

Mémoire d'actuariat

Construction de lois Best Estimate pour le risque
décès en assurance emprunteur

Lucille Delbet

Master 2 Actuariat du CNAM

Directeur de mémoire en entreprise : Gauthier Ait Braham

REMERCIEMENTS

Je tiens tout d'abord à remercier Gauthier Ait Braham, *Experience Analysis Actuary*, de m'avoir accompagnée tout au long de cette aventure et de m'avoir donné de précieux conseils pour la rédaction de ce mémoire.

Je remercie également mon manager Nicolas Heinrich pour m'avoir permis de réaliser ce mémoire et d'avoir toujours fait preuve de soutien et de disponibilité à mon égard.

Je remercie Nathanaël Abecera pour sa relecture et ses conseils avisés pour ce mémoire.

Je remercie mes collègues chez SCOR et plus particulièrement l'équipe *Pricing & Experience Analysis* qui m'ont aidée par leurs conseils, leur relecture mais aussi motivée par leur bonne humeur et bienveillance.

Enfin, un merci particulier à mes proches qui, eux aussi, ce sont portés volontaires pour la relecture de mon mémoire et m'ont soutenu pendant cette période de rédaction.

Résumé

Mots clés : Assurance, Emprunteur, Décès, Tarifs, Tables de mortalité d'expérience, Immobilier, Consommation, Machine Learning

L'assurance emprunteur garantit la prise en charge du capital restant dû en cas de décès de l'assuré ou des échéances de remboursement en cas d'arrêt de travail. Alors qu'elle représente un des marchés les plus importants en France avec environ 11 Mds d'euros de chiffre d'affaires 2021, l'assurance emprunteur est en mouvance perpétuelle, notamment du fait des nombreuses évolutions réglementaires qui accentuent toujours un peu plus la protection de l'assuré. Dans ce contexte, l'ajustement des tarifs au plus près du risque assuré est essentiel pour rester compétitif.

En exploitant les données d'un portefeuille emprunteur réassuré par SCOR, nous avons pu répondre à cette exigence de finesse des tarifs proposés grâce à la création de nouvelles tables de mortalité d'expérience. Après un travail conséquent de préparation des données, notamment à travers un exercice de classification des prêts entre prêts immobiliers et prêts à la consommation à l'aide d'un modèle machine learning, nous avons calculé des taux bruts de mortalité, qui ont ensuite été lissés puis extrapolés.

Une application des nouvelles lois de mortalité créées a finalement été réalisée pour mesurer leur impact sur les tarifs d'une nouvelle génération de souscription.

ABSTRACT

Key words : Insurance, Loan, Death, Rates, Experience mortality tables, Real estate loan, Personal loan, Machine Learning

Loan insurance guarantees the payment of the remaining capital due in the event of the insured's death or of the monthly payments in the event of work stoppage. While it represents one of the most important markets in France, with revenues of around 11 billion euros in 2021, loan insurance is in constant flux, particularly due to the many regulatory changes that are constantly increasing the protection of the insured. In this context, adjusting rates as closely as possible to the insured risk is essential to remain competitive.

By using data from a loan portfolio reinsured by SCOR, we have been able to meet this need for finesse in the proposed rates thanks to the creation of new experience mortality tables. After a substantial work of data preparation, in particular through a classification exercise of the loans between real estate and personal loans using a machine learning model, we first calculated gross mortality rates, which were then smoothed and extrapolated.

An application of the newly created mortality laws was finally performed to measure their impact on the rates of the new generation of underwriting.

TABLE DES MATIERES

Remerciements	2
Abstract	4
Introduction	8
Partie 1 – Présentation de l'assurance emprunteur	9
1.1. Présentation de l'assurance emprunteur	9
1.1.1. Définition.....	9
1.1.2. Présentation des principales garanties.....	10
1.1.3. Les différents types de crédit.....	11
1.1.4. Les modalités de remboursement.....	12
1.1.5. Les différents types de contrats d'assurance	15
1.2. La tarification de l'assurance emprunteur	16
1.2.1. Principe de la tarification emprunteur	16
1.2.2. Les tables de mortalité.....	20
1.2.3. Le calcul du taux de prime du risque décès	22
1.2.4. Tarification d'un contrat emprunteur pour les prêts immobiliers.....	25
1.3. Présentation de l'assurance emprunteur en France	26
1.3.1. Quelques chiffres.....	26
1.3.2. Les évolutions réglementaires.....	27
1.3.3. Les différentes logiques de tarification.....	31
Partie 2 – Construction de la base de donnees	33
2.1. Présentation des données	33
2.1.1. Le fichier des encours.....	33
2.1.1. Le fichier des sinistres	34
2.2. Analyse et retraitements	34
2.2.1. Regroupement des informations pour un même assuré	34

2.2.2.	Contrôles et retraitements	35
2.3.	Structure des données	35
2.4.	Statistiques descriptives.....	36
2.4.1.	Analyse des adhésions	37
2.4.2.	Analyse des sinistres.....	39
Partie 3 – Création d’une variable à l’aide d’un modèle machine learning		41
3.1.	Contexte.....	41
3.2.	Pré-processing des données.....	43
3.2.1.	Variables sélectionnées et traitements effectués	43
3.2.2.	Découpage.....	47
3.3.	Création d’un modèle Machine Learning et application.....	47
3.3.1.	Arbre de décision	47
3.3.2.	Random Forest.....	48
3.4.	Application du modèle.....	52
3.4.1.	Ajustement du montant des capitaux assurés	52
3.4.2.	Résultat sur la base de données étudiée.....	53
Partie 4 – Construction de lois d’expérience du risque décès		56
4.1.	Détermination de la période d’observation.....	56
4.1.1.	Pré-testing sur les données du portefeuille	56
4.1.2.	Analyse des expositions et des sinistres.....	58
4.2.	Estimation des taux de mortalité bruts	61
4.3.	Méthodes de lissage.....	63
4.3.1.	Utilisation du package R <i>Mortality Smooth</i>	63
4.3.2.	Lissage relationnel de Brass	64
4.4.	Validation des modèles	67
4.4.1.	Intervalles de confiance ponctuels	67
4.4.2.	Etude des résidus	69
4.5.	Gestion des durations courtes.....	72

4.6. Backtesting des lois créées	74
Partie 5 – Application des nouvelles lois et impacts sur les engagements	78
5.1. Méthodologie.....	78
5.1.1. Création des <i>model points</i>	78
5.1.2. Modélisation des flux projetés.....	79
5.2. Impact de la nouvelle table sur les flux projetés.....	80
5.2.1. Comparaison des tables de mortalité	80
5.2.1. Résultats des impacts du changement de tables	82
Conclusion	85
Bibliographie.....	87
Annexes	88

INTRODUCTION

L'assurance emprunteur permet, lors de la souscription d'un prêt auprès d'un établissement bancaire, de se protéger face aux risques de décès, d'arrêt de travail et/ou de perte d'emploi.

Avec une croissance des encours de crédit aux particuliers de +5,7% en 2021 malgré la crise sanitaire¹, l'assurance emprunteur représente l'un des domaines de l'assurance de personnes les plus importants sur le marché français. Elle atteint un chiffre d'affaires en 2021 d'environ 11 milliards d'euros, dont 67% représenteraient la couverture de prêts immobiliers².

Le risque décès, systématiquement proposé lors de la souscription d'une assurance emprunteur, représente environ 71% des primes selon les chiffres 2021². La quantification de ce risque est déterminante pour sa tarification et nécessite l'utilisation de tables de mortalité, qui permettent de calculer les probabilités de décès annuelles par âge. Des tables réglementaires peuvent être utilisées par défaut, construites à partir des données de l'Institut National de la Statistique et des Etudes Economiques (INSEE).

Dans un univers de plus en plus compétitif du fait des nombreuses évolutions réglementaires ces dernières années, mais également de l'apparition de nouveaux acteurs qui proposent un parcours de souscription de plus en plus simplifié, il est primordial d'utiliser pour la tarification d'un produit des tables de mortalité au plus proche du risque de la population cible du produit. Ce dernier présentant des caractéristiques spécifiques, les assureurs peuvent avoir recours à l'utilisation de tables d'expérience, construites à partir des données du portefeuille sous-jacent.

Ce mémoire présente l'étude d'un portefeuille emprunteur réassuré par SCOR dans le but de construire de nouvelles tables d'expérience. Après avoir traité les problématiques liées aux données du portefeuille, notamment avec le recours à un modèle de machine learning pour classifier les prêts selon leur type (prêts immobiliers ou prêts à la consommation), des tables de mortalité seront créées. L'impact sur les projections Best Estimate sera calculé pour prendre en compte l'implémentation de ces nouvelles tables d'expérience.

¹ Source : Banque de France, « *Crédits aux particuliers – Déc 2021* »

² Source : Fédération française de l'assurance (FFA, aujourd'hui France Assureurs), « *Les données clés de l'assurance française en 2021* »

1.1.2. Présentation des principales garanties

1.1.2.1. *Les garanties décès et PTIA*

La garantie décès est la garantie socle de l'assurance emprunteur. Cette garantie intervient en cas de décès de l'assuré emprunteur pendant la durée de remboursement de crédit : dans ce cas, l'organisme assureur rembourse à l'établissement de crédit le montant du capital restant dû (CRD) au jour du décès de l'assuré. Le contrat peut toutefois prévoir des exclusions de garantie (comme par exemple le suicide la première année du contrat ou la pratique d'un sport dangereux).

La garantie décès est accompagnée de la garantie PTIA (Perte Totale et Irréversible d'Autonomie), qui permet à l'assuré souscripteur le remboursement du capital restant dû si celui-ci est dans l'incapacité totale et définitive d'exercer toute occupation ou activité professionnelle et que l'assistance d'une tierce personne est requise pour accomplir les actes de la vie quotidienne.

1.1.2.2. *La garantie incapacité*

La garantie incapacité temporaire totale de travail (ITT) n'est pas obligatoire mais est fréquemment souscrite dans le cadre de l'assurance emprunteur. Elle permet la prise en charge par l'assureur des échéances du prêt dans le cas où, à la suite d'un accident ou à une maladie, l'assuré serait inapte à exercer totalement et temporairement une activité professionnelle. Le contrat peut préciser qu'il s'agisse de l'activité exercée au moment de l'accident ou de la maladie ou, dans certains cas, toute activité professionnelle.

Le contrat d'assurance peut là aussi prévoir des exclusions de garantie, ou encore un délai de franchise, c'est-à-dire un délai à partir de la date de survenance du sinistre durant lequel l'assuré n'est pas indemnisé.

Le contrat d'assurance peut également proposer une garantie optionnelle qui couvre l'incapacité temporaire partielle (ITP), c'est-à-dire le cas où, à la suite d'un accident ou d'une maladie, l'assuré serait temporairement et partiellement dans l'impossibilité d'exercer une activité professionnelle. Un pourcentage des échéances du prêt sera alors pris en charge par l'assureur.

1.1.2.3. La garantie invalidité

Comme la garantie ITT, la garantie invalidité permanente totale (IPT) n'est pas obligatoire mais est fréquemment souscrite dans le cadre de l'assurance emprunteur. Elle permet la prise en charge par l'assureur des échéances du prêt en cas d'inaptitude permanente de l'assuré à exercer une activité professionnelle à la suite d'une maladie ou un accident, et après consolidation de son état de santé, c'est-à-dire une fois que les lésions ou séquelles sont supposées définitives.

L'invalidité de l'assuré est définie selon un taux évalué par un médecin, à l'aide d'un barème croisé (professionnel/fonctionnel) indiqué au contrat. Dans la majorité des contrats d'assurance emprunteur, l'invalidité est considérée comme totale lorsque le taux d'invalidité est supérieur à 66%.

En cas d'invalidité permanente et totale, le contrat peut prévoir le remboursement des échéances de l'emprunt, ou le versement du capital restant dû de l'emprunt à la date de déclaration de l'invalidité.

L'assurance emprunteur peut également couvrir l'invalidité permanente partielle (IPP), qui permet la prise en charge par l'assureur des échéances du prêt en cas d'invalidité dont le taux serait généralement compris entre 33% et 66%. Si l'assuré se trouve en IPP, une fraction de la mensualité assurée du prêt est remboursée.

Les contrats d'assurance prévoient en général une liste d'exclusions, comme la pratique d'activités dangereuses ou d'un métier à risques.

1.1.3. Les différents types de crédit

Il existe trois grandes familles de crédit :

- Le crédit à la consommation : il permet l'achat de biens de consommation (acquisition d'une voiture, achat d'équipement électroménager...) ou la mise à disposition d'une trésorerie. L'article L312-1 du Code de la Consommation précise que le montant du crédit à la consommation doit être « égal ou supérieur à 200 euros et inférieur ou égal à 75 000 euros ». La durée minimale d'un crédit à la consommation est de trois mois. En revanche,

il n'existe pas d'obligation légale sur sa durée maximale. En pratique, les établissements bancaires proposent les crédits à la consommation sur au plus 7 ans.

- Le crédit immobilier : il permet l'achat de biens immobiliers. Fin 2021, le Haut Conseil de Stabilité Financière (HCSF) a annoncé que les banques seraient désormais contraintes de limiter la durée des prêts immobiliers à 25 ans, et de fixer à 35% le taux maximum d'endettement (contre 33% auparavant), avec une application prévue au 1^{er} janvier 2022.
- Le crédit professionnel : il est réservé aux professionnels uniquement, et permet de financer l'achat de matériel, l'immobilier ou d'avoir de la trésorerie à disposition dans le cadre d'une activité professionnelle, quelle que soit cette activité et quelle que soit la taille de l'entreprise.

1.1.4. Les modalités de remboursement

Lors de la mise en place du prêt, l'établissement de crédit fournit à l'emprunteur les modalités de remboursement du prêt, sous la forme d'un tableau d'amortissement, ou échéancier de prêt.

Ce tableau d'amortissement fait apparaître les informations suivantes :

- Le capital initial, noté CI
- La durée de l'emprunt en années, noté n
- Le taux d'intérêt annuel du prêt, noté i
- Le capital restant dû à l'échéance annuelle k , noté CRD_k
- Le montant des intérêts versés à chaque échéance k , noté I_k
- Le montant du capital amorti à chaque échéance k , noté A_k
- Le montant des mensualités (on considèrera ici la somme des mensualités de l'échéance k), noté M_k
- Le capital restant dû après remboursement des mensualités de l'échéance k , noté V_k

En règle générale, le prêt se rembourse par échéances constantes, avec des mensualités fixes sur toute la durée du prêt ($M_k = M$). Chaque mensualité est composée d'une part de capital (croissante au fur et à mesure du remboursement) et d'une part d'intérêts (décroissante au fur et à mesure du remboursement).

M est donc un montant fixe tel que :

$$CI = \sum_{k=1}^n \frac{M}{(1+i)^k}$$

$$CI = M \times \sum_{k=1}^n \frac{1}{(1+i)^k}$$

$$CI = M \times \frac{1 - (1+i)^{-n}}{i}$$

$$M = CI \times \frac{i}{1 - (1+i)^{-n}}$$

Le tableau d'amortissement se présente alors de la manière suivante :

Echéance	CRD _k	I _k	A _k	M _k	V _k	
1	CI	$i \times CI$	$M - I_0$	$CI \times \frac{i}{1 - (1+i)^{-n}}$	$CI \times (1+i) - M$	
...	
k	$\frac{CRD_{k-1} \times (1+i)}{-M}$	$\frac{i \times CI}{\times \frac{1 - (1+i)^{-n+k-1}}{1 - (1+i)^{-n}}}$	$M - I_k$			$\frac{CRD_k \times (1+i)}{-M}$
...
n	$\frac{CRD_{n-1} \times (1+i)}{-M}$	$i \times CI \times \frac{1 - (1+i)^{-1}}{1 - (1+i)^{-n}}$	$M - I_n$			0

Exemple : Calcul du tableau d'amortissement pour un prêt de durée 5 ans, d'un montant de 50 000€, à un taux d'emprunt de 1,5%, avec un remboursement par échéances constantes.

Echéance	CRD _k	I _k	A _k	M _k	V _k
1	50 000,00	750,00	9 704,47	10 454,47	40 295,53
2	40 295,53	604,43	9 850,03	10 454,47	30 445,50
3	30 445,50	456,68	9 997,78	10 454,47	20 447,72
4	20 447,72	306,72	10 147,75	10 454,47	10 299,97
5	10 299,97	154,50	10 299,97	10 454,47	0,00

Il est également possible d'opter pour un remboursement « in fine », qui permet de ne payer que les intérêts de l'emprunt lors des échéances mensuelles, avant de rembourser l'intégralité du capital emprunté en une seule fois à la fin du prêt. Ce type de remboursement est réservé aux investisseurs présentant de solides garanties, puisque le capital emprunté doit être remboursé en une fois à la fin du prêt. Ce type de prêt est également le plus coûteux pour l'assureur, puisqu'en cas de décès au cours du prêt, l'assureur devra rembourser le montant du capital initial.

Le tableau d'amortissement se présente alors de la manière suivante :

Echéance	CRD_k	I_k	A_k	M_k	V_k
0	CI	$i \times CI$	0	$i \times CI$	CI
...
k	CI	$i \times CI$	0	$i \times CI$	CI
...
n	CI	$i \times CI$	CI	$i \times CI + CI$	0

Exemple : Calcul du tableau d'amortissement pour un prêt de durée 5 ans, d'un montant de 50 000€, à un taux d'emprunt de 1,5%, avec remboursement « in fine ».

Echéance	CRD_k	I_k	A_k	M_k	V_k
1	50 000,00	750,00	-	750,00	50 000,00
2	50 000,00	750,00	-	750,00	50 000,00
3	50 000,00	750,00	-	750,00	50 000,00
4	50 000,00	750,00	-	750,00	50 000,00
5	50 000,00	750,00	50 000,00	50 750,00	-

D'autres modalités de remboursement sont possibles :

- Un remboursement par échéances modulables, qui permet de modifier le montant des mensualités à verser chaque mois, à la baisse ou à la hausse, par exemple en cas de changement de salaire.
- Un remboursement par différé d'amortissement, qui permet de rembourser dans un premier temps les intérêts avant de rembourser le capital.

- Un remboursement par anticipation, qui permet, en plus des mensualités, de rembourser tout ou partie du capital restant dû.

1.1.5. Les différents types de contrats d'assurance

Au moment de l'octroi du prêt, un contrat d'assurance collectif est proposé par l'établissement bancaire à l'emprunteur, qu'on appelle une assurance de groupe. L'article L141-1 du Code des Assurances définit cette assurance de groupe comme « *le contrat souscrit par une personne morale ou un chef d'entreprise en vue de l'adhésion d'un ensemble de personnes répondant à des conditions définies au contrat, pour la couverture des risques dépendant de la durée de la vie humaine, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité, des risques d'incapacité de travail ou d'invalidité ou du risque de chômage.* »

Le contrat collectif est basé sur le principe de mutualisation entre tous les assurés du groupe. De ce fait, les assurés présentant un risque moins important compensent les assurés ayant un risque plus défavorable.

Même si sa facilité d'adhésion est avérée, notamment grâce aux formalités médicales allégées, le contrat collectif est peu flexible. De plus, son tarif peut être tiré vers le haut pour les populations qualifiées de « bons risques » ou vers le bas pour les populations qualifiées de « mauvais risques » du fait de la mutualisation.

Le contrat individuel, quant à lui, permet d'avoir un contrat d'assurance emprunteur sur-mesure. Le coût de l'assurance est adapté selon le profil de l'emprunteur (âge, fumeur ou non, état de santé...). Mécaniquement, les contrats individuels présenteront un tarif très attractif pour les assurés considérés comme de « bons risques » alors qu'ils seront plus chers pour les assurés définis comme « mauvais risques ».

Un phénomène d'antisélection peut apparaître au moment de la souscription du contrat : alors que les « mauvais risques » auront tendance à souscrire au contrat collectif, plus avantageux pour eux en termes de tarif du fait de la mutualisation souhaitée, les « bons risques » se tourneront vers le contrat individuel, plus adapté à leur profil et donc plus attractif au niveau tarifaire. Le phénomène d'antisélection limite ainsi le principe de mutualisation et peut remettre en cause l'équilibre d'une structure tarifaire donnée.

1.2. La tarification de l'assurance emprunteur

1.2.1. Principe de la tarification emprunteur

Comme définie précédemment, l'assurance emprunteur garantit pendant toute la durée du prêt le remboursement du capital restant dû en cas de décès de l'assuré emprunteur, ou le paiement des échéances du prêt en cas d'incapacité ou d'invalidité de l'assuré emprunteur, moyennant le paiement d'une cotisation d'assurance.

Nous étudierons dans ces travaux la tarification du risque décès uniquement.

1.2.1.1. Différents types de tarification

Plusieurs types de tarification sont aujourd'hui proposées sur le marché :

- Le tarif en pourcentage du capital initial :

$$Cotisation = Taux_{CI} \times Capital\ Initial$$

Une tarification en pourcentage du capital initial est fonction de l'âge à l'adhésion et de la durée initiale de l'emprunt.

Dans ce cas, les primes sont constantes tout au long du contrat.

- Le tarif en pourcentage du capital restant dû :

$$Cotisation = Taux_{CRD} \times Capital\ Restant\ Dû$$

Les primes en pourcentage du capital restant dû peuvent être fonction :

- De l'âge atteint de l'assuré : Les primes reflètent le profil de risque de l'assuré. Le taux de primes est recalculé chaque année selon l'âge atteint de l'assuré. La cotisation évolue tout au long du contrat puisqu'elle suit l'évolution du CRD mais également l'évolution du risque lié à l'âge atteint de l'assuré.
- De l'âge à l'adhésion de l'assuré : Le taux de primes est le même toute la durée du prêt. Le montant de primes va décroître en fonction du capital restant dû.

En calculant les primes d'assurance décès pour un même profil³ selon les trois types de tarification différentes, nous observons l'évolution suivante :

- La tarification en pourcentage du capital restant dû est la plus élevée la première année du prêt.
- Comme le risque décès augmente avec l'âge, la tarification en pourcentage du capital restant dû en fonction de l'âge atteint devient la plus élevée.
- La tarification en pourcentage du capital initial étant constante sur toute la durée du prêt, elle est généralement moins élevée sur les premières années mais devient la plus élevée sur les dernières années du prêt.

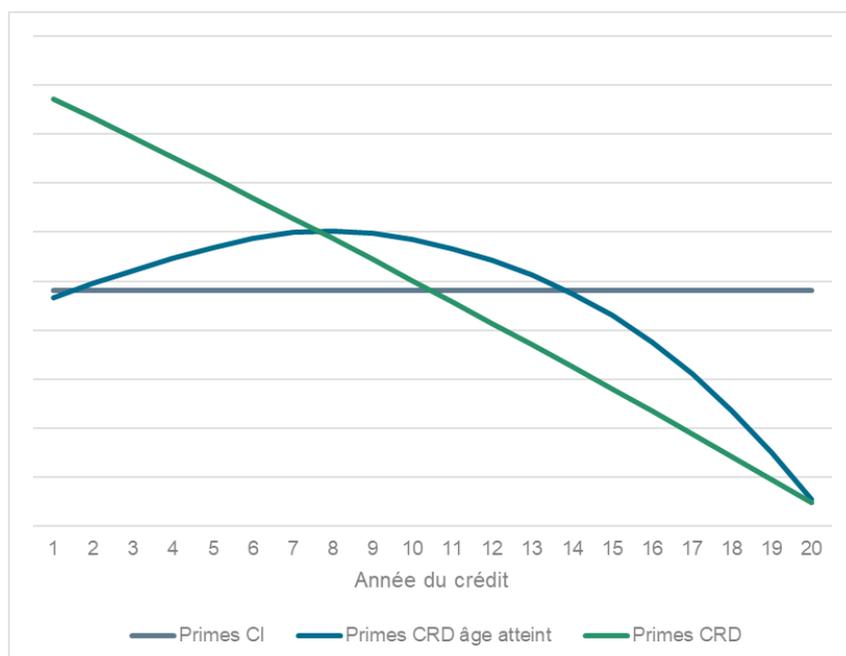


Figure 2: Evolution de la prime d'assurance en fonction de l'année du prêt selon les différentes méthodes de tarification

³ Nous négligeons ici la probabilité de rachat anticipé du prêt et nous basons sur le calcul d'une prime pure, hors chargements et hors taxe.

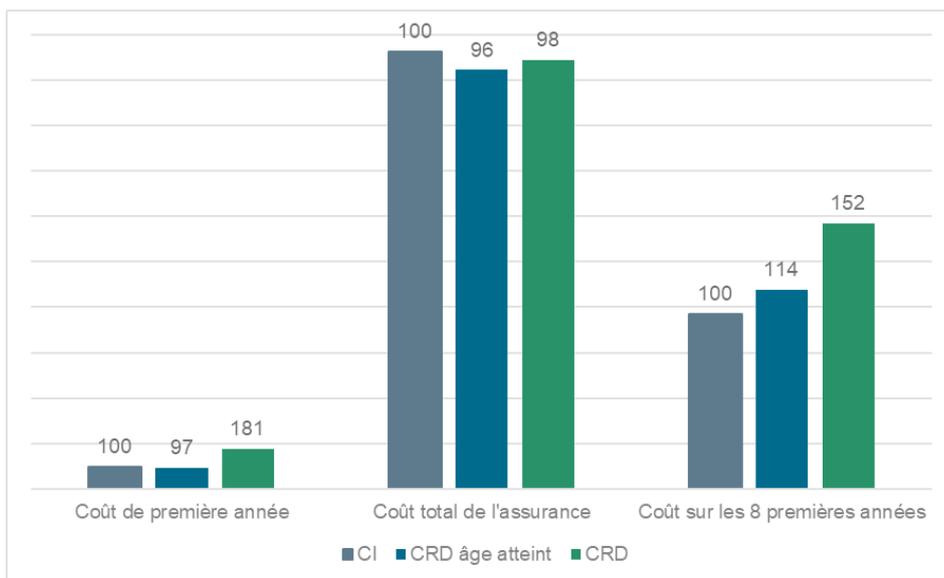


Figure 3: Calcul du coût d'assurance⁴ selon les différentes méthodes de tarification

1.2.1.2. Les hypothèses de tarification

Plusieurs hypothèses sont déterminantes dans la tarification d'un contrat emprunteur :

- Les caractéristiques relatives au prêt :
 - Le montant du prêt,
 - Le taux d'intérêt du prêt,
 - La durée de prêt,
 - Le type de remboursement,
 - Le fractionnement de remboursement,
 - Le type de tarification.
- Les caractéristiques relatives à l'assuré :
 - L'âge à l'adhésion de l'assuré : Le tarif est fonction de l'âge à l'adhésion de l'assuré dans le cadre d'une tarification en CI et de l'âge atteint de l'assuré dans le cadre d'une tarification en CRD. Le calcul du risque décès sous-jacent dépend également de l'âge atteint de l'assuré.
 - Le sexe de l'assuré : Bien que le tarif ne soit pas différencié par sexe, le calcul du risque décès sous-jacent dépend du sexe de l'assuré à travers le choix de la table de mortalité, que nous décrivons en section 1.2.2.

⁴ Coût d'assurance hors actualisation, hypothèses de rachats ou hypothèses de décès

1.2.1.3. *Evolution du risque au fil du temps*

Si l'on se place côté risque :

- La probabilité de survenance du sinistre (décès ou arrêt de travail) augmente avec le temps,
- Le montant de la prestation en cas de décès (remboursement du CRD) diminue avec le temps.

Pour permettre d'équilibrer les engagements pris par l'assureur (remboursement du CRD en cas de décès) et les engagements pris par l'assuré (paiement des cotisations) tout au long de la vie du prêt, des provisions doivent être calculées.

Ces provisions sont calculées de telle sorte que la valeur actuelle probable des engagements de l'assureur soit équivalente à la valeur actuelle probable des engagements de l'assuré.

Deux composantes interviennent alors dans le calcul de la valeur actuelle probable : les tables de mortalité, qui permettent de modéliser les espérances de durée de vie à chaque âge, et le taux technique utilisé pour l'actualisation des flux.

1.2.2. Les tables de mortalité

Le calcul des engagements servant à tarifier et provisionner un contrat d'assurance emprunteur nécessite le recours à l'utilisation de tables de mortalité. Ces tables fournissent par âge le nombre de personnes survivantes, et permettent donc de calculer les probabilités de décès annuelles par âge.

On note ainsi :

l_x : le nombre de personnes toujours en vie à l'âge x parmi 100 000 personnes.

