

Mémoire présenté devant le Conservatoire National des Arts et Métiers pour l'obtention du Master Droit Economie Gestion mention Actuariat et l'admission à l'Institut des Actuaires

le mercredi 7 juin 2023

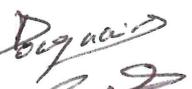
Par : Gildas HENAFF

Titre: Modélisation dynamique des résiliations sur un portefeuille d'assurance emprunteur en contrat groupe

Confidentialité : NON

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

Membres présents du jury de l'Institut des Actuaires signatures

Mme Edith BOCQUAIRE 

Mme Célya MEUNIER 

Entreprise :

Nom :

PREDICA (Crédit Agricole Assurances)

Membres présents du jury du Cnam

M. David FAURE (Président) 

M. Olivier DESMETTRE 

M. François WEISS 

Directeur de mémoire en entreprise :

Nom : Omblin de BAUDINIÈRE/Thibaut MULLER

Signature : 

Invité :

Nom :

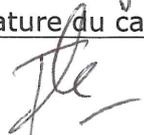
Signature :

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

Signature du responsable entreprise



Signature du candidat



Secrétariat

Bibliothèque :

MÉMOIRE D'ACTUARIAT

**Modélisation dynamique des résiliations sur
un portefeuille d'assurance emprunteur en
contrat groupe**

GILDAS HENAFF *

MASTER 2 ACTUARIAT - LE CNAM

—
7 juin 2023

Résumé

L'assurance des emprunteurs (ADE) représente l'un des domaines de l'assurance de personnes les plus importants sur le marché français, avec un chiffre d'affaires en 2020 d'un peu plus de 10 milliards d'euros, dont près de 7 milliards d'euros au titre de la couverture des prêts immobiliers, marché en croissance régulière depuis de nombreuses années (+55% depuis 2007), en lien avec les encours crédit. En 2021, Crédit Agricole Assurance représente près de 21% du chiffre d'affaire de l'ADE en France.

Le contrat d'assurance emprunteur consiste en la couverture d'une ou plusieurs personnes physiques (assuré, co-assuré voire caution) contre un risque potentiel d'insolvabilité lié à un événement (décès, arrêt de travail, invalidité ou, plus rarement perte d'emploi) venant potentiellement réduire leurs capacités de remboursement d'un prêt souscrit auprès d'un organisme prêteur.

Depuis 2011, le marché de l'assurance emprunteur a connu de nombreuses évolutions techniques et réglementaires, visant à ouvrir le marché à la concurrence pour permettre aux assurés de bénéficier de tarifs plus attractifs. Sur la même période, le marché du crédit immobilier a été dynamisé par une baisse significative des taux – même si la tendance s'est inversée depuis un peu plus d'un an, et par des incitations fiscales d'investissement sur le marché locatif.

Ce contexte favorable aux assurés emprunteurs les a incités à revoir à la fois leurs conditions d'emprunt et leur ADE et s'est traduit par une volatilité des renégociations et des résiliations directes des contrats d'assurance, ou indirectes via le rachat du crédit.

La modélisation actuelle des résiliations, basée sur des données rétrospectives capte insuffisamment cette volatilité. L'objectif visé est d'utiliser le potentiel des méthodes statistiques prédictives pour modéliser dynamiquement le risque de rachat au cours du temps. Le travail proposé présente la construction du modèle à partir des bases d'étude jusqu'à son back-testing dans un cadre solvabilité 2, en analysant ses forces et ses limites.

Mots-clés : SAS, méthodes de classification supervisées, DISQUAL (discrimination sur données qualitatives), LOGIT, ROC (receiver operating characteristic), AUC (Area Under the (ROC) Curve), Score, Hoem, Whittaker-Henderson, Solvabilité 2, Prophet, Model-Points, BEL (Best Estimate Liability), SCR (Solvency Capital Requirement)

Abstract

Borrower insurance (ADE) is one of the most important areas of personal insurance on the French market, with a turnover in 2020 of €10 billion, including almost €7 billion for the coverage of real estate loans, a market that has been steadily growing for many years (+55% since 2007), in relation to outstanding loans. In 2021, Crédit Agricole Assurance represents almost 21% of the turnover of the ADE in France.

