

**Mémoire présenté devant l'Institut du Risk Management
pour la validation du cursus à la Formation d'Actuaire
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires**

le

10 mars 2023

Par : Arnaud LE SERGENT

Titre : Étude du risque de modèle en prévoyance et application à un risque dépendance

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

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

Membres présents du jury de l'Institut des
actuaires :

Entreprise :

Nom : BAILEY

Signature et Cachet :

LA BANQUE POSTALE PRÉVOYANCE
90, boulevard Gallieni
92130 Issy-les-Moulineaux
RCS Nanterre n° 419 901 269
N° TVA intra-communautaire FR30419901269

Membres présents du jury de l'Institut du Risk
Management :

Directeur de mémoire en entreprise :

Nom : BAILEY Florence

Signature :

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



Secrétariat :

Bibliothèque :

Signature(s) du candidat(s)



Table des matières

Résumé.....	4
Remerciements.....	6
1. Introduction.....	7
2. Chapitre 1 - Présentation du marché de l'assurance prévoyance et dépendance en France.....	10
2.1. Présentation du marché de la prévoyance en France.....	11
2.2. Présentation des enjeux de dépendance en France.....	12
3. Chapitre 2 - Présentation de l'activité de prévoyance au sein de La Banque Postale et présentation des deux approches retenues d'évaluation et de gestion du risque de modèle.....	15
3.1. Contexte.....	15
3.2. Présentation de l'activité de LBP Prévoyance.....	16
3.3. Présentation de l'approche de scoring.....	21
3.4. Présentation de l'approche Worst Case.....	29
4. Chapitre 3 - Mise en place de la méthode de scoring pour un panel de modèles.....	37
4.1. Application de l'Outil d'aide à la décision, permettant de cartographier les modèles de LBP Prévoyance.....	37
4.2. Evaluation du risque brut.....	38
4.3. Evaluation des dispositifs d'atténuation du risque brut des modèles.....	41
4.4. Evaluation de la pertinence des données et méthodes et de la complexité des modèles étudiés.....	44
4.5. Evaluation du risque net des modèles étudiés.....	44
5. Chapitre 4 – Mise en place de l'approche Worst Case au niveau de l'estimation du best estimate du produit Protectys Autonomie.....	49
5.1. Etude des lois d'expérience construites dans le cadre du mémoire « Construction de lois d'expérience en assurance dépendance.....	49
5.2. Définition de l'appétence au risque de modèle.....	57
5.3. Comparaison des lois de référence et des lois d'expérience.....	58
5.4. Résultats de l'approche <i>Worst Case</i> : <i>comparaison des best estimate obtenus entre les lois de référence et les lois de comparaison</i>	71
5.5. Mise en place de sensibilités.....	72
6. Chapitre 5 – Mise en place d'un plan d'actions pour réduire la volatilité du BE : étude des lois de réduction du produit Protectys autonomie.....	76
6.1. Contexte réglementaire.....	78
6.2. Littérature.....	78
6.3. Description et traitement des données relatives aux résiliations et réductions.....	78
6.4. Analyse univariée.....	83
6.5. Application de méthodes de machines learning de prédiction.....	90
6.6. Conclusions et limites de la troisième approche.....	92
7. Conclusion.....	94
8. Bibliographie.....	99
9. Annexes.....	101

Résumé

L'activité d'assurance nécessite de réaliser des anticipations. Celles-ci sont réalisées grâce à des modèles prédictifs, qui induisent un risque d'erreur pouvant provoquer des décisions inadaptées. Dans ce cadre, ce mémoire propose trois approches complémentaires pour évaluer et gérer ce risque de modèle.

La 1^{ère} se base sur des applications d'Enterprise Risk Management. Elle consiste à cartographier les modèles d'une entité, à évaluer leur risque brut puis net et à conclure sur l'acceptabilité de ce dernier. Des plans d'action sont proposés si celle-ci n'est pas respectée.

La 2^{ème} s'inspire de la méthode Worst Case, de Kerkhof et al (2002). Elle consiste à mesurer l'écart entre les résultats d'un modèle de référence et les résultats d'autres modèles. Le risque de modèle est alors évalué comme l'écart entre les résultats du modèle de référence et les résultats les plus éloignés du modèle de référence.

La 3^{ème} approche consiste à étudier un plan d'action de réduction du risque de modèle par identification des profils d'assurés réalisant des réductions en utilisant des méthodes d'analyses univariées et de machine learning de prédiction. Cette identification permet de déterminer des actions afin de réduire les réductions, ce qui permet de réduire le risque de sous-estimation du best estimate, les réductions ayant un impact positif sur ce dernier.

Les trois approches étudiées sont complémentaires. La 1^{ère} permet d'évaluer, en grande maille, le risque de plusieurs modèles. La 2^{ème} permet d'évaluer, de manière plus fine, le risque d'un modèle en particulier. La 3^{ème} permet de mettre en place un plan d'action visant à réduire le risque du modèle préétudié.

Mots clés : risque de modèle, machine learning, enterprise risk management, gestion des risques, approche worst case, classification and regression tree, forêts aléatoires, prévoyance, dépendance.

Abstract

The insurance business requires anticipating costs. These anticipations are carried out thanks to predictive models, inducing a risk of error which can trigger unsuitable decisions. In this context, this thesis proposes three complementary approaches to assess model risk.

The 1st is based on Enterprise Risk Management theories. It consists of referencing the models of an entity, evaluating their gross and net risk and to conclude on the acceptability of the net risk. Action plans are proposed if the risk appetite is not respected.

The 2nd is inspired by the Worst-Case method, by Kerkhof and al (2002). It consists of measuring the difference between the results of a reference model and the results of other methods. The model risk is assessed as the difference between the results of the reference model and the results the furthest from the reference model. This method is here applied on a dependency best estimate valuation model.

The 3rd approach consists in studying a model risk reduction action plan by identifying the profiles of policyholders making reductions, using univariate analysis and machine learning prediction methods. This identification makes it possible to determine actions to reduce the reductions, which makes it possible to reduce the risk of underestimation of the best estimate, reductions being positively correlated to the latter.

The three approaches are complementary. The 1st makes it possible to evaluate, in a macro way, the risk of several models. The 2nd makes it possible to assess, in a more precise way, the risk of a particular model. The 3rd makes it possible to set up an action plan aimed at reducing the risk of the pre-studied model.

Keywords: model risk, machine learning, enterprise risk management, worst case approach, classification and regression tree, random forest, dependency.

Remerciements

Je remercie ma Directrice de mémoire, Florence Bailly, qui m'a aidé dans l'élaboration de ce mémoire.

Je remercie plus généralement La Banque Postale, La Banque Postale Prévoyance, l'Inspection Générale et l'ensemble des personnes qui m'ont soutenu dans ma formation d'actuariat et dans la rédaction de ce mémoire, en particulier, Sébastien Lepage, Antoine Imhoff, Pierre Antraygues et Jérôme Fischer.

Je remercie Marion Tsalalos qui m'a aidé à obtenir et à comprendre les données utilisées dans ce mémoire.

Je remercie également l'ensemble de l'équipe pédagogique de la formation Certificat d'Expertise Actuarielle et de Sorbonne Université, pour leur enseignement et leur accompagnement.

Enfin, merci à ma famille, qui m'a soutenu.

1. Introduction

1.1. Définition du risque de modèle en assurance

L'activité d'assurance dispose d'un cycle de production inversé. En effet, les assureurs ne connaissent leurs coûts qu'après avoir fixé le prix de leurs produits. Dans ce cadre, ces derniers doivent anticiper ces coûts grâce à des modèles prédictifs. On utilise par exemple des modèles de tarification, de provisionnement ou encore de solvabilité. Ces modèles induisent des risques de modélisation.

Plusieurs définitions du risque de modèle existent dans la littérature. On peut citer la définition de l'Arrêté du 3 novembre 2014, relatif au contrôle interne du secteur bancaire, qui le définit comme le risque de pertes générées par des décisions basées sur les résultats des modèles. A cette définition bancaire, on peut ajouter une définition plus assurantielle de l'erreur de modèle importante, issue du Règlement délégué (UE) 2015/35 complétant la Directive Solvabilité II : « [une erreur de modèle] est considérée comme importante lorsqu'elle est susceptible d'influer sur la prise de décision ou le jugement des utilisateurs de cette information ». A titre d'illustration, l'utilisation d'un modèle erroné peut amener à une sous-estimation des sinistres futurs d'un assureur et donc à un risque de perte ou de faillite.

Ces définitions induisent de nombreuses causes racines différentes. Le risque ou l'erreur de modèle peut ainsi intervenir à n'importe quelle étape du cycle de vie du modèle et peut être lié à des défaillances et faiblesses dans les processus opérationnels, par exemple lors des phases de conception ou de mise en œuvre des modèles, ou à la fiabilité intrinsèque du modèle.

Dans ce cadre, l'Arrêté du 3 novembre 2014, relatif au contrôle interne du secteur bancaire et auquel est soumis La Banque Postale, dispose que les établissements financiers sont dans l'obligation de mettre en place une gouvernance solide dans le cadre de la gestion et du suivi des modèles.

1.2. Problématique

Face à ces enjeux, le présent mémoire a pour objectif de répondre à la problématique suivante :

Comment évaluer et gérer le risque de modèle au niveau de l'entité La Banque Postale Prévoyance ?

1.3. Approches retenues

Afin de répondre à cette problématique, trois approches ont été retenues :

Une première approche, basée sur les concepts d'Enterprise Risk Management (cf. cours de gestion des risques de la formation 2^{nde} année du Certificat Expertise Actuarielle de D. Dubois - 2021, Sorbonne Université - et formation Actuaire Expert ERM de E. Renouf et G. Le Rhun – 2014 - « Le risque de modèle – Mise en pratique dans un contexte de réassurance-vie ») **et s'appuyant sur une méthode de scoring :**

La méthodologie correspondante a été construite dans le cadre de ce mémoire. L'approche consiste à (i) cartographier les modèles de La Banque Postale Prévoyance grâce à un outil

d'aide à la décision ; (ii) proposer une évaluation du risque brut de chaque modèle de La Banque Postale Prévoyance ; (iii) évaluer le dispositif d'atténuation du risque brut de chaque modèle ; (iv) évaluer le risque net des modèles étudiés ; (v) conclure sur l'acceptabilité de ce dernier en fonction de l'appétence au risque et mettre en place, si nécessaire, des plans d'action afin de le réduire.

Une deuxième approche, basée sur la méthode Worst Case, proposée par Kerkhof et al (2002), reprise dans le papier de Nicole Branger et al (2004) puis dans un article du Journal of Banking & Finance (2010) :

Cette méthode, appliquée spécifiquement dans le cadre de ce mémoire sur un modèle de dépendance de La Banque Postale Prévoyance, consiste à mesurer l'écart entre les résultats d'un modèle utilisé par l'entreprise, le modèle de référence, et les résultats qui seraient induits par l'utilisation d'autres méthodes. Le risque de modèle est alors évalué comme l'écart entre les résultats du modèle de référence et de la méthode qui donne les résultats les plus éloignés du modèle de référence.

Un modèle de dépendance est, en particulier étudié dans ce mémoire. L'objectif est de mesurer, pour le modèle choisi, les écarts en termes de best estimate entre le modèle de référence et des modèles de comparaison, et de conclure, en fonction de l'appétence au risque, sur la nécessité de revoir le modèle en place. Un précédent mémoire, intitulé « Construction de lois d'expérience en assurance dépendance », réalisé par Marion Tsagalos (2020) au sein de La Banque Postale, qui a consisté à proposer des lois d'expérience, sera utilisé pour proposer des lois de comparaison au niveau des lois de dépendance. De plus, des sensibilités sur les lois de dépendance du modèle de référence seront étudiées afin de compléter l'approche.

Une troisième approche, inspirée (i) du mémoire d'actuariat « Etude des profils de résiliation sur un portefeuille santé individuelle » (A. Gehler, 2009), qui présente et utilise différentes méthodes d'analyses appliquées afin de déterminer des profils de résiliation ; complétée (ii) par l'application de méthodes de machine learning de prédiction inspirées des cours d'O. Lopez donnés dans le cadre du Certificat d'Expertise Actuarielle (La Sorbonne Université, 2021) et de l'étude « Tarification des risques en assurance-non-vie, une approche par modèle d'apprentissage statistique » (A. Paglia et al., 2011) :

L'objectif de cette troisième approche est d'étudier la mise en place d'un plan d'action visant à réduire le risque de modèle relatif au modèle de dépendance préétudié en identifiant les profils des assurés réalisant des réductions de leur couverture d'assurance afin d'être en mesure de mettre en place des actions commerciales afin de réduire ces réductions, ce qui permettrait de réduire le risque de sous-estimation du best estimate, les réductions ayant un impact positif sur ce dernier.

1.4. Complémentarité des approches

Ces trois approches sont complémentaires et fonctionnent en entonnoir, c'est-à-dire avec un périmètre d'application d'abord large puis de plus en plus précis.

La première, qui est celle qui est utilisée en pratique en entreprise, permet d'évaluer le risque de modèle de manière simple, en grande maille et sur plusieurs modèles. On l'applique ici aux modèles les plus significatifs de La Banque Postale Prévoyance.

La deuxième est plus précise, mais plus chronophage, et permet d'évaluer le risque d'un modèle en particulier. On choisit d'étudier ici, en particulier, un modèle de dépendance qui est un des modèles induisant le plus de risque au sein de La Banque Postale Prévoyance.

La troisième méthode est appliquée sur un périmètre encore plus précis et permet l'étude d'un plan d'action visant à réduire le risque du modèle de dépendance, par application d'analyses univariées et méthodes de machine learning de prédiction.

1.5. Plan du mémoire

Ce mémoire est composé de cinq chapitres : (i) Présentation du marché de l'assurance prévoyance ; (ii) Présentation de l'activité de prévoyance au sein de La Banque Postale et présentation des deux premières approches d'évaluation et de gestion du risque de modèle ; (iii) Mise en place de la méthode de scoring pour un panel de modèles ; (iv) Mise en place de l'approche Worst Case au niveau de l'estimation du best estimate d'un produit dépendance ; (v) Etude d'un plan d'action de réduction du risque de modèle par identification des profils d'assurés réalisant des réductions en utilisant des méthodes d'analyses univariées et de machine learning de prédiction.

1. Chapitre 1 - Présentation du marché de l'assurance prévoyance en France et focus sur la dépendance

1.1. Présentation du marché de la prévoyance en France

Le marché de l'assurance de personne se répartit en France entre, d'une part, l'épargne et la retraite et, d'autre part, la santé et la prévoyance, cf. figure ci-dessous (la rédaction du présent mémoire ayant débuté en 2021, les données ont été arrêtées à 2020).

Cotisations en France en 2020 (Mds €)

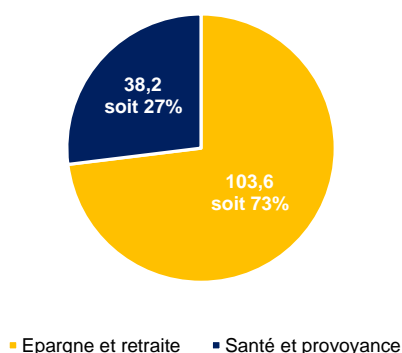


Figure 1 - Titre : répartition des cotisations d'assurance de personne en France en 2020
Source : France Assureurs, L'assurance française, données clés 2020

Les contrats d'assurance santé et de prévoyance couvrent notamment des prestations liées à des frais de maladies et d'incapacité, invalidité et dépendance.

Ce marché est en croissance, cf. figure ci-dessous.

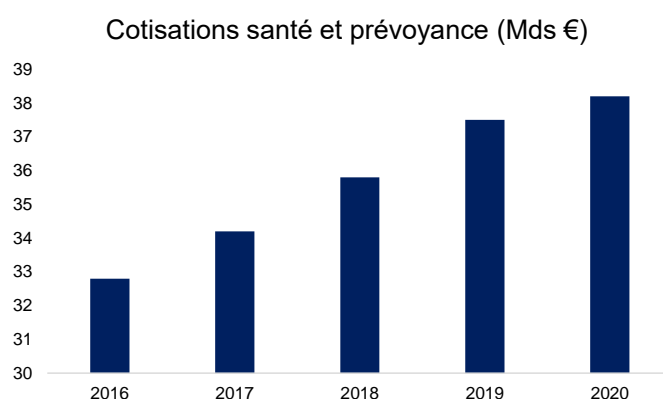


Figure 2 - Titre : évolution des montants de cotisations en assurance santé et prévoyance en France
Source : France Assureurs, L'assurance française, données clés 2020

Mécaniquement, les prestations de ce marché augmentent également, cf. figure ci-dessous.

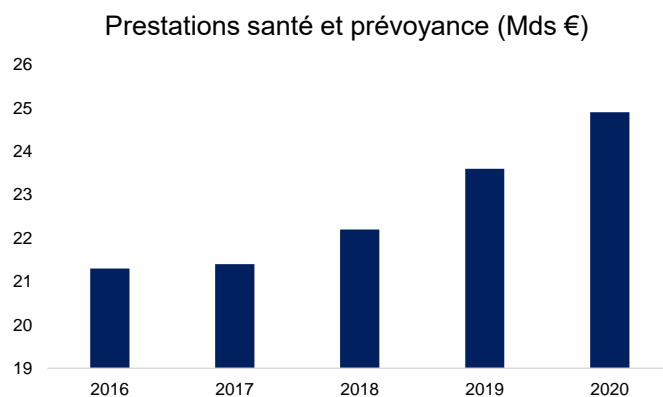


Figure 3 - Titre : évolution des montants de prestations en assurance prévoyance en France
Source : France Assureurs, L'assurance française, données clés 2020

En particulier, au sein de l'assurance prévoyance, les contrats couvrent notamment (i) le versement d'un capital ou d'une rente en cas d'incapacité temporaire de travail, d'invalidité ou de décès de l'assuré ou (ii) le versement de prestations pour couvrir en partie les frais liés à une perte d'autonomie.

La prévoyance, en particulier, a connu la croissance suivante, en termes de cotisations en France, cf. figure ci-dessous.

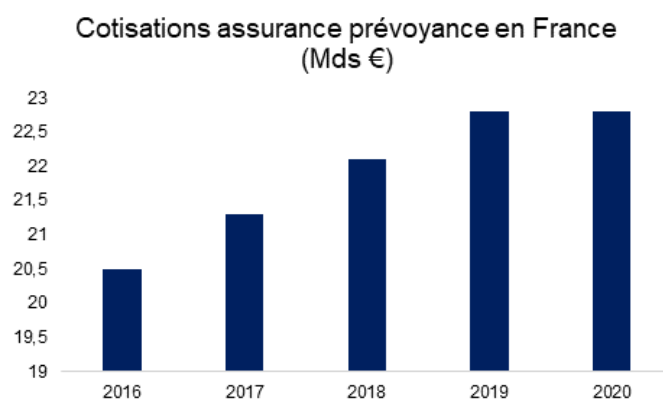


Figure 4 - Titre : évolution des montants de cotisations en assurance prévoyance en France
Source : France Assureurs, L'assurance française, données clés 2020

Le marché de la prévoyance fait face à des enjeux de croissance en France, qu'il est nécessaire d'accompagner afin d'en limiter les risques. Ainsi, ce mémoire s'inscrit dans ce contexte et propose une approche afin d'évaluer le risque de modèle au sein de l'activité Prévoyance de La Banque Postale.

La Banque Postale Prévoyance propose notamment des contrats d'assurance dépendance dont les enjeux sont présentés ci-dessous.

1.2. Présentation des enjeux de dépendance en France

D'après l'INSEE, « la population française continue de vieillir », comme le montre le graphique ci-dessous.

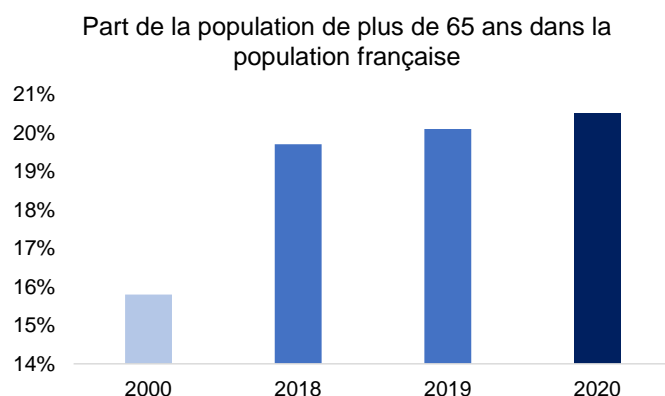


Figure 5 - Titre : évolution de la part de la population française de plus de 65 ans
Source : INSEE, Tableaux de l'économie française, édition 2020

Ce vieillissement de la population ne va faire que s'accroître, cf. figure ci-dessous.

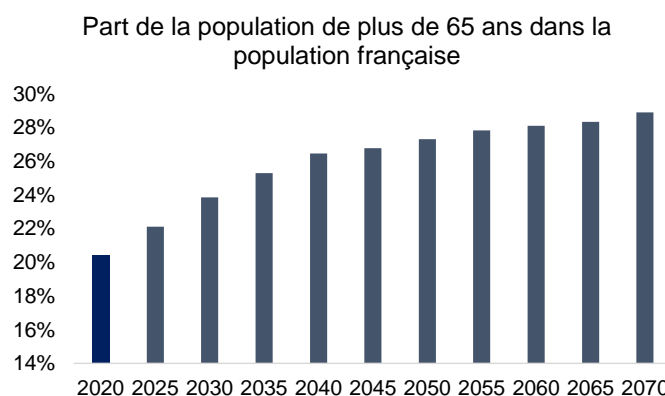


Figure 6 - Titre : projection, en scénario central, de la part de la population française de plus de 65 ans
Source : INSEE, Projections de population 2021-2070

Cette population vieillissante conduit mécaniquement à une augmentation de la dépendance.

Ce constat est connu depuis longtemps. Ainsi en 2006, Michel DUEE et Cyril REBILLARD indiquaient dans un article qu'« en supposant une stabilité de la durée de vie moyenne en dépendance, 1 200 000 personnes seront dépendantes en 2040, contre 800 000 actuellement. » (Michel DUEE, Cyril REBILLARD, 2006, La dépendance des personnes âgées : une projection en 2040).

En 2011, dans une étude soutenue par le gouvernement français (Axel RAHOA, 2011), il était estimé que le nombre de personnes âgées dépendantes allait plus que doubler entre 2010 et 2060, avec un nombre de personnes dépendante de près de 2 millions en 2040 (contre 1,2 million d'après l'article ci-dessus), cf. figure ci-dessous.

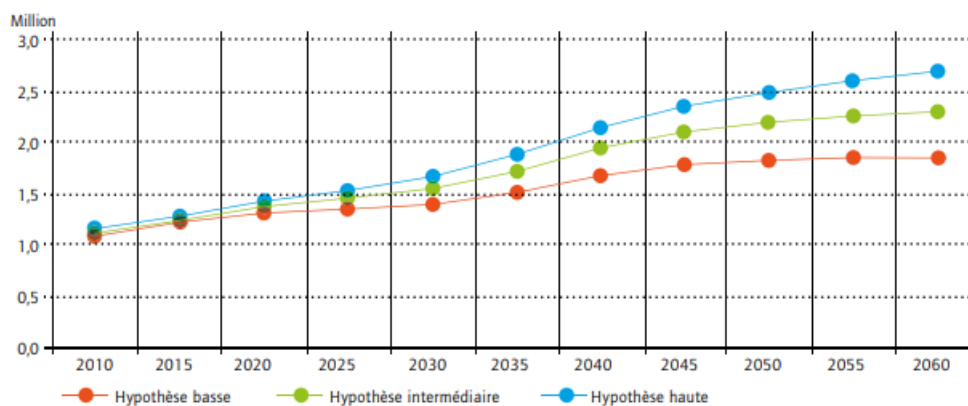


Figure 7 - Titre : projection du nombre de personnes âgées dépendantes
 Source : Synthèse du débat national sur la dépendance, Axel RAHOLA, juin 2011

Face à cette augmentation de la dépendance, il est nécessaire de mettre en place des mécanismes d'accompagnement. Un de ces mécanismes est l'assurance. Or l'assurance comporte, de manière intrinsèque, des risques qu'il est nécessaire de piloter. Un de ces risques est le risque de modèle, qui peut par exemple amener à une estimation erronée du nombre de dépendants. A titre d'illustration, le nombre de dépendants estimés en 2040 dans les deux études présentées ci-dessus était respectivement de 1,2 M dans la première étude versus près de 2 M dans la seconde.

Les estimations du nombre de dépendants, ou plus largement des sinistres en prévoyance, impactent les estimations réalisées au sein des entreprises d'assurance, par exemple en termes de tarification ou de provisionnement. Il est donc intéressant d'encadrer ces modèles d'estimation et de mettre en place des mécanismes d'atténuation du risque de modèle. Ce mémoire s'inscrit dans ce cadre et vise à évaluer le risque de modèle au sein de La Banque Postale Prévoyance.

2. Chapitre 2 - Présentation de l'activité de prévoyance au sein de La Banque Postale et présentation des deux approches retenues d'évaluation du risque de modèle

2.1. Contexte

La Banque Postale est présente, de manière historique, dans le secteur de l'assurance via trois filiales : La Banque Postale IARD, La Banque Postale Assurance Santé et La Banque Postale Prévoyance.

Depuis mars 2020, La Banque Postale s'est rapprochée de CNP Assurances et en détient désormais la majorité du capital, cf. figure ci-dessous.

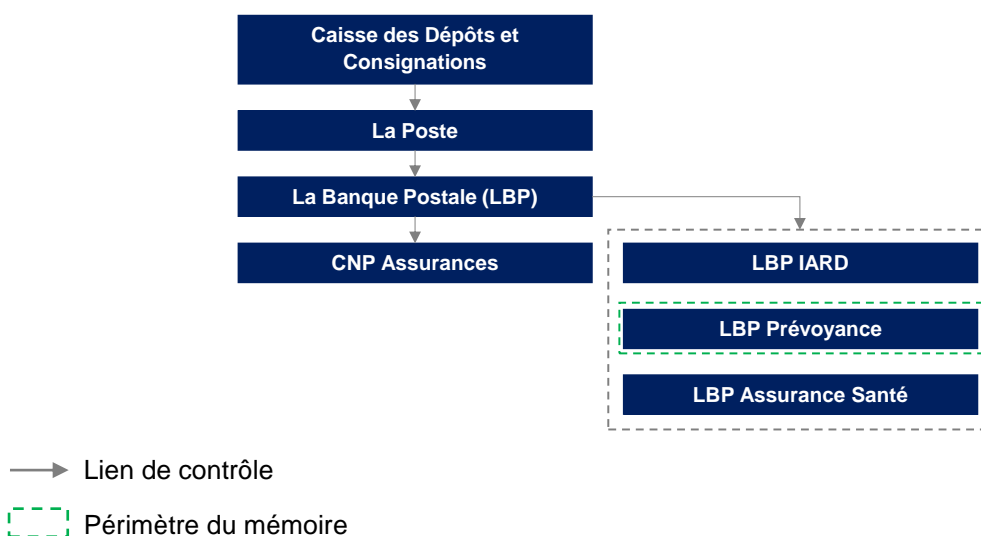


Figure 8 - Titre : Présentation des principales sociétés, en lien avec le sujet du mémoire, ayant un lien capitalistique avec La Banque Postale (avril 2022)

En décembre 2021, La Banque Postale a renforcé son taux de détention de CNP Assurances, en rachetant la totalité des 16,1% du capital de cette dernière, détenus auparavant par le Groupe BPCE. La participation totale de La Banque Postale, dans le Groupe CNP Assurances, s'élève ainsi à 78,9% en décembre 2021. Dans le même temps, La Banque Postale a annoncé vouloir acquérir les 21,1% restants, qui sont cotés en bourse.

À la suite de ce rapprochement entre La Banque Postale et CNP Assurances, l'activité d'assurance est devenue plus significative au sein du Groupe La Banque Postale et représentait 33,4% du Produit Net Bancaire en 2020.

Cette opération induit des travaux de convergence des méthodes, outils et processus utilisés dans les différentes entités. En particulier, des travaux sont menés afin de construire une vision commune au niveau de la gestion des risques et notamment, du risque de modèle.

Cette étude devrait permettre, d'une part, de mieux appréhender ce risque et, d'autre part, d'enrichir les discussions avec les autres entités du Groupe (notamment la Caisse des Dépôts et Consignations, la Direction des risques Groupe de La Banque Postale et CNP Assurances). Par conséquent, en plus de faire référence à la littérature actuarielle, ce mémoire professionnel fait également référence à des approches méthodologiques utilisées dans certaines de ces entités.

2.2. Présentation de l'activité de LBP Prévoyance

La Banque Postale est présente sur le segment de la prévoyance individuelle via sa filiale La Banque Postale Prévoyance. Cette filiale est présente sur le secteur de l'assurance de personnes mixte, en France uniquement.

Les principaux produits commercialisés par La Banque Postale Prévoyance sont des produits temporaire décès, temporaire décès accidentel, obsèques, dépendance, accidents de la vie, incapacité / invalidité et santé.

Ces produits sont distribués grâce à quatre canaux :

- Les agences bancaires, logées pour la majorité dans les bureaux de poste ;
- Un canal à distance, « La Banque Postale Chez Soi » ;
- Les équipes de middle et back office du Réseau de La Banque Postale (« centres financiers ») ;
- Le site internet de La Banque Postale.

En 2020, d'après son SFCR (Solvency and Financial Condition Reports), LBP Prévoyance cumulait 481 M € de primes émises brutes et 13 M € de résultat de souscription, cf. tableau ci-dessous.

Résultat de souscription	2019 (M €)	2020 (M €)	%
Primes émises brutes	568,0	480,9	-15,3%
Primes acquises brutes	557,2	481,2	-13,6%
Primes acquises cédées	23,7	24,5	3,7%
Sinistres bruts	256,5	245,3	-4,4%
Variation des autres provisions techniques	8,0	45,9	476,7%
Sinistres et autres provisions techniques cédés	19,2	14,8	-22,9%
Dépenses engagées	221,5	167,3	-24,5%
Résultat de souscription	66,9	13,0	-80,5%

Tableau 1 - Titre : Evolution du résultat de souscription de La Banque Postale Prévoyance
Source : SFCR 2020 de La Banque Postale Prévoyance

La variation du résultat de souscription s'explique notamment par la diminution des primes émises et par la hausse des charges de sinistre brute provenant d'un ajustement des paramètres de provisionnement.

La répartition des primes brutes, par produit, était la suivante :

Primes émises brutes de réassurance	2020 (M €)	%
Temporaire décès	91,7	19,1%
Assurance emprunteur	83,1	17,3%
Dépendance	74,8	15,6%
Contrat collectif personnel de La Poste	66,7	13,9%

Primes émises brutes de réassurance	2020 (M €)	%
Obsèques	61,6	12,8%
Assurance-crédit à la consommation	52,0	10,8%
Accidents de la vie	26,9	5,6%
Temporaire décès accidentel	20,9	4,3%
Santé	2,8	0,6%
Incapacité / invalidité	0,2	0,0%

 Périmètre d'application de l'approche de *scoring*


 Périmètre d'application de l'approche *Worst Case*

Tableau 2 - Titre : Répartition des primes, en 2020 et par produits, de La Banque Postale Prévoyance
Source : SFCR 2020 de La Banque Postale Prévoyance

La dépendance est un des produits les plus significatifs, en termes de primes.

En 2020, la répartition du résultat de souscription par ligne d'activité réglementaires était la suivante :

Compte de résultat	Assurance non-vie			Assurance vie				Total
	Assurance de revenus	Pertes pécuniaires	Frais médicaux	Assurance vie avec participation	Assurance santé	Réassurance vie	Réassurance santé	
Primes émises brutes	27,2	4,0	47,6	164,0	112,9	73,5	51,7	480,9
Primes acquises brutes	27,3	4,0	47,6	164,0	113,1	73,5	51,7	481,2
Primes acquises cédées	0,8	0,0	0,0	4,1	19,7	0,0	0,0	24,5
Sinistres bruts	8,5	0,9	40,9	98,3	33,3	21,6	41,7	245,3
Variation des autres provisions techniques	0,0	0,0	30,9	7,4	32,3	-0,1	-24,6	45,9
Sinistres et autres provisions techniques cédés	1,3	0,0	0,0	0,9	12,6	0,0	0,0	14,8
Dépenses engagées	3,4	1,4	4,2	76,4	21,4	47,2	13,3	167,3
Résultat de souscription	15,9	1,7	-28,5	-21,3	19,0	4,8	21,3	13,0

Tableau 3 - Titre : Résultat de souscription, par ligne d'activité réglementaire, de La Banque Postale Prévoyance
Source : SFCR 2020 de La Banque Postale Prévoyance

La dépendance fait partie de la ligne d'activité « Assurance santé ». Son résultat de souscription est le plus important au sein de La Banque Postale Prévoyance. Il était de 26,5 M € en 2019. Sa diminution vient d'une augmentation des provisions techniques.

L'actif de La Banque Postale Prévoyance se répartit tel que suit :

Types d'actifs	Montants en 2019 (Solvabilité 2)	Montants en 2020 (Solvabilité 2)
Détentions dans des entreprises liées	-	-
Actions	144	106
Obligations	1895	2021
Fonds d'investissement	243	238
Total placements (hors UC)	2282	2365

Tableau 4 - Titre : Répartition des actifs de La Banque Postale Prévoyance
Source : SFCR 2020 de La Banque Postale

De manière globale, les placements sont réalisés sur des actifs à risque modéré et/ou cotés. Ils sont donc peu exposés au risque de modèle.

Le SFCR 2020 expose également la répartition, par lignes d'activité, des provisions Best-Estimate de l'entité, cf. tableau ci-dessous.

Lignes d'activité	BE nets de réassurance au 31/12/2020 (M €)	%
Assurance avec participation aux bénéfices	968,6	73,0%
Assurance santé	198,2	14,9%
Assurance de protection de revenu	14,6	1,1%
Réassurance santé	124,9	9,4%
Réassurance proportionnelle - Assurance des frais médicaux	41,5	3,1%
Réassurance proportionnelle - Assurance pertes pécuniaires diverses	1,7	0,1%
Assurance des frais médicaux	0,3	0,0%
Réassurance vie	-23,8	-1,8%

Périmètre d'étude du mémoire

Tableau 5 - Titre : Répartition des BE, en 2020 et par lignes d'activité, de La Banque Postale Prévoyance
Source : SFCR 2020 de La Banque Postale Prévoyance

Il est indiqué dans le SFCR 2020 que le principal risque auquel est exposée La Banque Postale Prévoyance est le risque de souscription sur la ligne d'activité santé, qui représente 52% du BSCR (Basic Solvency Capital Requirement) avant diversification.