${}_t p_x = \frac{l_{x+t}}{l_x}$: la probabilité d'être en vie pour un assuré d'âge x au début de la période t

${}_t q_x = 1 - {}_t p_x$: la probabilité d'être en vie pour un assuré d'âge x au début de la période t

L'utilisation des tables de mortalité est encadrée par le Code des Assurances. L'article A132-18 précisent que :

« Les tarifs pratiqués par les entreprises d'assurance sur la vie et de capitalisation comprennent la rémunération de l'entreprise et sont établis d'après les éléments suivants :

1° Un taux d'intérêt technique fixé dans les conditions prévues à l'article A. 132-1.

2° Une des tables suivantes :

a) Tables homologuées par arrêté du ministre de l'Économie et des finances, établies par sexe, sur la base de populations d'assurés pour les contrats de rente viagère, et sur la base de données publiées par l'Institut national de la statistique et des études économiques pour les autres contrats ;

b) Tables établies ou non par sexe par l'entreprise d'assurance et certifiées par un actuinaire indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire reconnues par l'autorité mentionnée à l'article L. 310-12. »

1.2.2.1. Les tables réglementaires

Les tables réglementaires sont établies à partir de données relatives à la population française générale. Les tables actuellement en vigueur sont la TH 00-02 et la TF 00-02. Elles ont été

construites à partir des données de l'Institut National de la Statistique et des Etudes Economiques (INSEE) :

- La TH 00-02 calcule l'espérance de vie d'un homme à partir des données observées entre 2000 et 2002.
- La TF 00-02 calcule l'espérance de vie d'une femme à partir des données observées entre 2000 et 2002.

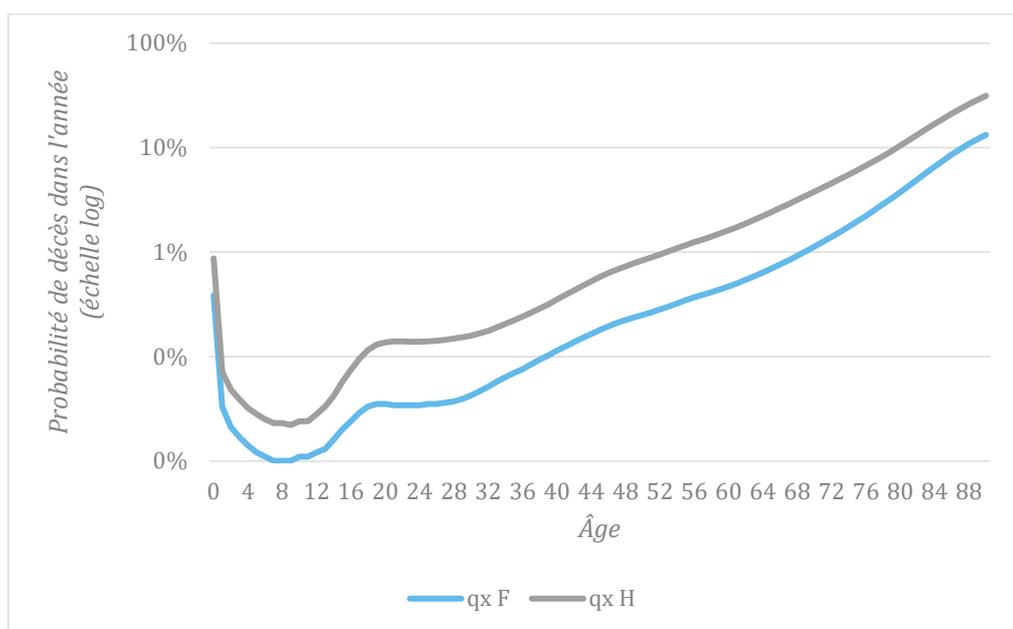


Figure 4: Probabilité de décès dans l'année suivant l'âge d'un individu avec la TH 00-02 (qx H) et la TF 00-02 (qx F) en échelle logarithmique

En comparant les deux tables, on observe que la mortalité des hommes est supérieure à celle des femmes, notamment pour les âges supérieurs à 50 ans. Avant 2012, les femmes payaient donc une prime d'assurance inférieure à celle des hommes, puisqu'on considérait qu'elles représentaient un « meilleur risque » par rapport aux hommes.

Depuis 2012, suite à une décision de la Cour de Justice de l'Union Européenne, il est interdit de prendre le sexe de l'assuré en compte dans le calcul de la prime d'assurance.

Par prudence, c'est donc la table TH 00-02 qui est utilisée pour la tarification du risque décès des contrats d'assurance emprunteur, puisqu'elle est plus défavorable dans l'évaluation de la probabilité de décès.

1.2.2.2. Les tables d'expérience

Un portefeuille assuré en emprunteur présente des caractéristiques spécifiques :

- Les personnes contractant un emprunt font partie d'une sous-population ayant des revenus plus élevés, ce qui influe à la baisse sur la probabilité de décès.
- De plus, la sélection médicale mise en place à la souscription du contrat d'assurance et qui permet de lutter contre le phénomène d'antisélection permet d'identifier les « mauvais risques » et les « bons risques ». De ce fait, les populations assurées sont généralement en meilleure santé que la population générale. Cela s'observe particulièrement sur les premières années du contrat.

De ce fait, la création de tables d'expérience, établies à partir des données historiques des portefeuilles assurés, s'est imposée pour pallier les différences de caractéristiques entre les populations sous-jacentes.

La tarification des contrats d'assurance emprunteur peut s'appuyer sur des tables d'expérience sans condition de certification. Toutefois, pour être utilisées pour les travaux de provisionnement en accord avec le régulateur, les tables d'expérience doivent être certifiées par un actuinaire agréé par l'Institut des Actuaires. Cette certification permet de s'assurer :

- De la cohérence des données utilisées pour la création de la table d'expérience (un certain volume de données et une profondeur d'historique sont nécessaires),
- De la pertinence des tables construites,
- De la prudence sur l'utilisation de ces tables.

1.2.3. Le calcul du taux de prime du risque décès

Lors de la souscription du contrat d'assurance, un taux de prime est calculé de telle sorte que la valeur actuelle probable de l'engagement de l'assuré (le paiement des primes pures) soit égale à la valeur actuelle probable de l'engagement de l'assureur (le paiement des prestations en cas de sinistre). Il détermine la prime pure due par l'assuré.

Comme décrit au paragraphe 1.2.1, le taux de primes peut être appliqué :

- Soit sur le capital initial (CI) : dans ce cas, la prime payable par l'assuré sera constante.
- Soit sur le capital restant dû (CRD) : dans ce cas, la prime payable par l'assuré sera décroissante.

Généralement, le paiement des primes en assurance emprunteur s'effectue de manière mensuelle en début de période.

A t=0 :

$$VAP(0, Assureur) = VAP(0, Assuré)$$

L'engagement de l'assureur correspond à l'espérance de paiement du capital restant dû du prêt à chaque échéance (qu'on considère ici comme mensuelle) en cas de décès de l'assuré. On supposera que le décès survient en milieu de mois.

Ainsi :

$$VAP(0, Assureur) = \sum_{k=0}^{n \times 12 - 1} CRD_k \times \frac{1}{12} q_{x+\frac{k}{12}} \times \frac{k}{12} p_x \times \frac{1}{(1+i_m)^{k+0,5}}$$

Avec :

$\frac{1}{12} q_{x+\frac{k}{12}}$ la probabilité mensuelle de décès de l'assuré entre l'âge $x + \frac{k}{12}$ et $x + \frac{k+1}{12}$

$\frac{k}{12} p_x$ la probabilité mensuelle de survie de l'assuré entre l'âge x et $x + \frac{k}{12}$

i_m le taux d'actualisation mensuel, tel que $i_m = (1 + i_{annuel})^{\frac{1}{12}} - 1$

1.2.3.1. Tarification en fonction du capital initial

La tarification en fonction du capital initial est calculée selon l'âge à l'adhésion de l'assuré et la durée initiale du prêt.

Le taux de prime qui sera appliqué au capital initial sera constant tout au long de la durée du prêt :

$$VAP(0, Assuré) = Tx_{CI} \times CI \times \sum_{k=0}^{n \times 12 - 1} \frac{k}{12} p_x \times \frac{1}{(1+i_m)^k}$$

Ainsi, en égalisant la VAP de l'assureur et la VAP de l'assuré, on obtient :

$$VAP(0, Assureur) = Tx_{CI} \times CI \times \sum_{k=0}^{n \times 12 - 1} \frac{k}{12} p_x \times \frac{1}{(1+i_m)^k}$$

$$Tx_{CI} = \frac{VAP(0, Assureur)}{CI \times \sum_{k=0}^{n \times 12 - 1} \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^k}}$$

1.2.3.2. Tarification en fonction du capital restant dû

Comme précisé dans la section 1.2.1.1, la tarification en fonction du capital restant dû peut dépendre de l'âge atteint de l'assuré ou de son âge à l'adhésion.

Dans le cas où le tarif dépend de l'âge atteint de l'assuré, le taux de prime va évoluer au fil du temps selon l'évolution du risque décès de l'assuré à chaque âge.

A chaque pas de temps k , l'équation à résoudre est la suivante :

$$CRD_k \times \frac{1}{12} q_{x+\frac{k}{12}} \times \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^{k+0,5}} = Tx_{CRD_k} \times CRD_k \times \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^k}$$

D'où :

$$Tx_{CRD_k} = \frac{1}{12} q_{x+\frac{k}{12}} \times \frac{1}{(1 + i_m)^{0,5}}$$

Dans le cas où le tarif est fonction de l'âge à l'adhésion, le taux de prime sera le même tout au long de la durée du prêt. On peut alors poser l'équation suivante :

$$VAP(0, Assuré) = Tx_{CRD} \times \sum_{k=0}^{n \times 12 - 1} CRD_k \times \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^k}$$

Ainsi, en égalisant la VAP de l'assureur et la VAP de l'assuré, on obtient :

$$VAP(0, Assureur) = Tx_{CRD} \times \sum_{k=0}^{n \times 12 - 1} CRD_k \times \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^k}$$

$$Tx_{CRD} = \frac{VAP(0, Assureur)}{\sum_{k=0}^{n \times 12 - 1} CRD_k \times \frac{k}{12} p_x \times \frac{1}{(1 + i_m)^k}}$$

1.2.4. Tarification d'un contrat emprunteur pour les prêts immobiliers

Pour illustrer les éléments décrits ci-dessus, nous présentons un exemple de tarification d'un contrat emprunteur dans le cadre d'un emprunt immobilier, en tarification en pourcentage du CI puis en pourcentage du CRD.

Les caractéristiques du prêt sont les suivantes :

- Le montant du prêt : 200 000€
- Le taux d'intérêt du prêt : 1%
- La durée de prêt : 20 ans
- Le type de remboursement : à échéances constantes
- Le fractionnement de remboursement : mensuel

Les caractéristiques relatives à l'assuré sont les suivantes :

- L'âge à l'adhésion de l'assuré : 40 ans
- La table de mortalité utilisée : TH00-02

On suppose un taux d'actualisation nul.

En construisant le tableau d'amortissement du prêt et à l'aide des formules explicitées en section 1.2.3, nous obtenons les taux de primes pures décès mensuels suivants :

- En tarification en fonction du CI : 0,0211%
- En tarification en fonction du CRD : 0,0383%
- En tarification en fonction du CRD âge atteint :

Âge atteint	Taux de prime mensuel
40	0,0197%
41	0,0220%
42	0,0244%
43	0,0271%
44	0,0300%
45	0,0331%
46	0,0362%
47	0,0393%
48	0,0423%
49	0,0454%
50	0,0485%
51	0,0520%
52	0,0557%
53	0,0596%
54	0,0639%
55	0,0684%

56	0,0729%
57	0,0777%
58	0,0829%
59	0,0889%

A cette prime pure viennent s'ajouter les chargements pour former la prime commerciale hors taxe. Ces chargements correspondent aux frais liés à l'acquisition et à la gestion du contrat.

$$Prime\ commerciale\ HT = \frac{Prime\ pure}{(1 - Chargements)}$$

On applique ensuite la taxe d'assurance pour obtenir la prime commerciale toutes taxes comprises :

$$Prime\ commerciale\ TTC = \frac{Prime\ pure}{(1 - Chargements)} \times (1 + Taxe)$$

La taxe d'assurance est de 9% pour le risque décès.

1.3. Présentation de l'assurance emprunteur en France

1.3.1. Quelques chiffres

Avec une croissance des encours de crédit aux particuliers de +5,7% en 2021 malgré la crise sanitaire, l'assurance emprunteur représente l'un des domaines de l'assurance de personnes les plus importants sur le marché français⁵. Elle atteint un chiffre d'affaires en 2021 de 11 milliards d'euros, dont 67% représenteraient la couverture de prêts immobiliers⁶. Le risque décès, systématiquement proposé lors de la souscription d'une assurance emprunteur, représente environ 71% des primes selon les chiffres 2021⁵.

⁵ Source : Banque de France, « *Crédits aux particuliers – Déc 2021* »

⁶ Source : Fédération française de l'assurance (FFA, aujourd'hui France Assureurs), « *Les données clés de l'assurance française en 2021* »

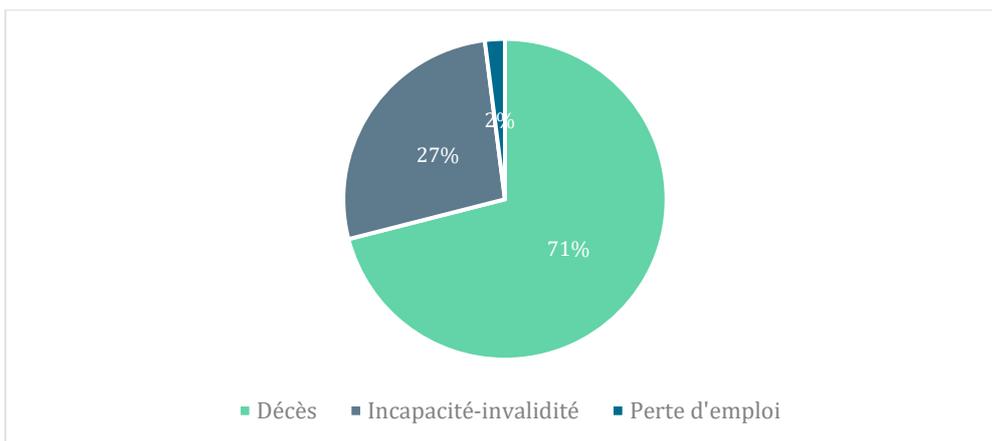


Figure 5: Répartition des cotisations par type de garantie

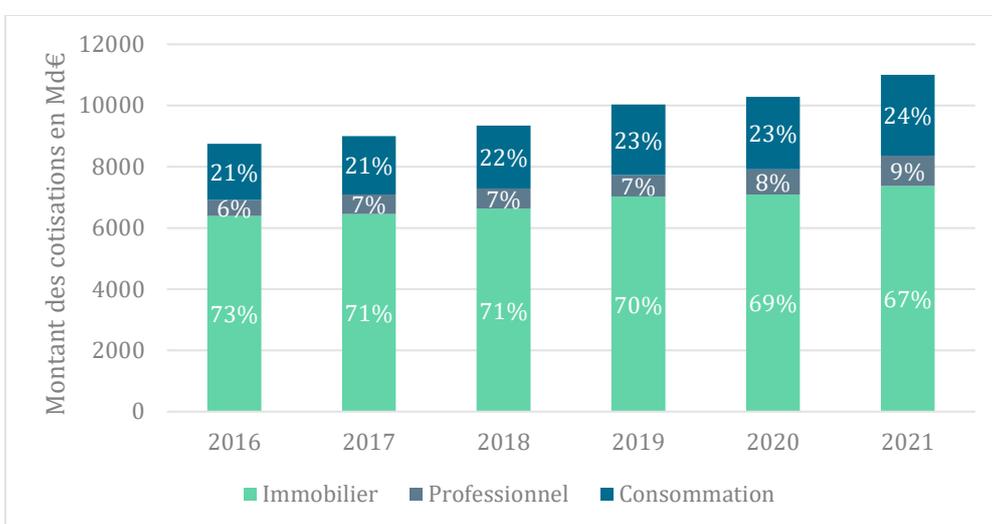


Figure 6: Répartition des cotisations par type de prêt

1.3.2. Les évolutions réglementaires

Le marché de l'assurance emprunteur a été longtemps concentré chez les établissements bancaires, qui proposaient des contrats simples avec peu de sélection médicale au moment de l'octroi du prêt à l'assuré.

Le marché s'est complexifié au fil du temps et a vu naître de nombreuses évolutions réglementaires qui lui ont permis de s'ouvrir à la concurrence.

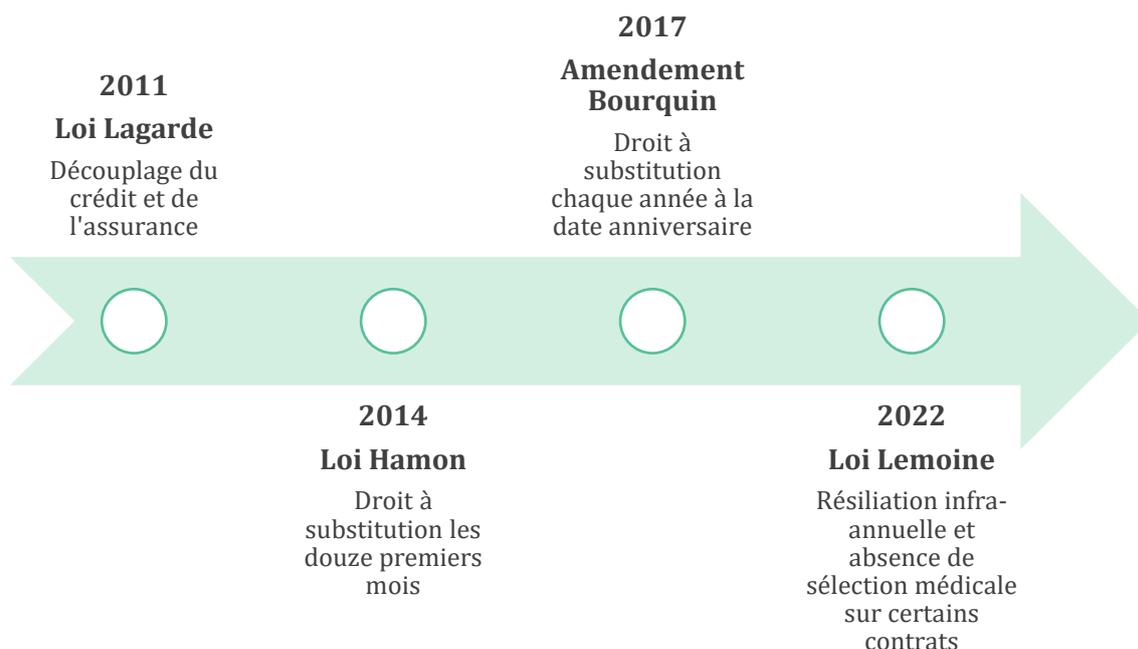


Figure 7: Les dates-clés de l'assurance emprunteur en France

En 2011 tout d'abord, la loi Lagarde a introduit sur le marché la délégation d'assurance de prêt : au moment de la souscription du prêt, l'assuré peut choisir une assurance autre que celle proposée par la banque prêteuse. Une seule condition doit être respectée : les garanties souscrites avec l'assurance externe doivent être au moins équivalentes à celles incluses dans le contrat d'assurance de l'établissement bancaire. La loi Lagarde précise également que l'établissement bancaire ne peut pas modifier les conditions de prêt si l'assuré choisit d'avoir recours à la délégation d'assurance.

En mars 2014, la loi Hamon vient ouvrir encore davantage le marché à la concurrence en permettant à l'assuré de changer son assurance emprunteur dans les douze mois après signature de son contrat, toujours sous la condition d'équivalence de garanties.

Enfin, l'amendement Bourquin en 2017 vient renforcer la loi Hamon et prévoit le droit pour un assuré de substituer son contrat emprunteur chaque année à la date anniversaire du contrat. Ce dispositif est applicable pour les nouveaux contrats souscrits mais également pour le stock d'encours de contrats.

Bien que ces évolutions réglementaires soient fortes, elles ont eu un impact modéré sur la distribution des contrats emprunteurs. Les parts de marché des bancassureurs sont en baisse ces

dernières années, au profit des assureurs, courtiers grossistes et intermédiaires d'assurance mais restent prépondérantes avec environ 88% des cotisations en 2019.

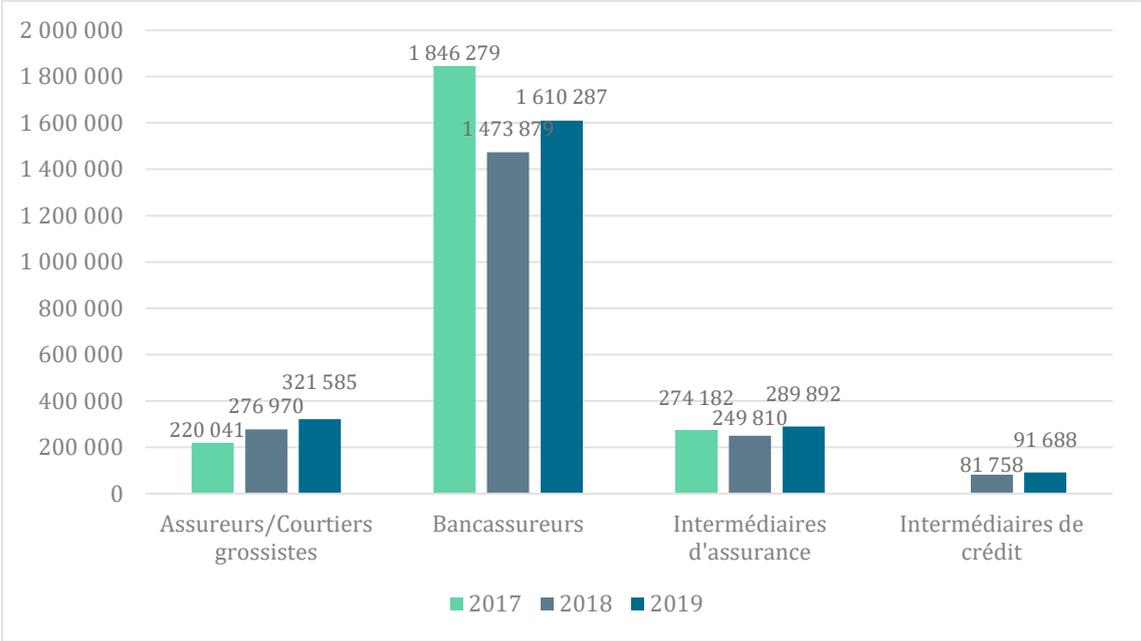


Figure 8: Répartition de la distribution en assurance emprunteur

Pour contrer la concurrence des contrats alternatifs proposés par les assureurs externes (contrats alternatifs externes), les banques ont, elles aussi, proposé des contrats alternatifs (contrats alternatifs internes). Là-aussi, bien que la part des contrats alternatifs soient en hausse chaque année, la part des contrats groupes bancaires reste prépondérante.



Figure 9: Répartition de la distribution des ventes en assurance emprunteur⁷

Bien qu'elles n'aient pas transformé la répartition de la distribution, ces évolutions réglementaires ont eu un impact notable sur les tarifs des contrats d'assurance emprunteur. L'ouverture à la concurrence a tiré les tarifs vers le bas : une baisse tarifaire globale (sur l'ensemble des profils assurés) entre 20% et 41%⁸ a été mesurée depuis 2010.

De plus, elles ont conduit à des produits plus sophistiqués, avec un étoffement des garanties proposées (perte d'emploi, maladies non objectivables...) et une segmentation plus fine des contrats par sous-populations.

En 2022, l'adoption d'une nouvelle loi, la loi Lemoine, vient bouleverser le marché de l'assurance emprunteur.

Cette loi prévoit plusieurs volets, avec deux mesures particulièrement impactantes pour la tarification des contrats :

- La suppression de la sélection médicale pour les prêts de moins de 200 000€ si l'assuré n'atteint pas les 60 ans à la fin de son prêt : en volume, cette mesure concernerait environ 52% des prêts en nombre, et 36% en capital.

⁷ Source : Comité Consultatif du Secteur Financier (CCSF), « Rapport – Bilan de l'assurance emprunteur », 2020

⁸ Source : Comité Consultatif du Secteur Financier (CCSF), « Rapport – Bilan de l'assurance emprunteur », 2020 – L'analyse porte uniquement sur les contrats alternatifs externes.

- La résiliation infra-annuelle des contrats d'assurance à compter de la signature de l'offre de prêt.

Elle est applicable dès le 1^{er} juin 2022 pour les nouveaux contrats et à partir du 1^{er} septembre 2022 pour le stock de contrats en cours.

Des impacts importants engendrés par la loi Lemoine ont d'ores et déjà été identifiés sur la tarification des contrats emprunteurs. Plusieurs effets pourraient pousser les acteurs du marché à revoir leurs tarifs vers le haut :

- L'absence de sélection médicale va engendrer une hausse de la sinistralité sur le segment des prêts concernés : les assurés qui devaient à l'époque faire l'objet d'une surprime paieront désormais le tarif standard.
- Les assurés qui se voyaient refusés lors de la sélection médicale du fait d'un surrisque trop important pourront désormais être assurés, à un tarif sans surprime eux-aussi.
- On peut également identifier un « effet d'aubaine » puisque certains individus malades, pour qui l'octroi d'un prêt n'était pas accessible pourraient désormais décider d'emprunter et bénéficier d'un tarif sans surprime.

Bien que mise en place pour faciliter l'accès à la propriété aux personnes en situation de risque aggravé, la loi Lemoine pourrait finalement engendrer une hausse des tarifs de l'assurance emprunteur pour tous les assurés, avec des produits possiblement plus « standard », et une limitation des options possibles pour limiter le risque d'antisélection chez les assurés.

1.3.3. Les différentes logiques de tarification

Les logiques de tarification diffèrent selon qu'il s'agisse d'un contrat groupe bancaire ou d'un contrat alternatif. Les contrats groupes bancaires doivent être en mesure de couvrir l'ensemble des assurés ayant souscrits un prêt auprès de la banque. Le tarif doit être mutualisé pour répondre à ce besoin.

Les contrats alternatifs sont plus libres dans la construction de leur tarif selon les risques sélectionnés. Le tarif peut donc être plus segmenté selon le risque de chaque individu.

Si les contrats groupes bancaires proposent généralement une tarification par tranche d'âge (89% des contrats groupes bancaires), les contrats alternatifs, quant à eux, proposent principalement une tarification âge par âge (87% des contrats alternatifs). Ces contrats vont également plus loin en proposant généralement une segmentation par catégorie socio-professionnelle (80% des

contrats alternatifs), par statut fumeur/non-fumeur (93% des contrats alternatifs⁵) ou par critère sports et activités à risque (67% des contrats alternatifs) ⁹.

De par cette segmentation plus fine, la dispersion des tarifs alternatifs est nettement plus importante que celle des tarifs groupes bancaires. Alors que les tarifs groupes bancaires sont lissés et donc homogènes, les tarifs des contrats alternatifs reflètent au plus près le profil de risque de chaque assuré. Ainsi, alors que les tarifs groupes bancaires seront plus intéressants pour les profils plus âgés, ils seront en revanche moins avantageux pour des profils plus jeunes et moins risqués, qui se tourneront davantage vers les contrats alternatifs.

Nous avons montré dans cette première partie le poids important de l'assurance emprunteur au sein du marché français de l'assurance. Particulièrement sujet aux évolutions réglementaires pour favoriser l'accès à la propriété pour les assurés, le marché subit une forte mise en concurrence. Même s'il reste très concentré chez les bancassureurs, notamment grâce à leur proposition de contrats alternatifs internes, la part des contrats alternatifs externes ne cesse d'augmenter, en particulier chez les tranches d'âges jeunes. Les évolutions réglementaires auront tout de même eu des bénéfices certains du point de vue de l'assuré : des tarifs en baisse depuis plusieurs années et des garanties renforcées. Pour permettre d'être toujours plus compétitifs, les assureurs ont dû ajuster leur modélisation du risque au plus près de leur portefeuille assuré. L'évaluation du risque décès reposant majoritairement sur l'utilisation de tables de mortalité, le recours à des tables dites « d'expérience » s'est vite imposé comme une solution efficace. Après avoir présenté et analysé la base de données utilisée, nous détaillerons dans les prochaines parties les différentes étapes de la construction d'une table d'expérience.