The borrower insurance contract consists of the cover of one or more natural persons (insured, co-insured or guarantor) against a potential risk of insolvency linked to an event (death, stoppage of work, disability or, more rarely, loss of employment) which potentially reduces their ability to repay a loan taken out with a lender.

Since 2011, the market for borrower insurance has undergone numerous technical and regulatory developments, aimed at opening up the market to competition to allow policyholders to benefit from more attractive prices. Meanwhile, the mortgage market has been buoyed by a significant rate cut - though one that has been reversing for a little more than a year - and by tax incentives for investment in rental housing.

This favorable environment for insured borrowers has prompted them to review both their borrowing conditions and their ADE, and has resulted in volatility in renegotiations and direct or indirect cancelations of insurance contracts through the repurchase of crédit.

Current modeling of the resilience, based on retrospective data, does not sufficiently capture this volatility. The aim is to use the potential of predictive statistical methods to dynamically model buy-back risk over time. The proposed work presents the construction of the model from the study bases to its back-testing in a solvency 2 framework, analyzing its strengths and limitations.

Keywords : SAS, supervised classification methods, DISQUAL (qualitative discrimination), LOGIT, ROC (receiver operating characteristic), AUC (Area Under the (ROC) Curve), Score, Hoem, Whittaker-Henderson, Solvency 2, Prophet, Model-Points, BEL (Best Estimate Liability), SCR (Solvency Capital Requirement)

Note de synthèse

1 Le contexte

L'assurance des emprunteurs (ADE) représente l'un des domaines de l'assurance de personnes les plus importants sur le marché français, avec un chiffre d'affaires en 2020 d'un peu plus de 10 milliards d'euros, dont près de 7 milliards d'euros au titre de la couverture des prêts immobiliers, marché en croissance régulière depuis de nombreuses années (+55% depuis 2007), en lien avec les encours crédit. En 2021, Crédit Agricole Assurance représente près de 21% du chiffre d'affaire de l'ADE en France.

Le contrat d'assurance emprunteur consiste en la couverture d'une ou plusieurs personnes physiques (assuré, co-assuré voire caution) contre un risque potentiel d'insolvabilité lié à un événement (décès, arrêt de travail, invalidité ou, plus rarement perte d'emploi) venant potentiellement réduire leurs capacités de remboursement d'un prêt souscrit auprès d'un organisme prêteur.

Depuis 2011, le marché de l'assurance emprunteur a connu de nombreuses évolutions techniques et réglementaires, visant à ouvrir le marché à la concurrence pour permettre aux assurés de bénéficier de tarifs plus attractifs. La loi Lagarde (2011), complétée par la loi bancaire (2013) permet le découplage du crédit immobilier et de l'assurance emprunteur. La loi Hamon (2014) met en place un droit au changement d'assurance emprunteur dans les 12 mois suivant la signature du crédit. La définition de l'Équivalence du niveau des garanties par le CCSF (2015) clarifie les offres du marché. L'Amendement Bourquin (2019) donne un droit de substitution annuel des contrats d'assurance emprunteur à la date anniversaire de la signature de l'offre de prêt.

Depuis 2007, le marché du crédit immobilier a été de son côté dynamisé par une baisse significative des taux (-78%), même si une inversion de la tendance est constatée à partir de 2022, et des incitations fiscales d'investissement sur le marché locatif, avec successivement les lois Scellier (2009-12), Duflot (2013-14) et le dispositif Pinel à partir de 2015.

Ce contexte favorable aux assurés emprunteurs les a incités à revoir à la fois leurs conditions d'emprunt et leur ADE et s'est traduit par une augmentation des renégociations et des résiliations directes des contrats d'assurance, ou indirectes via le rachat du crédit.

La méthode de modélisation actuelle des résiliations est basée sur des études rétrospectives actualisées dès lors que les indicateurs de suivi du risque mettent en avant une défaillance de ces dernières. Elle sous-tend l'hypothèse que les rachats des contrats emprunteur adossés à un prêt immobilier peuvent avoir deux origines différentes, une composante, liée au rachat du prêt sous-jacent et une composante spécifique à l'assurance.

Cependant, la méthode utilisée paraît insuffisante pour capter la volatilité des rachats au cours du temps, liée à celles des deux composantes identifiées ci-dessus, compte tenu des volatilités de taux constatés.