Face à ces enjeux en matière de risque, La Banque Postale Prévoyance a mis en place une stratégie de gestion des risques. Tout d'abord, l'objectif est de déterminer les risques que l'entité souhaite prendre et à quel niveau. Puis, pour piloter ces derniers, des dispositifs d'atténuation sont établis, par exemple : règles de souscription, de provisionnement et réassurance. De plus, des indicateurs de suivi sont utilisés, selon les trois critères suivants : impacts sur le capital, la liquidité ou le résultat.

Enfin, comme requis par la réglementation, La Banque Postale Prévoyance a mis en place un processus ORSA (Own Risk Solvency Assesment). D'après son SFCR, le processus ORSA de La Banque Postale Prévoyance « doit permettre :

- D'identifier les risques auxquels La Banque Postale Prévoyance est exposée,
- De quantifier son besoin global de solvabilité sur l'horizon du Business Plan,
- De vérifier sa capacité à respecter les exigences réglementaires du SCR et du MCR et des exigences concernant le calcul des provisions techniques sur l'horizon du Business Plan,
- D'évaluer si son profil de risque s'écarte des hypothèses utilisées pour calibrer les calculs de SCR par la formule standard. »

La Banque Postale réalise l'évaluation de son SCR (Solvency Capital Requirement) en formule standard. Le risque de modèle est considéré comme partie du risque opérationnel. Or, dans l'évaluation du SCR en formule standard, le risque opérationnel est calculé de manière forfaitaire. En effet, d'après la présentation de l'Institut des Actuaire, « Le risque opérationnel, un nouveau challenge pour l'actuaire » (2016) :

$$SCR_{op} = \min(0,3 * BSCR ; Op) + 0,25 * Exp_{ul};$$

Avec :

- *BSCR : capital de solvabilité requis de base ;*
- *Op : capital de base pour risque opérationnel (hors UC), calculé comme étant le max (Oppremiums ; Opprovisions), autrement dit : - Oppremiums : capital requis pour risque opérationnel sur base des primes acquises vie (hors UC) et non-vie (des 12 derniers mois et éventuellement des 12 mois précédents si augmentation significative (+20%) des primes). Avec coefficient ≈ 4% des primes en vie et 3% pour la non-vie. - Opprovisions : capital requis pour risque opérationnel sur base des provisions techniques vie (hors UC) et non-vie Avec coefficient ≈ 0.45% des PT vie et 3% pour la non-vie. o Expul : dépenses UC annuelles (12 derniers mois).*

Ainsi, l'évaluation du risque de modèle réalisée dans ce mémoire pourra permettre d'alimenter, pour cet item et sur le périmètre étudié, l'étude de l'écart entre le profil de risque de La Banque Postale Prévoyance et les hypothèses utilisées pour l'évaluation du risque opérationnel par la formule standard.

Ce mémoire entre plus globalement dans le cadre de la gestion des risques de La Banque Postale Prévoyance et propose différentes approches pour piloter le risque de modèle, qui peut avoir un impact sur le montant des provisions et donc sur l'ensemble des critères suscités (capital, liquidité, résultat).

2.2.1. Focus sur le produit de dépendance étudié dans le cadre de l'approche Worst Case

Au sein de LBP Prévoyance, trois produits dépendance sont ou ont été commercialisés : Protectys Autonomie, Assurance Autonomie et Forfait Autonomie.

En particulier, Protectys Autonomie, contrat collectif à adhésion facultative, offre, en cas de dépendance de l'assuré, une garantie de prestations forfaitaires et/ou une rente mensuelle viagère et/ou un montant de capital premiers frais. L'état de dépendance est évalué par référence à quatre Groupes-Iso Ressources (GIR) de la grille AGGIR, élaborée par l'administration française (cf. annexe 2 pour détail) et par référence aux actes de la vie quotidienne (toilette, habillement, alimentation, continence, déplacement et transferts).

Protectys Autonomie est un contrat en run-off qui représentait 93% du nombre de contrats en stock de LBP Prévoyance à fin 2018, avec 120 453 contrats, cf. tableau ci-dessous.

Année	2014	2015	2016	2017	2018
Nombre de contrats	141 503	135 320	129 745	124 670	120 453

Tableau 6 - Titre : Evolution du nombre de contrats Protectys Autonomie, à fin 2018

Ce produit n'est plus commercialisé depuis le 31 décembre 2014.

Etant donné ces chiffres et l'importance du nombre de contrats Protectys Autonomie, au sein de LBP Prévoyance, ce mémoire s'attache à appliquer l'approche Worst Case sur ce produit.

Ses caractéristiques sont les suivantes :

Caractéristiques	
Définition de la Dépendance	L'état de Dépendance est défini par référence à quatre Groupes Iso Ressources (GIR) de la grille AGGIR
Age limite	Pour la souscription, l'assuré doit être âgé de 18 ans au moins et ne pas avoir atteint son 75 ^{ème} anniversaire au jour de l'adhésion
Délai d'attente	Pour les cas de Dépendance dus à un accident : pas de délais d'attente Pour les cas de Dépendance fonctionnelle non-accidentelle : délai d'attente d'un an Pour les cas de Dépendance psychique non-accidentelle : délai d'attente de trois ans
Franchise	Franchise relative de 3 mois sur la rente Franchise relative de 3 mois sur le capital
Indexation	Chaque année au 1 ^{er} avril, la cotisation/les garanties varient selon un taux fixé dans la limite de l'évolution annuelle du plafond annuel de la Sécurité sociale au 1 ^{er} janvier de l'année en cours. Le nouveau montant de la cotisation s'applique à la date anniversaire de l'adhésion qui suit le 1 ^{er} avril
Revalorisation	Au 31 décembre de chaque année, La Banque Postale Prévoyance établit le compte de résultat de l'ensemble des opérations effectuées au titre des rentes en cours de service. Ce compte est alimenté par au moins 90 % des produits financiers nets et par le solde du compte au 31 décembre de l'exercice précédent. Une partie ou l'intégralité de ce compte sert à la revalorisation des rentes en cours de service. Ces revalorisations s'appliquent au 1 ^{er} janvier suivant
Révision	Indépendamment de l'indexation, la cotisation est révisable à la date anniversaire de l'adhésion qui suit le 1 ^{er} avril en fonction des résultats techniques du contrat d'assurance de groupe
Exonération de la cotisation	Exonération du paiement des cotisations liées au capital (respectivement rentes) dès le versement du capital (respectivement rentes)

Tableau 7 - Titre : Caractéristiques de produit Protectys Autonomie

Les états de dépendance garantis par Protectys Autonomie dépendent du choix qu'a fait l'assuré de souscrire, ou non, à la garantie complémentaire optionnelle, cf. tableau ci-dessous.

	Etats de dépendance garanties	
	Rente	Capital 1ers frais
Garantie principale	GIR1 et GIR2	GIR1, GIR2 et GIR3
Garantie complémentaire	GIR3	GIR4 + 3AVQ6

Tableau 8 - Titre : Etats de dépendance garantis, selon la souscription, ou non, de la garantie complémentaire optionnelles

Aussi, l'état de dépendance influe le type et le montant de la prestation versé, cf. tableau ci-dessous

Prestation	GIR1	GIR2	GIR3	GIR4
60% de la rente mensuelle	X	X	X (*)	
40% de la rente mensuelle	X	X		
Capital 1ers frais (**)	X	X	X	X (*)

(*) Si la garantie complémentaire a été souscrite.

(**) Le capital n'est versé qu'une seule fois.

Tableau 9 - Titre : Types et montants de prestations, selon l'état de dépendance

2.3. Présentation de l'approche de scoring

Pour rappel, ce mémoire est structuré en trois approches, allant de la plus large à la plus précise : (i) méthode large d'évaluation du risque de modèle appliquée sur plusieurs modèles (chapitres 2 - théorie - et 3 - application), (ii) méthode d'évaluation précise du risque de modèle appliquée à un modèle déterminé (chapitres 2 - théorie - et 4 - application), (iii) étude d'un plan d'action visant à réduire le risque de modèle par application d'analyses univariées et méthodes de machine learning de prédiction (chapitre 5).

Ainsi, le présent mémoire propose une approche de scoring, s'appuyant sur les concepts de gestion des risques. Elle s'inspire de plusieurs papiers, notamment le cours de gestion des risques de D. Dubois, dispensé dans le cadre du Certificat d'Expertise Actuarielle¹ et le support de la formation Actuaire Expert ERM de E. Renouf et G. Le Rhun (2014) « Le risque de modèle – Mise en pratique dans un contexte de réassurance-vie »².

La méthodologie retenue s'articule selon les étapes suivantes :

- Cartographier les modèles de La Banque Postale Prévoyance grâce à un outil d'aide à la décision ;
- Proposer une évaluation du risque brut de chaque modèle de La Banque Postale Prévoyance ;
- Evaluer le dispositif d'atténuation (gouvernance, contrôles, etc.) du risque brut de chaque modèle en se basant sur les attendus de l'Arrêté du 3 novembre 2014, relatif au contrôle interne des banques, ainsi que sur les attendus internes ;
- Evaluer le risque net du panel de modèles étudiés ;
- Conclure sur l'acceptabilité de ce dernier en fonction de l'appétence au risque et mettre en place, si nécessaire, des plans d'action afin de le réduire

Pour rappel et comme décrit dans le papier « Le risque de modèle » de Thomas Lallement et al (2014), le risque de modèle peut intervenir à chaque étape de la vie des modèles, cf. figure ci-dessous.

¹ DUBOIS D. 2021. Cours de Gestion des risques (ERM). Formation 2nde année Certificat Expertise Actuarielle. Sorbonne Université.

² RENOUF E., LE RHUN G. 2014. Le risque de modèle – Mise en pratique dans un contexte de réassurance-vie. Formation Expert ERM.

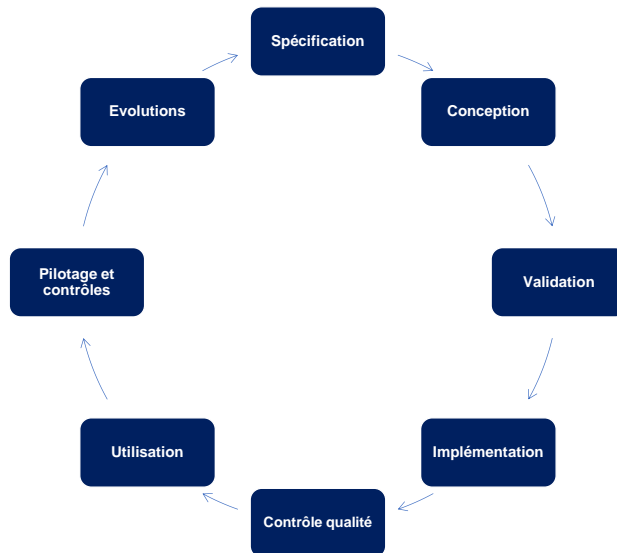


Figure 9 - Titre : Cycle de vie d'un modèle
 Source : Model risk management | Building supervisory confidence, 2018, Deloitte ; Gran Thornton. ORSA requirements Model risk management for insurance companies.

Dans ce cadre, l'approche de scoring s'inscrit dans la méthodologie de gestion des risques de La Banque Postale Prévoyance, cf. figure ci-dessous.

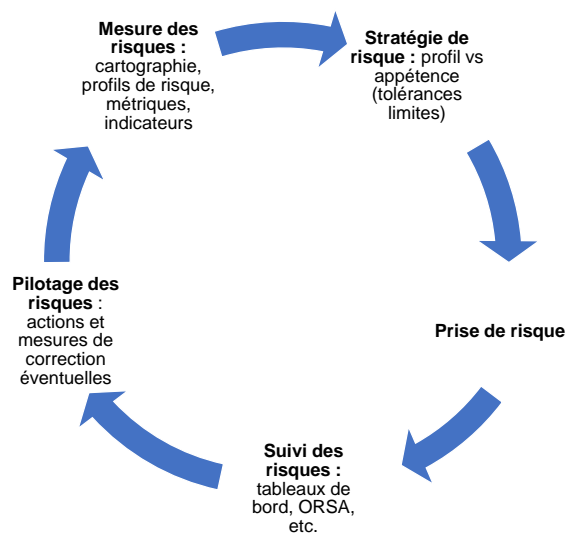


Figure 10 - Titre : Stratégie de gestion des risques de La Banque Postale Prévoyance
 Source : SFCR 2020 de La Banque Postale Prévoyance

La présentation de la méthodologie de scoring, pour chaque étape, est détaillée ci-dessous. L'application, réalisée sur un panel de modèle de LBP Prévoyance, sera présentée dans le chapitre 3 du présent mémoire.

2.3.1. Outil d'aide à la décision, permettant de cartographier les modèles de LBP Prévoyance

La première étape consiste à cartographier les modèles de LBP Prévoyance, en se basant sur un outil d'aide à la décision, présenté ci-dessous.

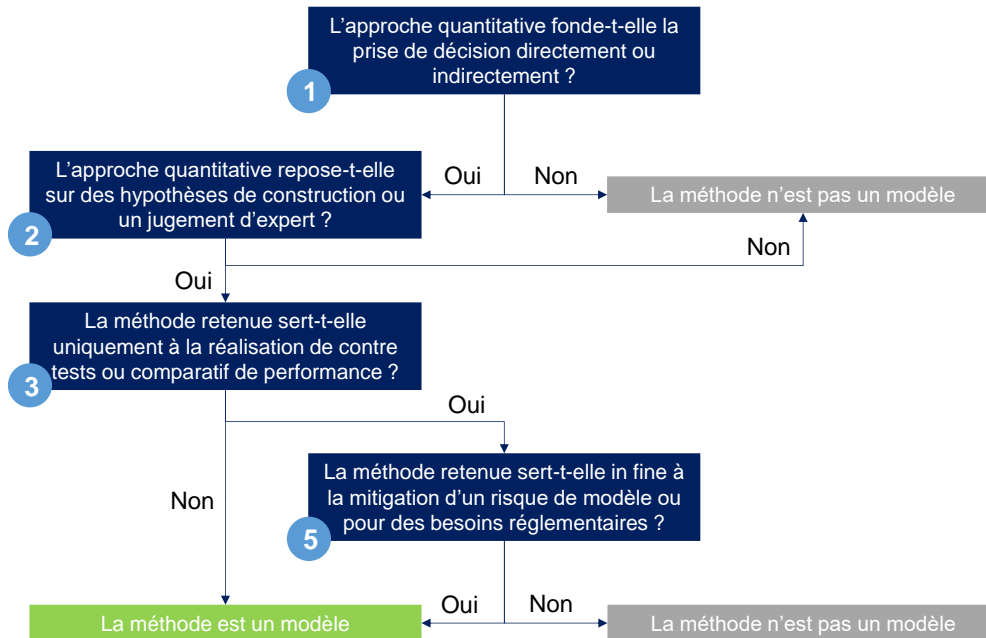


Figure 11 - Titre : Outil d'aide à la décision permettant de cartographier les modèles de LBP Prévoyance

Une fois les modèles cartographiés, le risque brut, inhérent à ces derniers, est évalué.

2.3.2. Evaluation du risque brut de chaque modèle

Au sein de LBP Prévoyance, le risque brut est évalué selon une échelle de 1 à 4 (1 = risque mineur ; 2 = risque modéré ; 3 = risque majeur ; 4 = risque critique).

Dans ce cadre, en s'inspirant des méthodologies du cours de gestion des risques de D. Dubois, dispensé dans le cadre du Certificat d'Expertise Actuarielle³, pour évaluer le risque brut dans l'outil de scoring élaboré dans le cadre de ce mémoire, la fréquence d'utilisation du modèle est croisée avec l'impact de la variable modélisée, ici estimé par son poids relatif dans le bilan ou le compte de résultat :

$$\text{Risque brut} = \text{Fréquence} * \text{Impact} ;$$

avec :

- La fréquence estimée grâce à la table de correspondance suivante.

³ DUBOIS D. 2021. Cours de Gestion des risques (ERM). Formation 2nde année Certificat Expertise Actuarielle. Sorbonne Université.

Fréquence	
Fréquence d'utilisation du modèle	Niveau de fréquence correspondant
Plusieurs fois par an	4
Annuelle	3
Moins d'une fois par an	2
Unique	1

Tableau 10 - Titre : Table de correspondance entre la fréquence d'utilisation du modèle et le niveau de fréquence correspondant

- L'impact est estimé via une évaluation du poids de la variable modélisée. Cette évaluation est réalisée :
 - o Pour les variables de bilan modélisées, en évaluant la part de la variable modélisée par rapport à un agrégat du bilan. Par exemple, le poids de la variable estimant les provisions peut être évalué en rapportant la part des provisions estimées par rapport au total bilan ou aux fonds propres ;
 - o Pour les variables de compte de résultat, en évaluant la part de la variable modélisée par rapport à un agrégat du compte de résultat. Par exemple, le poids des variables de tarification peut être évalué en rapportant la part des primes estimées par rapport au total des primes.

Une fois ce rapport obtenu, le poids est estimé grâce à une table de correspondance, cf. tableaux ci-dessous.

Variable de bilan	
Poids de la variable modélisée	Niveau d'impact correspondant
Poids < 10% du bilan de l'entité	1
$10\% \leq \text{poids} < 40\%$ du bilan de l'entité	2
$40\% \leq \text{poids} < 60\%$ du bilan de l'entité	3
60% du bilan de l'entité \leq poids	4

Tableau 11 - Titre : Table de correspondance entre poids de la variable de bilan modélisée et impact correspondant

Variable de compte de résultat	
Poids de la variable modélisée	Niveau d'impact correspondant
Poids < 10% des primes brutes d'assurance	1
$10\% \leq \text{poids} < 40\%$ des primes brutes d'assurance	2
$40\% \leq \text{poids} < 60\%$ des primes brutes d'assurance	3
60% des primes brutes d'assurance \leq poids	4

Tableau 12 - Titre : Table de correspondance entre poids de la variable de compte de résultat modélisée et impact correspondant

Une fois la fréquence et l'impact estimés, on obtient le risque brut en croisant ces deux évaluations, cf. figure ci-dessous.

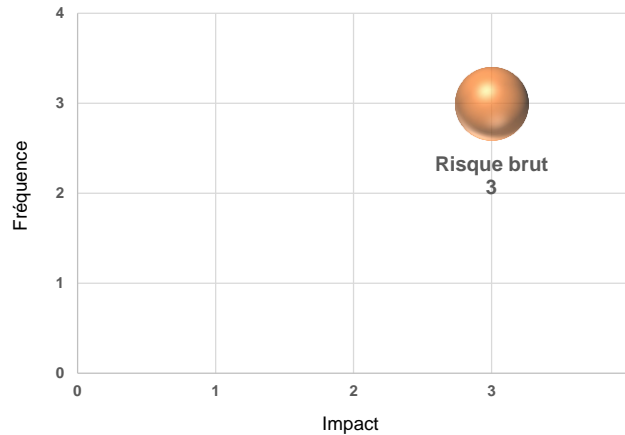


Figure 12 - Titre : Evaluation du risque brut

En fonction de l'appétence au risque de l'entité, ce risque brut est considéré comme acceptable ou non. S'il n'est pas acceptable, on étudie les dispositifs d'atténuation visant à le piloter.

2.3.3. Evaluation des dispositifs d'atténuation du risque brut des modèles étudiés

Au regard des concepts d'Enterprise Risk Management, le risque de modèle peut être réduit par la mise en place de dispositifs d'atténuation. Cela peut notamment se matérialiser par une organisation et une gouvernance des modèles appropriées, conformément à l'Arrêté du 3 novembre 2014 relatif au contrôle interne du secteur bancaire.

Dans ce cadre, le présent mémoire vise à évaluer ces dispositifs d'atténuation, pour chaque modèle étudié, en répondant aux questions suivantes.

Thématiques	Questions
Encadrement	Le modèle est-il encadré par une politique de gestion des modèles ?
	Les principales étapes de la vie du modèle font-elles l'objet de procédures à jour ?
	Une cotation du modèle a-t-elle été fournie ?
	Le modèle fait-il l'objet d'un inventaire annuel, porté par l'équipe de validation, recensant l'ensemble des modèles et leurs informations ?
Définition des rôles et responsabilités opérationnelles	Le responsable du modèle est-il désigné ?
	Le propriétaire du modèle est-il désigné ?
	Le développeur du modèle est-il désigné ?
	L'implémenteur du modèle est-il désigné ?
	L'utilisateur du modèle est-il désigné ?
	Le valideur du modèle est-il désigné et en charge de l'inventaire des modèles ?
	Le gestionnaire des risques du modèle est-il désigné ?
Des instances de gouvernance sont-elles désignées pour valider la gouvernance des modèles et notamment le dispositif de gestion des risques ?	
Pilotage	Une instance / équipe de validation a-t-elle formellement validé la mise en place du modèle ?
	Le cas échéant, cette instance revoit-elle les hypothèses liées au modèle ?
	Les réserves liées aux modèles font-elles l'objet d'un suivi ?

Thématiques	Questions
	Une revue périodique du modèle est-elle réalisée et portée par l'équipe de validation ou une instance, comprenant le suivi des contrôles continus sur la performance du modèle et des <i>back tests</i> ?
	Le modèle est-il revu annuellement en comité ou par l'équipe de validation ?
	Une instance ou équipe de validation valide-t-elle formellement les évolutions du modèle ?
Système d'information et données	Les systèmes d'informations et données utilisés pour les modèles sont-ils évalués ?
	Les données, hypothèses et modélisations sont-elles tracées et conservées ?
Contrôle permanent	Un dispositif de contrôle permanent de 1er niveau est-il en place ?
	Le cas échéant, les résultats des contrôles de 1er niveau sont-ils formalisés, suivis dans une instance et font-ils l'objet de plans d'actions mis en place ?
	Un dispositif de contrôle permanent de 2ème niveau est-il en place ?
	Le cas échéant, les résultats de contrôle de 2ème niveau sont-ils formalisés, suivis dans une instance et font-ils l'objet de plans d'actions mis en place ?
	L'audit interne a-t-il vérifié que le contrôle de premier niveau et le contrôle de deuxième niveau respectaient les différents principes et procédures de gestion du risque de modèle ?
	D'autres fonctions ont-elles un rôle de contrôles (fonctions actuarielle / risque par exemple) ?
	Le cas échéant, les résultats correspondants ont-ils fait l'objet de plans d'actions suivis ?

Tableau 13 - Titre : Méthodologie d'évaluation des dispositifs d'atténuation du risque de modèle, Sources : Arrêté du 3 novembre 2014 relatif au contrôle interne des banques, Normes internes

A ces questions, une importance de 1 à 3 est attribuée, ce qui permet de les hiérarchiser. Les réponses sont apportées à l'aide de la documentation à disposition et en réalisant des échanges avec les responsables des modèles étudiés.

Afin de pouvoir y répondre de manière homogène, des définitions ont été formulées, par exemple sur les rôles et responsabilités (cf. annexe 3 du présent mémoire, « définition des rôles »).

Au total, une note sur 100, sur le caractère complet ou non des dispositifs d'atténuation de chaque modèle étudié, est établie.

A cette note sur 100, on attribue ou non une pénalité en fonction de la pertinence et de la complexité du modèle étudié. L'idée consiste à dire que si un modèle n'est pas tout à fait pertinent, en fonction de la qualité des données utilisées par exemple, ou s'il est très complexe, alors les dispositifs d'atténuation seront moins efficaces.

Les items évalués pour la pertinence du modèle sont les suivants.

Au niveau de la qualité des données et méthodes utilisées, cf. tableau ci-dessous.

Items	Définition
Qualité des données utilisées	Si la qualité des données a été évaluée (exhaustivité, véracité, valeurs aberrantes) et ne fait pas l'objet de réserves, indiquer 1
	Si la qualité des données a été évaluée et fait l'objet de réserves non bloquantes, indiquer 2
	Si la qualité des données n'a pas été évaluée ou fait l'objet de réserves bloquantes, indiquer 3

Items	Définition
Qualité des méthodes utilisées	Si les méthodes utilisées ont été évaluées et ne font pas l'objet de réserves (exemple : méthodes classiques issues de la littérature actuarielle et ayant fait l'objet d'études sur leur caractère adapté et leur pertinence), indiquer 1
	Si les méthodes utilisées ont été évaluées et font l'objet de réserves non bloquantes (exemple : méthodes non classiques, telles que l'utilisation de nombreux proxys), indiquer 2
	Si les méthodes utilisées n'ont pas été évaluées ou font l'objet de réserves bloquantes, indiquer 3

Tableau 14 - Titre : Evaluation de la pertinence des modèles étudiés

Au niveau de la complexité du modèle, cf. tableau ci-dessous.

Items	Définition
Durée d'implémentation du modèle	Si la durée d'implémentation du modèle est inférieure à 2 mois, indiquer 1
	Si la durée d'implémentation du modèle est comprise entre 2 mois et 6 mois, indiquer 2
	Si la durée d'implémentation du modèle est supérieure à 6 mois, indiquer 3
Temps de calcul	Si le temps de calcul est inférieur à 10 minutes, indiquer 1
	Si le temps de calcul est compris entre 10 minutes et 1 heure, indiquer 2
	Si le temps de calcul est supérieur à 1 heure, indiquer 3
Imbrication / interactions avec d'autres modèles	Si le modèle n'est pas en lien avec d'autres modèles (modèle indépendant), indiquer 1
	Si le modèle dispose de plusieurs modules (exemple : modèle ALM santé), indiquer 2
	Si le modèle fait l'objet d'imbrications et d'interactions (ex : modèle ALM prévoyance), indiquer 3
Granularité du modèle (ligne à ligne versus modèle point)	Si le modèle nécessite une granularité Model point, indiquer 1
	Si le modèle nécessite une granularité très fine (exemple : ligne à ligne), indiquer 2
Projections et chocs/règles de gestion	Si le modèle ne nécessite pas de projection (calcul en t=0), indiquer 1
	Si le modèle nécessite des projections dans le temps, avec un ou des choc(s) simple(s) (c'est à dire ne nécessitant que la modification d'une ou plusieurs hypothèse(s) et pas de modifications supplémentaires des formules de calculs (exemple : choc de mortalité), indiquer 2
	Si le modèle nécessite des projections dans le temps et l'implémentation de chocs complexes (c'est à dire avec des modifications supplémentaires des formules / règles de gestion en plus de l'hypothèse), indiquer 3

Tableau 15 - Titre : Evaluation de la pertinence des modèles étudiés

Une note est ainsi calculée en croisant l'évaluation de la pertinence et celle de la complexité. On multiplie cette note par un pourcentage de pénalité défini puis on réduit d'autant l'évaluation dispositif d'atténuation réalisée au préalable :

$$\text{Pénalité} = \text{Pourcentage de pénalité défini} * \text{Pertinence} * \text{Complexité} ;$$

Evaluation finale du dispositif d'atténuation = Evaluation initiale du dispositif d'atténuation – pénalité.

2.3.4. Evaluation du risque de modèle net pour chaque modèle étudié

Pour obtenir le risque net, l'évaluation du risque brut est croisée, dans une matrice, avec l'évaluation finale des dispositifs d'atténuation, cf. figure ci-dessous.

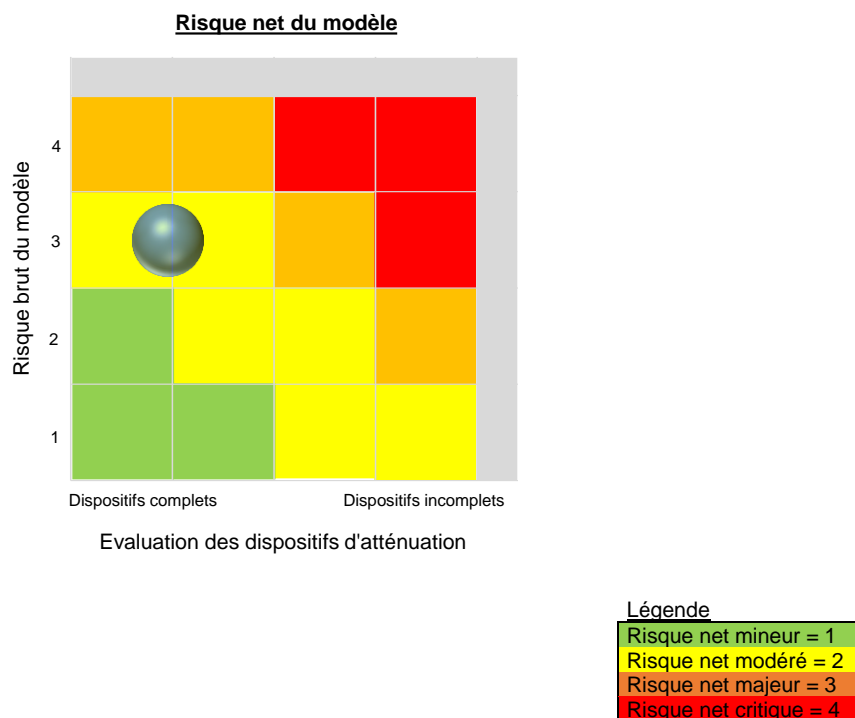


Figure 13 - Titre : Matrice d'évaluation du risque de modèle net

Ce croisement permet d'obtenir le risque net.

Enfin, en fonction de l'appétence au risque, ce risque net est considéré comme acceptable ou non. S'il ne l'est pas, des plans d'action pourront être mis en place pour le réduire, par exemple via un renforcement des dispositifs d'atténuation.

L'appétence au risque peut par exemple être défini comme la volonté de n'avoir que des modèles dont le risque net estimé est mineur ou modéré (ce qui signifierait qu'un modèle dont le risque brut est modéré, est acceptable, donc que l'évaluation des dispositifs d'atténuation et du risque net, pour ce modèle, ne serait, le cas échéant, pas nécessaire).

L'application de la méthode de scoring, au sein de LBP Prévoyance, est présentée dans le chapitre 3 du présent mémoire.

2.4. Présentation de l'approche Worst Case

Pour rappel, ce mémoire est structuré en trois approches, allant de la plus large à la plus précise : (i) méthode large d'évaluation du risque de modèle appliquée sur plusieurs modèles (chapitres 2 et 3), (ii) méthode d'évaluation précise du risque de modèle appliquée à un modèle déterminé (chapitres 2 et 4), (iii) étude d'un plan d'action visant à réduire le risque de modèle par application d'analyses univariées et méthodes de machine learning de prédiction (chapitre 5).

Proposée par Kerkhof et al (2002), l'approche Worst Case consiste à mesurer l'écart entre les résultats du modèle utilisé, le modèle de référence, et les résultats qui seraient engendrés par l'utilisation d'autres modèles.

Le risque de modèle est alors évalué comme l'écart entre les résultats du modèle de référence et de celui qui donne les résultats les plus éloignés du modèle de référence :

$$\rho (M_{réf.}) = \sup_{M_i \in M} R(M_i) - R(M_{réf.}) ;$$

avec :

- ρ = mesure du risque de modèle ;
- $M_{réf.}$ = modèle de référence ;
- M_i = modèles de comparaison ;
- R = Résultat donnée par le modèle.

Autrement dit :

Risque de modèle (modèle de référence) = sup (résultats obtenus grâce aux méthodes de comparaison – résultat obtenu grâce au modèle de référence).

On considère ici l'écart en valeur, et non pas l'écart absolu ou au carré, car le signe du résultat est important pour la présente étude. En effet, on réalise l'étude sur le best estimate, donc une sous-évaluation de ce dernier, correspondant à un résultat positif de ρ , est plus risquée que le cas contraire.

Les modèles de comparaison peuvent différer du modèle de référence de différentes manières. A titre d'illustration, les hypothèses peuvent être modifiées : données utilisées, plages de référence, variables de segmentation, hyper paramètres, etc. ; les méthodes peuvent également être modifiées : méthode de calcul des taux bruts en analyse de survie, méthodes de lissage ou encore de prolongation des lois.

L'approche Worst Case permet ainsi d'estimer le risque de modèle, notamment le risque de spécification.

D'autres approches auraient pu être retenues. Ainsi, l'approche bayésienne, proposée par DRAPER [1995] permet d'inclure des pondérations à chaque modèle de comparaison, selon sa pertinence :

$$\rho' (M_{réf.}) = \sum_i w_i [R(M_i) - R(M_{réf.})] ;$$

avec :

- ρ' = mesure du risque de modèle ;
- $M_{réf.}$ = modèle de référence ;
- w_i = poids accordé au modèle M_i en fonction de sa pertinence ;
- M_i = modèles de comparaison ;

- $R = \text{Résultat donnée par le modèle.}$

Ces deux approches, présentées également dans le papier « Measuring model risk » de Philipp Sibbertsen et al (2008) et ayant été utilisés dans d'autres mémoire d'actuariat⁴, ont des avantages et des inconvénients :

- L'approche Worst Case est plus simple à mettre en œuvre mais davantage soumise aux modèles extrêmes. Elle peut donc être trop prudente ;
- L'approche bayésienne permet de diminuer la sensibilité de l'exercice aux modèles de comparaison extrêmes. Cependant, elle est plus complexe à mettre en œuvre, du fait de la recherche des poids w_i .

Le présent mémoire retiendra l'approche Worst Case pour l'étude du risque de modèle.

2.4.1. Présentation du modèle sur lequel sera appliquée l'approche Worst Case

L'approche Worst Case est appliquée, dans le présent mémoire, sur le modèle de dépendance du produit Protectys Autonomie.

Le modèle de dépendance de Protectys Autonomie est un modèle semi-markovien c'est-à-dire à multi-états. Les états sont composés de trois lois :

- La mortalité des assurés autonomes, c'est-à-dire des non-dépendants au sens de la définition retenue (cf. ci-dessous pour définition) ;
- L'incidence de la dépendance ;
- La mortalité des assurés dépendants au sens de cette même définition.

La construction de ces lois est réalisée sur les données du portefeuille Protectys Autonomie. Leur extrapolation fait appel à des données de marché. Elles ont une sensibilité importante sur les âges élevés, du fait du caractère récent des données (le produit Protectys Autonomie a été lancé en 2004) et comportent ainsi un risque de dérive de la longévité.

Dans ce cadre, le risque de modèle de dépendance du produit Protectys Autonomie est notamment porté au sein de ces lois. L'approche Worst Case consistera ici à partir de lois d'expérience de comparaison et à en évaluer l'impact sur le best estimate.

A ces lois structurelles, s'ajoutent des lois de résiliations et réductions. Nous étudierons les réductions dans le chapitre 5 du présent mémoire et leur appliquerons des méthodes de machine learning de prédiction.