⁹ Source : Comité Consultatif du Secteur Financier (CCSF), « *Rapport – Bilan de l'assurance emprunteur* », 2020

PARTIE 2 – CONSTRUCTION DE LA BASE DE DONNEES

La construction d'une table d'expérience robuste repose sur une base de données de bonne qualité, correctement analysée et structurée. Plusieurs étapes sont essentielles pour s'assurer de la qualité et de la cohérence de l'information à disposition. Après avoir effectué ces vérifications, les données seront restructurées puis analysées à travers le calcul de statistiques descriptives du portefeuille.

2.1. Présentation des données

2.1.1. Le fichier des encours

Le fichier des encours utilisé dans le cadre de cette étude contient une importante quantité d'information : il s'agit d'un fichier « tête par tête », qui donne les caractéristiques de chaque prêt, en cours mais également clos, depuis l'origine des produits.

Ce fichier projette dans le temps (depuis la date d'effet jusqu'à la date de fin prévisionnelle ou réelle) les cotisations techniques ainsi que les chargements de gestion et commerciaux de chacun des contrats. En plus des cotisations, on y trouve aussi l'évolution du capital restant dû.

On trouve également dans ce fichier des informations non structurées mais qui donnent une information complémentaire sur le contrat (produit, durée, apporteur, etc), la tête assurée (sexe, âge, etc) ou la garantie (détail garantie, formule, etc).

Il existe cinq segments de prêts différents, dont les prêts sociaux et les prêts bancaires qui représentent 91% des prêts de la base de données. Les prêts sociaux sont des prêts accordés aux personnes à revenus modestes. L'objectif premier de ce type de prêt est de faciliter l'accès à la propriété, mais des prêts à la consommation sont également proposés.

Avant tout retraitement, on observe 1 315 134 assurés dans cette base.

2.1.1. Le fichier des sinistres

Le fichier des sinistres nous a été transmis séparément du fichier des encours. Il contient, lui aussi, des informations détaillées sur les sinistres ayant eu lieu sur les prêts en cours ou clos, depuis l'origine du produit (date de survenance du sinistre, date de déclaration, montant réglé, cause du sinistre, etc).

Le fichier reprend également des informations présentes dans le fichier des encours, telles que le numéro de contrat ou le numéro de personne, ce qui va nous permettre de rassembler l'information des deux fichiers de données.

Avant tout retraitement, on observe 16 846 sinistres décès dans cette base.

2.2. Analyse et retraitements

2.2.1. Regroupement des informations pour un même assuré

La première étape de la construction de notre base de données est de synthétiser les informations du fichier des encours et du fichier des sinistres pour chaque assuré.

L'objectif est de connaître le profil de chaque assuré du portefeuille et de pouvoir suivre leur parcours : âge à la souscription, sexe, date de souscription d'un prêt, date de fin d'un prêt, date du sinistre, surprimes, etc.

Nous retenons donc dans notre base de données :

- Les informations liées au risque biométrique de l'assuré : date de naissance, sexe, département de résidence.
- Les informations liées à son prêt : date de début de prêt, date de fin de prêt, réseau de distribution, type de prêt, produit, capital emprunté, taux ou montant de surprime.
- Les informations liées au sinistre : date de survenance du sinistre, cause du sinistre.

Après vérification, les informations du fichier des encours et du fichier des sinistres étaient dans la majorité des cas cohérentes. Pour les quelques exceptions, des décisions ont été prises pour choisir de retenir l'information la plus pertinente.

2.2.2. Contrôles et retraitements

L'étape de préparation des données est cruciale pour le reste de l'étude. Il s'agit d'une étape d'analyse des données et de potentiels retraitements des incohérences observées. Bien qu'elle soit chronophage, cette étape doit être soignée pour s'assurer de la robustesse des résultats obtenus au cours de l'étude.

Plusieurs vérifications sont nécessaires dans la base d'étude :

- La base ne doit pas contenir de doublons.
- Les dates de naissance, dates de début de prêt, dates de fin de prêt, dates de survenance du sinistre ne doivent pas être aberrantes et doivent être cohérentes entre elles : la date de naissance de l'assuré doit être antérieure à la date de début de prêt, elle-même antérieure à la date de fin de prêt ou date de survenance du sinistre.
- Lorsqu'un assuré a plusieurs prêts, les informations relatives à l'assuré doivent être cohérentes pour tous les prêts. Si un décès survient, la date de survenance du sinistre doit être la même pour l'ensemble des prêts souscrits.

En cas d'incohérence des informations, plusieurs solutions sont envisageables :

- Supprimer les lignes concernées de la base de données, ce qui signifie perdre de l'information disponible mais privilégier la qualité de l'information.
- Corriger les anomalies en prenant des hypothèses pour conserver l'information, même si elle peut être biaisée. C'est en général l'option retenue si les hypothèses sous-jacentes sont suffisamment fiables.

Après retraitement des données, on observe 1 314 990 assurés dans la base nettoyée.

2.3. Structure des données

Une fois le travail de préparation des données effectué, nous créons une structure homogène pour intégrer de manière synthétique et exhaustif l'ensemble des caractéristiques nécessaires pour l'étude.

La structure des données contient les variables ci-dessous.

Nom de la variable	Description
Clé_Assuré	Clé pour définir un assuré
Sexe	1 si l'assuré est un homme, 2 si l'assuré est une femme
Date_Naissance	Date de naissance de l'assuré
Date_Début	Date de début de l'exposition
Date_Fin	Date de fin de la ligne, NA si pas de date de fin
Cause_Fin	0 en cours (Date_End doit être NA), 1 ajout d'un prêt, 2 chute , 6 décès
Prêteur	Organisme prêteur
Réseau	Réseau de distribution
Produit	Nom du produit
Ind_Medical	1 si un nouveau prêt débute à la période de la ligne, 0 sinon
Montant_SA	Montant assuré du capital initial Si plusieurs prêts dans la ligne, somme des prêts
Surprime	Taux de surprime
Groupe_Indiv	GROUPE ou INDIVIDUEL
Statut	Statut du contrat
Departement	Département de résidence
Adh_Co_Adh	A si adhérent, C si co-adhérent

Figure 10: Présentation de la structure des données

2.4. Statistiques descriptives

La base de données contient les années d'exposition allant de 1970 à 2019. Elle compte 1 314 990 assurés à fin 2019.

Les chiffres-clés sont les suivants :

Age moyen à l'adhésion	36 ans
Pourcentage d'hommes	59%
Nombre total de décès	15 552

2.4.1. Analyse des adhésions

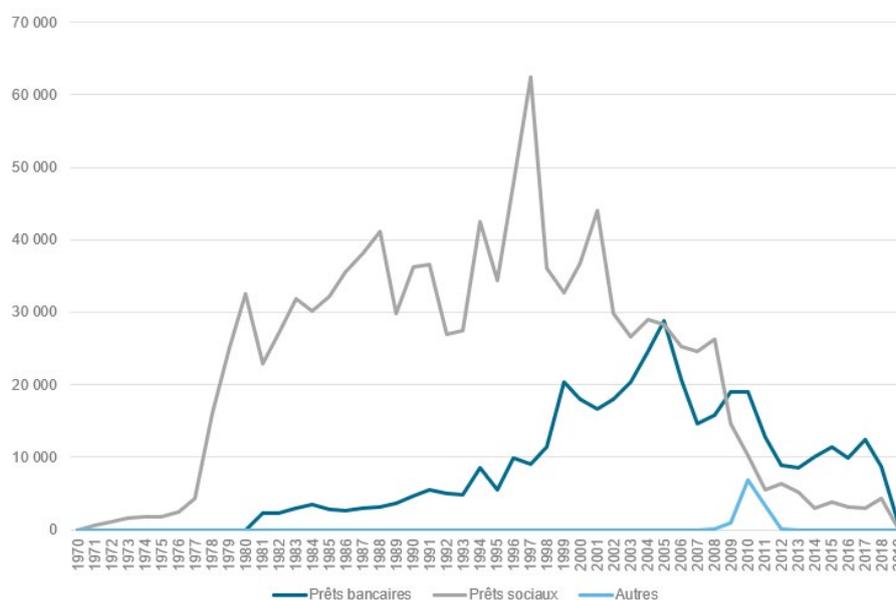


Figure 11: Analyse des adhésions par segment de prêts

La répartition des adhésions par segment de prêts met en avant un bloc important de prêts sociaux souscrits dans les années 1990. Les adhésions aux prêts bancaires se sont développées plus tard dans les années 2000.

Le segment des prêts sociaux représente la majorité du portefeuille. De plus, il s'agit d'un segment de prêts couvrant une population spécifique : les prêts sociaux sont accordés aux personnes ayant des revenus modestes. L'objectif premier de ce type de prêt est l'accès à la propriété, même si des prêts à la consommation sont également proposés. De par ces spécificités, il nous a paru opportun de construire une table de mortalité d'expérience pour ce type de prêts. Notre étude se concentrera donc sur le segment des prêts sociaux uniquement. Toutefois, il serait opportun pour les autres segments de prêts (bancaires et autres) d'avoir également leurs propres tables d'expérience. Pour cela, plusieurs méthodes pourraient être envisagées :

- Suivre la même démarche que pour les prêts sociaux en créant des tables basées sur les données de ces différents segments – cette méthode est la plus précise mais nécessite des conditions de volumétrie de données,
- Effectuer un positionnement de ces segments par rapport à une table de référence (tables d'expérience créées pour le segment des prêts sociaux ou tables réglementaires TH 00-02

et TF 00-02 par exemple) – cette méthode permet de pallier le manque de données le cas échéant.

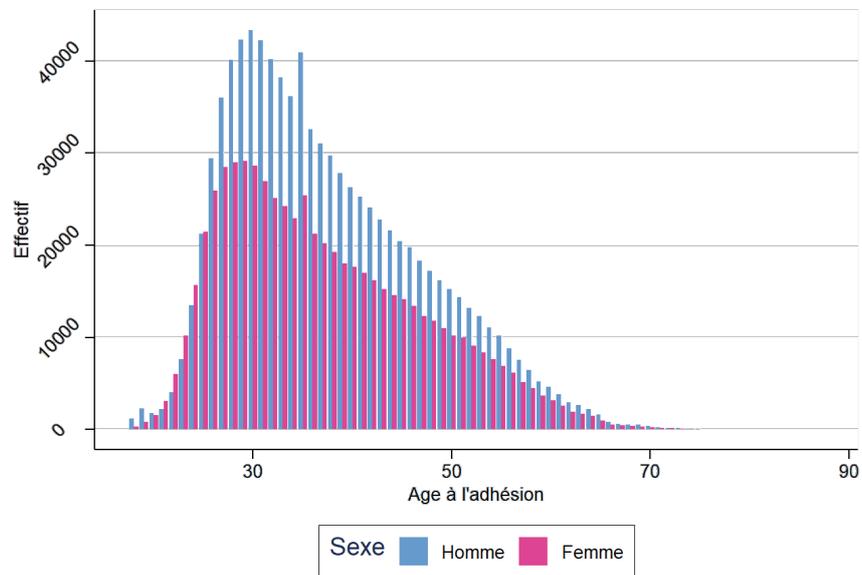


Figure 12: Analyse des âges à l'adhésion

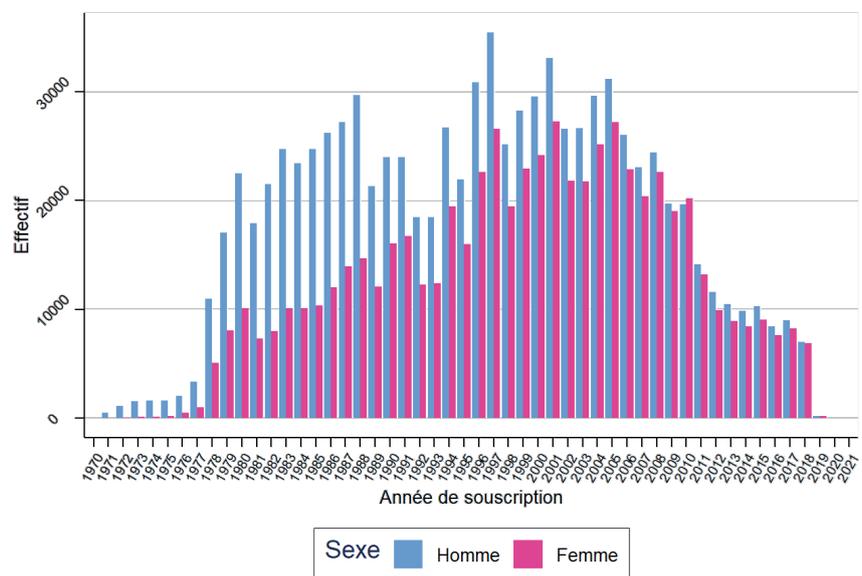


Figure 13: Analyse des années de souscription

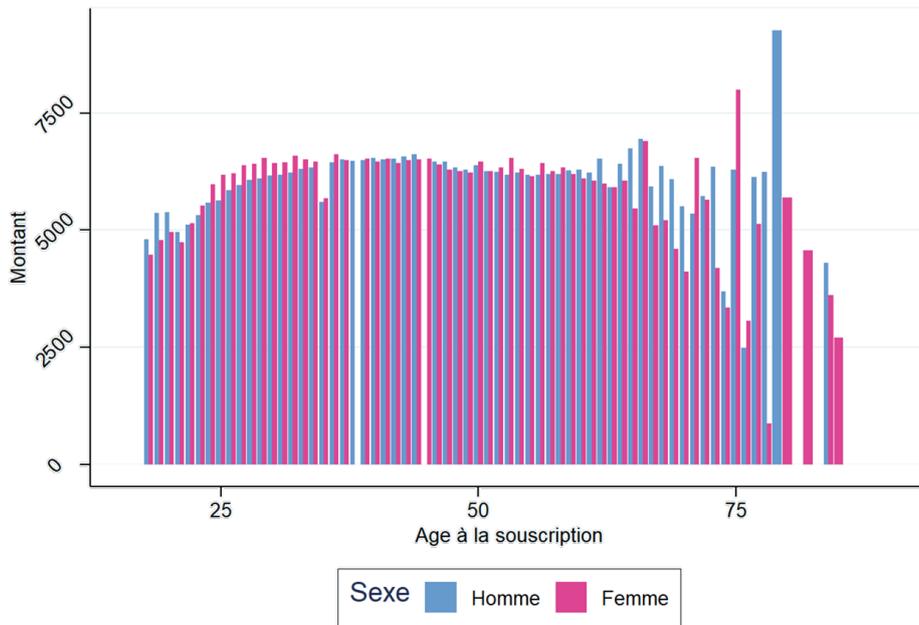


Figure 14: Analyse des montants moyens assurés par âge à la souscription

2.4.2. Analyse des sinistres

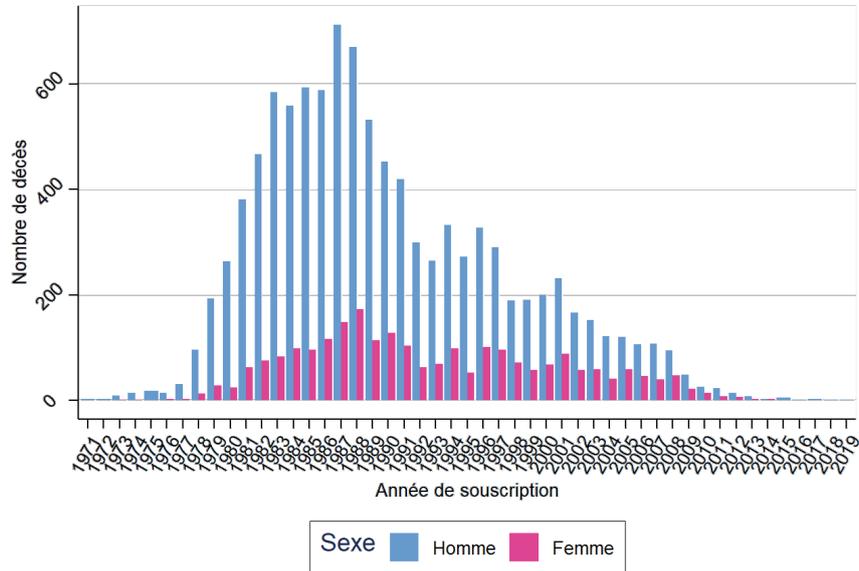


Figure 15: Analyse du nombre de décès par année de souscription

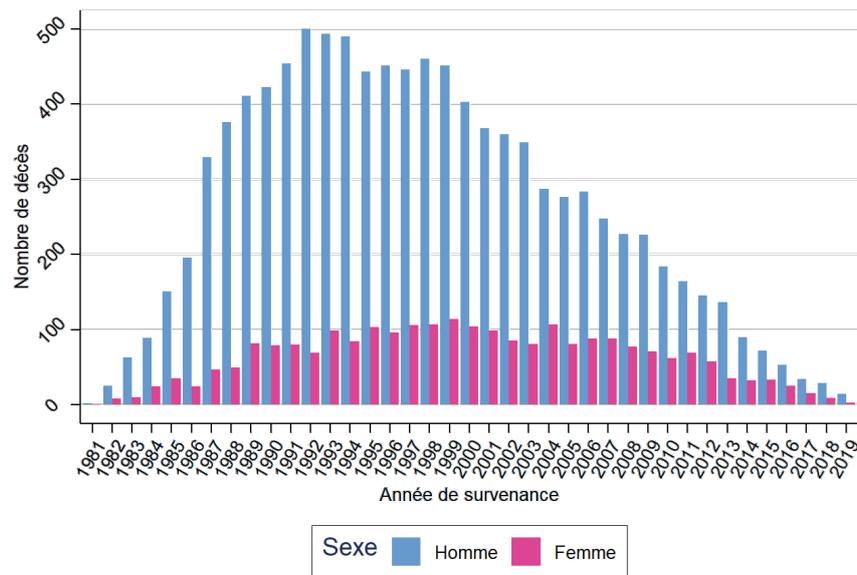


Figure 16: Analyse du nombre de décès par année d'occurrence

Nous avons préparé dans cette seconde partie la base de données à partir de laquelle nous construirons notre table d'expérience : nous retenons pour notre étude le segment des prêts sociaux uniquement pour prendre en compte les spécificités de ce segment. Avec près d'un million d'assurés, cette base est une richesse certaine d'informations et représente un socle robuste pour la création de notre table. Néanmoins, pour affiner notre modélisation, nous verrons dans la troisième partie que le type de prêt, immobilier ou à la consommation, est une variable clé. L'utilisation d'un modèle machine learning nous est apparue comme la solution adéquate pour faire la distinction entre prêt immobilier et prêt à la consommation pour chaque ligne de notre base de données.

PARTIE 3 – CREATION D'UNE VARIABLE A L'AIDE D'UN MODELE MACHINE LEARNING

La base de données construite est fiable et correctement structurée. Même si elle contient une quantité importante d'information, le type de prêt (prêt immobilier ou prêt à la consommation) n'y figure pas. Pourtant, cette variable serait une vraie valeur ajoutée dans le cadre de notre projet, de précédentes études chez SCOR ayant mis en évidence une différence notable sur le risque de mortalité entre ces deux types de prêts. La construction d'un modèle machine learning de type classification nous a semblé tout à fait adapté pour répartir les données entre ces deux types de prêts. L'algorithme de classification est une technique d'apprentissage supervisée, utilisée pour identifier la classe de nouvelles observations à partir d'une base de données d'apprentissage. Nous avons retenu pour notre étude le modèle *Random Forest* qui est une technique de machine learning très populaire pour sa facilité d'interprétation et la transparence de son utilisation.

3.1. Contexte

Même si la base contient un grand nombre de données, l'information relative au type de prêt (immobilier ou consommation) n'est pas disponible. C'est pourtant une information capitale pour notre étude. Nous chercherons donc par la suite à modéliser cette information à l'aide d'un modèle de machine learning.

Les précédentes études chez SCOR sur les portefeuilles emprunteur ont mis en évidence une différence notable sur le risque décès entre les prêts immobiliers et les prêts à la consommation. Il s'agit donc une variable indispensable pour notre étude.

Cette différence est illustrée par exemple dans le graphique ci-dessous. Dans le cadre de l'étude d'un de nos portefeuilles emprunteur, nous avons comparé le risque décès des prêts immobiliers puis des prêts à la consommation par rapport au risque décès moyen du portefeuille.

Pour cela, nous avons calculé des ratios Observés/Attendus ($\frac{A}{E} = \text{Actual/Expected}$; nous retiendrons cette notation pour la suite de ce mémoire) qui permettent d'estimer le risque décès pour les prêts immobiliers puis pour les prêts à la consommation (A) par rapport au risque décès moyen du portefeuille global (E).

Un ratio inférieur à 100% signifie que le risque décès du sous-segment est inférieur au risque moyen du portefeuille. A l'inverse, un ratio supérieur à 100% signifie que le risque décès du sous-segment est supérieur au risque moyen du portefeuille. Le risque décès moyen du portefeuille est bien entendu calculé en tenant compte de la répartition par âge et sexe de chaque modalité étudiée. En effet, cela permet de normaliser les effets vieillissement/rajeunissement d'une population ou un changement de répartition entre homme et femme (le risque de mortalité étant connu pour être plus favorable pour les femmes que pour les hommes).

Dans notre exemple ci-dessous, on voit que le ratio $\frac{A}{E}$ des prêts immobiliers est inférieur à 100% alors que celui des prêts à la consommation est supérieur à 100%.

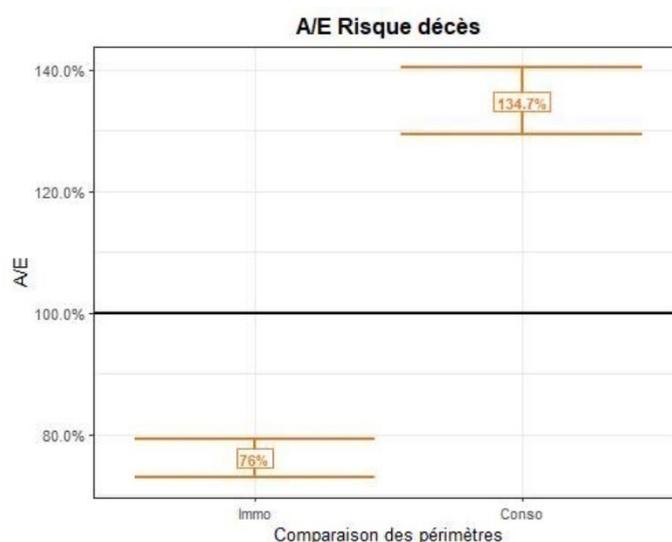


Figure 17: Comparaison des risques décès selon le type de prêt (immobilier ou consommation) sur la base de référence

Le risque décès apparaît plus important pour les prêts à la consommation que pour les prêts immobiliers. Cela peut s'expliquer notamment :

- Par l'accessibilité beaucoup plus large au prêt à la consommation pour la population : les populations sous-jacentes ne sont pas identiques selon le type de prêt. En effet, l'octroi de crédits à la consommation est plus simple que celui des crédits immobiliers et est souscrit par toutes les catégories socio-professionnelles. Les crédits immobiliers concernent majoritairement les catégories socio-professionnelles les plus favorisées, qui sont souvent considérées comme un « meilleur risque » puisqu'ils occupent des activités à moindre risque.

- Par l'absence de sélection médicale pour les prêts à la consommation contrairement aux prêts immobiliers.

Les données du portefeuille étudié dans le cadre de ce mémoire ne permettent pas de distinguer ces deux types de prêts. L'objectif consiste donc à trouver un modèle efficace qui permettra de prédire pour chaque prêt si celui-ci est un prêt immobilier ou un prêt à la consommation.

Pour ce faire, nous appliquerons sur un portefeuille déjà étudié et segmenté selon ce critère le modèle Random Forest en choisissant les hyperparamètres optimaux.

3.2. Pré-processing des données

3.2.1. Variables sélectionnées et traitements effectués

Pour construire le modèle qui nous permettra de déterminer le type de chaque prêt, nous avons sélectionné les variables suivantes, communes à la base de données étudiée et la base de données de référence :

- Le sexe de l'assuré

Les études menées sur la distinction hommes/femmes au regard des prêts mettent en évidence une capacité d'endettement plus faible pour les femmes que pour les hommes, impliquant des montants d'emprunts limités ou un allongement de la durée du prêt.

Selon une étude du site Magnolia.fr réalisée en février 2019, le montant moyen emprunté serait de 174 337€ pour les hommes, contre 156 674€ pour les femmes.

- L'âge de l'assuré à l'adhésion

D'après une étude du courtier Cafpi, l'âge moyen de l'emprunteur primo-accédant est de 34 ans en 2020.

C'est, en effet, ce que l'on retrouve lorsqu'on analyse la partie Immobilier du portefeuille emprunteur de référence, comme le montre le graphique ci-dessous.

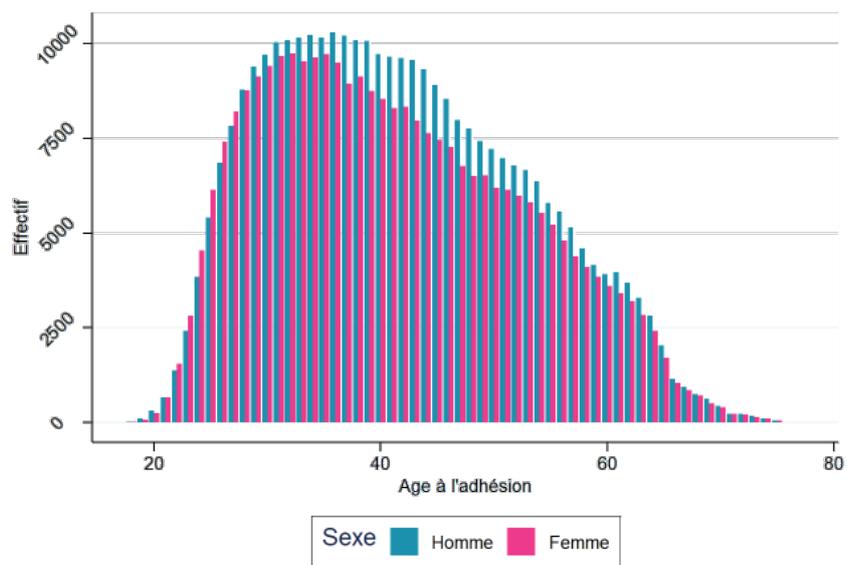


Figure 18: Analyse des âges à l'adhésion pour les prêts immobiliers du portefeuille de référence

Lorsqu'on analyse les âges à l'adhésion pour les prêts à la consommation de ce même portefeuille, on remarque que les adhésions sont plus uniformément réparties entre 25 ans et 50 ans.

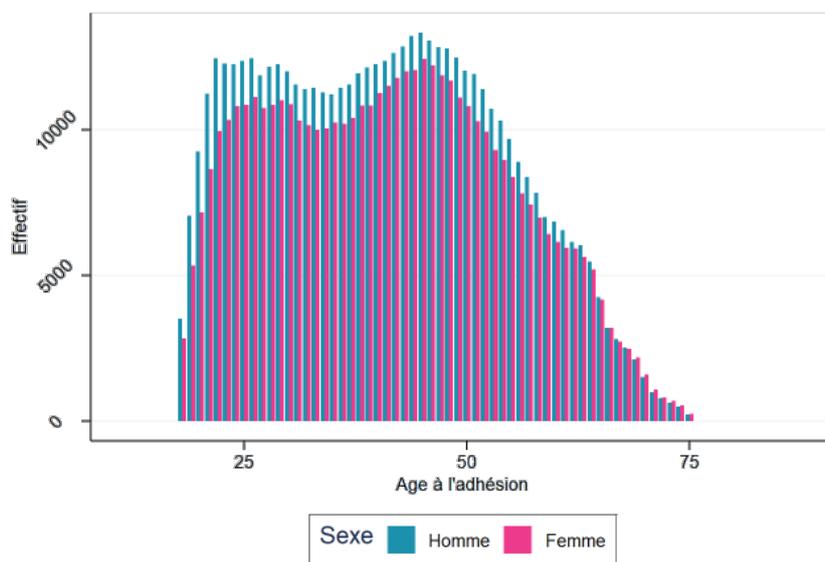


Figure 19: Analyse des âges à l'adhésion pour les prêts à la consommation du portefeuille de référence

- L'année de naissance de l'assuré

En intégrant aux variables l'année de naissance de l'assuré, le modèle peut en déduire la période de souscription, qui peut être plus ou moins propice à la souscription d'un prêt.