Une approche alternative de modélisation dynamique, prenant en compte les effets combinés de l'évolution du contexte économique et du comportement des assurés est proposée avec comme objectif d'améliorer l'ajustement des lois du modèle.

2 La construction des lois de résiliation

La définition des nouvelles lois de résiliation nécessite dans un 1^{er} temps de construire les bases d'étude avec les données pertinentes disponibles, de procéder à une analyse statistique prédictive des bases et d'utiliser ensuite les résultats de l'étude dans la modélisation des résiliations qui seront mises en œuvre dans un cadre Solvabilité 2.

2.1 Les étapes de construction des bases

Les bases s'appuient à la fois des données tête par tête de primes mensuelles des Caisses régionales du Crédit Agricole, extraites trimestriellement et sur les bases de survénances de sinistres tête par tête (décès, PTIA¹, ITD²) de Crédit Agricole Assurances (CAA).

Les résiliations sont déduites de manière différentielle entre 2 photos trimestrielles consécutives (T et $T + 1$). Elles correspondent aux observations qui disparaissent (rachats du crédit) entre T et $T + 1$ sans que cette disparition ne soit liée à la fin contractuelle du prêt de l'ADE, ou à l'occurrence d'un sinistre (décès, PTIA ou ITD) et aux observations dont la prime tombe entre T et $T + 1$ (résiliations ADE directes).

Deux bases annuelles agrégées sont utilisées pour la modélisation. Une 3^e base annuelle sert à back-tester la modélisation.

Le choix de la période d'étude répond à 2 contraintes. Les offres et les modes distributifs doivent être stabilisés et les sinistres décès, PTIA et ITD doivent être connus de manière exhaustive afin d'éviter d'assimiler des résiliations à des sinistres (cas des IBNR).

L'analyse de la distribution des effectifs par ancienneté conduit à regrouper les anciennetés supérieures ou égales à 15 ans.

2.2 La modélisation

Une étude statistique prédictive par ancienneté des contrats est réalisée. Les résultats des segmentations de l'étude sont ensuite utilisés pour construire les lois de résiliations.

2.2.1 L'étude statistique prédictive

Compte tenu des volumétries en jeu l'étude est réalisée sur un échantillonnage à 10% des bases annuelles construites précédemment, stratifié par Caisse régionale et ancienneté de contrat.

Elle a pour objectif d'expliquer, pour chaque ancienneté a , la variable (Y) qualitative « cible » de résiliation annuelle (présence ou départ du client dans l'année) à partir d'un certain nombre de variables explicatives ou prédictives ($\{X_k\}_{1 \leq k \leq p}$) à la fois quantitatives et qualitatives (données d'assurance de crédit ou socio-démographiques INSEE...).

1. Perte Totale et Irreversible d'Autonomie

2. Invalidité Totale Définitive

Le risque de résiliation dans l'année est quantifié par une note de risque S , ou score qui est calculé comme combinaison linéaire des variables explicatives ($\{X_k\}_{1 \leq k \leq p}$).

$$S_a = \sum_{1 \leq k \leq p} a_k X_k$$

les coefficients $\{a_k\}_{1 \leq k \leq p}$ étant optimisés pour la prédiction de Y .

Les méthodes supervisées de classement choisies pour l'étude sont la DIScrimination sur variables QUALitatives » (DISQUAL) et la régression logistique binaire (LOGIT).

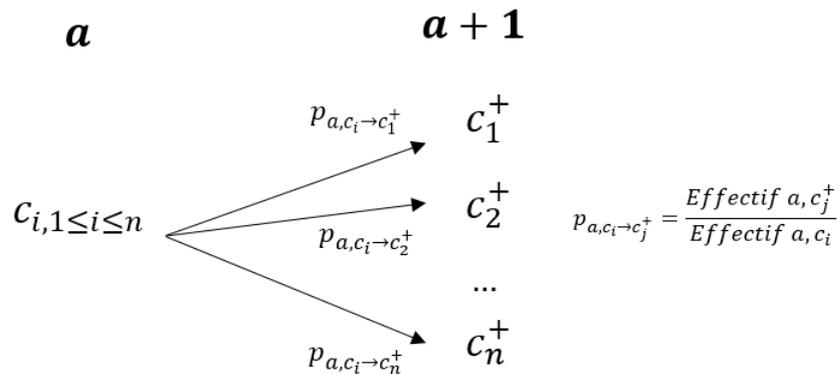
La performance des modèles issus de ces méthodes est évaluée par l'analyse des AUC et des courbes ROC. Pour chaque ancienneté, le modèle de la méthode la plus performante est retenu pour définir les scores de résiliation.