2.4.1.1. Définition de la dépendance retenue pour le produit Protectys Autonomie et modélisation correspondante :

Le modèle de dépendance du produit Protectys autonomie dispose de lois imbriquées, ce qui induit la représentation suivante de la population assurée :

- Cette dernière est d'abord séparée en deux : les assurés en vie et les assurés décédés ;
- Puis, les assurés en vie sont divisés selon la méthode suivante :

⁴ CADIOT A. *Une approche du risque de modèle lors de la tarification en réassurance non-vie.* ISFA. ; DAVESNE C. (2015). *Etude du risque de modèle dans le cadre d'un modèle interne.* ENSAE. ; DUSSAULE S. (2014). *L'erreur de modèle dans le dispositif Solvabilité II, problématiques et méthodologies.* Mémoire d'actuariat présenté à l'IA. EURIA.

- Séparation des assurés en GIR 1 ou en GIR 2, considérés en dépendance totale ;
- Séparation des assurés en GIR 1 ou en GIR 2 ou en GIR 3, considérés en dépendance totale ou partielle ;
- Séparation des assurés en GIR 1 ou en GIR 2 ou en GIR 3 ou en GIR 4, considérés en dépendance totale, partielle ou légère.

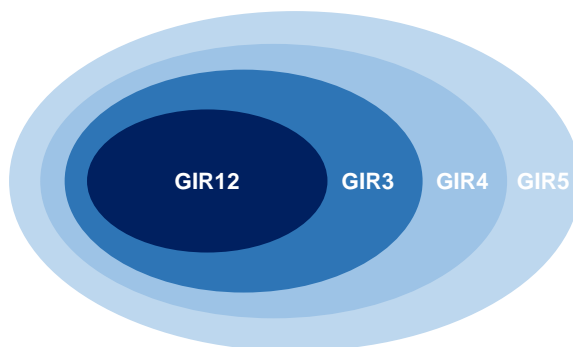


Figure 14 - Titre : Représentation de la population assurés en vie

On note dans ce cadre :

- GIR12 les assurés GIR1 ou GIR2, donc considérés en dépendance totale ;
- GIR123 les assurés GIR1 ou GIR2 ou GIR3, donc considérés en dépendance totale ou partielle ;
- GIR1234 les assurés GIR1 ou GIR2 ou GIR3 ou GIR4, donc considérés en dépendance totale, partielle ou légère ;
- Non GIR12 les assurés GIR3 ou GIR4 ou GIR5 ou GIR6 ;
- Non GIR123 les assurés GIR4 ou GIR5 ou GIR6 ;
- Non GIR1234 les assurés GIR5 ou GIR6.

Protectys Autonomie propose des garanties en fonction de ces trois niveaux de dépendance : dépendance totale (GIR12), dépendance partielle (GIR3) et dépendance légère (assurés en GIR4 et ne pouvant réaliser de manière autonome et complète trois des six actes de la vie quotidienne (toilette, habillage, alimentation, continence, déplacement et transferts)).

Dans ce cadre, trois modèles sont établis pour ce produit : c'est à dire un pour chaque niveau de garantie et de dépendance.

De plus, le modèle sépare les assurés selon leur sexe car ce dernier a un impact significatif sur la mortalité et la dépendance.

Enfin, une hypothèse du modèle dispose qu'il ne peut y avoir d'amélioration de l'état de l'assuré.

En synthèse, pour chaque sexe, trois modèles sont établis, c'est à dire un pour chaque niveau de garantie et de dépendance. Ces trois modèles disposent chacun de trois lois : mortalité des autonomes (GIR 5,6), incidence et mortalité des dépendants, cf. tableau ci-dessous.

GIR12		GIR123		GIR1234	
Homme	Femme	Homme	Femme	Homme	Femme
Taux d'incidence : par âge	Taux d'incidence : par âge	Taux d'incidence : par âge	Taux d'incidence : par âge	Taux d'incidence en dépendance légère : par âge	Taux d'incidence en dépendance légère : par âge
Mortalité des valides : par âge	Mortalité des valides : par âge	Mortalité des valides : par âge	Mortalité des valides : par âge	Mortalité des valides : par âge	Mortalité des valides : par âge
Mortalité des dépendants : par âge x ancienneté (en annuel)	Mortalité des dépendants : par âge x ancienneté (en annuel)	Mortalité des dépendants : par âge x ancienneté (en annuel)	Mortalité des dépendants : par âge x ancienneté (en annuel)	N/A	N/A

Tableau 16 - Titre : Présentation des modèles de dépendance du produit Protectys Autonomie

2.4.1.2. Présentation des lois de résiliation / réduction du produit Protectys Autonomie :

Les contrats relatifs au produit Protectys Autonomie peuvent faire l'objet de résiliations et de réductions :

- Si la résiliation intervient avant huit ans d'ancienneté, la garantie n'est plus applicable ;
- Si la résiliation intervient après huit ans d'ancienneté, la garantie peut être réduite.

En cas de réduction, le montant de la prestation est calculé en appliquant un coefficient de réduction des garanties. Celui est uniquement fonction de l'ancienneté du contrat.

Les chiffres correspondants sont les suivants.

Année	2014	2015	2016	2017	2018
Nombre de contrats	141 503	135 320	129 745	124 670	120 453
Nombre de résiliations	6 020	4 802	4 011	3 361	2 281
Nombre de réductions	271	380	689	1 313	2 257

Tableau 17 - Titre : Chiffres clés concernant les résiliations / réductions du produit Protectys Autonomie

La méthodologie retenue pour la modélisation des résiliations et réductions est la méthode de Hoem (Planchet et Thérond, 2006), qui permet de déterminer un estimateur paramétrique considérant les troncatures et les censures.

#	Etapes
1	Calcul de la durée d'exposition totale (notamment détermination de la durée d'exposition totale au risque de résiliation, du nombre de résiliations et réductions, de l'âge de l'assuré et de l'ancienneté du contrat en cas de résiliation)
2	Segmentation de la durée d'exposition par année d'observation, ancienneté, âge (détermination des taux d'entrée selon l'année d'observation, selon l'âge de l'assuré et selon l'ancienneté du contrat)
3	Calcul des taux bruts selon la méthode de Hoem

#	Etapes
4	Lissage, prolongation en début et en fin de loi
5	Comparaison du nombre de sinistres observés sur le nombre de sinistres attendus
6	Etablissement d'un coefficient de majoration de la loi de réduction en fonction de la revalorisation tarifaire

Tableau 18 - Titre : Principales étapes de modélisation des résiliations réductions en dépendance

Nous étudierons les réductions dans le chapitre 5 du présent mémoire et leur appliquerons des méthodes de machine learning de prédiction.

2.4.2. Méthodologie retenue pour application de l'approche Worst Case

Un mémoire, intitulé « Construction de lois d'expérience en assurance dépendance », réalisé par Marion Tsagalos (2020) au sein de LBP Prévoyance, a consisté à proposer des lois d'expérience comparatives, hors résiliations et réductions, du risque dépendance du produit Protectys Autonomie afin de les comparer aux lois de référence (cf. principales étapes ci-dessous).

2.4.2.1. Les principales étapes de modélisation de ces lois de comparaison sont les suivantes.

Etapes	Commentaires
Traitement des données	Détermination des dates de reconnaissance de sinistre Sélection des contrats Traitement des données manquantes Traitement des données aberrantes Ajout de variables Validation de la base
Choix de la fenêtre et des âges d'observation	
Calcul des taux bruts	Etude de différentes méthodes : Hoem, Kaplan Meier, Meira-Machado Traitement des risques concurrents
Lissage	Etude de différentes méthodes : paramétrique / non-paramétrique, moyennes mobiles, méthode de Whittaker-Henderson
Extrapolation	Etude de différentes méthodes : Modèles : logistique, Beard, Kannisto, Perk Méthodes de Brass, Kannisto, Gompertz-Makeham, de Denuit et Goderniaux, de Coale Choix de la tranche d'âge Traitement des risques concurrents
Contrôles de cohérence	Trois principales méthodes de contrôle de cohérence basées sur le calcul de ratios « constaté/attendu » ont été mises en place : - Les ratios « constaté/attendu » par âge et par année de survenance basés sur les données en portefeuille. Les « constatés » correspondent aux observations du portefeuille. - Les ratios « constaté/attendu » en simulant plusieurs années consécutives à partir de la population en portefeuille la première année considérée. - La comparaison de l'application d'une loi de mortalité générale (« constaté ») et de l'application des trois taux utilisés dans le modèle (« attendu ») : taux de mortalité des autonomes, taux de mortalité des dépendants et incidence.

Étapes	Commentaires
	Des contrôles complémentaires sont réalisés : - Par exemple, pour le lissage, sont mis en place des tests de validation, des critères de régularité et de fidélité ; - Intervalles de confiance.

Tableau 19 - Titre : Principales étapes de modélisation du mémoire Construction de lois d'expérience en assurance dépendance, réalisé par Marion Tsagalos (2020)

Ces étapes sont notamment appliquées pour construire des lois d'expérience de la mortalité des autonomes, d'incidence et de mortalité des dépendants.

Afin d'en estimer la qualité, un premier critère d'acceptation peut consister à mesurer les rapports constatés sur attendus.

Un second critère peut consister à étudier l'usage des lois. Dans ce cadre, l'objectif, ici, est de tester l'utilisation qui est faite des lois dépendance. De plus, dans le modèle de dépendance étudié, les futures décisions de gestion sont modélisées au sens de Solvabilité 2, c'est-à-dire des décisions du management. Ces futures décisions vont consister, par exemple, à modifier le provisionnement ou à appliquer des révisions tarifaires en fonction des potentielles dérives de sinistralité.

Dans ce cadre, les lois entrent dans les calculs Solvabilité 2 et un des critères pour mesurer le risque de modèle est le best estimate et ses composantes.

Le best estimate fait partie des provisions techniques prudentielles. La valeur des provisions techniques prudentielles, évaluée conformément à l'article L. 351-1 du Code des Assurances, correspond au montant actuel que les entreprises devraient payer si elles transféraient immédiatement leurs engagements à une autre entité agréée pour pratiquer des opérations d'assurance ou de réassurance⁵.

La valeur des provisions techniques prudentielles, mentionnées à l'article L. 351-2 du Code des Assurances, est égale à la somme de la meilleure estimation et de la marge de risque. La meilleure estimation, ou best estimate, correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs compte tenu de la valeur temporelle de l'argent estimée sur la base de la courbe des taux sans risque pertinente, soit la valeur actuelle attendue des flux de trésorerie futurs.

En assurance dépendance, les principales composantes du best estimate sont les primes, les frais et les sinistres.

Pour calculer le best estimate, le modèle d'interactions actif-passif réalise des projections à 50 ans, ce qui est cohérent avec la durée de vie d'un produit dépendance.

Le test sur l'utilisation des lois dépendance se fera donc au niveau des montants évalués de best estimate brut (avant réassurance, qui est de toute façon en quote-part) et visera à estimer si le best estimate de référence est suffisamment prudent par rapport aux évaluations de ce dernier qui découlent de l'utilisation des lois de comparaison. Le best estimate de référence sera considéré comme prudent si l'utilisation des nouvelles lois n'induit pas une nouvelle évaluation du best estimate ayant un écart positif de plus de 5% avec l'évaluation de référence.

Dans ce cadre, le présent mémoire :

⁵ Confère cours de GORRAND R. (2021). Bilan Prudentiel. Support de cours Certificat d'Expertise Actuarielle. Sorbonne Université, Paris.

- Repart des travaux réalisés dans le cadre du mémoire suscité afin de les challenger et d'estimer les écarts, selon la méthode Worst Case, en termes de best estimate par rapport aux résultats obtenus avec les lois de référence ;
 - o A cette fin, des travaux sur les lois d'expérience sont nécessaires, afin de s'assurer de leur comparabilité aux lois de référence (comparaison des extrapolations, du lissage et mise en forme des lois) ;
- Propose des analyses de sensibilités sur les lois de référence, afin d'en évaluer l'impact sur le best estimate.

Les résultats obtenus permettent d'évaluer, de manière quantitative, le risque de modèle au niveau de l'estimation du best estimate du produit Protectys Autonomie et de conclure, en fonction de l'appétence au risque de LBP Prévoyance, sur son acceptabilité ou la nécessité de mettre en place des plans d'action correctifs.

Les résultats correspondants sont présentés dans le chapitre 4 du présent mémoire.

3. Chapitre 3 - Mise en place de la méthode de scoring pour un panel de modèles

Pour rappel, ce mémoire est structuré en trois approches, allant de la plus large à la plus précise : (i) méthode large d'évaluation du risque de modèle appliquée sur plusieurs modèles (chapitres 2 – théorie - et 3 - application), (ii) méthode d'évaluation précise du risque de modèle appliquée à un modèle déterminé (chapitres 2 – théorie - et 4 - application), (iii) étude d'un plan d'action visant à réduire le risque de modèle par application d'analyses univariées et méthodes de machine learning de prédiction (chapitre 5).

La méthodologie de scoring et ses différentes étapes sont décrites dans le chapitre 2 du présent mémoire.

L'application de la méthode de scoring sera présentée ici sur six exemples, afin de présenter différents cas possibles et conserver de la lisibilité.

Les six exemples retenus sont les suivants.

#	Nom de l'exercice de modélisation	Objectifs de la modélisation
1	LBPP - Projection Interactions Actif-Passif, qui comprend les lois dépendance Protectys Autonomie	Projection des comportements des assurés et de leurs impacts sur les différents items des comptes de résultats sur les produits de prévoyance
2	LBPP - Provisions de sinistres	Calcul des provisions de sinistres des produits de prévoyance
3	LBPP – Tarification GAV	Calcul du tarif du produit GAV
4	LBPP – Provision risque croissant des produits en dépendance	Calcul de la Provision pour Risque Croissant des produits de dépendance
5	Lois d'expérience Protectys de comparaison, non utilisées à date, par LBP Prévoyance (issues du mémoire Construction de lois d'expérience en assurance dépendance)	Comparer les lois utilisées de référence avec les lois d'expérience
6	LBPP – Provision globale de gestion	Calcul de la provision S1 globale de gestion

Tableau 20 - Titre : Modélisations retenues pour la présentation de l'approche de scoring du risque de modèle

En amont de cette application, il est nécessaire de définir l'appétence au risque :

Elle est considérée ici comme la volonté de n'avoir que des modèles dont le niveau de risque est inférieur à majeur (soit mineur ou modéré au maximum). Cette appétence est définie en cohérence avec l'appétence de La Banque Postale Prévoyance.

3.1. Application de l'Outil d'aide à la décision, permettant de cartographier les modèles de LBP Prévoyance

Dans un premier temps, l'Outil d'aide à la décision, présenté dans le chapitre 1, est appliqué sur les six exemples retenus, cf. tableau ci-dessous.

#	Nom de l'exercice de modélisation	L'approche quantitative fonde-t-elle la prise de décision directement ou indirectement ?	Si la réponse précédente est « Oui » : L'approche quantitative repose-t-elle sur des hypothèses de construction ou un jugement d'expert ?	Si la réponse précédente est « Oui » : La méthode retenue sert-t-elle uniquement à la réalisation de contre tests ou comparatif de performance ?	Si la réponse précédente est « Non » : La méthode retenue sert-t-elle in fine à la mitigation d'un risque de modèle ou pour des besoins réglementaires ?	Conclusion de l'outil d'aide à la décision
1	LBPP - Projection Interactions Actif-Passif, qui comprend les lois dépendance Protectys Autonomie	<i>Oui</i>	<i>Oui</i>	<i>Non</i>	<i>Non applicable</i>	La méthode est un modèle
2	LBPP - Provisions de sinistres	<i>Oui</i>	<i>Oui</i>	<i>Non</i>	<i>Non applicable</i>	La méthode est un modèle
3	LBPP – Tarification GAV	<i>Oui</i>	<i>Oui</i>	<i>Non</i>	<i>Non applicable</i>	La méthode est un modèle
4	LBPP – Provision risque croissant en dépendance	<i>Oui</i>	<i>Oui</i>	<i>Non</i>	<i>Non applicable</i>	La méthode est un modèle
5	Lois d'expérience Protectys non utilisées à date, par LBP Prévoyance (issues du mémoire Construction de lois d'expérience en assurance dépendance)	<i>Non</i>	<i>NA</i>	<i>NA</i>	<i>NA (à date : un des objectifs de ce mémoire est d'ailleurs de proposer une utilisation de ces lois pour évaluer le risque de modèle par l'approche worst case)</i>	La méthode n'est pas un « modèle » au sens de l'arbre de décision
6	LBPP – Provision globale de gestion	<i>Oui</i>	<i>Oui</i>	<i>Non</i>	<i>Non applicable</i>	La méthode est un modèle

Tableau 21 - Titre : Exemples d'application de l'Outil d'aide à la décision

Cinq des six exercices de modélisation présentés ci-dessus peuvent, selon cette méthode, être considérés comme des modèles. Leur risque est donc évalué ci-dessous.

3.2. Evaluation du risque brut

Comme présenté dans le chapitre 1, le risque brut est évalué en croisant la fréquence d'utilisation du modèle et son poids financier. On réalise cette analyse en deux dimensions car c'est une méthode d'évaluation et de gestion des risques fréquente qui est utilisée au sein de LBP.

Pour des problématiques de confidentialité et de représentativité des cas possibles, les résultats, ici, ont été modifiés, cf. tableau ci-dessous.

Informations à compléter									Calcul automatique		
#	Modèles	Type de variable modélisée	Fréquence d'utilisation du modèle	Indicateur quantitatif	Valeur (M€)	Total bilan 2020 en S2 (SFCCR)	Total primes émises brutes 2020 (SFCCR)	Poids financier modélisé	Fréquence	Impact	Risque brut
1	Projection Interactions Actif-Passif	Bilan	Annuelle	Montant des provisions BE bruts	1255	2784,2	480,9	45%	3	3	3
2	Provisions de sinistres	Bilan	Annuelle	Montant de la provision de sinistre comptable	402	2784,2	480,9	14%	3	2	3
3	Tarification GAV	Compte de résultat	Moins d'une fois par an	Montant des primes annuel	53,6	2784,2	480,9	11%	2	2	2
4	LBPP – Provision risque croissant en dépendance	Bilan	Annuelle	Montant de la provision	520	2784,2	480,9	19%	2	3	3
6	LBPP – Provision globale de gestion	Bilan	Annuelle	Montant de la provision	25	2784,2	480,9	1%	3	1	2

Tableau 22 - Titre : Exemples d'application de l'Outil d'aide à la décision

Ce qui donne les matrices suivantes.

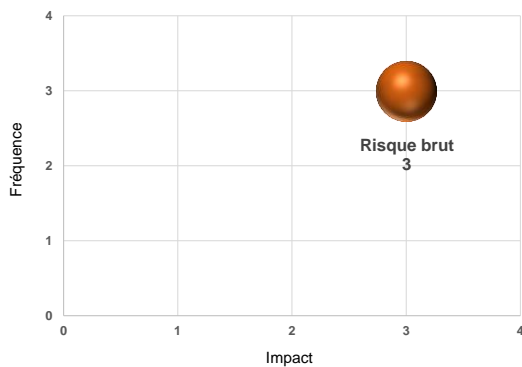


Figure 15 - Titre : Evaluation du risque brut du modèle 1

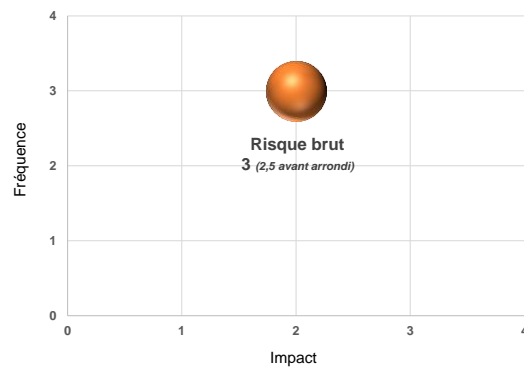


Figure 16 - Titre : Evaluation du risque brut du modèle 2

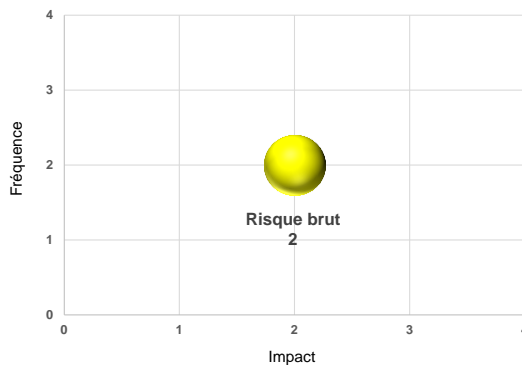


Figure 17 - Titre : Evaluation du risque brut du modèle 3

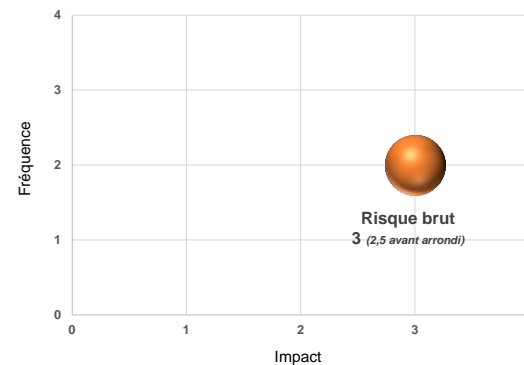
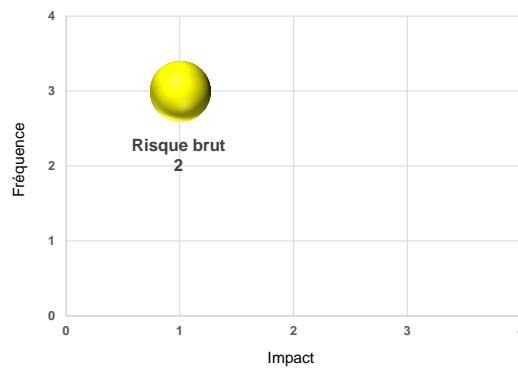


Figure 18 - Titre : Evaluation du risque brut du modèle 4



Légende :

Risque brut 1 = mineur ; risque brut 2 = modéré ; risque brut 3 = majeur ; risque brut 4 = critique.

Figure 29 - Titre : Evaluation du risque brut du modèle 5

Le risque brut estimé est acceptable pour le troisième et pour le cinquième modèle, selon l'appétence au risque définie ci-dessus.

Cependant, ce n'est pas le cas pour les autres modèles (modèles un, deux et quatre). Aussi, l'analyse du risque sera approfondie pour ces derniers.

3.3. Evaluation des dispositifs d'atténuation du risque brut des modèles

Les résultats présentés ci-dessous ont été modifiés pour des raisons de confidentialité et afin d'illustrer au mieux l'exercice et les différents cas possibles.

3.3.1. Evaluation des dispositifs d'atténuation du risque brut du Modèle 1 : Projection Interactions Actif-Passif

Thématiques	Questions	Evaluation des modèles		
		1	2	4
Encadrement	Le modèle est-il encadré par une politique de gestion des modèles ?	Oui	Oui	Oui
	Les principales étapes de la vie du modèle font-elles l'objet de procédures à jour ?	Partiel	Partiel	Partiel
	Une cotation du modèle a-t-elle été fournie ?	Oui	Partiel	Oui
	Le modèle fait-il l'objet d'un inventaire annuel, porté par l'équipe de validation, recensant l'ensemble des modèles et leurs informations ?	Oui	Partiel	Oui
Définition des rôles et responsabilités opérationnelles	Le responsable du modèle est-il désigné ?	Oui	Non	Oui
	Le propriétaire du modèle est-il désigné ?	Oui	Non	Oui
	Le développeur du modèle est-il désigné ?	Oui	Oui	Oui
	L'implémenteur du modèle est-il désigné ?	Oui	Oui	Oui
	L'utilisateur du modèle est-il désigné ?	Oui	Oui	Oui
	Le valideur du modèle est-il désigné et en charge de l'inventaire des modèles ?	Oui	Oui	Oui
	Le gestionnaire des risques du modèle est-il désigné ?	Oui	Oui	Oui
	Des instances de gouvernance sont-elles désignées pour valider la gouvernance des modèles et notamment le dispositif de gestion des risques ?	Oui	Oui	Oui
Pilotage	Une instance / équipe de validation a-t-elle formellement validé la mise en place du modèle ?	Oui	Oui	Oui
	Le cas échéant, cette instance revoit-elle les hypothèses liées au modèle ?	Oui	Non	Oui
	Les réserves liées aux modèles font-elles l'objet d'un suivi ?	Oui	Partiel	Oui
	Une revue périodique du modèle est-elle réalisée et portée par l'équipe de validation ou une instance, comprenant le suivi des contrôles continus sur la performance du modèle et des <i>back tests</i> ?	Oui	Partiel	Oui
	Le modèle est-il revu annuellement en comité ou par l'équipe de validation ?	Oui	Partiel	Oui
	Une instance ou équipe de validation valide-t-elle formellement les évolutions du modèle ?	Oui	Non	Oui
Système	Les systèmes d'informations et données	Non	Non	Non

Thématiques	Questions	Evaluation des modèles		
		1	2	4
d'information et données	utilisés pour les modèles sont-ils évalués ?			
	Les données, hypothèses et modélisations sont-elles tracées et conservées ?	Oui	Oui	Oui
Contrôle permanent	Un dispositif de contrôle permanent de 1er niveau est-il en place ?	Non	Non	Non
	Le cas échéant, les résultats des contrôles de 1er niveau sont-ils formalisés, suivis dans une instance et font-ils l'objet de plans d'actions mis en place ?	N/A	N/A	N/A
	Un dispositif de contrôle permanent de 2ème niveau est-il en place ?	Non	Non	Non
	Le cas échéant, les résultats de contrôle de 2ème niveau sont-ils formalisés, suivis dans une instance et font-ils l'objet de plans d'actions mis en place ?	N/A	N/A	N/A
	L'audit interne a-t-il vérifié que le contrôle de premier niveau et le contrôle de deuxième niveau respectaient les différents principes et procédures de gestion du risque de modèle ?	Oui	Non	Non
	D'autres fonctions ont-elles un rôle de contrôles (fonctions actuarielle / risque par exemple) ?	Oui	Oui	Oui
	Le cas échéant, les résultats correspondants ont-ils fait l'objet de plans d'actions suivis ?	Oui	Oui	Oui
Taux de conformité		86%	48%	79%

Tableau 23 - Titre : Evaluation des dispositifs d'atténuation pour les modèles 1 à 5

Une fois les dispositifs d'atténuation évalués, on analyse la pertinence et la complexité des modèles. Si un modèle est très complexe ou utilise des données de faible qualité par exemple, alors les dispositifs d'atténuation sont considérés comme moins efficaces et une pénalité leur sera alors attribué.

3.4. Evaluation de la pertinence des données et méthodes et de la complexité des modèles étudiés

En amont de cette évaluation, le niveau de pénalité à attribuer à l'évaluation des dispositifs d'atténuation, est à déterminer. On considérera ici, un niveau de 2,5%, cf. tableau ci-dessous.

Valeur de la pénalité
2,5%

$Pénalité = 2,5\% * (évaluation\ de\ la\ pertinence\ et\ de\ la\ complexité - 1)$:
- Un niveau de pertinence et complexité 1 donne une pénalité de 0% ;

- Un niveau de pertinence et complexité 2 donne une pénalité valant 2,5% ;
- Un niveau de pertinence et complexité 3 donne une pénalité de 5,0%.

Tableau 24 - Titre : Valeur de la pénalité à appliquer à l'évaluation des dispositifs d'atténuation

L'analyse de la pertinence des données et méthodes utilisées est réalisée, conformément à la méthodologie présentée dans le chapitre 1, cf. tableau ci-dessous.

#	Modèles	Qualité des données utilisées	Qualité des méthodes utilisées	Evaluation de la pertinence des données et méthodes, recalculée sur base 4
1	Projection Interactions Actif-Passif	1	2	2,0
2	Provisions de sinistres	2	2	2,7
4	Provisions risque croissant en dépendance	2	1	2,7

Tableau 25 - Titre : Evaluation de la pertinence des données et méthodes utilisées pour les modèles étudiés

Pour rappel, ces items sont évalués grâce à la table de correspondance suivante.

Items	Définition
Qualité des données utilisées	Si la qualité des données a été évaluée (exhaustivité, véracité, valeurs aberrantes) et ne fait pas l'objet de réserves, indiquer 1 Si la qualité des données a été évaluée et fait l'objet de réserves non bloquantes, indiquer 2 Si la qualité des données n'a pas été évaluée ou fait l'objet de réserves bloquantes, indiquer 3
Qualité des méthodes utilisées	Si les méthodes utilisées ont été évaluées et ne font pas l'objet de réserves (exemple : méthodes classiques issues de la littérature actuarielle et ayant fait l'objet d'études sur leur caractère adapté et leur pertinence), indiquer 1 Si les méthodes utilisées ont été évaluées et font l'objet de réserves non bloquantes (exemple : méthodes non classiques, telles que l'utilisation de nombreux proxys), indiquer 2 Si les méthodes utilisées n'ont pas été évaluées ou font l'objet de réserves bloquantes, indiquer 3

Tableau 26 - Titre : Evaluation de la pertinence des données et méthodes utilisées pour les modèles étudiés

La complexité est ensuite évaluée, cf. tableau ci-dessous.

#	Modèles	Durée d'implémentation du modèle	Temps de calcul	Imbrication / interactions avec d'autres modèles	Granularité du modèle (ligne à ligne vs modèle point)	Projections et chocs/règles de gestion	Evaluation de la complexité du modèle, recalculée sur base 4
1	LBPP - Projection Interactions Actif-Passif	3	3	3	2	2	4,0
2	LBPP - Provisions de sinistres	1	1	1	1	1	1,6
4	Provisions risque croissant en dépendance	2	1	1	2	1	2,3

Tableau 27 - Titre : Evaluation de la complexité des modèles étudiés

Pour rappel, la complexité est évaluée grâce à la table de correspondance suivante.

Items	Définition
Durée d'implémentation du modèle	Si la durée d'implémentation du modèle est inférieure à 2 mois, indiquer 1 Si la durée d'implémentation du modèle est comprise entre 2 mois et 6 mois, indiquer 2 Si la durée d'implémentation du modèle est supérieure à 6 mois, indiquer 3
Temps de calcul	Si le temps de calcul est inférieur à 10 minutes, indiquer 1 Si le temps de calcul est compris entre 10 minutes et 1 heure, indiquer 2 Si le temps de calcul est supérieur à 1 heure, indiquer 3
Imbrication / interactions avec d'autres modèles	Si le modèle n'est pas en lien avec d'autres modèles (modèle indépendant), indiquer 1 Si le modèle dispose de plusieurs modules (exemple : modèle ALM santé), indiquer 2 Si le modèle fait l'objet d'imbrications et d'interactions (ex : modèle ALM prévoyance), indiquer 3
Granularité du modèle (ligne à ligne vs modèle point)	Si le modèle nécessite une granularité Model point, indiquer 1 Si le modèle nécessite une granularité très fine (exemple : ligne à ligne), indiquer 2
Projections et chocs/règles de gestion	Si le modèle ne nécessite pas de projection (calcul en t=0), indiquer 1 Si le modèle nécessite des projections dans le temps, avec un ou des choc(s) simple(s) (c'est à dire ne nécessitant que la modification d'une ou plusieurs hypothèse(s) et pas de modifications supplémentaires des formules de calculs (exemple : choc de mortalité), indiquer 2 Si le modèle nécessite des projections dans le temps et l'implémentation de chocs complexes (c'est à dire avec des modifications supplémentaires des formules / règles de gestion en plus de l'hypothèse), indiquer 3

Tableau 28 - Titre : Table de correspondance pour évaluer la complexité des modèles

Enfin les deux évaluations sont croisées, « pertinence * complexité », cf. tableau ci-dessous.

#	Modèles	Evaluation de la pertinence et de la complexité du modèle
1	LBPP - Projection Interactions Actif-Passif	3,0
2	LBPP - Provisions de sinistres	2,1
4	Provisions risque croissant en dépendance	2,5

Tableau 29 - Titre : Evaluation de la pertinence X complexité pour les modèles étudiés

Enfin, la pénalité correspondante est appliquée, par différence, à l'évaluation des dispositifs d'atténuation, cf. tableau ci-dessous.

#	Modèles	Evaluation de la pertinence et de la complexité du modèle	Pénalité correspondante (cf. légende pour explication)	Evaluation finale des dispositifs d'atténuation
1	LBPP - Projection Interactions Actif-Passif	3,0	5,0%	81,0%
2	LBPP - Provisions de sinistres	2,1	2,8%	45,2%
4	Provisions risques croissant en dépendance	2,5	3,8%	76,2%

- $Pénalité = 2,5\% * (évaluation de la pertinence et de la complexité - 1)$:
- Un niveau de pertinence et complexité 1 donne une pénalité de 0% ;
 - Un niveau de pertinence et complexité 2 donne une pénalité valant 2,5% ;
 - Un niveau de pertinence et complexité 3 donne une pénalité de 5,0%.

Tableau 30 - Titre : Evaluation des dispositifs d'atténuation, après application de la pénalité pertinence X complexité

3.5. Evaluation du risque net des modèles étudiés

Une fois le risque brut et la conformité des dispositifs d'atténuation, après pénalité, évalués, on obtient le risque net grâce à la matrice suivante, qui croise les deux évaluations, cf. figures ci-dessous.