Le niveau du taux d'intérêt d'emprunt est, par exemple, un indicateur clé dans le niveau de demandes de prêts. Alors qu'on assiste depuis plusieurs mois à une hausse des taux d'intérêt (70 points de base en moyenne après 6 mois consécutifs de hausse entre février et juillet 2022), l'Observatoire Credit Logement/CSA note un recul du nombre de crédits accordés de 12,8% du mois d'avril à juin 2022 par rapport au même trimestre 2021. En effet, la remontée des taux engendre un coût du crédit plus élevé puisque la part des intérêts augmente. Mécaniquement, cela exclut davantage d'emprunteurs qui dépasseront alors le taux d'endettement de 35% dans le cas où ils n'auraient pas la possibilité d'allonger davantage leur durée de prêt. Dans une période de hausse de l'inflation, les taux d'intérêt devraient continuer à augmenter et influencer encore à la baisse le nombre de prêts souscrits.

- Le montant des capitaux assurés

Le montant des capitaux assurés est une variable essentielle du modèle que l'on va créer pour distinguer les prêts immobiliers des prêts à la consommation.

Sur le portefeuille de référence qui servira de base à notre modèle, le montant moyen assuré pour les prêts immobiliers s'élève à 84 492€ pour les hommes et 79 622€ pour les femmes.

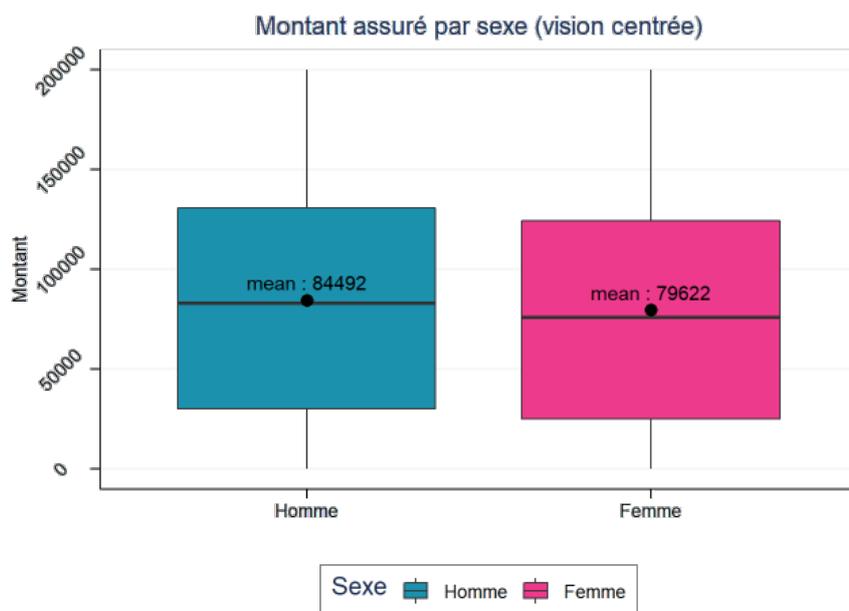


Figure 20: Analyse du montant assuré par sexe pour les prêts immobiliers sur le portefeuille de référence

Pour les prêts à la consommation, le montant assuré moyen est de 7 832€ pour les hommes et 7 390€ pour les femmes.

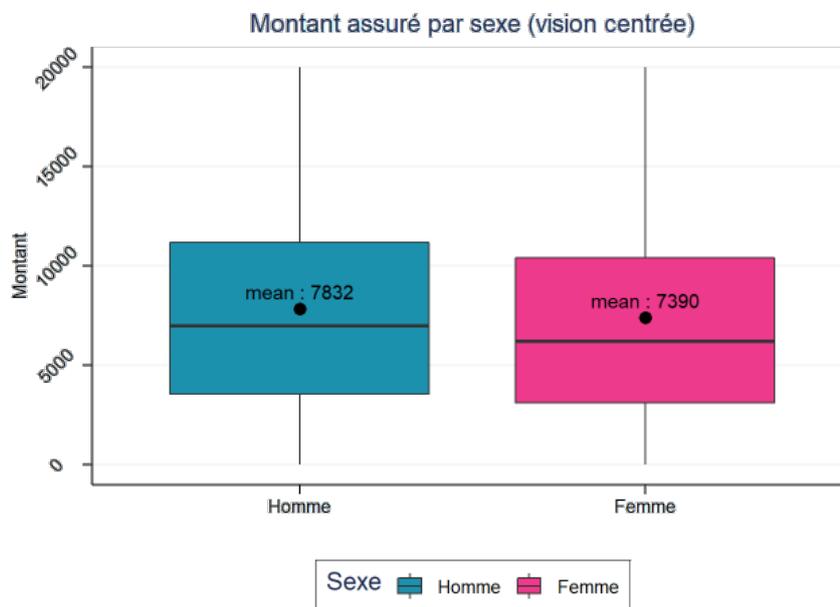


Figure 21: Analyse du montant assuré par sexe pour les prêts à la consommation sur le portefeuille de référence

Le montant du prêt est une variable fortement discriminante dans le modèle machine learning.

- La durée du prêt

Comme le montant des capitaux assurés, la durée du prêt est également une variable clé dans la distinction entre prêts immobiliers et prêts à la consommation.

La durée maximale d'un prêt à la consommation est de 12 ans, contre 25 ans (depuis le 1^{er} janvier 2022) pour les prêts immobiliers. Cependant, la durée des prêts à la consommation couplés à des prêts immobiliers peut aller jusqu'à 35 ans.

La durée de prêt moyen sur le portefeuille des prêts immobiliers de la base de référence de notre étude est de 122 mois, contre une durée moyenne de 37 mois pour les prêts à la consommation.

3.2.2. Découpage

Les informations du portefeuille de référence sont au niveau du prêt : il existe une ligne pour chaque prêt de chaque assuré.

Pour construire notre modèle, le portefeuille de référence est découpé en 2 parties :

- Base Train : base utilisée pour la calibration du modèle,
- Base Validation : base utilisée pour la calibration des hyperparamètres.

Le découpage est effectué de manière aléatoire selon les assurés : 75% dans la base « Train » et 25% dans la base « Validation ».

Le découpage a été effectué avec 100 « graines » différentes. Elles sont utilisées pour s'assurer que les résultats soient reproductibles, c'est-à-dire que les sorties soient les mêmes en cas de réexécution du code. Le découpage final retenu est celui qui répartit les modalités des variables de la manière la plus homogène entre les 2 bases.

3.3. Création d'un modèle Machine Learning et application

3.3.1. Arbre de décision

Avant de construire un modèle Random Forest, nous avons d'abord analysé les résultats donnés par un arbre de décision.

Les arbres de décision sont simples à comprendre, à interpréter, à utiliser, adaptables et puissants. Les arbres simples nécessitent peu de préparations (pas de besoin de normaliser les données numériques).

Une version sans normalisation des données numériques permet une meilleure lecture de l'arbre. Les arbres de Scikit Learn utilisent l'algorithme CART (Classification And Regression Tree) qui produit des arbres binaires : les nœuds non terminaux ont deux fils.

La méthode de plantation d'un arbre est la suivante : à chaque nœud, l'algorithme CART sépare la base « Train » en deux sous-ensembles en utilisant une variable k et un seuil t_k de telle sorte que le couplet (k, t_k) produit le sous-ensemble le plus pur (c'est-à-dire avec le moins de classes possibles dans sa population) pondéré par la taille.

Minimisation de la fonction de coût :

$$J(k, t_k) = \frac{m_{gauche}}{m} * G_{gauche} + \frac{m_{droite}}{m} * G_{droite}$$

G_{gauche} : impurété du sous ensemble de gauche

m_{gauche} : nombre d'observations du sous-ensemble de gauche

L'algorithme CART est un algorithme GLOUTON : le partage optimal s'effectue à chaque niveau mais ne se soucie pas de savoir si le partage des niveaux plus bas conduit à une impurété aussi faible que possible.

Un algorithme glouton ne garantit pas la solution optimale mais produit des résultats très satisfaisants.

Pour effectuer les prédictions, l'algorithme traverse l'arbre modélisé et donne les probabilités associées au nœud terminal.

Pour limiter le « surapprentissage » des arbres, appelé overfitting, il faut contrôler les hyperparamètres :

- critère : utilisation d'une des deux mesures de la variance totale pour les k classes à chaque étape de l'élaboration de l'arbre - Gini index ou entropie
- max_depth : profondeur maximale de l'arbre
- min_samples_split : nombre d'observations minimum que doit comporter un nœud pour que celui-ci puisse être divisé
- min_sample_leaf : nombre minimal d'observations qu'un nœud terminal doit avoir

3.3.2. Random Forest

3.3.2.1. Présentation

Les forêts aléatoires (ou Random forest) génèrent un grand nombre d'arbres de décision, combinés à la fin du processus. Elles appartiennent à la famille des modèles de bagging. Le bagging permet de réduire la variance des modèles en faisant la moyenne des prévisions de plusieurs modèles indépendants.

Un modèle avec une variance trop élevée s'adapte aux données sur lesquelles il est paramétré mais ne se généralise pas bien (cela engendre de l'overfitting).

Random Forest est composé d'un nombre d'arbres défini. Chaque arbre est modélisé de manière indépendante (l'algorithme est donc parallélisable) avec un échantillon « bootstrap » des données, c'est-à-dire un échantillonnage basé sur le prélèvement de plusieurs échantillons avec remise (appelés rééchantillonnages).

A noter qu'à chaque étape, la construction d'un nœud de l'arbre se fait sur un sous-ensemble de variables tirées aléatoirement : il y a donc une part de hasard supplémentaire par rapport au bagging standard, ce qui induit une diminution de la variance mais une augmentation du biais.

Les hyperparamètres optimisés sont :

- num.trees: nombre d'arbres dans la forêt
- mtry: nombre de variables à potentiellement séparer dans chaque nœud
- criterion: critère de partage d'un nœud
- max.depth: profondeur maximale de chaque arbre

3.3.2.2. *Choix des hyperparamètres*

Pour sélectionner les hyperparamètres optimaux de notre modèle qui permettra de répartir les prêts entre consommation et immobilier, nous lançons un nombre multiple de scénarios :

- Nombre d'arbres : de 10 à 200 avec un pas de 10,
- Nombre de variables à séparer : de 2 à 5,
- Profondeur maximale : de 5 à 100 avec un pas de 5.

Nous appelons dans notre étude « positif » un prêt qui est immobilier et « négatif » un prêt qui est à la consommation.

Un taux de précision de la classification est calculé pour chaque combinaison du modèle, ainsi qu'un nombre de positifs, un nombre de négatifs et un nombre de faux positifs et de faux négatifs.

Les résultats de ces combinaisons sont enregistrés dans un tableau qui se présente sous la forme suivante :

nb_arbres	nb_noeuds	prof	taux_err	nb_positifs	nb_negatifs	f_positifs	f_negatifs
10	2	5	12,44	215 209	200 892	15 877	35 893
10	2	10	11,88	214 317	201 784	14 270	35 178
10	2	15	11,46	215 845	200 256	14 160	33 540
10	2	20	11,31	216 307	199 794	14 078	32 996
10	2	25	11,44	217 969	198 132	15 178	32 434
10	2	30	11,75	218 422	197 679	16 037	32 840
10	2	35	11,75	219 967	196 134	16 821	32 079
10	2	40	11,83	219 354	196 747	16 684	32 555
10	2	45	11,76	219 001	197 100	16 347	32 571
10	2	50	11,71	219 528	196 573	16 521	32 218
10	2	55	11,77	219 026	197 075	16 386	32 585
10	2	60	11,77	220 127	195 974	16 937	32 035
10	2	65	11,78	218 686	197 415	16 244	32 783
10	2	70	11,8	220 563	195 538	17 209	31 871
10	2	75	11,83	219 629	196 472	16 822	32 418
10	2	80	11,76	219 797	196 304	16 756	32 184
10	2	85	11,8	219 552	196 549	16 707	32 380
10	2	90	11,77	219 698	196 403	16 721	32 248
10	2	95	11,78	219 427	196 674	16 600	32 398
10	2	100	11,82	218 697	197 404	16 331	32 859
10	3	5	12,36	216 720	199 381	16 461	34 966
10	3	10	11,74	215 604	200 497	14 611	34 232
10	3	15	11,41	216 846	199 255	14 559	32 938
10	3	20	11,31	217 973	198 128	14 913	32 165
10	3	25	11,76	220 138	195 963	16 919	32 006
10	3	30	12,52	224 038	192 063	20 461	31 648
10	3	35	12,77	226 536	189 565	22 229	30 918
10	3	40	12,91	227 034	189 067	22 754	30 945
10	3	45	12,92	226 819	189 282	22 682	31 088
10	3	50	12,9	227 012	189 089	22 731	30 944
10	3	55	12,89	226 755	189 346	22 593	31 063
10	3	60	12,89	227 049	189 052	22 735	30 911
10	3	65	12,94	227 054	189 047	22 828	30 999
10	3	70	12,94	226 898	189 203	22 748	31 075
10	3	75	12,92	227 101	189 000	22 809	30 933

Figure 22: Tableau de résultats des combinaisons des hyperparamètres

Le taux d'erreur optimal est obtenu avec les hyperparamètres suivants :

- Nombre d'arbres : 150
- Nombre de nœuds : 4
- Profondeur : 20

Si l'on étudie l'évolution du taux d'erreur en fonction de l'évolution de chaque hyperparamètre, on obtient les graphiques suivants :

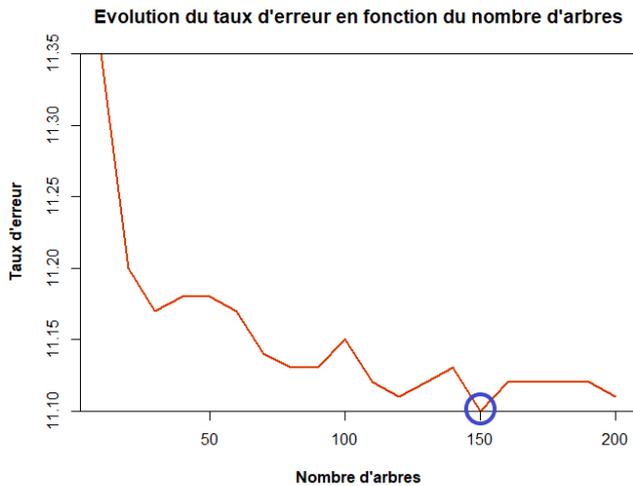


Figure 23: Evolution du taux d'erreur en fonction du nombre d'arbres

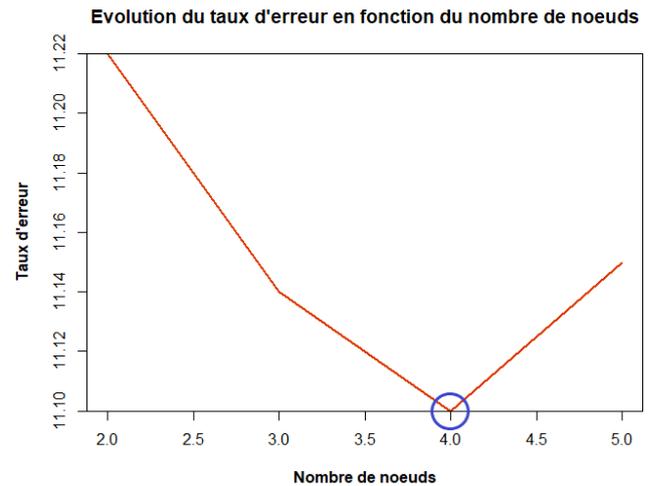


Figure 24: Evolution du taux d'erreur en fonction du nombre de noeuds

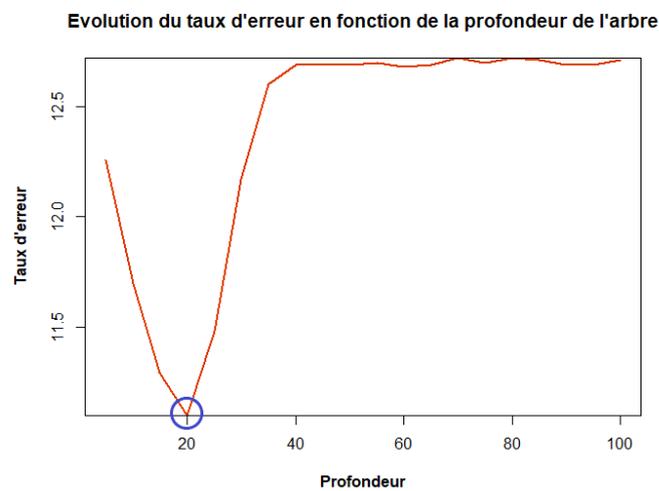


Figure 25: Evolution du taux d'erreur en fonction de la profondeur de l'arbre

3.3.2.3. Doubles conditions de la modélisation

A travers le choix des hyperparamètres du modèle, nous cherchons à maximiser le taux de précision de la classification entre prêts immobiliers et prêts à la consommation. Il s'agit d'un indicateur de performance du modèle qui donne le pourcentage de prêts correctement classés par le modèle.

Ce n'est toutefois pas la seule condition pour affirmer que notre modèle est robuste. En effet, une autre variable rentre en considération a posteriori : le décès ou non de l'assuré.

Comme expliqué dans le paragraphe 2.1, les taux de décès sur le segment des prêts à la consommation sont plus élevés que sur le segment des prêts immobiliers. De ce fait, une mauvaise classification des prêts entre immobilier et consommation pourrait avoir des conséquences ensuite sur la construction de notre table de mortalité :

- Trop de « faux positifs » pourraient détériorer les lois du segment prêts immobiliers car un nombre plus important de décès liés aux prêts à la consommation seraient pris en compte dans les lois du segment prêts immobiliers,
- Trop de « faux négatifs » pourraient améliorer à tort les lois du segment prêts à la consommation puisque des prêts immobiliers donc potentiellement moins touchés par le décès seraient intégrés dans les lois du segment prêts à la consommation.

Une fois notre modèle correctement calibré, nous obtenons sur la base « Validation » un taux de précision de 89% avec 7% de « faux positifs » et 16% de « faux négatifs ». En faisant tourner le modèle sur la base entière, nous obtenons un taux de précision de 92%. Le taux est plus élevé que pour la base « Train » seule puisque la base entière contient la base « Train » et la base « Validation » ; il est donc biaisé puisqu'il est appliqué à la base sur laquelle le modèle a appris.

3.4. Application du modèle

3.4.1. Ajustement du montant des capitaux assurés

Les prêts souscrits dans la base de données étudiée sont largement composés de prêts sociaux (72%) et de prêts bancaires (19%). Les prêts sociaux ont, en général, un montant moyen de capitaux assurés beaucoup plus faible que celui des prêts bancaires ; il s'agit également en majorité de prêts immobiliers, puisque ce type de prêts est mis en place pour permettre aux foyers à revenus modestes d'accéder à la propriété.

Etant donné que le portefeuille de référence, sur lequel notre modèle est construit, se compose uniquement de prêts bancaires, le montant moyen des capitaux assurés sera probablement plus élevé que sur notre base.

Après avoir comparé ce montant moyen de capitaux assurés entre les deux bases, nous devons appliquer aux capitaux assurés de la base étudiée un coefficient permettant d'avoir des montants de capitaux assurés à niveau équivalent entre les deux bases.

Pour calculer ce coefficient, qui sera égal au rapport entre le montant moyen des capitaux assurés sur le portefeuille de référence (a) et le montant moyen des capitaux assurés sur la base étudiée

(b), nous devons prendre en compte la composition de ces deux portefeuilles : comme indiqué ci-dessus, les prêts de la base de données étudiée sont en majorité des prêts immobiliers, tandis qu'à contrario, l'étude sur la base de données de référence révèle qu'elle est plus largement composée de prêts à la consommation (62%) que de prêts immobiliers (38%).

Un calcul du rapport des montants moyens de capitaux assurés sans retraitement engendrera un biais puisque les capitaux assurés des prêts immobiliers sont en général plus élevés que ceux des prêts à la consommation.

A dire d'expert, nous pouvons considérer que la base étudiée est composée à 90% de prêts immobiliers, et 10% de prêts à la consommation. Pour obtenir des bases homogènes et ainsi pouvoir calculer notre coefficient, nous créons une base de référence qui respectent cette même proportion.

Pour cela, nous sélectionnons la totalité des prêts immobiliers de la base de référence, et considérons que cela représente 90% de notre base. Pour les 10% restants, qui seront donc des prêts à la consommation, nous faisons un tirage aléatoire parmi ces prêts à la consommation. Nous répétons cette opération 100 fois pour avoir 100 bases de comparaison différentes, en calculant à chaque fois une moyenne des montants assurés sur ces bases.

Le montant moyen de capitaux assurés de la base de référence retenu pour le calcul du coefficient est calculé comme étant la moyenne des moyennes issues des 100 simulations précédentes.

Nous appliquons finalement le coefficient a/b au montant des capitaux assurés de la base étudiée.

3.4.2. Résultat sur la base de données étudiée

Nous appliquons le modèle Random Forest construit à partir de la base de données de référence sur la base de données étudiée, en utilisant le montant des capitaux auquel on applique le coefficient précédemment calculé.

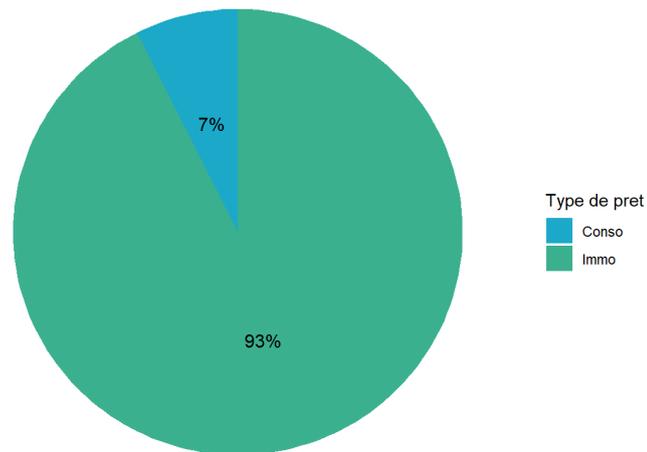


Figure 26: Répartition par type de prêt suite à l'application du modèle Random Forest

Le portefeuille est composé majoritairement de prêts immobiliers. Cela est cohérent avec l'essence même des prêts sociaux qui consiste à faciliter l'accès à la propriété pour tous.

Comme sur la base de données de référence, nous calculons sur notre base de données étudiée après application du modèle des ratios Observés/Attendus ($\frac{A}{E}$) qui permettent d'estimer le risque décès pour les prêts immobiliers puis pour les prêts à la consommation (A) par rapport au risque décès moyen du portefeuille global (E). Le graphique ci-dessous montre sur cette base de données également que le ratio $\frac{A}{E}$ des prêts immobiliers est inférieur à 100% alors que celui des prêts à la consommation est supérieur à 100%. Nous pouvons donc confirmer notre hypothèse de départ selon laquelle les prêts à la consommation ont un risque moins favorable que les prêts immobiliers.

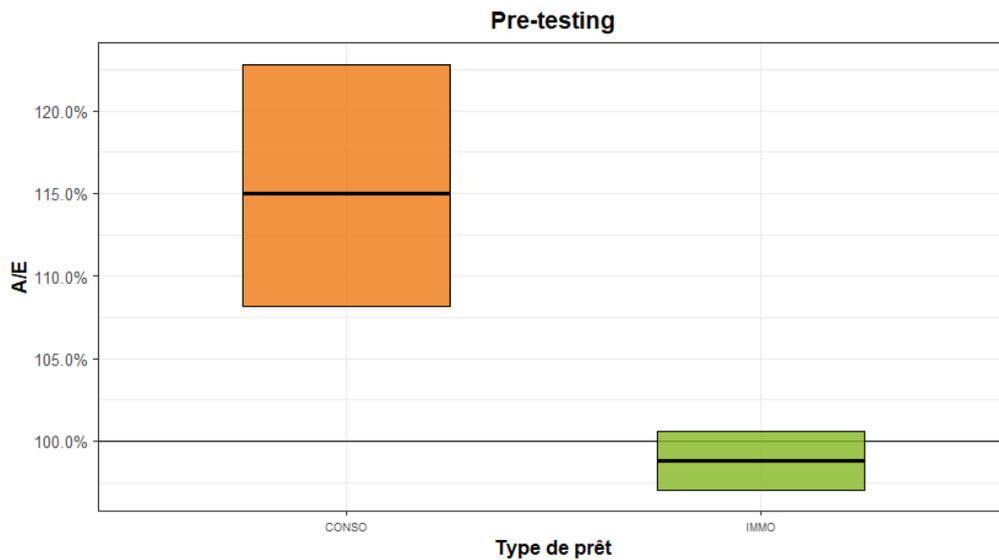


Figure 27: Comparaison des risques décès selon le type de prêt (immobilier ou consommation) sur la base de données étudiée

L'application du modèle de machine learning *Random Forest* a été particulièrement efficace pour permettre la classification de chaque prêt entre prêt immobilier et prêt à la consommation, avec un taux de précision élevé de 89% sur la base « Validation ». Ce modèle basé sur un assemblage d'arbres de décision nous a permis de mettre en exergue la prépondérance des prêts immobiliers dans le portefeuille étudié. Etant donné qu'ils possèdent des niveaux de risques différents, les prêts immobiliers et les prêts à la consommation seront séparés en deux bases distinctes pour ne pas biaiser l'analyse du risque de mortalité dans la suite de l'étude.

PARTIE 4 – CONSTRUCTION DE LOIS D’EXPERIENCE DU RISQUE DECES

La construction de nos tables d’expérience est réalisée en cinq étapes. Après avoir déterminé la période d’observation sur laquelle nous nous baserons pour l’étude, des taux de mortalité bruts issus directement des données à notre disposition seront calculés. Pour obtenir une courbe de taux lissée, nous utiliserons le package R *Mortality Smooth*, qui permet de lisser les taux bruts à l’aide d’un empilement de courbes, appelées « splines ». L’extrapolation des taux sera ensuite faite en utilisant le modèle de Brass. Enfin, nous vérifierons la qualité d’ajustements à travers le calcul d’intervalles de confiance et l’étude des résidus.

Il nous est possible désormais de séparer les prêts immobiliers des prêts à la consommation. Comme les produits créés sur le segment des prêts sociaux sont majoritairement des produits couvrant les prêts immobiliers et que nous avons mis en évidence un risque décès moins favorable pour les prêts à la consommation que pour les prêts immobiliers, nous nous concentrerons dans la suite de cette partie sur le segment immobilier qui représente la majeure partie de notre portefeuille et excluons de la base de données les prêts à la consommation.

4.1. Détermination de la période d’observation

4.1.1. Pré-testing sur les données du portefeuille

Les données disponibles ne sont pas uniformes selon les âges à l’adhésion, les années calendaires ou encore les années de souscription. Une étape de « pré-testing » est utile pour évaluer l’impact des modalités d’une variable sur le risque décès étudié. Nous pourrions par la suite calibrer le périmètre d’étude du portefeuille dans lequel les comportements sont homogènes.

Pour cela, des ratios $\frac{A}{E}$ sont calculés et permettent de mettre en évidence de potentielles irrégularités du portefeuille selon les modalités de chaque variable. Des intervalles de confiance à 95% sont également calculés : moins il y a d’observations par modalité plus l’intervalle de confiance sera important.

Si l’on prend comme exemple la variable duration du prêt, on observe que les trois premières années de duration des prêts présentent un risque de mortalité plus faible que le risque de mortalité moyen du portefeuille. Cela s’explique notamment par la sélection médicale à la

souscription, qui permet de par les questionnaires médicaux et autres procédures à remplir pour souscrire à une assurance emprunteur, de contrôler les risques assurés du portefeuille. Ces trois premières durations feront l'objet d'un retraitement spécifique dans les lois de mortalités finales pour prendre en compte leur risque de mortalité moindre par rapport au risque moyen du portefeuille.

