propres (OCI) : Les gains et les pertes proviennent des changements des taux d'intérêts afin de limiter la sensibilité du résultat à la variation des taux d'intérêt. Néanmoins, nous retrouverons cette volatilité dans les fonds propres.

Si les gains et les pertes ne sont pas reconnus entièrement en résultat, il faudra calculer le passif au taux figé puis au taux courant. Cela augmentera la difficulté d'implémentation et allongera les temps de calculs.

4.3.5.b Le modèle Premium Allocation Approach

Cette approche est une simplification du modèle BBA car elle ne nécessite pas de décomposer les provisions en différents blocs, elle ne peut être appliquée que si l'une des deux conditions suivantes est vérifiée :

- Le contrat d'assurance a une durée de couverture courte (maximum un an).
- Le modèle PAA donne une approximation raisonnable du modèle BBA.

Un contrat doit donc soit vérifier le critère numéro 1 soit vérifier le critère numéro 2. Utiliser le modèle PAA pour des contrats pluri-annuels, c'est-à-dire, des contrats ne respectant pas le critère numéro 1, pourrait nécessiter une grande quantité de travail pour valider le second critère d'éligibilité "Le modèle PAA donne une approximation raisonnable du modèle BBA".

Selon le paragraphe 54 d'IFRS 17 Insurance Contracts, le second critère n'est pas vérifié si, à la création du groupe de contrats, l'entité s'attend à une variabilité significative des FCF qui affecterait l'évaluation du LRC pendant la période précédant la survenance d'un sinistre. La variabilité des FCF augmente, par exemple, avec :

- L'importance des flux de trésorerie futurs liés à tout dérivé incorporé dans les contrats.
- La durée de la période de couverture du groupe de contrats.

Pour les contrats éligibles :

- Le modèle PAA offre des résultats similaires au modèle général à moindre coût.
- Le modèle PAA présente des similitudes avec l'approche comptes sociaux pour les contrats d'assurance non-vie.

Dans cette simplification du modèle BBA, l'amortissement de la CSM est remplacé par un calcul de prime acquise qui ressemble à la méthode utilisée en normes françaises. Cette approche permet à une entité d'évaluer le montant de la couverture restante en répartissant la prime sur la période de couverture. Un organisme d'assurance peut donc utiliser le modèle PAA pour évaluer approximativement les contrats sur la période de couverture restante mais il s'agit d'un modèle facultatif, il peut donc préférer l'utilisation du modèle général s'il le souhaite. Bien qu'il ne soit pas demandé explicitement d'effectuer un test pour démontrer que le modèle PAA est une approximation raisonnable du modèle BBA, il semble bon de présumer qu'un certain niveau de validation sera requis afin de satisfaire toutes les parties prenantes de l'entreprise et notamment son auditeur.

Le modèle PAA s'applique principalement au LRC. Cependant, la CSM et le RA ne seront pas calculés explicitement. Effectivement, dans le modèle PAA, la Provision pour Primes Non Acquises (PPNA) remplace le LRC du modèle BBA. Quant au LIC, il sera évalué comme sous le modèle BBA car la simplification ne peut s'appliquer qu'aux sinistres non survenus. Néanmoins, il n'est pas obligatoire d'actualiser les flux de sinistres et de frais lorsque les délais de liquidation sont inférieurs à un an. De plus, l'actualisation des primes perçues n'est pas obligatoire si celles-ci servent à couvrir une période qui dure moins d'un an. L'assureur reconnaît, en résultat, la part des primes reçues allouées à la période. Il s'agit d'un amortissement de ces primes et il doit être fait au fur et à mesure du temps sauf si nous savons que le risque n'évolue pas linéairement.

Le modèle PAA ne peut pas être utilisé pour les contrats onéreux. Ainsi, sous ce modèle, s'il n'y a pas de faits et de circonstances indiquant le contraire, l'assureur peut supposer qu'aucun des contrats du portefeuille n'est onéreux à l'origine. Cependant, si, à un moment de la période de couverture, les faits et circonstances indiquent qu'un groupe de contrats est onéreux, il sera nécessaire de calculer la différence entre (paragraphe 57 d'IFRS 17 Insurance Contracts) :

- La valeur comptable du LRC sous le modèle PAA.
- Les FCF relatifs à la couverture restante du groupe sous le modèle BBA. Néanmoins, si l'entité n'ajuste pas le LIC pour la valeur temps de l'argent et l'effet du risque financier dans l'application du modèle PAA, elle ne doit pas inclure ces ajustements dans les FCF.

Dans la mesure où les FCF sous le modèle BBA dépassent la valeur comptable du LRC sous le modèle PAA, l'entité doit comptabiliser une perte en résultat et augmenter le LRC (paragraphe 58 d'IFRS 17 Insurance Contracts). Par ailleurs, l'assureur peut négliger la valeur temporelle des flux si ceux-ci sont encaissés ou versés dans une période de 12 mois maximum.

De plus, sous le modèle PAA, il existe une option permettant de comptabiliser, en charges, les flux de trésorerie liés à l'acquisition de contrats d'assurance lorsque ces coûts sont engagés, à condition que la période de couverture de chaque contrat du groupe ne dépasse pas un an. Si l'entité n'est pas en mesure de le faire ou choisit de ne pas le faire, les flux de trésorerie liés à l'acquisition sont alors inclus dans le LRC.

La comptabilisation, en charges, des flux de trésorerie liés à l'acquisition de contrats d'assurance lors de la comptabilisation initiale d'un groupe de contrats d'assurance a pour effet d'augmenter le LRC à la comptabilisation initiale et de réduire la probabilité de toute perte ultérieure d'un contrat onéreux. Ainsi, nous observons une augmentation de la charge au résultat lors de la comptabilisation initiale due à la comptabilisation, en charges, des flux de trésorerie liés à l'acquisition de contrats d'assurance. Cette augmentation est compensée par une augmentation du profit relâché durant la période de couverture.

Le modèle PAA consiste à calculer une provision technique en $t = 0$:

$$\text{Provision technique initiale} = \text{Prime initiale} - \text{Flux de pré-ouverture} - \text{Frais d'acquisition}.$$

Ainsi, le modèle PAA devrait permettre de limiter l'impact de la mise en oeuvre d'IFRS 17, en particulier, pour les assureurs non-vie. Cependant, même si un assureur a l'intention de faire usage de cette méthode, sa mise en oeuvre ne sera pas aussi simple qu'elle peut le paraître à première vue car il faut effectuer des changements importants sur les processus utilisés.

Lorsqu'un assureur applique le modèle PAA, il doit évaluer le LRC de la manière suivante (paragraphe 55 d'IFRS 17 Insurance Contracts) :

- Lors de la comptabilisation initiale, la valeur comptable du passif est égale à :
 - + Les primes reçues à la date de comptabilisation (s'il y en a).
 - - Le montant des flux de trésorerie liés aux frais d'acquisition à cette date sauf si l'entité choisit de comptabiliser ces paiements en charges.
 - +/- Tout montant découlant de la décomptabilisation, à cette date, de l'actif ou du passif ayant été comptabilisé au titre des flux de trésorerie liés aux frais d'acquisition.
- À la fin de chaque période de présentation de l'information financière ultérieure, la valeur comptable du passif est égale à :
 - + Sa valeur comptable à l'ouverture de la période de présentation de l'information financière.
 - + Les primes reçues au cours de la période.
 - - Les flux de trésorerie liés aux frais d'acquisition sauf si l'entité choisit de comptabiliser ces paiements en charges.
 - +/- Tout montant relatif à l'amortissement des flux de trésorerie liés aux frais d'acquisition qui a été comptabilisé en charges dans la période de présentation de l'information financière.
 - + Tout ajustement apporté à une composante de financement.
 - - Le montant comptabilisé en revenu d'assurance pour la couverture fournie durant cette période.
 - - Toute composante d'investissement payée ou transférée au LIC.

Selon le paragraphe 56 d'IFRS 17 Insurance Contracts, si les contrats d'assurance du groupe ont une importante composante de financement, l'entité doit ajuster la valeur comptable du LRC afin de refléter la valeur temps de l'argent et l'effet du risque financier en utilisant des taux d'actualisation qui reflètent les caractéristiques des flux de trésorerie du groupe de contrats d'assurance au moment de la comptabilisation initiale. L'entité n'est pas tenue d'ajuster la valeur comptable du LRC afin de refléter la valeur temps de l'argent et l'effet du risque financier si, lors de la comptabilisation initiale, l'entité s'attend à ce que le délai entre la fourniture de chaque partie de la couverture et la date d'échéance de la prime correspondante ne dépasse pas un an.

Des différences entre les modèles BBA et PAA peuvent survenir après la reconnaissance initiale pour les groupes de contrats qui ne sont pas onéreux. Les principales différences dans l'évaluation du LRC peuvent être dues à :

- Des différences dans le modèle de libération du LRC en résultat en tant que revenu d'assurance. L'allocation du LRC pour un groupe de contrats appliquant le modèle PAA sur la base des dates de versements prévues pour les dépenses de services d'assurance encourues peut donner un résultat différent de celui obtenu avec le modèle général.
- L'effet des changements du BE qui ajuste la libération de la CSM durant la période en cours (voir exemple ci-après).
- Des variations des taux d'actualisation : Sous le modèle BBA, le LRC est mesuré en utilisant les taux actuels à chaque date de clôture tandis que, sous le modèle PAA, les taux d'actualisation ne sont pas mis à jour.

Nous pouvons comparer, dans un exemple, le LRC sous le modèle BBA et sous le modèle PAA en cas de variation des flux de trésorerie attendus. Soit un groupe de contrats valorisé selon le modèle général, nous supposons qu'une prime de 3 000 € est perçue au début d'une période de

couverture de deux ans. L'assureur estime que la somme du BE et du RA sera de 1 400 € pour chaque année de couverture. La CSM à l'ouverture vaut : $3\,000 - 1\,400 \times 2 = 200$. Lors de la première année, des sinistres surviennent pour un total de 1 400 € comme prévu. À la fin de la première année, l'assureur suppose que les flux de trésorerie de l'année de couverture suivante seront de 1 500 € au lieu des 1 400 € prévus. Nous avons le tableau suivant pour la CSM :

	CSM
Au début de la première année	200
Ajustement pour les services futurs	$1400 - 1500 = -100$
Relâchement de la CSM	$-\frac{200-100}{2} = -50$
A la fin de la première année	$200 - 100 - 50 = 50$

Sous le modèle général, le LRC à la fin de la première année serait de $1\,500 + 50 = 1\,550$ €. Le revenu d'assurance de la première année serait de $1\,400 + 50 = 1\,450$ (somme des prestations d'assurance prévues et du relâchement de la CSM). Le revenu d'assurance de la deuxième année serait de $1\,500 + 50 = 1\,550$.

Si l'entité avait appliqué le modèle PAA, elle aurait alloué 1 500 € au résultat de la première année car le revenu d'assurance et le LRC à la fin de la première année seraient de 1 500 €, un montant qui est donc différent de ceux obtenus sous le modèle général (LRC = 1 550 € et revenu d'assurance = 1 450 € à la fin de la première année).

L'obligation dans le modèle général de relâcher une partie de la CSM, après avoir procédé à des ajustements à cause de variations des flux de trésorerie attendus liés aux services futurs, peut donc causer une différence entre le LRC du modèle général et le LRC du modèle PAA.

4.3.5.c Le modèle Variable Fee Approach

Le modèle VFA a pour objectif d'adapter le modèle BBA aux contrats d'assurance dont la caractéristique principale est une prestation de gestion d'actifs, ces contrats sont dits "directement participatifs" ou "contrat avec participation aux bénéfices directe". Sur ce type de contrats, le profit provient de la marge financière que l'assureur prélève sur le portefeuille de placements sous-jacents : il s'agit de la "variable fee", d'où le nom de ce modèle. En fait, l'assureur a la possibilité d'ajuster les charges qui s'appliquent à la gestion des contrats pendant la durée de l'engagement. Ainsi, la rentabilité de ce type de contrats est sensible à l'évolution des conditions de marché. Le modèle VFA va essayer de capter cette sensibilité.

Il n'y a pas de différence entre le modèle BBA et le modèle VFA à la première comptabilisation, c'est lors des comptabilisations futures que les modèles vont différer car les modifications d'hypothèses ne sont pas comptabilisées de la même façon. La plus grande différence entre ces deux méthodes est la manière de calculer la CSM. En effet, le modèle VFA considère que le service fourni par le contrat est proche d'une activité de gestion d'actifs. Dès lors, l'impact d'une variation de la juste valeur des éléments sous-jacents, mais également, l'impact de l'environnement économique, sont considérés comme étant liés au service fourni par le contrat, ils viennent donc ajuster la CSM.

4.3.6 La rentabilité sous IFRS 17

4.3.6.a Le chiffre d'affaires

Sous IFRS 17, nous déterminons le chiffre d'affaires en analysant la diminution du LRC entre deux bilans successifs :

- Les flux de trésorerie du BE diminuent à la suite de la sortie des sinistres, prestations et frais initialement prévus pour la période écoulée.
- La réalisation du risque durant la période vient diminuer le montant du RA.
- La CSM, quant à elle, diminue car elle est amortie tout au long de la période de couverture.

Les trois éléments listés vont permettre de reconstituer la prime nette de frais d'acquisition (sous réserve des effets d'actualisation) sur laquelle nous avons calibré le LRC pour sa valorisation initiale. Nous allons ensuite ajouter une allocation de frais d'acquisition à ces éléments afin d'aboutir à un montant de CA exhaustif. Ainsi, le chiffre d'affaires est constitué des éléments suivants :

- Les sinistres, prestations et frais initialement prévus.
- Une reprise du RA.
- L'amortissement de la CSM.
- Une allocation de frais d'acquisition.

La mise en oeuvre d'IFRS 17 va nécessiter d'anticiper son impact sur la communication financière et le pilotage car elle modifie les indicateurs d'activité en profondeur. Par ailleurs, afin de rester cohérent avec les autres secteurs d'activité, le CA et le montant des prestations ne comprennent pas l'investissement des primes d'assurance vie sous IFRS 17.

4.3.6.b L'option Other Comprehensive Income

Sous IFRS 17, il faut identifier et inclure les variations du passif d'assurance qui sont liées à son actualisation dans le résultat financier. Il ne faut pas confondre ces variations liées à l'actualisation du passif avec les variations de provisions techniques qui alimentent le résultat opérationnel. Cependant, IFRS 17 permet d'enregistrer l'impact des variations de taux d'actualisation (et d'autres hypothèses relatives aux risques financiers) directement en capitaux propres et non en résultat financier. Cette comptabilisation est effectuée par le mécanisme des OCI. Cette méthode de comptabilisation est optionnelle et fait référence à la norme IFRS 9. Cette norme de comptabilisation des instruments financiers prévoit, dans certains cas, que les fluctuations de la valeur de marché des instruments financiers soient enregistrées directement en capitaux propres selon la même méthode.

En effet, si le compte de résultat IFRS 17 devait comprendre tous les revenus et les charges financières, les assureurs devraient l'évaluer en juste valeur afin de garantir l'adéquation avec l'actif. De plus, les profits et les pertes venant des activités de souscription et d'investissement pourraient être masqués par des profits et des pertes plus volatils venant du taux d'actualisation appliqué aux flux de trésorerie liés aux contrats d'assurance. Ainsi, cela pourrait augmenter la volatilité du résultat et biaiser la performance des assureurs. L'assureur peut donc retenir un taux qui va neutraliser totalement ou partiellement la variation du taux d'actualisation, l'effet ainsi masqué en résultat est passé en OCI.

L'option OCI consiste donc en la comptabilisation d'un profit ou d'une perte directement en capitaux propres plutôt qu'en résultat. Le montant enregistré en OCI au cours d'une période est la différence entre le total des charges ou des produits financiers d'assurance et le montant inclus dans le résultat. Le décompte des écarts enregistrés en OCI doit être nul au terme peu importe la situation. De plus, l'option OCI est choisie par groupe de contrats.

Dès lors, la comptabilisation via le mécanisme des OCI permet de diminuer la sensibilité du résultat financier à la variation des marchés que ce soit au niveau des contrats d'assurance ou des placements financiers.

Dans les normes IFRS, l'accroissement de richesse de l'année est distingué entre :

- Le résultat net qui rend compte de la performance du management.
- Les OCI qui rendent compte de l'effet des variables externes sur lesquelles le management n'a pas d'impact (par exemple, la fluctuation des marchés financiers). Les OCI sont également des éléments de résultat mais elles apparaissent dans un compte séparé.

4.3.6.c Les frais d'acquisition

Sous IFRS 17, nous ne pouvons plus reconnaître les frais d'acquisition reportés comme un actif. En effet l'IASB estime que ces frais seront pris en compte dans le tarif et qu'ils seront donc récupérés au fur et à mesure de la reconnaissance de la prime en produit. Nous devons donc prendre en compte ces frais dans les flux de trésorerie futurs. De ce fait, en alourdissant les flux de trésorerie, ils vont réduire d'autant la CSM. Au lieu de constituer un actif comme cela était fait auparavant, nous réduisons un passif (la CSM). Dès lors, les frais d'acquisition ne sont plus isolés au bilan mais ils restent identifiés au compte de résultat. Il faut noter que cela concerne uniquement les coûts attribuables aux groupes de contrats.

Par ailleurs, les frais d'acquisition et de gestion qui viennent minorer le montant de la CSM sont les frais qui sont directement rattachables aux portefeuilles de contrats. Cependant, l'IASB ne donne pas de critère précis permettant de caractériser le terme "directement rattachable".

Un assureur doit comptabiliser un actif ou un passif pour tout flux de trésorerie lié aux frais d'acquisition d'un groupe de contrats d'assurance émis que l'assureur paie ou reçoit avant que le groupe ne soit reconnu. Cependant, lorsque le contrat a une durée inférieure à un an, l'assureur peut choisir de les comptabiliser en charge ou en revenu en application du paragraphe 59a) d'IFRS 17 Insurance Contracts.

4.3.7 Conclusion IFRS 17

Afin de mettre en place IFRS 17, il faudra mobiliser des compétences très diverses présentes dans différentes fonctions des organismes d'assurance. Cette mise en place devrait, par ailleurs, être organisée sous forme d'un projet qui reflète l'importance des enjeux sur les fonctions suivantes : Finance, Risque et Actuariat, Données et Systèmes. En effet, une collaboration efficace de ces fonctions est identifiée comme l'un des principaux facteurs clés de succès. Les travaux de transition effectués combinent, en particulier, les périmètres de l'actuariat, de la comptabilité et du contrôle de gestion :

- L'actuariat peut apporter son domaine d'expertise dans la valorisation en juste valeur.
- La comptabilité apporte la connaissance de la donnée historique ce qui est utile pour l'utilisation de méthodes rétrospectives de calcul de la CSM.
- Le contrôle de gestion pourra ajouter son expérience en modèles de coûts et en pilotage.

Il est nécessaire de définir un pilotage efficace, d'organiser les différents chantiers et de mettre en place une certaine transversalité permettant de suivre la progression du projet et de faire le lien entre les parties prenantes.

En plus des enjeux de la multidisciplinarité, la mise en place de cette norme implique des changements très importants à effectuer dans des délais très courts pour certains assureurs et, en particulier, pour les assureurs vie. Dès lors, l'organisation du projet IFRS 17 devient la clé d'une mise en oeuvre réussie. C'est dans ce cadre que la tribu IFRS 17 a été mise en place chez Natixis Assurances.

Selon IFRS 17, les passifs relatifs aux contrats d'assurance seront calculés comme une valeur actuelle des flux de trésorerie futurs avec un ajustement pour risque. De plus, le taux d'actualisation utilisé sera représentatif des taux d'intérêt actuels. Si la valeur actuelle des flux de trésorerie futurs produit un gain à l'émission du contrat, le modèle exigera une marge de service contractuelle afin de compenser le gain du premier jour. Cette marge sera amortie sur la durée du contrat. Il faudra également réaliser une nouvelle présentation du compte de résultat.

Finalement, les enjeux de la première application de cette norme sont fortement liés au montant de la marge de service contractuelle qui sera calculé au 1er janvier 2023 car ce montant conditionnera les résultats futurs et la capacité des assureurs à faire face à des déviations de la sinistralité.

4.4 Évaluation de la rentabilité des contrats

4.4.1 Première comptabilisation

Le Best Estimate de sinistres en 0 est égal à :

$$BE_0 = \sum_{i>0} \frac{F_i}{(1+r_i)^i}, \text{ où :}$$

- F_i : Les flux de trésorerie relatifs aux sinistres en i .
- r_i : Le taux d'actualisation de maturité i qui est issu de la courbe des taux à déterminer.

La Contractual Service Margin en $t = 0$ est égale à :

$$CSM_0 = \text{Max}(0; CF_0 - (BE_0 + RA_0)).$$

Où :

- CF_0 : Les flux de trésorerie relatifs aux primes à $t = 0$.
- BE_0 : Le Best Estimate à $t = 0$.
- RA_0 : Le Risk Adjustment à $t = 0$.

La CSM à la première comptabilisation est calculée afin d'étaler le résultat attendu tout au long de la période de couverture, il n'y a donc pas de profit à l'origine.

4.4.2 Modèle Building Block Approach : Exemple calculatoire

Soit un portefeuille de contrats qui couvre un risque quelconque pour une durée de deux ans. Nous supposons que l'assureur encaisse 1 100 € de primes et les sinistres sont estimés à 400 € par an (soit 800 € pour les deux ans). Nous valorisons un montant de RA à 150 €. Nous prenons un taux d'actualisation à 3%.

Nous allons maintenant estimer les flux futurs, i.e., nous allons calculer un BE en actualisant les montants des prestations versées chaque année avec un taux d'actualisation à 3%. Nous obtenons les résultats suivants :

BE	Année 1	Année 2
$388 + 377 = 765$	388	377

Le BE de l'année 1 a été obtenu de la façon suivante : $BE_1 = \frac{400}{(1+3\%)^1}$. Ainsi, le BE total vaut 765. Si nous ajoutons, à ce montant, le montant du RA, nous obtenons : $BE+RA = 765+150 = 915$.

La CSM est la différence entre les primes encaissées et la somme du BE et du RA. Ainsi, la CSM en $t = 0$ vaut : $CSM = 1\ 100 - 915 = 185$.

Nous pouvons en déduire le bilan à $t = 0$:

Actif	Passif
1 100	CSM = 185
	RA = 150
	BE = 765

Au moment de la clôture des comptes la première année, il faut recalculer le BE à partir des flux futurs actualisés au taux de 3%. Nous obtenons ceci :

BE	Année 2
388	388

Le BE vaut 388 car il ne reste plus qu'une année à considérer. Nous pouvons ensuite analyser la variation du BE :

BE en date d'enregistrement des contrats	765
"Désactualisation"	$765 \times 3\% = 23$
Sinistres prévus la première année	-400
BE à la clôture	$765 + 23 - 400 = 388$

Il faut maintenant réévaluer le montant du RA, si nous supposons qu'il est amorti de façon linéaire, nous obtenons ceci :

RA en date d'enregistrement	150
Amortissement de la période	$-\frac{150}{2} = -75$
RA à la clôture	$150 - 75 = 75$

Nous pouvons ensuite calculer la CSM en $t = 1$:

CSM en date d'enregistrement des contrats	185
"Désactualisation"	$185 \times 3\% = 6$
Amortissement de la période	$-\frac{185+6}{2} = -95.5$
CSM à la clôture	$185 + 6 - 95.5 = 95.5$

Ainsi, nous obtenons le tableau des variations du passif d'assurance :

	BE	RA	CSM	Total
En date d'enregistrement des contrats	765	150	185	1 100
"Désactualisation"	23	/	6	29
Amortissement de la période	/	-75	-95.5	-170.5
Sinistres prévus pour la première année	-400	/	/	-400
A la clôture	388	75	95.5	558.5

Nous pouvons ainsi alimenter le compte de résultat sous IFRS 17 en $t = 1$:

Sinistres prévus	400
Amortissement du RA	75
Amortissement de la CSM	95.5
Produits d'assurance	$400 + 75 + 95.5 = 570.5$
Sinistres effectivement encourus	400
Charges d'assurance	400
Résultat d'assurance	$570.5 - 400 = 170.5$
Charges financières d'assurance (Désactualisation)	29
Résultat net	$170.5 - 29 = 141.5$

4.5 Calcul du résultat IFRS 17

Pour le calcul du résultat IFRS 17 sur le périmètre DCAT, nous avons décidé d'utiliser la méthode Premium Allocation Approach car les contrats étudiés ont une durée d'un an et sont à tacite reconduction. L'objectif de cette sous-partie est de montrer comment nous sommes parvenus à calculer le résultat IFRS 17 pour l'année 2019.

Afin d'obtenir les taux de Risk Adjustment d'ouverture et de clôture présentés dans ce mémoire, nous avons utilisé la méthode de Monte-Carlo. C'est une méthode qui repose sur la loi des grands nombres, elle permet d'estimer une quantité numérique en utilisant la simulation de variables aléatoires. Ainsi, en simulant différentes trajectoires équiprobables, nous pouvons obtenir un estimateur convergent de la moyenne des flux probables.

Grâce aux simulations de Monte-Carlo effectuées, nous avons pu établir la distribution du Best Estimate de sinistres d'ouverture et de clôture et nous avons pu définir une Value-at-Risk à 95%. Ainsi, nous avons été en mesure de déterminer le RA permettant d'arriver à cette VaR et cela nous a donné le taux de RA d'ouverture :

$$\text{Taux de RA d'ouverture} = \frac{\text{RA d'ouverture}}{\text{BE d'ouverture}}$$

Nous avons ensuite obtenu le taux de RA de clôture de la même manière. Le RA dépend donc de la VaR et celle-ci est choisie en fonction de la marge de prudence que l'on souhaite avoir. Sous Solvabilité 2, une VaR à 99,5% est utilisée, cependant, IFRS 17 n'est pas une norme prudentielle donc nous ne sommes pas obligés d'utiliser une VaR aussi prudente. Dès lors, en prenant une VaR à 95%, le compte de résultat IFRS 17 n'est pas trop pénalisé car le RA n'est pas trop élevé, toutefois, nous restons assez prudents. Cette VaR nous a donc semblé être un bon compromis. Les taux de RA d'ouverture et de clôture ont été déterminés au global pour ensuite être appliqués aux BE d'ouverture et de clôture par famille de produits et par HRG afin d'obtenir des RA à cette même maille.

Le calcul du BE S2 évoqué dans la sous-partie 4.2.3 n'a pas uniquement été utile à la comparaison des normes Solvabilité 2 et French GAAP, il a également été important pour le calcul du

résultat IFRS 17. En effet, pour calculer le résultat IFRS 17 2019 avec la méthode Premium Allocation Approach, il est nécessaire d'avoir le BE de sinistres calculé à fin 2018 et le BE de sinistres calculé à fin 2019. Ces deux BE constituent respectivement le BE de sinistres de clôture 2018, i.e., d'ouverture 2019 et le BE de sinistres de clôture 2019. Nous avons besoin de ces deux BE pour calculer le résultat IFRS 17 car il faut calculer une variation de Best Estimate sur 2019 pour arriver à ce résultat, il était donc nécessaire d'avoir en notre possession les BE d'ouverture et de clôture contrairement à ce qui est nécessaire pour calculer le BE total de S2. Pour obtenir le Best Estimate de 2019, il a fallu projeter les hypothèses S2 à fin 2018, les primes et les survenances de sinistres sont projetées du 01/01/2019 au 31/12/2019 et les prestations générées par ces survenances sont projetées plus longtemps selon les lois de maintien en arrêt de travail, par exemple. En effet, une survenance en arrêt de travail peut générer des prestations pendant plusieurs dizaines d'années. Les prestations relatives aux sinistres antérieurs au 01/01/2019 sont également projetées sur les années à venir. Ainsi, le BE utilisé sous Solvabilité 2 est un BE d'ouverture mais total, c'est-à-dire que nous ne regardons pas seulement le BE de sinistres comme c'est le cas sous IFRS 17, mais également le BE de primes.

Ainsi, pour obtenir le résultat IFRS 17, nous avons effectué nos calculs directement à la maille famille de produits, risque, réseau et code société. En effet, nous avons des données comptables par Produit Commercial (PC), ce qui nous a permis de les avoir à la maille souhaitée. Toutefois, comme nous utilisons les BE de S2 qui ne sont disponibles que par HRG et par famille de produits, nous avons dû récupérer les colonnes "Risque", "Code société" et "Réseau" à partir du HRG (voir description de la variable dans la sous-partie 3.1.1).

Nous avons donc commencé par importer les données nécessaires au calcul du résultat IFRS 17, voici les bases dont nous avons eu besoin :

- Données comptables par produit commercial pour l'année 2019.
- Données comptables de réassurance par produit commercial pour l'année 2019.
- Best Estimate de sinistres bruts et cédés par famille de produits et par HRG. Il s'agit des BE de sinistres actualisés mais également non actualisés. Dans nos calculs, lorsque cela n'est pas précisé, cela signifie que les Best Estimate ont été actualisés au taux d'ouverture ou de clôture selon qu'il s'agisse du BE d'ouverture ou de clôture. En effet, lorsqu'ils n'ont pas été actualisés au taux d'ouverture ou de clôture, nous le préciserons.
- Base contenant les frais généraux totaux par produit commercial pour l'année 2019.

Afin d'obtenir un tableau récapitulatif des BE de sinistres bruts et cédés par famille de produits et par HRG, nous avons dû agréger les données que nous possédions pour chaque famille de produits. En effet, le modèle de la Direction des Risques permettant la projection sous Solvabilité 2 ne nous renvoie pas cela directement. Ce modèle projette une base de données et renvoie les flux agrégés par pas de temps et HRG. Ainsi, pour obtenir les BE par famille de produits et HRG, il faut créer autant de bases que de familles de produits et les projeter à l'aide du modèle développé par la DR. Une fois les projections obtenues, il a donc fallu calculer les BE pour chacune des familles avec les formules données dans la sous-partie 4.2.3 et, par la suite, mettre toutes les données dans un seul fichier afin de simplifier les jointures entre les BE et les données comptables.

Tout d'abord, pour les données comptables brutes et cédées, nous n'avions pas toutes les variables nécessaires mais avons pu les récupérer en joignant nos données, par PC, à des tables de correspondance. Nous avons ainsi pu récupérer la famille de produits, le risque et le réseau. Après cela, nous avons agrégé nos données par société, réseau, risque et famille de produits car les codes de PC ne nous étaient plus utiles étant donné que nous ne les possédions pas dans les données de BE.

Pour déterminer le résultat brut de réassurance, nous avons eu besoin des BE de sinistres bruts, des données comptables et de la base contenant les frais généraux totaux. Puis, pour le calcul du résultat de réassurance, nous avons eu besoin des BE de sinistres cédés et des données comptables de réassurance.

4.5.1 Ajout des postes nécessaires au calcul du résultat brut de réassurance IFRS 17

Pour calculer le résultat brut de réassurance IFRS 17, nous avons donc utilisé les données comptables French GAAP ainsi que les BE de sinistres S2 déterminés à fin 2018 et à fin 2019,

correspondant respectivement au BE de sinistres d'ouverture 2019 et de clôture 2019. La première chose que nous avons dû faire était de calculer une variation de Best Estimate de sinistres en soustrayant le BE de sinistres d'ouverture au BE de sinistres de clôture.