3.5.1. Evaluation du risque net du Modèle 1 : Projection Interactions Actif-Passif

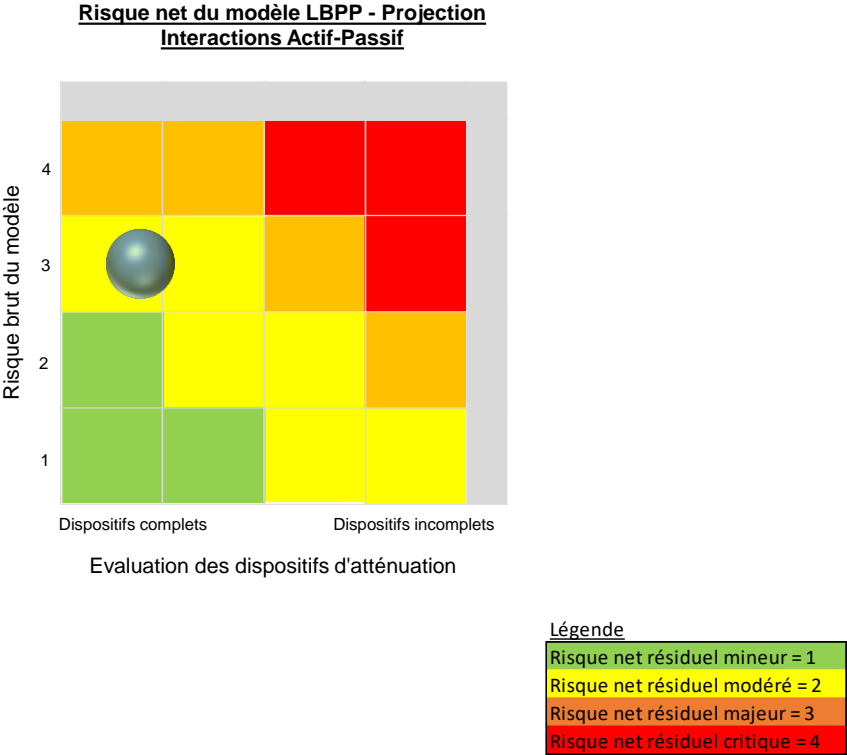


Figure 3 - Titre : Evaluation du risque net du modèle 1

3.5.2. Evaluation du risque net du Modèle 2 : Provisions de sinistres

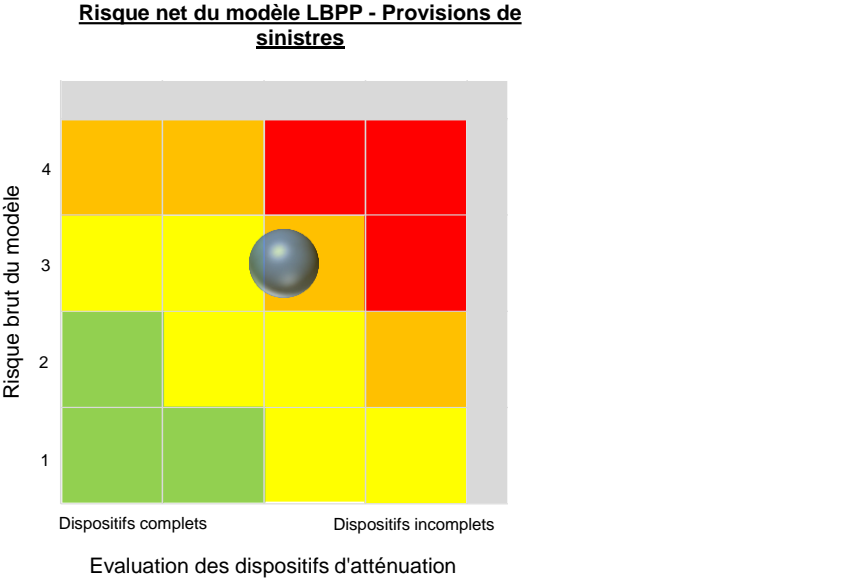


Figure 21 - Titre : Evaluation du risque net du modèle 2

3.5.3. Evaluation du risque net du Modèle 4 : Provisions risques et charges dépendance

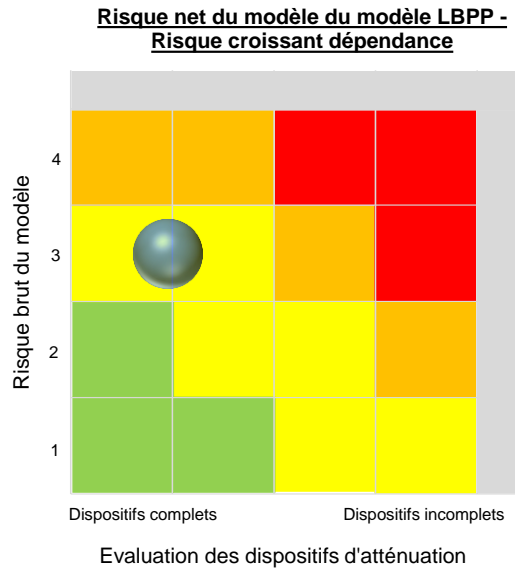


Figure 22 - Titre : Evaluation du risque net du modèle 4

3.5.4. Conclusions sur l'acceptabilité des risques évalués

Le modèle 1 et le modèle 4 induisent un risque de modèle net modéré. En cohérence par rapport à l'appétence au risque définie, ce niveau est considéré comme acceptable.

Le modèle 2 induit un risque de modèle net, selon cette méthode, majeur. En cohérence par rapport à l'appétence au risque définie, ce niveau n'est pas acceptable. Il est donc nécessaire de mettre en place des plans d'actions afin de le réduire. Des propositions sont présentées ci-dessous.

Renforcer les dispositifs d'atténuation
Documenter l'ensemble des étapes de la vie du modèle
Réaliser un travail de cotation documenté et argumenté
Recenser toutes les informations relatives au modèle et les mettre régulièrement à jour
Désigner un responsable et un propriétaire du modèle
Valider les hypothèses du modèle dans un comité ad hoc
Réaliser une revue périodique du modèle
Valider les évolutions du modèle de manière formelle
Réaliser des tests sur les systèmes d'informations liés au modèle (régression, etc.)
Mettre en place un dispositif de contrôle permanent en lien avec les étapes de la vie du modèle

Tableau 31 - Titre : Propositions de plans d'actions possible pour diminuer le risque net du modèle 2 et respecter l'appétence au risque

Des réflexions sur la faisabilité et sur l'investissement relatif à la mise en œuvre de ces plans d'actions doivent être menées afin de mettre en place les actions les plus efficaces.

A titre d'illustration, le fait de désigner un responsable et un propriétaire du modèle et de revoir formellement les hypothèses de ce dernier permet d'améliorer convenablement les dispositifs d'atténuation jusqu'à un risque net modéré, donc acceptable selon l'appétence au risque considérée, cf. figure ci-dessous.

Risque net du modèle - Provisions de sinistres, après mise en place des plans d'actions mentionnés ci-dessus

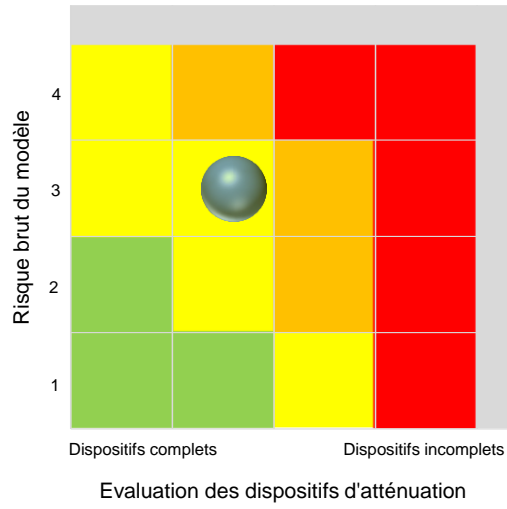


Figure 23 - Titre : Risque net du modèle - Provisions de sinistres, après mise en place des plans d'actions

4. Chapitre 4 – Mise en place de l’approche Worst Case au niveau de l’estimation du best estimate du produit Protectys Autonomie

Pour rappel, ce mémoire est structuré en trois approches, allant de la plus large à la plus précise : (i) méthode large d’évaluation du risque de modèle appliquée sur plusieurs modèles (chapitres 2 – théorie - et 3 – application), (ii) méthode d’évaluation précise du risque de modèle appliquée à un modèle déterminé (chapitres 2 – théorie - et 4 - application), (iii) étude d’un plan d’action visant à réduire le risque de modèle par application d’analyses univariées et méthodes de machine learning de prédiction (chapitre 5).

4.1. Etude des lois d’expérience construites dans le cadre du mémoire « Construction de lois d’expérience en assurance dépendance »

Un des objectifs du présent mémoire est de contrôler et d’utiliser les lois d’expérience construites dans le cadre du mémoire « Construction de lois d’expérience en assurance dépendance » (M. Tsagalos, 2020), en les prenant comme données source afin d’alimenter un modèle « Prophet », préexistant, de calcul de best estimate. Ceci, afin de comparer les résultats obtenus avec ceux obtenus à partir des lois de référence utilisées par LBPP. Les données étudiées ici ont été arrêtées en septembre 2019.

Pour rappel, le best estimate « correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs compte tenu de la valeur temporelle de l’argent estimée sur la base de la courbe des taux sans risque pertinente, soit la valeur actuelle attendue des flux de trésorerie futurs » (article R351-2 du Code des Assurances, transposition en droit français de l’article 77 de la Directive Solvabilité 2).

Dans ce cadre, afin d’estimer le best estimate brut, le modèle « Prophet » projette différents scénarios, notamment à partir des lois de dépendance estimées, des entrées en termes de primes et des sorties en termes de paiements de sinistres et de frais de gestion, puis pondère ces flux par leur probabilité avant de les actualiser au taux sans risque. L’objectif ici est de s’assurer des méthodes utilisées pour estimer les lois de dépendance puis de les utiliser dans le modèle « Prophet », afin de comparer les résultats obtenus en termes de best estimate avec les résultats préexistants.

Les travaux réalisés dans le présent mémoire et sur la deuxième approche sont schématisés ci-dessous.

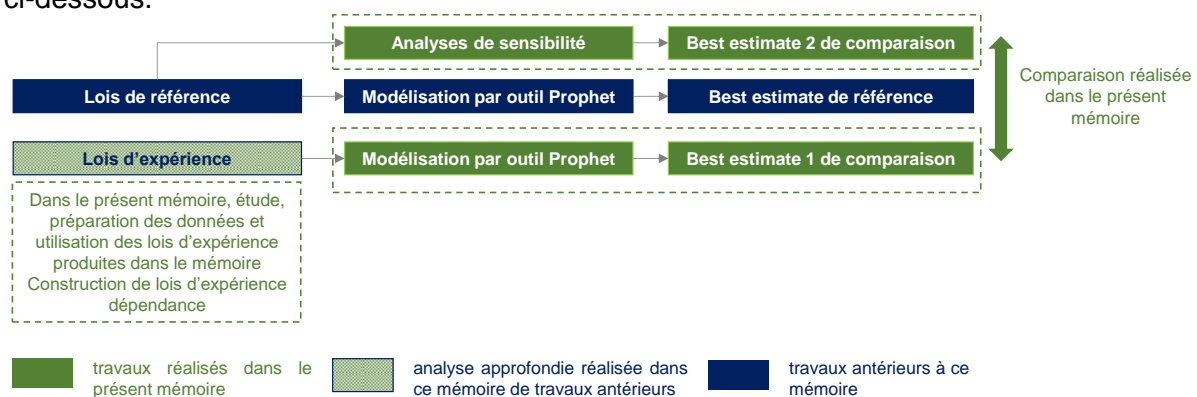


Figure 24 - Titre : Périmètre d’étude du présent mémoire

Les lois de référence utilisées par LBP Prévoyance ont été produites par un réassureur. LBPP n'a pas accès au détail de la modélisation. L'objectif de ce mémoire d'évaluer le risque de modèle correspondant, en comparant le best estimate obtenu avec les lois de référence et celles d'expérience de comparaison, a donc de l'importance. Ce mémoire s'intéresse en particulier à étudier la façon dont ont été construites les lois d'expérience, afin de s'assurer qu'elles sont exploitables. En effet, elles n'ont pas fait l'objet d'une validation formelle et officielle comme les autres modèles. Ensuite, l'objectif est de comparer les résultats obtenus, en termes de best estimate, en utilisant ces lois d'expérience par rapport aux lois de référence. La modélisation par l'outil Prophet faisant l'objet d'études et de validations officielles de manière annuelle, ce mémoire ne l'étudie pas.

Les principales étapes de construction de lois d'expérience de comparaison sont **(i) le traitement des données, (ii) le choix de la fenêtre et des âges d'observation, (iii) le calcul des taux bruts, (iv) le lissage et (v) l'extrapolation**. Des **contrôles** doivent être réalisés afin de s'assurer de la cohérence des données d'entrée, de sortie et des méthodologies.

Les bases de données utilisées pour les lois d'expérience portent sur l'ensemble des contrats Protectys Autonomie, avec trois types d'informations : assuré (identifiant, sexe, date de naissance, date de décès le cas échéant), contrat (date de souscription, d'effet du contrat, de clôture du contrat le cas échéant, de mise en réduction le cas échéant) et garantie (début, fin, date du sinistre, montant et type de garantie, montant effectivement versé).

Ces données ont fait l'objet de retraitements en cas de données manquantes ou aberrantes. Ces choix ont fait l'objet d'une validation par les métiers. La base obtenue a été validée dans sa globalité en la comparant avec les données comptables.

En l'espèce, les données comptables portent sur le montant des sinistres effectivement réglés chaque année. La base a ainsi été validée en confrontant les montants réglés dans la base avec les montants comptables.

Le choix de la fenêtre d'observation semble cohérent. En effet, les mois récents font l'objet d'une exclusion car la déclaration d'un sinistre de dépendance ou d'un décès peut être tardive. Prendre en compte les mois récents pourrait ainsi sous-estimer la sinistralité. Ensuite, la plage d'observation la plus longue est choisie afin d'avoir un maximum d'observations et donc une base pertinente. Le choix des âges d'observation a été fait par rapport à la significativité du risque observé.

Ensuite, le calcul des taux bruts a d'abord porté sur les lois d'incidence et de mortalité des autonomes que nous détaillerons ici pour les lois de mortalité des femmes autonomes non-GIR 12. Une approche similaire a été réalisée pour les lois d'incidence ainsi que sur les autres niveaux d'autonomies, pour les hommes et pour les dépendants. Il existe quand même une différence au niveau des dépendants : la construction des lois de mortalité doit prendre en compte l'ancienneté dans l'état en dépendance.

4.1.1. Etude des lois d'expérience : étapes principales de construction des lois de mortalité et d'incidence, focus sur les lois de mortalité des femmes autonomes non-GIR 12

Avant d'utiliser les lois d'expérience de comparaison, nous étudions les modalités selon lesquelles elles ont été construites. Ci-dessous, l'étude est détaillée pour les lois de mortalité des femmes autonomes.

La première étape est d'étudier les censures et les troncatures. En effet, la base de données comporte des censures et troncatures dues au fait que les assurés ne font pas l'objet d'une observation dès leur naissance et jusqu'à la concrétisation du risque. Ils entrent ainsi dans le portefeuille d'observation à un âge donné et peuvent en sortir pour différentes raisons autres que la survenance du risque, par exemple leur décès ou la résiliation de leur contrat. Les observations portefeuille ne rendent donc pas compte de l'exposition complète au risque, d'où une étude nécessaire des censures et troncatures.

Si l'entrée en dépendance ou la mortalité advient avant que débute l'observation, les données correspondantes ne figurent pas dans la base. On dit qu'elles sont tronquées à gauche. A droite, les observations ne sont pas complètes dans le sens où l'événement n'a peut-être pas encore eu lieu. On dit qu'elles sont censurées à droite. L'étude des censures et troncatures a été réalisée de manière pertinente. Leur prise en compte dans la modélisation est vérifiée par la suite.

La deuxième étape consiste à étudier la présence de risques concurrents. Il y a concurrence des risques s'il y a différentes sorties d'un état donné, par exemple un autonome peut soit décéder soit devenir dépendant, ce qui est notre cas ici. Or, certains estimateurs de mortalité et d'incidence étudiés par la suite ne prennent pas en considération cette concurrence des risques. Il est donc nécessaire de les adapter. Pour cela, différentes méthodes sont envisagées, dont l'attribution d'un ordre de priorité qui est retenue.

Ordre de priorité :

$$\hat{i}(x) = i(x) ;$$

$$\hat{q}(x) = (1 - i(x)) * q(x) ;$$

avec $\hat{i}(x)$ incidence transformée, fonction de $i(x)$, incidence ; x variable représentant le temps ; $\hat{q}(x)$ mortalité transformée, fonction de $i(x)$ et de $q(x)$ probabilité de décès en fonction du temps.

Ici, la priorité est donnée à l'incidence, ce qui est prudent en termes de provisionnement.

Les méthodes sont comparées grâce à des ratios « constaté sur attendu ».

La méthode par ordre de priorité, avec la priorité donnée à l'incidence, est choisie du fait (i) de meilleurs ratios « constaté sur attendu », (ii) car elle est plus prudente en termes d'estimation de la sinistralité et (iii) car c'est la méthode choisie par LBPP.

Pour le choix de l'estimateur des taux brut, différents estimateurs, prenant en compte les censures et troncatures, sont étudiés : l'estimateur des moments de Hoem, celui de Kaplan-Meier et celui de Meira-Machado. A chaque fois, des contrôles par intervalles de confiance sont réalisés.

Ces méthodologies font l'objet de plusieurs papiers, par exemple celui de Philippe Saint-Pierre (2021) ou de Quentin Guibert et Frédéric Planchet (2017).

La méthodologie de l'estimateur de Hoem consiste à calculer à tous les âges considérés, le niveau d'exposition ainsi que les occurrences de décès ou d'entrées en dépendance. Ainsi, la probabilité d'occurrence sur une durée i est :

$$p_i = \frac{d_i}{e_i} ;$$

avec p_i probabilité d'occurrence sur la durée i , d_i nombre d'occurrences sur la durée i et e_i exposition au risque sur i .

L'intervalle de confiance est ensuite construit en supposant que le risque suit une loi binomiale.

L'estimateur de Kaplan-Meier est calculé sur la base des fonctions de survie aux dates $T(i)$ d'occurrence des sinistres ou des censures.

On a la fonction de survie :

$$S(t) = P(X > t) \text{ qui est la probabilité de survie ou de rester autonome jusqu'à } t.$$

Ce qui donne la fonction de répartition, probabilité de décès ou d'occurrence de l'événement, suivante :

$$F(t) = P(X \leq t) = 1 - S(t).$$

La densité de probabilité est f tel que :

$$F(t) = \int_0^t f(x) dx ;$$

Donc :

$$f(t) = F'(t) = -S'(t).$$

Le risque instantané de décès $h(t)$ ou d'incidence est la probabilité d'occurrence du décès ou de l'événement juste après un instant t , sachant que l'individu a survécu jusque-là. Il s'exprime tel que suit :

$$h(t) = \lim_{h \rightarrow 0} \frac{P(t \leq X < t + h | X > t)}{h} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -(\ln S(t))'.$$

Le taux de hasard cumulé $H(t)$ est égal à :

$$H(t) = \int_0^t h(x) dx = -\ln(S(t)) ;$$

On a donc :

$$S(t) = \exp(-H(t)) \text{ et comme } H(t) = \int_0^t h(x) dx \text{ et } f(t) = h(t) * S(t) \text{ alors :}$$

$$f(t) = h(t) * \exp\left(-\int_0^t h(x) dx\right).$$

En pratique, on a l'estimateur H_{NA} de Nelson Aalen du taux de hasard cumulé tel que :

$$H_{NA}(t) = \sum_{T_i < t} \frac{d_i}{Y_i} ;$$

avec d_i nombre d'occurrences de i (mortalité ou incidence), Y_i nombre d'individus en vie / autonome avant l'occurrence de i .

L'estimateur de Kaplan-Meier part du postulat qu'être en vie un instant après t correspond à survivre jusqu'à t et également en t , donc avec $t_1 < t_2 < t$:

$$P(X > t) = P(X > t | X > t_2) * P(X > t_2) = P(X > t | X > t_2) * P(X > t_2 | X > t_3) * P(X > t_3);$$

Ce qui donne, avec des temps d'occurrences distincts et en ordre croissant :

$$P(X > T_{(i)}) = \prod_{k=1}^i P(X > T_{(k)} | X > T_{(k-1)});$$

On a in fine l'estimateur de Kaplan-Meier S_k de la fonction de survie :

$$S_k(t) = \prod_{k=1}^{i < t} \left(1 - \frac{d_i}{Y_i}\right);$$

Enfin, on retrouve la probabilité p_i de survenance de sinistre sur une période $[t_i, t_{i+1}]$ par différence : $p_i = 1 - \frac{S(t_{i+1})}{S(t_i)}$.

La durée moyenne de survie $E(X)$ et sa variance $V(X)$ sont :

$$E(X) = \int_0^{\infty} S(t) dt;$$

$$V(X) = 2 * \int_0^{\infty} t S(t) dt - (E(X))^2.$$

L'intervalle de confiance $IC(\alpha)$ est construit en se basant sur l'estimateur de Greenwood V_G de la variance de $S_k(t)$:

$$V_G(S_k(t)) = S_k(t)^2 \sum_{T_i < t} \frac{d_i}{Y_i * (Y_i - d_i)};$$

$$IC(\alpha) = [S_k(t) \pm u_{\frac{\alpha}{2}} * \sqrt{V_G(S_k(t))}] ,$$

avec $u_{\frac{\alpha}{2}}$ quantile d'ordre $\frac{\alpha}{2}$ de la loi normale centrée réduite.

In fine, pour le taux de mortalité, l'estimateur proposé par Kaplan-Meier est choisi pour les raisons suivantes : meilleurs ratios constaté sur attendu, taux de mortalité inférieurs ce qui est prudent au regard du risque et prise en compte des troncatures et censures.

Pour le taux d'incidence, l'estimateur de Kaplan-Meier est également choisi car les ratios « constaté sur attendu » sont meilleurs, que les taux obtenus sont plus stables sur les grands âges et qu'il est plus prudent car induit une incidence plus élevée.

Deux méthodes de lissage sont étudiées : moyennes mobiles et méthode de Whittaker-Henderson. Ces méthodes font l'objet de tests de validation. Ces méthodes sont présentées dans le cours de Frédéric Planchet (2022-2023) et dans les Lignes Directrices Mortalité publiées en 2006 par l'Institut des Actuaire.

Avec les moyennes mobiles, la valeur lissée M_x du taux est calculée en considérant la moyenne mobile de $2n + 1$ estimations consécutives non lissées :

$$M_x = \frac{1}{2n + 1} * \sum_{i=-n}^n q_{x+i} ; \text{avec } n \in \mathbb{N}.$$

Concernant les moyennes mobiles, le choix de l'ordre $2n + 1$ est justifié. En effet, avec cette méthode, plus l'ordre est significatif, moins les résultats sur les âges faibles et élevés seront disponibles. Or, en dépendance, les âges élevés sont ceux ayant le plus d'impact en termes de risque. Par conséquent, dans le but de conserver une information maximale au niveau des âges élevés, les ordres 3 et 4 sont choisis (correspondant respectivement à $n = 1$ et $n = 2$).

Concernant la méthode de Whittaker-Henderson, il est nécessaire de déterminer trois paramètres : h , le poids de la régularité dans le lissage, z l'ordre auquel on calcule le critère de régularité et w_i les poids du critère de fidélité pour chaque âge.

On a ainsi :

$$W = F + hR,$$

avec W mesure de Whittaker-Henderson qu'on cherche à minimiser, F mesure de la fidélité entre les taux lissés et les taux bruts, R mesure de la régularité de la courbe lissée ; $h \in R^+$ est le poids que l'on souhaite attribuer à R par rapport à F ;

$$F = \sum w_i (\text{valeurs lissées} - \text{valeurs brutes})^2 ;$$

$$S = \sum_{x=x_{inf}}^{x=x_{sup}} (\Delta^z g_x)^2,$$

qui correspond à la somme des différences, à l'ordre z , des taux lissés.

En pratique, ces paramètres sont choisis de manière justifiée. Les poids w_i sont choisis en fonction de l'exposition à chaque âge. Au niveau de z , différentes valeurs sont étudiées. Les résultats sont présentés en Annexe 4. z est choisi dans une optique de compromis entre ratio constaté sur attendu proche de 100% et minimisation du critère de régularité. Pour h , il n'y a pas de différence significative entre les différentes valeurs étudiées.

Les lissages font l'objet de validations par les tests de Kolmogorov-Smirnov et du Chi 2. Ces tests visent à savoir si les taux lissés suivent la même loi que les taux bruts.

Les p-value correspondantes sont supérieures à 0,05. Par conséquent, l'hypothèse H_0 , qui veut que les taux lissés suivent la même loi que les taux bruts, n'est pas rejetée.

Le choix de la méthode est réalisé grâce à une comparaison des ratios constaté / attendu et régularité du lissage, dont les résultats sont présentés en annexe 5. Le choix de retenir la méthode Whittaker-Henderson est justifié dans un compromis entre régularité et fidélité. Les taux lissés obtenus sont de plus dans l'intervalle de confiance relatif aux taux bruts.

A noter que la mortalité des hommes et autres niveaux de dépendance induisent des méthodes et conclusions similaires.

Enfin, les méthodes d'extrapolation sont étudiées. En effet, les données à disposition ne sont pas suffisantes pour calculer des taux après 87 ans (moins de 15 observations). Or, il est nécessaire de projeter les taux jusqu'à extinction du portefeuille.

Pour cela, différents modèles sont étudiés : les modèles logistiques de Beard, Kannisto et Perk. D'autres méthodes sont également étudiées : Gompertz-Makeham, Brass et Denuit. Les modèles logistiques sont intéressants car induisent une extrapolation des taux croissante, lente au début, puis qui s'accélère avec l'âge, avec une limite à 1; ce qui est cohérent par rapport au risque de dépendance qui n'est pas certain d'arriver.

Les paramètres de ces modèles sont déterminés grâce aux méthodes des moindres critères ordinaires et pondérés :

$$\arg \min \sum_x (q^{brut}(x) - q^{estimé}(x))^2 ;$$

$$\arg \min \sum_x w_x (q^{brut}(x) - q^{estimé}(x))^2 ;$$

avec w_x pondération qui est calculée par l'exposition au risque.

Avec les modèles logistiques, le taux de hasard est :

$$h(x) = a + \frac{c * e^{bx}}{1 + d * e^{bx}} ;$$

avec x l'âge et a, b, c et d les paramètres à déterminer.

On a enfin $q(x)$, probabilité d'occurrence du risque sur 1 an, telle que :

$$q(x) = 1 - e^{-\int_x^{x+1} h(u) du} .$$

On remplace ensuite h par l'expression ci-dessus puis, en faisant l'hypothèse que le taux de hasard est constant sur une période d'un an, on obtient :

$$q(x) = 1 - e^{-\int_x^{x+1} h(u) du} = 1 - e^{-h(x)} = 1 - e^{-a - \frac{c * e^{bx}}{1 + d * e^{bx}}} .$$

Les modèles logistiques de Beard, Kannisto et Perk sont des variantes du modèle logistique.

Le modèle de Beard suppose que la constante peut être supprimée pour les âges élevés. On a :

$$h(x) = \frac{a * e^{bx}}{1 + c * e^{bx}} .$$

Le modèle de Kannisto suppose que c et d sont proches et constituent un seul paramètre :

$$h(x) = \frac{a * e^{bx}}{1 + a * e^{bx}} .$$

Le modèle de Perk dispose d'une constante au numérateur :

$$h(x) = \frac{a * e^{bx} + d}{1 + c * e^{bx}} .$$

La loi de Gompertz-Makeham dispose de 3 paramètres :

$$h(x) = \alpha + \beta * \gamma ;$$

avec α risque d'occurrence de l'événement liée à un accident donc non lié à l'âge ; $\beta * \gamma$ influence croissante de l'âge sur le risque avec $\beta > 0$ et $\gamma > 1$.

La méthode de Brass fait le lien entre les logits des taux estimés et ceux des logits des taux issus d'une table de référence :

$$\ln\left(\frac{q(x)}{1-q(x)}\right) = a * \ln\left(\frac{q^{réf}(x)}{1-q^{réf}(x)}\right) + b ;$$

avec a, b paramètres à déterminer, q(x) taux estimés et $q^{réf}(x)$ taux de référence.

La méthode de Denuit et Goderniaux dispose que les logarithmes des taux de hasard peuvent être simulé par un polynôme :

$$\ln(q_x) = a + b * x + c * x^2;$$

avec pour les âges importants, des contraintes sur la croissance et l'occurrence certaine du risque sont déterminées.

Afin de choisir la tranche d'âge à partir de laquelle les données lissées seront extrapolées, les méthodes présentées ci-dessus sont appliquées sur différentes tranches d'âge. Le choix final semble pertinent : l'âge minimal a une influence non significative, contrairement à l'âge maximal. L'âge maximal le plus élevé est choisi dans le but de conserver le maximum d'informations sur les âges élevés.

La base contient des données jusqu'à 87 ans. Ainsi, pour choisir la méthode d'extrapolation, une autre base est utilisée (assurance obsèques sans sélection médicale donc dont les taux doivent être supérieurs à ceux extrapolés). Le choix des méthodes retenues semble pertinent : Beard, Perk et Gompertz-Makeham ne sont pas choisies car les extrapolations sont supérieures aux résultats du Portefeuille obsèques de comparaison. Or, des taux inférieurs sont recherchés car le portefeuille dépendance induit une sélection médicale. Denuit est écartée car l'extrapolation est trop « plate ».

Ainsi, les méthodes de Brass et Kannisto sont retenues. In fine, on compare les résultats obtenus avec ces deux méthodes, pour les lois extrapolées, avec la mortalité générale du portefeuille (on souhaite notamment avoir des taux supérieurs pour les hommes que pour les femmes).

4.1.2. Etude de la construction des lois d'expérience de mortalité des dépendants

La construction des lois de mortalité des dépendants doit prendre en compte une dimension supplémentaire par rapport aux autonomes : l'ancienneté dans l'état en dépendance.

Or, du fait que la base de données relative aux assurés Protectys Autonomie soit trop faiblement fournie pour avoir des taux de dépendance fiable à chaque âge, des consolidations sont réalisées par tranche de 5 ans pour calculer les lois d'expérience de comparaison.

Or, les lois utilisées par LBPP sont calculées à l'année. Par conséquent, la comparaison des lois LBPP et des lois d'expérience ne sera pas réalisée sur le périmètre des dépendants. L'approche Worst Case portera donc sur le périmètre des autonomes.

4.1.3. Analyse des contrôles réalisés sur les lois d'expérience

Afin de s'assurer de la pertinence de la construction des lois d'expérience de comparaison, des contrôles ont été réalisés : calculs et comparaison de ratios « constaté sur attendu », avec en « constaté » les données issues des observations réelles et en « attendu » les lois d'expérience de comparaison, selon différents périmètres (année, âge, portefeuille global). In fine, ces ratios sont estimés prudents quand ils sont supérieurs à 1, que ce soit pour la mortalité relative aux autonomes ou aux dépendants. En effet, cela induit moins de paiements de rentes de dépendance. Pour la dépendance, ils sont prudents si inférieurs à 1 car on estime alors que plus de rentes de dépendance seront payés que ce qui est constaté. Les résultats de ces contrôles, réalisés tout d'abord par âge et année d'occurrence, disposent une prudence pour la mortalité des autonomes et pour l'incident de 75 ans et plus. Cependant, ce n'est pas le cas pour les âges de 60 à 74 ans. Etant donnée qu'il y a peu d'entrées en dépendance pour ces âges, l'impact est faible.

Enfin, une étude similaire est également réalisée sur plusieurs années de suite.

4.1.4. Synthèse et limites de la construction des lois d'expérience de comparaison

En synthèse, comme présenté ci-dessus, les lois d'expérience de comparaison ont été construites selon une méthodologie éprouvée grâce à plusieurs facteurs :

- Etude du portefeuille, des données et de leurs spécificités (censures, troncatures, risques concurrents) ;
- Etudes de plusieurs méthodes pour chaque étape de la modélisation ;
- Contrôles par intervalles de confiance ;
- Contrôles par différentes comparaisons de calculs de ratios.

Comme toute modélisation, celle-ci comporte néanmoins des limites dont nous pouvons énumérer certaines ci-après :

- La base de données est peu fournie sur les âges élevés et les passages en dépendance, induisant des risques de projection non pertinente ; de plus, les passages en dépendance peuvent être chronophages car nécessitent la consolidation et la stabilisation de l'état de dépendance de l'assuré, ce qui induit que pour avoir un nombre exhaustif d'entrée en dépendance sur une période donnée, il est nécessaire d'avoir un temps significatif de recul ;
- Même si, pour chaque étape de modélisation, plusieurs méthodes ont été étudiées, elles peuvent toujours être complétées ;
- Au niveau des contrôles « constaté sur attendu », la comparaison n'est pas possible pour les âges élevés, donc les taux extrapolés, car le portefeuille de base ne comporte pas assez de données au-delà de 84 ans.

4.2. Définition de l'appétence au risque de modèle

En amont de l'analyse par l'approche Worst Case, il est nécessaire de définir l'appétence au risque, c'est-à-dire la limite partir de laquelle la mise en place d'un plan d'action est nécessaire. Ici, la limite est établie à 5%. Ce choix de 5%, bien qu'arbitraire, vise à analyser une variation suffisamment significative du best estimate.

Cela signifie que si l'approche Worst Case donne une estimation du risque de modèle supérieure à 5% (plus de 5% de variation entre le résultat initial et les résultats des modèles de comparaison) alors il sera nécessaire de mettre en place des plans d'actions afin de vérifier que le risque est maîtrisé.

4.3. Comparaison des lois de référence et des lois d'expérience

Comme présenté dans le chapitre II, Protectys Autonomie, contrat collectif à adhésion facultative, offre, en cas de dépendance de l'assuré, une garantie de prestations forfaitaires, et/ou une rente mensuelle viagère et/ou un montant de capital premiers frais.

Trois lois interviennent pour modéliser le risque dépendance du produit Protectys Autonomie :

- La mortalité des assurés autonomes ;
- L'incidence de la dépendance ;
- La mortalité des assurés dépendants au sens de cette même définition.

La modélisation de ces lois est notamment utilisée par LBP Prévoyance dans le calcul du best estimate.

L'objectif ici sera de comparer les lois de référence à des lois d'expérience et d'analyser, grâce à la méthode Worst Case présentée en chapitre 2, l'écart en termes de best estimate, obtenu en utilisant les lois d'expérience plutôt que les lois de référence et d'estimer le risque de modèle correspondant.

Par ailleurs, les ratios constatés/attendus relatifs aux lois de référence indiquent que celles-ci sont prudentes. Aussi, des sensibilités seront également appliquées à ces dernières afin de mesurer l'impact d'une réduction de cette prudence.

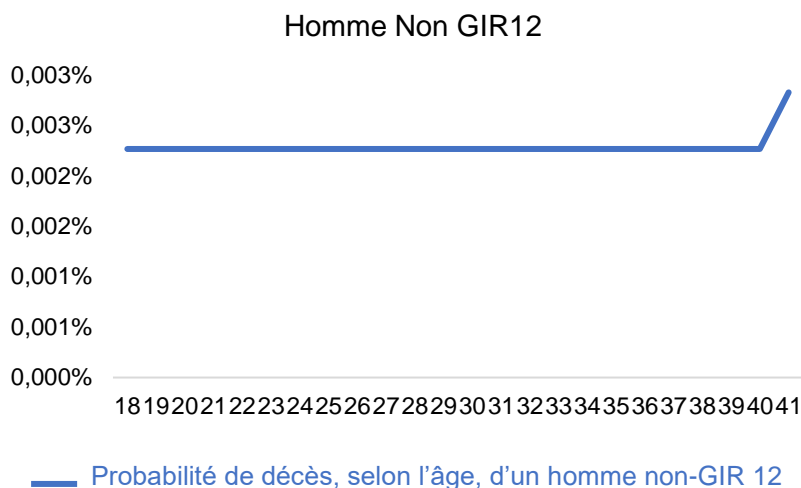
4.3.1. Analyses préliminaires à la comparaison des lois obtenues

Afin de respecter la confidentialité des données utilisées dans le cadre de ce mémoire, les lois présentées ci-dessous ont fait l'objet d'une transformation qui n'impacte pas le raisonnement.