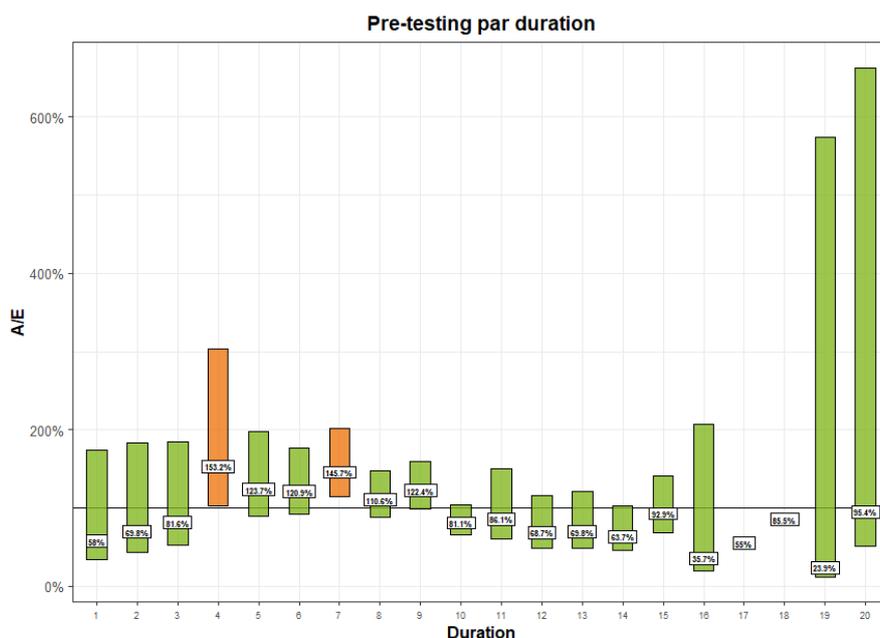


Figure 28: Comparaison des risques décès selon la durée du prêt sur la base de données étudiée

Nous réalisons cette même analyse pour les modalités de chaque variable impactante dans la construction de nos lois, avec et sans segmentation par sexe :

- Années calendaires,
- Années de souscription du prêt.

Nous retenons finalement le périmètre d'étude suivant :

- Années calendaires : 2013-2017,
- Durations : supérieures à 3 ans.

Les graphiques de pré-testing du segment immobilier des prêts sociaux sont disponibles en Annexe 1.

4.1.2. Analyse des expositions et des sinistres

Nous notons $E_x = \sum_i E_{i,x}$ l'exposition au risque décès à l'âge x avec, pour chaque individu i , $E_{i,x}$ le nombre de jours pendant lesquels l'individu i est observé à l'âge x .

Pour la période d'observation retenue (voir section 4.1.1), l'exposition du portefeuille est représentée dans les graphiques suivants.

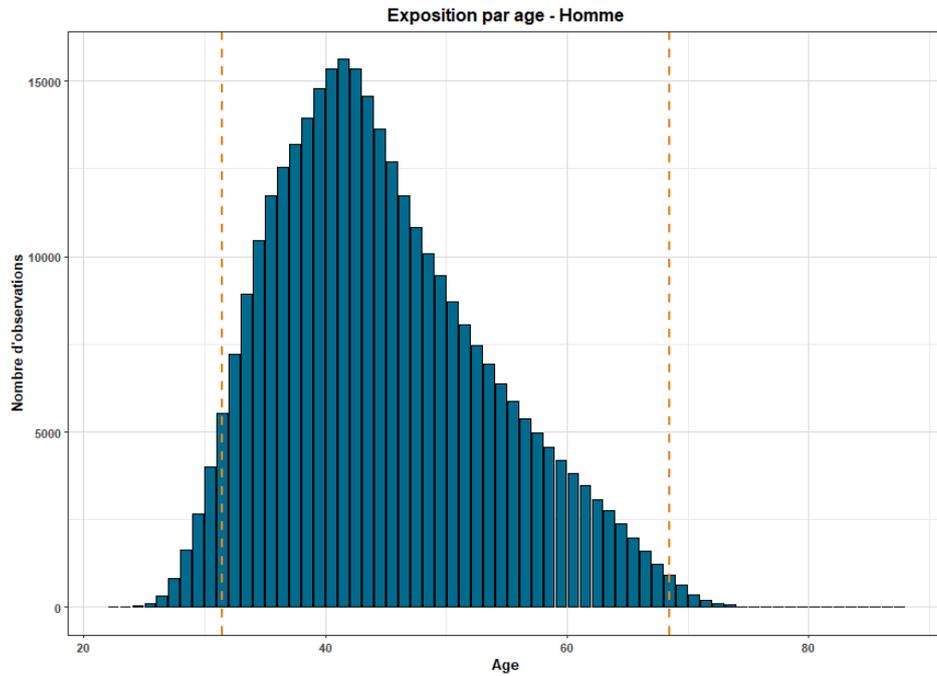


Figure 29: Analyse de l'exposition par âge - Homme

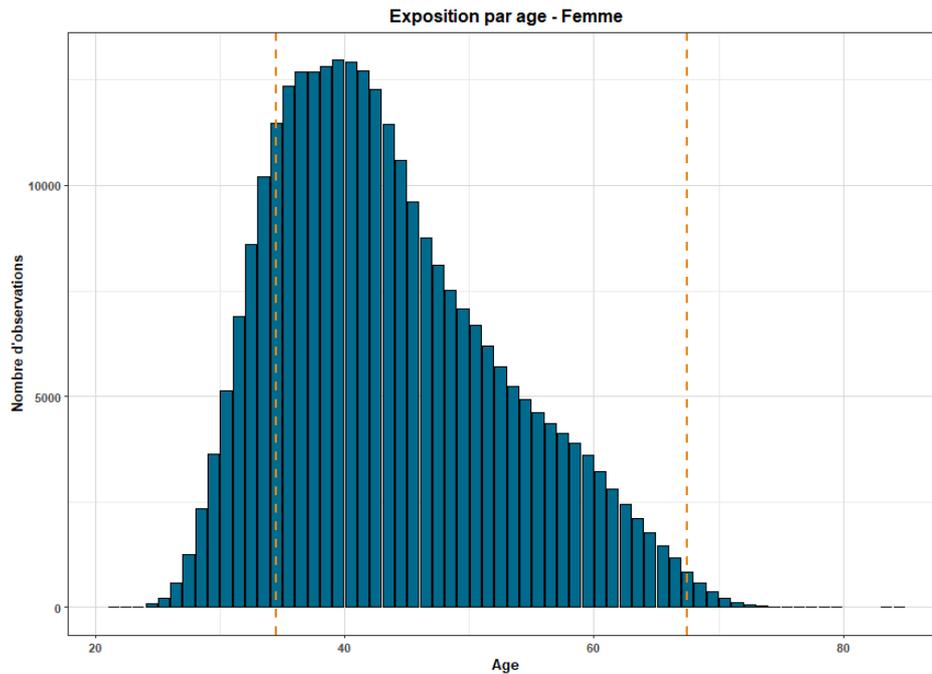


Figure 30: Analyse de l'exposition par âge - Femme

Nous observons que la majeure partie de l'exposition se situe sur les âges 30-70 ans. Sur les âges jeunes inférieurs à 30 ans et les grands âges supérieurs à 65 ans, nous utiliserons des méthodes d'extrapolation pour pallier au manque d'observations.

Nous notons $L_{x,t}$ le nombre d'individus d'âge x vivant au 1^{er} janvier de l'année calendaire t . Le nombre d'individus d'âge x sur l'ensemble de la plage d'observation est donc noté $L_x = \sum_t L_{x,t}$. Dans la même logique, nous notons $D_{x,t}$ le nombre de décès d'individus d'âge x durant l'année calendaire t . Le nombre de décès d'individus d'âge x sur l'ensemble de la plage d'observation est donc noté $D_x = \sum_t D_{x,t}$.

Pour la période d'observation retenue (voir section 4.1.1), le nombre de décès est représenté dans les graphiques suivants.

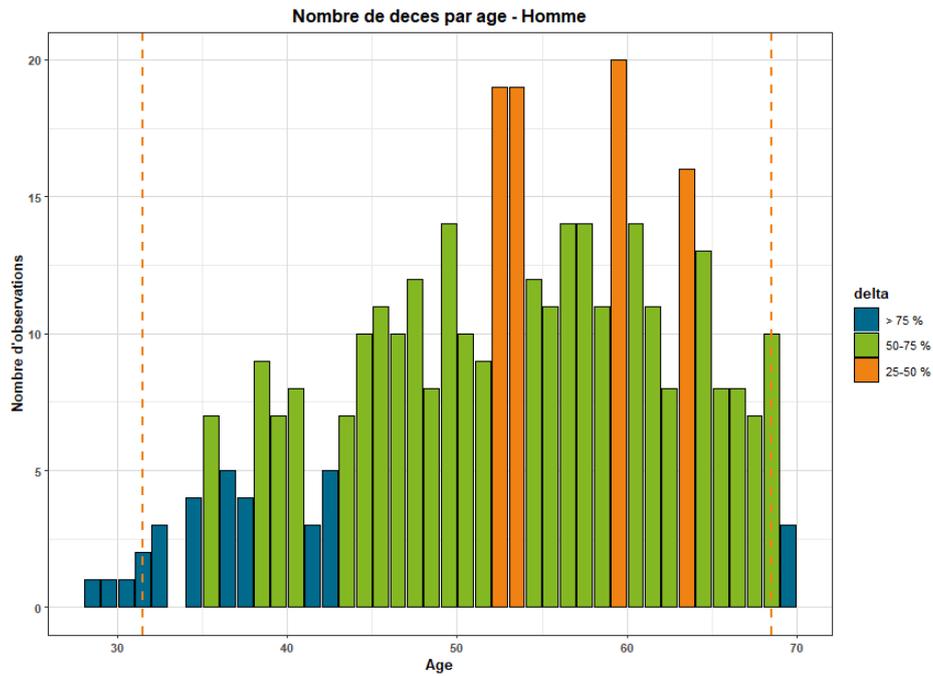


Figure 31: Analyse du nombre de décès par âge - Homme

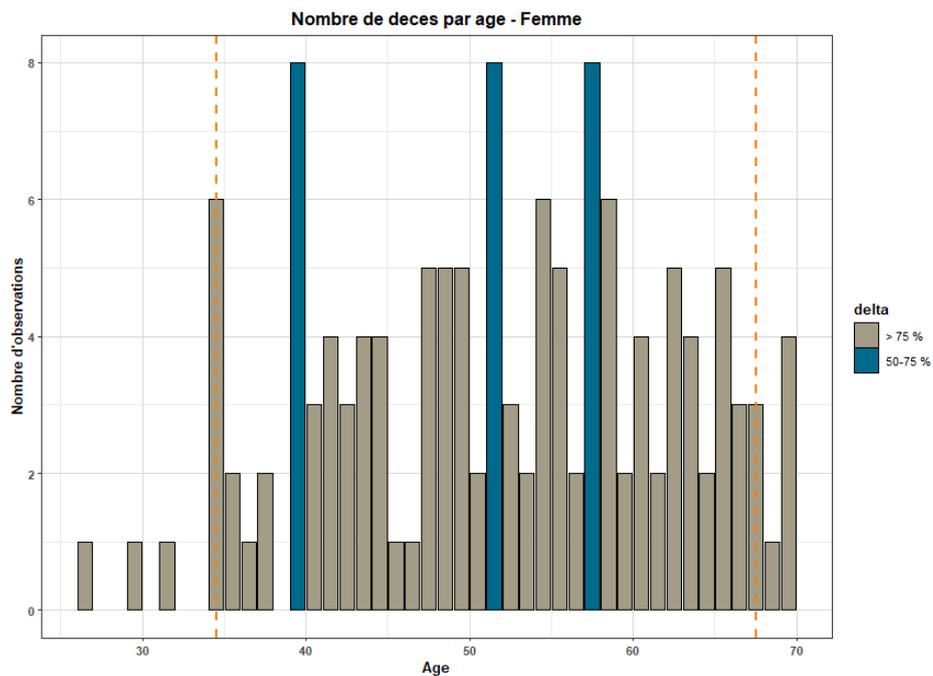


Figure 32: Analyse du nombre de décès par âge - Femme

Pour les hommes comme pour les femmes, 90% des décès observés se situent sur la tranche d'âge 34-65 ans. Nous observons chez les hommes seulement 3 décès avant 31 ans, et chez les femmes seulement 3 décès avant 34 ans. Nous n'observons plus aucun décès après 70 ans.

Nous avons toutefois beaucoup moins d'observations chez les femmes que chez les hommes, ce qui pourra entraîner une variation plus forte dans les taux de mortalité calculés chez les femmes dans la suite de notre étude.

Après analyse, nous sélectionnons comme tranches d'âges :

- 31-68 ans pour les hommes,
- 34-67 ans pour les femmes.

4.2. Estimation des taux de mortalité bruts

A partir des données du portefeuille retenues, nous calculons des taux de mortalité bruts grâce à l'estimateur de Hoem.

Le principe de l'estimateur de Hoem consiste à construire des tranches d'observation constantes : dans notre cas, nous retiendrons les tranches fonction de l'âge.

Sous l'hypothèse que l'intensité de mortalité est constante par morceaux entre deux âges entiers consécutifs, les estimateurs du maximum de vraisemblance des intensités de mortalités sont définis par :

$$\widehat{\mu}_x = \frac{D_x}{E_x}$$

En utilisant la constance des taux de mortalités instantanés, nous obtenons :

$$\widehat{q}_x = 1 - \exp\left(-\frac{D_x}{E_x}\right)$$

$\widehat{\mu}_x$ est l'estimateur de vraisemblance. Il est asymptotiquement sans biais et distribué selon une loi normale.

Nous pouvons en déduire l'espérance et la variance suivante :

$$E[\widehat{\mu}_x] = E\left[\frac{D_x}{E_x}\right] = \frac{E[D_x]}{E_x} = \mu_x$$

$$Var[\widehat{\mu}_x] = Var\left[\frac{D_x}{E_x}\right] = \frac{Var[D_x]}{E_x^2} = \frac{\mu_x}{E_x}$$

L'intervalle de confiance à 95% de cet estimateur est obtenu grâce au théorème central limite :

$$IC_{95\%} = \frac{D_x}{E_x} \pm 1.96 \times \frac{\sqrt{D_x}}{E_x}$$

Cet estimateur est particulièrement pertinent pour les bases de données importantes. Il permet d'estimer les taux bruts à l'aide d'un rapport entre le nombre de sorties observées et l'exposition.

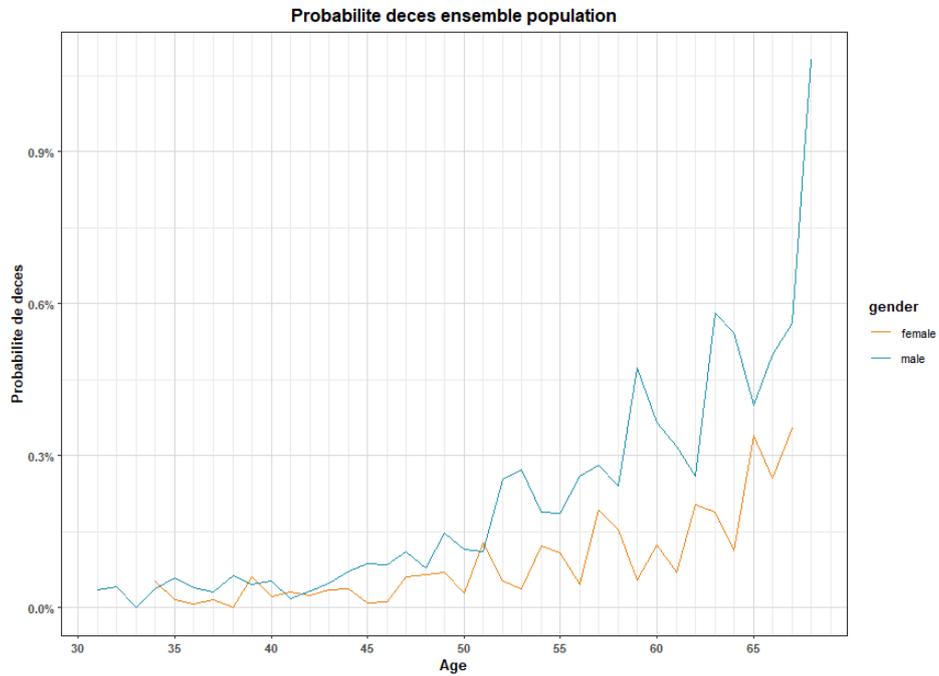


Figure 33: Taux bruts de mortalité par l'estimateur de Hoem

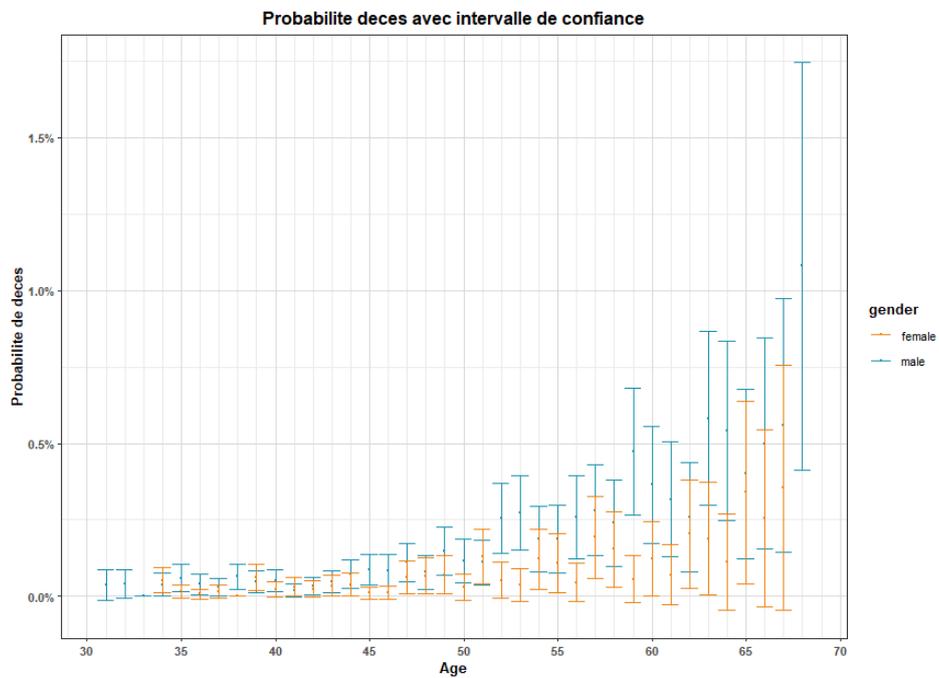


Figure 34: Taux bruts de mortalité par l'estimateur de Hoem avec intervalle de confiance

On observe globalement une croissance des taux de mortalité avec l'âge : plus l'assuré vieillit, plus sa probabilité de décès augmente.

L'irrégularité des taux bruts de mortalité nous conduit à l'utilisation d'une méthode de lissage pour obtenir une courbe de taux de mortalité exploitable.

4.3. Méthodes de lissage

Les taux de mortalité bruts d'expérience sont calculés pour les âges où nous avons des observations de décès. Ils forment une courbe de mortalité irrégulière qui ne peut être exploitée directement. Plusieurs ajustements vont donc être effectués.

4.3.1. Utilisation du package R *Mortality Smooth*

Nous utilisons dans un premier temps le package R *Mortality Smooth*. Cette méthode permet de lisser à l'aide d'un empilement de courbes, appelées « splines », les nombres de décès modélisés par une loi de Poisson :

$$D_x \sim P(E_x \times \mu_x)$$

Mortality Smooth fractionne la courbe des taux bruts en sous-intervalles puis ajuste les taux de chaque morceau grâce à un polynôme, tout en respectant des contraintes de continuité et d'égalité des pentes et des courbures pour les raccordements aux points de jonction.

Par défaut, le polynôme est cubique, une pénalité est appliquée en fonction du nombre de degrés du polynôme et un nœud apparaît tous les cinq points de données. Ces paramètres peuvent être ajustés.

Cette méthode démontre plusieurs avantages pour notre étude par rapport à d'autres méthodes de lissage comme Whittaker Henderson par exemple :

- Elle contraint le nombre de sinistres obtenu avec la loi lissée à être égal au nombre de sinistres observés.
- Son approche est continue.
- Elle est paramétrique et les paramètres peuvent être calibrés de manière optimale. Cette calibration est automatisée dans le package sans action de l'utilisateur.

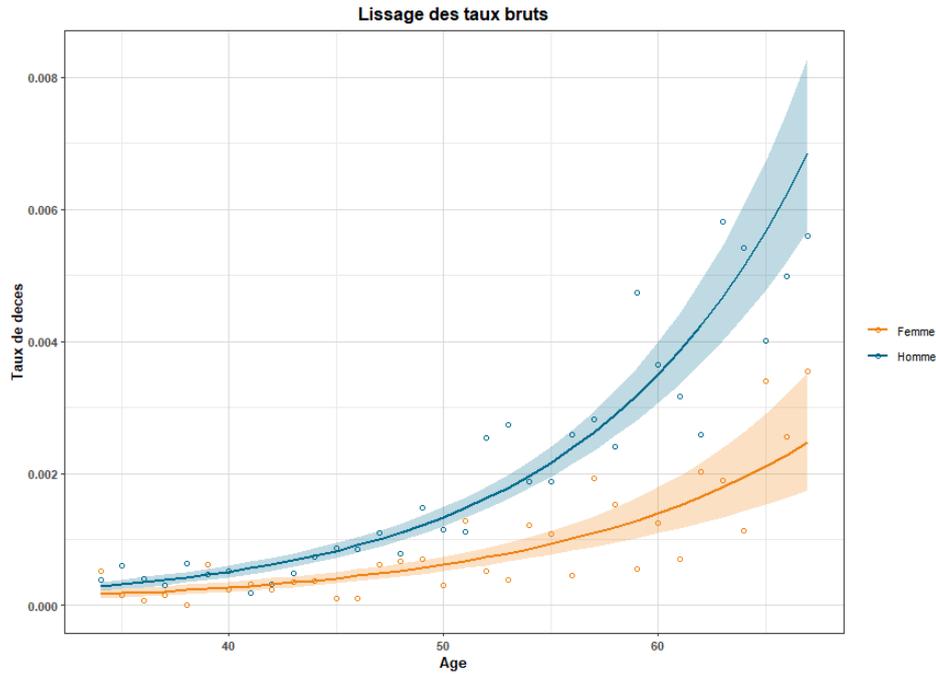


Figure 35: Lissage des taux bruts de mortalité à l'aide du package Mortality Smooth

4.3.2. Lissage relationnel de Brass

Pour permettre ensuite l'extrapolation aux âges élevés des lois de décès, nous avons recours à l'utilisation du modèle de Brass. Cette méthode s'appuie sur un positionnement des lois de décès par rapport à une table de référence. Elle effectue pour cela une régression linéaire des taux bruts sur les logit de la table de référence.

Les \widehat{q}_x et les q_x^{ref} sont ainsi liés par la formule :

$$\text{logit } \widehat{q}_x = \alpha + \beta \times \text{logit } q_x^{ref} + \epsilon_x$$

Avec ϵ_x le terme d'erreur.

Les paramètres α et β sont estimés grâce à la méthode des moindres carrés. Nous pouvons ensuite calculer les taux ajustés de la manière suivante :

$$\widehat{q}_x = \frac{\exp(\alpha + \beta \times \text{logit } q_x^{ref})}{1 + \exp(\alpha + \beta \times \text{logit } q_x^{ref})}$$

Dans notre cas, nous utiliserons comme table de référence la loi de mortalité de la population française disponible dans la « Human Mortality Database » (HMD) sur la période 2016-2018. La base de données HMD fournit par pays les expositions et les décès âge par âge et par sexe.

Avant utilisation, cette loi de référence est dans un premier temps lissée à l'aide du package *Mortality Smooth* pour corriger les potentielles irrégularités qu'on peut observer, notamment aux âges élevés. Toutefois, dans notre cas, au vu de la taille importante de la base de données à notre disposition, la volatilité des taux est plus faible et les courbes quasiment lissées.

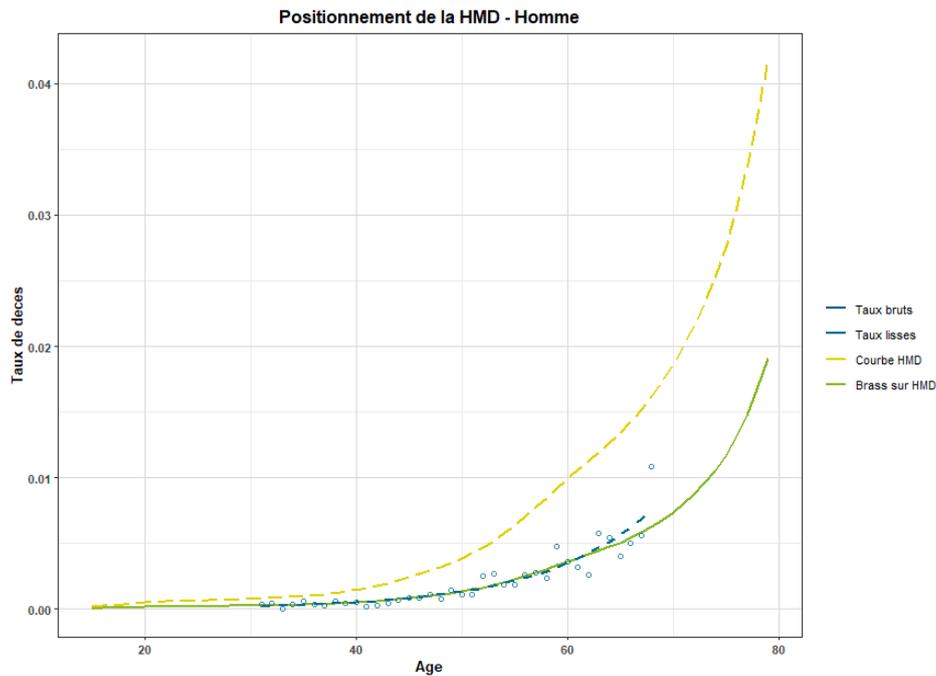


Figure 36: Positionnement des lois créées par rapport à la HMD (homme)

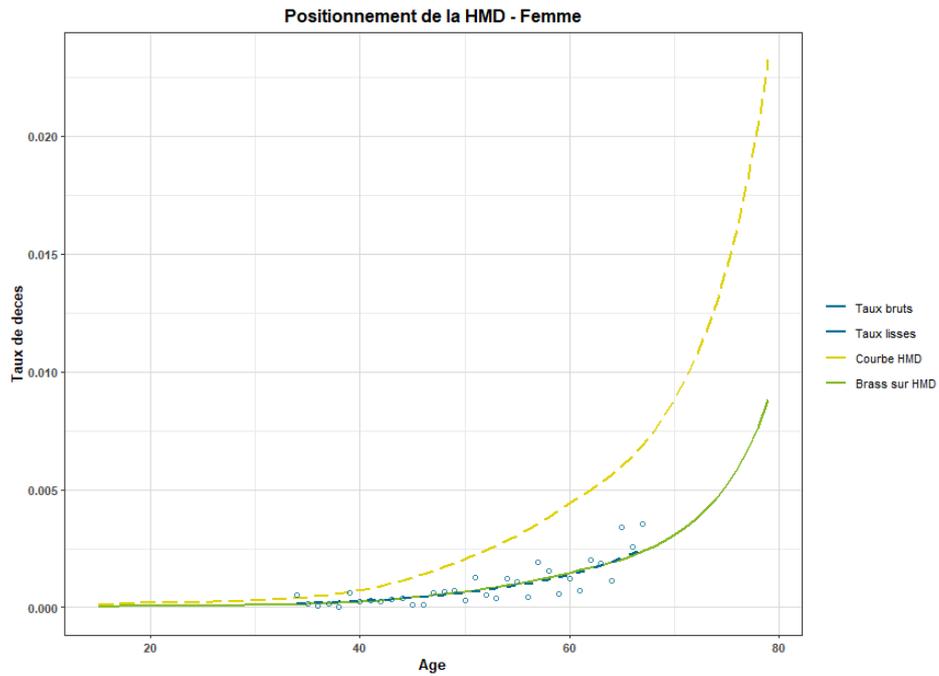


Figure 37: Positionnement des lois créés par rapport à la HMD (femme)

Les taux lissés et extrapolés à l'aide de la méthode de Brass sont très proches des taux lissés obtenus après application du package *Mortality Smooth*, puis sont extrapolés ensuite selon la loi de référence.