Nous avons commencé par joindre nos données avec les taux de Risk Adjustment que nous avons déterminés à partir de la méthode de Monte-Carlo :

- Taux de Risk Adjustment d'ouverture
- Taux de Risk Adjustment de clôture

Voici ce que nous avons obtenu :

- $Risk\ Adjustment\ d'ouverture = Best\ Estimate\ d'ouverture \times Taux\ de\ Risk\ Adjustment\ d'ouverture$
- $Risk\ Adjustment\ de\ clôture = Best\ Estimate\ de\ clôture \times Taux\ de\ Risk\ Adjustment\ de\ clôture$
- $Variation\ de\ Risk\ Adjustment = Risk\ Adjustment\ de\ clôture - Risk\ Adjustment\ d'ouverture$

Nous avons ensuite calculé la variation de provision totale en calculant préalablement les provisions totales d'ouverture et de clôture. En effet, ce calcul était nécessaire afin de déterminer le solde d'écart entre les normes French GAAP et IFRS 17. Les provisions d'ouverture (respectivement de clôture) comprennent notamment les provisions de sinistres d'ouverture (respectivement de clôture) ainsi que la provision de primes pures et commerciales d'ouverture (respectivement de clôture). Le lecteur intéressé par la liste totale des provisions pourra se référer à l'annexe A. Ainsi, nous avons pu déterminer la variation de provision totale en calculant la différence entre les provisions totales de clôture et d'ouverture.

Afin de calculer la Provision pour Frais de Gestion de Sinistres (PFGS) d'ouverture et de clôture, nous avons joint, à nos données, les taux suivants, par code société, réseau et risque :

- Taux de PFGS d'ouverture
- Taux de PFGS de clôture

Cela nous a permis de réaliser les calculs suivants :

- $PFGS\ d'ouverture = (Provisions\ de\ sinistres\ d'ouverture + Provision\ mathématiques\ de\ rentes\ d'ouverture) \times Taux\ de\ PFGS\ d'ouverture$
- $PFGS\ de\ clôture = (Provisions\ de\ sinistres\ de\ clôture + Provision\ mathématiques\ de\ rentes\ de\ clôture) \times Taux\ de\ PFGS\ de\ clôture$
- $Variation\ de\ PFGS = PFGS\ de\ clôture - PFGS\ d'ouverture$

À partir de cela, nous avons pu calculer la variation de provision French GAAP de la façon suivante :

- $Provision\ French\ GAAP\ d'ouverture = Provisions\ d'ouverture + PFGS\ d'ouverture$
- $Provision\ French\ GAAP\ de\ clôture = Provisions\ de\ clôture + PFGS\ de\ clôture$
- $Variation\ de\ provision\ French\ GAAP = Provision\ French\ GAAP\ de\ clôture - Provision\ French\ GAAP\ d'ouverture$

Par ailleurs, trois autres variations ont été calculées à partir des variables présentes dans les données comptables :

- $Variation\ de\ provision\ pour\ égalisation = Provision\ pour\ égalisation\ de\ fin - Provision\ pour\ égalisation\ de\ début$
- $Variation\ du\ fonds\ de\ revalorisation = Fonds\ de\ revalorisation\ de\ fin - Fonds\ de\ revalorisation\ de\ début$
- $Variation\ de\ provision\ de\ primes\ totale = (Provision\ de\ primes\ pures\ de\ clôture + Provision\ de\ primes\ commerciales\ de\ clôture) - (Provision\ de\ primes\ pures\ d'ouverture + Provision\ de\ primes\ commerciales\ d'ouverture)$

Enfin, pour calculer le solde d'écart entre les normes French GAAP et IFRS 17, nous avons utilisé les précédents calculs réalisés :

$$\begin{aligned} \text{Solde d'écart entre les normes} = & \text{Variation de provision French GAAP} - \text{Variation de Best} \\ & \text{Estimate de sinistres} - \text{Variation de Risk Adjustment} + \text{Variation de provision pour égalisation} \\ & + \text{Variation du fonds de revalorisation} + \text{Variation de provision de primes totale.} \end{aligned}$$

Nous étions donc en mesure de calculer la CSR différée grâce à la formule suivante :

$$CSR_{diff} = \begin{cases} \text{Solde écart normes} \times \text{Taux}_{com \text{ sur rés}} & \text{si Solde écart normes} > 0 \\ 0 & \text{sinon.} \end{cases}$$

En réalité, le taux de commission sur résultat ($\text{Taux}_{com \text{ sur rés}}$) correspond au taux de commission sur résultat déduit, celui-ci étant déterminé en divisant la commission sur résultat par le solde technique, nous avons donc utilisé les pourcentages par risque disponibles dans la maquette IFRS 17 réalisée par la tribu IFRS 17. Par ailleurs, le taux vaut en moyenne 50% ce qui est le taux de commission sur résultat utilisé au global sous French GAAP.

Nous avons ensuite dû calculer l'Unwinding et l'Insurance Revenue Update afin de les éliminer du BE de clôture :

- *Unwinding* = *Best Estimate de sinistres d'ouverture non actualisé* × *Courbe des taux à l'ouverture*.
- *Insurance Revenue Update* : Il s'agit de la différence entre les flux d'ouverture actualisés avec la courbe des taux à l'ouverture et les flux d'ouverture actualisés avec la courbe des taux à la clôture.

Le taux i en t de la courbe des taux vaut : $(\text{Facteur d'actualisation}_t)^{-\frac{1}{t}} - 1$, t représente le temps en mois.

$$\text{Facteur d'actualisation} = \frac{1}{(1+i)^t}.$$

Si $t = 1$, cela veut dire qu'il s'agit du taux de janvier de l'année N. Dans notre cas, 0 représente donc le mois de décembre de l'année N-1 (moment auquel nous projetons sur les années à venir).

La CSR différée, l'Unwinding et l'Insurance Revenue Update vont venir impacter le BE de sinistres de clôture. Ainsi, nous avons calculé le BE de sinistres de clôture retraité et le RA de clôture retraité :

- *Best Estimate de sinistres de clôture retraité* = *Best Estimate de sinistres de clôture* + *Commission sur résultat différée* - *Unwinding* - *Insurance Revenue Update*
- *Variation de Best Estimate de sinistres retraité* = *Best Estimate de sinistres de clôture retraité* - *Best Estimate de sinistres d'ouverture*
- *Risk Adjustment de clôture retraité* = $(\text{Best Estimate de sinistres de clôture retraité} - \text{Commission sur résultat différée}) \times \text{Taux de Risk Adjustment de clôture}$
- *Variation de Risk Adjustment retraité* = *Risk Adjustment de clôture retraité* - *Risk Adjustment d'ouverture*

Nous avons ensuite pu calculer le Liability for Remaining Coverage d'ouverture et de clôture et la variation de ce passif :

- *Liability for Remaining Coverage d'ouverture* = *Provision de primes pures d'ouverture* + *Provision de primes commerciales d'ouverture*
- *Liability for Remaining Coverage de clôture* = *Provision de primes pures de clôture* + *Provision de primes commerciales de clôture*
- *Variation de Liability for Remaining Coverage* = *Liability for Remaining Coverage de clôture* - *Liability for Remaining Coverage d'ouverture*

Les frais généraux se décomposent en :

- Frais d'acquisition
- Frais d'administration
- Frais de prestations

- Frais de placements
- Autres charges techniques

Les quatre premiers types de frais sont des frais dits « rattachables » tandis que les autres charges techniques sont des frais « non rattachables ». Ainsi, en moyenne 75% des frais sont rattachables et 25% ne le sont pas, voici les taux de rattachement des frais par risque :

- AT BP & CE : 77, 70%
- DC ACC BP : 64, 34%
- DC ACC CE : 81, 14%
- DC TC BP : 76, 11%
- DC TC CE : 77, 98%

Grâce aux variables que nous avons ajoutées, nous avons pu calculer les différents éléments du compte de résultat brut de réassurance sous IFRS 17 :

- $Produits = Primes HT$
- $Charges = Prestations + Variation de Best Estimate de sinistres retraité + Variation de Risk Adjustment retraité + Variation du Liability for Remaining Coverage + Commission sur résultat + Frais rattachables$
- $Résultat d'assurance = Produits - Commissions d'acquisition - Charges$
- $Autres charges techniques = Frais non rattachables$
- $Résultat technique = Résultat d'assurance - Autres charges techniques$
- $Charges financières = Unwinding + Insurance Revenue Update$
- $Résultat financier = - Charges financières$
- $Profit and Loss = Résultat technique + Résultat financier$

4.5.2 Ajout des postes nécessaires au calcul du résultat de réassurance IFRS 17

Les variables permettant d'obtenir le résultat de réassurance IFRS 17 ont été calculées de manière analogue à ce qui a été fait pour le résultat brut de réassurance IFRS 17.

Ainsi, nous avons pu calculer les éléments principaux du compte de résultat de réassurance IFRS 17 :

- $Produits cédés = Primes pures cédées$
- $Commissions de cession = Commissions d'acquisition cédées + Frais de gestion cédés$
- $Charges cédées = Prestations cédées + Variation de Best Estimate de sinistres cédé retraité + Variation de Risk Adjustment cédé retraité + Variation du Liability for Remaining Coverage cédé + Commission sur résultat cédé$
- $Résultat de réassurance = Produits cédés - Commissions de cession - Charges cédées$
- $Résultat technique de réassurance = Résultat de réassurance$
- $Résultat financier de réassurance = - Charges financières de réassurance$
- $Profit and Loss de réassurance = Résultat technique de réassurance + Résultat financier de réassurance$

Nous pouvons remarquer qu'il y a quelques différences entre les deux comptes de résultat notamment au niveau des commissions de cession car elles comprennent les frais de gestion qui n'étaient pas comptabilisés pour le résultat brut de réassurance. Par ailleurs, pour le calcul des charges cédées, les frais rattachables n'ont pas été pris en compte alors que nous les avons comptabilisés dans le calcul des charges du résultat brut de réassurance. En effet, les frais rattachables représentent une partie des frais généraux et ces frais ne sont pas comptabilisés dans le résultat de réassurance.

Il a ensuite été possible de calculer le résultat IFRS 17 de la façon suivante :

$$Résultat IFRS 17 = (Profit and Loss - Profit and Loss de réassurance)$$

Par ailleurs, le résultat IFRS 17 peut être assimilé au Résultat Brut d'Exploitation.

À partir de nos calculs, nous avons pu obtenir le tableau suivant :

IFRS17 brut de réassurance		IFRS17 Réassurance	
(+) Produits	115 997 371	(+) Produits	29 389 133
Primes	115 997 371	Primes	29 389 133
(-) Frais d'acquisition	16 823 927	(-) Frais d'acquisition	7 707 828
Commissions sur cotisations	16 823 927	Commission de cession	7 707 828
(-) Charges	69 663 370	(-) Charges	20 219 086
Sinistres réglés	20 112 011	Sinistres réglés (y.c frais de gestion)	4 153 577
Variation BE (hors effet actu)	11 217 999	Variations du BE (hors effet actu)	4 769 810
dont BE ouverture	- 52 974 526	dont BE ouverture	- 11 818 874
dont BE Sinistre		dont BE clôture (hors effet actu)	16 588 684
dont BE Frais		dont BE Frais	
dont BE CSR		dont CSR différée IFRS 17	
dont BE clôture (hors effet actu)	64 192 525	dont PPNA clôture	
dont BE Sinistre		Variation RA	- 616 265
dont BE Frais		dont RA ouverture	- 1 382 784
dont BE CSR		dont RA clôture	766 519
dont CSR différée IFRS 17		Commission sur résultat FG	11 911 964
élimination Unwinding & IR Update		Frais rattachables	-
Variation RA	- 1 104 041	Résultat de réassurance	1 462 218
dont RA ouverture	- 6 185 436	(-) Autres charges techniques	-
dont RA clôture	5 081 394	Frais non rattachables	-
Variation LRC	43 315	Résultat technique	1 462 218
dont LRC ouverture	- 1 454 980	(+) Produits financiers	
dont LRC clôture	1 498 295	(-) Charges financiers	34 347
Commission sur résultat French GAAP	26 566 542	Résultat financier	- 34 347
Frais rattachables	12 827 545	Profit & Loss	1 427 872
Résultat d'assurance	29 510 074	(+) Delta OCI Assets	
(-) Autres charges techniques	4 323 171	(-) Delta OCI liabilities	-
Frais non rattachables	4 323 171	Total comprehensive income	1 427 872
Résultat technique	25 186 904		
(+) Produits financiers			
(-) Charges financiers	129 678		
Résultat financier	- 129 678		
Profit & Loss	25 057 226		
(+) Delta OCI Assets			
(-) Delta OCI liabilities	-		
Total comprehensive income	25 057 226		

FIGURE 20 – Résultat DCAT IFRS 17

Ainsi, pour l'année 2019, le résultat IFRS global était de :

$$25\,057\,226 - 1\,427\,872 = 23\,629\,354.$$

4.6 Comparaison IFRS 17 et Solvabilité 2

Tout d'abord, IFRS 17 et Solvabilité 2 ont des objectifs différents. D'une part, Solvabilité 2 est une norme prudentielle qui a pour objectifs la gestion des risques et la prévention des situations d'insolvabilité. D'autre part, IFRS 17 est une norme comptable qui a un objectif de valorisation à destination des analystes financiers dans une optique liquidative.

Malgré ces objectifs différents, ces normes ont certains points communs :

- Elles évaluent le passif d'assurance en se basant sur une approche prospective des flux de trésorerie en juste valeur.
- Elles procèdent à l'évaluation des engagements en espérance (Best Estimate).

- Elles sont utilisées au niveau européen.
- Elles ont un objectif de comparabilité.
- Elles sont toutes les deux difficiles à mettre en oeuvre.
- Elles prennent en compte un ajustement pour risque.
- Elles valorisent les options et garanties.

Par les ressemblances entre ces deux normes, il est possible de capitaliser sur Solvabilité 2.

Il y a des similitudes dans la mesure des passifs, ceux-ci sont calculés comme la somme d'une valeur actuelle probable des flux de trésorerie futurs et d'une dotation pour risque. Il est donc possible de capitaliser sur :

- La modélisation des flux de trésorerie. En effet, nous pouvons regarder les flux pris en compte (cependant, il y a quelques écarts sur certains frais généraux, par exemple) et le taux d'actualisation (néanmoins, il y a potentiellement plus de courbes à utiliser sous IFRS 17, ce qui va donc nécessiter des capacités de calcul plus importantes).
- La méthode CoC. Cependant, ce n'est pas forcément la meilleure approche pour afficher un niveau de confiance sous IFRS 17.

Concernant les informations à fournir et les reporting, il est possible de capitaliser sur :

- Les processus de production mis en oeuvre sous S2 pour le QRT et pour les rapports narratifs, en particulier, sur les analyses produites dans le cadre de ces rapports.
- L'analyse des variations pour produire les réconciliations.
- La réduction des délais entreprise dans le cadre de S2 (fast close).

Il est également possible de capitaliser sur la qualité des données (architecture, historisation, stockage, piste d'audit et processus de contrôles).

Nous pouvons donc remarquer que les principes de ces deux normes sont proches. Cependant, à cause de leurs différences, il est nécessaire d'adapter en profondeur les modèles actuariels et leur alimentation.

Voici quelques différences entre les deux normes :

- IFRS 17 fait des regroupements par groupe de contrats tandis que S2 regroupe par type de risque.
- IFRS 17 invente une nouvelle définition du résultat d'assurance alors que S2 ne cherche pas à mesurer la rentabilité.
- La granularité sous IFRS 17 est plus fine (portefeuilles \times génération \times profitabilité).
- Les taux d'actualisation de S2 s'éloignent des taux de marché.
- IFRS 17 ne définit pas précisément les frais rattachables au contrat.
- Le passif IFRS 17 introduit un nouvel élément : la CSM.
- Sous S2, toutes les composantes d'un contrat sont traitées avec S2, ce n'est pas le cas sous IFRS 17, comme nous l'avons vu, certaines composantes sont, par exemple, traitées sous IFRS 9.
- Sous S2, les contrats sont comptabilisés à la date d'engagement, tandis que, sous IFRS 17, nous avons vu que nous les comptabilisons à la date la plus ancienne parmi le début de la couverture, la réception de la première prime et la date à laquelle le contrat devient onéreux. La date de reconnaissance des contrats peut donc différer sous IFRS 17 et sous S2.

		Solvabilité 2	IFRS 17
Zone géographique concernée		Europe	Monde
Objectifs		<u>Cadre prudentiel</u> : Limiter la probabilité de ruine de l'assureur à 0,5% et garantir aux assurés que l'assureur honorera bien ses engagements (SCR/MCR)	<u>Cadre comptable</u> : Harmonisation de l'information financière
Comptabilisation des contrats		Comptabilisation à la signature du contrat	Comptabilisation au début de la couverture
Granularité		Line of Business x Homogeneous Risk Group	Portefeuilles x génération x profitabilité
Provisions techniques		Best Estimate + Risk Margin	Best Estimate + Risk Adjustment + Contractual Service Margin
Frontière des contrats		Projection des primes futures jusqu'à la date à partir de laquelle l'assureur a la possibilité unilatérale de résilier	Projection des primes futures dès lors que l'entité d'assurance peut exiger le paiement de la prime ou qu'elle est dans l'obligation de délivrer le service
Reconnaissance des profits futurs		Reconnus directement dans les capitaux propres	Contractual Service Margin amortie à chaque période
Courbe des taux	Calcul	Courbe des taux sans risque fournie par l'EIOPA	Calculée par l'assureur à l'aide de l'approche Top-Down ou Bottom-Up
	Traitement des écarts	Impacts des variations sur les fonds propres	Impacts des variations sur les fonds propres ou en résultat si l'assureur décide d'activer l'option OCI
Reporting		QRT et rapports narratifs	États financiers

TABLE 14 – Solvabilité 2 VS IFRS 17

Dans le cadre de ce mémoire, nous avons pu comparer la norme Solvabilité 2 avec la norme IFRS 17 grâce au ROE de ces deux normes par famille de produits.

Le Return On Equity (ROE) est calculé selon la définition suivante :

$$ROE = \frac{\text{Résultat Net}}{SCR \times \text{Exigence de solvabilité}}$$

En effet, la formule originale du ROE est la suivante :

$$ROE = \frac{\text{Résultat Net}}{\text{Fonds propres}}$$

Solvabilité 2 exige que les entreprises aient une marge de couverture des fonds propres, appelée exigence de solvabilité, et calculée de la façon suivante :

$$\text{Exigence de solvabilité} = \frac{\text{Fonds propres}}{SCR}$$

Par ailleurs, ce ratio doit être supérieur à 100%. Ainsi, nous obtenons l'égalité suivante : $\text{Fonds propres} = SCR \times \text{Exigence de solvabilité}$ et la formule définie ci-dessus est vérifiée.

Pour le calcul des SCR, il faut projeter chaque base autant de fois qu'il y a de chocs standards à appliquer. Pour chaque choc, nous obtenons un BE choqué qui est retranché au BE central pour obtenir le SCR sous-modulaire.

Dans le département Actuariat Produits Prévoyance, le processus de calcul des SCR est organisé de la façon suivante :

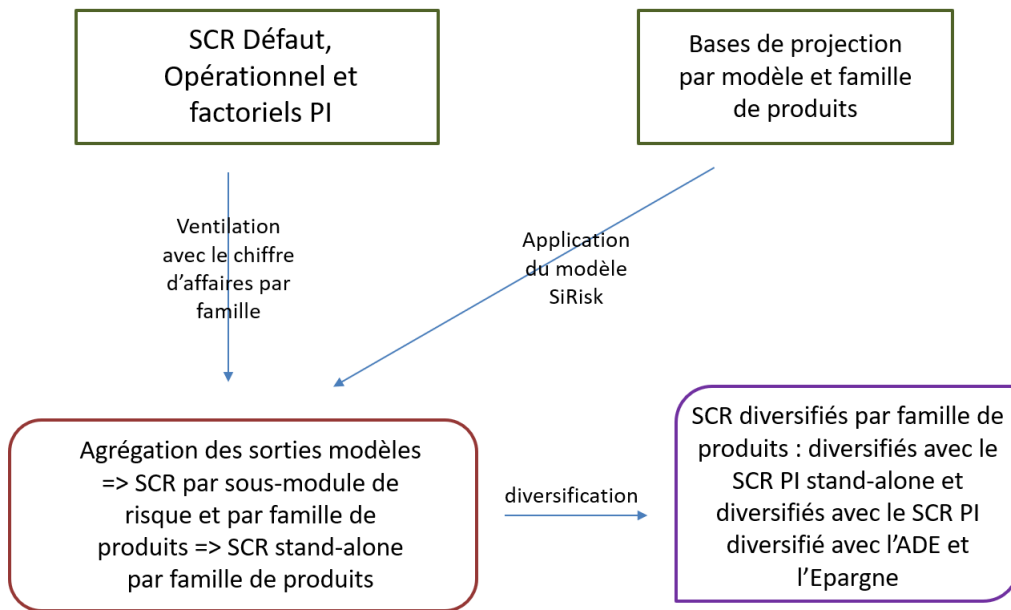


FIGURE 21 – Processus de calcul des SCR en prévoyance individuelle

Sur le schéma ci-dessus, nous pouvons observer le processus de calcul des SCR en Prévoyance Individuelle en partant des bases de projection par modèle et famille de produits pour arriver aux SCR diversifiés par famille de produits. Le SCR obtenu pour la PI est un SCR diversifié avec l'assurance des emprunteurs et l'Épargne. En effet, nous commençons par calculer trois SCR stand-alone : un pour la PI, un pour l'Épargne et un pour l'ADE. Nous appliquons ensuite des matrices de corrélation, ce qui nous permet d'avoir un gain de diversification par activité. Les trois SCR sont ensuite agrégés pour former un SCR global pour les trois activités. Cependant, grâce au gain de diversification, nous obtenons l'inégalité suivante :

$$SCR_{global} < SCR_{PI} + SCR_{Épargne} + SCR_{ADE}$$

Voici un schéma représentant le gain de diversification :

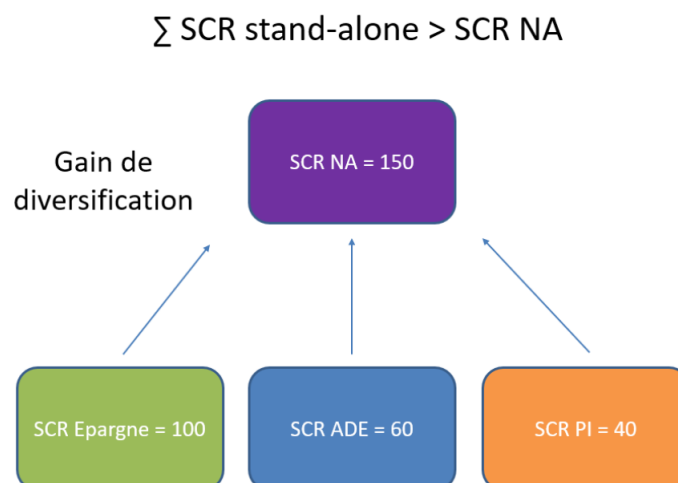


FIGURE 22 – Gain de diversification

Or, le SCR global doit être égal à la somme des SCR des trois activités, il est donc nécessaire de réduire les trois SCR stand-alone par un mécanisme de réallocation afin que leur somme soit égale au SCR global, cela nous permet d'obtenir des SCR diversifiés. Dès lors, le SCR de la PI est diminué donc le besoin en fonds propres diminue également. Le modèle SiRisk est le modèle de projection des flux S2 de la DR. Pour le calcul des ROE, c'est le SCR stand-alone qui a été utilisé

car il est plus prudent que le SCR diversifié.

Les ROE ont été calculés pour une exigence de solvabilité de 110%. Le dénominateur du ROE est commun aux normes Solvabilité 2 et IFRS 17, seul le numérateur change. En effet, le résultat net S2 est le résultat net projeté S2. Il s'agit du Best Estimate S2 que nous avons transformé en RBE à partir de la formule vue dans la sous-partie 4.2.4, puis, nous avons obtenu le résultat net en retranchant l'impôt sur les sociétés. Le résultat net IFRS 17 correspond au résultat IFRS 17 diminué de l'impôt sur les sociétés. Notre comparaison des ROE revient donc à une comparaison des résultats nets car le dénominateur est commun. La simple comparaison du résultat net des deux normes aurait été suffisante mais calculer un ROE nous permet d'avoir un indicateur de rentabilité par famille de produits pour ces deux normes. Voici les ROE que nous avons obtenus en utilisant le SCR stand-alone par famille de produits :

Famille de produits	ROE S2	ROE IFRS 17	Ecart ROE (en points)	Poids du CA (2019)
FAMILLE FRUCTI-FAMILLE	32,41 %	50,36 %	17,95 %	41,0%
SECUR FAMILLE CE DC PTIA	16,56 %	34,33 %	17,77 %	19,9%
FRUCTI-HOMME CLE	38,92 %	52,00 %	13,08 %	18,4%
FRUCTI-PROFESSIONNEL	64,75 %	96,70 %	31,95 %	10,1%
FRUCTI-PATRIMOINE	65,92 %	99,50 %	33,58 %	3,8%
SECUR URGENCE CE	12,98 %	20,09 %	7,10 %	2,5%
FAMILLE FRUCTI-BUDGET	73,06 %	184,76 %	111,70 %	0,9%
DEDICACE	84,56 %	242,08 %	157,52 %	0,9%
D.S.EPARGNE SALARIALE	-233,14 %	-222,37 %	10,77 %	0,6%
ASSURANCE HOMME-CLE C.E.	150,83 %	37,80 %	-113,03 %	0,3%
FRUCTI-FACILITES AGRI	-115,20 %	-90,29 %	24,91 %	0,3%
CREDIT MARITIME VIE	40,15 %	57,16 %	17,01 %	0,2%
FAMILLE FRUCTI-FACILITES	14,28 %	36,28 %	22,00 %	0,1%
CERTI CAPITAL	99,43 %	253,86 %	154,43 %	0,1%
FRUCTI-FACILITES PRO	-28,44 %	-44,23 %	-15,80 %	0,1%
SECURI-CREDIT BPROP	18,93 %	45,75 %	26,83 %	0,1%
FAMILLE DOUBLE SECURITE	9,51 %	336,36 %	326,85 %	0,0%
FAMILLE FRUCTI-SECURITE	-895,00 %	305,58 %	1200,59 %	0,0%

TABLE 15 – Comparaison ROE Solvabilité 2 et IFRS 17

Sur ce tableau, nous pouvons remarquer que les ROE des deux normes sont parfois très proches et parfois plutôt éloignés. De manière générale, pour les principales familles de produits, les ROE sont assez proches alors que, pour les petites familles de produits, ils sont beaucoup plus éloignés à cause des faibles volumes de données.

4.7 Comparaison IFRS 17 et French GAAP

La comparaison entre les normes French GAAP (FG) et IFRS 17 est très importante car elle permet de voir si les hypothèses de projection utilisées pour le calcul des BE de sinistres S2 permettent d'être proche du réel. Afin de comparer la norme IFRS 17 à French GAAP, nous avons réalisé une analyse de passage.

En analysant les postes des comptes French GAAP et IFRS 17, nous avons observé une différence sur les éléments suivants :

- La variation de provisions de sinistres : Sous French GAAP, elle est déterminée à l'aide du provisionnement French GAAP tandis que, sous IFRS 17, il s'agit d'une estimation des flux de sinistralité futurs utilisant des hypothèses de projection.
- Le compte de réassurance : Sous French GAAP, il est déterminé dans un outil prévu à cet effet tandis que, sous IFRS 17, il s'agit d'une estimation des flux de sinistres futurs cédés avec des taux de cession définis à l'aide de proxys car la maille modèle est différente de la maille des traités de réassurance.

Voici ce que nous avons obtenu par famille de produits :

Famille de produits	Variation BE+RA IFRS 17 (k€)	Variation de provision de sinistres FG (k€)	Ecart de variation de provision de sinistres	Résultat réassurance IFRS 17 (k€)	Solde de réassurance FG (k€)	Ecart des comptes de réassurance	Poids du CA
FAMILLE FRUCTI-FAMILLE	5 586	6 481	14 %	353	1 527	77 %	41,0%
SECUR FAMILLE CE DC PTIA	3 676	3 211	14 %	150	911	83 %	19,9%
FRUCTI-HOMME CLE	1 862	2 336	20 %	322	655	51 %	18,4%
FRUCTI-PROFESSIONNEL	-452	-428	6 %	631	505	25 %	10,1%
FRUCTI-PATRIMOINE	-409	-144	185 %	-23	142	116 %	3,8%
SECUR URGENCE CE	458	226	103 %	-4	144	102 %	2,5%
FAMILLE ASSURANCE FAMILLE	202	127	59 %	3	43	92 %	0,9%
FAMILLE FRUCTI-BUDGET	-730	-161	354 %	54	-2	2 655 %	0,9%
DEDICACE	-526	-32	1 550 %	10	-2	637 %	0,9%
D.S.EPARGNE SALARIALE	107	53	104 %	-62	21	399 %	0,6%
ASSURANCE HOMME-CLE C.E.	48	6	742 %	94	14	554 %	0,3%
FRUCTI-FACILITES AGRI	249	231	7 %	-46	10	567 %	0,3%
CREDIT MARITIME VIE	40	46	14 %	3	10	68 %	0,2%
FAMILLE FRUCTI-FACILITES	-15	-12	29 %	1	5	74 %	0,1%
CERTI CAPITAL	-27	15	282 %	6	0	3 282 %	0,1%
FRUCTI-FACILITES PRO	6	-5	221 %	1	3	67 %	0,1%
SECURI-CREDIT BPROP	-30	-19	53 %	3	3	9 %	0,1%
FAMILLE DOUBLE SECURITE	-5	17	128 %	-14	0	17 185 %	0,0%
FAMILLE FRUCTI-SECURITE	7	15	51 %	-10	0	100 883 %	0,0%

TABLE 16 – Analyse de passage de French GAAP à IFRS 17 par famille de produits

Les familles de produits les plus importantes ont un écart de variation de provisions de sinistres globalement moins élevé que les autres. Il en va de même pour l'écart des comptes de réassurance malgré quelques petites familles de produits ayant un faible écart telles que Securi-Credit Bprop.

Afin de pouvoir comparer le résultat net French GAAP et IFRS 17, nous l'avons calculé, pour les deux normes, à la maille famille de produits. Le résultat net est égal au Résultat Brut d'Exploitation auquel nous retirons l'impôt sur les sociétés. Voici la comparaison que nous avons obtenue entre ces deux résultats nets :

Famille de produits	Résultat Net FG (k€)	Résultat Net IFRS 17 (k€)	Ecart Résultat Net	Poids Résultat Net FG	Poids Résultat Net IFRS 17
FAMILLE FRUCTI-FAMILLE	4 044	6 285	55 %	39,7 %	38,1 %
FRUCTI-HOMME CLE	2 055	2 746	34 %	20,2 %	16,6 %
FRUCTI-PROFESSIONNEL	1 652	2 468	49 %	16,2 %	15,0 %
SECUR FAMILLE CE DC PTIA	1 270	2 633	107 %	12,5 %	16,0 %
FRUCTI-PATRIMOINE	669	1 009	51 %	6,6 %	6,1 %
FAMILLE FRUCTI-BUDGET	232	586	153 %	2,3 %	3,6 %
FAMILLE ASSURANCE FAMILLE	193	220	14 %	1,9 %	1,3 %
DEDICACE	179	511	186 %	1,8 %	3,1 %
SECUR URGENCE CE	165	255	55 %	1,6 %	1,5 %
ASSURANCE HOMME-CLE C.E.	74	19	75 %	0,7 %	0,1 %
CREDIT MARITIME VIE	21	29	42 %	0,2 %	0,2 %
CERTI CAPITAL	16	40	155 %	0,2 %	0,2 %
SECURI-CREDIT BPROP	8	19	142 %	0,1 %	0,1 %
FAMILLE FRUCTI-FACILITES	7	18	154 %	0,1 %	0,1 %
FAMILLE DOUBLE SECURITE	1	26	3 437 %	0,0 %	0,2 %
FAMILLE FRUCTI-SECURITE	-9	3	134 %	-0,1 %	0,0 %
FRUCTI-FACILITES PRO	-9	-15	56 %	-0,1 %	-0,1 %
FRUCTI-FACILITES AGRI	-81	-63	22 %	-0,8 %	-0,4 %
D.S.EPARGNE SALARIALE	-305	-291	5 %	-3,0 %	-1,8 %

TABLE 17 – Comparaison des résultats nets French GAAP et IFRS 17

L'écart entre les résultats nets est proche de 50% pour beaucoup de familles de produits. Néanmoins, il est, par exemple, très faible pour D.S. Épargne Salariale et très élevé pour Famille Double Sécurité. Il n'en reste pas moins que l'ordre d'importance au sens du résultat net est presque le même pour French GAAP et IFRS 17.