Avant de comparer les lois de référence et les lois d'expérience, il est nécessaire de travailler les lois d'expérience afin que leur format et certaines hypothèses soient cohérents avec les lois de référence.

- Au niveau des lois de mortalité des autonomes et des lois d'incidence :

Les lois de référence sont stables de 18 à 40 ans compris ; cf. figures ci-dessous.



Figures 25 et 26 - Titre : Lois de référence de mortalité des autonomes non-GIR 12

Les lois de référence sont stables de 18 à 40 ans compris.

Par conséquent, la même hypothèse est appliquée aux lois d'expérience de comparaison.

Par ailleurs, les lois de référence sont modélisées jusqu'à 120 ans alors que les lois d'expérience de comparaison sont modélisées jusqu'à 109 ans. Dans ce cadre, afin d'assurer la comparabilité des lois, les taux de référence sont repris dans les lois d'expérience de comparaison au-delà de 109 ans.

Enfin, au niveau des lois d'expérience de comparaison, 30 taux, sur un total de 1 236, soit 2,4%, nécessitent, après analyse graphique, un re-lissage. La méthode des moyennes mobiles est appliquée.

La valeur lissée M_x du taux est calculée en considérant la moyenne mobile de $2n + 1$ estimations consécutives non lissées :

$$M_x = \frac{1}{2n + 1} * \sum_{i=-n}^n q_{x+i}.$$

Cette méthode a l'avantage d'être facile à mettre en œuvre et étant donnée la faible volumétrie des points à retraiter, il est considéré que le risque lié à son utilisation n'est pas significatif. L'ordre 3, qui correspond à $n = 1$, est choisi car il y a peu de données aux extrêmes de la base de données donc il est plus prudent de choisir l'ordre le plus faible possible afin d'utiliser les données disponibles et les plus pertinentes possibles. Enfin, comme le montre l'annexe 5, les ratios « constatés sur attendus » calculés sur l'ensemble des taux bruts, sont proche de 1.

- Au niveau des lois de mortalité des dépendants :

Les lois de référence de mortalité des dépendants, sont modélisées selon deux critères : l'âge d'entrée en dépendance et la durée de vie en dépendance. Pour une durée de vie en dépendance donnée, elles sont stables jusqu'à un âge d'entrée en dépendance de 44 ans puis varient ensuite pour chaque âge au-delà de 44 ans.

Les lois d'expérience de comparaison, du fait de la faible volumétrie des données, sont calculées par tranche de dix ans. Par exemple, pour une durée de vie en dépendance donnée, le taux est stable pour les entrées en dépendance entre 70 et 79 ans, entre 80 et 89 ans et ainsi de suite.

A titre d'illustration, cf. figures ci-dessous.

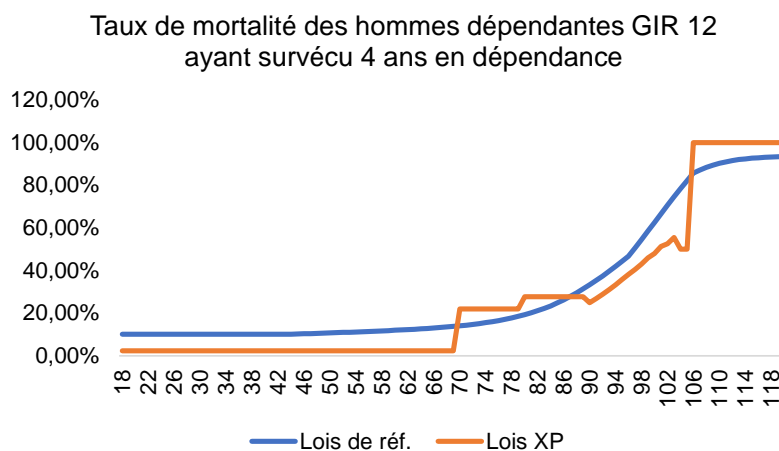
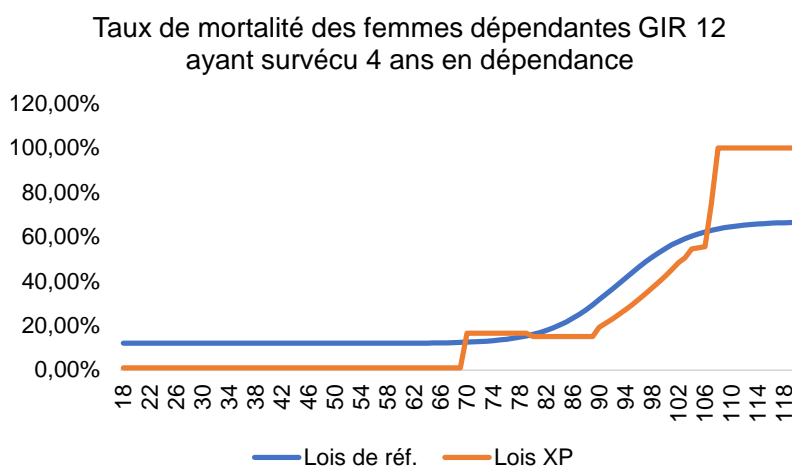


Figure 27 et 28 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des dépendants GIR 12 entrée en dépendance il y a 4 ans

Le manque de données, qui induit cette construction « par palier » des lois d'expérience de mortalité des dépendants, ne permet pas une comparabilité suffisante avec les lois de référence de mortalité des dépendants. Aussi, l'étude du risque de modèle selon l'approche Worst Case n'inclura pas cette partie.

4.3.2. Comparaison des lois en amont de l'analyse de l'approche Worst Case

En amont de l'approche Worst Case, il est intéressant de comparer les lois de référence et d'expérience. En effet, l'objectif est d'appréhender les écarts, entre ces deux jeux de lois, afin d'essayer de comprendre les résultats qui seront obtenus, avec l'approche Worst Case.

Pour cela, on réalise une comparaison graphique et également on calcule les sommes des différences entre les taux pour différentes tranches d'âges.

Nous n'avons pas accès au détail de la méthode de modélisation des lois de référence. Cependant, comme on peut le voir ci-dessous, l'observation graphique nous laisse penser que les lois de référence n'ont pas fait l'objet de lissage sur les âges élevés alors qu'un lissage a été appliqué aux lois de comparaison.

- Au niveau des lois de mortalité des autonomes, on obtient les graphes de comparaison suivants.

Comparaison n°1.1.

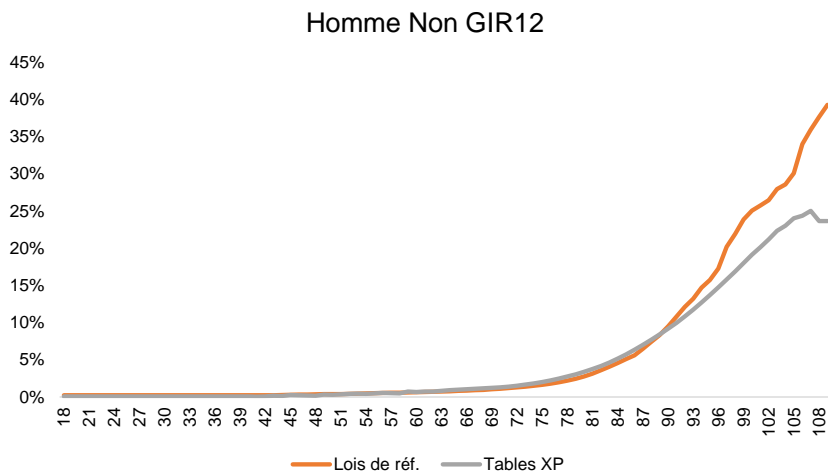


Figure 29 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des hommes autonomes non-GIR 12

	Somme des différences XP – Réf.	Somme des carrés des différences XP – Réf.
De 18 à 48 ans	-4,29%	0,01%
De 49 à 79 ans	4,72%	0,02%
80 ans et plus	-104,70%	9,45%
Au global	-104,26%	9,48%

Tableau 32 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une mortalité plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°1.2.

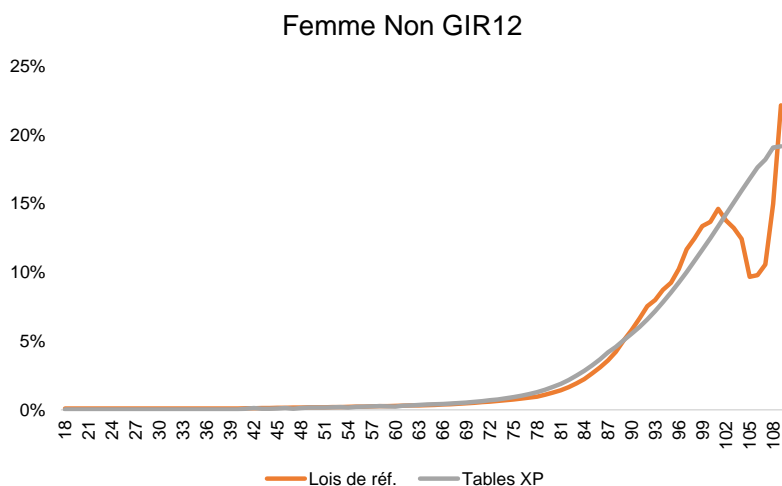


Figure 30 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des femmes autonomes non-GIR 12

	Somme des différences XP – Réf.	Somme des carrés des différences XP – Réf.
De 18 à 48 ans	-1,68%	0,00%
De 49 à 79 ans	1,93%	0,00%
80 ans et plus	21,58%	2,31%
Au global	21,83%	2,31%

Tableau 33 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Du fait des variations de la loi de référence, la comparaison est complexe à réaliser. Les lois de référence induisent une mortalité moins importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°1.3.

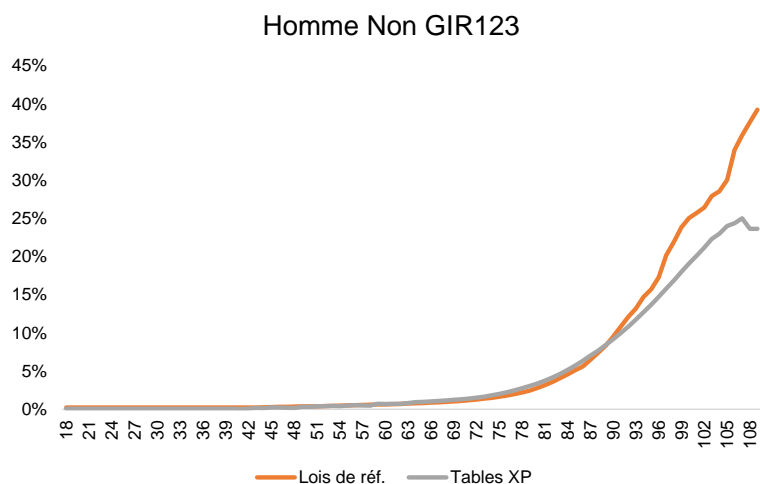


Figure 31 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des femmes autonomes non-GIR 12

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	-3,80%	0,00%
De 49 à 79 ans	3,79%	0,01%
80 ans et plus	-11,97%	7,71%
Au global	-11,98%	7,73%

Tableau 34 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une mortalité plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°1.4.

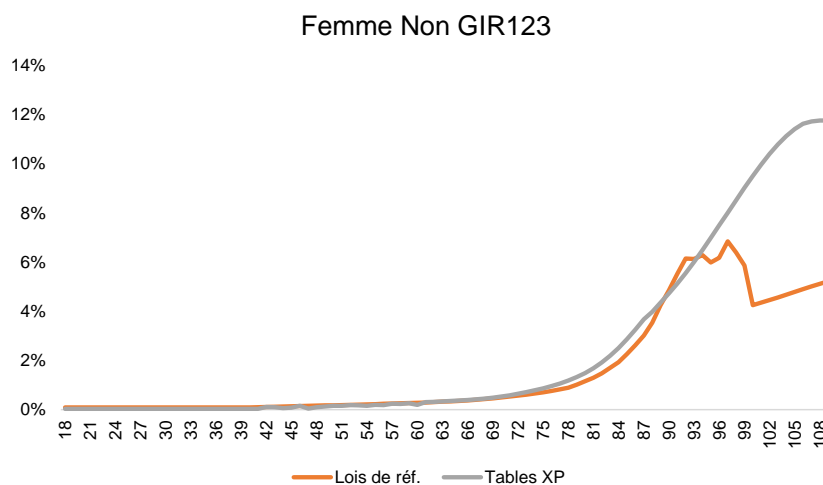


Figure 32 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des femmes autonomes non-GIR 123

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	-1,72%	0,00%
De 49 à 79 ans	1,39%	0,00%
80 ans et plus	75,11%	4,17%
Au global	74,79%	4,18%

Tableau 35 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une mortalité moins importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°1.5.

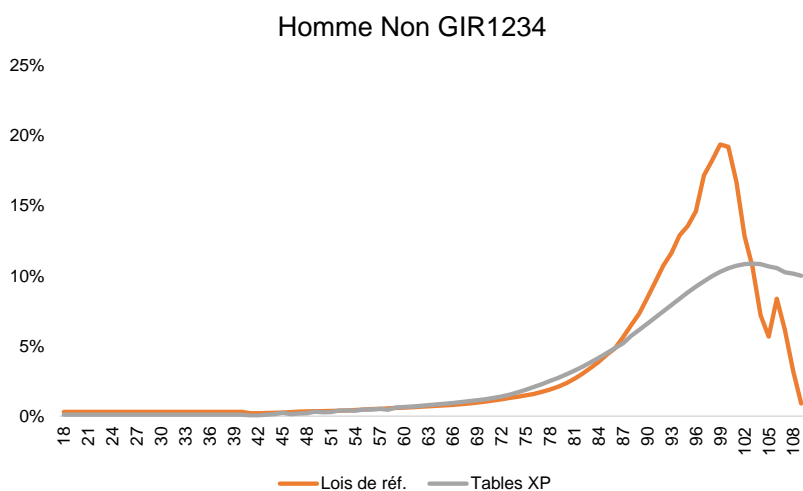


Figure 33 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des hommes autonomes non-GIR 1234

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	-5,56%	0,01%
De 49 à 79 ans	4,57%	0,02%
80 ans et plus	-36,16%	6,23%
Au global	-37,15%	6,26%

Tableau 36 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une mortalité plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°1.6.

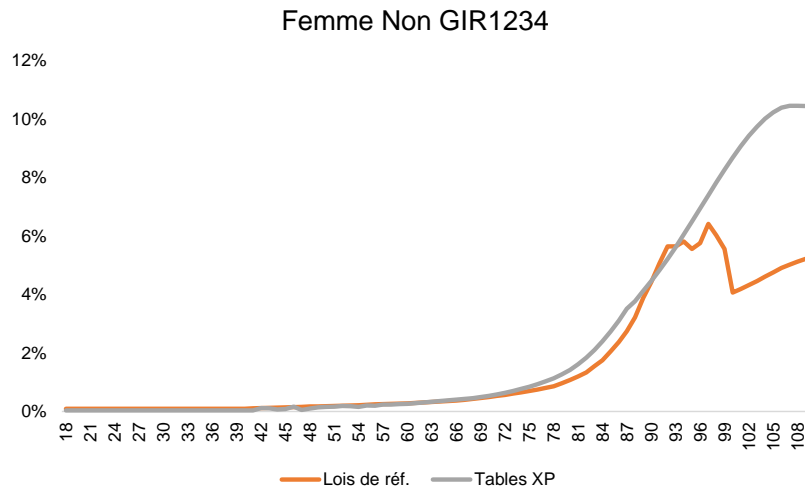


Figure 34 - Titre : Lois de référence et lois d'expérience de comparaison (XP) de mortalité des femmes autonomes non-GIR 1234

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	-1,67%	0,00%
De 49 à 79 ans	1,33%	0,00%
80 ans et plus	64,91%	2,91%
Au global	67,49%	2,92%

Tableau 37 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une mortalité moins importante, notamment au niveau des âges élevés mais également au global.

En synthèse, il est difficile de conclure car les écarts entre les lois de référence et les lois d'expérience de la mortalité des autonomes varient dans un sens comme dans l'autre. Dans tous les cas, l'impact sur le best estimate est complexe à anticiper car une mortalité plus importante induit un risque de dépendance réduit mais également moins de primes.

- Au niveau des lois d'incidence, on obtient les graphes de comparaison suivants.

Comparaison n°2.1.

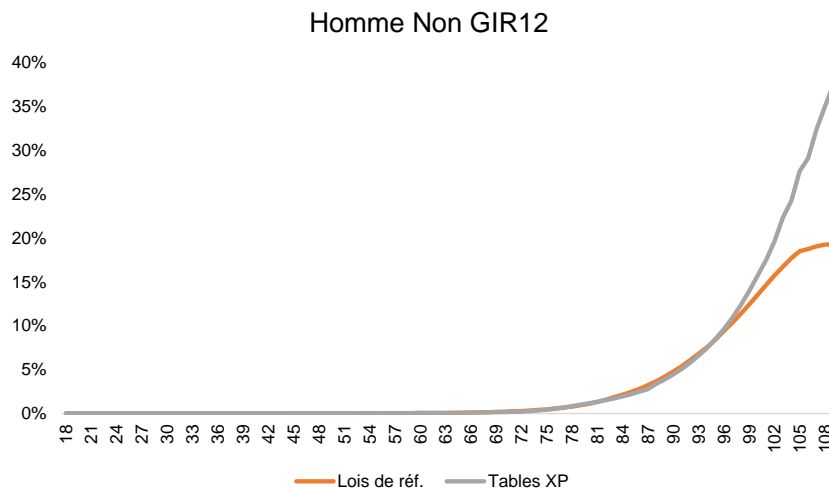


Figure 35 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des hommes non-GIR 12 (passage en GIR 12)

	Somme des différences XP – Réf.	Somme des carrés des différences XP – Réf.
De 18 à 48 ans	0,68%	0,00%
De 49 à 79 ans	-0,55%	0,00%
80 ans et plus	88,05%	10,59%
Au global	88,18%	10,59%

Tableau 38 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°2.2.

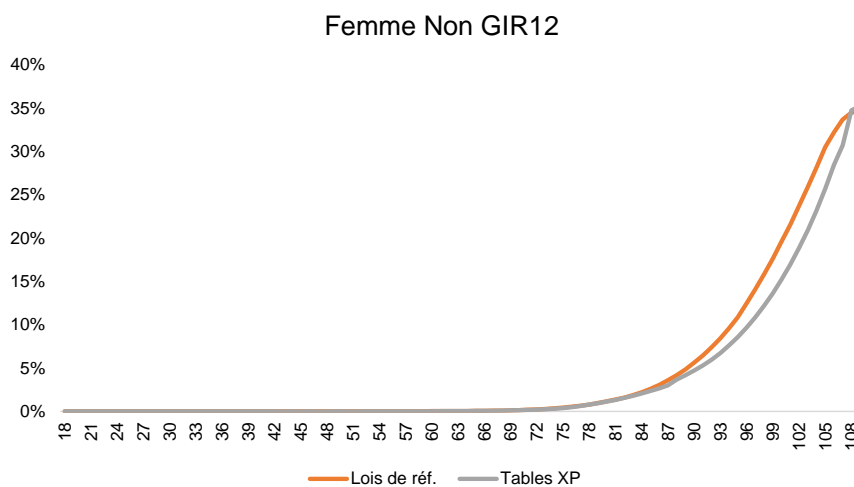


Figure 36 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des femmes non-GIR 12 (passage en GIR 12)

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	0,42%	0,00%
De 49 à 79 ans	-0,43%	0,00%
80 ans et plus	-60,31%	2,25%
Au global	-60,33%	2,25%

Tableau 39 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence moins importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°2.3.

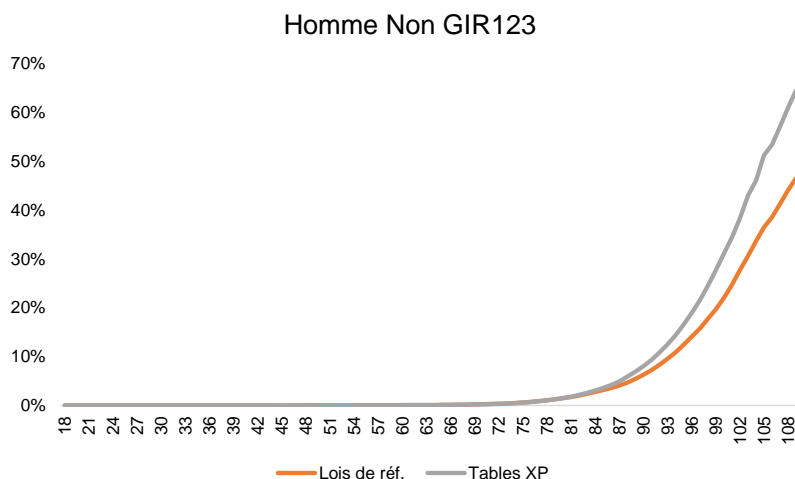


Figure 37 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des hommes non-GIR 123 (passage en GIR 123)

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	0,74%	0,00%
De 49 à 79 ans	-0,59%	0,00%
80 ans et plus	181,92%	21,17%
Au global	182,07%	21,17%

Tableau 40 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°2.4.

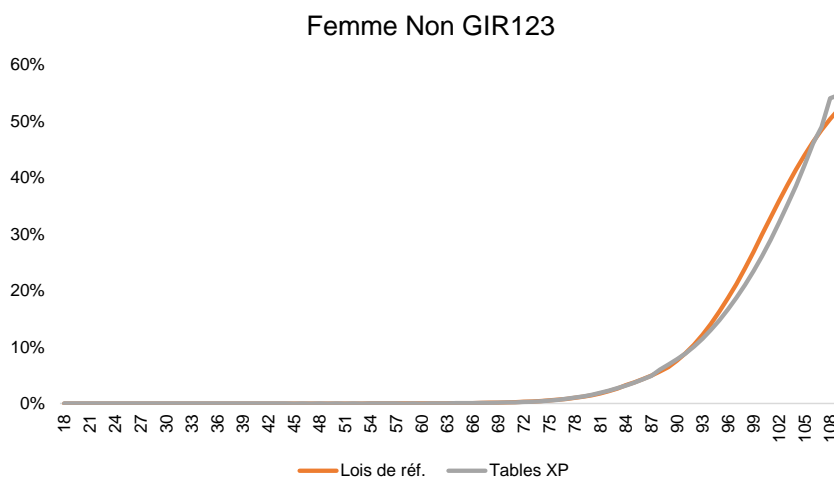


Figure 38 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des hommes non-GIR 123 (passage en GIR 123)

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	0,51%	0,00%
De 49 à 79 ans	-0,15%	0,00%
80 ans et plus	-26,67%	1,24%
Au global	-26,31%	1,24%

Tableau 41 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence moins importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°2.5.

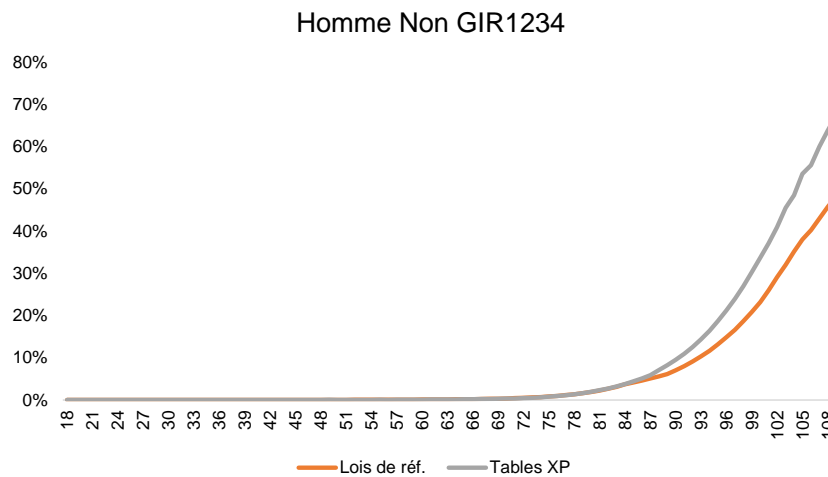


Figure 39 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des hommes non-GIR 1234 (passage en GIR 1234)

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	0,67%	0,00%
De 49 à 79 ans	-0,75%	0,00%
80 ans et plus	205,44%	25,48%
Au global	205,36%	25,48%

Tableau 42 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence plus importante, notamment au niveau des âges élevés mais également au global.

Comparaison n°2.6.

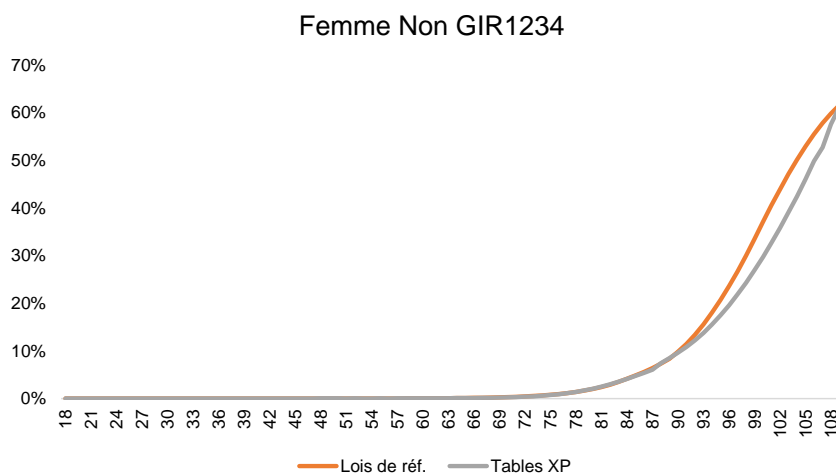


Figure 40 - Titre : Lois de référence et lois d'expérience (XP) d'incidence des femmes non-GIR 1234 (passage en GIR 1234)

	Somme des différences XP - Réf	Somme des carrés des différences XP - Réf
De 18 à 48 ans	0,45%	0,00%
De 49 à 79 ans	-0,31%	0,00%
80 ans et plus	-90,19%	5,54%
Au global	-90,05%	5,54%

Tableau 43 - Titre : Comparaison des lois de référence et d'expérience par calcul des différences des taux obtenus sur différentes tranches d'âges

Les lois de référence induisent une incidence moins importante, notamment au niveau des âges élevés mais également au global.

En synthèse, le sens des écarts, pour l'incidence, varie en fonction du sexe. En effet, pour la population des hommes, les taux d'incidence des lois de référence sont inférieurs aux taux d'incidence des lois d'expérience de comparaison. Inversement, pour la population des femmes, les taux d'incidence des lois de référence sont supérieurs aux taux d'incidence des lois d'expérience de comparaison.

Or, sachant que :

- Plus l'incidence est estimée significative, plus le risque estimé est important et plus le best estimate le sera ;
- Le portefeuille client du produit Protectys Autonomie est composé à un tiers d'hommes et à deux tiers de femmes ;

Les lois de référence semblent donc plus prudentes que les lois d'expérience car induisent une incidence plus importante pour les femmes.

4.4. Résultats de l'approche Worst Case : comparaison des best estimate obtenus entre les lois de référence et les lois de comparaison

Les lois de référence et les lois d'expérience de comparaison sont implémentées dans un outil de projection qui permet de calculer le best estimate. Le modèle de référence et les modèles de comparaison sont les suivants.

Numéro de modèle	Description
Modèle 1 de référence	Lois de référence
Modèle 1 de comparaison	Lois d'expérience de mortalité des autonomes non-GIR 12 et toutes choses égales par ailleurs
Modèle 2 de comparaison	Lois d'expérience d'incidence des autonomes non-GIR 12 et toutes choses égales par ailleurs
Modèle 3 de comparaison	Lois d'expérience de mortalité des autonomes non-GIR 123 et toutes choses égales par ailleurs
Modèle 4 de comparaison	Lois d'expérience d'incidence des autonomes non-GIR 123 et toutes choses égales par ailleurs
Modèle 5 de comparaison	Lois d'expérience de mortalité des autonomes non-GIR 1234 et toutes choses égales par ailleurs
Modèle 6 de comparaison	Lois d'expérience d'incidence des autonomes non-GIR 1234 et toutes choses égales par ailleurs

Tableau 44 - Titre : Présentation des modèles de référence et de comparaison

4.4.1. Résultats obtenus avec le modèle de référence et avec les modèles de comparaison

Ci-dessous, le tableau de synthèse des résultats obtenus selon l'approche Worst Case.

Composantes (M €)	Modèles						
	Référence	1	2	3	4	5	6
Primes Chargées	738	737	739	730	736	737	738
Prestations Sinistres	1 070	1 062	1 044	1 057	1 080	1 068	1 071
Frais	143	142	143	141	143	143	142
Best Estimate brut	475	467	448	468	487	474	477
Approche WC - Variations par rapport au modèle 1	N/A	-1,7%	-5,7%	-1,5%	2,5%	-0,2%	0,4%
Sens de la variation	N/A	Baisse	Baisse	Baisse	Hausse	Baisse	Hausse

Tableau 45 - Titre : Synthèse de l'approche Worst Case, comparant les lois de référence et les lois d'expérience de comparaison

En termes de *best estimate*, les lois de référence sont, dans 4 cas sur 6, plus prudentes au sens de la définition retenue, que les lois d'expérience de comparaison. En l'occurrence, le dépassement de l'appétence au risque pour le modèle 3 de

comparaison (lois d'expérience d'incidence des autonomes non-GIR 12) ne comporte pas un risque significatif car les lois de référence sont plus prudentes.

D'après les analyses réalisées précédemment, il est possible que la variation soit issue des lois d'incidence des femmes qui sont plus prudentes au niveau des lois de référence (incidence plus importante donc une évaluation du best estimate plus importante) que des lois d'expérience, cumulé à un portefeuille composé à deux tiers de femmes.

Par conséquent, il n'est pas nécessaire, selon l'appétence au risque, de mettre en place de plan d'action. Il pourra tout de même être intéressant de suivre les ratios constaté / attendu pour s'assurer (i) d'une part que les lois utilisées restent prudentes et (ii) d'autre part que les lois de référence ne sont pas « trop » prudentes, ce qui pourrait induire un positionnement concurrentiel non adapté. Afin de compléter l'analyse, nous étudions tout de même, dans le chapitre 5, la mise en place d'un plan d'action de réduction du risque de modèle par identification des profils d'assurés réalisant des réductions en utilisant des méthodes d'analyses univariées et de machine learning de prédiction.

4.5. Mise en place de sensibilités

Les résultats de l'approche Worst Case induisent que le best estimate de référence est prudent, au sens de la définition retenue. Sur les six modèles de comparaison étudiés, les variations maximales des résultats par rapport au modèle de référence sont de - 5,7%. Ces variations vont ainsi dans le sens d'une réduction du best estimate.

Dans ce cadre, on réalise des analyses complémentaires sur les lois dépendance. Il est par exemple intéressant d'analyser les rapports « constaté sur attendus » entre les lois de référence et ce qui est réellement observé et de réaliser des analyses de sensibilité en conséquence, afin d'en mesurer l'impact sur le best estimate.

4.5.1. Au niveau des lois de référence de mortalité des autonomes

Les rapports « constaté sur attendu », segmentés par âge, sont les suivants.

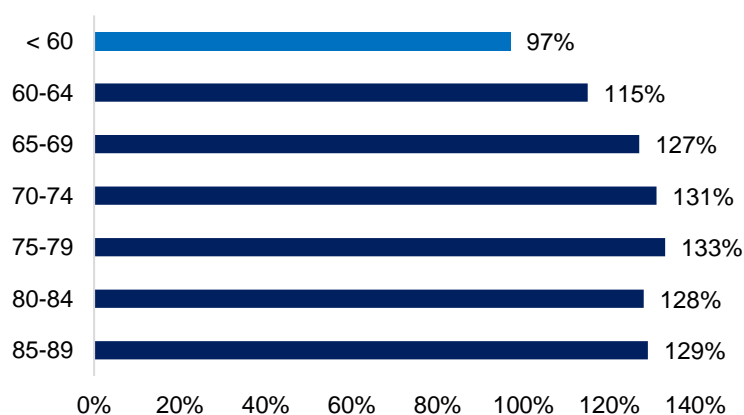


Figure 41 - Titre : rapports « constaté sur attendu », segmentés par âge, au niveau des lois de mortalité des autonomes

On constate que les lois de mortalité des valides utilisées pour le provisionnement ont un rapport constaté sur attendu supérieur à 100%, excepté pour les moins de 60 ans (Non GIR12 et Non GIR123), ce qui induit une hypothèse de mortalité inférieure à ce qui est observé réellement. Or, une hypothèse de mortalité moins importante que la réalité a deux conséquences : une surestimation des primes et une surestimation des personnes pouvant devenir dépendantes. In fine, le deuxième facteur l'emporte et cela est prudent. Les lois de mortalité des valides utilisées pour le provisionnement sont donc prudentes excepté pour les moins de 60 ans.

Les plages d'âge les plus significatives, en termes de volumétrie d'assurés, sont les plages 65-83 ans (plus de 3000 assurés).

Etant donné les rapports « constaté sur attendu », les sensibilités appliquées sur les lois de mortalité des autonomes, le sont à la hausse et sur ces tranches d'âge. En synthèse, les sensibilités finales retenues sur les lois de mortalité des autonomes sont de 5% à la hausse sur la tranche 65-83 ans. Ce choix de 5%, bien qu'arbitraire, vise à analyser une variation suffisamment significative pour avoir un impact sur le best estimate mais pas trop éloignée des lois de référence.

4.5.2. Au niveau des lois de référence de mortalité des dépendants

Les rapports « constaté sur attendu », segmentés par âge, sont les suivants.