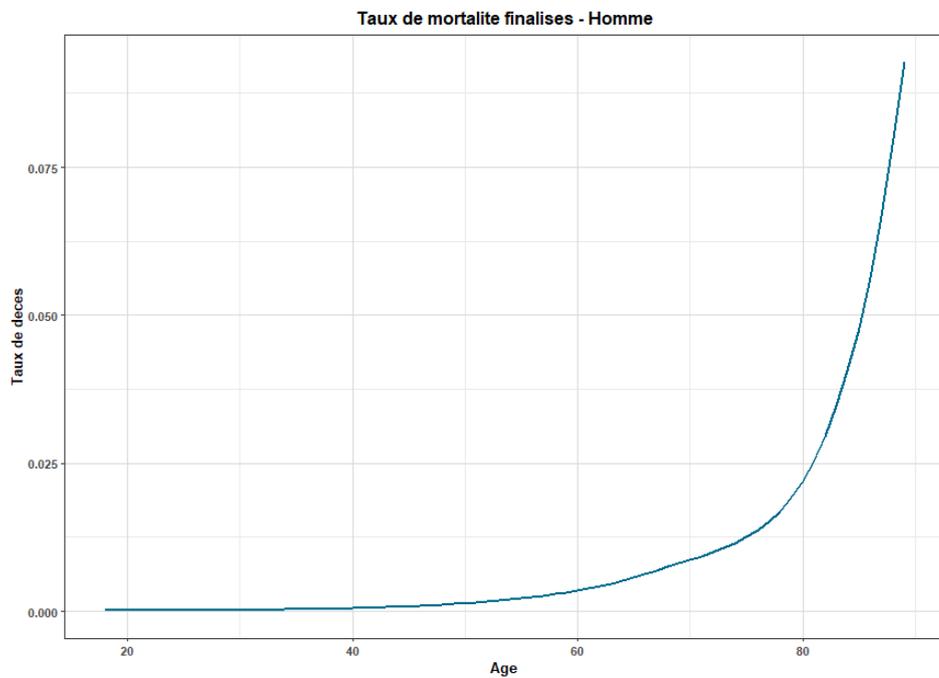


Figure 38: Lois finales (homme)

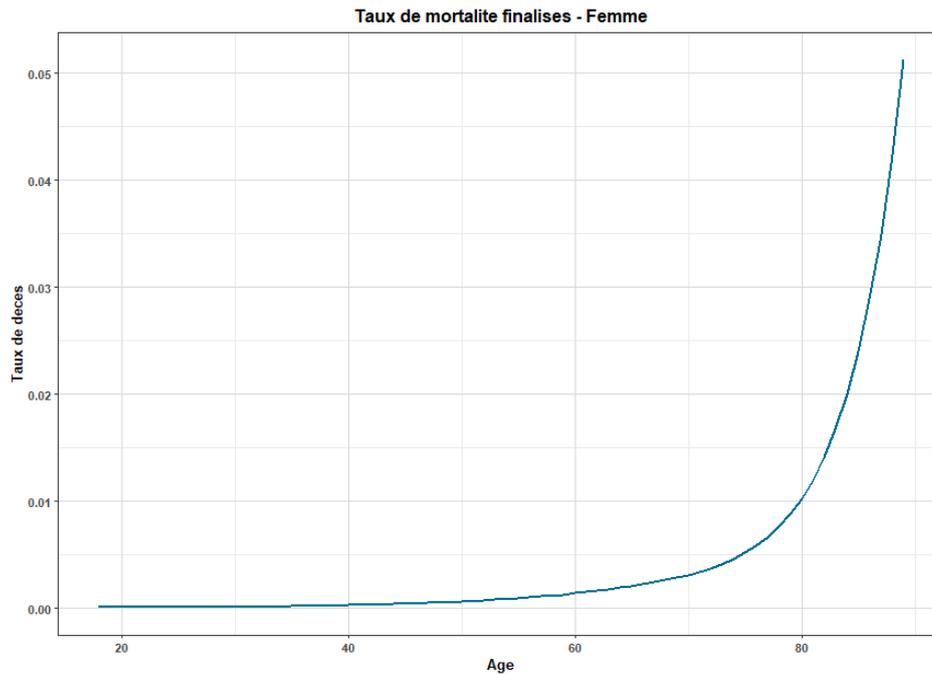


Figure 39: Lois finales (femme)

4.4. Validation des modèles

Pour valider la qualité des ajustements, nous réalisons plusieurs tests statistiques. Ils nous permettront de nous assurer que les taux de mortalité lissés ne soient pas trop éloignés des taux de mortalité bruts et d'évaluer la pertinence des approches concernant la construction de la table de mortalité d'expérience.

4.4.1. Intervalles de confiance ponctuels

Nous comparons dans un premier temps les décès attendus avec la modélisation et les décès observés âge par âge. Pour cela, nous utilisons l'approximation de la loi de décès par une normale et l'estimation d'un intervalle de confiance à 95%.

Nous supposons que le nombre de décès suit une loi normale tel que :

$$D_x \sim \mathcal{N}(E_x q_x, E_x q_x (1 - q_x))$$

Nous pouvons alors approximer des intervalles de confiance ponctuels à 95% avec la formule suivante :

$$(E_x \widehat{q}_x - 1,96 \sqrt{E_x \widehat{q}_x (1 - \widehat{q}_x)}; E_x \widehat{q}_x + 1,96 \sqrt{E_x \widehat{q}_x (1 - \widehat{q}_x)})$$

Pour affirmer la pertinence de notre table d'expérience, les décès observés doivent se situer à l'intérieur des intervalles de confiance théoriques sur l'ensemble de la plage d'âges estimée.

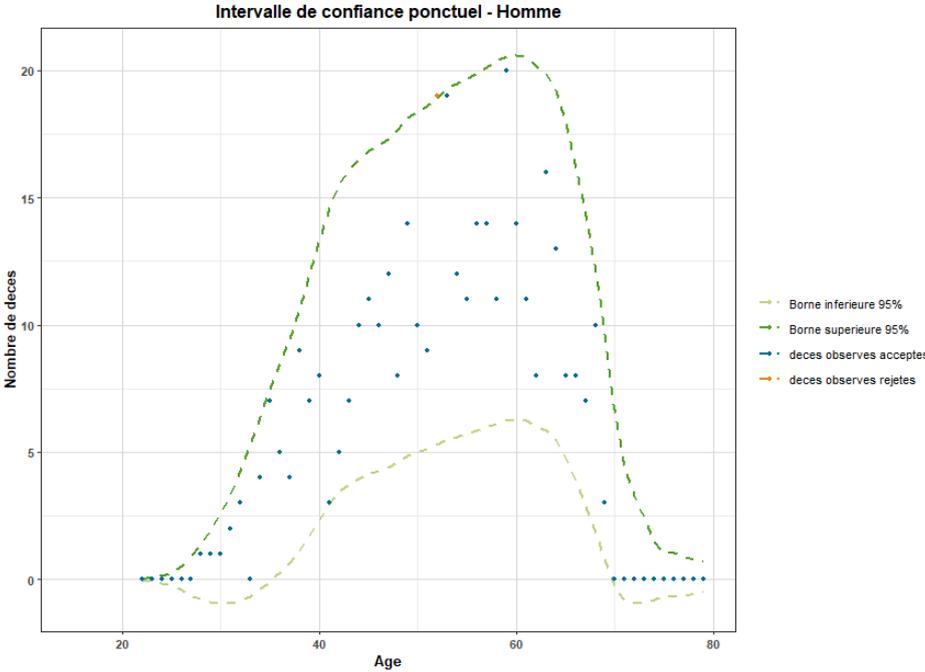


Figure 40: Intervalles de confiance ponctuels sur les décès (homme)

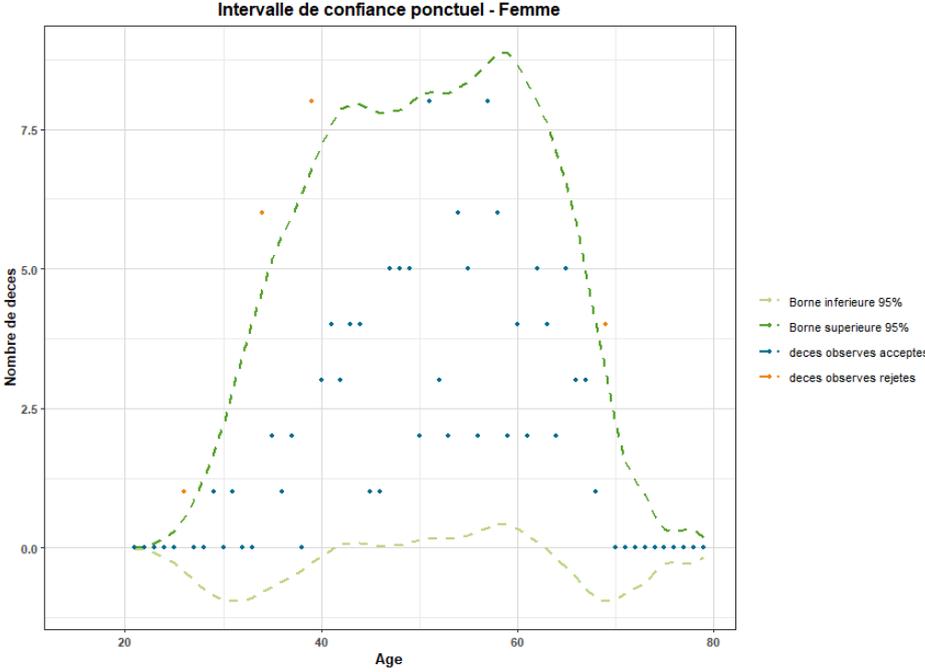


Figure 41: Intervalles de confiance ponctuels sur les décès (femme)

Nous observons un âge chez les hommes et quatre âges chez les femmes pour lesquels le nombre de décès observés est en dehors de l'intervalle de confiance. Ces observations restent toutefois proches de l'intervalle de confiance.

Ces écarts ne sont pas considérés comme significatifs et prouvent l'adéquation des méthodes utilisées pour notre modélisation.

4.4.2. Etude des résidus

Pour juger la qualité du lissage, nous étudions deux résidus : les résidus de réponse et les résidus de Pearson. Il s'agit d'une représentation graphique des ajustements par rapport à la mortalité observée pour un âge atteint.

Nous notons q_x les taux de mortalité bruts obtenus par l'estimateur de Hoem, \widehat{q}_x les taux de mortalité lissés obtenus grâce aux méthodes d'ajustement.

4.4.2.1. Les résidus de réponse

Les résidus de réponse sont définis comme suit :

$$r_x = q_x - \widehat{q}_x$$

Les résidus de réponse doivent être le plus proche possible de zéro, avec une alternance homogène des signes.

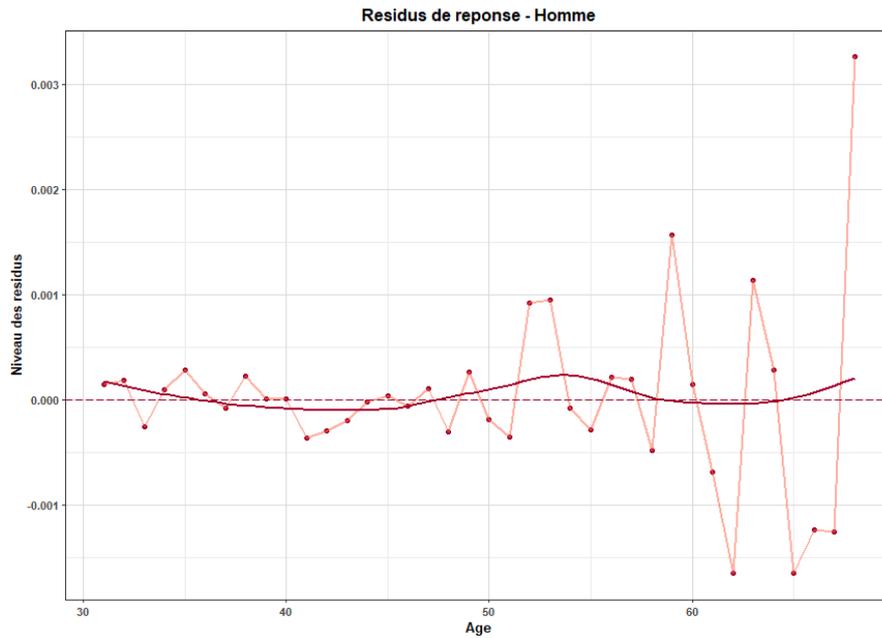


Figure 42: Résidus de réponse (homme)

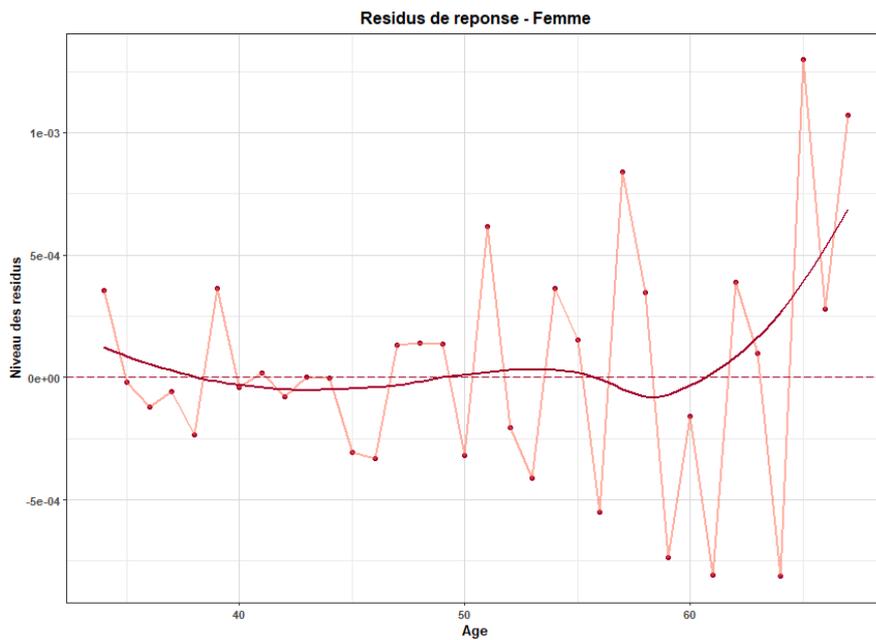


Figure 43: Résidus de réponse (femme)

Les résidus de réponse chez les hommes sont globalement satisfaisant car proches de 0.

Chez les femmes, le manque d'observations sur les âges élevés entraîne un niveau de résidus plus élevé, sans toutefois remettre en question la robustesse de notre modèle.

4.4.2.2. Les résidus de Pearson

Les résidus de Pearson sont définis comme suit :

$$r_x = \frac{D_x - E_x \widehat{q}_x}{\sqrt{\text{Var}[E_x \widehat{q}_x]}}$$

Les résidus de Pearson, quant à eux, doivent osciller dans l'intervalle $[-2 ; 2]$, avec une alternance homogène des signes.

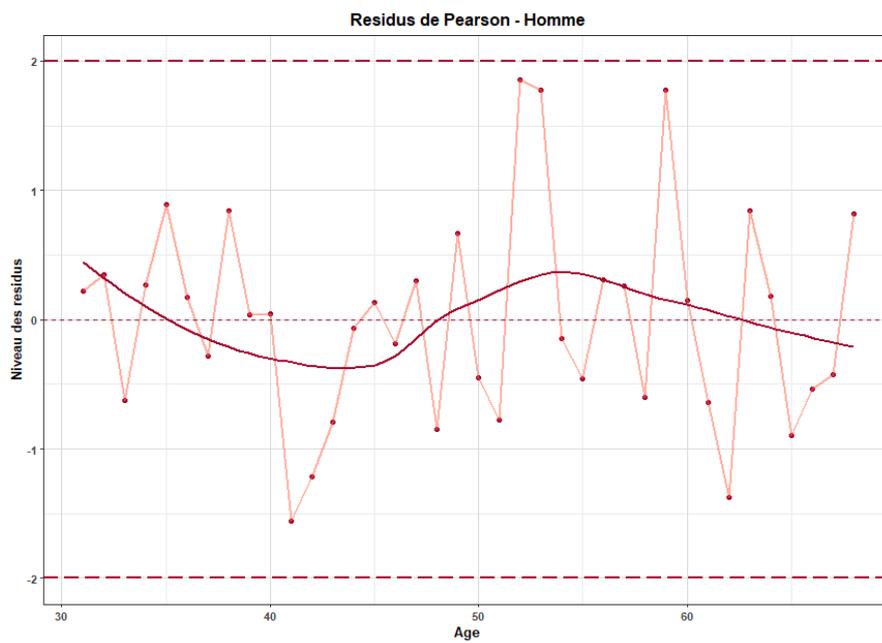


Figure 44: Résidus de Pearson (homme)

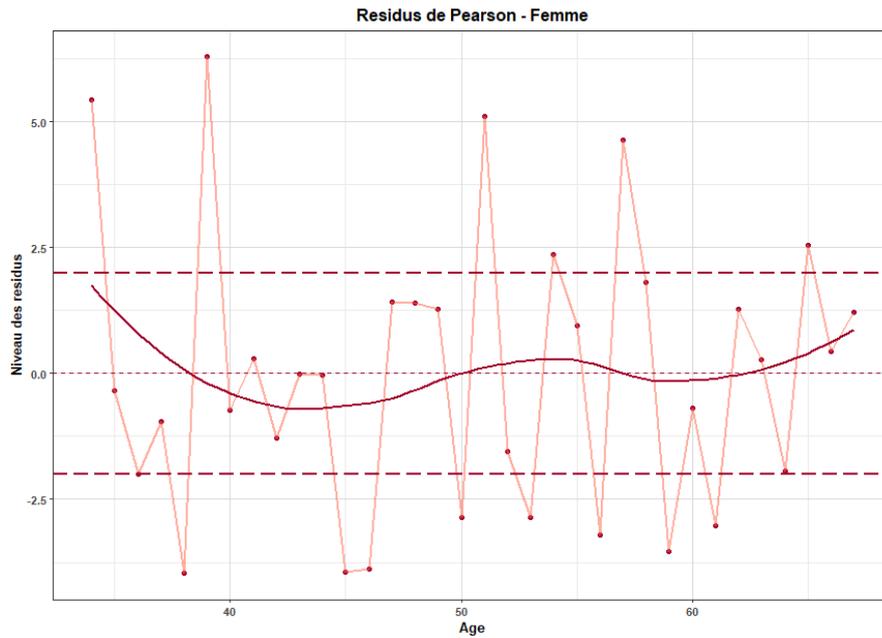


Figure 45: Résidus de Pearson (femme)

Les résidus de Pearson sont globalement satisfaisants chez les hommes et chez les femmes. Là encore, nous notons chez les femmes une variation du niveau des résidus plus élevée du fait du plus faible nombre d'observations.

4.5. Gestion des durations courtes

Lors du pré-testing effectué dans la partie 4.1.1., nous avons décidé de ne pas retenir les durations 1, 2, 3 qui présentent un risque de mortalité moindre par rapport au risque de mortalité moyen du portefeuille étudié.

Des coefficients d'ajustements vont être calculés pour modéliser le risque décès sur ces trois durations afin de prendre en compte leur spécificité dans l'utilisation des lois construites. Pour ce faire, nous calculons des ratios A/E pour chacune des trois durations par tranches d'âge pour comparer leur niveau de risque par rapport au risque moyen des lois construites.

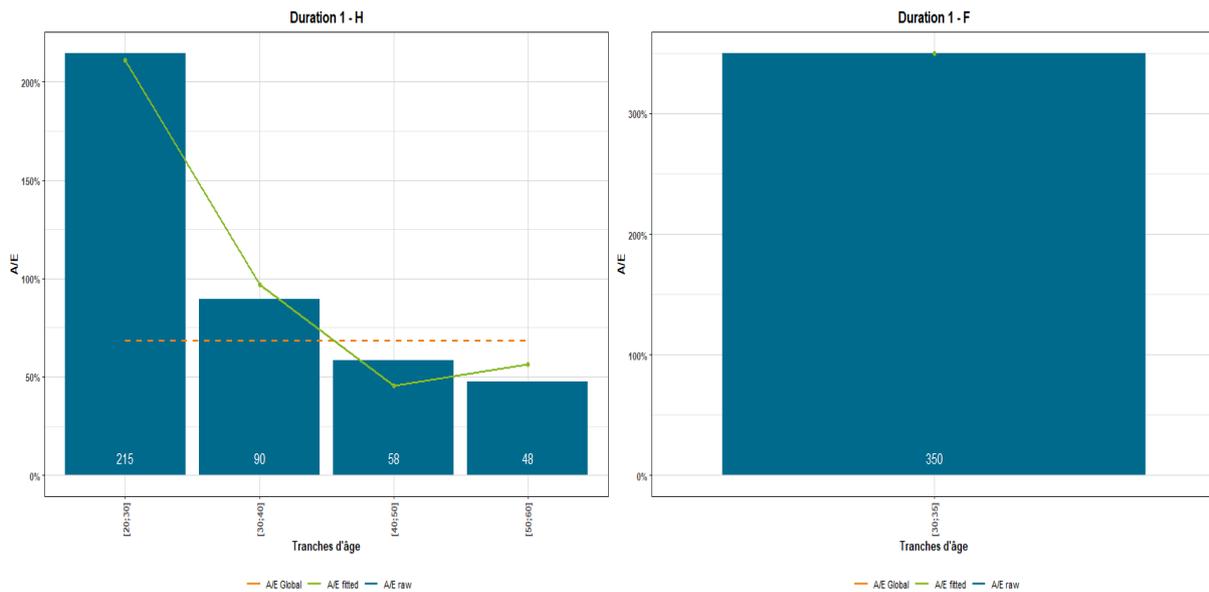


Figure 46: Ratios A/E pour les durations 1 par tranches d'âge

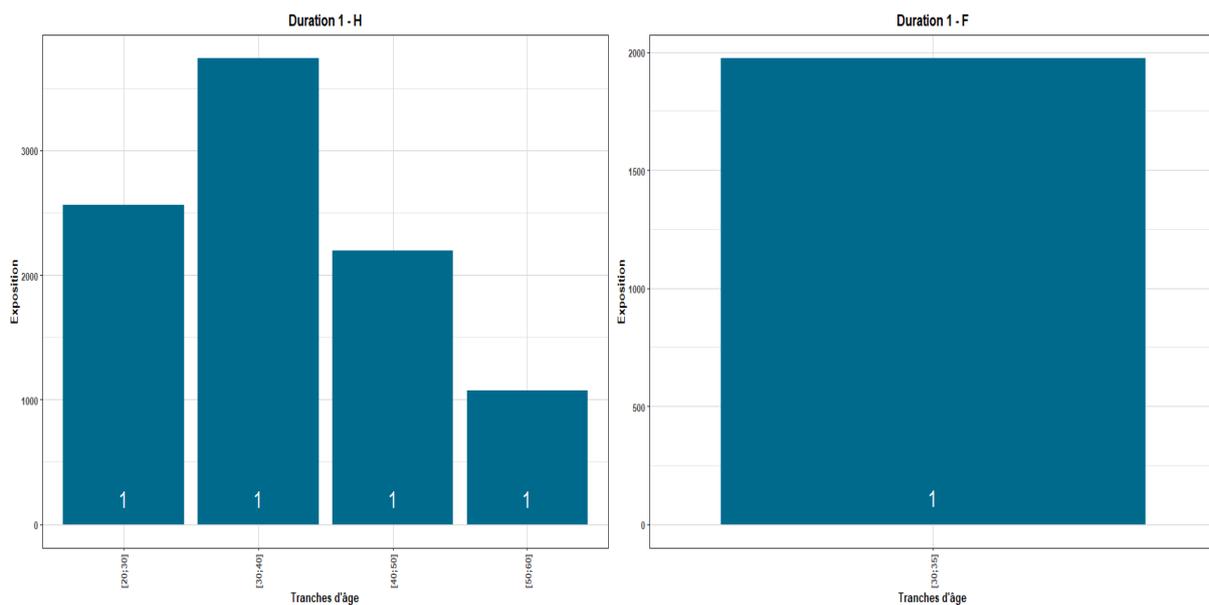


Figure 47: Nombre d'expositions pour les durations 1 par tranches d'âge

Les mêmes graphiques ont été construits pour les durations 2 et 3 et sont disponibles en Annexe 4.

L'exposition et le nombre de décès sont très faibles pour les durations 1, 2, 3, notamment pour les femmes.

Pour avoir le plus d'expositions et de décès possibles et fiabiliser nos coefficients, nous décidons de regrouper les données disponibles sur les trois durations pour calculer un coefficient unique.

Nous obtenons les coefficients suivants :

- Pour les hommes : 0,7648,
- Pour les femmes : 0,6205.

4.6. Backtesting des lois créées

Un backtesting (c'est-à-dire un test à posteriori des lois sur les données historiques) final sur l'ensemble de la base de données de prêts immobiliers (tous segments de prêts confondus) a été réalisé pour vérifier la cohérence des lois de mortalité créées après extrapolations et ajustements. Des ratios $\frac{A}{E}$ ont été calculés : cette fois-ci, les « Expected » représentent le nombre de décès modélisé à partir de l'exposition et des lois créées, avec une granularité de calcul identique à celle des lois (sexe, âge et duration). Les « Actuals » sont toujours le nombre de décès observé à partir de la base de données totale des prêts immobiliers pour la modalité étudiée.

Le but est de comparer le nombre de décès réellement observés dans les données par rapport au nombre de décès estimé à l'aide des lois créées, selon plusieurs variables. Si les lois sont correctement créées, notamment sur le périmètre homogène défini pendant le pré-testing, les ratios doivent être proche de 100%. Cela signifie que la mortalité sous-jacente est bien retranscrite dans les lois.

Pour rappel, le périmètre retenu pour la création de nos lois était le suivant :

- Prêts sociaux immobiliers,
- Années calendaires : 2013-2017,
- Durations : supérieures à 3 ans.

Nous avons lissé et extrapolé les lois, puis appliqué un coefficient d'ajustement pour les durations 1 à 3 pour prendre en compte l'impact de la sélection médicale à l'entrée du contrat.

Le backtesting des lois créées a été réalisé selon plusieurs variables : l'année calendaire, l'âge atteint mais aussi le segment de prêts (prêts sociaux, prêts bancaires, autres).

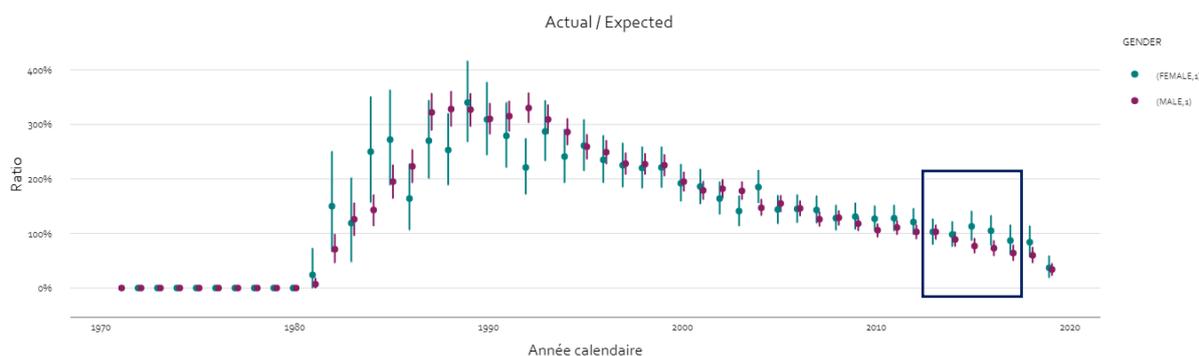


Figure 48 : Backtesting des lois selon l'année calendaire sur la base entière

En observant les ratios $\frac{A}{E}$ sur la totalité de la base de données des prêts immobiliers selon l'année calendaire, nous retrouvons la même tendance que celle observée au moment du pré-testing (cf Annexe 1), à savoir une amélioration de la mortalité au fil des années calendaires. Nous encadrons le périmètre finalement retenu pour la construction des lois.