Nous avons donc pu réaliser l'analyse de passage à travers la variation de provision de sinistres, le compte de réassurance et le résultat net. Voici le graphique de l'analyse de passage réalisée :

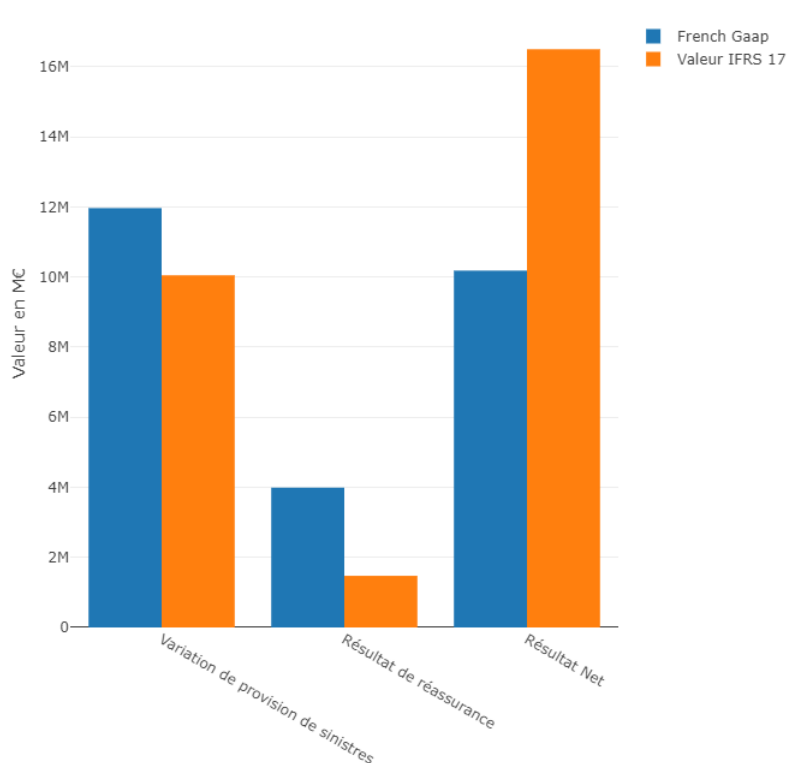


FIGURE 23 – Analyse de passage French GAAP IFRS 17

Sur cette analyse, nous pouvons voir que la variation de provision de sinistres est plus élevée sous French GAAP que sous IFRS 17 et il en va de même pour le compte de réassurance. Toutefois, au niveau du résultat net, c'est celui d'IFRS 17 qui est le plus élevé. Cela s'explique par le fait que la variation de provision de sinistres et le compte de réassurance viennent en diminution du résultat net. Dès lors, étant donné que ces deux éléments sont plus faibles sous IFRS 17, ils diminuent moins le résultat net qui en devient plus élevé que sous French GAAP. Au vu des écarts du compte de réassurance, nous pouvons nous questionner sur la modélisation des flux de réassurance effectuée sous Solvabilité 2.

Cette partie nous aura donc permis de comprendre les enjeux d'IFRS 17 ainsi que l'importance de la comparaison des projections avec le réel. En effet, grâce à ces comparaisons, nous avons pu remarquer un écart entre les comptes de réassurance French GAAP et IFRS 17 dû, en partie, à la modélisation des flux de réassurance sous Solvabilité 2. Dans la partie suivante, nous serons amenés à comparer le BE total de Solvabilité 2 avec le Produit Net Assurance réel et prédit afin de voir si les prédictions de Machine Learning permettent d'être plus proche de la réalité que les projections S2.

5 Machine Learning

L'objectif de cette partie est de montrer l'utilité des méthodes de Machine Learning pour prédire un Produit Net Assurance par tête et de présenter les différents modèles que nous avons construits. Par ailleurs, cette partie vise également à backtester les résultats des projections S2 grâce au modèle de Machine Learning que nous avons retenu.

Afin d'obtenir les résultats des projections S2 par famille de produits, il est nécessaire d'utiliser le modèle développé au sein de la Direction des Risques en modifiant le paramétrage : il s'agit d'effectuer un paramétrage pour chaque famille de produits et de lancer le nombre de scénarios en conséquence. La procédure est longue car elle nécessite un gros travail sur les données en terme de préparation et de contrôle. En effet, sous Solvabilité 2, il faut préparer tous les inputs modèle, puis, tout projeter et calculer les résultats de tous les scénarios paramétrés. De plus, quelques étapes sont manuelles et sont donc soumises à l'erreur.

Dès lors, nous avons décidé de comparer les résultats obtenus par les projections S2 avec des prédictions du Produit Net Assurance réalisées par des méthodes de Machine Learning, mais qu'est ce que le Machine Learning ? Nous pourrions penser qu'il s'agit d'une invention récente, et pourtant, ce n'est pas le cas. En effet, dès 1959, Arthur Samuel, l'un des pionniers de l'intelligence artificielle a défini le Machine Learning comme le "field of study that gives computers the ability to learn without being explicitly programmed", c'est-à-dire, comme le champ d'étude visant à donner la capacité à une machine d'apprendre sans explicitement être programmée. Après cette première définition, Tom Michell a proposé, en 1997, une définition plus précise qui est la suivante : "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". Afin d'illustrer cette définition, Tom Mitchell a utilisé quelques exemples, nous avons retenu celui-ci concernant la reconnaissance d'écriture manuscrite. Dans ce cas là, nous avons :

- Tâche T (Task) : Reconnaître et classer des mots manuscrits dans des images.
- Mesure de performance P (Performance measure) : Déterminer le pourcentage de mots correctement classés.
- Apprentissage par l'expérience E (training Experience) : Apprendre grâce à une base de données contenant des mots manuscrits déjà classés.

C'est grâce à la période d'apprentissage qu'un modèle de Machine Learning va déduire certains critères de classification. Par exemple, il va remarquer que certaines personnes écrivent plutôt en italique, avec des grandes lettres ou en détaché.

Ainsi, en utilisant des méthodes de Machine Learning et un script plus automatique permettant de calculer le PNA par tête sur notre base de données, nous sommes capables d'enrichir la base de données pour les modèles de Machine Learning des années suivantes et de mettre à jour le modèle. Cela permet alors :

- De réduire le temps de calcul de manière conséquente.
- D'obtenir une estimation des Best Estimate S2 pour contrôler de manière macro les résultats réglementaires.
- D'obtenir un résultat tête par tête permettant d'agréger les flux aux mailles souhaitées. Ainsi, nous pouvons agréger les proxys obtenus par famille de produits car cela est important pour la vision métier et le pilotage des offres de Natixis Assurances. A contrario, cela est impossible sous S2 car, pour obtenir les résultats tête par tête, il faut projeter chaque tête une à une ce qui fait autant de paramétrages que de nombre de têtes et c'est impossible. Le paramétrage est très lourd pour chaque base de données donc il n'est pas possible de paramétrer des millions de bases.
- De prédire la situation à 1 an avec une erreur relative maîtrisée.
- D'être plus proche de la réalité. Le Machine Learning évite les biais d'hypothèses du modèle des risques qui projette les hypothèses S2.

Il est, en effet, très difficile de projeter la totalité des scénarios, avec le modèle mis en place par la DR, afin d'obtenir les projections par famille de produits, car cela nécessite un paramétrage

conséquent avec l'utilisation de plusieurs logiciels.

Dès lors, afin d'obtenir un proxy par famille de produits, nous avons mis en place des modèles de Machine Learning, pour lesquels, la prédiction par tête n'est pas nécessairement satisfaisante mais l'objectif est d'obtenir un résultat cohérent par famille et nous allons voir comment nous avons pu l'obtenir. Par ailleurs, l'automatisme de notre script permettra d'enrichir la base de données chaque année.

5.1 Les méthodes de Machine Learning

La data science est l'étude des données. Cela implique de développer des méthodes d'enregistrement, de stockage et d'analyse des données afin d'extraire efficacement les informations utiles. La data science intègre divers éléments et s'appuie sur des techniques et des théories venant de nombreux domaines : le traitement du signal, les mathématiques, l'optimisation, les modèles de probabilité, l'apprentissage statistique, la programmation informatique, l'ingénierie des données, la reconnaissance de formes, la modélisation de l'incertitude et le stockage de données. C'est donc une discipline s'appuyant sur des outils de statistiques mais également sur des outils mathématiques et informatiques.

Le data mining aussi appelé fouille de données consiste en l'utilisation de méthodes automatiques ou semi-automatiques afin d'extraire des données que nous allons ensuite analyser dans le but de faire apparaître des corrélations non encore identifiées entre ces données : cela crée des connaissances.

5.1.1 Les principaux types de données

Durant la création de nos modèles de Machine Learning, nous avons rencontré deux types de données :

- Les données quantitatives : Ce sont des valeurs décrivant une quantité qui est mesurable. Ce sont donc des nombres et il est possible d'en calculer la moyenne, l'écart-type et de les comparer si nous le souhaitons. Elles peuvent être continues, i.e., elles peuvent prendre n'importe quelle valeur dans un ensemble de valeurs, c'est par exemple le cas de la température. Elles peuvent également être discrètes, i.e., elles ne peuvent prendre qu'un nombre limité de valeurs dans un ensemble de valeurs, c'est par exemple le cas du nombre de personnes composant un foyer.
- Les données qualitatives : Elles permettent de décrire des caractéristiques. Il ne s'agit pas de nombres, mais d'un ensemble de modalités, donc nous ne pouvons pas effectuer de calculs sur ces valeurs, toutefois, il est possible de les comparer ou éventuellement de les trier. Elles peuvent être nominales (ou catégorielles), i.e., les modalités ne peuvent pas être ordonnées, c'est par exemple le cas du genre (homme, femme). Elles peuvent être ordinales, i.e., les modalités sont ordonnées selon un ordre "logique", c'est par exemple le cas de la taille des vêtements (S, M, L).

5.1.2 Data pre-processing

Lorsque nous avons une base de données, nous commençons par un pré-traitement de celle-ci (data pre-processing), i.e., il faut :

- Remplacer les valeurs manquantes par une méthode choisie.
- Grouper par modalités les variables qualitatives qui en ont beaucoup sinon, lors de la discrétisation, nous aurons des variables avec une variance très faible et celles-ci apporteront peu au modèle. C'est, par exemple, le cas d'une variable créée par une discrétisation, si la modalité qui lui a été associée était très peu présente dans la variable d'origine. Cependant, en couplant cette modalité avec d'autres, la variable résultante peut être significative pour le modèle, d'où l'importance du regroupement des modalités.
- Discrétiser les variables qualitatives.
- Supprimer les prédicteurs de variance nulle.
- Supprimer les variables redondantes et celles corrélées à d'autres variables.
- Vérifier que les couples de prédicteurs ne sont pas dépendants linéairement.
- Normaliser les variables quantitatives.

Les valeurs manquantes peuvent être un réel problème :

- Si nous supprimons les lignes où des valeurs manquent pour au moins une variable, nous perdons potentiellement de l'information.
- Conserver une variable pour laquelle il manque énormément de données peut être pénalisant selon le modèle utilisé. De plus, il faut bien choisir la méthode de remplacement des valeurs manquantes. En effet, si nous décidons de remplacer les valeurs manquantes par la moyenne des autres valeurs, nous pourrions rendre la variable quasi constante.

Afin de remplacer les valeurs manquantes, nous pouvons procéder à l'imputation simple ou multiple :

- Imputation simple : Si la distribution est centrée, il est préférable de prendre la moyenne de la variable pour remplacer les valeurs manquantes. S'il s'agit d'une distribution qui n'est pas centrée, dans ce cas, il est préférable de prendre la médiane de la variable. Si la variable dans laquelle il manque des données dépend d'autres variables de la base de données, il est préférable d'effectuer une régression linéaire afin de remplacer les valeurs manquantes.
- Imputation multiple : Dans ce processus, nous remplaçons chaque valeur manquante par $M > 1$ valeurs tirées d'une distribution appropriée. Ensuite, nous analysons indépendamment et avec la même méthode standard les $M > 1$ bases de données complètes. Enfin, nous combinons les résultats des analyses afin de refléter la variabilité supplémentaire due aux données manquantes et d'obtenir une estimation unique de toutes les valeurs manquantes.

5.1.3 L'apprentissage non-supervisé

L'apprentissage non-supervisé consiste en la déduction de connaissances à partir des données des échantillons d'apprentissage, et ceci, sans avoir de cible, il s'agit de rechercher des structures naturelles dans les données. Dans ce type de méthodes, il n'y a pas de valeur cible à atteindre à partir d'une prédiction, contrairement à ce qui est fait dans les méthodes d'apprentissage supervisé.

Le fait de créer des échantillons d'apprentissage étiquetés peut parfois être très coûteux, c'est pour cela que l'apprentissage non-supervisé peut s'avérer utile. Par ailleurs, il permet d'effectuer des analyses exploratoires afin d'en apprendre plus sur la structure et la nature des données. Il permet d'effectuer un pré-traitement avant d'appliquer une autre analyse de données et est utile pour l'étude de caractéristiques pertinentes. Ainsi, grâce à l'apprentissage non-supervisé, il est possible d'obtenir des modèles descriptifs permettant de mieux connaître ses données et de découvrir des informations "cachées" au milieu d'un grand nombre de données.

Nous pouvons associer diverses tâches à l'apprentissage non-supervisé :

- Le clustering (ou regroupement en français) : Il permet de construire des classes automatiquement à partir des exemples disponibles. Les techniques de clustering ont pour objectif de décomposer un ensemble de données en plusieurs sous-ensembles homogènes autant que possible. Nous ne connaissons pas la taille, la forme ou le nombre de classes créées. Il existe de nombreuses méthodes de clustering qui sont classifiées ainsi : méthodes de partitionnement, méthodes hiérarchiques, avec ou sans recouvrement, incrémentales ou non incrémentales. Avec le clustering, nous cherchons la maximisation de la similarité intra-classe et la minimisation de la similarité inter-classe.
- Règles d'association : Elles permettent de détecter des associations entre les variables ou d'analyser leurs relations.
- Réduction de dimensions.

5.1.3.a Méthodes de partitionnement

L'objectif des méthodes de partitionnement est de parvenir à décomposer un ensemble de données en un ensemble de clusters disjoints : si nous possédons un ensemble de N données, la méthode de partitionnement va construire K partitions de données avec $N \geq K$, chaque partition représentera un cluster. Ainsi, ces méthodes consistent en la création d'une partition initiale comprenant K classes, puis, en l'itération d'un processus permettant d'optimiser le partitionnement en déplaçant les objets d'une classe à une autre. Il y a ensuite deux possibilités :

- Méthode exacte : Elle consiste à évaluer toutes les partitions possibles.
- Méthodes approchées/heuristiques :

- Méthode K-moyennes (centres mobiles) : Dans cette méthode, nous représentons une classe par son centre de gravité. Un objet à classer appartiendra à la classe dont le centre de gravité est le plus proche de cet objet.
- Méthode K-medoids : Dans cette méthode, nous représentons une classe par son élément le plus représentatif.

5.1.3.b Méthodes hiérarchiques

L'objectif de ces méthodes est d'organiser les clusters sous la forme d'une structure d'arbre. Il existe, pour cela, deux approches :

- Approche hiérarchique ascendante (ou agglomération) : Cette approche commence avec un seul objet dans chaque classe. Ensuite, elle fusionne successivement les deux classes les plus proches. Elle s'arrête lorsqu'il n'y a plus qu'une seule classe (cette unique classe contient toutes les observations).
- Approche hiérarchique descendante (ou par division) : Initialement, cette approche commence avec tous les objets dans la même classe. Ensuite, elle divise l'unique classe en sous-classes jusqu'à ce qu'il y ait suffisamment de niveaux ou qu'il ne reste plus que des classes contenant une seule observation.

Dans ces méthodes, nous n'avons pas besoin de prédéfinir le nombre de clusters. Ces méthodes sont faciles à implémenter et fournissent une structure contrairement aux méthodes de partitionnement ce qui les rend préférables pour une analyse détaillée. Cependant, il n'est pas possible de remettre en question les divisions ou agglomérations, i.e., deux classes qui sont agglomérées à un niveau ne peuvent pas être séparées à un autre niveau.

5.1.4 L'apprentissage supervisé

L'apprentissage supervisé est constitué de deux étapes :

1. Nous commençons par estimer un modèle à partir des données disponibles, c'est une phase d'apprentissage : le modèle va apprendre comment labelliser les données dans le cas d'une classification et, dans le cas d'une régression, le modèle va apprendre comment renvoyer une valeur aussi proche de la réalité que possible.
2. Il s'ensuit une phase de test où nous vérifions que le modèle labellise les données comme il le faut ou attribue les bonnes valeurs.

Afin d'entraîner correctement un modèle, il est nécessaire de séparer la base de données originale en deux parties distinctes :

- La base d'apprentissage ("training set") : Elle est constituée de 70% à 80% des données et permet d'entraîner le modèle.
- La base de test ("test set") : Elle est constituée de 20% à 30% des données et permet de tester les performances du modèle.

Il est important de bien définir la taille des deux bases de données. En effet, une base d'apprentissage trop petite apprendra moins bien qu'une base d'apprentissage plus riche. Cependant, une base de test trop petite ne nous offrira pas assez de données afin de tester la performance du modèle.

Dans le cas de la classification, le modèle retourne les probabilités d'appartenir à chaque classe, ces probabilités sont également appelées "score". Dans le cas d'une classification entre deux classes (par exemple, l'évènement attendu a eu lieu et l'évènement attendu n'a pas eu lieu), nous obtenons des probabilités conditionnelles d'appartenir à une classe et non pas des variables binaires (0 ou 1). Ces probabilités sont arrondies à 0 ou 1 selon le seuil fixé par l'utilisateur afin d'obtenir les classes auxquelles les données appartiennent (le seuil de classification correspond à la probabilité à partir de laquelle nous considérons que l'évènement a eu lieu : par exemple, l'individu appartient à la classe 1). En classification, nous utilisons des données que nous avons associées à un label ou à une classe cible et nous voulons que l'algorithme soit capable, après un entraînement, de prédire ce label sur de nouvelles données qui n'ont pas encore été associées à un label.

D'autre part, dans le cas de la régression, le modèle retourne une valeur pour chaque ligne, il s'agit d'une valeur prédite. Nous pouvons évaluer l'efficacité du modèle en regardant l'écart entre la valeur réelle et la valeur prédite, par exemple, avec une métrique de type Root-Mean-Square Error (RMSE).

Nous disposons de certaines variables :

- Les variables explicatives : $X = (X_1, X_2, \dots, X_K)$.
- La variable à expliquer : Y . Elle peut être catégorielle (dans le cas d'une classification) ou numérique (dans le cas d'une régression).

Avec l'apprentissage supervisé, nous cherchons à déterminer Y_1 à partir des caractéristiques présentes dans $X_{1,k}$, avec $k \in \{1, \dots, K\}$. Cela est permis grâce à une fonction qui associe Y à X . L'apprentissage supervisé permet de déduire cette fonction à partir de données labellisées. Nous avons donc la relation qui suit : $f(X) = Y$. Il faut ensuite tester la performance de prédiction de f selon un ou plusieurs critères d'évaluation, ce test doit être effectué sur les données d'apprentissage mais surtout sur les données de test.

Il faut faire attention avec l'apprentissage supervisé, il y a "apprendre", "sur-apprendre" et "sous-apprendre" :

- Apprendre : Il s'agit de modéliser les régularités "réelles" qui sont présentes dans les données, c'est donc une modélisation pertinente.
- Sur-apprendre : Il s'agit de modéliser les régularités "accidentelles" (les spécificités) présentes dans l'échantillon d'apprentissage. Ainsi, le modèle prédictif s'adaptera très bien aux données d'apprentissage mais il ne pourra pas bien prédire les autres données.
- Sous-apprendre : Il s'agit de ne pas correctement modéliser, cela cause un mauvais apprentissage des données par l'algorithme.

Voici quelques méthodes d'apprentissage supervisé :

- Generalized Linear Model (GLM)
- Arbre de décision
- Bagging
- Random Forest
- Boosting
- Gradient Boosting
- XGBoost

Le biais et la variance sont des éléments importants concernant les modèles de Machine Learning :

- Le biais : C'est l'espérance de la différence entre la prédiction obtenue par le modèle et les vraies valeurs : $Biais(\hat{f}(x)) = \mathbb{E}(\hat{f}(x) - f(x))$.
- La variance : C'est la variabilité des prédictions du modèle pour des observations données : $Var(\hat{f}(x)) = \mathbb{E}(\hat{f}(x)^2) - \mathbb{E}(\hat{f}(x))^2$.
- Compromis biais/variance : Plus un modèle est complexe, plus il arrive à capter la variance de la base d'apprentissage mais donc plus il risque de sur-apprendre et ainsi d'augmenter le biais lors de l'application du modèle sur la base de test.

Nous pouvons donc définir l'erreur d'un modèle de la façon suivante : $Err(x) = Biais^2 + Variance + Erreur irréductible$. L'objectif est de minimiser cette erreur en tenant compte du comportement du biais et de la variance. En effet, ces deux éléments varient de manière opposée en fonction de la complexité du modèle.

Afin de trouver un bon compromis, nous pouvons :

- Réduire le nombre de variables du modèle : Lorsque la variance est trop élevée, i.e., le modèle est très dépendant de la base d'apprentissage, le fait de réduire le nombre de variables du modèle permet de simplifier le modèle et donc d'en réduire la variance.

- Utiliser des méthodes ensemblistes : Il existe des algorithmes qui se basent sur la combinaison de plusieurs modèles à haute variance en les agrégeant (par exemple, en les moyennant) afin de réduire la variance finale.

Une méthode ensembliste est une technique de Machine Learning qui consiste à entraîner plusieurs modèles en utilisant pour chacun d'eux le même algorithme d'apprentissage. Par exemple, entraîner plusieurs arbres de décision est une méthode ensembliste.

Les principales causes d'erreurs durant l'apprentissage sont, en particulier, dues au biais et à la variance. Les méthodes ensemblistes aident à minimiser ces facteurs, elles ont été conçues pour améliorer la stabilité et la précision des algorithmes de ML.

5.1.4.a Generalized Linear Model

Les GLM sont constitués de trois composantes qui sont les suivantes :

- La variable réponse nommée Y : C'est une composante aléatoire à laquelle une loi de probabilité a été associée.
- Les variables explicatives X_1, X_2, \dots, X_K : Elles sont utilisées comme des prédicteurs dans le modèle, elles définissent la composante déterministe sous forme d'une combinaison linéaire.
- La dernière composante est le lien : Il décrit la relation fonctionnelle entre l'espérance mathématique de la variable réponse et la combinaison linéaire des variables explicatives.

Un modèle statistique $(\Omega, \mathcal{F}, (\mathbb{P}_{\theta, \phi})_{\theta \in \Theta, \phi > 0})$ est appelé famille exponentielle si les probabilités $\mathbb{P}_{\theta, \phi}$ admettent une densité f par rapport à une mesure dominante avec :

$$f_{\theta, \phi}(y) = c_{\phi}(y) \exp\left(\frac{y\theta - a(\theta)}{\phi}\right), \text{ où :}$$

- θ est le paramètre canonique et ϕ est le paramètre de dispersion.
- $a(\theta)$ est de classe C^2 et est convexe.
- $c_{\phi}(y)$ ne dépend pas de θ .

Un modèle est un GLM s'il vérifie les hypothèses suivantes :

1. $Y|X = x \sim \mathbb{P}_{\theta, \phi}$ appartient à une famille exponentielle.
2. $g(\mu(X)) = g(\mathbb{E}[Y|X]) = X\beta$ pour une certaine fonction g bijective, appelée fonction de lien.

$Y|X = x$ est souvent choisie parmi la loi de Poisson, le logit, la loi gaussienne, la loi binomiale ou la loi Gamma. Par exemple, la loi gaussienne de paramètres m et σ^2 , avec σ^2 connu, appartient à une famille exponentielle où : $\theta = m$, $a(\theta) = \frac{\theta^2}{2}$ et $\phi = \sigma^2$.

La formule générale du GLM est donc de la forme suivante : $g(\mathbb{E}(Y|X)) = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K$, où :

- g est la fonction de lien. La fonction de lien est différente selon la loi utilisée. Par exemple, il s'agit de la fonction identité pour une loi gaussienne et de la fonction inverse pour une loi Gamma.
- $\beta_1, \beta_2, \dots, \beta_K$ sont les poids respectivement associés aux prédicteurs X_1, X_2, \dots, X_K .

5.1.4.b Arbre de décision

Il s'agit d'une des méthodes d'apprentissage supervisé les plus populaires en ce qui concerne les problèmes de classification des données. Le but de ce type de modèles est d'expliquer une valeur grâce à une série de variables discrètes ou continues. C'est une méthode d'apprentissage non-paramétrique et non-linéaire. Les arbres de décision modélisent une hiérarchisation des tests avec l'objectif de prédire un résultat, il en existe deux types :

- Les arbres de régression (Regression Tree) : Ces arbres permettent de prédire une valeur numérique (par exemple, le prix d'une maison). Dans ce cas, la variable réponse Y est quantitative continue.
- Les arbres de classification (Classification Tree) : Ces arbres permettent de prédire la classe à laquelle la variable de sortie appartient (cela peut, par exemple, permettre de répartir des assurés en différents types de profils). Dans ce cas, la variable réponse Y est qualitative.

Les arbres de décision vont partitionner les individus en produisant plusieurs groupes d'individus de la manière la plus homogène possible du point de vue de la variable à prédire.

Nous définissons ainsi la hiérarchie des décisions élémentaires :

- Noeud : À chaque noeud, une variable de décision est choisie.
- Arc : Il s'agit d'une condition sur la variable choisie.
- Feuille : Au bout d'un chemin de l'arbre, résultat de la décision.

Lors de la création d'un arbre de décision, il est nécessaire d'avoir plusieurs itérations. À chaque itération, les données sont séparées en k sous-ensembles (souvent, $k = 2$) afin d'expliquer la variable de sortie. La première division des individus est obtenue en choisissant la variable explicative qui offrira la meilleure séparation possible. Cette première division permet de définir les sous-populations d'individus, celles-ci sont représentées par les noeuds de l'arbre de décision. Une mesure de proportion est rattachée à chaque noeud afin d'expliquer l'appartenance à une classe ou bien la signification d'une variable de sortie. De nouvelles itérations sont effectuées jusqu'à ce qu'il ne soit plus possible de séparer les données. C'est alors que nous obtenons des noeuds terminaux qui sont nommés "feuilles". Ainsi, chaque feuille est caractérisée par un certain chemin à travers l'arbre qui est appelé "règle" et l'ensemble des règles d'un arbre constitue le modèle de Machine Learning.

L'indice de diversité de Gini est un indice utilisé par les arbres de décision. Il permet de mesurer l'"impureté" ou le "désordre" à un noeud, i.e., avec quelle fréquence un élément aléatoire serait mal classé si nous choissions son étiquette au hasard selon la distribution des étiquettes dans le sous-ensemble. En effet, la qualité d'un noeud et son pouvoir discriminant peuvent être évalués par son "impureté". L'indice de Gini est calculé à chaque noeud. Pour un noeud S , il est calculé de la manière suivante :

- Nous partitionnons S sur les valeurs de la cible en n groupes : C_1, C_2, \dots, C_n .
- Nous calculons la probabilité estimée qu'un élément de S se retrouve dans C_i : $\frac{|C_i|}{|S|}$.
- $I_G(S) = \sum_{i=1}^n \frac{|C_i|}{|S|} (1 - \frac{|C_i|}{|S|}) = \sum_{i \neq j} \frac{|C_i||C_j|}{|S|^2}$.

Sur un arbre de décision, les différentes décisions possibles se situent sur les feuilles de l'arbre à l'extrémité des branches. Ces feuilles sont atteintes en fonction des décisions prises à chaque noeud. Il s'agit d'appliquer de manière itérative des règles logiques et chacune de ces règles est choisie en fonction du résultat de la règle précédente.

Voici les avantages de ce type de modèles :

- Ils sont simples à interpréter.
- Ils sont rapides à entraîner.

Toutefois, ce type de modèles peut très vite mener à un sur-apprentissage des données, il prédira parfaitement les données sur lesquelles il s'entraînera mais aura beaucoup de mal à prédire d'autres données. Ainsi, les arbres de décision sont rarement utilisés seuls, ils sont souvent regroupés, par exemple, pour la création d'une forêt aléatoire (voir sous-partie 5.1.4.d).

5.1.4.c Bagging

Le bootstrap est une méthode statistique puissante permettant d'estimer une quantité à partir d'un échantillon de données. Ceci est plus facile à comprendre si la quantité est une statistique descriptive telle qu'une moyenne ou un écart-type. Le bootstrap consiste à créer de nouveaux

échantillons (échantillons bootstrap) en réalisant des tirages aléatoires avec remise dans la base de données.

Le mot "bagging" vient de la contraction des mots "bootstrap" et "aggregating", c'est une méthode ensembliste simple et très puissante. C'est une procédure générale qui peut être utilisée pour réduire la variance des algorithmes qui en ont une élevée, c'est, par exemple, le cas des arbres de décision. Le bagging permet alors de diminuer la variance de l'estimateur en corrigeant l'instabilité des arbres de décision. Le bagging est l'application de la procédure bootstrap à un algorithme de Machine Learning à haute variance. C'est une méthode permettant d'améliorer la classification, en particulier, dans le cas des arbres de décisions. En effet, les arbres de décisions ont beaucoup d'avantages mais ce sont des "classifieurs (modèles) faibles", c'est-à-dire qu'ils ne sont pas très efficaces. Le bagging est également utilisé pour les régressions.

Les arbres de décision sont sensibles aux données spécifiques sur lesquelles ils sont entraînés. Si les données d'apprentissage sont modifiées (par exemple, si l'arbre de décision est entraîné sur un sous-ensemble des données d'apprentissage), l'arbre de décision peut être très différent donc les prévisions peuvent être très différentes également.

Le bagging fonctionne ainsi :

1. Créer B échantillons de données de manière aléatoire grâce à un tirage des données avec remise.
2. Entraîner chaque modèle sur un échantillon de données différent. Si nous avons créé B nouveaux échantillons, alors, B modèles sont ajustés en utilisant ces B échantillons bootstrap.
3. À partir des données qui n'ont pas servi à entraîner les modèles, combiner les B modèles, soit en calculant la moyenne de ces modèles dans le cas d'une régression, soit en procédant par vote dans le cas d'une classification. C'est le fait de combiner ces modèles "indépendants" qui permet de diminuer la variance.

Lors du bagging sur des arbres de décision, le seul paramètre à inclure est le nombre d'échantillons et donc le nombre d'arbres. Ce paramètre peut être déterminé en augmentant le nombre d'arbres petit à petit jusqu'à ce que la précision de la forêt d'arbres commence à ne plus s'améliorer. Toutefois, plus le nombre de modèles (arbres de décision, dans ce cas) augmente et plus l'algorithme prend du temps.

5.1.4.d Random Forest

Nous pourrions nous attendre à ce que l'ajout de données dans la base d'apprentissage ne modifie que peu le modèle et plutôt en l'améliorant. Or, dans le cas des arbres de décision, le modèle peut totalement changer avec l'ajout de nouvelles observations car les arbres ont tendance à se sur-ajuster aux données. Dès lors, afin de pouvoir utiliser les arbres de décision en réduisant leurs défauts, Leo Breiman a eu l'idée d'utiliser plusieurs arbres dans le but de créer des forêts d'arbres aléatoires. Cela a permis de garder la polyvalence des arbres de décision, i.e., les forêts aléatoires peuvent être utilisées en classification (le résultat final est obtenu en faisant voter chaque arbre) et en régression (le résultat final est obtenu en prenant la moyenne des résultats de chaque arbre).