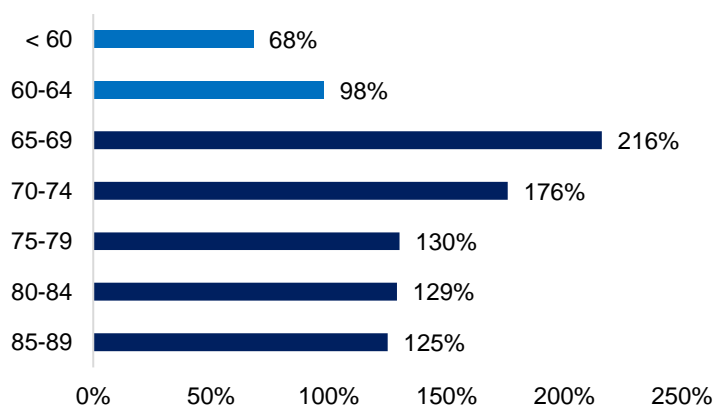


Figure 42 - Titre : rapports « constaté sur attendu », segmentés par âge, au niveau des lois de mortalité des dépendants

Les lois de mortalité en GIR12 et en GIR123 utilisées pour le provisionnement sont prudentes excepté pour les moins de 60 ans. En effet, une mortalité réelle plus importante des dépendants que prévu, induit un sur-provisionnement.

Les plages d'âge les plus significatives, en termes de volumétrie d'assurés, sont les plages 75-79 ans (15% de la population assurée), 80-84 ans (29% de la population assurée) et 85-89 ans (35% de la population assurée).

Etant donné les rapports « constaté sur attendu », les sensibilités appliquées sur les lois de mortalité des dépendants, le sont à la hausse. En synthèse, les sensibilités finales retenues sur les lois de mortalité des dépendants sont de 5% à la hausse sur la tranche 75-89 ans.

4.5.3. Au niveau des lois d'incidence de référence

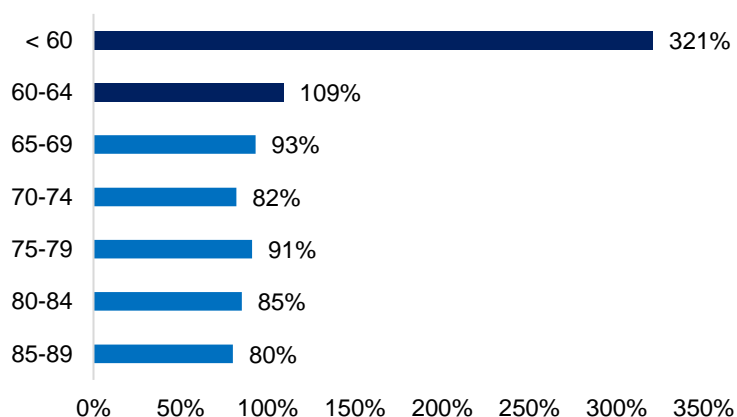


Figure 43 - Titre : rapports « constaté sur attendu », segmentés par âge, au niveau des lois d'incidence

Les lois d'incidence en GIR12 et en GIR123 utilisées pour le provisionnement sont prudentes excepté pour les moins de 65 ans. En effet, un taux réel d'incidence inférieur au taux prévu induit un sur-provisionnement.

Les observations sont significatives sur la plage d'âges 63-87 ans.

Etant donné les rapports « constaté sur attendu », les sensibilités appliquées sur les lois d'incidence, le sont à la baisse. En synthèse, les sensibilités finales retenues sur les lois de d'incidence sont de 5% à la baisse sur la tranche 65-87 ans.

4.5.4. Résultats

L'utilisation des lois, dans l'estimation du best estimate, est ici étudiée, cf. tableau ci-dessous.

	BE de référence	Sensibilité 1	Sensibilité 2	Sensibilité 3	Les 3 sensibilités ensemble
Best estimate	475	479	453	456	438
Ecart au BE de référence	N/A	+1%	-5%	-4%	-8%

Tableau 46 - Titre : résultats, en termes de BE, des sensibilités sur les lois de référence

Les résultats montrent que le best estimate est sensible aux variations des lois étudiées. Les sensibilités vont en moyenne dans le sens d'une diminution du best estimate, ce qui corrobore les résultats obtenus avec la méthode Worst Case et induit donc que le best estimate actuel, c'est-à-dire calculé à partir des lois de référence, peut être considéré comme prudent.

Ces résultats, issus de travaux de modélisation, sont à considérer avec prudence et devront, comme présenté dans le papier « Les modèles de risque et la prise de décision » de Hélène N'Diaye et al, être complétés d'autres indicateurs de suivi, comme le rapport « sinistres sur primes ».

Afin de compléter l'analyse, nous étudions, dans le chapitre 5, la mise en place d'un plan d'action de réduction du risque de modèle par identification des profils d'assurés réalisant des réductions en utilisant des méthodes d'analyses univariées et de machine learning de prédiction.

5. Chapitre 5 – Mise en place d'un plan d'actions pour réduire la volatilité du BE : étude des lois de réduction du produit Protectys autonomie

Pour rappel, ce mémoire est structuré en trois approches, allant de la plus large à la plus précise : (i) méthode large d'évaluation du risque de modèle appliquée sur plusieurs modèles (chapitres 2 - théorie - et 3 - application), (ii) méthode d'évaluation précise du risque de modèle appliquée à un modèle déterminé (chapitres 2 - théorie - et 4 - application), (iii) étude d'un plan d'action visant à réduire le risque de modèle par application d'analyses univariées et méthodes de machine learning de prédiction (chapitre 5), cf. figure ci-dessous.

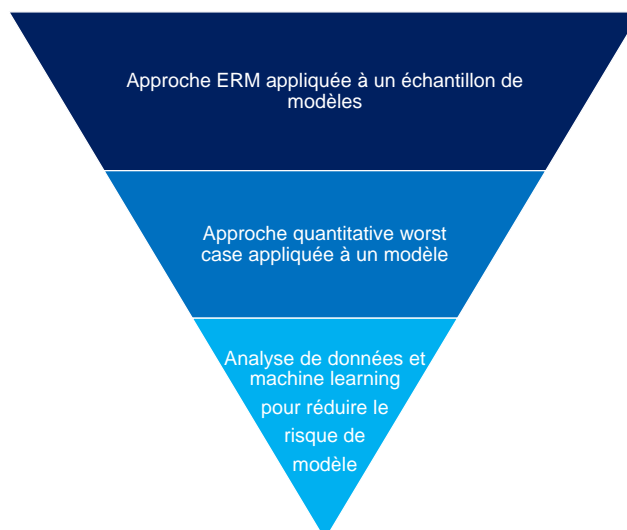


Figure 44 - Titre : approches étudiées dans ce mémoire

L'objectif de ce chapitre est la mise en œuvre d'un plan d'action afin de réduire le risque de modèle lié à une sous-estimation du best estimate du produit dépendance Protectys. On s'intéresse ici en particulier aux résiliations et réductions du produit Protectys. En effet, si un assuré résilie son contrat au-delà de 8 ans après la souscription, alors il ne paiera plus de primes mais sera potentiellement toujours couvert par une garantie réduite. Si un assuré résilie avant, il ne sera en revanche plus couvert. Les réductions ont un impact positif sur le best estimate. Ainsi, réduire dans la réalité les réductions permet de réduire le risque de sous-estimation du best estimate.

L'estimation de la réduction induit un risque au sein de la modélisation du best estimate. Dans ce cadre, l'objectif de ce chapitre est d'identifier, par analyses de données et méthode de machine learning de prédiction, les caractéristiques des assurés réalisant des réductions, afin de mettre en place des actions, commerciales par exemple, pour réduire leur nombre. Cela permettra, d'une part, de réduire le risque de modèle, mais également de mieux connaître la population assurée.

5.1. Contexte réglementaire

La loi Chatel, datant de 2005, a instauré la communication obligatoire, par l'assureur vis-à-vis de l'assuré, d'un avis d'échéance entre 15 jours et 3 mois avant chaque date de reconduction tacite. Elle s'applique aux contrats de prévoyance et de dépendance.

De plus, la Directive de distribution d'assurance, applicable également à l'assurance prévoyance, a instauré davantage de transparence et de devoir de conseil dans la vente de produits d'assurance. Dans ce cadre, les assurés sont plus enclins à résilier leurs contrats d'assurance.

En outre, Solvabilité 2 induit que les assureurs doivent anticiper les comportements de leurs clients, dont les résiliations, et les prendre en compte dans le calcul du best estimate.

Ainsi, en l'espèce, les contrats Protectys sont des contrats d'assurance dépendance à reconduction tacite, ils sont donc soumis à ces réglementations et à ces enjeux.

5.2. Littérature

En ce qui concerne l'analyse univariée, nous nous sommes inspirés du mémoire d'actuariat « Etude des profils de résiliation sur un portefeuille santé individuelle » (A. GEHLER, 2009), qui présente et utilise différentes méthodes d'analyse appliquées afin de déterminer des profils de résiliation.

Pour ce qui est de l'application de méthodes de machine learning de prédiction nous permettant d'identifier les variables déterminantes, nous nous sommes inspirés des cours dispensés dans le cadre du Certificat d'Expertise Actuarielle d'O. LOPEZ (La Sorbonne Université, 2021), de l'étude « Tarification des risques en assurance-non-vie, une approche par modèle d'apprentissage statistique » (A. PAGLIA et al., 2011), complétés par des bibliothèques Python (notamment scikit-learn.org et stat4decision.com/fr/foret-aleatoire-avec-python).

Différentes méthodes issues de ces papiers ont ainsi été utilisées ici et appliquées au Portefeuille Protectys afin d'identifier des caractéristiques déterminantes des assurés réalisant des résiliations/réductions.

5.3. Description et traitement des données relatives aux résiliations et réductions

5.3.1. Liste des données

Les bases de données ont été créées à l'aide de SAS Enterprise Guide puis exportées au format « table sas ». Les données ont été arrêtées à fin 2021. Une attention particulière est portée au traitement de ces données, qui, bien qu'anonymisées, peuvent contenir des informations sensibles.

Avant pré-traitements, il y a 272 071 lignes avec une ligne par contrat, avec les colonnes de données suivantes :

Libellé donnée	Description
NUM_CONTRAT	Numéro de contrat
ID_PERSONNE	Identifiant du client 1
num_sirocco	Identifiant du client 2
CD_PRODUI	Code produit
produit	Libellé du produit
canal_vente	Canal de vente
date_sou	Date de souscription
date_effet_entree	Date d'effet du contrat
date_premier_anniv	Date du premier anniversaire du contrat
date_effet_sortie	Date de fin du contrat
dtdeces	Date du décès du client, le cas échéant
dtresil	Date de résiliation du contrat, le cas échéant
CD_SOR_CONTRAT	Code indiquant si le contrat est terminé et pourquoi
LIB_SORT_CONTRAT	Libellé correspondant à la sortie ou non du contrat et pourquoi
CD_SEXE	Sexe du client
dtnai	Date de naissance
situ_fam	Situation familiale
dt_sit_famil	Date de mise à jour de la situation familiale
CD_POSTAL	Code postal domicile
DEP_CD_POSTAL_v2	Département de domicile
MONTANT_GARANTIE	Montant garantie
prime_TTC_an_theo	Prime TTC par an à payer
frac	Fractionnement du paiement de la prime (mensuel, bimensuel, quadrimestriel, annuel)
formule	Formule souscrite (dépendance totale ou dépendance totale + partielle)
taux_tech	Taux technique
option_acci	Souscription de la garantie option accidents
LIB_TYPE_VENTE	Souscription simple ou avec conjoint
date_red2	Date de la réduction, le cas échéant
top_red	Indicateur = 1 si le contrat a été réduit
date_surv_sin_deces	Date de la survenance du décès
date_1er_reglement_sin_deces	Date du premier règlement du sinistre décès
mnt_global_prest_DC	Montant total de la prestation décès
date_rente_dep_debut_couv	Date de début de la rente de dépendance

Tableau 47 - Titre : description de la base de données des résiliations et réductions Protectys

5.3.2. Pré-traitements des données

5.3.2.1. Données aberrantes

Pour chaque type de données, nous avons vérifié par des analyses de données réalisées grâce à des tableaux croisés dynamiques ou des filtres, s'il existait des données aberrantes.

Nous avons trouvé les incohérences suivantes :

- Des dates de sorties avant 2004, ce qui n'est pas possible car le produit a été commercialisé en 2004 (3 lignes) ;
- Des dates de décès avant 2004, ce qui n'est pas possible car le produit a été commercialisé en 2004 (4 lignes) ;
- Des dates de décès antérieures à des dates de souscription (4 lignes dont deux sont comprises dans l'incohérence citée juste avant) ;
- Des dates de souscription après 2014, ce qui n'est pas possible car la commercialisation du produit a été arrêtée en 2014 (5 lignes) ;
- Des réductions avant 8 ans d'ancienneté du contrat (14 lignes) ;
- Des primes TTC après réduction non égale à 0 (32 lignes).

Le reste des types de données ne présente a priori pas d'incohérence. Au total, cela concernait une cinquantaine de lignes. On a soit supprimé les lignes aberrantes, soit remplacé les valeurs aberrantes.

5.3.2.2. Ajouts de variables

A des fins d'analyse, trois variables ont été ajoutées à celles initialement présentes dans la base. Elles ont été construites par combinaison des variables préexistantes.

La « durée entre la souscription et la résiliation », la « durée entre la souscription et la réduction » et l'« ancienneté du contrat au moment de la sortie », ont ainsi été créées afin d'avoir une analyse plus complète des données. En particulier, les résiliations et réductions peuvent être analysés selon une variable « taux de l'occurrence » ou une variable « durée avant occurrence ».

Deux variables de taux d'occurrence sont ainsi analysées :

$$\frac{\text{nombre de résiliations ou réduction par année en cours}}{\text{Stock à date}} ;$$

$$\frac{\text{nombre de réductions par année de production}}{\text{Production de l'année concernée}} .$$

Une variable liée à la durée avant occurrence est également analysée : évolution des délais de résiliation et réduction pour chaque année de production.

5.3.2.3. Traitements des données manquantes et autres traitements

Les données manquantes ont été traitées dans Python. Le code correspondant est présenté en Annexe 8. Des bibliothèques existantes ont été utilisées et sont identifiées dans la bibliographie du présent mémoire.

Tout d'abord, la base de données, après traitement des données aberrantes et création de nouvelles variables, a été importée dans Anaconda.

On a dans un premier temps vérifié la présence et supprimé les éventuels doublons. Une seule ligne était concernée.

On supprime ensuite les variables qui ne nous intéressent pas pour l'étude des réductions : NUM_CONTRAT, ID_PERSONNE, num_sirocco, CD_PRODUI, produit, date_premier_anniv et dt_sit_famil, date_1er_reglement_sin_deces, mnt_global_prest_DC, date_rente_dep_debut_couv.

Les variables date_effet_entree, CD_SOR_CONTRAT, DEP_CD_POSTAL_v2, date_effet_sortie, dtdeces, LIB_SORT_CONTRAT, dtresil, Durée entre la souscription et la résiliation, Durée entre la souscription et la réduction et date_red2, date_surv_sin_deces, sont supprimées également car redondantes avec d'autres variables ou comprises dans les variables créées.

Les variables MONTANT_GARANTIE et prime_TTC_an_theo sont également supprimées car les montants concernés sont après réduction donc non pertinents pour l'analyse.

Dans un deuxième temps, on a identifié les variables restantes contenant des données manquantes. Il en ressort le résultat suivant :

Libellé donnée	Présence de données manquantes	Nombre de données manquantes
canal_vente	True	272 065
date_sou	False	0
CD_SEXE	False	0
dtnai	False	0
situ_fam	True	81 621
CD_POSTAL	True	81 010
frac	False	0
formule	True	528
taux_tech	True	272 070
option_acci	True	272 070
LIB_TYPE_VENTE	True	81 008
top_red	True	259 958

Tableau 48 - Titre : nombre de données manquantes selon différentes variables de la base de données résiliations et réductions Protectys

On supprime les variables contenant plus de 99% de données manquantes.

Les variables restantes contiennent pour certaines des données manquantes, ce qui est normal car tous les assurés ne sont pas sortis, décédés, n'ont pas résilié, réduit ou ne touchent pas une rente.

En revanche, certaines données sont vraiment manquantes, concernant les colonnes situ_fam, CD_POSTAL, LIB_TYPE_VENTE et formule.

Pour ces données manquantes, on remplace les valeurs manquantes par les valeurs des lignes précédentes. On aurait pu utiliser des méthodes plus sophistiquées, comme les plus proches voisins ou la médiane ; mais par un souci de simplification et pour limiter le risque d'erreurs, on utilise la méthode susmentionnée.

Enfin, une fois les données manquantes traitées, on convertit les données restantes dont les modalités sont sous forme de chaînes de caractères en chiffres. Cela, afin de pouvoir les exploiter par la suite.

5.3.2.4. Contrôles de cohérence

Afin de vérifier la qualité et la bonne représentativité des données étudiées, une validation a été réalisée sur le nombre de contrats en cours et sur les sorties dans la base étudiée par rapport au nombre de contrats en cours et sorties issus des reportings de 2016, 2017 et 2018 de l'Actuariat Inventaire (cf. annexe 7 pour détail).

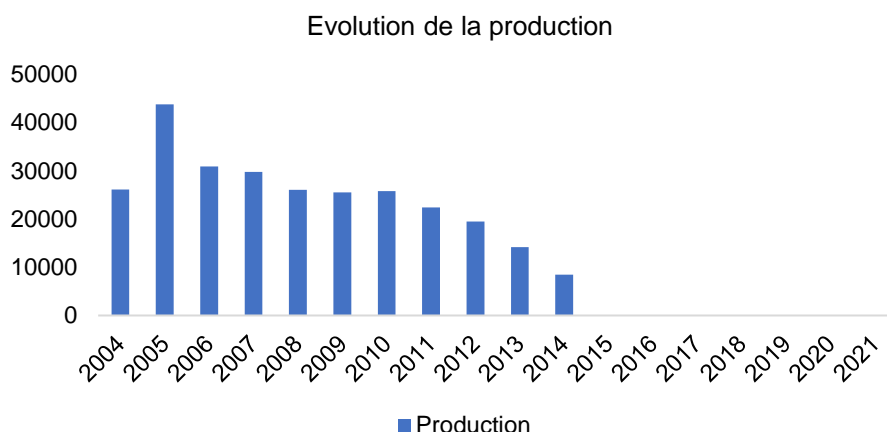
	2016	2017	2018
Écarts sur le nombre de contrats en cours	0,1%	0,1%	0,1%
Écarts sur le nombre de résiliations	0,1%	0,1%	0,2%
Écarts sur le nombre de résiliations	0,3%	0,2%	0,1%

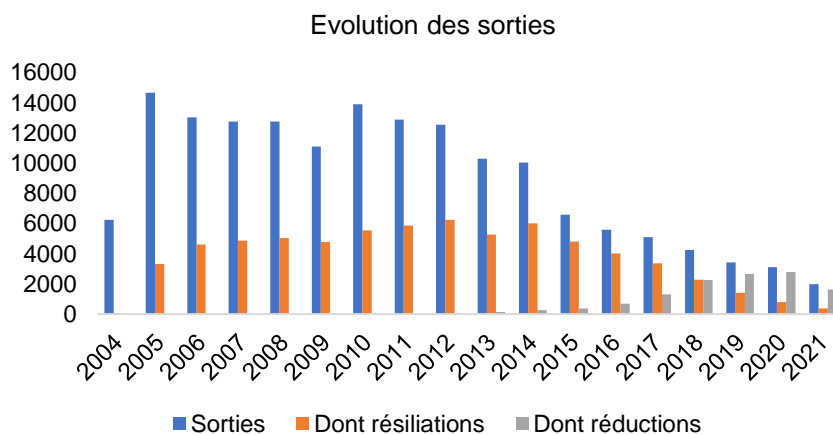
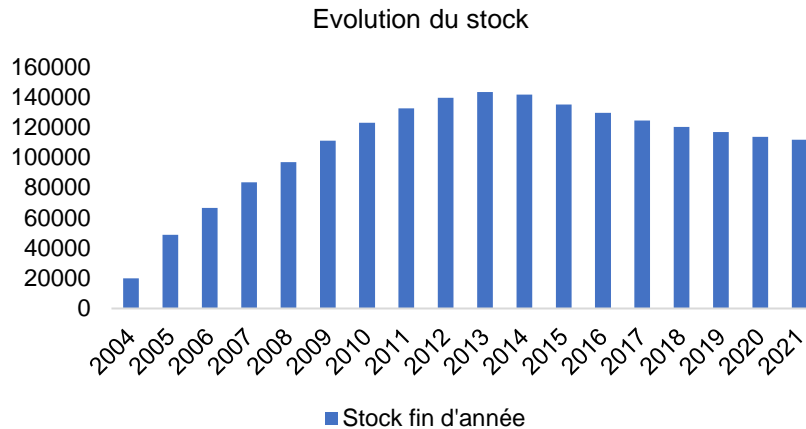
Tableau 49 - Titre : contrôles des écarts entre la base de données des résiliations réductions Protectys et les reportings de l'actuariat inventaire

Les résultats sont convenables. Les écarts peuvent être expliqués par des dates d'extraction différentes. Evidemment, d'autres contrôles auraient été possible, par exemple sur d'autres périodes ou avec un focus sur les réductions. Un seuil de matérialité pourrait également être déterminé.

5.3.3. Statistiques descriptives

Le produit Protectys a été commercialisé à partir de 2004 jusqu'à fin 2014. Ci-après sont représentées les évolutions de sa production, de son stock, des résiliations et des réductions (cf. annexe 9 pour détail des chiffres clés).





Figures 45, 46 et 47 - Titre : évolution de la production, du stock et des sorties du produit Protectys

Comme on peut l'observer, la production de ce produit a été arrêtée en 2014. Le stock s'élevait alors à c. 140 000 contrats. Ensuite, il est en légère diminution du fait des sorties pour sinistres, pour résiliations et pour réductions.

En début de période d'observation, notamment en 2005, le nombre de sorties par rapport au stock est très important du fait d'un nombre de sorties pour renoncements, impayés première prime et sans suite très significatif comparé aux périodes postérieures.

Le nombre de résiliations est en hausse de 2005 à 2012, se stabilise ensuite puis diminue à partir des 2015, ce qui est logique car les résiliations sont progressivement remplacées par des réductions. Ainsi, à partir de 2015, les réductions sont en augmentation.

5.4. Analyse univariée

L'objectif global de l'analyse de données est de déterminer, à partir des données disponibles, les caractéristiques des assurés réalisant des réductions. Pour ce faire, nous avons tout d'abord réalisé une analyse univariée des variables caractérisant ces assurés en considérant soit les taux de réduction, soit le délai entre la souscription et la réduction. Nous réaliserons dans un second temps une application de méthodes de machine learning de prédiction qui nous permettront d'infirmer ou confirmer l'analyse univariée en identifiant les variables les plus discriminantes vis-à-vis du taux de réduction.

A titre de comparaison et de compréhension du portefeuille, nous avons également réalisé une partie des analyses, notamment celles en fonction du temps, sur les résiliations.

5.4.1. Analyse univariée en fonction du temps

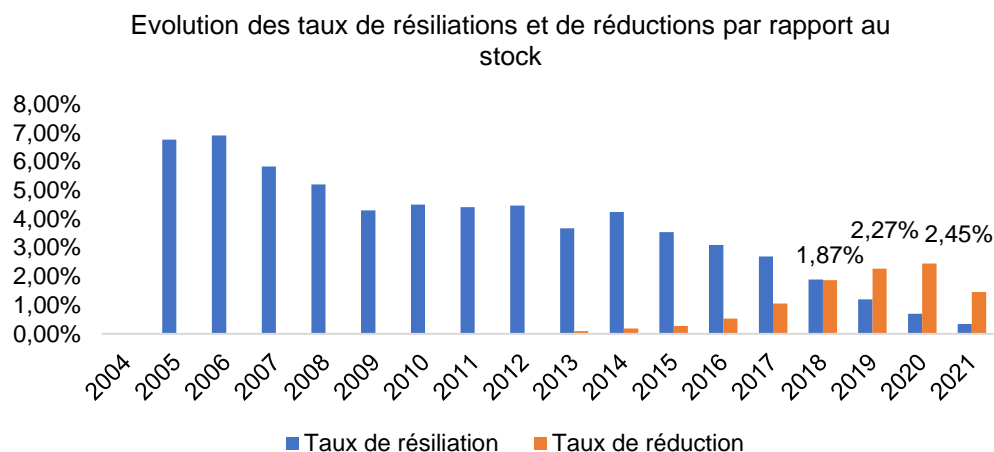


Figure 48 - Titre : Evolution des taux de résiliations du produit Protectys et de réductions par rapport au stock

Les taux de résiliations par rapport au stock sont en diminution sur l'ensemble de la période observée, du fait d'un stock en hausse. Les taux de réduction augmentent à partir de 2013 jusqu'en 2021, ce qui est logique car les réductions remplacent peu à peu les résiliations.

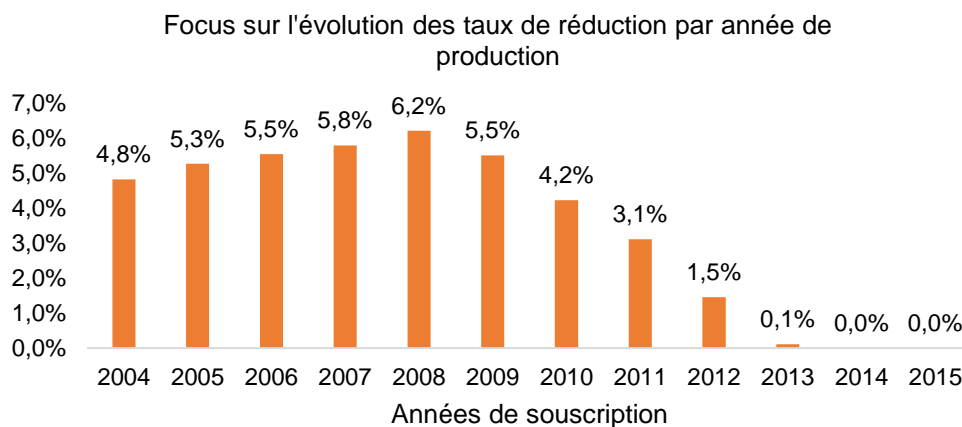
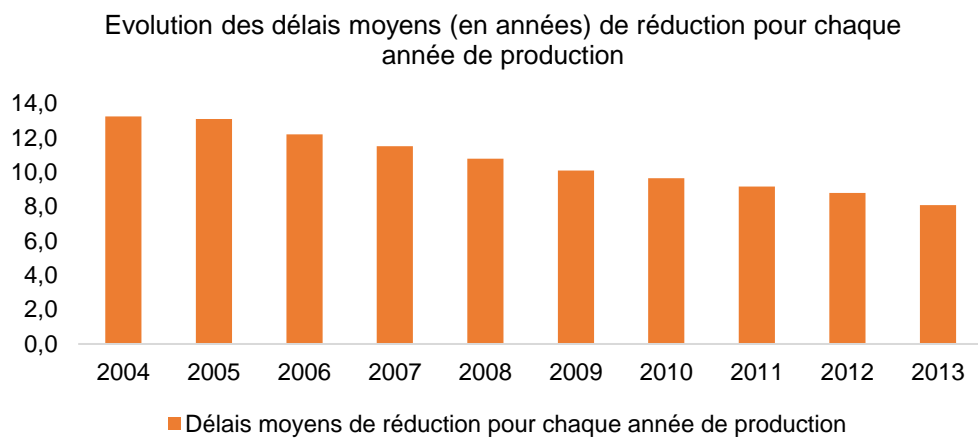
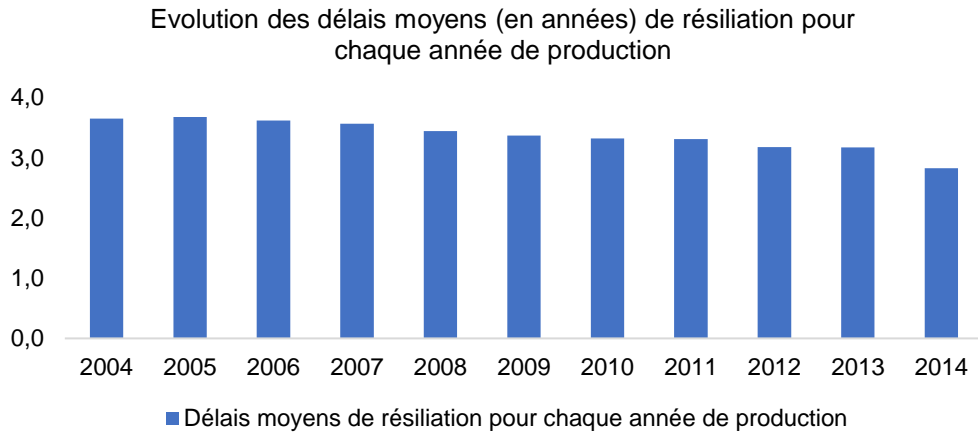


Figure 49 - Titre : Focus sur l'évolution des taux de réduction du produit Protectys par année de production

Si on se focalise sur les réductions, le taux de réduction par année de production augmente entre les années de souscription 2004 à 2008, cohorte qui atteint un pic, puis diminue ensuite, ce qui répond au mécanisme de réduction possible à partir de 8 ans d'ancienneté des contrats.



Figures 50 et 51 - Titre : Evolution des délais moyens (en années) de résiliation et réduction du produit Protectys pour chaque année de production

Si on observe les délais moyens de réduction, ceux-ci sont en constante diminution entre 2004 et 2014. Comme cela répond au mécanisme de réduction possible à partir de 8 ans d'ancienneté et aux différentes durées d'exposition au risque selon les années de souscription (les souscriptions de 2009 ont été moins exposées à l'occurrence du risque que les souscriptions de 2004), il n'est pas possible de statuer, avec ces informations, sur une potentielle augmentation du risque.

5.4.2. Analyse univariée en fonction des autres variables

Nous nous focaliserons ici sur les taux de réduction. Les taux moyens peuvent varier car nous avons travaillé, pour l'analyse univariée, après retraitement des données aberrantes mais sans remplacer les données manquantes afin de ne pas biaiser l'analyse. L'objectif étant d'identifier les variables discriminantes, nous avons estimé que cela ne posait pas de problème pour l'analyse.

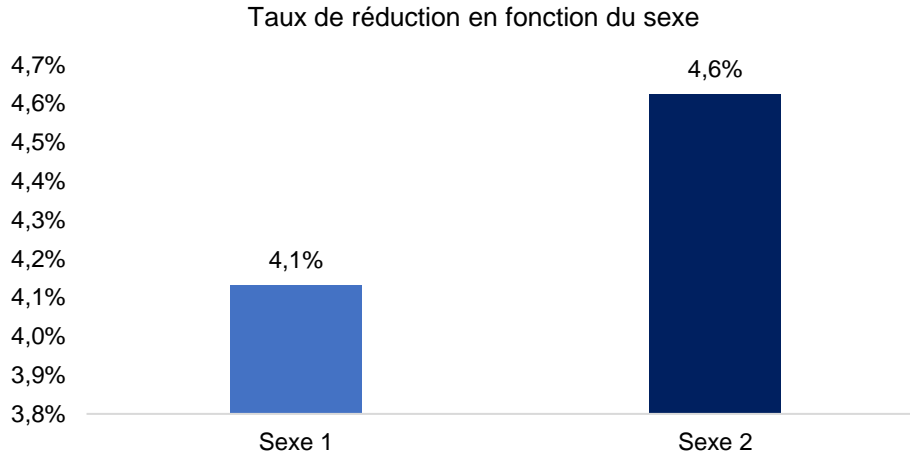


Figure 52 - Titre : Taux de réduction du produit Protectys en fonction du sexe

Le sexe 2 induit un plus fort taux de réduction. Cependant, l'écart type entre les deux catégories, de 0,35%, peut être considéré comme faible.

Nous avons ensuite étudié la corrélation entre l'année de naissance et le taux de réduction. Cela a nécessité un travail sur les données dont le nombre, pour certaines catégories, n'était pas forcément suffisant pour être représentatif. Nous avons sélectionné les catégories pour lesquelles le nombre d'observations était supérieur à 1000.

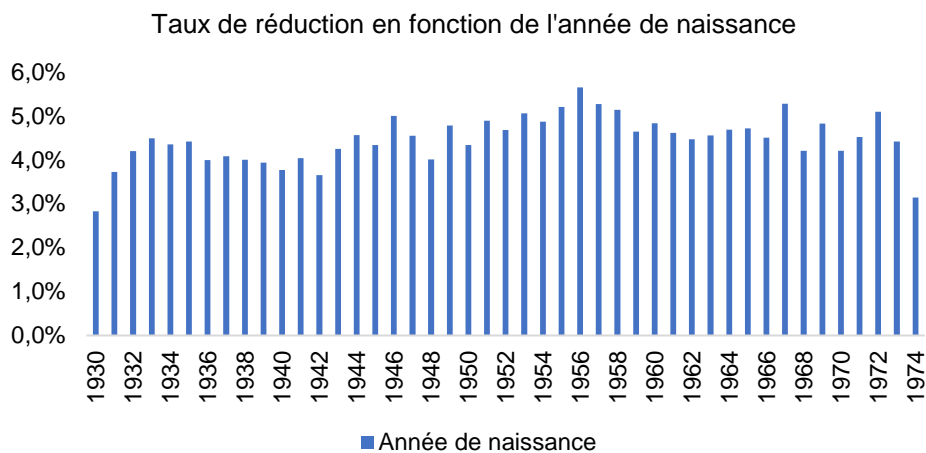


Figure 53 - Titre : Taux de réduction du produit Protectys en fonction de l'année de naissance

Le taux de réduction varie en fonction de l'année de naissance et suit une courbe sinusoïdale. Ainsi, on se demande si on peut le considérer comme globalement stable ou si au contraire, l'année de naissance est déterminante.

Pour compléter l'analyse, nous avons étudié le nuage de point entre le nombre de réduction et le nombre de contrats par année de production, ainsi que la pente de la régression linéaire correspondante.

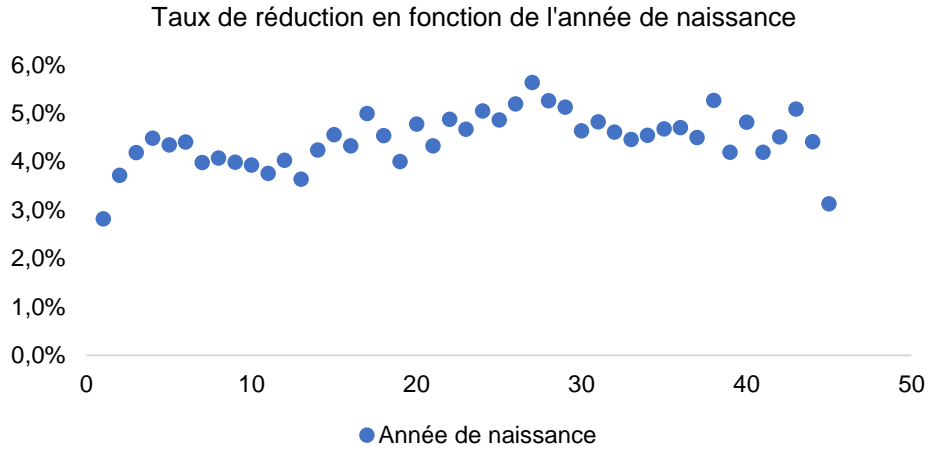


Figure 54 - Titre : nuage de points du produit Protectys du taux de réduction en fonction de l'année de naissance

La pente de la régression linéaire correspondante est faible, à c. 4%. Nous aurons besoin de compléter cette analyse en appliquant des techniques de machine learning de prédiction qui nous permettront de confirmer les variables discriminantes ou non.