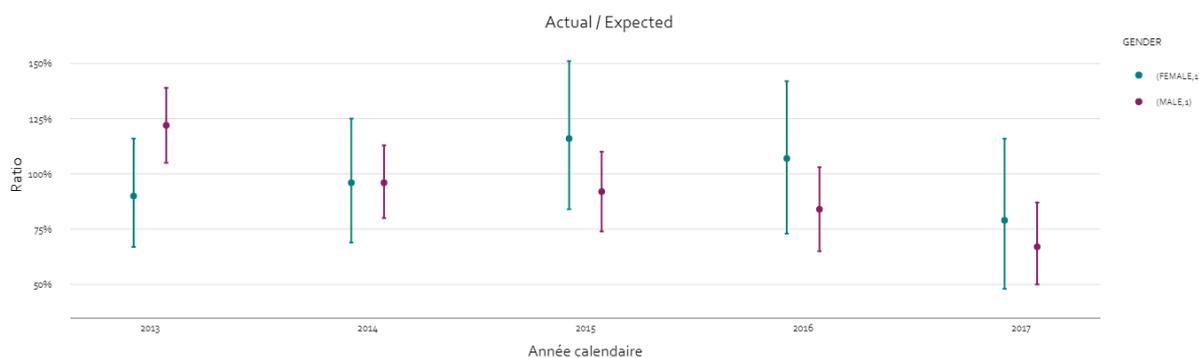


Figure 49 : Backtesting des lois selon l'année calendaire sur le périmètre retenu pour la création des lois

Si l'on s'intéresse au périmètre retenu pour la création des lois uniquement, nous retrouvons des ratios aux alentours de 100%. La tendance d'amélioration de mortalité sur l'année calendaire 2017 observée au moment du pré-testing (cf Annexe 1) est de nouveau visible dans le graphique ci-dessus.

Les ratios calculés selon l'âge atteint sur le périmètre retenu pour la création des lois oscillent bien aux alentours de 100%. Quelques ratios peuvent s'éloigner de 100%, dans le cas où le nombre d'observations est plus faible (âges jeunes ou élevés), mais les intervalles de confiance sont plus larges et permettent de valider l'adéquation des lois avec le risque observé (100% est inclus dans l'intervalle de confiance). Le travail d'extrapolation des lois nous a permis d'obtenir des taux de

mortalité même dans le cas où le nombre d’observations était nul (pour les âges supérieurs à 70 ans par exemple).

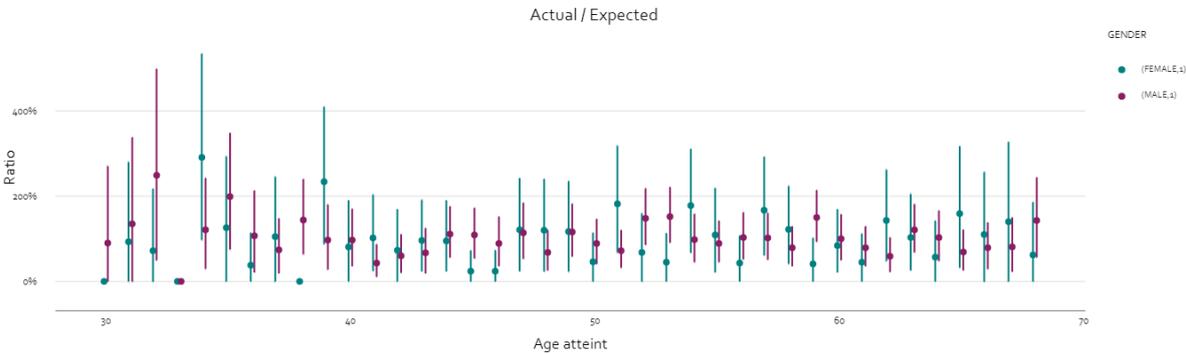


Figure 50 : Backtesting des lois selon l'âge atteint sur le périmètre retenu pour la création des lois

Enfin, le backtesting réalisé selon le segment de prêts sur le périmètre retenu pour la création des lois (ici en gardant toutefois tous les segments de prêts pour comparaison), nous permet de vérifier que le segment des prêts sociaux affiche un ratio cohérent proche de 100%.

Le segment des prêts bancaires affiche un ratio aux alentours de 70% pour les hommes et 105% pour les femmes (avec toutefois moins d’observations pour les femmes). Le risque de mortalité du segment des prêts bancaires semble donc plus favorable que le risque de mortalité des prêts sociaux.

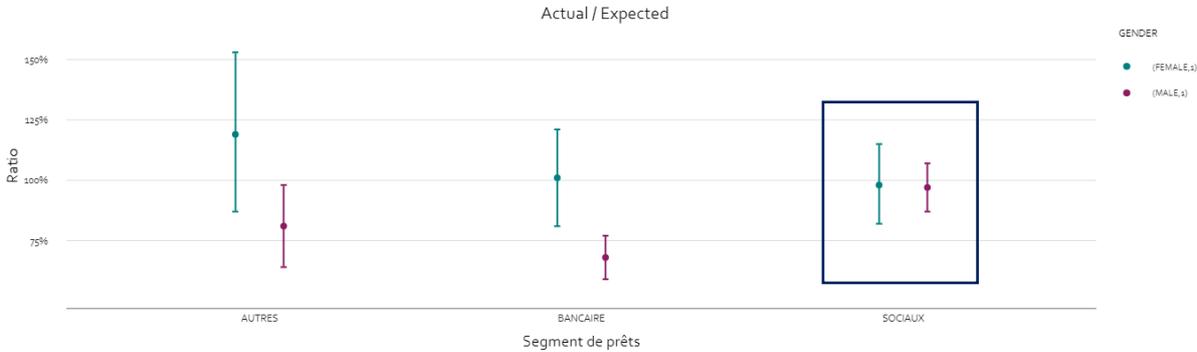


Figure 51 : Backtesting des lois selon le segment de prêts sur le périmètre retenu pour la création des lois

De façon générale, nous observons un ratio moyen de 98% pour les femmes et 97% pour les hommes sur le périmètre retenu pour la création des lois, ce qui nous permet d’affirmer que nous avons correctement exploité l’information disponible dans les données utilisées pour la construction des lois.

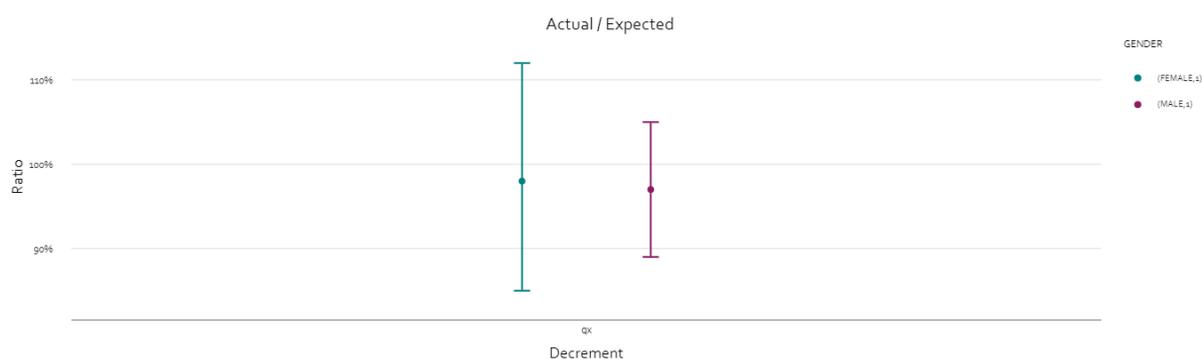


Figure 52 : Backtesting des lois sur le q_x moyen sur le périmètre retenu pour la création des lois

Dans cette quatrième partie, nous avons construit des tables de mortalité pour le segment des prêts sociaux immobiliers, par sexe, en vision « ultimate » et ajustées pour les durations 1, 2 et 3. La robustesse de nos lois a été testée et les résultats des tests sont satisfaisants.

Ces nouvelles tables d'expérience vont avoir un impact sur les projections Best Estimate des flux : les taux de mortalité annuels par âge atteint étant différents de ceux des tables de mortalité actuellement utilisées dans les projections de nos engagements, les chroniques de primes et de sinistres vont être modélisées différemment. En les comparant, nous pouvons mesurer l'impact du changement de tables de mortalité sur le calcul de nos engagements.

PARTIE 5 – APPLICATION DES NOUVELLES LOIS ET IMPACTS SUR LES ENGAGEMENTS

Nous étudions dans cette dernière partie l'impact du changement de table de mortalité d'expérience sur les engagements modélisés pour une génération d'assurés.

Pour cela, nous projetons les flux de primes et de sinistres à l'aide de la table créée dans la partie 4 et les comparons aux flux projetés avec la table Best Estimate classiquement utilisée. Dans notre modèle, les flux projetés avec la table Best Estimate correspondent à la meilleure vision du risque. Cette définition prend sa source dans Solvabilité 2 qui définit le Best Estimate comme la juste valeur des flux, au plus proche du risque.

5.1. Méthodologie

Pour calculer l'impact du changement de table sur les flux Best Estimate, nous devons projeter, pour une nouvelle génération d'assurés, les flux de primes et de sinistres à chaque pas de temps du prêt. Nous décrivons par la suite les étapes de cette modélisation.

*5.1.1. Création des *model points**

Une base de données regroupant l'ensemble des assurés du portefeuille, appelée *model points*, est utilisée pour projeter les flux de primes et de sinistres. Ces *model points* présentent plusieurs caractéristiques :

- L'âge à l'adhésion de l'assuré,
- Le sexe de l'assuré,
- La durée du prêt,
- Le montant du prêt,
- Le taux d'intérêt du prêt.

Ils doivent répondre aux hypothèses prises au moment de la souscription sur l'âge moyen, la durée moyenne, le montant de capitaux moyen, etc. Ils peuvent être regroupés s'ils possèdent des caractéristiques similaires en termes d'âge et de durée de prêt. Les capitaux empruntés des sous-

groupes sont alors sommés, et une moyenne est faite sur la répartition hommes/femmes du sous-groupe et sur leur taux d'intérêt.

Pour notre calcul d'impact, nous prenons comme base de *model points* la dernière génération, c'est-à-dire la génération 2020, du nouveau produit lancé par notre client en ajustant à la hausse le taux d'intérêt moyen afin de refléter le contexte économique actuel. Les données de la génération 2020 nous ont été transmises récemment par le client et n'étaient pas incluses dans les données de la base d'encours étudiée pour la création des lois.

Il s'agit d'un produit assurance emprunteur pour les prêts immobiliers. Nous estimerons donc l'impact de la nouvelle table de mortalité des prêts immobiliers uniquement.

Les statistiques de cette génération avant tout ajustement sont les suivantes :

Nombre d'assurés	2 386
Age moyen à l'adhésion	37,82
Durée moyenne des prêts	18,44
Capitaux moyens assurés	66 845
Taux d'intérêt moyen	1,30%
Pourcentage d'hommes	52%

Après ajustement du taux d'intérêt moyen à la hausse, celui-ci devient 1,80%.

5.1.2. Modélisation des flux projetés

La modélisation consiste en un déroulement des flux probables de primes et de sinistres pour chaque *model point* jusqu'à extinction des engagements.

A chaque pas de temps, la modélisation va prendre en compte la probabilité de sortie du portefeuille de l'assuré. La sortie peut être due :

- Au décès de l'assuré dans l'année,
- Au rachat du contrat par l'assuré.

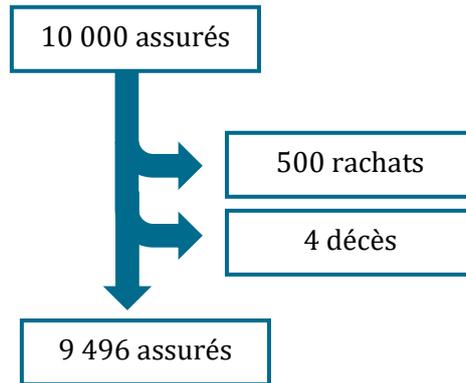


Figure 53: Illustration des deux types de sorties du portefeuille d'assurés

Ainsi, le nombre d'assurés d'âge x encore présents dans le portefeuille chaque année est caractérisé par la formule suivante :

$$L_x = L_{x-1} \times (1 - r) \times (1 - q_x)$$

Avec r le taux de chute annuel estimé. Pour notre étude, nous prendrons un taux de chute annuel de 5%.

A partir de ces différentes hypothèses, nous calculerons à chaque pas de duration i pour des assurés d'âge x les métriques suivantes :

- Le montant probable de primes :

$$P_i = L_x \times P$$

- Le montant probable de sinistres :

$$S_i = q_x \times CRD_i$$

- Le montant probable de profits :

$$R_i = P_i - S_i$$

5.2. Impact de la nouvelle table sur les flux projetés

5.2.1. Comparaison des tables de mortalité

Nous comparons dans un premier temps les probabilités de décès des deux tables.

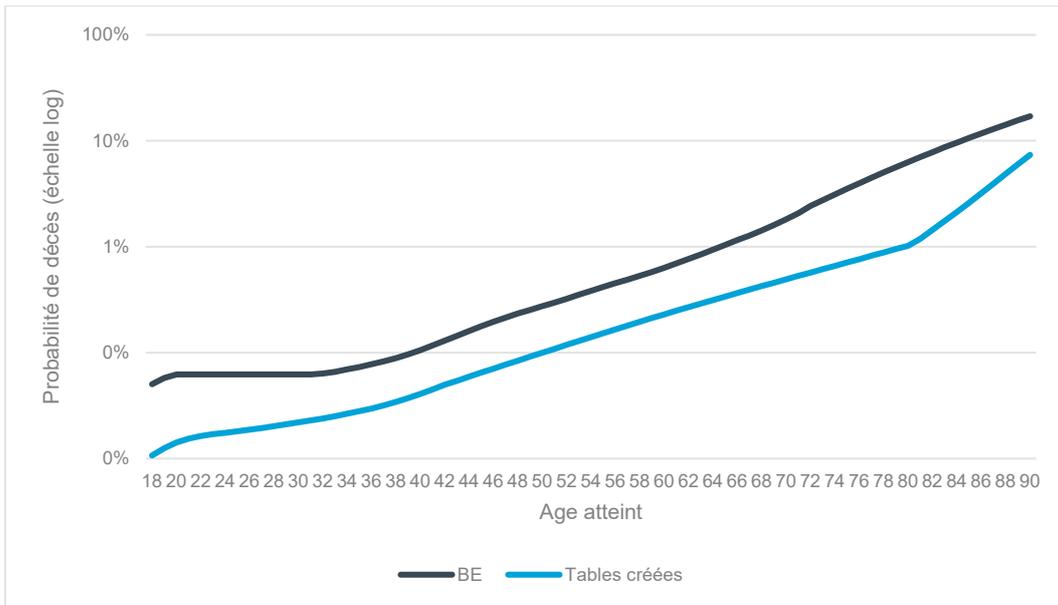


Figure 54: Comparaison des taux de mortalité entre la table Best Estimate et la table créée¹⁰ en échelle logarithmique Homme

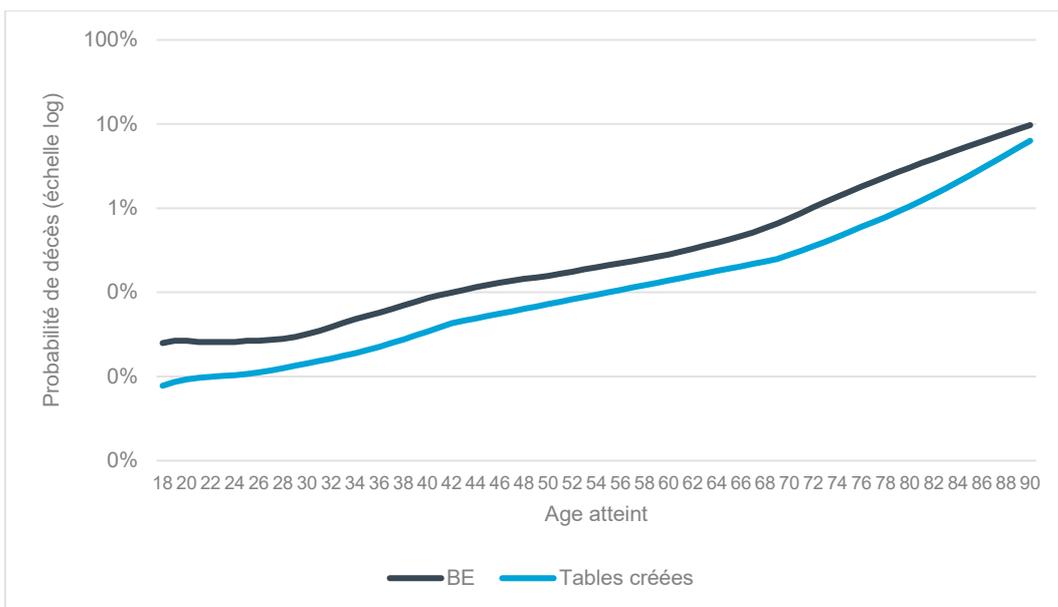


Figure 55: Comparaison des taux de mortalité entre la table Best Estimate et la table créée¹¹ en échelle logarithmique Femme

Nous observons sur les graphiques ci-dessus que les nouvelles lois créées présentent des taux de mortalité plus faibles que les lois Best Estimate. On peut dire que la population d'assurés qui

¹⁰ Les taux de mortalité présentés sont les taux ultime, sans effet de sélection médicale

compose notre base d'étude pour la création des nouvelles lois présentent un meilleur risque que la population des tables Best Estimate.

5.2.1. Résultats des impacts du changement de tables

<i>Résultats en €</i>	Valeur actualisée des primes	Valeur actualisée des sinistres	Valeur actualisée des profits
Table Best Estimate	1 575 250	1 103 693	187 600
Table d'expérience créée	1 576 263	990 108	279 027
Rapport	<i>100,1%</i>	<i>89,7%</i>	<i>148,7%</i>

Figure 56: Comparaison des projections de flux avec la table Best Estimate et avec la table d'expérience créée

Les résultats d'impact illustrent le constat fait dans la partie 5.2.1. Les probabilités de décès étant plus faibles avec les nouvelles tables créées, le volume de sinistres attendus est plus faible et les profits plus élevés. Le volume de primes est légèrement plus élevé avec les nouvelles lois puisqu'il y a moins de décès projetés et donc moins de sorties du portefeuille.

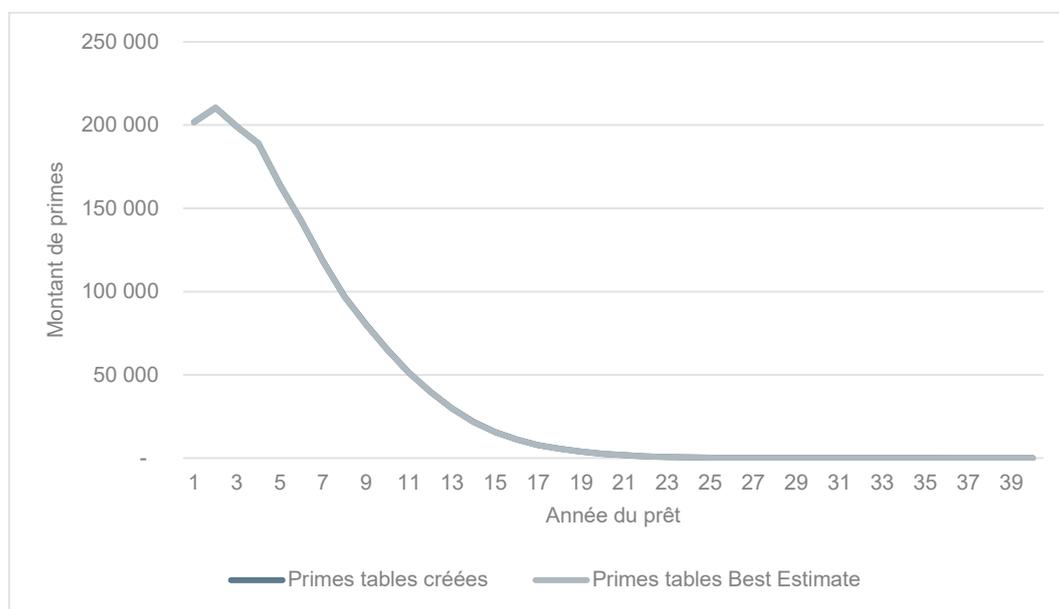


Figure 57: Evolution des primes par année de projection du prêt

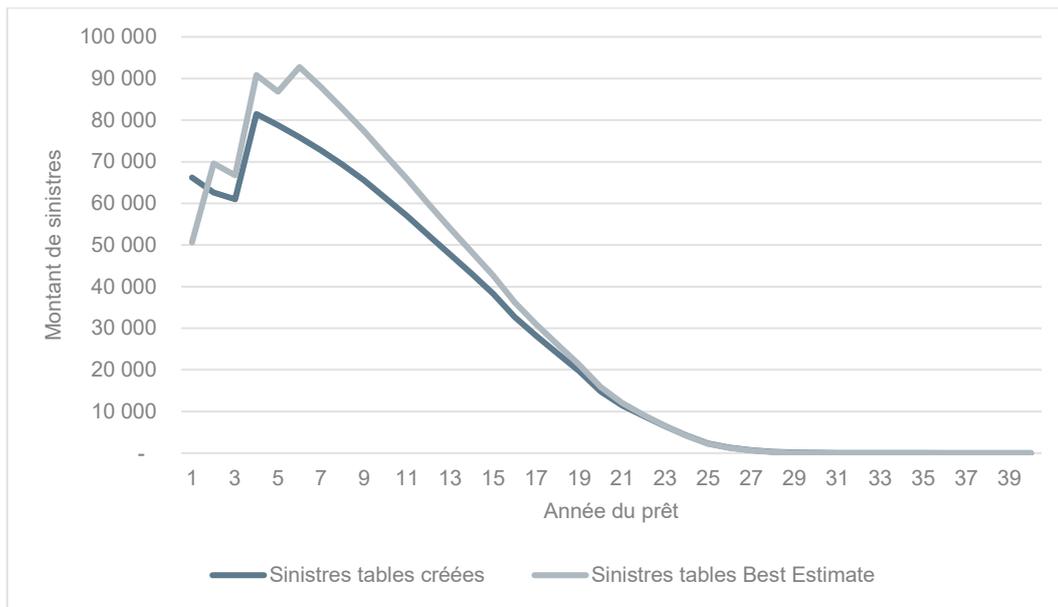


Figure 58: Evolution des sinistres par année de projection du prêt

Nous remarquons dans le graphique ci-dessus que l'effet sélection médicale est modélisé de manière différente entre les deux tables : nous avons retenu, par souci de robustesse, pour nos tables créées un taux unique appliqué sur les durations 1, 2 et 3 pour refléter le risque moindre engendré par la sélection médicale à l'adhésion. Sur les tables Best Estimate, des coefficients d'ajustement sont appliqués jusqu'à la durée 6 pour refléter l'effet sélection médicale.

Le paiement des sinistres correspondant au remboursement du CRD du prêt, une fois l'effet sélection médicale absorbé, le volume de sinistres décroît au fil du temps.

Après une comparaison brute de la table de mortalité initialement utilisée pour projeter les flux de primes et de sinistres et la table de mortalité d'expérience construite en partie 4, nous avons constaté que le risque décès de notre table d'expérience apparaissait plus favorable que dans la table initialement exploitée. En comparant les chroniques de flux projetés avec les deux tables, en toute logique nous constatons un résultat plus important avec l'utilisation de la nouvelle table d'expérience qu'avec l'ancienne table.

Au-delà de l'impact du changement de table sur les engagements de l'assureur, il serait opportun d'ajuster le tarif des nouvelles générations d'adhésion pour prendre en compte l'évolution du risque du portefeuille. Le risque décès des nouvelles lois étant plus favorable que celui des lois précédemment utilisées, les tarifs pourraient être potentiellement revus à la baisse, toujours dans un souci d'apparaître plus compétitifs dans un environnement très concurrentiel.

Le changement de table de mortalité peut également impacter les générations déjà souscrites à travers le recalcul des provisions pour risque croissant. Ces provisions, modélisées pour prendre en compte le décalage entre le risque assuré et la prime perçue par l'assureur, sont calculées comme la différence des valeurs actuelles des engagements pris par l'assureur et par l'assuré. Suite à l'implémentation de la nouvelle table de mortalité, les engagements des deux parties vont être recalculés et le montant des provisions pourra être ajusté.

CONCLUSION

La base de données à notre disposition était riche en informations et disposait d'un historique suffisamment profond pour construire des tables de mortalité d'expérience robustes. Alors que le marché de l'assurance emprunteur est en perpétuel mouvement, notamment du fait des nombreuses évolutions réglementaires qui accentuent toujours un peu plus la protection de l'assuré, l'utilisation de tables de mortalité d'expérience est un levier important dans l'ajustement des tarifs au plus près du risque assuré.

Après un travail important de préparation des données pour s'assurer de la cohérence des informations de la base, nous avons retenu pour notre étude le segment des prêts sociaux. Dans un premier temps, nous avons construit un modèle machine learning robuste pour dissocier les prêts immobiliers et les prêts à la consommation. A travers de précédentes études, le type de prêt est apparu comme une variable particulièrement discriminante pour le risque de mortalité. Comme les produits créés sur le segment des prêts sociaux sont majoritairement des produits couvrant les prêts immobiliers et que nous avons mis en évidence un risque décès moins favorable pour les prêts à la consommation que pour les prêts immobiliers, nous avons concentré notre étude sur la base de prêts immobiliers.

La construction des tables d'expérience a été réalisée en cinq étapes. Après avoir déterminé la période d'observation de l'étude, des taux de mortalité bruts issus directement des données à notre disposition ont été calculés. Pour obtenir ensuite une courbe de taux lissée, nous avons utilisé le package R *Mortality Smooth*. L'extrapolation des taux a ensuite été faite en utilisant le modèle de Brass. Enfin, la qualité d'ajustements a été vérifiée à travers le calcul d'intervalles de confiance et l'étude des résidus.

Ces travaux nous ont permis d'obtenir des tables de mortalité pour le segment des prêts sociaux immobilier, par sexe, en vision « ultime » et ajustées pour les durations 1, 2 et 3. L'utilisation de ces tables va avoir un impact sur les projections Best Estimate des flux : les chroniques de primes et de sinistres seront différentes de celles projetées avec les anciennes tables de mortalité. Cet impact dû au changement de table de mortalité a été évalué aux alentours de -10% sur la valeur actualisée des sinistres probables. Cette diminution de l'estimation du risque de mortalité du portefeuille laisse une marge de manœuvre confortable à l'assureur pour ajuster ses tarifs au plus proche du risque et apparaître encore plus compétitif sur un marché où l'assuré possède une

vision exhaustive des produits et des tarifs pratiqués notamment grâce aux comparateurs sur internet.

Toutefois, bien que reflétant au plus juste le risque de mortalité aujourd'hui, il conviendra de vérifier l'adéquation de ces nouvelles tables d'ici quelques années lorsque le marché aura le recul nécessaire sur le potentiel impact de la loi Lemoine, mise en application à partir du 1^{er} juin 2022. En effet, l'absence de sélection médicale que prévoit cette loi pour les assurés n'ayant pas atteint les 60 ans à la fin de leur prêt et dont l'encours ne dépasse pas 200 000€ risque d'avoir un impact défavorable sur le risque décès. Jusqu'à présent, on constatait à travers les études menées sur les portefeuilles emprunteur, un risque de mortalité plus favorable sur les deux ou trois premières durations du contrat grâce à la sélection médicale mise en place au moment de la souscription du prêt. Cet effet sur les premières durations risque de s'amoinrir sur la partie du portefeuille concernée par la loi Lemoine. Le marché a d'ores et déjà anticipé cette hausse de sinistralité et proposé parfois des majorations tarifaires pour l'absorber. Toutefois, ce n'est qu'avec un recul de plusieurs années et les données suffisantes que nous serons en mesure d'évaluer la justesse des estimations d'impacts faites jusqu'ici.