Les Random Forest sont une amélioration par rapport au bagging appliqué aux arbres de décision. En effet, il y a un problème avec les arbres de décision : ils choisissent la variable à diviser en utilisant un algorithme qui minimise les erreurs. Ainsi, même avec le bagging, les arbres de décision peuvent avoir de nombreuses similitudes structurelles et donc avoir une forte corrélation dans leurs prédictions. Or, la combinaison des prédictions de plusieurs modèles fonctionne mieux si les prédictions des modèles ne sont pas corrélées ou alors faiblement corrélées.

Ainsi, les Random Forest modifient la façon d'apprendre des arbres afin que les prédictions résultant de tous les arbres et de toutes les branches soient moins corrélées, il s'agit d'un simple ajustement. Dans les arbres de décision, lorsque nous arrivons à un noeud, l'algorithme d'apprentissage est autorisé à parcourir toutes les variables et toutes les valeurs que peuvent prendre les variables afin de sélectionner la segmentation optimale. La méthode Random Forest modifie cette procédure afin que l'algorithme d'apprentissage soit limité à un échantillon aléatoire de variables à parcourir, i.e., il n'aura pas accès à toutes les variables.

Le nombre de variables pouvant être recherchées à chaque noeud doit être spécifié en tant que paramètre de l'algorithme. Pour la classification, une bonne valeur par défaut est : $m = \sqrt{p}$ et, pour la régression, une bonne valeur par défaut est : $m = \frac{p}{3}$, où m est le nombre de variables sélectionnées au hasard qui peuvent être recherchées à un noeud et p est le nombre de variables d'entrée. Par exemple, si un ensemble de données a 16 variables d'entrée pour un problème de classification, alors, nous aurons : $m = \sqrt{16} = 4$.

Les Random Forest permettent, en utilisant plusieurs arbres de décision, de diminuer la variance des prévisions d'un seul arbre de décision par l'introduction de facteurs aléatoires dans leur construction (échantillon aléatoire limité de variables à parcourir à chaque noeud). En combinant les arbres, les Random Forest permettent d'augmenter leur performance et de diminuer leur surajustement aux données. Cet algorithme consiste à effectuer un apprentissage sur un grand nombre d'arbres de décision qui ont été construits aléatoirement et entraînés sur des échantillons de données différents selon le principe du bagging. Pour les Random Forest, les échantillons bootstrap comprennent tous n données comme l'échantillon d'origine mais il y a plusieurs lignes identiques et certaines lignes de l'échantillon d'origine ne sont pas présentes car il s'agit d'un tirage des données avec remise. Cependant, plus il y a d'arbres dans la forêt, plus l'algorithme prend du temps, il est donc nécessaire de trouver le nombre d'arbres idéal.

La méthode Random Forest est l'un des classifieurs les plus efficaces contrairement à la méthode de l'arbre de décision unique. Néanmoins, pour que le bagging soit efficace dans le cas des Random Forest, il faut que les arbres de décision soient performants individuellement et qu'ils soient assez différents les uns des autres dans le but de se compléter.

5.1.4.e Boosting

Le mot "boosting" représente la transformation de plusieurs prédicteurs "faibles" en prédicteurs "forts", il s'agit d'une méthode non-linéaire extrêmement performante. Le boosting permet de réduire le biais des algorithmes sur lesquels il s'appuie. En effet, le boosting regroupe beaucoup d'algorithmes s'appuyant sur des ensembles de classifieurs binaires, il permet d'optimiser les performances de ces algorithmes. C'est une technique itérative permettant d'approcher la solution d'un problème d'optimisation. En effet, en apprentissage supervisé, les modèles sont souvent construits en déterminant les paramètres qui permettent d'optimiser une fonction objectif. Le boosting consiste en la combinaison de ces classifieurs faibles. Ainsi, par des itérations successives, la connaissance d'un classifieur faible s'ajoute au classifieur final qui est plus efficace. Les classifieurs sont pondérés par la qualité de leur classification, i.e., plus il est efficace, plus il aura un poids important dans le classifieur final.

Au vu des définitions respectives du bagging et du boosting, nous remarquons que ces deux méthodes se ressemblent. En effet, elles sont toutes les deux basées sur la combinaison de classifieurs faibles dans le but de créer un classifieur plus performant. Par ailleurs, afin d'utiliser une de ces méthodes, il faut choisir un algorithme d'apprentissage de base (le classifieur faible).

Cependant, elles ont des différences. En effet, dans le cas du bagging, n'importe quel élément a la même probabilité d'apparaître dans le nouvel échantillon d'apprentissage. A contrario, pour le boosting, les observations sont associées à un poids, ainsi, certaines d'entre elles seront présentes plus souvent dans les nouveaux échantillons.

Alors que l'étape d'apprentissage se fait en parallèle pour le bagging, i.e., chaque modèle est construit indépendamment des autres, pour le boosting, chaque classifieur est entraîné sur des données en tenant compte du succès des précédents modèles, après chaque étape d'entraînement, les poids sont redistribués. Ainsi, le poids des données mal classées est augmenté de manière à ce que les modèles suivants se concentrent sur ces données durant leur entraînement.

Afin de prédire de nouvelles données, il suffit d'appliquer les N apprenants créés à ces données. Pour le bagging, la prédiction est obtenue en faisant la moyenne des réponses des N apprenants. Le boosting, quant à lui, attribue un deuxième ensemble de poids, cette fois-ci, pour les N classifieurs, afin de prendre une moyenne pondérée de leurs estimations. Un classifieur prédisant bien les résultats aura un poids plus élevé qu'un mauvais classifieur.

Les deux techniques permettent de réduire la variance, cependant, seul le boosting essaie de

réduire le biais et seul le bagging peut résoudre un problème de sur-apprentissage alors que le boosting peut, au contraire, l'augmenter.

5.1.4.f Gradient Boosting

Souvent, les équations ne sont pas linéaires. Dès lors, les logiciels calculent les estimations en utilisant des algorithmes dits itératifs. Parmi ces algorithmes, le plus connu et le plus utilisé est celui de Newton-Raphson, il approxime le logarithme de la fonction de vraisemblance par une fonction polynomiale, qui a la forme d'une parabole concave, dans un voisinage du paramètre initial. Les approximations successives de cet algorithme itératif convergent rapidement vers les estimations au sens du maximum de vraisemblance. De plus, chaque étape de cet algorithme constitue un ajustement de type "moindres carrés pondérés".

Le gradient boosting est une technique utile pour les problèmes de régression et de classification. Elle utilise la "descente de gradient" et le boosting. La descente de gradient est une technique permettant d'approximer la solution d'un problème d'optimisation avec une approche itérative. Cette problématique revient à minimiser la fonction de perte $j(y, h(x))$.

Le gradient boosting permet de produire un modèle de prédiction sous la forme d'un ensemble de modèles de prédiction dits "faibles" (il s'agit généralement d'arbres de décision). Les modèles sont construits étape par étape comme pour les autres méthodes de boosting et il les généralise en permettant l'optimisation d'une fonction de perte arbitrairement différentiable. C'est une méthode qui permet de nombreux choix de pertes et de nombreux choix d'apprenants faibles.

Une fonction de coût, que nous allons chercher à minimiser, est définie de la façon suivante : Soit S un ensemble de n observations : $S = x_1, x_2, \dots, x_n$, les données de S sont labellisées par le vecteur $Y = y_1, y_2, \dots, y_n$. Si h est la fonction d'hypothèse qui cherche à approximer le vecteur Y , nous définissons alors la fonction de coût $J : J(h) = \sum_{i=1}^n j(y_i, h(x_i))$. Dans le cas du boosting, la notion d'erreur est représentée par la notion de coût. Pour les méthodes classiques, c'est l'écart entre la fonction d'hypothèse h et la cible Y qui est qualifiée d'erreur. A contrario, pour le boosting, nous comparons les résultats de chaque nouvelle fonction d'hypothèse h_i aux résidus qui ont été "laissés" par h_{i-1} . Il s'agit donc de construire itérativement B fonctions d'hypothèse h qui vont ensuite être assemblées dans l'algorithme final H . La première fonction d'hypothèse h_1 va réaliser un premier apprentissage des données, à ce stade, H sera égale à h_1 , puis, h_2 va tenter d'améliorer la fonction H telle que :

$$\forall i \in \{1, \dots, n\}, H(x_i) + h_2(x_i) \approx y_i \iff h_2(x_i) \approx y_i - H(x_i).$$

Le rôle de h_2 est bien d'approximer les données d'entrée par rapport aux erreurs de h_1 . Par ailleurs, le terme $y_i - H(x_i)$ correspond aux résidus de l'étape précédente. Nous allons maintenant tenter d'exprimer ce terme comme un gradient négatif d'une certaine fonction de coût. Pour cela, il suffit d'utiliser la fonction de coût des moindres carrés pour un problème de régression :

$$j(y_i, H(x_i)) = \frac{(y_i - H(x_i))^2}{2},$$

$$J = \sum_{i=1}^n j(y_i, H(x_i)).$$

Nous calculons ensuite le gradient de J par rapport à $H(x_i)$:

$$\frac{\partial J}{\partial H(x_i)} = \frac{\partial \sum_k j(y_k, H(x_k))}{\partial H(x_i)} = H(x_i) - y_i.$$

Ainsi, h_2 doit minimiser la fonction de coût J . La descente de gradient est utilisée à chaque itération permettant de construire une nouvelle fonction d'hypothèse h_i . Nous exprimons donc les résidus comme le gradient négatif de J .

Ensuite, nous construisons H itérativement en mettant à jour notre fonction H de la manière suivante :

$$H(x_i) := H(x_i) - \frac{\partial J}{\partial H(x_i)}, \forall i \in \{1, \dots, n\}.$$

L'idée du gradient boosting est donc de remplacer la minimisation exacte par un "gradient step" puis d'approximer celui-ci par un élément de H .

Ce sont les "outliers" dans nos données (s'il y en a) qui vont conditionner le choix de la fonction de coût. Les outliers, également appelés "données aberrantes", sont des données qui ne sont pas proches des autres. Utiliser la descente de gradient permet l'utilisation de multiples fonctions de coût. Par ailleurs, même si nous souhaitons traiter un problème de classification avec un gradient boosting, le modèle utilisera des algorithmes de régression faibles.

Le gradient boosting est donc une méthode extrêmement performante qui permet beaucoup de choses. Cependant, l'un de ses défauts est le temps qu'il prend pour s'entraîner.

La méthode de gradient boosting la plus utilisée est ADAPtive BOOSTing (AdaBoost) pour la classification binaire.

Pour plus de détails concernant le gradient boosting, nous invitons le lecteur à regarder le livre référencé [DATA-SCIENCE-2015] dans la bibliographie.

5.1.4.g XGBoost

L'Extreme Gradient Boosting (XGBoost) est une implémentation spécifique de la méthode gradient boosting mais il utilise des approximations plus précises pour trouver le meilleur modèle d'arbre. Il apporte de nouveaux éléments par rapport au gradient boosting :

- Il permet de calculer des gradients du second ordre, c'est-à-dire, des dérivées secondes partielles de la fonction de perte (similaire à la méthode de Newton-Raphson). Cela fournit plus d'informations sur la direction des gradients et comment atteindre le minimum de notre fonction de perte. Ainsi, alors que le gradient boosting basique utilise la fonction de perte de notre modèle de base comme proxy pour minimiser l'erreur globale du modèle, le XGBoost utilise la dérivée du second ordre comme approximation.
- L'implémentation du XGBoost est parallèle, cela permet de l'entraîner beaucoup plus vite.
- Il permet, par exemple, l'utilisation de modèles linéaires comme algorithmes faibles alors que le gradient boosting ne permet d'implémenter que des algorithmes de régression.

Le XGBoost s'attaque au principal soucis du gradient boosting. En effet, le gradient boosting tient compte de la perte potentielle de tous les splits possibles avant de créer une nouvelle branche à l'arbre. Le XGBoost, quant à lui, examine la distribution des variables sur toutes les données d'un noeud et utilise ces informations pour réduire l'espace de recherche des splits.

Le XGBoost permet d'étudier rapidement de nombreux hyperparamètres conçus pour limiter le sur-apprentissage des données. En effet, le fait de regrouper beaucoup de modèles faibles cause du sur-apprentissage. Ainsi, le paramétrage est assez complexe avec ces algorithmes et il est important de bien les paramétrer afin d'éviter le sur-apprentissage.

Après la création des modèles, il est nécessaire de les évaluer afin de les valider ou non. En effet, il ne va pas être difficile de construire un modèle qui saura bien prédire ses données d'apprentissage étant donné qu'il s'est appuyé sur celles-ci lors de sa construction. Ce qui va être bien plus complexe, c'est de créer un modèle pouvant prédire correctement des données qu'il ne connaît pas, il faut que le modèle soit capable de se généraliser.

5.1.5 Indicateurs de performance

Il existe plusieurs indicateurs de performance nous permettant d'évaluer la performance des méthodes de data science testées. Il faut tout de même faire attention à ne jamais évaluer un modèle sur des données que nous avons utilisées pour l'entraîner.

Lors de l'évaluation d'un modèle, nous utilisons une métrique de performance sur la base d'apprentissage et sur la base test. La performance de la base test est inférieure à la performance de la base d'apprentissage car le modèle a été construit à l'aide de cette base, il est donc normal que la performance soit meilleure sur la base d'apprentissage.

Afin d'obtenir le meilleur modèle possible, nous essayons donc d'augmenter la performance sur la base test et de diminuer celle sur la base d'apprentissage dans le but de limiter le sur-apprentissage du modèle. Pour cela, il est possible d'essayer différents choix de variables ou d'utiliser plusieurs

paramétrages sur la base d'apprentissage afin de trouver celui qui maximise la performance sur la base test. C'est d'ailleurs ce que nous verrons plus tard lors de l'utilisation d'un XGBoost sur nos données. Cependant, étant donné que le modèle est choisi de manière à maximiser la qualité des prédictions sur la base test, il s'ensuit que la base test n'est plus tout à fait une nouvelle base de données que le modèle ne connaît pas, cela peut donc créer un sur-apprentissage du modèle. Pour contrer ce problème, il est possible de séparer la base de données en trois bases au lieu de deux (apprentissage, validation et test).

5.1.5.a La validation croisée

La validation croisée nous permet d'utiliser toutes les données dont nous disposons pour l'entraînement et pour la validation du modèle, elle permet donc d'éviter un biais potentiel lié à une validation unique. En effet, au lieu de simplement séparer l'ensemble des données en deux sous-ensembles permettant, d'une part, d'entraîner les données et, d'autre part, de les valider, nous allons séparer l'ensemble des données plusieurs fois. Il existe trois principales méthodes de validation croisée :

- La méthode Leave-One-Out cross-Validation (LOOV) : Il s'agit de sortir une observation i de notre base de données d'apprentissage et d'entraîner le modèle sur les $n-1$ données restantes. Le modèle est alors utilisé pour prédire i et nous calculons ensuite l'erreur de prédiction. Le même processus est répété pour toutes les valeurs de $i = \{1, \dots, n\}$. Enfin, les n erreurs de prédiction sont utilisées dans le but d'évaluer la performance du modèle en validation croisée.
- La méthode Leave-K-Out cross-Validation (LKOV) : Celle-ci fonctionne de la même manière que la méthode LOOV mais il faut sortir k observations à prédire au lieu d'une seule. Le processus est répété de manière à avoir réalisé tous les découpages possibles pour les données de modélisation et de prédiction.
- La méthode k-fold cross-validation : Dans cette méthode, les données sont aléatoirement séparées en k sous-échantillons contenant à peu près le même nombre de données. Nous nous retrouvons donc avec k jeux de données à entraîner et à tester contenant chacun k parties. Nous allons ensuite entraîner $k-1$ parties et garder la partie qui reste pour la prédiction. Nous procédons de même pour les k jeux de données. Ainsi, toutes les parties seront testées une par une et il y a bien indépendance entre la base d'apprentissage et la base de test. Contrairement à ce qui est fait avec la méthode LKOV, le processus n'est répété que k fois. Finalement, toutes les parties auront servi une fois dans un jeu de test et $k-1$ fois dans un jeu d'entraînement. Il est ensuite possible d'évaluer la performance du modèle soit en évaluant les prédictions que nous avons faites sur l'ensemble des données car toutes les données ont été prédites une fois, soit en faisant la moyenne des performances obtenues sur les k parties. Si nous voulons évaluer un modèle de classification, il faudra faire en sorte que les k parties créées contiennent à peu près la même proportion de chaque classe que le jeu de données complet afin d'éviter de biaiser les résultats, cette variante s'appelle alors la stratified k-fold cross-validation.

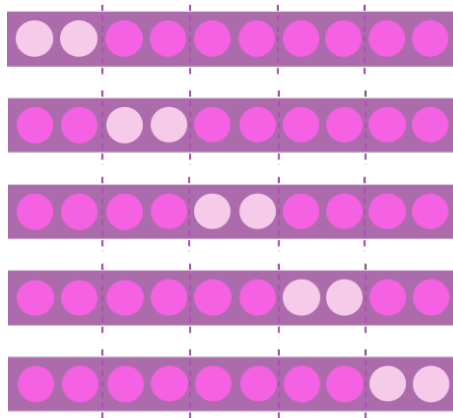


FIGURE 24 – Représentation de la méthode k-fold cross-validation

Sur le schéma ci-dessus, chaque point appartient à un des cinq jeux de test (en rose clair) et aux quatre autres jeux d'entraînement (en rose foncé).

Les trois méthodes que nous avons décrites peuvent être séparées en deux familles : les méthodes exhaustives (LOOV et LKOV) et les méthodes non-exhaustives (k-fold cross-validation). Une méthode est dite exhaustive si elle a divisé le jeu de données en fonction de toutes les combinaisons possibles. En effet, la k-fold cross-validation ne l'est pas car elle sépare les données en un nombre limité de sous-ensembles d'observations possibles. Il s'agit donc d'une approximation des méthodes exhaustives et elle présente l'avantage d'être plus rapide que celles-ci.

Il existe de nombreuses mesures permettant d'évaluer la qualité d'un modèle de régression, elles sont toutes basées sur des calculs réalisés à l'aide de trois grandeurs :

- La valeur observée d'une série à prédire (y_i).
- La valeur prédite par le modèle pour l'observation de y_i (\hat{y}_i).
- La moyenne empirique de y (\bar{y}), c'est la prédiction "naïve" de référence.

Ces différentes grandeurs permettent de calculer, pour tout i :

- L'erreur de prédiction du modèle : $y_i - \hat{y}_i$.
- L'erreur de prédiction "naïve" : $y_i - \bar{y}$.

Tout cela permet alors de définir des indicateurs de performance du modèle.

5.1.5.b La Root-Mean-Square Error

La RMSE permet d'obtenir une indication quant à la dispersion ou la variabilité de la prédiction d'un modèle, elle peut donc être reliée à la variance du modèle ou plutôt à son écart-type. Il faut, cependant, faire attention à l'évaluation de sa valeur. Par exemple, une RMSE de 20 est relativement faible si la moyenne des observations est de 700. En revanche, un modèle a une variance élevée s'il a une RMSE de 20 et que la moyenne des observations est de 25. En effet, dans le premier cas, la RMSE du modèle correspond à seulement 3% de la moyenne des observations alors que, dans le second cas, la RMSE correspond à 80% de la moyenne des observations.

La RMSE s'écrit ainsi : $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$.

La RMSE est un indicateur important pour les modèles de régression en Machine Learning, elle correspond à une indication agrégée de l'erreur de prédiction. C'est cette méthode, en particulier, qui nous a permis d'évaluer la qualité de la prédiction de nos différents modèles

5.1.5.c Le critère d'information d'Akaike

Le critère d'information d'Akaike (AIC) s'écrit comme suit :

- Formulation générale : $-2\log(L) + 2(k + 1)$, où L est la vraisemblance du modèle.
- Formulation dans le cas d'erreurs distribuées normalement : $n\log(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2) + 2(k + 1)$.

Ainsi, l'AIC évalue un modèle en fonction de sa vraisemblance et pénalise les modèles avec trop de variables. Plus l'AIC est petit et plus le modèle sera meilleur.

Maintenant que nous avons vu plusieurs méthodes d'apprentissage supervisé et que nous avons compris l'importance des indicateurs de performance afin de vérifier la qualité des prédictions des modèles, nous allons présenter les résultats que nous avons obtenus.

5.2 Présentation des modèles construits et des résultats obtenus

L'objectif de cette sous-partie est de présenter les différents modèles de Machine Learning que nous avons construits et de regarder les prédictions que nous avons obtenues. Par la suite, nous avons également comparé le meilleur modèle de ML au Best Estimate S2 transformé en PNA.

Avec les garanties du périmètre DCAT sur la période 2016-2019, nous avons pu accumuler une grande base de données permettant alors d'utiliser des méthodes de Machine Learning.

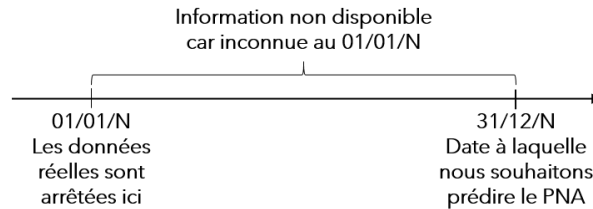


FIGURE 25 – Explication des données pouvant être utilisées pour la prédiction du PNA

Le but des modèles de Machine Learning était de prédire le PNA à la fin de l'année N à partir des données présentes dans la base au début de l'année N , i.e., à la fin de l'année $N-1$. Par exemple, si nous souhaitons prédire le PNA pour un client au 31/12/2017, nous allons uniquement utiliser les données présentes au 01/01/2017, soit les données présentes au 31/12/2016. Ainsi, pour la prédiction du PNA au 31/12/2017, si un sinistre est survenu en 2017, l'ancienneté depuis la survenance sera nulle car nous n'aurons aucune connaissance du sinistre à venir au 01/01/2017. Nous allons donc utiliser les provisions de sinistres de début d'année N et un top illustrant si un sinistre était en cours l'année $N-1$ à la place des provisions de fin d'année N et du top illustrant si un sinistre était en cours l'année N . Finalement, nous avons supprimé les affaires nouvelles de l'année N car, au début de l'année N , nous n'avons pas connaissance de celles-ci.

Avant de commencer à entraîner nos modèles sur la base de données, il a été important de regarder les variables que nous allons garder. Afin de décider cela, nous avons regardé les variables présentes dans la base Solvency permettant de calculer le Best Estimate de 2019 au 31/12/2018 sous Solvabilité 2. Tout d'abord, nous avons préparé cette base afin qu'elle puisse, par la suite, être utilisée dans le but de prédire le PNA de l'année 2019 avec les différents modèles de Machine Learning créés. Il a donc fallu séparer les lignes par risque car tous les risques étaient agrégés dans une même ligne par client. La base avait une ligne par client mais plusieurs colonnes de montants assurés et de primes HT, par exemple (une pour chaque risque).

Une fois les lignes séparées, nous avons pu calculer de nouvelles variables pour la base Solvency afin d'obtenir un maximum de variables présentes dans la base construite (voir sous-partie 3.1.1). Ainsi, nous avons calculé des variables à ajouter telles que l'ancienneté en mois ou l'ancienneté depuis la survenance en mois car nous avons les dates de survenance dans la base Solvency.

Afin d'ajouter d'autres variables manquantes, nous avons joint la base Solvency à la base que nous avons construite. En effet, les clients présents dans la base Solvency au 31/12/2018 étaient également présents dans notre base portefeuille à cette même date. En joignant les deux bases, nous avons donc été en mesure de récupérer certaines variables utiles au modèle. En effet, nous avons pu récupérer les provisions de fin d'année mais il aurait également été possible de les récupérer avec la base de suivi des sinistres. Par ailleurs, nous avons également pu récupérer les nom des établissements bancaires et le top indiquant si un sinistre était en cours dans l'année ou non.

Si nous n'avions pas ajouté toutes ces variables à la base Solvency, nous n'aurions pas pu les utiliser dans le modèle, sinon, le modèle aurait appris à partir de ces variables et n'aurait pas pu prédire un PNA sans celles-ci.

Nous noterons que les BE S2 ont été déterminés par famille de produits et par HRG, nous avons donc pu les agréger uniquement par famille de produits d'une part et uniquement par risque d'autre part, ceci, dans le but de comparer les indicateurs de rentabilité par famille de produits ou par risque. Au sein du département Actuariat Produits Prévoyance, il est très long de déterminer les BE de S2 car nous avons environ 20 familles de produits, deux modèles différents (gestion centralisée et gestion déléguée) et qu'il y a 18 chocs à effectuer soit 18 scénarios. Au total, nous arrivons donc à $20 \times 2 \times 18 = 720$ scénarios. Ainsi, comme nous l'avons précisé dans les avantages du Machine Learning, il aurait été trop chronophage et donc impossible de calculer des BE par tête sous S2, c'est pour cela que les BE sont toujours calculés à la maille famille de produits et HRG et, c'est ici, que les modèles de Machine Learning montrent tout leur intérêt. En effet, certains modèles ont pris un peu de temps à se construire étant donné le nombre d'observations que nous avons, cependant, ils ont mis bien moins de temps que ce dont nous aurions eu besoin pour déterminer des BE par tête avec S2. Par ailleurs, l'an prochain, il sera possible de reconstruire un modèle avec les mêmes paramètres que ceux utilisés pour le modèle retenu et celui-ci pourra tourner sur la base

contenant les données que nous avons à disposition cette année ainsi que de nouvelles données, il sera donc très rapide d'obtenir une prédiction tête par tête par Machine Learning l'an prochain.

Le travail effectué sur la base Solvency nous a permis de comparer, in fine, le PNA prédit sur l'année 2019 pour les données de cette base avec le BE S2 de 2019. Cependant, comme nous l'avons vu dans la sous-partie 4.2.4, nous avons d'abord dû transformer le BE S2 en PNA afin de pouvoir le comparer au PNA prédit par ML.

Nous avons décidé d'entraîner les modèles sur une sous-base contenant les années 2016, 2017 et 2018 en divisant bien cette sous-base en une partie "train" et une partie "test". Étant donné la taille de notre base de données, nous avons décidé de garder 80% des données pour l'entraînement et 20% pour le test. Pour la séparation, nous avons veillé à ce que les lignes concernant des clients présents plusieurs années (par exemple, en 2016 et en 2017) soient bien toutes dans la même base afin de ne pas influencer les résultats de la base test si l'une des lignes était dans la base test et l'autre dans la base d'apprentissage. Le fait d'avoir plusieurs années différentes dans les deux bases ne pose pas de problème car nous n'étudions pas un phénomène chronologique, i.e., il n'y a pas de dépendance temporelle entre les différentes années d'observation.

5.2.1 Pre-processing

Notre première étape a été de réaliser un pre-processing de la base de données. Or, comme nous l'avons vu plus tôt, le data pre-processing est constitué de plusieurs étapes. Nous avons donc commencé par discrétiser nos variables qualitatives telles que la famille de produits, le risque ou le nom de l'établissement bancaire. Afin de les discrétiser, nous avons utilisé une fonction permettant de séparer la variable en plusieurs colonnes portant le nom des différentes modalités (seule une modalité n'est pas représentée car elle peut se déduire des autres). Par exemple, nous avons trois risques dans notre base de données, ainsi, en discrétisant la variable "Risque", nous avons obtenu deux colonnes : "Risque DC TC" et "Risque DC ACC". Dans le cas des variables de sexe, code société et réseau, qui ne comportent que deux modalités, nous avons effectué le traitement suivant :

- Création d'une variable nommée sexe.H qui vaut 1 si le client est un homme et 0 sinon.
- Création d'une variable soc.109 qui vaut 1 s'il s'agit de la société 109 et 0 sinon.
- Création d'une variable Reseau.BP qui vaut 1 si le client appartient au réseau Banque Populaire et 0 sinon.

Après la discrétisation de nos variables qualitatives, nous avons regardé s'il manquait des valeurs dans notre base de données et il n'en manquait pas, nous n'avons donc pas eu besoin d'utiliser les différentes méthodes permettant de remplacer les valeurs manquantes. Ensuite, nous avons pu séparer notre base de données en une base d'apprentissage et une base de test comme mentionné ci-dessus. Lors de ce split, nous avons veillé à avoir une répartition similaire des familles de produits et des risques. Pour cela, nous avons regroupé nos données par famille de produits, risque et année comptable et nous les avons ordonnées par client afin de les numéroter pour faciliter un split. Ceci nous a, par exemple, permis d'obtenir 80% des données de la famille de produits X pour le risque DC TC et l'année comptable 2018 à placer dans la base apprentissage et de garder les 20% restants dans la base test. Nous avons, par ailleurs, vérifié la distribution des variables des bases train et test, nous avons regardé si les moyennes et les écart-types des données de ces deux bases n'étaient pas trop éloignés. Il n'y avait pas d'écarts significatifs donc nous avons gardé la séparation de nos données telle quelle.

Par la suite, nous avons créé une fonction regroupant plusieurs petites fonctions et permettant alors de :

- Regarder les variables sans variance significative afin de les supprimer si nécessaire.
- Regarder les variables corrélées à d'autres afin de les supprimer si nécessaire.
- Normaliser les variables quantitatives entre 0 et 1.

Nous avons ensuite appliqué cette fonction à notre base d'apprentissage. Cela nous a permis d'obtenir la liste des colonnes à supprimer, mais également, la moyenne, l'écart-type, le minimum et le maximum de chaque variable présente dans la base d'apprentissage afin de pouvoir appliquer le même traitement à la base d'apprentissage et à la base de test. Nous avons donc créé une autre

fonction permettant de supprimer les colonnes à supprimer et permettant de normaliser, avec les mêmes coefficients, les données de la base d'apprentissage et de la base de test.

Cependant, toutes les variables que la fonction demandait de supprimer n'étaient pas nécessairement à supprimer. Voici un tableau indiquant les variables à supprimer que nous avons décidé de garder :

Variables sans variance	Variables corrélées
Provision de sinistres de début d'année	Risque DC TC
Ancienneté depuis la survenance en mois	Primes pures acquises
Top sinistre en cours l'année précédente	

TABLE 18 – Variables sans variance ou corrélées gardées

Le montant de provisions de sinistres de début d'année est une variable qui peut fortement influencer le PNA à la fin de l'année mais il a une faible variance car il est souvent nul, il n'est positif que lorsqu'il y a un sinistre en cours. En effet, si la provision de sinistres de début d'année est positive, cela signifie que le client avait un sinistre en cours l'année N-1 et il existe alors deux possibilités :

- Soit le sinistre est encore en cours à la fin de l'année N, auquel cas, il demeurera une provision ou une prestation de fin d'année N et le signe du PNA dépendra de celui de la charge de sinistre comptable. Celle-ci sera positive si la provision/prestation de fin d'année est supérieure à la provision de début d'année (le PNA sera alors négatif ou très proche de 0) et négative sinon (le PNA sera alors positif).
- Soit le sinistre n'est plus en cours à la fin de l'année N parce qu'il a été refusé, auquel cas, la charge de sinistre comptable pour l'année N sera négative et le PNA sera positif.

Les variables "Ancienneté depuis la survenance en mois" et "Top sinistre en cours l'année précédente" ont également été indiquées à supprimer car elles valent souvent 0 comme la provision de sinistres de début d'année. Nous avons décidé de garder ces deux variables pour les mêmes raisons que l'autre. En effet, toutes ces variables sans variance peuvent beaucoup aider le modèle dans sa prédiction. Le PNA de l'année N à la maille famille de produits, pour les clients ayant un sinistre en cours l'année N-1, est négatif pour la plupart des familles.