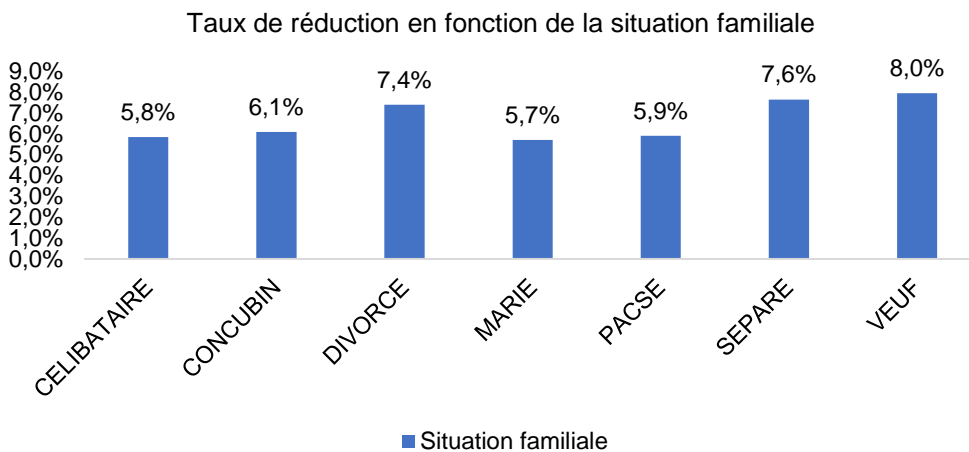


Figure 55 - Titre : Taux de réduction du produit Protectys en fonction de la situation familiale

On observe que la situation familiale est corrélée au taux de réduction. Les assurés divorcés, séparés ou veufs induisant des taux plus élevés.

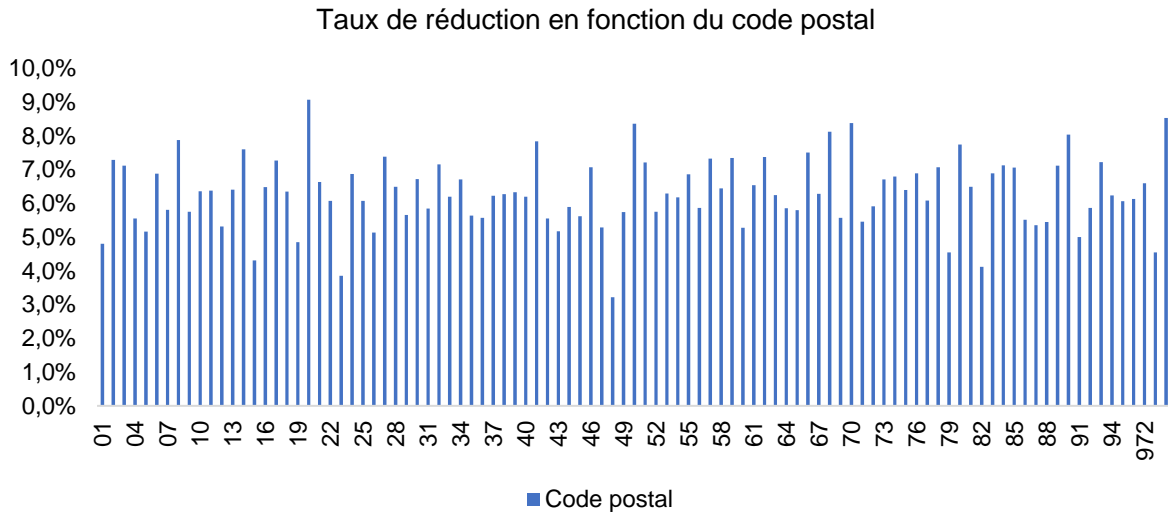


Figure 56 - Titre : Taux de réduction du produit Protectys en fonction du code postal

Le code postal semble avoir un impact sur le taux de réduction.

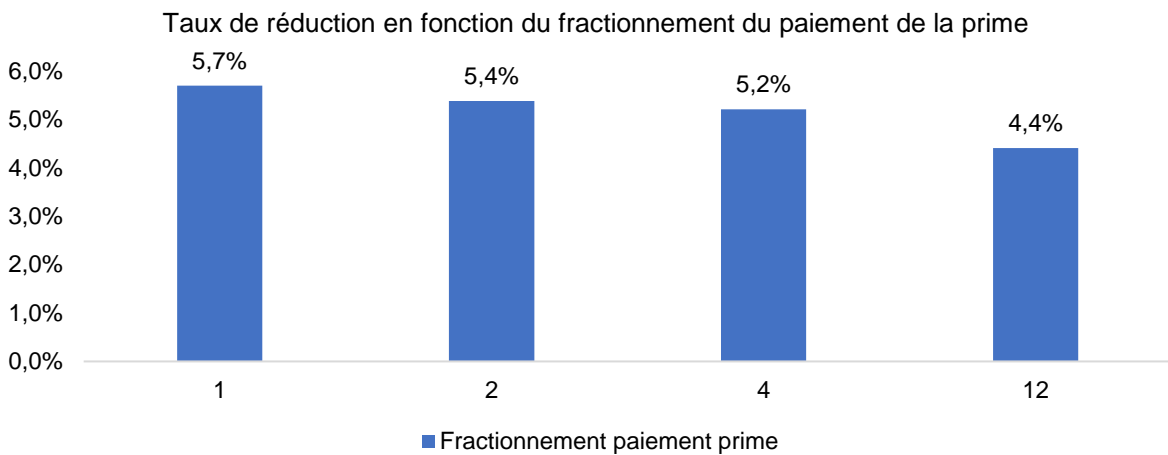


Figure 57 - Titre : Taux de réduction en fonction du fractionnement du paiement de la prime

La typologie de fractionnement du paiement de la prime a une incidence sur le taux de réduction. Ainsi, plus le paiement est fractionné, plus, le taux est élevé.

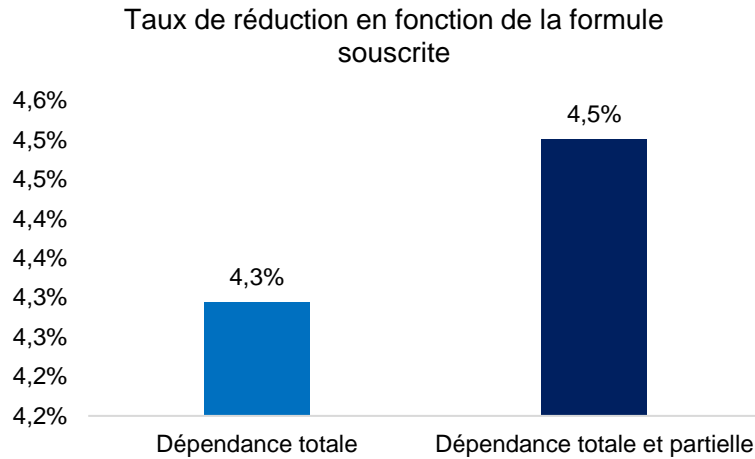


Figure 58 - Titre : Taux de réduction en fonction de la formule souscrite

La typologie de formule souscrite a un faible impact sur le taux de réduction. On peut noter tout de même que les formules les plus fournies sont celles qui disposent d'un taux plus élevé.

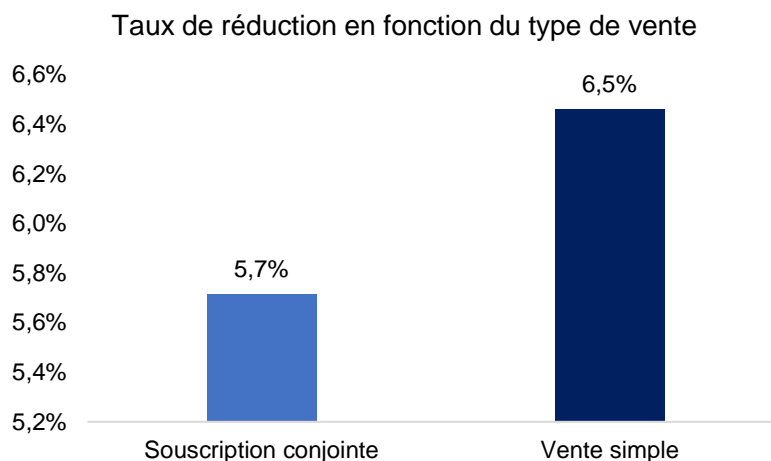


Figure 59 - Titre : Taux de réduction en fonction du type de vente

Le type de vente a un impact sur le taux de réduction, avec un taux de réduction plus important pour les ventes simples.

N.B. : Les variables garanties et taux technique ne sont pas exploitables car correspondent à des montants post-réduction donc à des garanties réduites et à des taux à 0%.

5.4.3. Conclusion de l'analyse univariée

L'analyse univariée a permis d'identifier des variables corrélées au taux de réduction.

Variables	Niveau de discrimination par rapport au taux de réduction	Ecart-type des taux de réduction	Catégorie induisant le taux de réduction le plus élevé
Année de souscription		2,46%	2008
Sexe		0,35%	2
Année de naissance	?	NA	NA
Situation familiale		0,97%	Divorcé, séparé ou veuf
Code postal		1,27%	20
Fractionnement prime		0,55%	1 (mensuel)
Formule souscrite		0,15%	Dépendance totale et partielle
Type de vente		0,53%	Vente simple

Légende :




	Forte discrimination
	Discrimination moyenne
	Faible discrimination

Tableau 50 - Titre : conclusions de l'analyse univariée : niveaux de discrimination par rapport au taux de réduction

On identifie grâce à l'analyse univariée des variables ayant une corrélation positive significative avec le taux de réduction. Cette analyse comporte des limites qui sont explicitées ci-dessous après l'application de méthodes de prédiction.

Elle permet d'identifier parmi les modalités de chaque variable, celles qui sont déterminantes. Cependant, afin d'avoir une analyse complète et plus solide, il est nécessaire de la compléter par des applications de méthodes de prédiction qui permettront d'identifier les variables les plus discriminantes.

5.5. Application de méthodes de machines learning de prédiction

5.5.1. Présentation des méthodes

Deux méthodes de machine learning vont être étudiés ici : Classification and Regression Tree (CART) et les forêts aléatoires ou random forest.

D'après les cours de Sciences des données d'Olivier Lopez dispensé dans le cadre du Certificat d'Expertise actuarielle (2021), Classification and Regression Tree est une méthode non paramétrique permettant de faire de la classification binaire en séparant plusieurs fois l'espace en 2, souvent par une minimisation de l'indice de Gini. L'indice de Gini se calcule en divisant, sur la partie positif du graphe, l'aire entre la bissectrice d'un repère orthonormée et la courbe représentant la répartition réelle de la variable, puis en divisant cette aire par l'aire entre la bissectrice et l'axe des abscisses. Plus l'indice est proche de 0, plus la répartition de la variable dans la population est égale, et inversement. Une minimisation de l'indice de Gini

correspond à une volonté d'avoir une répartition la plus égale possible d'une variable au sein d'une population, afin de séparer les groupes efficacement, cf. illustrations ci-dessous. Cela permet d'avoir aussi des résultats interprétables et exploitables, par exemple pour un questionnaire d'assurance.

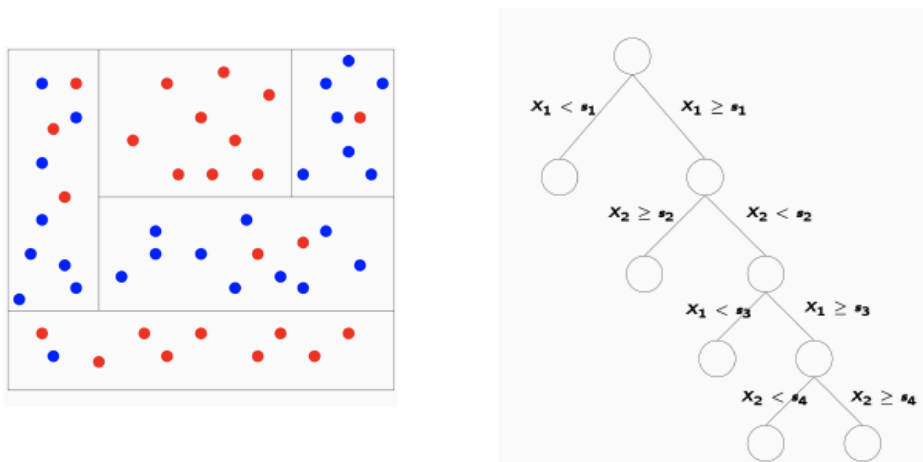


Figure 60 - Titre : illustrations de la méthode CART, source : cours de Sciences des données d'Olivier Lopez dispensé dans le cadre du Certificat d'Expertise actuarielle (2021)

Le nombre de nœuds maximum est un hyper paramètre à déterminer. Nous avons choisi par défaut celui proposé par la librairie Python utilisée. Les autres paramètres choisis sont également ceux proposés par défaut⁶.

La méthode des Forêts aléatoires combine les arbres de classification, le « bagging » (utilisation de plusieurs arbres ce qui permet de réduire le bruit et la variance d'un unique arbre) et le « Columns subsampling » (utilisation, à chaque fois, de différents jeux de caractéristiques aléatoires pour les séparations des données, ce qui permet de réduire la corrélation entre les arbres produits).

Le nombre d'arbres est un hyper paramètre à déterminer. Nous avons choisi par défaut celui proposé par la librairie (100). Les autres paramètres choisis sont également ceux proposés par défaut⁷.

5.5.2. Application

5.5.3. Pré-traitements des données

Les pré-traitements réalisés sont ceux présentés ci-dessus, juste avant l'analyse univariée.

5.5.4. Résultats

Le code Python correspondant est présenté en annexe 8. Des librairies existantes ont été utilisées et sont identifiées dans la bibliographie du présent mémoire.

Le modèle CART entraîné dispose d'une précision de 83% sur l'échantillon de test. C'est-à-dire que dans 83% des cas, il réussit à prédire de manière correcte la réduction ou non la réduction. C'est un bon score.

Les variables les plus discriminantes d'après le modèle sont les suivantes.

⁶ sklearn.tree.DecisionTreeClassifier — scikit-learn 0.24.2 documentation

⁷ sklearn.ensemble.RandomForestClassifier — scikit-learn 0.24.2 documentation

Variables	Importance sur une échelle de 1
dtnai	0,29
CD_POSTAL	0,28
date_sou	0,24
Ancienneté contrat au moment de la sortie	0,09
situ_fam	0,04
CD_SEXE	0,02
formule	0,02
frac	0,01
LIB_TYPE_VENTE	0,01

Légende :

 Forte discrimination

Tableau 51 - Titre : conclusions de l'application du modèle CART : niveaux de discrimination par rapport au taux de réduction

Le modèle des forêts aléatoires entraîné dispose d'une précision supérieure à 87%.

Les variables les plus discriminantes d'après ce modèle sont les suivantes.

Variables	Importance sur une échelle de 1
dtnai	0,29
CD_POSTAL	0,28
date_sou	0,26
Ancienneté contrat au moment de la sortie	0,10
situ_fam	0,03
CD_SEXE	0,02
formule	0,01
frac	0,01
LIB_TYPE_VENTE	0,01

Légende :

 Forte discrimination

Tableau 52 - Titre : conclusions de l'application du modèle des forêts aléatoires : niveaux de discrimination par rapport au taux de réduction

On retrouve les mêmes résultats entre les deux modèles de prédictions. Les résultats sont aussi assez cohérents avec les résultats de l'analyse univariée : on retrouve les mêmes variables les plus discriminantes vis-à-vis du taux de réduction. En particulier, nous avons un doute sur la variable date de naissance qui s'avère être une variable discriminante.

5.6. Conclusions et limites de la troisième approche

Pour rappel, l'objectif de ce chapitre est la mise en œuvre d'un plan d'action afin de réduire le risque de modèle lié à une sous-estimation du best estimate du produit dépendance Protectys. Les réductions ont un impact positif sur le best estimate. Ainsi, réduire dans la réalité les réductions permet de réduire le risque de sous-estimation du best estimate. Dans ce cadre, nous avons cherché à déterminer, par analyses de données et méthode de prédiction, les caractéristiques des assurés réalisant des réductions, afin de mettre en place des actions, commerciales par exemple, pour réduire leur nombre. Cela permettra, d'une part, de réduire le risque de modèle, mais également de mieux connaître la population assurée.

Les deux types d'analyses, univariée et par modèles de prédiction, sont complémentaires. Elles permettent d'identifier les variables les plus discriminantes vis-à-vis du taux de réduction et les modalités correspondantes. La date de naissance, le lieu d'habitation et la date de souscription sont ainsi les variables qui ressortent comme les plus discriminantes. Des actions pourront être réalisées auprès des assurés concernés afin de limiter, si souhaité, les réductions et réduire le risque de sous-estimation du best estimate.

Cette analyse comporte plusieurs limites.

Premièrement, les données utilisées sont limitées. Des données complémentaires auraient pu être intéressantes, par exemple : CSP, autres produits détenus, ancienneté client au niveau de LBP, tarification, actions commerciales, niveau de revenus, situation familiale, note LBP de risque du client.

Deuxièmement, pour les analyses univariées, nous avons réalisé l'analyse par rapport au taux de réduction. Nous aurions pu également nous intéresser à la durée entre la souscription et la réduction en fonction des différentes catégories.

Quatrièmement, au niveau des contrôles sur la base de données, d'autres contrôles auraient été possible, par exemple sur d'autres période ou avec un focus sur les réductions. Un seuil de matérialité devrait également être déterminé. Nous aurions pu également faire des tests de représentativité au niveau des données disponibles, afin de nous assurer que leur nombre soit suffisant. Enfin, on aurait pu utiliser des méthodes plus sophistiquées pour remplacer les données manquantes, comme les plus proches voisins ou la médiane.

Enfin, au niveau des applications de modèles de prédiction, nous avons utilisé deux méthodes. D'autres méthodes existent et auraient pu être étudiées. De plus, nous avons utilisé, pour les deux méthodes étudiées, les hyper paramètres par défaut, des travaux plus précis auraient pu être réalisés.

Ces analyses pourront ainsi être complétées après la rédaction de ce mémoire.

6. Conclusion

6.1. Rappel du contexte

L'activité d'assurance est soumise au risque du modèle. Le risque de modèle peut être « considéré comme [une erreur] susceptible d'influer sur la prise de décision ou le jugement des utilisateurs de cette information ». Cette erreur peut intervenir à toute étape du cycle de vie d'un modèle.

Dans ce cadre, ce mémoire avait pour objectif d'étudier ce risque au niveau de La Banque Postale Prévoyance, filiale de La Banque Postale spécialisée dans les produits d'assurance prévoyance (notamment garanties décès, dépendance, accident de la vie).

La problématique était la suivante :

Comment évaluer et gérer le risque de modèle au niveau de l'entité La Banque Postale Prévoyance ?

6.2. Approches retenues

Trois approches ont été étudiées dans ce mémoire, une première basée sur les méthodologies ERM (Enterprise Risk Management ou Gestion des Risques) étudiées dans le cadre du Certificat d'Expertise Actuarielle, une deuxième sur la méthode Worst Case, proposée par Kerkhof et al (2002) et une troisième qui présente et utilise différentes méthodes d'analyses (univariées et machine learning de prédiction) appliquées afin de déterminer des profils de réduction ; méthodes inspirées des cours d'O. Lopez donnés dans le cadre du Certificat d'Expertise Actuarielle (La Sorbonne Université, 2021).

La première méthode, construite dans le cadre de ce mémoire consiste à (i) cartographier les modèles de La Banque Postale Prévoyance ; (ii) proposer une évaluation du risque brut d'un panel de modèles de La Banque Postale Prévoyance ; (iii) évaluer le dispositif d'atténuation du risque brut puis le risque net de ces modèles ; (iv) conclure sur l'acceptabilité de ce dernier en fonction de l'appétence au risque et mettre en place, si nécessaire, des plans d'action afin de le réduire.

La deuxième méthode consiste à mesurer l'écart entre les résultats d'un modèle utilisé par l'entreprise, le modèle de référence, et les résultats qui seraient induits par l'utilisation d'autres méthodes. Le risque de modèle est alors évalué comme l'écart entre les résultats du modèle de référence et de la méthode qui donne les résultats les plus éloignés du modèle de référence. Cette méthode est appliquée, dans ce mémoire, à un modèle de dépendance de La Banque Postale Prévoyance.

La troisième méthode consiste en l'étude d'un plan d'action de réduction du risque de sous-estimation du best estimate par identification des profils d'assurés réalisant des réductions, qui ont un impact positif sur le best estimate, en utilisant des méthodes d'analyses univariées et de machine learning de prédiction.

6.3. Résultats

Avec la première approche : sur les six exercices de modélisation étudiés, cinq peuvent être considérés comme des modèles au sens de l'arbre de décision retenu. Parmi ces cinq modèles, deux ont un risque brut acceptable (modéré) et trois non acceptable, selon l'appétence au risque prédéfinie. Après évaluation des dispositifs de maîtrise des risques relatifs à ces modèles, un seul modèle a un risque net majeur. Grâce à la méthodologie d'évaluation des dispositifs de maîtrise des risques, il est aisé de trouver un plan d'action pour le réduire, par exemple, désigner un responsable et un propriétaire du modèle et revoir formellement les hypothèses de ce dernier.

La première approche permet d'évaluer et de suivre le risque de plusieurs modèles et de mettre à jour cette évaluation à fréquence régulière. Elle permet également de mettre facilement les modèles en comparaison les uns par rapport aux autres et de suivre l'évolution de cette hiérarchie.

Concernant la deuxième approche : sur les six modèles de comparaison étudiés, les variations maximales des résultats par rapport au modèle de référence sont de - 5,7%. Ces variations vont dans le sens d'une réduction du best estimate. Ainsi, le modèle de référence peut être considéré comme prudent au sens de la définition retenue et il n'est pas nécessaire de mettre en place de plan d'action. Le cas échéant, un plan d'action possible aurait pu consister à retenir les nouvelles lois d'expérience, pouvant, en fonction du nombre de données disponibles, être plus proches de la réalité du portefeuille que le modèle de référence qui se base en partie sur des données d'autres portefeuilles.

La deuxième approche nécessite d'implémenter des modèles de comparaison par rapport au modèle de référence étudié. Ici, des lois d'expérience sont étudiées pour chaque niveau de dépendance (GIR 12, GIR 123 et GIR 1234) et selon le sexe. Elles sont ensuite introduites dans un modèle de projection des flux qui permet d'obtenir le best estimate.

La troisième approche combine deux méthodes complémentaires : analyses univariées et modèles de machine learning de prédiction. Elles permettent d'identifier les variables les plus discriminantes vis-à-vis du taux de réduction et les modalités correspondantes. La date de naissance, le lieu d'habitation et la date de souscription sont ainsi les variables qui ressortent comme les plus discriminantes. Des actions pourront être réalisées auprès des assurés concernés afin de limiter les réductions et réduire le risque de sous-estimation du best estimate.

6.4. Avantages et limites des méthodes étudiées dans ce mémoire

Les trois approches étudiées ont des avantages et des inconvénients.

Elles permettent d'étudier le risque de modèle et de l'estimer en le mettant au regard d'une appétence au risque préalablement définie et de mettre en place des plans d'actions si nécessaire.

La première approche dispose des avantages suivants. Même si la méthodologie de construction de cette dernière est chronophage, elle permet, une fois celle-ci comprise, de facilement l'implémenter et de la mettre à jour, potentiellement par différents utilisateurs. De plus, elle permet de hiérarchiser les modèles les uns par rapport aux autres et de suivre l'évaluation du risque à fréquence régulière. En cas de non-respect de l'appétence au risque,

il est simple de définir des plans d'actions à mettre en place, grâce à la méthodologie d'évaluation des dispositifs de maîtrise des risques réalisée au sein de cette approche.

Toutefois, elle comporte des limites : par exemple, il est nécessaire de déterminer des éléments à dire d'expert tels que l'appétence au risque ou les niveaux d'importance des dispositifs de maîtrise des risques. L'évaluation de ces derniers peut également comporter des biais car, même si des définitions ont été formulées afin de les encadrer, les questions d'évaluation des dispositifs peuvent être sujettes à interprétation. Enfin, la méthode n'est pas aussi précise que l'approche Worst Case.

En synthèse, un des intérêts principaux de cette méthode est de pouvoir suivre l'évolution du risque de modèle d'une année à l'autre et de pouvoir hiérarchiser les modèles les uns par rapport aux autres.

La deuxième approche dispose des avantages suivants. Elle est plus précise que la première méthode. Elle permet d'évaluer, de manière quantitative le risque de modèle et peut permettre des utilisations telle que mesurer une partie de l'écart entre la formule standard du SCR opérationnel et la réalité.

Elle est cependant plus complexe et plus chronophage que la 1^{ère} approche en termes d'implémentation et est soumise aux limites suivantes : (i) il est nécessaire d'élaborer des modèles de comparaison, qui sont eux-mêmes soumis à des biais d'implémentation ; (ii) le nombre de modèles de comparaison est un facteur significatif dans cette méthode, il est en effet nécessaire d'en avoir suffisamment pour pouvoir avoir une analyse étayée. Une des difficultés est de déterminer ce nombre de modèles suffisants ; (iii) un seul modèle peut amener à une évaluation extrême du risque de modèle. Ainsi, l'approche Worst Case peut induire des résultats extrêmes et d'autres approches, telle que l'approche bayésienne, peuvent lui être préférée ; (iv) la comparaison ici a porté sur les lois d'expériences. Il aurait également pu être envisagé de proposer des modèles de comparaison au niveau des autres inputs entrant dans le calcul du best estimate, tels que les rachats ou les charges ; (v) enfin, la mise en place de plans d'actions, le cas échéant, est plus complexe que pour la première approche présentée dans ce mémoire et nécessite des analyses approfondies.

La troisième permet de déterminer, de manière sophistiquée et actuarielle, les caractéristiques des assurés réalisant des réductions et de mettre en place des actions pour réduire ces dernières et donc réduire le risque de sous-estimation du best estimate.

Toutefois, elle comporte des limites : nombre de variables utilisées, fiabilité et traitement des données et méthodes appliquées complexes et non exhaustives.

En synthèse, les trois approches étudiées dans ce mémoire sont complémentaires. La première permet d'évaluer le risque de modèle de manière simple, en grande maille et sur plusieurs modèles. La deuxième est plus précise mais plus chronophage et permet d'évaluer le risque d'un modèle en particulier. La troisième permet l'étude d'un plan d'action visant à réduire le risque de modèle par application d'analyses univariées et méthodes de machine learning de prédiction.

Des études et méthodes de pilotage additionnelles peuvent être utilisées. A titre d'illustration, la revue annuelle des modèles et hypothèses significatifs peut être réalisée ainsi que des études de sensibilité, telles que celles réalisées dans le présent mémoire. Enfin, des indicateurs, tels que le rapport constaté sur attendu, peuvent permettre de suivre le risque de modèle.

7. Bibliographie

ACPR_2015_Notice « Solvabilité II » Évaluation interne des risques et de la solvabilité (ORSA).

BRANGER N._SCHLAG C._2004_Model Risk: A Conceptual Framework for Risk Measurement and Hedging (1-19)_Faculty of Economics and Business Administration_Goethe University.

CADIOT A._date non communiquée_Une approche du risque de modèle lors de la tarification en réassurance non-vie (62-70)_ISFA.

DRAPER D._1995_Assessment and propagation of model uncertainty_Journal of the Royal Statistical Society (46-67)_University of Bath, UK.

DAVESNE C._2015_Etude du risque de modèle dans le cadre d'un modèle interne_ENSAE.

DELOITTE_2018_Model risk management | Building supervisory confidence.

DUBOIS D._2021_Cours de Gestion des risques (ERM)_Formation 2^{nde} année Certificat Expertise Actuarielle_Sorbonne Université, Paris.

DUEE M._REBILLARD C._2006_La dépendance des personnes âgées : une projection en 2040 (1-7)_Donnée sociales - La société française_édition 2006.

DUSSAULE S._2014_L'erreur de modèle dans le dispositif Solvabilité II : problématiques et méthodologies_Mémoire d'actuariat présenté à l'IA_EURIA.

FRANCE ASSUREURS_2021_L'assurance française, données clés 2020 (1-104).

FRANCE ASSUREURS_2021_Le marché des assurances santé et prévoyance en 2020 (1-6).

GEHLER A._2009_ Etude des profils de résiliation sur un portefeuille santé individuelle_Université Louis Pasteur Strasbourg

GORRAND R._2021_Bilan Prudentiel_Support de cours Certificat d'Expertise Actuarielle_Sorbonne Université, Paris.

GRAN THORTON_2016_ORSA requirements: Model risk management for insurance companies.

KERKHOF J._MELENNER B._SCHUMACHER J._2002_Model Risk and Regulatory Capital (Center Discussion Paper; Vol. 2002-27)_Finance.

KERKHOF J. and al_2010_Model Risk and capital reserve_Journal of Banking & Finance.

INSEE_2021_Estimations de population jusqu'en 2021 et projections de population 2021-2070 à partir de 2022.

INSTITUT DES ACTUAIRES_2006_Lignes Directrices Mortalité (29-35).

INSTITUT DES ACTUAIRES_2016_présentation page 7 « Le risque opérationnel, un nouveau challenge pour l'actuaire ».

JOURNAL OFFICIEL LOIS ET DECRETS_2014_Arrêté du 3 novembre 2014 relatif au contrôle interne des banques.

LA BANQUE POSTALE PREVOYANCE_2021_Rapport sur la solvabilité et la situation financière au 31 décembre 2020_Paris.

LALLEMENT T._NISIPASU E._TOPUZU M._ELBAHTOURI L._2014_Le risque de modèle (1-5).

LALLEMENT T._2014_Le risque de modèle_abstract_Mémoire d'actuariat présenté à l'Institut des Actuaire.

LOPEZ O._2021_Cours de Sciences des données donnés dans le cadre du Certificat d'Expertise Actuarielle_La Sorbonne Université

N'DIAYE H._FREZAL S._2016_Les modèles de risque et la prise de décision_Risques n° 105 (141-143).

PAGLIA A. et al_2011_ Tarification des risques en assurance-non-vie, une approche par modèle d'apprentissage statistique_ Euro-Institut d'Actuariat - Groupama

PARLEMENT EUROPEEN_2009_Directive 2009/138/CE du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II) (refonte).

PARLEMENT EUROPEEN. (2014). Règlement délégué (UE) 2015/35 DE LA COMMISSION du 10 octobre 2014 complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II).

QUIBERT Q._PLANCHET F._(2018)_Utilisation des estimateurs de Kaplan-Meier et de Hoem pour la construction de tables de mortalité prospectives (6-10)_ISFA – Laboratoire SAF. Université de Lyon 1_017. hal-01509483.

PLANCHET F._THEROND P.E._2006_Modèles de durée - Applications actuarielles_Economica, pp.250_Assurance Audit Actuariat_hal-00530877.

PLANCHET F._2022_Modèles de durée, méthodes de lissage et d'ajustement (16)_Support de cours 2022-2023_ISFA.

RAHOLA A._2011_Synthèse du débat national sur la dépendance_Ministère des Solidarités et de la cohésion sociale.

RENOUF E._LE RHUN G._2014_Le risque de modèle – Mise en pratique dans un contexte de réassurance-vie_Formation Expert ERM_Institut du risk management.

SAINT-PIERRE P._2021_Introduction à l'analyse des durées de survie_Institut de Mathématiques de Toulouse, Université Paul Sabatier - Toulouse III.

SCIKIT-LEARN.ORG_Librairies python

`sklearn.tree.DecisionTreeClassifier` — scikit-learn 0.24.2 documentation

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

`sklearn.ensemble.RandomForestClassifier` — scikit-learn 0.24.2 documentation

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

STAT4DECISION.COM/FR/FORET-ALEATOIRE-AVEC-PYTHON_Librairies python

SIBBERTSEN P._STAHL G._LUEDTKE C._2008_Measuring model risk_The Journal of Risk Model Validation (65–73)_ Volume 2/Number 4, Winter 2008/09.

TSAGALOS M._2020_Construction de lois d'expérience dépendance_ISFA, Université Lyon 1.

8. Annexes

8.1. Annexe 1 : principales définitions des composantes de l'erreur de modèle

Les principales composantes de l'erreur de modèle ont été définies dans l'étude de Thomas LALLEMENT et al (2014), Le risque de modèle :

« Le risque de spécification »

Le risque de spécification peut se définir comme le risque que le modèle choisi par l'actuaire pour évaluer un risque assurantiel ou financier soit inapproprié. Il y a deux cas de figures dans lesquels on peut considérer un modèle comme inapproprié :

- Si le modèle n'est pas adapté pour le type de risque que l'on souhaite étudier.
- Si le modèle est basé sur des hypothèses qui ne sont pas vérifiées en réalité.

C'est la raison pour laquelle l'actuaire doit se poser trois questions essentielles avant tout choix de modèle :

- Quel risque doit être modélisé ?
- Quelles sont les données à notre disposition ?
- Quels sont les modèles que nous pouvons appliquer pour modéliser ce risque ?

Le risque d'implémentation

Il existe deux formes de risque d'implémentation :

- Le risque qu'un mauvais choix d'algorithme soit effectué pour implémenter les concepts spécifiques au modèle.
- Le risque qu'un algorithme approprié soit choisi, mais qu'il contienne des erreurs de codage.

La plupart du temps, la modélisation d'un risque se fait par l'intermédiaire de logiciel informatique. Beaucoup d'actuares et de gestionnaires de risque implémentent les modèles dans leur département et ne jugent pas nécessaire d'avoir recours à un support informatique (IT) pour le faire.

Parfois, une revue approfondie du code ne peut être effectuée car elle nécessite tout simplement trop de temps. Des erreurs d'implémentation peuvent alors subsister et provoquer des résultats incohérents. De plus, la complexité des techniques de codage employées ne permet parfois pas de tracer rapidement la cause du problème.