BIBLIOGRAPHIE

- Banque de France, « *Crédits aux particuliers – Déc 2021* », site consulté le 6 juin 2022
<https://www.banque-france.fr/statistiques/credits-aux-particuliers-dec-2021>
- Cafpi, « *Les jeunes primo-accédants face à l'accès à la propriété en France* », site consulté le 9 juin 2022
<https://www.cafpi.fr/credit-immobilier/actualites/les-jeunes-primo-accedants-face-a-laces-a-la-propriete-en-france>
- Comité Consultatif du Secteur Financier, « *Bilan de l'assurance emprunteur* », 2020
https://www.ccsfin.fr/sites/default/files/media/2020/11/11/2020_ccsf_rapport_bilan_assurance_emprunteur.pdf
- Delwarde A. & Denuit M., « *Construction de tables de mortalité périodiques et Prospectives* », 2006
- France Assureurs, « *Le marché des assurances santé et prévoyance en 2020* », site consulté le 20 décembre 2021
<https://www.franceassureurs.fr/nos-chiffres-cles/lassurance-sante-et-prevoyance/>
- Lopez O., « *Modèles de durée* », Cours ISUP, 2016/2017
- Magnolia.fr, « *Les femmes paient plus cher leur crédit immobilier* », site consulté le 9 juin 2022
<https://www.magnolia.fr/actualites/assurance-emprunteur/femmes-paient-plus-cher-leur-credit-immobilier>
- Observatoire Credit Logement/CSA, « *La hausse rapide des taux continue* », site consulté le 10 octobre 2022
<https://www.observatoirecreditlogement.fr/derniere-publication>
- Planchet F., Thérond P., « *Modèle de durée : Applications actuarielles* », 2006
- PwC, « *Assurance emprunteur : une dynamique soutenue dans un contexte de plus en plus incertain* », 2021
<https://www.pwc.fr/fr/publications/assurance/assurance-emprunteur-dynamique-soutenue.html>

ANNEXES

Annexe 1 : Pré-testing du segment immobilier des prêts sociaux - Par année calendaire

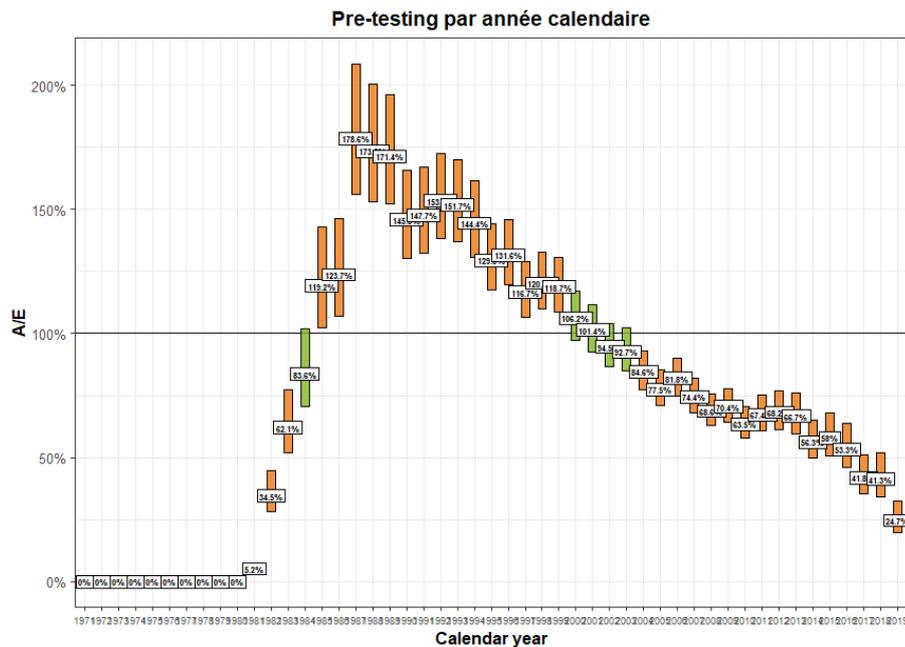


Figure 59: Pré-testing par année calendaire avant décision

On observe sur ce graphique une amélioration notable du risque de mortalité au fil des années calendaires par rapport au risque moyen du portefeuille. Pour prendre en compte cette amélioration des dernières années tout en gardant un recul suffisant pour avoir l'exhaustivité des sinistres survenus, nous gardons la période d'années calendaires 2013-2017.

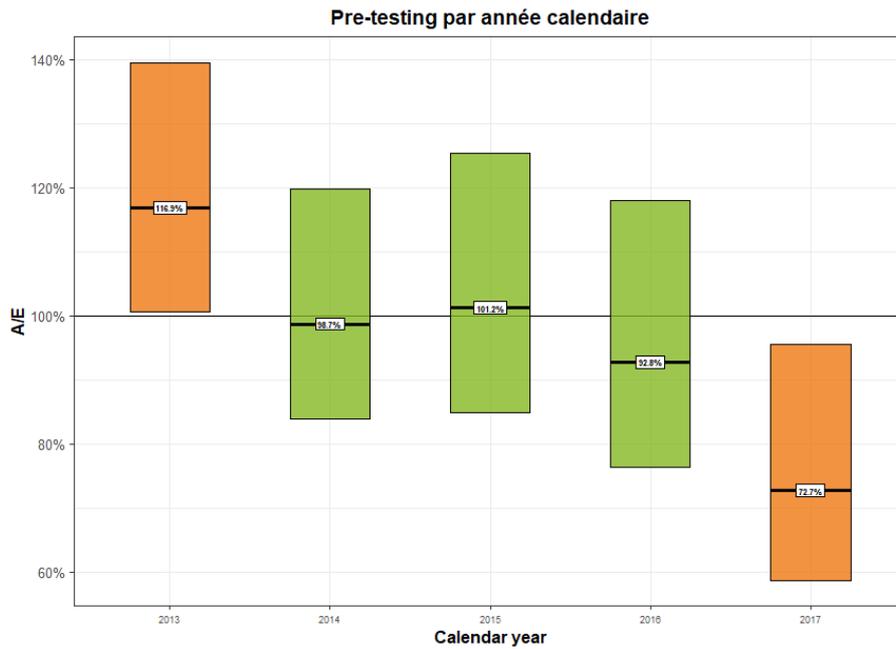


Figure 60: Pré-testing par année calendaire après décision

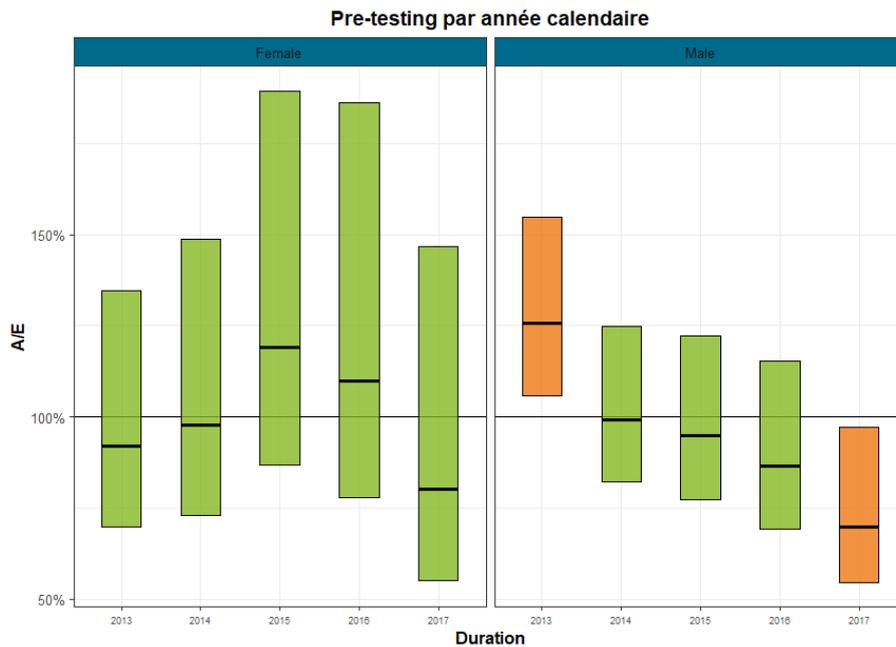


Figure 61: Pré-testing par année calendaire par sexe après décision

Bien que l'année calendaire 2013 présente un risque de mortalité supérieur au risque moyen du portefeuille et que l'année 2017 présente un risque de mortalité inférieur au risque moyen du portefeuille, nous retenons ces deux années dans notre périmètre d'observation pour avoir un risque moyen sur 5 ans.

Annexe 2 : Pré-testing du segment immobilier des prêts sociaux – Par duration

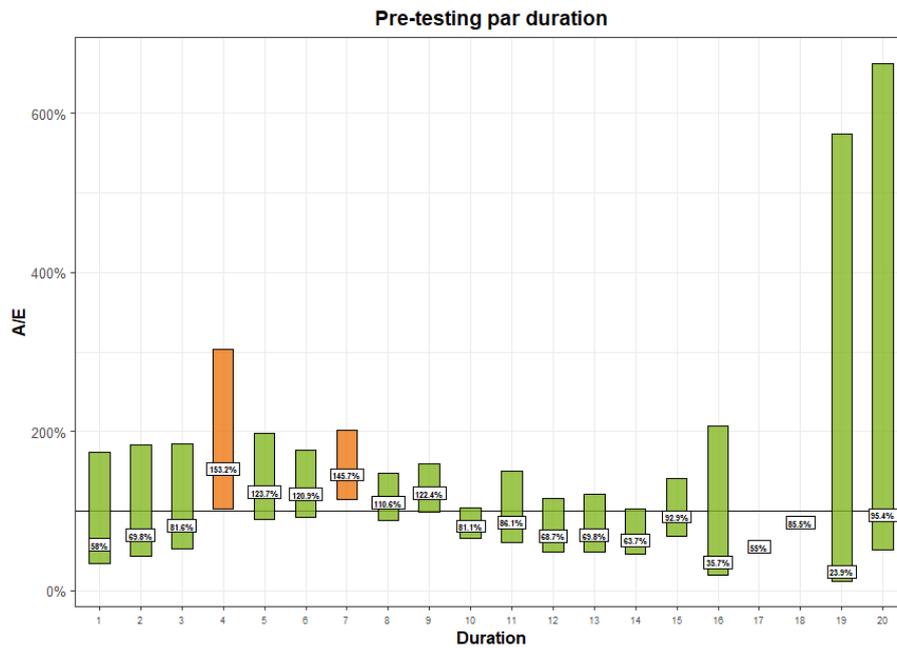


Figure 62: Pré-testing par duration avant décision

Nous observons sur les trois premières durations un risque de mortalité moins important que le risque moyen du portefeuille. Cela s'explique notamment par la sélection médicale à la souscription, qui permet de par les questionnaires médicaux et autres procédures à remplir pour souscrire à une assurance emprunteur, de contrôler les risques assurés du portefeuille. Ces trois premières durations feront l'objet d'un retraitement spécifique dans les lois de mortalités finales pour prendre en compte leur risque de mortalité moindre par rapport au risque moyen du portefeuille.

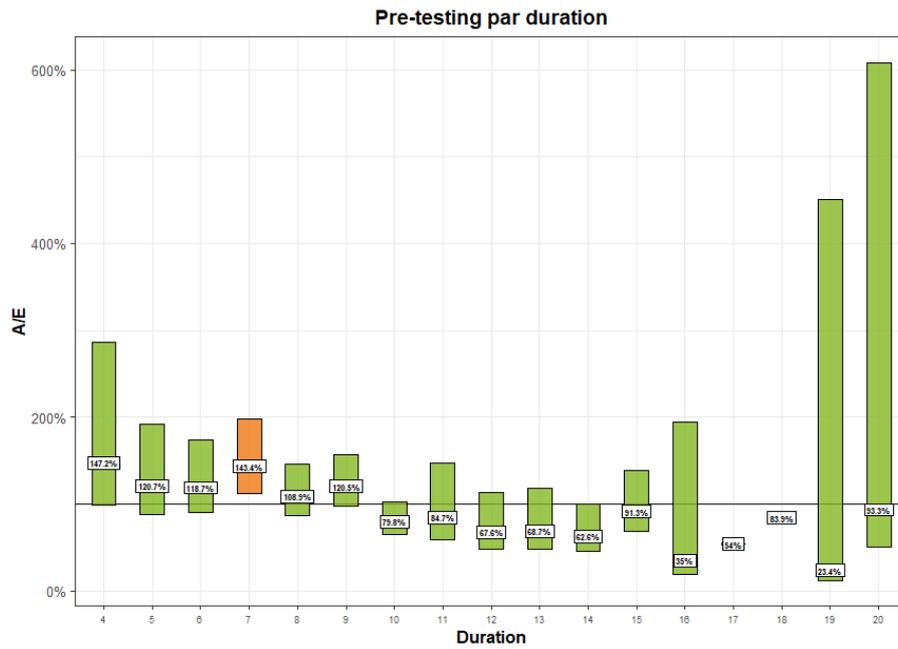


Figure 63: Pré-testing par duration après décision

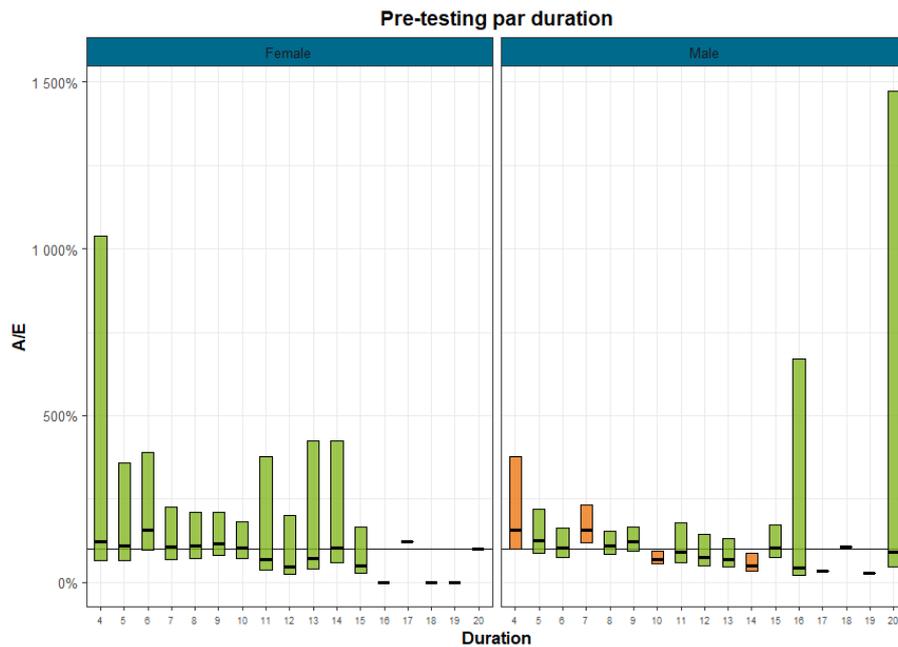


Figure 64: Pré-testing par duration par sexe après décision

Il n'y a que peu d'observations sur les durations supérieures à 15 ans. Nous les gardons tout de même dans notre périmètre d'étude puisqu'elles auront un poids faible dans l'étude.

Annexe 3 : Pré-testing du segment immobilier des prêts sociaux – Par année de souscription

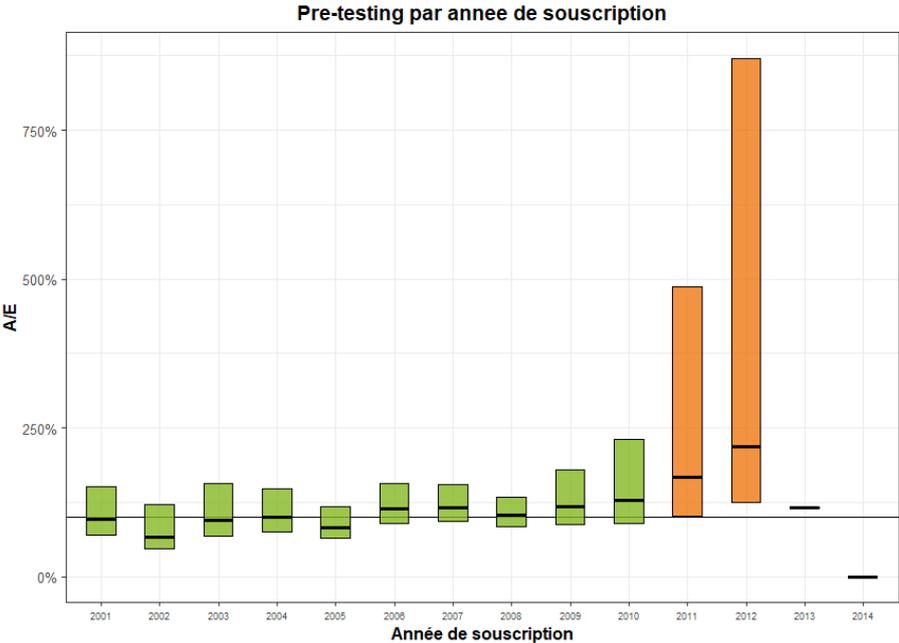


Figure 65: Pré-testing par année de souscription

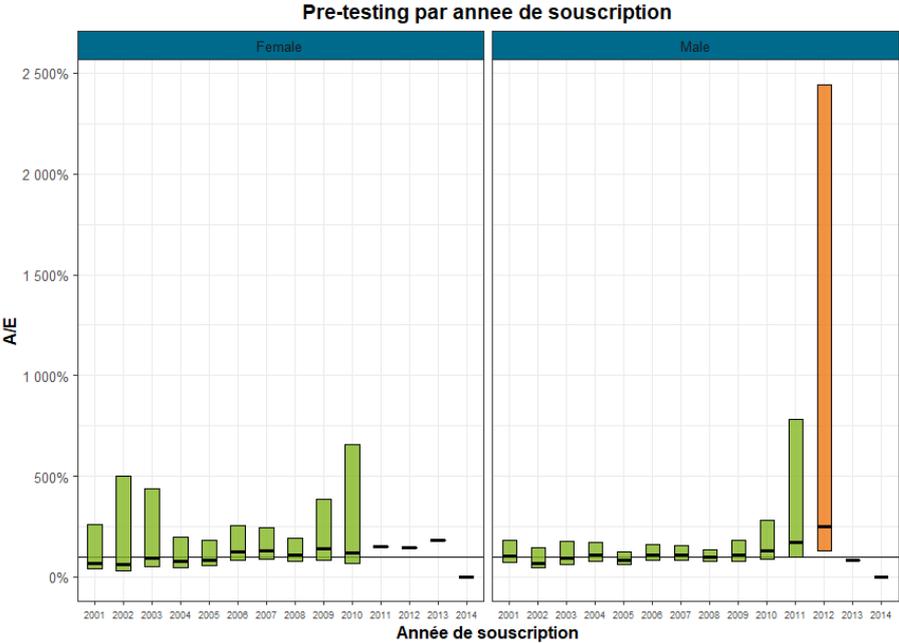


Figure 66: Pré-testing par année de souscription par sexe

Toutes les années de souscription ont un risque de mortalité proche du risque de mortalité moyen du portefeuille, exceptées les années de souscription 2011 et 2012. Toutefois, comme il y a peu d'observations sur ces années de souscription (au vu des intervalles de confiance importants), ces années ne pèseront que très peu dans la modélisation. Nous gardons l'ensemble des années de souscription.

Annexe 4 : Gestion des durations courtes (durations 2 et 3)

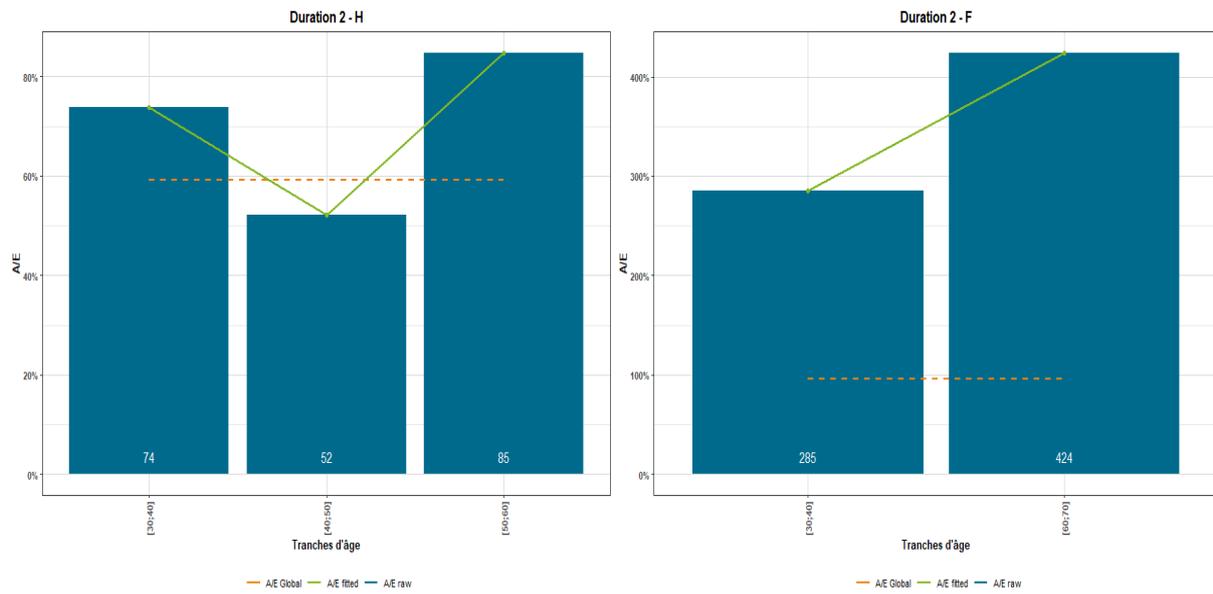


Figure 67: Ratios A/E pour les durations 2 par tranches d'âge

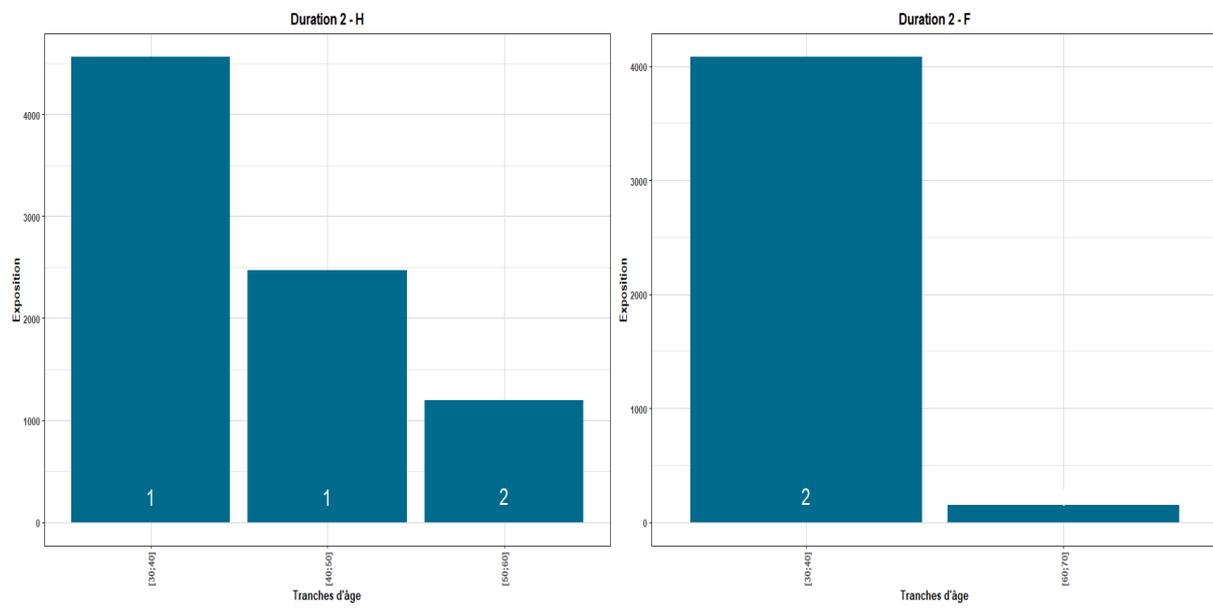


Figure 68: Nombre d'expositions pour les durations 2 par tranches d'âge

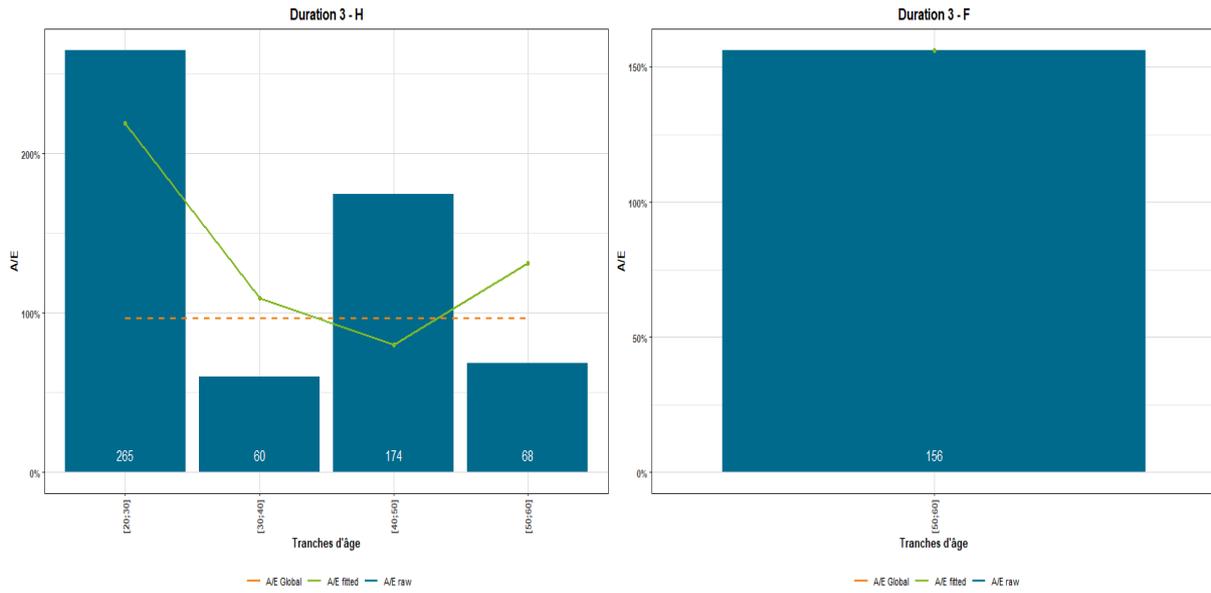


Figure 69: Ratios A/E pour les durations 3 par tranches d'âge

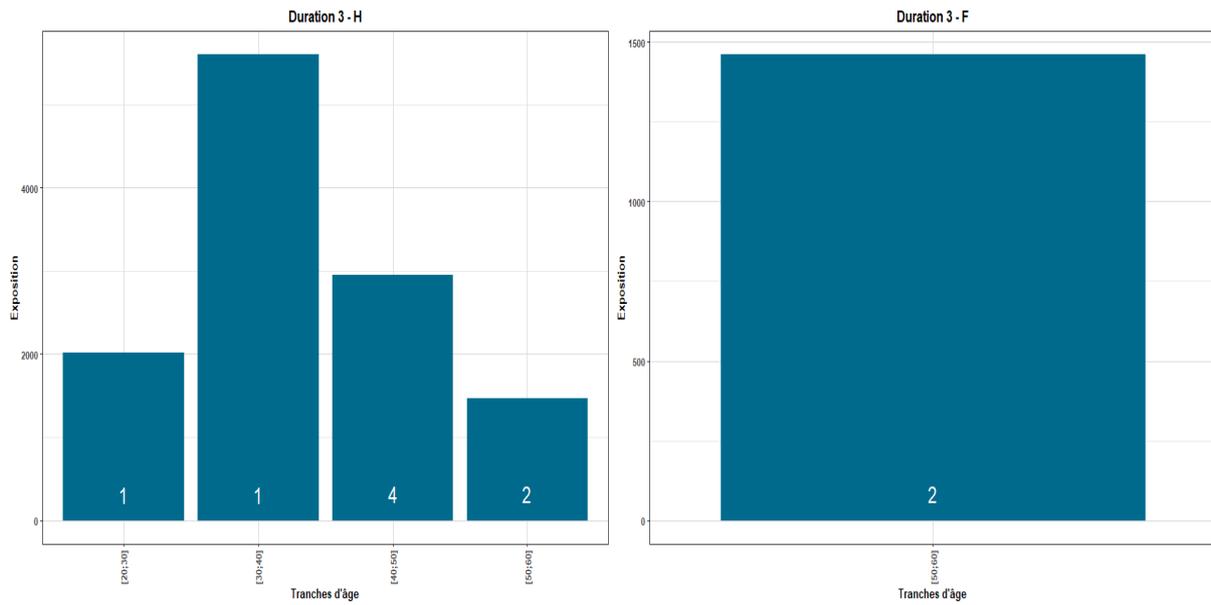


Figure 70: Nombre d'expositions pour les durations 3 par tranches d'âge