En ce qui concerne les variables corrélées, nous avons décidé de les garder car elles peuvent quand même avoir un certain impact sur le modèle malgré leur corrélation à d'autres variables. En effet, la variable "Risque DC TC" est corrélée à la variable "Risque DC ACC" et la variable "Primes pures acquises" est corrélée à la variable "Montant de primes HT" mais cela n'empêche pas qu'elles peuvent être très importantes pour le modèle. Par ailleurs, "corrélé" ne veut pas dire identique, les deux variables de risque sont corrélées car il n'y a pas énormément de clients assurés pour l'arrêt de travail dans notre base. En effet, il est toujours important d'utiliser, à la fois ce que les fonctions suggèrent et sa propre connaissance métier. Il vaut mieux garder plus de variables et laisser le modèle de ML dire qu'une variable n'est pas significative. Néanmoins, à garder beaucoup de variables non significatives, certains modèles tels que l'algorithme Random Forest, pourraient, à chaque noeud, ne tirer que des variables non significatives et ainsi, obtenir des arbres moins performants. Par ailleurs, si des variables non significatives sont beaucoup sélectionnées aux noeuds des arbres, cela augmente leur importance pour le modèle, qui en devient moins bon.

Nous allons maintenant aborder les différents modèles de Machine Learning que nous avons construits. Étant donné que nos prédictions concerneront l'année 2019, le variable "Top sinistre en cours l'année précédente" vaudra 1 si un sinistre était en cours au 31/12/2018, i.e., au 01/01/2019 et 0 si ce n'était pas le cas.

Afin de faire tourner nos modèles, nous avons une base de données contenant 29 variables explicatives telles que le montant de primes HT, les provisions de sinistres de début d'année, l'âge comptable et le top sinistre en cours l'année précédente.

5.2.2 Arbre de décision

En premier lieu, afin d'avoir une première idée de prédiction par Machine Learning, nous avons fait tourner un simple arbre de décision sur notre base d'apprentissage. Celui-ci nous a permis d'avoir une idée des variables importantes dans la détermination du PNA, en effet, nous avons pu voir les variables qu'il a utilisées pour prédire le PNA.

Voici les RMSE que nous avons obtenues sur la base d'apprentissage, de test et sur la base à prédire contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
508,41	346,96	512,07

TABLE 19 – RMSE train, test & base 2019 PNA Arbre de décision BPCE

Nous pouvons remarquer un grand écart entre la RMSE sur la base d'apprentissage et sur la base de test. En effet, dans la base de test, il y a moins de PNA négatifs en proportion que dans la base d'apprentissage, c'est pour cela que la RMSE est meilleure sur la base test. Par exemple, pour un PNA réel négatif de -2 000 et une prédiction de PNA de -1 800, le PNA prédit paraît assez proche du réel (10% d'écart), cependant, il y a tout de même un écart de 200 qui va venir fortement pénaliser la RMSE. Ainsi, la base d'apprentissage et la base de test sont comparables en RMSE seulement si la structure est la même dans ces deux bases, i.e., si nous avons la même proportion de PNA négatifs, or, ce n'est pas le cas ici.

Dès lors, ces RMSE ne présentent pas de grand intérêt seules car nous ne pouvons pas les comparer entre elles mais elles vont nous permettre de comparer les différents modèles entre eux et de retenir le meilleur modèle.

Grâce à la fonction "fancyRpartPlot" du package "rattle" de R, nous avons pu obtenir la représentation de l'arbre de décision que nous avons construit :

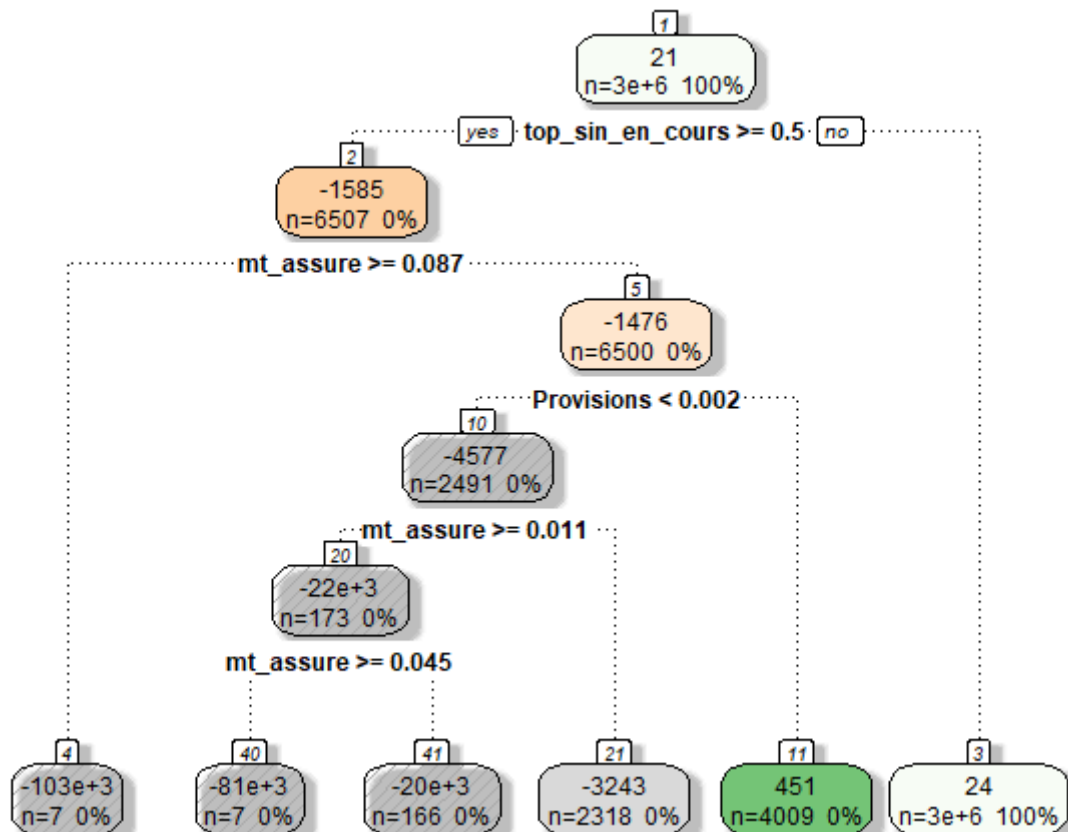


FIGURE 26 – Schéma PNA Arbre de décision BPCE

Ainsi, grâce à cette représentation de l'arbre de décision, nous pouvons voir que les variables qui ont aidé le modèle à prédire le PNA sont les suivantes :

- Top sinistre en cours l'année précédente (top_sin_en_cours sur le schéma).
- Montant assuré (mt_assure sur le schéma).
- Provision de sinistres de début d'année (Provisions sur le schéma).

Sur le schéma, nous voyons que les valeurs du montant assuré et des provisions sont très faibles, c'est parce qu'il s'agit de données qui ont été normalisées entre 0 et 1. Par ailleurs, nous pouvons remarquer que le modèle ne prédit que six valeurs différentes de PNA dont quatre négatives. Le nombre écrit dans "n=" représente le nombre d'observations présentes dans chaque noeud et feuille. Le pourcentage écrit représente le pourcentage d'observations, ici, il est toujours nul sauf au tout début de l'arbre et tout en bas à droite car il y a énormément d'observations et seule une petite quantité est gardée à chaque noeud. Ainsi, nous pouvons remarquer que la grande majorité des données a été associée à un PNA de 24.

Pour arriver à prédire un PNA de -20 000, voici le chemin que va suivre le modèle :

- Est-ce que la variable top sinistre en cours l'année précédente est supérieure ou égale à 0,5 ?
→ Oui.
- Est-ce que le montant assuré normalisé est supérieur ou égal à 0,087 ?
→ Non.
- Est-ce que la provision de sinistres de début d'année normalisée est inférieure à 0,002 ?
→ Oui.
- Est-ce que le montant assuré normalisé est supérieur ou égal à 0,011 ?
→ Oui.
- Est-ce que le montant assuré normalisé est supérieur ou égal à 0,045 ?
→ Non.

Nous remarquons que, pour le modèle, le fait d'avoir une provision de sinistres de début d'année élevée rend le PNA à la fin de l'année positif, cela veut dire que, selon le modèle, à la fin de l'année, il n'y aura plus de provisions ni de versement de prestations ou bien ils seront moindres vis à vis du montant de provisions de début d'année. Ainsi, le modèle considère que les sinistres en cours l'année N-1 se terminent souvent l'année N (soit par une prestation plus faible qu'attendue soit par un refus) ou alors la provision de fin d'année N est plus faible que la provision de début d'année N. Pour les décès, les sinistres sont rarement provisionnés, ainsi, si la provision de début d'année est positive, le sinistre a de grande chance d'être réglé pendant l'année ou d'être refusé. À contrario, les arrêts de travail peuvent s'étendre sur plusieurs années, les sinistres sont donc provisionnés pendant longtemps. En effet, la garantie AT comprend l'incapacité et l'invalidité, l'assureur peut donc être amené à verser des prestations à l'assuré jusqu'à la fin de sa vie s'il passe en invalidité.

Voici les résultats que nous avons obtenus :

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision (k€)	Ecart réel prédit
FAMILLE FRUCTI-FAMILLE	12 067	20 356	(69 %)
FRUCTI-HOMME CLE	6 131	1 541	(75 %)
FRUCTI-PROFESSIONNEL	3 522	-987	(128 %)
SECUR FAMILLE CE DC PTIA	3 500	19 819	(466 %)
FRUCTI-PATRIMOINE	1 326	-86	(107 %)
SECUR URGENCE CE	691	3 232	(367 %)
FAMILLE FRUCTI-BUDGET	522	1 069	(105 %)
DEDICACE	492	1 317	(168 %)
ASSURANCE HOMME-CLE C.E.	63	14	(78 %)
CREDIT MARITIME VIE	55	81	(48 %)
CERTI CAPITAL	36	134	(273 %)
FAMILLE FRUCTI-FACILITES	32	70	(122 %)
SECURI-CREDIT BPROP	22	42	(96 %)
FAMILLE DOUBLE SECURITE	18	49	(168 %)
FRUCTI-FACILITES PRO	7	30	(327 %)
FAMILLE FRUCTI-SECURITE	-6	2	(140 %)
FRUCTI-FACILITES AGRI	-71	41	(158 %)
D.S.EPARGNE SALARIALE	-101	199	(297 %)

TABLE 20 – Résultat PNA Arbre de décision BPCE

Comme nous pouvons le voir, le modèle n'effectue pas de bonnes prédictions mais cela est expliqué par l'inefficacité de l'utilisation d'un seul arbre. En effet, comme nous avons pu l'observer sur le schéma, le modèle n'est capable de prédire que six valeurs de PNA différentes ce qui le rend peu performant. Pour chaque famille de produits, l'écart n'est pas négligeable, cependant, le modèle prédit tout de même mieux certaines familles que d'autres. Afin de voir un peu plus pour quelles données le modèle est mauvais, nous avons décidé de regarder séparément les prédictions des données ayant un sinistre en cours l'année précédente et les prédictions des autres données.

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision BPCE (k€)	Ecart réel prédit	Nombre de sinistres en cours en 01/01/2019
FRUCTI-PROFESSIONNEL	128	-1 960	(1 635 %)	823
SECURI-CREDIT BPROP	0	6	(3 784 %)	14
FAMILLE FRUCTI-FACILITES	-4	-30	(594 %)	39
FAMILLE DOUBLE SECURITE	-5	-2	(65 %)	4
FAMILLE FRUCTI-SECURITE	-8	-3	(63 %)	2
CREDIT MARITIME VIE	-12	-6	(49 %)	19
FRUCTI-FACILITES PRO	-23	0	(102 %)	9
SECUR URGENCE CE	-48	-32	(32 %)	68
FRUCTI-FACILITES AGRI	-57	-31	(46 %)	9
DEDICACE	-60	-72	(20 %)	132
FAMILLE FRUCTI-BUDGET	-69	-37	(47 %)	181
D.S.EPARGNE SALARIALE	-197	-84	(57 %)	27
FRUCTI-PATRIMOINE	-222	-236	(6 %)	20
FRUCTI-HOMME CLE	-710	-1 092	(54 %)	124
SECUR FAMILLE CE DC PTIA	-1 078	-700	(35 %)	554
FAMILLE FRUCTI-FAMILLE	-2 098	-1 392	(34 %)	1438

TABLE 21 – Résultat PNA Arbre de décision BPCE (sinistre en cours en 2018)

Tout d'abord, pour les données avec un sinistre en cours l'année précédente, nous remarquons que le modèle capte plutôt bien que le PNA est souvent négatif même si les prédictions ne sont pas très bonnes pour autant, elles restent assez éloignées de la réalité pour certaines familles de produits.

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision BPCE (k€)	Ecart réel prédit
FAMILLE FRUCTI-FAMILLE	14 165	21 748	(54 %)
FRUCTI-HOMME CLE	6 841	2 632	(62 %)
SECUR FAMILLE CE DC PTIA	4 578	20 519	(348 %)
FRUCTI-PROFESSIONNEL	3 394	974	(71 %)
FRUCTI-PATRIMOINE	1 547	150	(90 %)
SECUR URGENCE CE	739	3 264	(342 %)
FAMILLE FRUCTI-BUDGET	591	1 105	(87 %)
DEDICACE	552	1 389	(152 %)
D.S.EPARGNE SALARIALE	96	283	(196 %)
CREDIT MARITIME VIE	67	87	(30 %)
ASSURANCE HOMME-CLE C.E.	63	14	(78 %)
FAMILLE FRUCTI-FACILITES	36	100	(179 %)
CERTI CAPITAL	36	134	(273 %)
FRUCTI-FACILITES PRO	30	29	(1 %)
FAMILLE DOUBLE SECURITE	24	51	(115 %)
SECURI-CREDIT BPROP	22	36	(66 %)
FAMILLE FRUCTI-SECURITE	2	5	(182 %)
FRUCTI-FACILITES AGRI	-13	72	(649 %)

TABLE 22 – Résultats PNA Arbre de décision BPCE (sans sinistre en cours en 2018)

En ce qui concerne les données sans sinistre en cours l'année précédente, le modèle prédit plutôt un PNA positif (tout comme le PNA réel). Ainsi, peu de clients ont eu un sinistre en 2019 et la somme des PNA réels par famille de produits est positive pour presque toutes les familles de produits, à l'exception de Fructi-Facilites Agri. Nous pouvons donc en déduire qu'un ou plusieurs sinistres sont survenus au cours de l'année 2019 pour cette famille de produits et cela a engendré des provisions et/ou des prestations de fin d'année 2019 et donc un PNA négatif. L'arbre de décision n'a pas su capter la négativité de ce PNA.

Nous allons maintenant regarder des modèles plus réputés pour leur qualité de prédictions tels que les XGBoost ou l'algorithme Random Forest.

5.2.3 XGBoost

Notre base d'apprentissage contenant un grand nombre de données, lors de la calibration des paramètres pour le XGBoost, beaucoup de modèles avec un paramétrage différent tournent et cela prend beaucoup de temps. Nous avons donc décidé de réduire le nombre de lignes en affectant des poids aux données et les paramètres optimaux ont été trouvés selon la "grid search". Après la

détermination des paramètres optimaux par le modèle, nous l'avons refait tourner avec les paramètres conservés sur l'ensemble des lignes.

Voici un schéma expliquant la "grid search" aléatoire, c'est-à-dire, la manière dont l'algorithme va piocher le meilleur jeu de paramètres :

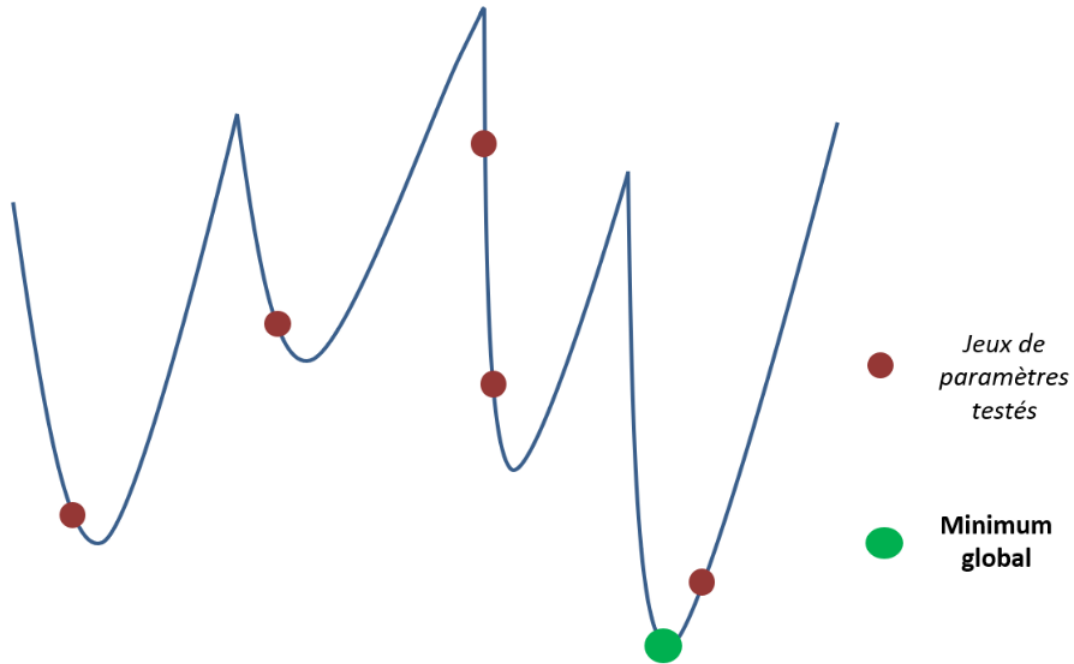


FIGURE 27 – Grid search aléatoire

Par ailleurs, voici un deuxième schéma expliquant la façon dont nous avons ajouté des poids à nos données :

Les données ayant un « top sinistre en cours année précédente » égal à 0 sont tirées sans remise

Age comptable	Ancienneté	Top sinistre en cours année précédente	Poids
26,51	1,52	0	
47,12	4,31	0	
59,31	8,22	1	1
28,47	5,17	0	2
62,03	6,53	0	2

FIGURE 28 – Attribution des poids aux données

Les données pour lesquelles la variable « Top sinistre en cours l'année précédente » valait 1 ont été associées à un poids de 1, i.e., elles n'ont pas été regroupées. À contrario, les données pour lesquelles, cette variable était nulle, ont été regroupées pour n'en conserver que la moitié. Pour ce faire, nous avons tiré aléatoirement et sans remise la moitié de ces données sans sinistre en cours l'année précédente et nous les avons associées à un poids de 2. Cela a permis de considérablement réduire le nombre d'observations (nous avons donc quasiment supprimé la moitié des observations étant donné qu'il y avait peu d'observations ayant un sinistre en cours l'année précédente).

Nous avons fait tourner un XGBoost avec un code permettant de trouver, parmi plusieurs paramètres, les paramètres optimaux pour le modèle. En entrée de cette fonction/modèle, nous avons donc dû mettre plusieurs paramètres :

- Nombre de cross-validation : Ce paramètre indique le nombre de validations croisées, i.e, le nombre de parties (folds) obtenues lors du découpage de la base de données.
- Nrounds : Il s'agit du nombre d'arbres (nombre d'itérations) à entraîner.
- Max_depth : C'est la profondeur maximale de chaque arbre. C'est un critère important dépendant du niveau d'interaction entre les variables.
- Colsample_bytree : Il s'agit de "column sampling", c'est le pourcentage des variables (colonnes) utilisées pour la construction de chaque arbre.

- **Eta** : Il s'agit du taux d'apprentissage, c'est le rétrécissement que nous faisons à chaque étape qu'il y a. Pour tirer le meilleur parti du XGBoost, le taux d'apprentissage doit être réglé aussi bas que possible. Cependant, à mesure que le taux d'apprentissage diminue, nous avons besoin de beaucoup plus d'étapes pour atteindre l'optimum. Il faut réduire la taille de pas utilisée dans la mise à jour pour éviter le sur-ajustement. Après chaque étape de "boosting", nous pouvons directement obtenir les poids des nouvelles variables et eta réduit les poids des variables pour rendre le processus de boost plus conservateur.
- **Gamma** : Il s'agit de la réduction minimale des pertes requise pour effectuer une partition supplémentaire sur un noeud de l'arbre. Plus le gamma est grand, plus l'algorithme sera conservateur. Un noeud est divisé uniquement lorsque la division résultante donne une réduction positive de la fonction de perte. Gamma spécifie la réduction de perte minimale requise pour effectuer un split.
- **Min_child_weight** : Dans une tâche de régression linéaire, cela correspond simplement au nombre minimum de données nécessaires à chaque noeud. Plus le min_child_weight est grand, plus l'algorithme sera conservateur. Cet élément est utilisé pour contrôler le sur-ajustement. Des valeurs plus élevées empêchent un modèle d'apprendre des relations qui pourraient être très spécifiques à l'échantillon particulier sélectionné pour un arbre. Cependant, des valeurs trop élevées peuvent entraîner un sous-ajustement, il doit donc être réglé à l'aide de la cross-validation.
- **Sub_sample** : Ce paramètre désigne la fraction des observations sélectionnées au hasard pour chaque arbre.
- **Bootstrap** : C'est le nombre constant de tirages que nous souhaitons effectuer sur chaque variable afin de moyenniser les performances. Le bootstrap rééchantillonne aléatoirement les données de chaque variable de test à chaque itération et regarde la perte de performance du modèle. Plus la perte de performance est élevée et plus la variable est importante pour le modèle. Il faut rééchantillonner plusieurs fois afin de le faire de manière aléatoire, un seul rééchantillonnage ne suffit pas.

Nous avons testé plusieurs jeux de paramètres différents :

`ModeleOptimalXGB_reg(App, Test, classif = FALSE, bootstrap = 20, vectpoids=vectpoids_fin, nb_cv = 10, nrounds=c(45,50), max_depth=c(2,3,4), colsample_bytree=c(1), eta=c(0.3), gamma=c(0.9,1), min_child_weight=c(5), sub_sample=c(1), chemin_stockage=chemin)`, où :

- **App** correspond à la base d'apprentissage.
- **Test** correspond à la base de test.
- **"vectpoids_fin"** correspond à un vecteur contenant les différents poids associés à nos données. Lors de la calibration, nous avons utilisé un poids de 1 sur les observations qui avaient un sinistre en cours l'année précédente et un poids de 2 sur les autres. Pour le modèle final, nous avons conservé toutes nos observations donc elles ont été associées à un poids de 1.
- **Chemin** correspond à l'arborescence où nous avons enregistré les sorties du modèle.

Nous noterons que nous avons mis des vecteurs pour certains paramètres afin que le modèle puisse tester plusieurs paramètres et choisir les meilleurs (cela correspond au grid search aléatoire mentionné au-dessus), voici le modèle qui a été retenu :

- **Bootstrap** = 20
- **Nombre de cross-validation** = 10
- **Nrounds** = 50
- **Max_depth** = 4
- **Colsample_bytree** = 1
- **Eta** = 0,3
- **Gamma** = 1
- **Min_child_weight** = 5
- **Sub_sample** = 1

Au départ, comme nous l'avons indiqué plus haut avant de rentrer dans le détail de nos modèles, nous avons 29 variables explicatives. Grâce aux paramètres choisis, le modèle a su, en quatre étapes, garder les onze variables qui lui semblaient les plus importantes. À chaque étape, un modèle

avec les meilleurs paramètres est appliqué afin d'effectuer la sélection de variables selon la perte de performance. C'est ainsi qu'il choisit les variables qu'il va conserver : Si la variable en étant rééchantillonnée aléatoirement permet d'obtenir de meilleures prédictions sur la base de test, c'est qu'elle n'est pas utile au modèle. En rééchantillonnant, nous calculons la RMSE moyenne générée sur les prédictions. En effet, une variable importante est présente dans plusieurs arbres, donc elle est souvent utilisée, si elle est rééchantillonnée, sa présence va être différente, elle va prendre des chemins différents, ainsi, la prédiction sera différente et moins bonne. Voici les variables qui ont été choisies par ordre d'importance :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	433,05	(35,87 %)
mt_assure	380,90	(19,51 %)
Provisions	347,23	(8,94 %)
soc.109	330,23	(3,61 %)
mtht_ann_complet	326,13	(2,32 %)
Anciennete_depuis_surv	324,90	(1,94 %)
RisqueDC_ACC	323,41	(1,47 %)
RisqueDC_TC	321,00	(0,71 %)
PP_acq_complet	320,74	(0,63 %)
mt_gestion_ann_complet	320,01	(0,40 %)
mt_com_ann_complet	319,18	(0,14 %)
Modèle	318,73	(0,00 %)

TABLE 23 – Variables d'importance PNA XGBoost BPCE

Sur ce tableau, nous pouvons observer l'impact de chacune des variables sur la RMSE de l'échantillon de test. Comme nous pouvions nous en douter, le modèle a gardé la variable « Top sinistre en cours l'année précédente », c'est la variable la plus importante pour le modèle. Si elle était rééchantillonnée aléatoirement, la RMSE de la base test serait beaucoup plus élevée (écart de 36% environ : 433,05 vs 318,73). La variable « Top sinistre en cours l'année précédente » est très importante car elle peut impacter le signe du PNA, ainsi, si nous rééchantillonons cette variable, cela a de fortes conséquences sur les prédictions du modèle.

Les différentes étapes du modèle sont détaillées ci-dessous au niveau de la RMSE et de la diminution du nombre de prédicteurs :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	483,22	323,37	29
Étape 2	503,64	321,84	16
Étape 3	507,15	322,76	13
Étape 4	501,54	318,73	11

TABLE 24 – Évolution de la RMSE PNA XGBoost BPCE

Nous pouvons remarquer que la RMSE de la base train a augmenté durant les trois premières étapes puis a légèrement diminué, cela signifie que le modèle ne sur-apprend pas au fur et à mesure des étapes, autrement, la RMSE train aurait diminué par rapport à sa valeur à l'étape 1 car, si un modèle prédit bien la base d'apprentissage, sa RMSE train est faible. La RMSE test, quant à elle, a diminué, ainsi, le modèle s'en sort très bien dans les prédictions malgré la diminution du nombre de variables. Dès lors, nous avons réduit le temps de calcul et la complexité du modèle sans impacter la qualité de sa prédiction.

Voici un schéma représentant l'évolution de l'erreur en fonction de la valeur d'un paramètre d'apprentissage :

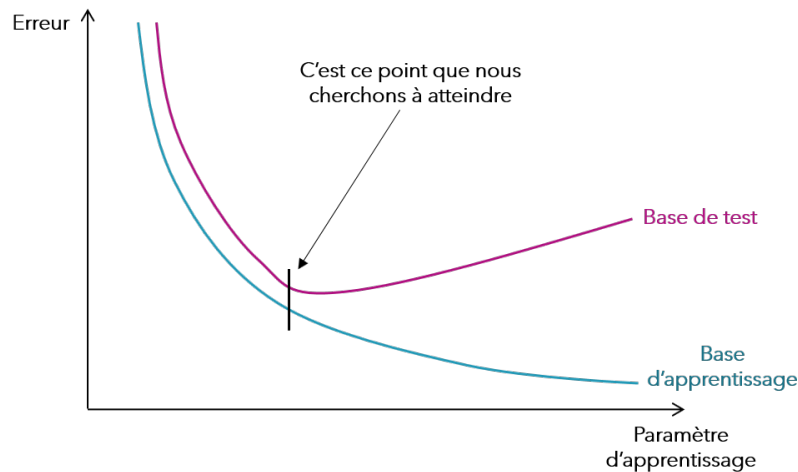


FIGURE 29 – Évolution de l'erreur en fonction de la valeur d'un paramètre d'apprentissage

Le paramètre d'apprentissage peut, par exemple, être le nombre d'arbres ou la profondeur des arbres. Ainsi, nous pouvons voir que si les arbres sont trop profonds, l'erreur sur la base d'apprentissage est très faible alors que l'erreur sur la base de test est élevée, cela représente donc le sur-ajustement aux données d'apprentissage. Il faut donc faire attention aux valeurs que nous prenons pour les paramètres. En effet, plus les arbres vont être profonds (respectivement plus il va y avoir d'arbres) et plus le modèle va risquer de se sur-ajuster aux données d'apprentissage. L'objectif est donc de diminuer l'erreur d'apprentissage jusqu'au point où l'erreur de test est minimisée.

Étant donné que le modèle avait tourné sur environ la moitié des observations seulement, nous l'avons de nouveau fait tourner sur l'ensemble des observations mais avec les paramètres retenus. Cela a permis au modèle de ne pas avoir à rechercher quels paramètres garder, il savait déjà lesquels utiliser, il a donc tourné très rapidement et n'a sélectionné que neuf variables d'importance cette fois-ci. Voici les variables qui ont été choisies par ordre d'importance :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	407,25	(27,12 %)
mt_assure	359,06	(12,08 %)
Provisions	348,47	(8,78 %)
soc.109	332,22	(3,70 %)
Anciennete_depuis_surv	329,54	(2,87 %)
PP_acq_complet	328,63	(2,58 %)
mtht_ann_complet	323,34	(0,93 %)
mt_gestion_ann_complet	323,28	(0,91 %)
RisqueDC_ACC	320,53	(0,05 %)
Modèle	320,36	(0,00 %)

TABLE 25 – Variables d'importance base totale PNA XGBoost BPCE

Ici, nous pouvons, encore une fois, observer l'impact des variables d'importance sur le modèle. Les neuf variables sélectionnées font bien partie des onze variables gardées lorsque le modèle avait tourné sur la moitié des observations mais elles ne sont pas tout à fait dans le même ordre. En effet, il en était de même lorsque les modèles ont tourné, au départ, les 29 variables étaient dans un certain ordre d'importance qui a évolué au fur et à mesure des étapes et de la suppression des variables les moins utiles. Nos modèles XGBoost ont tourné en plusieurs étapes et, à chacune de celles-ci, le modèle a cherché à éliminer un certain nombre de variables qu'il ne jugeait pas utiles pour ses prédictions, ainsi, les RMSE train et test ont évolué au fil des étapes.

Par ailleurs, les différentes étapes du modèle final sont détaillées ci-dessous au niveau des RMSE et de la diminution du nombre de prédicteurs :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	492,53	317,68	29
Étape 2	485,47	318,22	13
Étape 3	492,35	317,78	11
Étape 4	494,33	320,36	9

TABLE 26 – Évolution de la RMSE base totale PNA XGBoost BPCE

Nous remarquons que la RMSE de la base d'apprentissage augmente légèrement à partir de l'étape 2 et cela signifie que les prédictions sont un peu moins bonnes avec la diminution du nombre de prédicteurs. De plus, la RMSE de la base de test a tendance à rester à peu près stable, nous pouvons remarquer qu'elle diminue légèrement de l'étape 2 à l'étape 3 et ré-augmente un peu juste après, cela signifie que le modèle fait un peu plus d'erreurs mais il faut noter qu'il a beaucoup été simplifié par la réduction du nombre de variables car nous sommes passés de 29 variables à seulement 9 soit environ un tiers des variables d'origine sans pour autant avoir une énorme augmentation de la RMSE (l'augmentation est d'à peine 1%). Nous avons donc grandement simplifié le modèle en échange d'une très faible perte de précision.

Au final, voici les RMSE que nous avons obtenues sur la base d'apprentissage, de test et sur la base à prédire contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
494,33	320,36	426,23

TABLE 27 – RMSE train, test & base 2019 PNA XGBoost BPCE

Nous pouvons remarquer que toutes ces RMSE sont plus faibles que celles obtenues pour l'arbre de décision, dès lors, le XGBoost est plus performant que l'arbre de décision, comme nous pouvions nous en douter.