Le risque de traitement des données [Conception ou Utilisation dans le schéma présenté en Introduction du présent mémoire]

Une mauvaise compréhension et utilisation des données mises à la disposition des actuaires peuvent conduire à divers biais que ce soit dans le choix des modèles ou dans l'estimation de leur paramètre. Il est donc souvent nécessaire d'effectuer divers choix ou retraitements sur les données :

- Choix de l'historique de données : ce choix doit se faire de manière précise et adaptée. Souvent, certaines périodes historiques ne sont plus représentatives de notre futur.
- Retraitement des données : suppression d'évènements exceptionnels si nécessaire, correction d'erreur de saisie, lissage, etc.

Le risque d'estimation [Conception ou Utilisation dans le schéma présenté en Introduction du présent mémoire]

Le risque d'estimation intervient lors du calibrage d'un modèle, méthode qui consiste à ajuster les caractéristiques du modèle à des données disponibles et/ou à certaines contraintes imposées a priori. L'estimation des paramètres du modèle dépend alors de ces données initiales et des méthodes statistiques utilisées. D'une part, les résultats obtenus par ces méthodes se voient affectés d'une erreur d'échantillonnage qui induit un écart systématique par rapport à la "vraie" valeur des paramètres. D'autre part, ces méthodes de calibrage donnent parfois des valeurs très différentes pour les paramètres du modèle et il est difficile alors de définir laquelle sera la plus appropriée.

Le risque d'application [Conception ou Utilisation dans le schéma présenté en Introduction du présent mémoire]

Le risque d'application peut être issu de la complexité des modèles. La plupart de ces modèles sont basés sur des travaux de recherches, et des connaissances techniques et théoriques poussées sont parfois requises pour les comprendre.

Ainsi cette complexité, notamment au niveau des hypothèses du modèle, oblige les actuaires à définir et appliquer des approximations permettant l'utilisation des modèles à partir des données existantes.

Les risques d'implémentation, de traitement des données, d'estimation et d'application ne présentent pas de méthodes concrètes de quantification, et interviennent en général après le choix des modèles. Le choix s'est donc porté sur le risque de spécification qui est en général la principale cause de risque de modèle et qui présente dans la littérature quelques méthodes de quantification. »

8.2. Annexe 2 : grille AGGIR élaborée par l'administration française

Grille AGGIR	
Gir	Degrés de dépendance
Gir 1	- Assuré confiné au lit ou au fauteuil, dont les fonctions mentales sont gravement altérées et qui nécessite une présence indispensable et continue d'intervenants
Gir 2	- Assuré confiné au lit ou au fauteuil, dont les fonctions mentales ne sont pas totalement altérées et dont l'état exige une prise en charge pour la plupart des activités de la vie courante - Ou assuré dont les fonctions mentales sont altérées, mais qui est capable de se déplacer et qui nécessite une surveillance permanente
Gir 3	Assuré ayant conservé son autonomie mentale, partiellement son autonomie locomotrice, mais qui a besoin quotidiennement et plusieurs fois par jour d'une aide pour les soins corporels
Gir 4	- Assuré n'assumant pas seul ses transferts mais qui, une fois levé, peut se déplacer à l'intérieur de son logement, et qui a besoin d'aides pour la toilette et l'habillement - Ou assuré n'ayant pas de problèmes locomoteurs mais qui doit être aidé pour les soins corporels et les repas
Gir 5	Assuré ayant seulement besoin d'une aide ponctuelle pour la toilette, la préparation des repas et le ménage
Gir 6	Assuré encore autonome pour les actes essentiels de la vie courante

Tableau 53 - Titre : grille AGGIR élaborée par l'administration française

8.3. Annexe 3 : définition des rôles

Rôles	Définition
Responsable du modèle	<p>Le responsable du modèle est un membre expérimenté de l'équipe d'utilisateurs. Il supervise les différentes étapes du cycle de vie d'un modèle. Il coordonne et assure le bon développement, la bonne implémentation et le correct usage du modèle. Il assure par ailleurs la communication avec les différents organes de validation et est ainsi garant de la détention des validations nécessaires à l'utilisation du modèle. Le choix du responsable ne peut être générique et doit se faire au cas par cas.</p> <p>Chaque modèle doit posséder un responsable, les rôles de ce dernier étant d'assurer les points suivants :</p> <ul style="list-style-type: none"> ▪ La déclaration du modèle ; ▪ L'approbation des prérequis nécessaires à la mise en place du modèle ; ▪ La correcte mise en place des lignes budgétaires nécessaires à la mise en place du modèle ; ▪ Les autorisations d'accès aux systèmes et aux données nécessaires à la mise en place du modèle ; ▪ Le modèle est conçu pour des utilisations et des utilisateurs clairement identifiés tout au long de la vie du modèle ; ▪ Les ressources affectées à la conception, le développement et l'implémentation doivent posséder l'expérience et les compétences appropriées ; ▪ L'environnement d'implémentation du modèle (incluant les interfaces, les données utilisées et les dépendances) doit être clairement identifié et sécurisé. En particulier, toute restriction doit être précisément documentée, validée et contrôlée ; ▪ Les limitations, les faiblesses, les incertitudes et les imprécisions du modèle sont clairement documentées et communiquées à l'ensemble des utilisateurs ainsi qu'à l'équipe de gestion des risques de modèle ; ▪ Tout changement du modèle ou de l'environnement dans lequel il évolue doit être testé et documenté. L'équipe de gestion des risques de modèle doit être informée ; ▪ Les principaux acteurs doivent être consultés lors des phases de développement et de validation ; ▪ A chaque étape de la vie du modèle, des tests appropriés doivent être conduits afin d'assurer sa pertinence et sa robustesse pour les utilisations qui en découlent ; ▪ La documentation du modèle doit être précise, complète et doit permettre d'assurer des travaux de validation adéquats ; ▪ Le modèle doit posséder les validations nécessaires avant toute mise en production. Cela inclut le déploiement initial et les éventuelles évolutions ; ▪ Le contrôle continu de la performance du modèle doit être assuré et communiqué à l'équipe de gestion des risques de modèle ; ▪ Si un problème est identifié, le responsable doit assurer la correcte remédiation selon le calendrier communiqué à l'équipe de gestion des risques de modèle. <p>Le responsable appartient à la première ligne de défense.</p>
Propriétaire du modèle	<p>Généralement le responsable du modèle et le propriétaire sont confondus. Ils peuvent toutefois être dissociés pour les modèles externes ; dans ce cas le propriétaire se situe au sein de l'entreprise prestataire externe. Le responsable doit alors obtenir auprès du propriétaire externe les éléments permettant d'assurer les points listés ci-dessus.</p>
Développeur du modèle	<p>Le développeur est responsable de l'ensemble des activités de développement du modèle. Cela inclut la méthodologie, la conception mathématique et la construction d'un prototype d'implémentation.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ Le développeur doit avoir les compétences pour concevoir un modèle et en construire un prototype. L'implémentation dans le système cible peut-être à la charge du développeur du modèle ; ▪ Le développeur doit assurer la conception méthodologique basée sur une approche quantitative et/ou des travaux de recherche ; ▪ Le développeur doit assurer que l'approche mathématique théorique est robuste et appropriée (cela inclut les modèles externes) ;

Rôles	Définition
	<ul style="list-style-type: none"> ▪ Le développeur assure que le prototype est fiable et approprié ; ▪ Le développeur propose et réalise un plan de tests permettant de prouver la robustesse et la fiabilité du modèle (cela inclut les modèles externes) ; ▪ Le développeur définit les processus de contrôle continu (cela inclut les modèles externes) ; ▪ Le développeur assure la documentation lors de l'initiation et des éventuelles évolutions du modèle ou de l'environnement. <p>Le développeur appartient à la première ligne de défense.</p>
Implémenteur du modèle	<p>L'implémenteur est responsable de la correcte implémentation du modèle dans le système cible en se basant sur les travaux du développeur du modèle. L'implémenteur et le développeur peuvent, lorsque cela s'y prête, être la même personne. Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ L'implémenteur réalise la mise en place dans le système cible du modèle ; ▪ L'implémenteur réalise les tests nécessaires permettant d'assurer le bon comportement du modèle au sein du système cible ; ▪ Les caractéristiques et les tests d'implémentation sont documentés et fournis au propriétaire du modèle par l'implémenteur ; ▪ L'implémenteur assure que toute évolution ou changement dans le modèle doivent être reflétés dans les systèmes mais également audités, testés et documentés ; ▪ L'implémenteur assure la stabilité de l'environnement d'implémentation. <p>L'implémenteur appartient à la première ligne de défense.</p>
Utilisateur du modèle	<p>L'utilisateur du modèle est responsable de l'utilisation du modèle dans le cadre validé et documenté. Il doit, par conséquent, prendre connaissance de toute éventuelle faiblesse ou limitation du modèle et alerter le propriétaire en cas d'anomalie.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ L'utilisateur doit comprendre le modèle et connaître l'utilisation pour laquelle il a été développé ; ▪ L'utilisateur doit être informé de toute imprécision du modèle ; ▪ L'utilisateur doit échanger avec le développeur du modèle afin d'assurer la cohérence entre les travaux de développement et l'utilisation finale du modèle ; ▪ L'utilisateur doit vérifier la cohérence du modèle au cours de la vie de ce dernier et alerter le propriétaire s'il remarque une anomalie. <p>L'utilisateur appartient à la première ligne de défense.</p>

Rôles	Définition
<p>Valideur du modèle</p>	<p>Le valideur du modèle est responsable de la vérification indépendante et de la validité du modèle proposé par le responsable au vu de l'utilisation visée. ▪ Le valideur assure la validation du modèle pour l'objectif défini et documente précisément ses travaux ;</p> <ul style="list-style-type: none"> ▪ Le valideur assure la revue périodique du modèle afin de s'assurer de la cohérence du modèle dans le temps au vu de son utilisation. La documentation doit par conséquent être mise à jour au regard de ces travaux ; ▪ Le valideur assure que toute déficience constatée est suivie et encadrée ; ▪ Le valideur vérifie la correcte documentation fournie par le propriétaire du modèle ; ▪ Le valideur assure la validation de toute modification du modèle et met à jour en conséquence les documents de validation ; ▪ En cas de constatation d'une anomalie, le valideur a l'autorité pour demander des évolutions du modèle. <p>Le valideur appartient à la deuxième ligne de défense.</p> <p>Ainsi, il assure que les risques, les limitations et les faiblesses sont correctement identifiés et encadrés. Le valideur a ainsi l'autorité et l'indépendance pour remettre en question le modèle ou relever des limites, faiblesses ou incohérences additionnelles. Suivant l'importance de ces éléments, le valideur peut être amené à ne pas valider le modèle.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ Le valideur doit avoir les compétences et les ressources nécessaires permettant de conduire les travaux appropriés sur le modèle qu'il valide ; ▪ Le valideur doit analyser et tester la méthodologie, l'implémentation, la performance et la qualité des données et revoir les tests et analyses conduits par les autres acteurs afin d'évaluer le risque modèle ; <p>Le valideur appartient à la deuxième ligne de défense.</p> <p>L'équipe de validation doit avoir les compétences techniques afin de pouvoir porter une analyse et un regard critique sur le modèle, ses hypothèses et son utilisation.</p> <p>L'équipe de validation est responsable de la validation des éléments suivants :</p> <ul style="list-style-type: none"> ▪ La conception du modèle – théorie, hypothèses et démonstrations ; ▪ La qualité des données utilisées ; ▪ L'implémentation ; ▪ La performance du modèle. <p>À la suite de la validation initiale, l'équipe de validation assure la revue périodique. Celle-ci assure la cohérence du modèle vis-à-vis de son utilisation et de l'évolution du contexte. En cas d'incohérence, le propriétaire du modèle, en accord avec la validation, met en place un plan de remédiation.</p> <p>L'équipe de validation a autorité pour intervenir directement auprès du management opérationnel lorsque cela est nécessaire.</p>
<p>Gestionnaire des risques de modèle</p>	<p>Le gestionnaire des risques de modèle est responsable de l'établissement de l'ensemble des normes du dispositif de gestion des risques de modèle. Cette activité est menée au sein de l'équipe en charge de la validation.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ Mettre en place et valider les principes et les procédures de gestion des risques de modèle ; ▪ Définir les exigences de documentation (documents de conception et de validation) ; ▪ Etablir un dispositif de validation en cohérence avec les principes et procédures ; ▪ Mettre en place l'inventaire des modèles et le tenir à jour. A ce titre, il en est le propriétaire ; ▪ Superviser l'ensemble des activités de validation (programme et son exécution) ; ▪ Approuver la validation finale des modèles ; ▪ Superviser le contrôle continu afin d'identifier toute détérioration de modèle et ainsi veiller à la bonne remédiation lorsque cela est nécessaire ; ▪ Communiquer sur le risque de modèle auprès des instances de gouvernance. <p>Le gestionnaire des risques de modèle appartient à la deuxième ligne de défense.</p>

Rôles	Définition
Cotation du modèle	<p>La cotation de modèle permet d'optimiser la gouvernance des risques de modèle. Elle peut être utilisée afin de prioriser les activités de gestion des risques de modèle ou définir les attentes en matière de documentation de la part des propriétaires des modèles.</p> <p>Le responsable a la charge de la fourniture d'une première cotation du modèle lors de la validation initiale. La validation peut, si elle juge cela nécessaire, modifier cette cotation. La cotation est revue à fréquence annuelle permettant ainsi de prendre en compte les évolutions constatées sur le modèle, son utilisation, sa documentation et la gestion du risque qui lui est associée.</p> <p>La méthodologie de cotation est définie par la deuxième ligne de défense au sein d'une procédure dédiée. Elle peut être adaptée aux spécificités de certains modèles sur proposition de la première ligne en accord avec la deuxième ligne.</p> <p>Conséquence de la cotation Tous les modèles sont soumis au même dispositif de gestion des risques de modèle. Cependant, selon la cotation définie, les éléments suivants peuvent se distinguer :</p> <ul style="list-style-type: none"> ▪ La fréquence et l'ordre de planification des travaux de revue périodique ; ▪ Un niveau de détail plus important est attendu de la part du propriétaire du modèle sur les modèles présentant une cotation élevée ; ▪ L'ampleur des tests et des contrôles réalisés par l'équipe de validation ; ▪ Le degré de remontée hiérarchique pour les modèles de cotation élevée.
Contrôle permanent de 1er niveau	<p>La première ligne de défense est constituée du management opérationnel du modèle qui détient le modèle et mesure son risque. Le propriétaire assure que le modèle est construit de manière cohérente avec ses objectifs. Le management opérationnel assure un contrôle permanent du modèle, identifie et mitige les zones de risque. Le propriétaire coordonne directement les parties prenantes de la première ligne de défense : le développeur du modèle, l'implémenteur du modèle et les utilisateurs du modèle. Le propriétaire du modèle demeure responsable des valeurs de sortie du modèle.</p> <p>Le contrôle de 1er niveau intervient notamment au niveau des étapes suivantes : identification et qualification, documentation et tests, limitation et incertitudes, performance du modèle, ajustements.</p>

Rôles	Définition
<p>Contrôle permanent de 2ème niveau</p>	<p>La deuxième ligne de défense est constituée de l'équipe de validation indépendante et de l'équipe de contrôle permanent. Les première et deuxième lignes doivent être totalement indépendantes. La deuxième ligne doit posséder l'autorité nécessaire pour assurer le suivi des risques de modèle.</p> <p>Le dispositif de contrôle de deuxième niveau s'articule ainsi entre la validation et le contrôle permanent.</p> <p>La validation est chargée de la validation de la conception et de l'implémentation des modèles ainsi que de leur revue périodique.</p> <p>Le contrôle permanent évalue la couverture et l'efficacité des contrôles de premier niveau portant sur le processus de production des données. Pour ce faire, le contrôle permanent peut s'appuyer sur les travaux de la validation.</p> <p>Une réunion annuelle permet à la validation et au contrôle permanent d'échanger sur les différentes anomalies ou faiblesses détectées et, ainsi, coordonner leurs actions. Le contrôle permanent opère également un contrôle de deuxième niveau sur les processus de validation, implémentation et revue des modèles.</p> <p>Autrement dit :</p> <ul style="list-style-type: none"> - L'équipe de validation doit avoir les compétences techniques afin de pouvoir porter une analyse et un regard critique sur le modèle, ses hypothèses et son utilisation. <p>L'équipe de validation est responsable de la validation des éléments suivants :</p> <ul style="list-style-type: none"> ▪ La conception du modèle – théorie, hypothèses et démonstrations ; ▪ La qualité des données utilisées ; ▪ L'implémentation ; ▪ La performance du modèle. <p>À la suite de la validation initiale, l'équipe de validation assure la revue périodique. Celle-ci assure la cohérence du modèle vis-à-vis de son utilisation et de l'évolution du contexte. En cas d'incohérence, le propriétaire du modèle, en accord avec la validation, met en place un plan de remédiation.</p> <p>L'équipe de validation a autorité pour intervenir directement auprès du management opérationnel lorsque cela est nécessaire.</p> <ul style="list-style-type: none"> - Le contrôle permanent est responsable : <ul style="list-style-type: none"> ▪ De l'évaluation de la couverture et de l'efficacité des contrôles de premier niveau portant sur la production des données ; ▪ Du contrôle de deuxième niveau sur les processus d'implémentation, de validation et revue périodique des modèles. <p>Afin de mener à bien ses missions, le contrôle permanent peut s'appuyer sur les travaux réalisés par l'équipe de validation.</p>
<p>Audit interne (3ème ligne de défense)</p>	<p>Cette ligne, composée de l'audit interne, s'assure que les première et deuxième lignes de défense respectent les différents principes et procédures de gestion du risque de modèle. L'objectif est de vérifier la cohérence, l'application et la sécurité de la gouvernance du risque de modèle. Cette ligne dispose d'une indépendance importante et en réfère à la gouvernance.</p> <p>L'audit interne est garant de la bonne gestion des risques de modèle. De ce fait, il possède l'autorité et les compétences pour intervenir tout au long du cycle de vie d'un modèle afin de vérifier les pratiques des différents acteurs.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ Assurer le contrôle interne et la cohérence de la gestion des risques de modèle ; ▪ Vérifier que l'inventaire de modèle est en place et à jour ; ▪ Contrôler la cohérence des méthodes, hypothèses et sources d'information et de données utilisées dans les modèles ; ▪ S'assurer que les principes et procédures sont en accord avec les exigences réglementaires ; ▪ Contrôler les pratiques des première et deuxième lignes de défense ainsi que leur respect des principes et procédures ; ▪ Emettre des recommandations lorsque cela est nécessaire et en assurer le bon traitement ; ▪ Assurer un suivi auprès de la gouvernance de son contrôle.

Rôles	Définition
Instances de gouvernance	<p>La gouvernance des modèles doit être soumise à l'examen des instances de gouvernance. A ce titre, ces dernières doivent être impliquées dans la définition du dispositif de gestion des risques de modèle. Elles sont aussi responsables de la stratégie globale et doivent veiller à la cohérence et l'application du dispositif.</p> <p>Les responsabilités associées sont les suivantes :</p> <ul style="list-style-type: none"> ▪ Définir et suivre le niveau d'appétence ou de tolérance pour le risque opérationnel et notamment pour le risque de modèle ; ▪ Définir le cadre général du dispositif de gestion des risques de modèle ; ▪ Approuver les principes et mener une revue annuelle de la gouvernance ; ▪ Assurer la cohérence du dispositif avec les exigences réglementaires.

Tableau 54 - Titre : définition des rôles évalués dans le cadre de l'approche de scoring

8.4. Annexe 4 : critères de détermination du paramètre de lissage z

Z	Ecart absolu	Ratio « constaté sur attendu »	Critère de régularité d'ordre 2
1	0,0864	101,1%	0,07
2	0,0112	87,0%	0,30
3	0,0075	100,9%	0,49
4	0,0076	101,1%	0,52
5	0,0074	101,1%	0,55
6	0,0066	101,1%	2,49

Tableau 55 - Titre : critères de détermination du paramètre de lissage z (source : TSAGALOS M. (2020). Construction de lois d'expérience dépendance. ISFA, Université Lyon 1. Confidentiel jusqu'en 2022)

8.5. Annexe 5 : tests de validation des méthodes de lissage pour la mortalité des femmes Non GIR12

Critères	Moyenne-mobile ordre 3 (n=1)	Moyenne-mobile ordre 4 (n=2)	Whittaker-Henderson
Ecart en absolu	0,0073	0,0111	0,0096
Ecart quadratique	0,0370	0,1279	0,0868
Ratios « constatés sur attendus »	99,83%	98,80%	100,32%
Critère de régularité 1	0,6512	0,8231	0,7089
Critère de régularité 2	0,760	0,1174	0,0126
p-value Kolmogorov-Smirnov	1	1	1
p-value Chi2	1	1	1

Tableau 56 - Titre : tests de validation des méthodes de lissage pour la mortalité des femmes Non GIR12 (source : TSAGALOS M. (2020). Construction de lois d'expérience dépendance. ISFA, Université Lyon 1. Confidentiel jusqu'en 2022)

8.6. Annexe 6 : Chiffres clés résiliations du produit Protectys

Année de production	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Production	26104	43702	30867	29706	25995	25463	25773	22341	19484	14186	8445	0	0	0	0	0	0	0
Sorties	6236	14666	13033	12762	12753	11113	13909	12890	12542	10291	10035	6588	5597	5109	4260	3434	3102	1987
<i>Dont résiliations</i>	5	3308	4610	4873	5044	4783	5541	5857	6240	5268	6016	4797	4008	3357	2277	1408	797	378
<i>Dont réductions</i>	NA	NA	NA	NA	NA	NA	NA	NA	2	146	269	377	687	1310	2254	2657	2782	1629
Stock fin d'année	19868	48904	66738	83682	96924	111274	123138	132589	139531	143426	141836	135248	129651	124542	120282	116848	113746	111759
Taux de résiliation	0,03%	6,76%	6,91%	5,82%	5,20%	4,30%	4,50%	4,42%	4,47%	3,67%	4,24%	3,55%	3,09%	2,70%	1,89%	1,20%	0,70%	0,34%
<i>Délais moyens de résiliation pour chaque année de production</i>	3,7	3,7	3,6	3,6	3,4	3,4	3,3	3,3	3,2	3,2	2,8	0,7	NA	NA	NA	NA	NA	NA
Taux de réduction	NA	NA	NA	NA	NA	NA	NA	NA	0,00%	0,10%	0,19%	0,28%	0,53%	1,05%	1,87%	2,27%	2,45%	1,46%
<i>Délais moyens de réduction pour chaque année de production</i>	13,2	13,1	12,2	11,5	10,8	10,1	9,7	9,2	8,8	8,1	NA	NA	NA	NA	NA	NA	NA	NA

Tableau 57 - Titre : chiffres clés résiliations du produit Protectys

8.7. Annexe 7 : Contrôles des écarts entre la base de données Protectys et les reportings

	Année	2016	2017	2018
Reportings	Nombre de contrats	129 745	124 670	120 453
	Nombre de résiliations	4 011	3 361	2 281
	Nombre de réductions	689	1 313	2 257
Base de données	Nombre de contrats	129 651	124 542	120 282
	Nombre de résiliations	4 008	3 357	2 277
	Nombre de réductions	687	1 310	2 254
Ecart	Nombre de contrats	0,1%	0,1%	0,1%
	Nombre de résiliations	0,1%	0,1%	0,2%
	Nombre de réductions	0,3%	0,2%	0,1%

Tableau 58 - Titre : contrôles des écarts entre la base de données Protectys et les reportings

8.8. Annexe 8 : code Python de la partie 5 – Application de méthodes de machine learning de prédiction

```
In [1]: #Sources :
#https://scikit-learn.org
#https://scikit-learn.org/stable/modules/tree.html
#https://github.com
#https://fr.moonbooks.org/Articles/
#Comment-lire-un-fichier-excel-extension-xlsx-avec-pandas-
en-python-/
#https://www.delftstack.com/fr/howto/python-pandas/
#how-to-check-if-nan-exists-in-pandas-dataframe/
#https://www.geeksforgeeks.org/
#cart-classification-and-regression-tree-in-machine-
learning/
import pandas as pd
import numpy as np
#On nomme le fichier à lire
f = "/Users/Arnaud/Desktop/Taf final R et R/Lois R & R
source/Base R&R modifiée.#On le lit
df = pd.read_excel(f)
#On supprime les éventuels doublons
df.drop_duplicates(keep = 'first', inplace=True)
#Un seul doublon détecté
#Les variables NUM_CONTRAT, ID_PERSONNE, num_sirocco,
CD_PRODUIT, produit,
#date_premier_anniv, sont supprimées car inutiles pour
l'analyse
#Les variables CD_SOR_CONTRAT, DEP_CD_POSTAL_v2,
MONTANT_GARANTIE_sous
#sont redondantes avec d'autres variables donc supprimées
#Les variables relatives aux montants de primes et
garanties sont
#supprimées car les montants sont a posteriori de la
réduction donc
#inutilisable
del df['NUM_CONTRAT']
del df['ID_PERSONNE']
del df['num_sirocco']
del df['CD_PRODUIT']
del df['produit']
del df['date_premier_anniv']
del df['CD_SOR_CONTRAT']
del df['DEP_CD_POSTAL_v2']
del df['prime_TTC_an_theo']
del df['MONTANT_GARANTIE']
```

```

#On regarde si chaque colonne contient des valeurs nulles
df.isnull().any()
#On regarde le nombre de données manquantes par colonnes
df.isnull().sum()
#On supprime les colonnes qui ont plus de 99% de données
manquante
column_with_nan = df.columns[df.isnull().any()]
for column in column_with_nan:
if df[column].isnull().sum()*100.0/df.shape[0] > 99:
df.drop(column,1, inplace=True)
#On vérifie que les colonnes avec plus de 99% de données
manquantes
#ont bien été supprimées
df.isnull().sum()
#Les variables restantes contiennent pour certaines des
données
/var/folders/zs/19j845156kqc_83r535mlw7c0000gn/T/ipykernel
_4412/84435080.py
:53: FutureWarning: In a future version of pandas all
arguments of DataFram
e.drop except for the argument 'labels' will be keyword-
only
df.drop(column,1, inplace=True)
#manquantes, ce qui est normal car tous les assurés
#ne sont pas sortis, décédés, n'ont pas résilié, réduit ou
ne
#touchent pas une rente
#On peut supprimer les colonnes dtdeces, info_dc,
dt_sit_famil,
#date_rente_dep_debut_couv et date_rente_dep_dernier_couv
#qui ne sont pas utiles pour l'analyse des réductions
del df['dtdeces']
del df['date_rente_dep_debut_couv']
del df['dt_sit_famil']
#En revanche, certaines données sont vraiment manquantes
concernant
#les colonnes situ_fam, CD_POSTAL,
#LIB_TYPE_VENTE ou formule
#On remplace les valeurs manquantes pour ces variables en
utilisant
#les valeurs des lignes précédentes puis on vérifie qu'il
n'y ait
#plus de données manquante
df['situ_fam'].ffill(inplace=True)
df['CD_POSTAL'].ffill(inplace=True)
df['LIB_TYPE_VENTE'].ffill(inplace=True)
df['formule'].ffill(inplace=True)

```

```

df.isnull().any()
#On remplace les valeurs "string" par des chiffres
import numpy as np
df['situ_fam'] = df['situ_fam'].replace("CELIBATAIRE", 1)
df['situ_fam'] = df['situ_fam'].replace("CONCUBIN", 2)
df['situ_fam'] = df['situ_fam'].replace("DIVORCE", 3)
df['situ_fam'] = df['situ_fam'].replace("MARIE", 4)
df['situ_fam'] = df['situ_fam'].replace("PACSE", 5)
df['situ_fam'] = df['situ_fam'].replace("SEPRE", 6)
df['situ_fam'] = df['situ_fam'].replace("VEUF", 7)
df['formule'] = df['formule'].replace("DT", 1)
df['formule'] = df['formule'].replace("DT+DP", 2)
df['LIB_TYPE_VENTE'] =
df['LIB_TYPE_VENTE'].replace("SOUSCRIPTION CONJOINTE -
df['LIB_TYPE_VENTE'] = df['LIB_TYPE_VENTE'].replace("VENTE
SIMPLE", 2)
df['top_red'] = df['top_red'].fillna(0)

```

```

In [2]: #On introduit un modèle de prédiction CART
#On charge les librairies
import pandas as pd
from sklearn.tree import DecisionTreeClassifier # Import
du classifieur
from sklearn.model_selection import train_test_split #
Import fonction de
#séparation entre échantillon d'entraînement et de test
from sklearn import metrics #Import les mesures permettant
d'évaluer
#la précision du classifieur
#On partage les données en données source et cible
feature_cols =
['date_sou', 'CD_SEXE', 'dtnai', 'situ_fam', 'CD_POSTAL',
'frac']
X = df[feature_cols] # Données d'entrées
y = df.top_red # Variable cible
#On partage les données en bloc d'entraînement et de test
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_# Création du classifieur
clf = DecisionTreeClassifier()
# Entraînement
clf = clf.fit(X_train,y_train)
#Prédire la réponse avec les données de test
y_pred = clf.predict(X_test)
#Précision du modèle, fréquence à laquelle le classifieur
est correct
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

```

```

#On peut afficher l'arbre de décision avec les 3 lignes
ci-dessous
#from sklearn.tree import export_text
#r = export_text(clf, feature_names=feature_cols)
#print(r)
#On peut afficher l'importance des variables de notre
modèle :
pd.DataFrame(clf.feature_importances_,
index = X_train.columns,
columns = ["importance"]).sort_values(
"importance",
ascending = False)

```

Out[2]:

Accuracy: 0.8275457296529079

importance

dtnai 0.290091

CD_POSTAL 0.287662

date_sou 0.242118

Ancienneté contrat au moment de la sortie 0.088021

situ_fam 0.039038

CD_SEXE 0.022135

formule 0.017701

frac 0.010358

LIB_TYPE_VENTE 0.002877

In [3]: #On fait l'exercice avec les forêts aléatoires

```

from sklearn.ensemble import RandomForestClassifier

```

```

modele_rf = RandomForestClassifier(

```

```

n_estimators=100,

```

```

criterion='gini',

```

```

max_depth=None,

```

```

min_samples_split=2,

```

```

min_samples_leaf=1,

```

```

min_weight_fraction_leaf=0.0,

```

```

Accuracy: 87.41417037282072 %

```

importance

```

max_features='auto',

```

```

max_leaf_nodes=None,

```

```

min_impurity_decrease=0.0,

```

```

bootstrap=True,

```

```

oob_score=False,

```

```

n_jobs=None,

```

```

random_state=None,

```

```

verbose=0,

```

```

warm_start=False,

```

```

class_weight=None,

```

```

ccp_alpha=0.0,
max_samples=None,)
#Source : https://www.stat4decision.com/fr/foret-aleatoire-avec-python/ :
#"Tous ces hyperparamètres ont des valeurs par défaut, il est utile de
#les comprendre. En voici une description rapide :
#n_estimators : il s'agit du nombre d'arbres dans la forêt
#criterium : il s'agit du critère utilisé pour construire les arbres et
#séparer les branches des arbres
#max_depth : il s'agit de la profondeur maximale des arbres utilisés
#(le nombre de niveaux dans l'arbre de décision)
#min_sample_split : il s'agit du nombre d'échantillons minimal dans
#une feuille pour refaire une séparation
#min_samples_leaf : il s'agit du nombre d'échantillons minimal pour
#créer une feuille
#min_weight_fraction_leaf : il s'agit de la fraction du nombre total
#d'échantillon minimal pour créer une feuille
#max_features : il s'agit du nombre de colonnes sélectionnées pour
#chaque arbre (par défaut on prend la racine carré du nombre de colonnes)
#max_leaf_nodes : il s'agit du nombre maximal de feuilles
#min_impurity_decrease : il s'agit de la baisse minimale du critère
#d'impureté pour faire une séparation
#bootstrap : paramètre pour utiliser du bootstrap, si il est à False,
#le même échantillon est pris pour chaque arbre
#n_jobs ; nombre de traitements à effectuer en parallèle
#random_state : graine aléatoire
#warm_start : ceci permet de repartir du résultat du dernier apprentissage
#pour faire l'apprentissage
#class_weights : il s'agit des poids associés à chaque classe si cela a un
#sens
#max_samples : si vous voulez réduire le nombre d'observations dans vos
#échantillons bootstrap"
modele_rf.fit(X_train,y_train)

```

```

#Précision du modèle, fréquence à laquelle le classifieur
est correct
from sklearn.metrics import accuracy_score,
confusion_matrix
print(f"Accuracy: {accuracy_score(y_test,
modele_rf.predict(X_test))*100} %")
#On peut afficher l'importance des variables de notre
modèle :
pd.DataFrame(modele_rf.feature_importances_,
index = X_train.columns,
columns = ["importance"]).sort_values(
"importance",
ascending = False)

```

Out[3]:

```
Accuracy: 86.4815549919751 %
```

```

dtnai 0.290611
CD_POSTAL 0.282306
date_sou 0.262527
Ancienneté contrat au moment de la sortie 0.095903
situ_fam 0.026264
CD_SEXE 0.015021
formule 0.012878
frac 0.009214
LIB_TYPE_VENTE 0.005275

```

```

In [4]: #On modélise avec une régression logistique
from sklearn.linear_model import LogisticRegression
LR = LogisticRegression(random_state=0, solver='lbfgs',
multi_class='ovr').
print(f"Accuracy: {accuracy_score(y_test,
LR.predict(X_test))*100} %")

```

```
Accuracy: 83.4815549919751 %
```

```

In [5]: #On modélise avec un modèle SVM
import sklearn as sk
from sklearn import svm
import pandas as pd
import os
SVM = svm.LinearSVC()
SVM.fit(X_train, y_train)
print(f"Accuracy: {accuracy_score(y_test,
SVM.predict(X_test))*100} %")

```

Accuracy: 87.4815549919751 %

```
In [6]: #Plus proches voisins
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier(n_neighbors=3)
model.fit(X_train, y_train)
print(f"Accuracy: {accuracy_score(y_test,
model.predict(X_test))*100} %")
```

Accuracy: 85.80893397532498 %

```
In [7]: #Adaboost
# Import du classifieur AdaBoost
from sklearn.ensemble import AdaBoostClassifier
# Création du classifieur adaboost
Accuracy: 86.4815549919751 %
abc = AdaBoostClassifier(n_estimators=50, learning_rate=1,
random_state=0)
# Entraînement
modell1 = abc.fit(X_train, y_train)
print(f"Accuracy: {accuracy_score(y_test,
abc.predict(X_test))*100} %")
```

Accuracy: 87.4815549919751 %