2.2.2 Les trajectoires et les clusters

La segmentation des scores par ancienneté construits précédemment permet de définir les classes de risque (c_1, c_2, \dots, c_n) des observations du modèle. L'étude ne traite que les segmentations à 3 et 4 classes.

Les taux bruts de résiliations t_{a,c_i} associés à chaque classe (c_i) de et chaque ancienneté a sont estimés à partir du niveau de segmentation fixé ci-dessus.

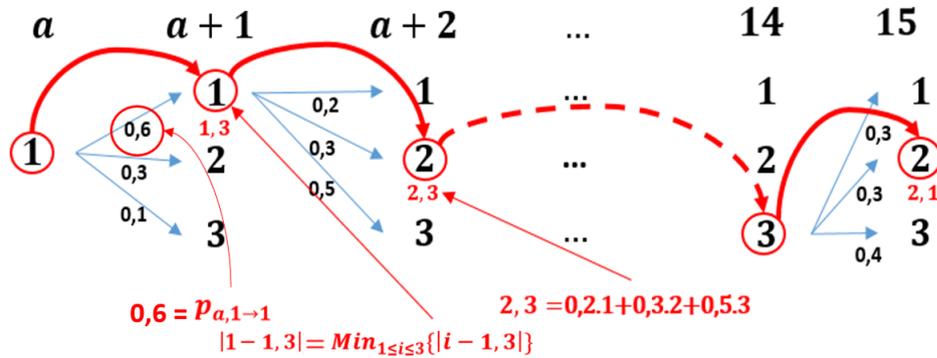
Les classes de risque ($c_1^+, c_2^+, \dots, c_n^+$) des observations d'ancienneté a « vieilles » d'un an (o_{a+1}) sont calculées à partir des scores des observations d'ancienneté $a+1$ avec des inducteurs ajustés ($S_{a+1}(o_{a+1})$). L'association des classes de risque en début d'année a aux classes des observations vieilles d'un an $a+1$ permet de définir une matrice de changement d'états :



La classe de transition entre a et $a+1$ ($c_{a+1,i}$), partant de la classe c_i est définie comme la classe de risque la plus « proche » de l'espérance de changement d'état entre a et $a+1$ ($\sum_{1 \leq k \leq n} p_{a,c_i \rightarrow c_k^+} c_k^+$).

La trajectoire entre a et 15 partant de la classe c_i est définie de manière itérative comme le $(15-a)$ -uplet $(c_i, c_{a+1,i}, c_{a+2,i}, \dots, c_{15,i})$.

L'exemple suivant, avec une granularité à 3 classes synthétise la construction d'une trajectoire (1, 1, 2... , 3, 2) à partir de a et de la classe 1 :



Enfin les clusters correspondent à des regroupements de trajectoires identiques, comme l'illustre l'exemple suivant dans le cas d'une granularité à 3 classes de risques :

	a	...	13	14	15	}	Cluster
<i>Traject₁</i>	3	...	3	3	2		
<i>Traject₂</i>			3	3	2		
<i>Traject₃</i>				3	2		
<i>Traject₄</i>					2		

Les lois de résiliations sont obtenues pour chaque cluster après une étape de lissage des taux bruts associés aux classes de risques calculées précédemment.

Les scénarii économiques sont intégrés à la construction des trajectoires en ajustant les seuils des inducteurs quantitatifs qui définissent les scores.

3 La mise en œuvre

3.1 Les résultats

Les lois dynamiques de résiliation sont analysées dans un cadre Solvabilité 2, sur le model-point de back-test construit précédemment et sont comparées aux lois historiques ainsi qu'aux résiliations réellement constatées. L'étude s'intéresse aux projections des effectifs de contrats, des flux financiers et enfin aux best estimate (BEL) et aux coûts du risque.