Nous avons obtenu les prédictions suivantes :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BPCE OLD (k€)	Ecart réel prédit
FAMILLE FRUCTI-FAMILLE	12 067	11 933	(1 %)
FRUCTI-HOMME CLE	6 131	5 940	(3 %)
FRUCTI-PROFESSIONNEL	3 522	3 225	(8 %)
SECUR FAMILLE CE DC PTIA	3 500	7 235	(107 %)
FRUCTI-PATRIMOINE	1 326	1 010	(24 %)
SECUR URGENGE CE	691	879	(27 %)
FAMILLE FRUCTI-BUDGET	522	391	(25 %)
DEDICACE	492	424	(14 %)
ASSURANCE HOMME-CLE C.E.	63	36	(43 %)
CREDIT MARITIME VIE	55	57	(5 %)
CERTI CAPITAL	36	31	(15 %)
FAMILLE FRUCTI-FACILITES	32	-6	(118 %)
SECURI-CREDIT BPROP	22	19	(12 %)
FAMILLE DOUBLE SECURITE	18	9	(50 %)
FRUCTI-FACILITES PRO	7	14	(94 %)
FAMILLE FRUCTI-SECURITE	-6	-7	(15 %)
FRUCTI-FACILITES AGRI	-71	60	(185 %)
D.S.EPARGNE SALARIALE	-101	51	(151 %)

TABLE 28 – Résultats PNA XGBoost BPCE

Nous avons remarqué un grand écart de prédiction pour l'une des familles de produits les plus importantes, il s'agit d'une famille appartenant au réseau Caisse d'Épargne (Secur Famille CE DC PTIA). Les écarts que nous avons constatés, que ce soit pour cette famille ou pour les autres, peuvent être expliqués par le manque d'historique. En effet, par exemple, pour les contrats commercialisés par les Caisses d'Épargne, l'historique est trop court, les premiers contrats datent de 2016. Dès lors, le modèle de Machine Learning pour le PNA n'utilisant pas les données des affaires nouvelles, ne prend en compte, dans sa création, seulement les contrats présents au 01/01/2017 et ceux présents au 01/01/2018. En effet, nous rappelons que les modèles de Machine Learning ont

été construits sur les années 2016 à 2018. Or, 2016 étant la première année de commercialisation des contrats par les Caisses d'Épargne, ces contrats sont des affaires nouvelles en 2016 et ne font donc pas partie des bases train et test des modèles de ML pour l'année 2016. En effet, le portefeuille n'est pas encore en régime de croisière sur ce réseau et il n'y a donc pas assez d'historique de sinistralité pour bien prédire, en particulier, à cause de la montée en charge du portefeuille. En ce qui concerne les écarts sur les autres familles, ils sont moins visibles et peuvent être dus au manque de données ou d'historique pour les familles de produits en question. Même si les Banques Populaires commercialisent depuis longtemps nos contrats de prévoyance individuelle, certaines familles de produits sont récentes ou possèdent un faible portefeuille de clients.

Afin d'essayer d'obtenir de meilleures prédictions pour les Caisses d'Épargne, nous avons décidé de faire tourner un modèle GLM ainsi qu'un modèle XGBoost exclusivement sur ce réseau (que nous verrons un peu plus loin). Par ailleurs, étant donné que nous voulions faire tourner un XGBoost uniquement sur le réseau Caisse d'Épargne, nous avons décidé d'en faire tourner un également sur les Banques Populaires afin d'analyser si les prédictions obtenues pouvaient être meilleures que celles du modèle XGBoost sur l'ensemble des réseaux. Avant de faire tourner ces nouveaux modèles, nous avons décidé d'ajouter de nouvelles variables pouvant aider les prédictions des modèles, voici les variables que nous avons ajoutées :

- Ratio S/PHT l'année N-1 : Il s'agit du ratio sinistres sur primes HT de l'année N-1.
- Ratio S/PPacq l'année N-1 : Il s'agit du ratio sinistres sur primes pures acquises de l'année N-1.
- Ratio C/PHT l'année N-1 : Il s'agit du ratio charge de sinistres comptable sur primes HT de l'année N-1.
- Ratio C/PPacq l'année N-1 : Il s'agit du ratio charge de sinistres comptable sur primes pures acquises de l'année N-1.
- Part du PNA par famille l'année N-1 : Par exemple, s'il y a 10 clients pour une famille de produits X et que le PNA total de cette famille est 150 pour l'année N-1, la part du PNA pour un client de cette famille sera égale à un dixième de 150 soit 15.
- Part du PNA par risque l'année N-1 : Il s'agit du même indicateur qu'au-dessus mais en prenant la part du PNA par risque et non pas par famille de produits.

Ainsi, les modèles auront maintenant 35 prédicteurs possibles au lieu de 29. Nous avons tenté de réaliser un nouveau XGBoost sur les réseaux BP et CE avec les nouvelles variables afin de regarder s'il décidait de garder les nouvelles variables ou non. Ceci nous a également permis de comparer, par la suite, les deux XGBoost BPCE afin de ne garder que le meilleur modèle.

5.2.4 XGBoost avec les nouvelles variables

Comme pour le premier modèle XGBoost sur l'ensemble des réseaux, nous avons donné plusieurs jeux de paramètres au modèle (les mêmes) afin de voir quels paramètres le modèle allait garder cette fois-ci et nous avons ajouté des poids à certaines de nos observations afin de réduire la taille de la base de données, de la même manière que pour le premier XGBoost. Nous avons appliqué le même algorithme de sélection de variables que précédemment. Le modèle a retenu les paramètres suivants :

- Bootstrap = 20
- Nombre de cross-validation = 10
- Nrounds = 50
- Max_depth = 4
- Colsample_bytree = 1
- Eta = 0,3
- Gamma = 0,9
- Min_child_weight = 5
- Sub_sample = 1

Le seul paramètre qui a changé est la valeur du gamma, cette fois-ci, le modèle a choisi un gamma égal à 0,9 alors qu'il l'avait placé à 1 sur le premier XGBoost. La valeur de gamma est plus petite ce qui va rendre le modèle moins conservateur. Lorsque le modèle a tourné sur environ la moitié des données, il a sélectionné les variables d'importance suivantes (parmi les 35 possibles) :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	430,38	(34,05 %)
mt_assure	352,50	(9,79 %)
Provisions	344,74	(7,37 %)
mtht_ann_complet	328,31	(2,25 %)
PP_acq_complet	327,70	(2,07 %)
ChargesPPacq	326,00	(1,54 %)
Anciennete	324,32	(1,01 %)
part_PNA_fam_N_1	324,16	(0,96 %)
SPHT	323,99	(0,91 %)
Anciennete_depuis_surv	323,66	(0,81 %)
soc.109	323,07	(0,63 %)
mt_gestion_ann_complet	322,22	(0,36 %)
part_PNA_risq_N_1	322,04	(0,30 %)
mt_com_ann_complet	322,02	(0,30 %)
ChargesPHT	321,95	(0,28 %)
RisqueDC_ACC	321,76	(0,21 %)
RisqueDC_TC	321,66	(0,18 %)
Modèle	321,07	(0,00 %)

TABLE 29 – Variables d'importance XGBoost PNA BPCE NEW

Nous pouvons voir que le modèle a gardé 17 variables, soit la moitié des variables, ce qui est non négligeable. La variable "Top sinistre en cours l'année précédente" est toujours la variable la plus importante pour le modèle. Parmi les variables que nous avons ajoutées, le modèle a gardé toutes les nouvelles variables à l'exception du ratio S/PPacq l'année N-1.

Voici les différentes étapes que le modèle a suivi afin de réduire au maximum le nombre de variables sans impacter la qualité des prédictions :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	523,87	317,15	35
Étape 2	523,18	320,14	18
Étape 3	523,6	321,07	17

TABLE 30 – Évolution de la RMSE PNA XGBoost BPCE NEW

Pour ce modèle, la RMSE train est restée très stable au fur et à mesure des étapes alors que la RMSE test a légèrement augmenté donc le modèle a un peu perdu en qualité de prédiction avec la réduction du nombre de variables.

Nous avons ensuite refait tourner le modèle sur toutes les observations avec les paramètres retenus et nous avons obtenu les variables d'importance suivantes :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	427,69	(34,59 %)
mt_assure	349,84	(10,09 %)
Provisions	345,08	(8,59 %)
SPHT	324,81	(2,21 %)
PP_acq_complet	322,69	(1,55 %)
part_PNA_risq_N_1	322,60	(1,52 %)
ChargesPPacq	322,61	(1,52 %)
soc.109	321,72	(1,24 %)
ChargesPHT	320,86	(0,97 %)
mtht_ann_complet	320,52	(0,86 %)
Anciennete_depuis_surv	320,25	(0,78 %)
mt_gestion_ann_complet	319,31	(0,48 %)
mt_com_ann_complet	318,35	(0,18 %)
RisqueDC_TC	318,06	(0,09 %)
part_PNA_fam_N_1	317,93	(0,05 %)
Modèle	317,78	(0,00 %)

TABLE 31 – Variables d'importance base totale PNA XGBoost BPCE NEW

Les 15 variables sélectionnées cette fois-ci font partie des 17 variables d'importance qui avaient été retenues lorsque le modèle avait tourné sur la moitié des observations. Cependant, encore une fois, l'ordre d'importance a été légèrement modifié.

Lors de la construction du modèle, les étapes suivantes ont été réalisées :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	489,61	319,95	35
Étape 2	491,37	317,77	16
Étape 3	491,37	317,77	15

TABLE 32 – Évolution de la RMSE base totale PNA XGBoost BPCE NEW

Pour ce modèle-ci, la RMSE train a légèrement augmenté et la RMSE test a légèrement diminué au fil des étapes. Ainsi, beaucoup de variables ont été supprimées et cela n'a pas impacté la performance du modèle.

Nous avons eu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
491,37	317,77	425,47

TABLE 33 – RMSE train, test & base 2019 PNA XGBoost BPCE NEW

Il y a peu de différences entre ces RMSE et celles obtenues avec le modèle XGBoost précédent, cependant, elles sont tout de même légèrement meilleures.

Grâce à ce modèle, nous avons obtenu les prédictions de PNA suivantes :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BPCE NEW (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	12 004	(1 %)
FRUCTI-HOMME CLE	6 131	6 153	(0 %)
FRUCTI-PROFESSIONNEL	3 522	4 029	(14 %)
SECUR FAMILLE CE DC PTIA	3 500	6 935	(98 %)
FRUCTI-PATRIMOINE	1 326	1 022	(23 %)
SECUR URGENCE CE	691	991	(43 %)
FAMILLE FRUCTI-BUDGET	522	336	(36 %)
DEDICACE	492	493	(0 %)
ASSURANCE HOMME-CLE C.E.	63	56	(12 %)
CREDIT MARITIME VIE	55	57	(4 %)
CERTI CAPITAL	36	40	(10 %)
FAMILLE FRUCTI-FACILITES	32	-15	(146 %)
SECURI-CREDIT BPROP	22	16	(25 %)
FAMILLE DOUBLE SECURITE	18	10	(45 %)
FRUCTI-FACILITES PRO	7	11	(59 %)
FAMILLE FRUCTI-SECURITE	-6	-8	(43 %)
FRUCTI-FACILITES AGRI	-71	59	(184 %)
D.S.EPARGNE SALARIALE	-101	38	(137 %)

TABLE 34 – Résultats PNA XGBoost BPCE NEW

À première vue, ce nouveau modèle sur les réseaux BP et CE a été à peu près aussi performant que le premier. Encore une fois, il n'a pas su bien prédire Secur Famille CE DC PTIA. Comme nous l'avons évoqué plus tôt, nous avons donc tenté de réaliser un GLM sur les deux réseaux confondus, dans un premier temps, et uniquement sur le réseau CE dans un second temps.

5.2.5 Generalized Linear Model loi gaussienne BPCE

Avant de réaliser un GLM sur les réseaux BP et CE, nous avons regardé comment était distribué le PNA, nous avons donc réalisé un plot de sa fréquence en fonction de sa valeur. Par ailleurs, pour créer les GLM, nous avons ajouté la variable "année comptable" aux variables explicatives car cela peut les aider. En effet, le GLM va essayer de capter une tendance linéaire pour faire de bonnes prédictions. Comme le portefeuille CE est en montée en charge, la sinistralité évolue et l'année

comptable pourrait capter cet effet. Pour le XGBoost, il n'aurait pas été possible d'utiliser cette variable car, sur les données d'apprentissage, il n'aurait eu que les années 2016 à 2018, ainsi, en lui fournissant les données de 2019 pour la prédiction, il aurait classé toutes les valeurs dans la branche de l'arbre " ≥ 2018 ", par exemple, et cela n'aurait donc rien apporté au modèle.

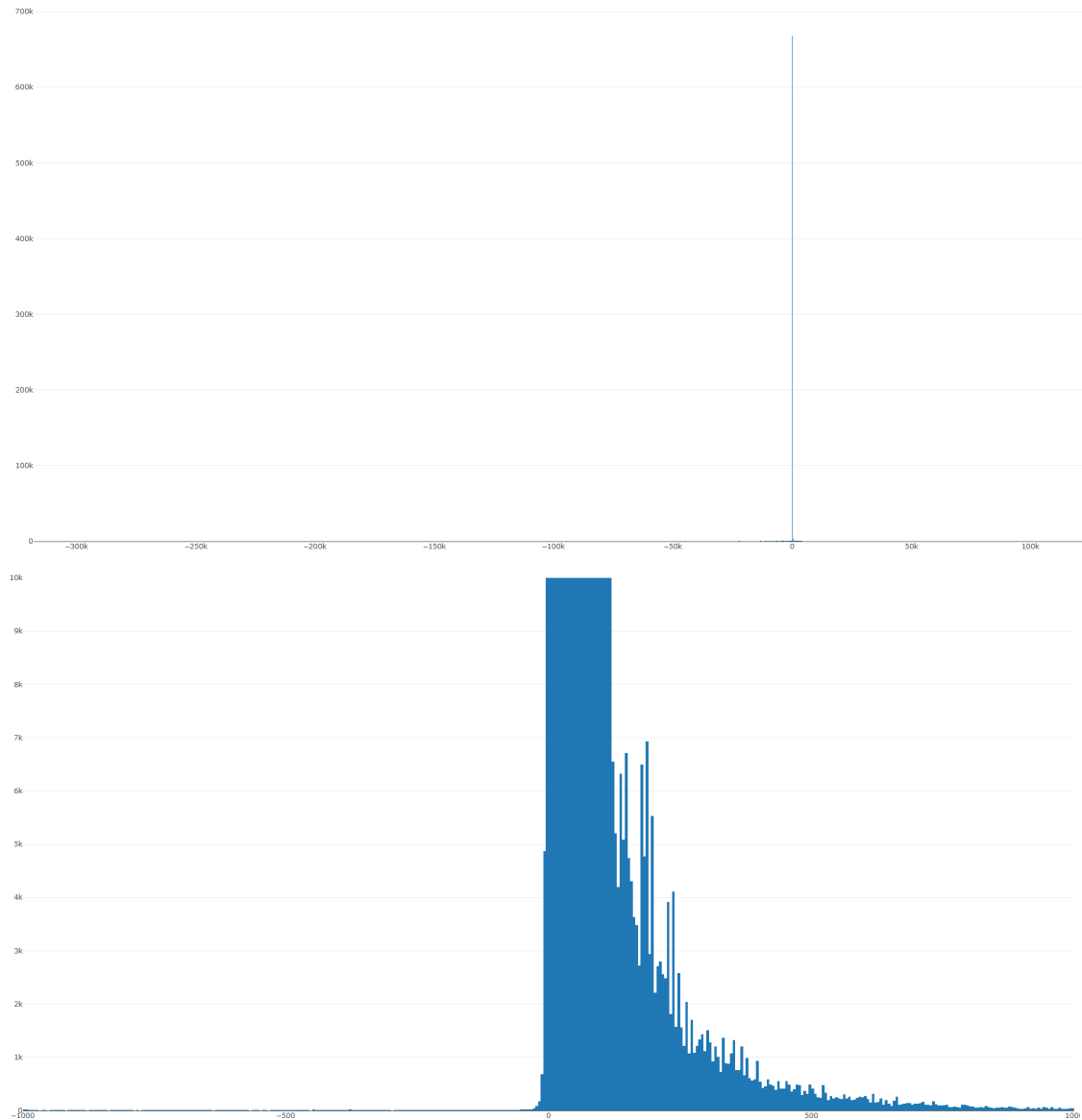


FIGURE 30 – Plot du PNA en fonction de sa fréquence BPCE

Le premier plot était trop éloigné pour réellement comprendre la forme que suit le PNA, nous avons donc effectué un zoom mais, étant donné que le PNA est souvent peu éloigné de 0, il y a un pic très élevé au niveau de la fréquence des valeurs autour de 0, on ne voit donc pas très bien le sommet de la courbe sur le plot zoomé. Cependant, nous pouvons tout de même remarquer la tendance que suit la courbe et cela pourrait, par exemple, être modélisé par une loi Gamma. Toutefois, pour ce type de lois, nous ne pouvons pas avoir de valeurs négatives, or, le PNA est parfois négatif, il faut donc effectuer une translation de nos données si nous voulons utiliser une autre loi que la loi gaussienne, qui est la loi standard utilisée pour les GLM. En effet, la fonction de lien associée à la loi Gamma est la fonction inverse et celle-ci n'est pas définie sur \mathbb{R}^- .

Ainsi, dans un premier temps, nous avons utilisé la loi gaussienne pour le GLM et nous avons obtenu le modèle suivant :

```
> summary(method_glm_model1)

call:
NULL

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-315496    -1         3        10    95801

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.038e+05  1.308e+06  -0.385  0.700033
annee_cpt    -4.062e+00  1.172e+00  -3.464  0.000532 ***
soc.109      -1.034e+01  1.349e+00  -7.663  1.82e-14 ***
Reseau_BP    -2.051e+00  3.046e+00  -0.673  0.500692
nom_fam_techDEDICACE  4.272e+00  2.156e+00  1.982  0.047516 *
nom_fam_techFAMILLE_FRUCTI_FAMILLE  2.834e+00  1.598e+00  1.774  0.076066 .
nom_fam_techFRUCTI_HOMME_CLE  8.241e+00  2.096e+00  3.932  8.44e-05 ***
nom_fam_techSECUR_FAMILLE_CE_DC_PTIA  8.231e-01  2.461e+00  0.334  0.738048
RisqueDC_ACC -5.510e+01  5.041e+00 -10.930 < 2e-16 ***
RisqueDC_TC  -5.617e+01  4.785e+00 -11.739 < 2e-16 ***
sexe.H       -5.652e+00  7.386e-01  -7.653  1.96e-14 ***
periodicite  2.508e+00  1.121e+00  2.236  0.025341 *
mt_assure    -1.330e+02  5.817e+01  -2.286  0.022248 *
mtht_ann_complet  2.499e+04  2.291e+02  109.105 < 2e-16 ***
mt_com_ann_complet -3.262e+03  1.217e+02  -26.798 < 2e-16 ***
mt_gestion_ann_complet  7.150e+02  1.782e+02  4.013  6.00e-05 ***
Anciennete   8.748e+05  2.271e+06  0.385  0.700068
AgeAdh       2.499e+06  6.487e+06  0.385  0.700069
AgeExe       -2.214e+06  5.746e+06  -0.385  0.700070
Provisions   3.146e+04  1.257e+02  250.302 < 2e-16 ***
Anciennete_depuis_surv -5.913e+02  7.016e+01  -8.428 < 2e-16 ***
top_sin_en_cours -2.701e+03  1.029e+01 -262.352 < 2e-16 ***
nom_etabBANQUE_POPULAIRE_ALSACE_LORRAINE_CHAMPAGNE  1.558e+00  1.580e+00  0.986  0.324338
nom_etabBANQUE_POPULAIRE_AUVERGNE_RHONE_ALPES -8.191e-01  1.489e+00  -0.550  0.582335
nom_etabBANQUE_POPULAIRE_BOURGOGNE_FRANCHE_COMTE  1.124e+00  1.602e+00  0.702  0.482780
nom_etabBANQUE_POPULAIRE_DU_SUD  2.899e-04  1.729e+00  0.000  0.999866
nom_etabBANQUE_POPULAIRE_GRAND_QUEST  8.549e-01  1.533e+00  0.558  0.576965
nom_etabBANQUE_POPULAIRE_OCCITANE  6.833e+00  1.751e+00  3.903  9.50e-05 ***
nom_etabBANQUE_POPULAIRE_RIVES_DE_PARIS  5.857e+00  1.581e+00  3.705  0.000211 ***
nom_etabBANQUE_POPULAIRE_VAL_DE_FRANCE  1.085e+00  1.721e+00  0.631  0.528344
SPHT         8.702e+04  1.984e+03  43.861 < 2e-16 ***
SPPacq      -8.969e+04  2.052e+03 -43.712 < 2e-16 ***
ChargesPHT  -5.163e+03  1.358e+03  -3.801  0.000144 ***
ChargesPPacq  1.111e+04  1.322e+03  8.403 < 2e-16 ***
part_PNA_fam_N_1  3.460e+01  7.666e+00  4.514  6.37e-06 ***
part_PNA_risq_N_1  6.520e+01  6.070e+00  10.740 < 2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 385774)

Null deviance: 1.2028e+12 on 2961137 degrees of freedom
Residual deviance: 1.1423e+12 on 2961102 degrees of freedom
AIC: 46492522

Number of Fisher Scoring iterations: 2
```

FIGURE 31 – Summary PNA GLM BPCE loi gaussienne

Sur cette image, nous pouvons voir les coefficients (colonne "Estimate") que le modèle a attribué à chaque variable et leur importance grâce aux symboles sur la droite du summary, pour les plus importantes, il y a le symbole : "****". Encore une fois, la variable "Top sinistre en cours l'année précédente" est retenue comme une variable très importante pour le modèle. Par ailleurs, nous pouvons remarquer que beaucoup de variables sont importantes pour ce modèle d'après les symboles, contrairement à ce que nous avons pu observer pour les XGBoost qui gardent souvent moins de la moitié des variables d'origine.

Nous avons réalisé ce modèle car nous voulions regarder s'il captait plutôt bien la valeur du PNA ou non. Nous avons commencé par comparer le PNA total prédit (environ 25 millions) et le PNA total réel en somme (environ 28 millions), il n'y avait pas beaucoup d'écart donc nous avons regardé les différents PNA un peu plus en détail. Pour cela, nous avons décidé de séparer la base 2019 en trois bases :

- Une première base contenant les contrats sinistrés avant l'année de prédiction du PNA (sinistres survenus avant l'année 2019 et supposés connus). En effet, peu de sinistres survenus avant 2019 ont été déclarés après le 01/01/2019 (et sont donc inconnus à cette date), ainsi, nous avons pris en compte tous les sinistres survenus avant 2019.

- Une deuxième base contenant les contrats sinistrés pendant l'année de prédiction du PNA (sinistres inconnus au 01/01/2019 car survenus pendant l'année 2019).
- Une troisième base contenant les contrats non sinistrés (ou dont le sinistre n'a pas encore été déclaré) au 31/12/2019.

Voici les résultats que nous avons obtenus pour ces trois différentes bases :

	PNA réel	PNA prédit par GLM loi gaussienne BPCE (k€)
Contrats sinistrés avant 2019	- 4 080	- 5 111
Contrats sinistrés en 2019	- 13 583	217
Contrats non sinistrés à fin 2019	45 969	29 520

TABLE 35 – Résultats PNA GLM BPCE loi gaussienne : Contrats sinistrés et non sinistrés

Nous avons ainsi pu comparer le PNA total réel et prédit pour ces trois différentes bases :

- Pour la base contenant les sinistres connus au 01/01/N, le PNA total prédit était négatif tout comme le PNA total réel. Par ailleurs, au niveau des valeurs, nous n'étions pas trop éloignés.
- Pour la base contenant les sinistres inconnus au 01/01/N et survenus l'année N, le PNA total prédit était positif mais proche de 0 alors que le PNA total réel était négatif, cela a résulté en un grand écart de valeurs.
- Pour la base contenant les contrats non sinistrés au 31/12/N, le PNA total prédit était positif tout comme le PNA total réel, cependant, les deux valeurs étaient très éloignées.

Nous avons donc pu remarquer que le modèle captait assez bien le signe du PNA. En effet, même si, pour les contrats sinistrés l'année 2019, il a prédit un PNA positif au total, il s'agissait d'un PNA proche de 0 contrairement au PNA obtenu pour les contrats non sinistrés qui était très éloigné de 0. Par ailleurs, en regardant la valeur du PNA par tête, la prédiction n'était pas très bonne mais c'est compréhensible, le modèle ne peut pas prédire le PNA parfait par tête.

Nous avons eu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
511,10	330,17	452,20

TABLE 36 – RMSE train, test & base 2019 PNA GLM BPCE loi gaussienne

Par rapport aux XGBoost BPCE précédents, toutes les RMSE sont plus élevées, cela signifie que les prédictions du GLM BPCE sont moins bonnes, au global, que les prédictions des XGBoost peu importe la base considérée.

Voici les résultats que nous avons obtenus par famille de produits :

Famille de produits	PNA réel (k€)	PNA prédit par GLM loi gaussienne BPCE (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	10 776	(11 %)
FRUCTI-HOMME CLE	6 131	5 969	(3 %)
FRUCTI-PROFESSIONNEL	3 522	4 153	(18 %)
SECUR FAMILLE CE DC PTIA	3 500	1 663	(52 %)
FRUCTI-PATRIMOINE	1 326	1 354	(2 %)
SECUR URGENCE CE	691	78	(89 %)
FAMILLE FRUCTI-BUDGET	522	22	(96 %)
DEDICACE	492	315	(36 %)
ASSURANCE HOMME-CLE C.E.	63	44	(30 %)
CREDIT MARITIME VIE	55	34	(39 %)
CERTI CAPITAL	36	25	(32 %)
FAMILLE FRUCTI-FACILITES	32	-57	(279 %)
SECURI-CREDIT BPROP	22	-3	(113 %)
FAMILLE DOUBLE SECURITE	18	7	(61 %)
FRUCTI-FACILITES PRO	7	1	(87 %)
FAMILLE FRUCTI-SECURITE	-6	-4	(34 %)
FRUCTI-FACILITES AGRI	-71	66	(194 %)
D.S.EPARGNE SALARIALE	-101	184	(281 %)

TABLE 37 – Résultats PNA GLM BPCE loi gaussienne

Sur ces résultats, nous remarquons que, même s'il ne s'agit pas d'un modèle uniquement sur le réseau CE, il a mieux prédit Secur Famille CE DC PTIA que nos deux XGBoost BPCE. Comme nous l'avions évoqué plus tôt, il y a une montée en charge des Caisses d'Épargne, c'est-à-dire que les sinistres vont augmenter pendant les premières années car les contrats ne sont pas ouverts depuis longtemps. Ainsi, il y a un effet "linéaire" de l'augmentation des sinistres ce qui permet probablement au GLM, grâce à la variable d'importance "année comptable", de mieux prédire ce réseau. Nous avons tout de même construit un GLM uniquement sur le réseau CE afin de voir si nous pouvions avoir de meilleures prédictions. En effet, la prédiction pour Secur Urgence CE n'était pas très bonne avec ce premier GLM.

5.2.6 Generalized Linear Model loi gaussienne CE

Les modèles qui ont tourné sur le réseau CE n'avaient pas tout à fait les mêmes variables explicatives en entrée que les modèles qui ont tourné sur les deux réseaux. En effet, les variables des familles de produits gardées commençant par "nom_fam_tech" ne pouvaient concerner que les familles de produits du réseau CE et les variables de nom d'établissement bancaire gardées commençant par "nom_etab" ne pouvaient concerner que les établissements bancaires du réseau CE. Les modèles qui ont tourné sur le réseau avaient donc 32 variables explicatives (+ l'année comptable pour les GLM).

Pour le GLM sur le réseau CE, nous avons tout d'abord regardé un plot du PNA comme ce que nous avons fait sur les réseaux BP et CE et nous avons pu remarquer, à nouveau, que la plupart des valeurs du PNA étaient proches de 0. Voici le plot que nous avons obtenu, en zoomant :

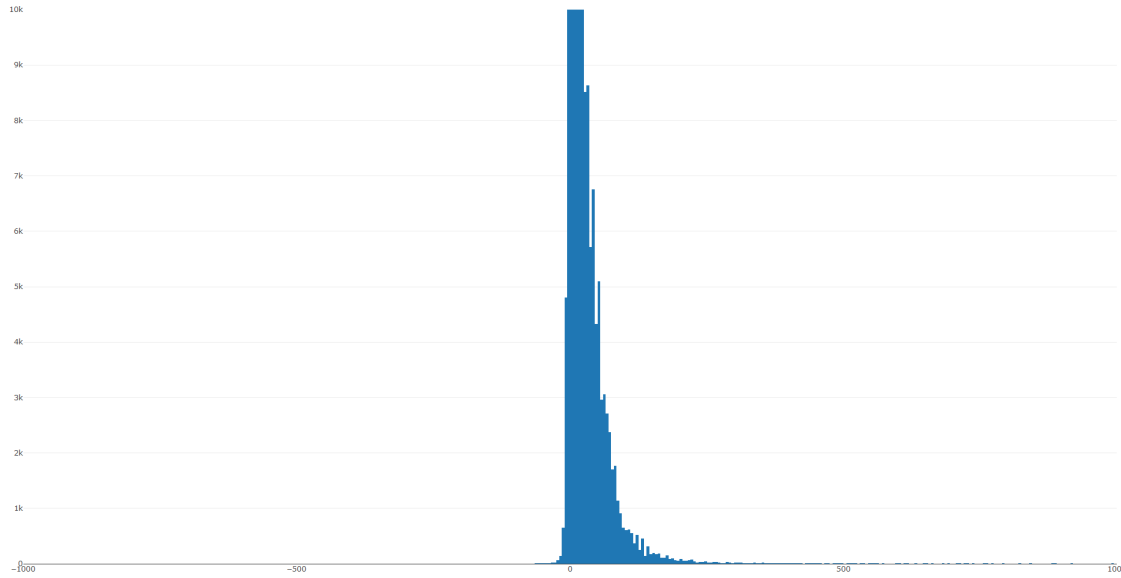


FIGURE 32 – Plot du PNA en fonction de sa fréquence CE

La distribution du PNA ressemble à celle que nous avons au total sur BPCE, ainsi, la distribution d'une loi Gamma correspondrait plus à notre distribution qu'une loi gaussienne donc nous avons décidé de faire tourner un GLM avec une loi gaussienne, dans un premier temps, puis de translater nos données pour faire tourner un GLM avec une loi Gamma.

Voici un exemple de loi Gamma :

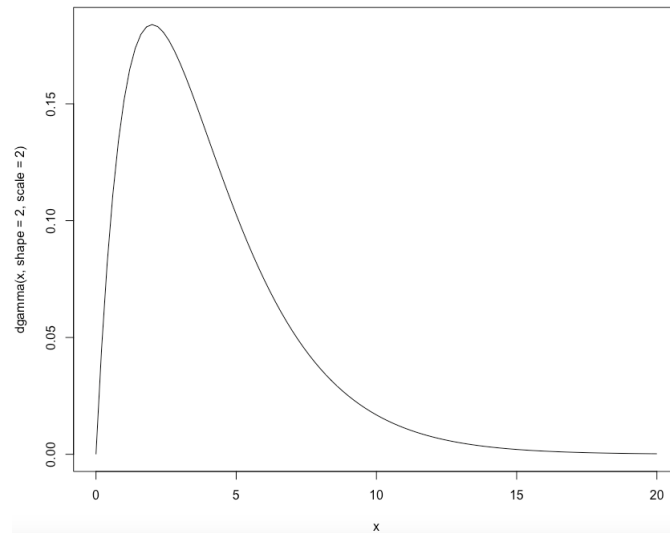


FIGURE 33 – Distribution loi Gamma

Nous pouvons voir une ressemblance entre notre plot du PNA et cette distribution de loi $Gamma(k = 2, \theta = 2)$. Nous verrons donc dans la sous-partie suivante les résultats obtenus avec un GLM associé à une loi Gamma.

Dans le cadre du GLM avec une loi gaussienne, nous avons obtenu le modèle suivant :

```
> set.seed(20)
> GLM <- train(formule, data = App, method="glm", trControl=trainControl(method="none"))
> summary(GLM)

Call:
NULL

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-44952    -1         2         5    4949

Coefficients: (2 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.035e+03  1.700e+04  -0.355  0.72263
annee_cpt    -2.405e+00  1.255e+00  -1.916  0.05543 .
soc.109      -7.497e+00  1.690e+01  -0.444  0.65734
nom_fam_techSECUR_FAMILLE_CE_DC_PTIA  5.495e+00  1.874e+00  2.932  0.00336 **
RisqueDC_ACC  1.195e+02  9.163e+01  1.304  0.19234
RisqueDC_TC   1.282e+02  9.139e+01  1.403  0.16073
sexe.H       -3.036e+00  5.237e-01  -5.797  6.76e-09 ***
periodicite  4.420e-01  1.544e+00  0.286  0.77472
mt_assure    -5.952e+02  7.945e+01  -7.492  6.78e-14 ***
mtht_ann_complet  4.960e+03  5.112e+02  9.704  < 2e-16 ***
mt_com_ann_complet  5.226e+02  2.515e+02  2.078  0.03768 *
mt_gestion_ann_complet -9.786e+02  6.062e+02  -1.614  0.10645
Anciennete   5.630e+04  1.623e+05  0.347  0.72871
AgeAdh       1.379e+06  3.977e+06  0.347  0.72871
AgeExe      -1.422e+06  4.099e+06  -0.347  0.72871
Provisions   4.542e+03  9.050e+01  50.184  < 2e-16 ***
Anciennete_depuis_surv  5.399e+03  7.261e+01  74.360  < 2e-16 ***
top_sin_en_cours -5.444e+03  2.389e+01 -227.900  < 2e-16 ***
nom_etabCAISSE_D_EPARGNE_AQUITAINE_POITOU_CHARENTES  3.110e+00  1.146e+00  2.714  0.00665 **
nom_etabCAISSE_D_EPARGNE_BRETAGNE_PAYS_DE_LOIRE -1.153e+00  9.762e-01  -1.181  0.23754
nom_etabCAISSE_D_EPARGNE_COTE_D_AZUR  1.121e+00  1.297e+00  0.865  0.38725
nom_etabCAISSE_D_EPARGNE_DE_BOURGOGNE_FRANCHE_COMTE  7.138e-01  1.126e+00  0.634  0.52616
nom_etabCAISSE_D_EPARGNE_DE_PROVENCE_ALPES_CORSE  2.704e+00  1.187e+00  2.279  0.02269 *
nom_etabCAISSE_D_EPARGNE_GRAND_EST_EUROPE  1.977e+00  1.095e+00  1.806  0.07091 .
nom_etabCAISSE_D_EPARGNE_ILE_DE_FRANCE  7.731e-01  9.092e-01  0.850  0.39516
nom_etabCAISSE_D_EPARGNE_LOIRE_CENTRE  2.011e+00  1.184e+00  1.698  0.08958 .
nom_etabCAISSE_D_EPARGNE_RHONE_ALPES  2.191e+00  1.002e+00  2.186  0.02882 *
SPHT         2.114e+03  2.010e+02  10.522  < 2e-16 ***
ChargesPHT   1.871e+03  8.595e+02  2.177  0.02946 *
ChargesPPacq  1.736e+03  8.592e+02  2.020  0.04338 *
part_PNA_fam_N_1  2.055e+01  1.319e+01  1.558  0.11922
part_PNA_risq_N_1  1.886e+01  1.393e+01  1.353  0.17596
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 39882.47)

Null deviance: 2.5997e+10 on 582380 degrees of freedom
Residual deviance: 2.3226e+10 on 582349 degrees of freedom
AIC: 7822325

Number of Fisher scoring iterations: 2
```

FIGURE 34 – Summary PNA GLM CE loi gaussienne

Comme pour les précédents modèles, nous remarquons que la variable "Top sinistre en cours l'année précédente" est toujours parmi les variables les plus importantes, cependant, ce modèle, contrairement à l'autre GLM sur BPCE a sélectionné moins de variables d'importance.

Nous avons eu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
199,7	141,54	188,24

TABLE 38 – RMSE train, test & base 2019 PNA GLM CE loi gaussienne

Ici, les RMSE obtenues ont été calculées sur des données où la proportion de PNA négatifs est plus faible que sur les deux réseaux car les CE ne sont pas encore en régime de croisière, comme nous l'avons déjà évoqué. Ainsi, ces RMSE ne pourront être comparées qu'aux modèles réalisés uniquement sur les CE.

Voici les résultats que nous avons obtenus par famille de produits :

Famille de produits	PNA réel (k€)	PNA prédit par GLM loi gaussienne CE (k€)	Ecart réel VS prédit
SECUR FAMILLE CE DC PTIA	3 500	1 373	(61 %)
SECUR URGENCE CE	691	270	(61 %)
ASSURANCE HOMME-CLE C.E.	63	48	(25 %)

TABLE 39 – Résultats PNA GLM CE loi gaussienne

Les prédictions obtenues par ce modèle sont globalement les meilleures que nous avons obtenues pour le réseau CE par rapport à l'arbre de décision BPCE, aux deux XGBoost BPCE et au GLM BPCE mais elles ne sont pas très bonnes pour autant.

5.2.7 Generalized Linear Model loi Gamma CE

Avant de pouvoir utiliser une loi Gamma pour faire tourner notre GLM, nous avons dû traduire nos PNA afin qu'ils soient tous positifs. Ainsi, nous avons regardé le PNA minimum et nous l'avons soustrait, car il était négatif, à tous les PNA. Nous avons ensuite obtenu le modèle suivant :

```
> set.seed(20)
> GLM <- glm(formule, family=Gamma(link="inverse"), data = App)
>
> summary(GLM)

Call:
glm(formula = formule, family = Gamma(link = "inverse"), data = App)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.4076  0.0000  0.0000  0.0001  0.1179

Coefficients: (2 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.546e-05  8.590e-06  2.963  0.00304 **
annee_cpt    1.193e-09  6.347e-10  1.879  0.06027 .
soc.109      2.944e-09  8.543e-09  0.345  0.73036
nom_fam_techSECUR_FAMILLE_CE_DC_PTIA -2.878e-09  9.470e-10 -3.039  0.00238 **
RisqueDC_ACC -5.804e-08  4.636e-08 -1.252  0.21053
RisqueDC_TC  -6.243e-08  4.624e-08 -1.350  0.17695
sexe.H       1.524e-09  2.649e-10  5.753  8.78e-09 ***
periodicite  -2.267e-10  7.811e-10 -0.290  0.77158
mt_assure    3.037e-07  4.012e-08  7.569  3.75e-14 ***
mtht_ann_complet -2.460e-06  2.550e-07 -9.646 < 2e-16 ***
mt_com_ann_complet -3.470e-07  1.250e-07 -2.776  0.00550 **
mt_gestion_ann_complet 6.209e-07  3.007e-07  2.065  0.03895 *
Anciennete   -2.861e-05  8.202e-05 -0.349  0.72719
AgeAdh       -7.010e-04  2.009e-03 -0.349  0.72718
AgeExe       7.226e-04  2.071e-03  0.349  0.72718
Provisions   -2.250e-06  4.414e-08 -50.960 < 2e-16 ***
Anciennete_depuis_surv -3.086e-06  3.804e-08 -81.131 < 2e-16 ***
top_sin_en_cours 3.027e-06  1.324e-08 228.682 < 2e-16 ***
nom_etabCAISSE_D_EPARGNE_AQUITAINE_POITOU_CHARENTES -1.571e-09  5.796e-10 -2.710  0.00673 **
nom_etabCAISSE_D_EPARGNE_BRETAGNE_PAYS_DE_LOIRE 5.835e-10  4.937e-10  1.182  0.23729
nom_etabCAISSE_D_EPARGNE_COTE_D_AZUR -5.542e-10  6.558e-10 -0.845  0.39809
nom_etabCAISSE_D_EPARGNE_DE_BOURGOGNE_FRANCHE_COMTE -3.614e-10  5.695e-10 -0.635  0.52569
nom_etabCAISSE_D_EPARGNE_DE_PROVENCE_ALPES_CORSE -1.363e-09  6.001e-10 -2.271  0.02316 *
nom_etabCAISSE_D_EPARGNE_GRAND_EST_EUROPE -9.920e-10  5.536e-10 -1.792  0.07315 .
nom_etabCAISSE_D_EPARGNE_ILE_DE_FRANCE -3.899e-10  4.598e-10 -0.848  0.39647
nom_etabCAISSE_D_EPARGNE_LOIRE_CENTRE -1.021e-09  5.990e-10 -1.705  0.08821 .
nom_etabCAISSE_D_EPARGNE_RHONE_ALPES -1.099e-09  5.070e-10 -2.168  0.03013 *
SPHT         -1.201e-06  1.017e-07 -11.810 < 2e-16 ***
ChargesPHT   -1.325e-06  4.378e-07 -3.027  0.00247 **
ChargesPPacq -6.527e-07  4.373e-07 -1.492  0.13558
part_PNA_fam_N_1 -1.017e-08  6.669e-09 -1.525  0.12714
part_PNA_risq_N_1 -1.023e-08  7.042e-09 -1.452  0.14647
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 2.035295e-05)

Null deviance: 36.486  on 582380  degrees of freedom
Residual deviance: 35.008  on 582349  degrees of freedom
AIC: 8463513

Number of Fisher Scoring iterations: 4
```

FIGURE 35 – Summary PNA GLM CE loi Gamma

Au niveau des variables d'importance, nous pouvons remarquer que le modèle a sélectionné à peu près les mêmes variables que le GLM avec une loi gaussienne.

Nous avons obtenu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
200,03	141,67	184,81

TABLE 40 – RMSE train, test & base 2019 PNA GLM CE loi Gamma

Au vu de la RMSE de la base 2019, le GLM CE avec une loi Gamma est légèrement meilleur que celui avec une loi gaussienne car sa RMSE sur la base 2019 est un peu plus faible (184,81 VS 188,24) et les RMSE train et test de ces deux modèles sont très proches.

Afin de pouvoir comparer nos données réelles avec nos prédictions, il a fallu prendre l'inverse de chaque prédiction car la fonction de lien que nous avons utilisée pour la loi Gamma était la fonction inverse. Ensuite, nous avons ajouté le PNA minimum à chaque PNA réel et à chaque PNA prédit afin de récupérer les PNA d'origine pour la comparaison et non les PNA translétés.

Nous avons obtenu les résultats suivants par famille de produits :

Famille de produits	PNA réel (k€)	PNA prédit par GLM loi Gamma CE (k€)	Ecart réel prédit
SECUR FAMILLE CE DC PTIA	3 500	1 573	(55 %)
SECUR URGENCE CE	691	283	(59 %)
ASSURANCE HOMME-CLE C.E.	63	48	(25 %)

TABLE 41 – Résultats PNA GLM CE loi Gamma

Ici, les prédictions sont légèrement meilleures que celles que nous avons obtenues avec le GLM CE associé à une loi gaussienne. Cela est sûrement dû au fait que la loi Gamma était plus en adéquation à la distribution de notre PNA que la loi gaussienne. Toutefois, les prédictions ne sont pas très bonnes pour autant mais ce sont les meilleures obtenues jusqu'ici. Nous avons ensuite voulu voir si le XGBoost CE pouvait nous offrir de meilleures prédictions pour les Caisses d'Épargne que les précédents modèles.

5.2.8 XGBoost sur les Caisses d'Épargne

Nous avons fait tourner un nouveau XGBoost uniquement sur les Caisses d'Épargne afin d'observer s'il pouvait mieux prédire le PNA des familles de produits des CE que les XGBoost BPCE et que les GLM précédemment construits.

Cette fois-ci, le modèle a tourné directement sur toutes les données car la volumétrie est moins importante lorsque nous considérons un seul réseau au lieu des deux. Voici les paramètres que le XGBoost CE a décidé de garder parmi le même jeu de paramètres que celui proposé aux deux précédents XGBoost :

- Bootstrap = 20
- Nombre de cross-validation = 10
- Nrounds = 50
- Max_depth = 2
- Colsample_bytree = 1
- Eta = 0,3
- Gamma = 0,9
- Min_child_weight = 5
- Sub_sample = 1

Le modèle a gardé les variables d'importance suivantes parmi les 32 variables explicatives présentes au début du modèle :

Variables	RMSE test	Ecart modèle
PP_acq_complet	171,31	(21,46 %)
mtht_ann_complet	166,29	(17,90 %)
top_sin_en_cours	157,19	(11,45 %)
Provisions	144,36	(2,35 %)
mt_assure	141,73	(0,49 %)
Anciennete_depuis_surv	141,36	(0,23 %)
RisqueDC_ACC	141,10	(0,04 %)
periodicite	141,05	(0,01 %)
Modèle	141,04	(0,00 %)

TABLE 42 – Variables d'importance base totale PNA XGBoost CE

Nous pouvons voir que le modèle a gardé 8 variables, cela ne représente qu'un quart de la totalité des variables, il faut donc vérifier que le modèle n'a pas perdu en performance de prédiction en supprimant autant de variables. Par ailleurs, la variable "Top sinistre en cours l'année précédente" fait bien partie des variables les plus importantes mais elle n'a pas été considérée avec la même importance que ce que nous avons pu observer pour les précédents XGBoost.

Le modèle a été construit en plusieurs étapes qui sont les suivantes :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	177,89	141,92	32
Étape 2	172,75	148,04	14
Étape 3	181,81	153,62	9
Étape 4	190,88	141,04	8

TABLE 43 – Évolution de la RMSE base totale PNA XGBoost CE

Ici, la RMSE train a diminué à la deuxième étape pour augmenter après donc le modèle est devenu moins performant sur la base d'apprentissage au fil des étapes. Quant à la RMSE test, elle a augmenté à la deuxième et à la troisième étape pour diminuer ensuite, ainsi, la RMSE test est quasiment identique à la première et à la quatrième étape, la qualité de la prédiction n'a donc pas diminué malgré la forte baisse du nombre de prédicteurs.

Finalement, nous avons eu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
190,88	141,04	172,4

TABLE 44 – RMSE train, test & base 2019 PNA XGBoost CE

Parmi les trois modèles uniquement sur les CE, c'est celui qui a la RMSE la plus faible pour la base 2019, c'est aussi celui qui a la RMSE train la plus faible alors que la RMSE test est presque identique pour les trois modèles.

Nous avons finalement obtenu les prédictions de PNA suivantes :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost CE (k€)	Ecart réel prédit
SECUR FAMILLE CE DC PTIA	3 500	4 286	(22 %)
SECUR URGENCE CE	691	620	(10 %)
ASSURANCE HOMME-CLE C.E.	63	42	(33 %)

TABLE 45 – Résultats PNA XGBoost CE

C'est indubitablement le modèle qui a le mieux prédit les CE. Ici, nous pouvons dire que les prédictions sont plutôt bonnes pour les trois familles du réseau Caisse d'Épargne.

Dès lors, pour la famille très mal prédite par les deux XGBoost BPCE (Secur Famille CE DC PTIA), les GLM ont été meilleurs que les XGBoost BPCE mais c'est le XGBoost CE qui a le mieux prédit cette famille.

Nous allons maintenant aborder le modèle XGBoost qui a uniquement tourné sur les Banques Populaires.

5.2.9 XGBoost sur les Banques Populaires

Nous avons encore une fois donné le même jeu de paramètres au modèle afin de voir lesquels le modèle allait choisir. Il a retenu les paramètres suivants :

- Bootstrap = 20
- Nombre de cross-validation = 10
- Nrounds = 50
- Max_depth = 4
- Colsample_bytree = 1
- Eta = 0,3
- Gamma = 0,9
- Min_child_weight = 5
- Sub_sample = 1

Comme pour les XGBoost BPCE, il y avait beaucoup d'observations donc nous avons ajouté des poids aux données. Lorsque le modèle a tourné sur environ la moitié des données, il a sélectionné les variables d'importance suivantes :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	503,66	(40,89 %)
mt_assure	398,46	(11,47 %)
Provisions	383,11	(7,17 %)
Anciennete_depuis_surv	362,03	(1,28 %)
mt_com_ann_complet	361,89	(1,24 %)
soc.109	361,02	(0,99 %)
RisqueDC_TC	360,94	(0,97 %)
mt_gestion_ann_complet	360,68	(0,90 %)
ChargesPHT	360,41	(0,82 %)
RisqueDC_ACC	360,23	(0,77 %)
mtht_ann_complet	359,51	(0,57 %)
PP_acq_complet	359,45	(0,55 %)
ChargesPPacq	358,94	(0,41 %)
part_PNA_fam_N_1	357,87	(0,11 %)
Modèle	357,47	(0,00 %)

TABLE 46 – Variables d'importance PNA XGBoost BP

Nous pouvons voir que le modèle a gardé 14 variables sur les 35 possibles au départ, cela correspond à un peu plus du tiers des variables, il en a donc éliminé beaucoup, nous allons regarder si cela a impacté la performance du modèle. Par ailleurs, contrairement au XGBoost sur les CE, il a considéré que c'était la variable "Top sinistre en cours l'année précédente" qui était la plus importante. Nous pouvons donc en déduire que cette variable a été considérée comme la plus importante pour les XGBoost BPCE en raison de son importance pour le réseau BP. En effet, il est compréhensible qu'elle n'ait pas la même importance sur les CE car les contrats n'étaient pas en vigueur depuis très longtemps au 01/01/2019 donc peu de sinistres sont survenus avant cette date, la variable "Top sinistre en cours l'année précédente" était donc très rarement égale à 1. Ci-dessous, nous pouvons observer les étapes de la création du modèle :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	546,83	356,15	35
Étape 2	564,45	355,87	17
Étape 3	556,07	357,47	16

TABLE 47 – Évolution de la RMSE PNA XGBoost BP

La RMSE train n'a pas beaucoup évolué entre la première et la troisième étape et la RMSE test est restée globalement stable donc nous pouvons en déduire que la qualité de prédiction n'a pas été affectée par la réduction du nombre de variables. Ainsi, nous avons bien réduit la complexité des modèles avec un très faible impact sur l'erreur.

Par la suite, nous avons refait tourner le modèle sur toutes les observations avec les paramètres retenus et nous avons obtenu les variables d'importance suivantes :

Variables	RMSE test	Ecart modèle
top_sin_en_cours	475,96	(37,68 %)
mt_assure	396,61	(14,72 %)
Provisions	378,94	(9,61 %)
soc.109	356,45	(3,11 %)
ChargesPHT	352,06	(1,84 %)
SPHT	351,92	(1,80 %)
PP_acq_complet	349,32	(1,04 %)
ChargesPPacq	348,81	(0,90 %)
mtht_ann_complet	348,72	(0,87 %)
RisqueDC_ACC	348,63	(0,85 %)
RisqueDC_TC	348,57	(0,83 %)
Anciennete_depuis_surv	348,28	(0,74 %)
Anciennete	348,21	(0,72 %)
part_PNA_fam_N_1	346,68	(0,28 %)
mt_gestion_ann_complet	346,39	(0,20 %)
mt_com_ann_complet	345,93	(0,06 %)
Modèle	345,71	(0,00 %)

TABLE 48 – Variables d'importance base totale PNA XGBoost BP

Cette fois-ci, le modèle a conservé 16 variables, soit deux de plus que celles sélectionnées lorsque le modèle avait tourné sur la moitié des observations. La totalité des observations lui a donc permis de comprendre l'utilité de certaines variables qu'il n'avait pas conservées à l'origine, il s'agit des variables "Ratio S/PHT l'année N-1" et "Ancienneté en mois".

Pour être construit, ce modèle a suivi trois étapes qui sont les suivantes :

Étapes modèle	RMSE train	RMSE test	Nombre de prédicteurs
Étape 1	535,99	345,49	35
Étape 2	536,89	348,06	16
Étape 3	541,21	345,71	15

TABLE 49 – Évolution de la RMSE base totale PNA XGBoost BP

Pour ce modèle de Machine Learning, la RMSE train a augmenté au fil des étapes et la RMSE test augmenté à la deuxième étape pour ensuite diminuer à peu près jusqu'à la RMSE de départ. Il s'ensuit que ce modèle a une performance presque identique avec la diminution du nombre de variables.

Au final, nous avons eu les RMSE suivantes sur la base d'apprentissage, de test et sur la base contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
541,21	345,71	563,99

TABLE 50 – RMSE train, test & base 2019 PNA XGBoost BP

Au vu des RMSE sur les bases train, test et sur la base 2019, il semblerait que ce modèle soit moins bon que les XGBoost BPCE car les RMSE sont plus élevées. Cependant, cela est dû au fait que la plupart des PNA négatifs que nous avons sont sur les BP car les CE ne sont pas encore en régime de croisière. Comme nous l'avons vu, le fait d'avoir beaucoup de PNA négatifs pénalise énormément la RMSE car les écarts avec le réel sont plus élevés pour ces PNA. Ainsi,

$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ est plus grand pour les BP qu'au global. Nous comparerons plus bas les prédictions par famille de produits de chaque modèle afin de décider quel(s) modèle(s) garder.

Voici les prédictions de PNA que nous avons obtenues à partir du XGBoost sur les Banques Populaires :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BP (k€)	Ecart réel prédit
FAMILLE FRUCTI-FAMILLE	12 067	14 317	(19 %)
FRUCTI-HOMME CLE	6 131	6 130	(0 %)
FRUCTI-PROFESSIONNEL	3 522	3 668	(4 %)
FRUCTI-PATRIMOINE	1 326	1 007	(24 %)
FAMILLE FRUCTI-BUDGET	522	451	(14 %)
DEDICACE	492	490	(1 %)
CREDIT MARITIME VIE	55	75	(37 %)
CERTI CAPITAL	36	54	(49 %)
FAMILLE FRUCTI-FACILITES	32	-20	(164 %)
SECURI-CREDIT BPROP	22	17	(19 %)
FAMILLE DOUBLE SECURITE	18	17	(8 %)
FRUCTI-FACILITES PRO	7	7	(1 %)
FAMILLE FRUCTI-SECURITE	-6	20	(437 %)
FRUCTI-FACILITES AGRI	-71	61	(186 %)
D.S.EPARGNE SALARIALE	-101	63	(163 %)

TABLE 51 – Résultats PNA XGBoost BP

Nous avons obtenu des prédictions assez bonnes sur les familles de produits les plus importantes à l'exception de Famille Fructi-Famille alors que cette famille est très bien prédite par les deux XGBoost BPCE.

Nous avons fait tourner un dernier modèle sur notre base de données avec les deux réseaux afin de pouvoir choisir le meilleur modèle, in fine.

5.2.10 Random Forest

Afin d'avoir une idée des prédictions que nous pourrions obtenir à l'aide d'un modèle de forêts aléatoires, nous avons décidé d'en faire tourner un avec 20 arbres et nous avons pris $\frac{35}{3}$ pour le nombre de variables pouvant être recherchées à chaque noeud et nous avons tronqué ce nombre à 11. En effet, comme nous l'avons vu précédemment dans la description des modèles de Machine Learning, dans le cas d'une régression, pour ce paramètre, il est préférable de choisir un tiers du nombre de variables alors que nous aurions choisi la racine carrée du nombre de variables si nous avions été dans le cas d'une classification. Nous avons, ensuite, fait tourner le modèle sur la base d'apprentissage contenant toutes les observations.

Nous avons obtenu les RMSE suivantes pour la base d'apprentissage, de test et pour la base à prédire contenant les données au 01/01/2019 :

RMSE train	RMSE test	RMSE base 2019
516,81	330,67	472,53

TABLE 52 – RMSE train, test & base 2019 PNA Random Forest BPCE

Par rapport aux XGBoost réalisés sur les deux réseaux, toutes ces RMSE sont plus élevées. Cela signifie que le modèle est moins bon que les XGBoost. Toutefois, les RMSE sont globalement plus faibles que pour l'arbre de décision donc notre algorithme Random Forest reste meilleur qu'un simple arbre de décision et cela correspond à ce que l'on attendait.

Les variables d'importance retenues par le modèle Random Forest sont présentées ci-dessous :

```
> method_rf_model1$importance[order(method_rf_model1$importance[, 1], decreasing = TRUE), ]
      mt_assure 147966365880 Ancienne 81040360780
      mt_gestion_ann_complet 66855790273 Ancienne 61176056718
      AgeAdh 54771113116 ChargesPPacq 41472827756
      top_sin_en_cours 20476827290 periodicite 13064060229
      part_PNA_fam_N_1 11112579814 part_PNA_risq_N_1 8880143114
      nom_etabBANQUE_POPULAIRE_GRAND_QUEST 5772751770
      sexe.H 4605063954 SPHT 4194414532
      nom_etabBANQUE_POPULAIRE_RIVES_DE_PARIS 3849637239
      nom_etabBANQUE_POPULAIRE_BOURGOGNE_FRANCHE_COMTE 2906682301
      nom_etabBANQUE_POPULAIRE_VAL_DE_FRANCE 2448511756
      nom_fam_techDEDICACE 308557543
      nom_fam_techSECUR_FAMILLE_CE_DC_PTIA 147111151
      PP_acq_complet 94506322775
      mtht_ann_complet 78839738272
      AgeExe 63202654402
      mt_com_ann_complet 59837913151
      Provisions 51675384241
      ChargesPHT 22015172126
      RisqueDC_ACC 16338078282
      soc.109 11801030114
      nom_etabBANQUE_POPULAIRE_AUVERGNE_RHONE_ALPES 9175049782
      nom_fam_techFRUCTI_HOMME_CLE 6216408088
      RisqueDC_TC 5370128025
      nom_etabBANQUE_POPULAIRE_DU_SUD 4289348780
      nom_etabBANQUE_POPULAIRE_ALSACE_LORRAINE_CHAMPAGNE 4074126688
      nom_fam_techFAMILLE_FRUCTI_FAMILLE 3100614100
      SPPacq 2716301274
      nom_etabBANQUE_POPULAIRE_OCCITANE 1492223690
      Reseau_BP 233017912
```

FIGURE 36 – Variables d'importance PNA Random Forest BPCE

Les variables sont triées de la plus importante (en haut à gauche) à la moins importante (en bas à gauche). Nous pouvons remarquer que, pour ce modèle, comme pour les autres, les variables les plus importantes sont assez similaires. Néanmoins, ce modèle a jugé que la variable "Top sinistre en cours l'année précédente" était moins importante que beaucoup d'autres variables alors que c'était la variable la plus importante pour la plupart des précédents modèles.

Nous avons ensuite regardé les prédictions que nous avons obtenues par famille de produits et voici ce que nous avons obtenu :

Famille de produits	PNA réel (k€)	PNA prédit par Random Forest (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	14 488	(20 %)
FRUCTI-HOMME CLE	6 131	4 410	(28 %)
FRUCTI-PROFESSIONNEL	3 522	1 430	(59 %)
SECUR FAMILLE CE DC PTIA	3 500	8 512	(143 %)
FRUCTI-PATRIMOINE	1 326	1 161	(12 %)
SECUR URGENCE CE	691	1 146	(66 %)
FAMILLE FRUCTI-BUDGET	522	994	(90 %)
DEDICACE	492	496	(1 %)
ASSURANCE HOMME-CLE C.E.	63	1	(98 %)
CREDIT MARITIME VIE	55	68	(25 %)
CERTI CAPITAL	36	29	(18 %)
FAMILLE FRUCTI-FACILITES	32	-29	(191 %)
SECURI-CREDIT BPROP	22	11	(47 %)
FAMILLE DOUBLE SECURITE	18	9	(53 %)
FRUCTI-FACILITES PRO	7	-7	(196 %)
FAMILLE FRUCTI-SECURITE	-6	-7	(17 %)
FRUCTI-FACILITES AGRI	-71	-86	(22 %)
D.S.EPARGNE SALARIALE	-101	-296	(193 %)

TABLE 53 – Résultats PNA Random Forest BPCE

Les prédictions sont assez bonnes pour les principales familles de produits à l'exception de Secur Famille CE DC PTIA. En effet, sur cette famille, les prédictions sont très mauvaises par rapport à ce que nous avons pu obtenir précédemment. Par ailleurs, Famille Fructi-Famille a été bien mieux prédite par les deux XGBoost BPCE que par ce modèle. Ainsi, comme nous avons pu le remarquer en regardant les RMSE, ce modèle est moins bon que les deux précédents XGBoost BPCE.

5.2.11 Comparaison des modèles et modèle retenu

Afin de déterminer le(s) meilleur(s) modèle(s), nous avons décidé de comparer les différentes RMSE de ceux-ci, mais également, les prédictions totales par famille de produits.

Tout d'abord, nous avons regardé les RMSE de tous les modèles. Il faut cependant noter que, pour les modèles uniquement sur les CE ou uniquement sur les BP, nous ne pouvons pas comparer ces RMSE à toutes les autres car elles n'ont pas été calculées sur les mêmes bases de données. En effet, comme nous l'avons vu, ce sont chez les BP que nous retrouvons le plus de PNA négatifs et ce sont eux qui augmentent la RMSE car les prédictions sont souvent plus éloignées de la réalité que pour les PNA positifs. Voici ce que nous avons obtenu :

Modèle	RMSE train	RMSE test	RMSE base 2019
Arbre de décision BPCE	508,41	346,96	512,07
XGBoost BPCE	494,33	320,36	426,23
XGBoost BPCE NEW	491,37	317,77	425,47
Random Forest BPCE	516,81	330,67	472,53
GLM BPCE	511,10	330,17	452,20
GLM CE loi gaussienne	199,7	141,54	188,24
GLM CE loi Gamma	200,03	141,67	184,81
XGBoost CE	190,88	141,04	172,4
XGBoost BP	541,21	345,71	563,99

TABLE 54 – Comparaison RMSE train, test & base 2019 PNA

Tout d'abord, en ce qui concerne les modèles BPCE, nous remarquons que le modèle XGBoost BPCE NEW a la RMSE train la plus faible, c'est donc le modèle qui prédit le mieux la base d'apprentissage tandis que l'algorithme Random Forest a la RMSE train la plus élevée. Quant à la RMSE test, c'est encore le XGBoost BPCE NEW qui a la plus faible, c'est donc lui qui a été le meilleur en prédiction sur la base test et c'est l'arbre de décision qui a la plus élevée et qui a donc le moins bien prédit l'échantillon de test. Enfin, le modèle qui a la plus petite RMSE pour la base 2019 est encore une fois le modèle XGBoost BPCE NEW et celui qui a la plus élevée est l'arbre de décision. Dès lors, au vu de la RMSE, il semble que, pour les modèles BPCE, le XGBoost BPCE NEW soit le meilleur modèle.

En ce qui concerne les Caisses d'Épargne, c'est le XGBoost CE qui a la RMSE train la plus faible et qui fait donc le moins d'erreurs sur les prédictions de la base d'apprentissage tandis que le GLM CE avec une loi Gamma a la RMSE train la plus élevée. Les trois modèles ont relativement la même RMSE test donc aucun modèle n'a été meilleur pour prédire la base test. Finalement, le modèle prédisant le mieux la base 2019 est le XGBoost CE et c'est le GLM CE avec une loi gaussienne qui a la RMSE la plus élevée pour la base 2019. Ainsi, c'est le XGBoost CE qui prédit le mieux les trois différentes bases.

Voici ce que nous avons obtenu en comparant toutes les prédictions :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BPCE (k€)	Ecart réel VS prédit	PNA prédit par XGBoost BPCE NEW (k€)	Ecart réel VS prédit	PNA prédit par GLM loi gaussienne BPCE (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	11 933	(1 %)	12 004	(1 %)	10 776	(11 %)
FRUCTI-HOMME CLE	6 131	5 940	(3 %)	6 153	(0 %)	5 969	(3 %)
FRUCTI-PROFESSIONNEL	3 522	3 225	(8 %)	4 029	(14 %)	4 153	(18 %)
SECUR FAMILLE CE DC PTIA	3 500	7 235	(107 %)	6 935	(98 %)	1 663	(52 %)
FRUCTI-PATRIMOINE	1 326	1 010	(24 %)	1 022	(23 %)	1 354	(2 %)
SECUR URGENCE CE	691	879	(27 %)	991	(43 %)	78	(89 %)
FAMILLE FRUCTI-BUDGET	522	391	(25 %)	336	(36 %)	22	(96 %)
DEDICACE	492	424	(14 %)	493	(0 %)	315	(36 %)
ASSURANCE HOMME-CLE C.E.	63	36	(43 %)	56	(12 %)	44	(30 %)
CREDIT MARITIME VIE	55	57	(5 %)	57	(4 %)	34	(39 %)
CERTI CAPITAL	36	31	(15 %)	40	(10 %)	25	(32 %)
FAMILLE FRUCTI-FACILITES	32	-6	(118 %)	-15	(146 %)	-57	(279 %)
SECURI-CREDIT BPROP	22	19	(12 %)	16	(25 %)	-3	(113 %)
FAMILLE DOUBLE SECURITE	18	9	(50 %)	10	(45 %)	7	(61 %)
FRUCTI-FACILITES PRO	7	14	(94 %)	11	(59 %)	1	(87 %)
FAMILLE FRUCTI-SECURITE	-6	-7	(15 %)	-8	(43 %)	-4	(34 %)
FRUCTI-FACILITES AGRI	-71	60	(185 %)	59	(184 %)	66	(194 %)
D.S.EPARGNE SALARIALE	-101	51	(151 %)	38	(137 %)	184	(281 %)

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision (k€)	Ecart réel VS prédit	PNA prédit par Random Forest (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	20 356	(69 %)	14 488	(20 %)
FRUCTI-HOMME CLE	6 131	1 541	(75 %)	4 410	(28 %)
FRUCTI-PROFESSIONNEL	3 522	-987	(128 %)	1 430	(59 %)
SECUR FAMILLE CE DC PTIA	3 500	19 819	(466 %)	8 512	(143 %)
FRUCTI-PATRIMOINE	1 326	-86	(107 %)	1 161	(12 %)
SECUR URGENCE CE	691	3 232	(367 %)	1 146	(66 %)
FAMILLE FRUCTI-BUDGET	522	1 069	(105 %)	994	(90 %)
DEDICACE	492	1 317	(168 %)	496	(1 %)
ASSURANCE HOMME-CLE C.E.	63	14	(78 %)	1	(98 %)
CREDIT MARITIME VIE	55	81	(48 %)	68	(25 %)
CERTI CAPITAL	36	134	(273 %)	29	(18 %)
FAMILLE FRUCTI-FACILITES	32	70	(122 %)	-29	(191 %)
SECURI-CREDIT BPROP	22	42	(96 %)	11	(47 %)
FAMILLE DOUBLE SECURITE	18	49	(168 %)	9	(53 %)
FRUCTI-FACILITES PRO	7	30	(327 %)	-7	(196 %)
FAMILLE FRUCTI-SECURITE	-6	2	(140 %)	-7	(17 %)
FRUCTI-FACILITES AGRI	-71	41	(158 %)	-86	(22 %)
D.S.EPARGNE SALARIALE	-101	199	(297 %)	-296	(193 %)

TABLE 55 – Comparaison PNA BPCE

Sur ce tableau, les écarts en gras sont les plus faibles, ainsi, si l'écart est en gras pour une famille de produits pour un certain modèle, cela signifie que c'est celui qui prédit le mieux la famille de produits en question. Le modèle qui prédit le mieux la plupart des familles de produits est le XGBoost BPCE NEW. Par ailleurs, c'est aussi ce modèle qui avait la RMSE la plus faible, c'est donc le modèle que nous allons retenir au global. Néanmoins, il reste à vérifier que c'est le modèle le plus adapté pour les deux réseaux. Nous allons donc maintenant regarder le meilleur modèle pour les BP et pour les CE.

Pour les Banques Populaires, nous avons obtenu ceci :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BP (k€)	Ecart réel VS prédit	PNA prédit par XGBoost BPCE (k€)	Ecart réel VS prédit	PNA prédit par XGBoost BPCE NEW (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	14 317	(19 %)	11 933	(1 %)	12 004	(1 %)
FRUCTI-HOMME CLE	6 131	6 130	(0 %)	5 940	(3 %)	6 153	(0 %)
FRUCTI-PROFESSIONNEL	3 522	3 668	(4 %)	3 225	(8 %)	4 029	(14 %)
FRUCTI-PATRIMOINE	1 326	1 007	(24 %)	1 010	(24 %)	1 022	(23 %)
FAMILLE FRUCTI-BUDGET	522	451	(14 %)	391	(25 %)	336	(36 %)
DEDICACE	492	492	(0 %)	424	(14 %)	493	(0 %)
CREDIT MARITIME VIE	55	75	(37 %)	57	(5 %)	57	(4 %)
CERTI CAPITAL	36	54	(49 %)	31	(15 %)	40	(10 %)
FAMILLE FRUCTI-FACILITES	32	-20	(164 %)	-6	(118 %)	-15	(146 %)
SECURI-CREDIT BPROP	22	17	(19 %)	19	(12 %)	16	(25 %)
FAMILLE DOUBLE SECURITE	18	17	(8 %)	9	(50 %)	10	(45 %)
FRUCTI-FACILITES PRO	7	7	(1 %)	14	(94 %)	11	(59 %)
FAMILLE FRUCTI-SECURITE	-6	20	(437 %)	-7	(15 %)	-8	(43 %)
FRUCTI-FACILITES AGRI	-71	61	(186 %)	60	(185 %)	59	(184 %)
D.S.EPARGNE SALARIALE	-101	63	(163 %)	51	(151 %)	38	(137 %)

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision (k€)	Ecart réel VS prédit	PNA prédit par Random Forest (k€)	Ecart réel VS prédit	PNA prédit par GLM loi gaussienne BPCE (k€)	Ecart réel VS prédit
FAMILLE FRUCTI-FAMILLE	12 067	20 356	(69 %)	14 488	(20 %)	10 776	(11 %)
FRUCTI-HOMME CLE	6 131	1 541	(75 %)	4 410	(28 %)	5 969	(3 %)
FRUCTI-PROFESSIONNEL	3 522	-987	(128 %)	1 430	(59 %)	4 153	(18 %)
FRUCTI-PATRIMOINE	1 326	-86	(107 %)	1 161	(12 %)	1 354	(2 %)
FAMILLE FRUCTI-BUDGET	522	1 069	(105 %)	994	(90 %)	22	(96 %)
DEDICACE	492	1 317	(168 %)	496	(1 %)	315	(36 %)
CREDIT MARITIME VIE	55	81	(48 %)	68	(25 %)	34	(39 %)
CERTI CAPITAL	36	134	(273 %)	29	(18 %)	25	(32 %)
FAMILLE FRUCTI-FACILITES	32	70	(122 %)	-29	(191 %)	-57	(279 %)
SECURI-CREDIT BPROP	22	42	(96 %)	11	(47 %)	-3	(113 %)
FAMILLE DOUBLE SECURITE	18	49	(168 %)	9	(53 %)	7	(61 %)
FRUCTI-FACILITES PRO	7	30	(327 %)	-7	(196 %)	1	(87 %)
FAMILLE FRUCTI-SECURITE	-6	2	(140 %)	-7	(17 %)	-4	(34 %)
FRUCTI-FACILITES AGRI	-71	41	(158 %)	-86	(22 %)	66	(194 %)
D.S.EPARGNE SALARIALE	-101	199	(297 %)	-296	(193 %)	184	(281 %)

TABLE 56 – Comparaison PNA BP

Nous pouvons remarquer que c'est le XGBoost BP qui offre les meilleures prédictions pour la plupart des familles de produits. Cependant, le meilleur modèle pour Famille Fructi-Famille est le XGBoost BPCE NEW. De plus, pour certaines familles de produits où le meilleur modèle est le XGBoost BP, le XGBoost BPCE NEW offre également de bonnes prédictions. Ainsi, les deux modèles offrent de bonnes prédictions pour le réseau BP.

Voici ce que nous avons obtenu pour les Caisses d'Épargne :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost CE (k€)	Ecart réel VS prédit	PNA prédit par XGBoost BPCE (k€)	Ecart réel VS prédit	PNA prédit par XGBoost BPCE NEW (k€)	Ecart réel VS prédit
SECUR FAMILLE CE DC PTIA	3 500	4 286	(22 %)	7 235	(107 %)	6 935	(98 %)
SECUR URGENCE CE	691	620	(10 %)	879	(27 %)	991	(43 %)
ASSURANCE HOMME-CLE C.E.	63	42	(33 %)	36	(43 %)	56	(12 %)

Famille de produits	PNA réel (k€)	PNA prédit par Arbre de décision (k€)	Ecart réel VS prédit	PNA prédit par Random Forest (k€)	Ecart réel VS prédit	PNA prédit par GLM loi gaussienne BPCE (k€)	Ecart réel VS prédit
SECUR FAMILLE CE DC PTIA	3 500	19 819	(466 %)	8 512	(143 %)	1 663	(52 %)
SECUR URGENCE CE	691	3 232	(367 %)	1 146	(66 %)	78	(89 %)
ASSURANCE HOMME-CLE C.E.	63	14	(78 %)	1	(98 %)	44	(30 %)

Famille de produits	PNA réel (k€)	PNA prédit par GLM loi gaussienne CE (k€)	Ecart réel VS prédit	PNA prédit par GLM loi Gamma CE (k€)	Ecart réel VS prédit
SECUR FAMILLE CE DC PTIA	3 500	1 373	(61 %)	1 573	(55 %)
SECUR URGENCE CE	691	270	(61 %)	283	(59 %)
ASSURANCE HOMME-CLE C.E.	63	48	(25 %)	48	(25 %)

TABLE 57 – Comparaison PNA CE

Ici, c'est le XGBoost CE qui offre les meilleures prédictions. Le XGBoost BPCE NEW est le meilleur modèle pour une des prédictions néanmoins, comme nous l'avons vu, la prédiction de Secur Famille CE DC PTIA n'est pas très bonne.

Dès lors, au vu des différents modèles réalisés et des observations faites par rapport aux RMSE et aux prédictions, nous avons décidé de retenir le XGBoost BPCE NEW pour les deux réseaux à des fins de simplifications. En effet, cela permettra de ne refaire tourner qu'un seul modèle l'an prochain lorsque nous ajouterons des données. Par ailleurs, les prédictions de ce modèle étaient plutôt bonnes pour les BP, en effet, les performances du XGBoost BP et du XGBoost BPCE NEW étaient assez proches. En ce qui concerne les CE, il est vrai que la prédiction de Secur Famille CE DC PTIA n'était pas très proche de la réalité avec le XGBoost BPCE NEW, cependant, les prédictions des deux autres familles de produits CE n'étaient pas très éloignées de celles obtenues avec le XGBoost CE. En outre, nous noterons que, même si ce modèle ne fonctionne pas très bien pour les CE pour le moment, plus tard, avec l'afflux de nouvelles données, il pourra mieux apprendre et donc mieux prédire.

Une fois le meilleur modèle choisi, nous avons pu comparer ses prédictions avec le BE S2 transformé en PNA, nous avons obtenu les résultats suivants :

Famille de produits	PNA réel (k€)	PNA prédit par XGBoost BPCE NEW (k€)	Ecart réel VS prédit	BE S2 transformé en PNA (k€)	Ecart réel VS projeté	ML meilleur que S2 ?
FAMILLE FRUCTI-FAMILLE	12 067	12 004	(1 %)	10 370	(14 %)	OUI
FRUCTI-HOMME CLE	6 131	6 153	(0 %)	3 229	(47 %)	OUI
FRUCTI-PROFESSIONNEL	3 522	4 029	(14 %)	8 348	(137 %)	OUI
SECUR FAMILLE CE DC PTIA	3 500	6 935	(98 %)	4 040	(15 %)	NON
FRUCTI-PATRIMOINE	1 326	1 022	(23 %)	1 173	(12 %)	NON
SECUR URGENCE CE	691	991	(43 %)	597	(14 %)	NON
FAMILLE FRUCTI-BUDGET	522	336	(36 %)	1 455	(179 %)	OUI
DEDICACE	492	493	(0 %)	617	(25 %)	OUI
ASSURANCE HOMME-CLE C.E.	63	56	(12 %)	23	(64 %)	OUI
CREDIT MARITIME VIE	55	57	(4 %)	93	(70 %)	OUI
CERTI CAPITAL	36	40	(10 %)	52	(45 %)	OUI
FAMILLE FRUCTI-FACILITES	32	-15	(146 %)	10	(67 %)	NON
SECURI-CREDIT BPROP	22	16	(25 %)	19	(10 %)	NON
FAMILLE DOUBLE SECURITE	18	10	(45 %)	11	(38 %)	NON
FRUCTI-FACILITES PRO	7	11	(59 %)	9	(23 %)	NON
FAMILLE FRUCTI-SECURITE	-6	-8	(43 %)	7	(219 %)	OUI
FRUCTI-FACILITES AGRI	-71	59	(184 %)	48	(168 %)	NON
D.S.EPARGNE SALARIALE	-101	38	(137 %)	206	(303 %)	OUI

TABLE 58 – Résultats PNA XGBoost BPCE NEW et comparaison avec le BE S2

Ainsi, il apparaît que les prédictions obtenues par Machine Learning sont souvent meilleures que le BE S2. En effet, pour les familles de produits les plus importantes, c'est le XGBoost BPCE NEW qui offre la meilleure approximation du PNA à fin 2019. Néanmoins, pour les plus petites familles de produits, le BE S2 est souvent plus proche du PNA réel que le PNA prédit par ML. Cela s'explique par le fait qu'il y a peu de données pour ces familles de produits, le modèle de

Machine Learning n'apprend donc pas bien comment prédire ces familles. Le ML permet donc un gain de temps mais il permet également d'obtenir de meilleures approximations du résultat réel que le BE S2. Ainsi, nous pouvons nous questionner sur le paramétrage qui est effectué sous S2 car les projections sont un peu écartées du réel.

Le lecteur souhaitant connaître les prédictions que nous avons obtenues pour l'année 2020 à partir du XGBoost BPCE NEW pourra regarder l'annexe B.

Dès lors, cette partie du mémoire nous a permis de voir l'utilité des modèles de Machine Learning. En effet, grâce aux prédictions de PNA obtenues par ces modèles, nous avons pu backtester le BE S2 et nous avons remarqué que le Machine Learning était souvent plus proche du PNA réel que le BE S2 transformé en PNA. Par ailleurs, nous avons déjà pu constater des écarts causés par S2 dans la sous-partie 4.7, le paramétrage S2 pourra donc être amélioré, à la suite de ce mémoire, notamment sur la réassurance.

Conclusion

Les deux principaux objectifs de ce mémoire étaient d'expliquer le partage de la valeur pour un bancassureur et de comparer les normes French GAAP, Solvabilité 2 et IFRS 17 par différents moyens.

Nous avons donc commencé par nous intéresser à deux indicateurs de rentabilité : Le Produit Net Bancaire qui indique la rentabilité de la banque et le Produit Net Assurance qui indique, quant à lui, la rentabilité de l'assureur. Afin d'évaluer ces indicateurs de différentes manières, nous avons construit une base de données permettant de les obtenir par tête. Ainsi, il a été possible d'agrèger ces indicateurs aux différentes mailles souhaitées, en particulier, à la maille famille de produits. Grâce à cette base, nous avons pu obtenir le PNA et le PNB de chaque établissement bancaire ce qui nous a permis de construire une cartographie représentant notamment le ratio PNA/PNB pour les différentes régions de France et les réseaux Banque Populaire et Caisse d'Épargne. Par ailleurs, nous avons pu construire des tableaux représentant l'évolution du PNA, du PNB et du ratio PNA/PNB au fil des années ainsi que l'évolution du PNA en fonction de l'ancienneté des contrats. Tout cela nous a permis de comprendre que le ratio PNA/PNB n'est pas toujours équitable, cela dépend de la famille de produits, de l'année comptable étudiée, ou même, de la région et du réseau considéré. Néanmoins, comme nous l'avons expliqué, un ratio défavorable à l'assureur, i.e., un ratio inférieur à 1, ne signifie pas pour autant que l'assureur aura très peu de chiffre d'affaires, en effet, la banque peut apporter beaucoup d'affaires et donc permettre à l'assureur de gagner plus. Ainsi, il est important de soigner le ratio PNA/PNB pour que les banques aient envie d'apporter un maximum d'affaires à l'assureur. Il n'en reste pas moins que le ratio est parfois très éloigné de 1, la plupart du temps en faveur des Banques Populaires et des Caisses d'Épargne (ratio < 1). En outre, les comparaisons que nous avons réalisées à la maille famille de produits vont permettre au département Actuariat Produits Prévoyance de classer les familles par rentabilité et d'effectuer d'éventuelles revues tarifaires.

Par la suite, nous sommes rentrés dans l'univers IFRS 17, la norme internationale qui entrera en vigueur au 1er janvier 2023 afin de comprendre ses enjeux mais également de pouvoir la comparer à Solvabilité 2 ainsi qu'à French GAAP. Nous avons donc calculé le résultat IFRS 17 de l'année 2019. Ceci nous a permis de réaliser une analyse de passage de la norme French GAAP à IFRS 17 puis de comparer le résultat net de ces deux normes. L'analyse de passage entre IFRS 17 (où les flux de réassurance sont estimés selon S2) et French GAAP a montré une divergence significative du compte de réassurance. En ce qui concerne la comparaison entre IFRS 17 et S2, nous avons décidé de comparer ces deux normes via le Return On Equity.

Enfin, la base que nous avons construite par tête nous a permis de réaliser plusieurs modèles de Machine Learning grâce auxquels nous avons pu comparer le PNA réel avec sa prédiction et nous avons décidé de retenir un modèle XGBoost qui avait tourné sur les deux réseaux. En outre, à partir du meilleur modèle de ML, nous avons pu comparer le PNA prédit avec le BE S2 que nous avons transformé en PNA, afin d'obtenir une comparaison des normes French GAAP et Solvabilité 2. Cette comparaison a démontré que les prédictions réalisées par Machine Learning étaient plus proches de la réalité que les Best Estimate de Solvabilité 2. Ainsi, nous pouvons nous questionner sur le paramétrage qui est effectué sous S2 car les projections sont un peu écartées du réel. En effet, il est important d'avoir un regard critique sur les projections S2 et d'analyser les sources d'écart afin d'améliorer la modélisation et la calibration des hypothèses. Au vu des différents écarts constatés sous Solvabilité 2, le paramétrage S2 pourra donc être amélioré, à la suite de ce mémoire, notamment au niveau de la réassurance. Par ailleurs, le modèle de Machine Learning retenu est un proxy qui sera utilisé, dans le futur, afin de backtester les résultats des projections S2 comme nous avons pu le faire dans ce mémoire pour l'année 2019.

Dès lors, ce mémoire nous a permis d'expliquer l'importance du partage de la valeur pour Natixis Assurances dans un premier temps. Il a également permis de comparer trois normes importantes (French GAAP, Solvabilité 2 et IFRS 17) par différents moyens tels que le Machine Learning. Ces algorithmes ont permis un gain de temps par rapport à ce qui est fait sous Solvabilité 2 car cette norme nécessite de préparer tous les inputs modèle, puis, de tout projeter et, enfin, de calculer les résultats de tous les scénarios paramétrés, cela prend donc un temps considérable. Ainsi, même si les modèles de ML sont considérés comme des "boîtes noires", il est intéressant de regarder les prédictions qu'ils nous permettent d'obtenir. En outre, en actuariat, nous devons souvent projeter

des données avec des modèles et comparer nos résultats d'une année sur l'autre. Toutefois, nous ne comparons que très peu les projections avec les données réelles. Ce mémoire nous a donc permis de comparer des résultats projetés avec des résultats réels et, ainsi, de développer un meilleur esprit critique et une meilleure analyse des résultats.

Table des figures

1	Natixis Assurances : Deux métiers au service des banques du groupe BPCE	7
2	Une organisation et quatre grands domaines	8
3	Ratio PNA/PNB pour toutes les familles de produits et tous les risques 2016-2019	19
4	Ratio PNA/PNB des plus grandes familles BP et CE 2019	20
5	Évolution du ratio PNA/PNB DC TC de 2016 à 2019	21
6	Pieuvre S2	27
7	Matrice de corrélation BSCR	28
8	Passage de Solvabilité 1 à Solvabilité 2	29
9	Les trois piliers de Solvabilité 2	31
10	Chaîne de Markov : Espace des états du modèle de projection	31
11	Schéma sinistres connus et inconnus	32
12	Périmètre IFRS 17	34
13	Bilan IFRS 9/IFRS 17	36
14	Une nouvelle granularité	39
15	Contrats déficitaires et profitables	40
16	Bottom-Up VS Top-Down	44
17	Risk Margin VS Risk Adjustment	45
18	Liability for Remaining Coverage et Liability for Incurred Claims	48
19	Fonctionnement du modèle BBA	51
20	Résultat DCAT IFRS 17	63
21	Processus de calcul des SCR en prévoyance individuelle	66
22	Gain de diversification	66
23	Analyse de passage French GAAP IFRS 17	69
24	Représentation de la méthode k-fold cross-validation	81
25	Explication des données pouvant être utilisées pour la prédiction du PNA	83
26	Schéma PNA Arbre de décision BPCE	86
27	Grid search aléatoire	89
28	Attribution des poids aux données	89
29	Évolution de l'erreur en fonction de la valeur d'un paramètre d'apprentissage	92
30	Plot du PNA en fonction de sa fréquence BPCE	97
31	Summary PNA GLM BPCE loi gaussienne	98
32	Plot du PNA en fonction de sa fréquence CE	101
33	Distribution loi Gamma	101
34	Summary PNA GLM CE loi gaussienne	102
35	Summary PNA GLM CE loi Gamma	103
36	Variables d'importance PNA Random Forest BPCE	109

Liste des tableaux

1	Chiffre d'affaires hors taxes et nombre de contrats par risque	9
2	Distribution du chiffre d'affaires par risque	17
3	Poids des différentes familles de produits en 2019	18
4	Évolution des règlements par risque	18
5	Évolution du ratio PNA/PNB par banque au fil du temps	22
6	Évolution du ratio PNB/CA par banque au fil du temps	22
7	Évolution du PNA en fonction de l'ancienneté des contrats	23
8	Évolution du PNA au fil du temps	23
9	Évolution du PNB au fil du temps	24
10	Évolution du ratio PNA/PNB au fil du temps	24
11	Évolution du RBE au fil du temps	25
12	Modèles disponibles par activité	50
13	Modèles disponibles en prévoyance individuelle	50
14	Solvabilité 2 VS IFRS 17	65
15	Comparaison ROE Solvabilité 2 et IFRS 17	67
16	Analyse de passage de French GAAP à IFRS 17 par famille de produits	68
17	Comparaison des résultats nets French GAAP et IFRS 17	68
18	Variables sans variance ou corrélées gardées	85
19	RMSE train, test & base 2019 PNA Arbre de décision BPCE	86
20	Résultat PNA Arbre de décision BPCE	87
21	Résultat PNA Arbre de décision BPCE (sinistre en cours en 2018)	88
22	Résultats PNA Arbre de décision BPCE (sans sinistre en cours en 2018)	88
23	Variables d'importance PNA XGBoost BPCE	91
24	Évolution de la RMSE PNA XGBoost BPCE	91
25	Variables d'importance base totale PNA XGBoost BPCE	92
26	Évolution de la RMSE base totale PNA XGBoost BPCE	93
27	RMSE train, test & base 2019 PNA XGBoost BPCE	93
28	Résultats PNA XGBoost BPCE	93
29	Variables d'importance XGBoost PNA BPCE NEW	95
30	Évolution de la RMSE PNA XGBoost BPCE NEW	95
31	Variables d'importance base totale PNA XGBoost BPCE NEW	95
32	Évolution de la RMSE base totale PNA XGBoost BPCE NEW	96
33	RMSE train, test & base 2019 PNA XGBoost BPCE NEW	96
34	Résultats PNA XGBoost BPCE NEW	96
35	Résultats PNA GLM BPCE loi gaussienne : Contrats sinistrés et non sinistrés	99
36	RMSE train, test & base 2019 PNA GLM BPCE loi gaussienne	99
37	Résultats PNA GLM BPCE loi gaussienne	100
38	RMSE train, test & base 2019 PNA GLM CE loi gaussienne	102
39	Résultats PNA GLM CE loi gaussienne	103
40	RMSE train, test & base 2019 PNA GLM CE loi Gamma	104
41	Résultats PNA GLM CE loi Gamma	104
42	Variables d'importance base totale PNA XGBoost CE	105
43	Évolution de la RMSE base totale PNA XGBoost CE	105
44	RMSE train, test & base 2019 PNA XGBoost CE	105
45	Résultats PNA XGBoost CE	105
46	Variables d'importance PNA XGBoost BP	106
47	Évolution de la RMSE PNA XGBoost BP	106
48	Variables d'importance base totale PNA XGBoost BP	107
49	Évolution de la RMSE base totale PNA XGBoost BP	107
50	RMSE train, test & base 2019 PNA XGBoost BP	107
51	Résultats PNA XGBoost BP	108
52	RMSE train, test & base 2019 PNA Random Forest BPCE	108
53	Résultats PNA Random Forest BPCE	109
54	Comparaison RMSE train, test & base 2019 PNA	110
55	Comparaison PNA BPCE	111
56	Comparaison PNA BP	111
57	Comparaison PNA CE	112
58	Résultats PNA XGBoost BPCE NEW et comparaison avec le BE S2	112

Bibliographie

- [EIOPA-2014] *Technical specifications for the preparatory phase (part 1)*, EIOPA 2014
- [KAMEGA-2014] *Introduction à solvabilité 2, applications de mesure de risque*, Aymric KAMEGA, 2014
- [DATA-SCIENCE-2015] *Data science : Fondamentaux et études de cas*, Eric Biernat Michel Lutz, 2015
- [GEORGES-2016] *Construction d'une loi de maintien en arrêt de travail et chocs de la formule standard*, Nicolas Georges, 2016
- [ACTUARIS-2017] *INFOTECH # 48 INTRODUCTION À LA NORME IFRS 17*, Actuaris, 2017
- [FORSIDES-2017] *MORNING FORSIDES DU 19 DÉCEMBRE 2017*, Forsides, 2017
- [IA-2017] *Attention nouvelle norme ! Comment appréhender IFRS 17 ?*, Institut des actuaires, 2017
- [IFRS17] *IFRS 17 Insurance Contracts*, IASB, 2017
- [PWC-2017] *IFRS 17 Insurance Contracts in brief*, PwC, 2017
- [AHMED-FERRERI-2018] *Norme IFRS 17 : Etude des facteurs explicatifs de la variation du résultat en assurance emprunteur*, Kamel Bent Ahmed Paul-Laurent Ferreri, 2018
- [EY-2018] *Applying IFRS 17 : A closer look at the new Insurance Contracts Standard*, EY, 2018
- [IA-2018] *IFRS 17 Point d'actualités et présentation des travaux du GT IFRS 17*, Institut des actuaires, 2018
- [KERNEIS-2018] *IFRS 17 : Enjeux et application en assurance emprunteur*, Julie Kernéis, 2018
- [MILLIMAN-2018] *IFRS 17 : How simple is the simplified approach*, Milliman, 2018
- [PELTIER-2018] *Présentation IFRS 17*, Claire Peltier, 2018
- [SEABIRD-2018] *IFRS 17 vous donne du fil à retordre ?*, SeaBird, 2018
- [ADDACTIS-2019] *LE RISK ADJUSTMENT SOUS IFRS 17*, addactis, 2019
- [APREF-2019] *Prise en compte de la Réassurance dans IFRS 17*, APREF, 2019
- [ASIGMA-2019] *IFRS 17 Quelle approche opérationnelle ?*, Asigma, 2019
- [OPTIMIND-2019] *IFRS 17 Transition : une appropriation des différents enjeux pour une meilleure maîtrise de la communication financière*, Optimind, 2019
- [SALMON-2019] *Méthodes de détermination du taux d'actualisation dans le cadre de la norme IFRS 17*, Virgile Salmon, 2019
- [SARDAR-2019] *CSM : Identification of coverage units*, Rebecca Sardar, 2019
- [HELARY-2020] *Modélisation du rachat en assurance des emprunteurs*, Adrien Helary, 2020
- [ACPR] <https://acpr.banque-france.fr>
- [ALGO-ML] <https://machinelearningmastery.com/>
- [BAGGING-BOOSTING] <https://quantdare.com/>
- [COURS-ML] <http://eric.univ-lyon2.fr/>
- [Focus-IFRS17] <http://www.focusifrs.com>
- [LEXIQUE-ML] <https://dataanalyticspost.com/Lexique>

Annexes

A Détail des provisions totales d'ouverture et de clôture

- Provisions d'ouverture : C'est la somme des différents éléments ci-dessous :
 - Les provisions de sinistres d'ouverture
 - L'IPM sur les provisions d'ouverture
 - Les provisions de primes pures et commerciales d'ouverture
 - Le transfert de provisions mathématiques d'ouverture
 - La provision mathématiques de rente d'ouverture
 - Les provisions pour contentieux d'ouverture
 - Les prévisions de recours restant à encaisser d'ouverture
 - Les intérêts sur les provisions
 - La participation aux bénéfices incorporée dans l'exercice et la variation d'IPM
 - Les provisions mathématiques pures pour risque en cours d'ouverture
- Provisions de clôture : Elle comprend des provisions en commun avec les provisions d'ouverture mais également d'autres provisions, c'est la somme des éléments ci-dessous :
 - Les provisions de sinistres de clôture
 - L'IPM sur les provisions de clôture
 - Les provisions de primes pures et commerciales de clôture
 - Le transfert de provisions mathématiques de clôture
 - La provision mathématiques de rente de clôture
 - La participation aux bénéfices de la provision mathématiques de rente
 - Les provisions pour contentieux de clôture
 - Les prévisions de recours restant à encaisser de clôture
 - La participation aux bénéfices incorporée dans l'exercice aux provisions de sinistres
 - Les provisions mathématiques pures pour risque en cours de clôture
 - Les recours encaissés
- *Variation de provision = Provisions de clôture – Provisions d'ouverture*

B Prédiction du PNA pour l'année 2020

Grâce au modèle retenu au global, le XGBoost BPCE NEW, nous avons décidé de prédire le PNA pour l'année 2020 à partir des données présentes au 31/12/2019. Voici les prédictions que nous avons obtenues :

Famille de produits	PNA réel 2019 (k€)	PNA prédit par XGBoost BPCE NEW 2020 (k€)
FAMILLE FRUCTI-FAMILLE	12 690	13 296
SECUR FAMILLE CE DC PTIA	6 682	11 128
FRUCTI-HOMME CLE	5 944	6 858
FRUCTI-PROFESSIONNEL	3 447	4 090
FRUCTI-PATRIMOINE	1 610	1 149
SECUR URGENCE CE	842	1 180
FAMILLE FRUCTI-BUDGET	501	495
DEDICACE	466	550
FAMILLE ASSURANCE FAMILLE	408	1 352
ASSURANCE HOMME-CLE C.E.	162	71
CREDIT MARITIME VIE	73	31
CERTI CAPITAL	36	37
FAMILLE FRUCTI-FACILITES	29	-49
SECURI-CREDIT BPROP	24	5
FAMILLE DOUBLE SECURITE	7	11
FRUCTI-FACILITES PRO	1	5
FAMILLE FRUCTI-SECURITE	-12	1
FRUCTI-FACILITES AGRI	-74	90
D.S.EPARGNE SALARIALE	-274	13

Résultats PNA XGBoost BPCE NEW 2020

Nous avons décidé d'observer les écarts entre les prédictions obtenues pour le PNA au 31/12/2020 et le PNA réel au 31/12/2019, y compris les affaires nouvelles. Nous avons remarqué que les prédictions obtenues pour la plupart des familles de produits importantes étaient assez proches du PNA réel 2019 excepté pour Secur Famille CE DC PTIA, où le PNA prédit pour 2020 était nettement supérieur au PNA réel 2019. En effet, c'est un produit en montée en charge donc l'écart peut être expliqué par les affaires nouvelles (celles de 2019 n'ont pas été présentes toute l'année donc nous n'avons pas reçu la totalité des primes alors qu'en 2020, la prime annuelle de ces affaires nouvelles sera comptabilisée entièrement). Toutefois, ces prédictions n'ont pas été réalisées pour être comparées au PNA réel de l'année précédente mais pour avoir une idée du PNA que nous obtiendrons l'an prochain.