La loi segmentée à 3 classes de risques est mise en œuvre dans un 1^{er} temps sans y intégrer de scénario d'évolution de contexte économique (scénario central); les flux de rachats induits sont comparés à ceux résultant des lois de rachats historiques, ainsi qu'au flux réel de rachats constatés à un an. Les effectifs résultant de la loi dynamique se rapprochent davantage du réel que ceux obtenus avec la loi historique. L'affinement de la granularité de la segmentation (passage de 3 à 4 classes de risque) est sans impact. significatif.

La loi segmentée à 3 classes de risques est ensuite étudiée en intégrant des scénarii d'évolutions économiques mono ou multifactoriels, impliquant des indicateurs crédit (taux de crédit) ou ADE (âges des assurés ou taux ADE). Les tests mettent en évidence le potentiel dynamique et la sensibilité de la loi segmentée aux contextes modélisés.

Enfin le Best Estimate (BEL) et le capital requis (SCR), pour les garanties décès et arrêt de travail, résultant de la loi segmentée sont comparés à ceux résultant de la loi historique. Les BEL sont globalement assez proches. En revanche, la loi segmentée dynamique est un peu plus exigeante en fonds propres que la loi historique avec des SCR plus élevés.

3.2 Les limites et les pistes d'optimisations

La modélisation alternative centrale présente une qualité d'ajustement modérée au réel, qui peut être optimisée.

Cette qualité peut être liée à la performance des prédictions définissant les scores. Une piste de réponse à ce problème serait de revoir l'échantillonnage sur lequel est réalisé la modélisation, en procédant à un rééquilibrage des effectifs. Ce rééquilibrage pourrait se faire, par exemple en sous-échantillonnant les contrats non rachetés en ne gardant pour chaque contrat racheté (caractérisé par sa date et son agence de souscription) que les contrats souscrits dans la même agence et au même moment.

Elle peut avoir pour origine le principe même de construction des trajectoires. La construction s'appuie sur des ajustements des seuils des inducteurs des scores qu'il conviendrait de back-tester et d'optimiser dans le temps.

L'écart au réel peut s'expliquer enfin par les probabilités de rachats héritées de la base de modélisation. Dans notre exemple la base du modèle est caractérisée par un volume de rachats particulièrement élevée par rapport à la base de back-test qui tend à augmenter globalement les probabilités de rachats, malgré la segmentation. Un paramétrage dynamique adapté des scénarii, ci-dessus, peut cependant corriger et améliorer les performances du modèle.

1 The context

Borrower insurance (ADE) is one of the most important areas of personal insurance on the French market, with a turnover in 2020 of €10 billion, including almost €7 billion for the coverage of real estate loans, a market that has been steadily growing for many years (+55% since 2007), in relation to outstanding loans. In 2021, Crédit Agricole Assurance represents almost 21% of the turnover of the ADE in France.

The borrower insurance contract consists of the cover of one or more natural persons (insured, co-insured or guarantor) against a potential risk of insolvency linked to an event (death, disability or, more rarely, loss of employment) which potentially reduces their ability to repay a loan taken out with a lender.

Since 2011, the market for borrower insurance has undergone numerous technical and regulatory developments, aimed at opening up the market to competition in order to allow policyholders to benefit from more attractive prices : the Lagarde Act (2011), supplemented by the Banking Act (2013), allows the decoupling of mortgage crédit and borrower insurance. The Hamon Act (2014) establishes a right to change borrower insurance within 12 months of signing the crédit. The CCSF definition of Equivalence in the Level of Collateral (2015) clarifies the market offers. The Bourquin Amendment (2019) gives an annual right of substitution for borrower insurance contracts on the anniversary date of the signature of the loan offer.

Since 2007, the mortgage market has also been boosted by a significant drop in rates (-78%), until a reversal of the trend from 2022, and tax incentives for investment in the rental market, with successive laws such as Scellier (2009-12), DufLOT (2013-14) and Pinel from 2015.

This favorable environment for insured borrowers has prompted them to review both their borrowing conditions and their ADE and has led to an increase in renegotiations and direct or indirect cancelations of insurance contracts through the repurchase of crédit.

The current method of modeling the resilience shall be based on retrospective studies updated whenever the risk monitoring indicators point to a failure of the latter. It is hypothesized that buybacks of mortgage-backed borrower contracts can have two different origins : a component, related to the buyback of the underlying loan, and an insurance-specific component.

However, the method used seems insufficient to capture the volatility of redemptions over time, linked to those of the two components identified above, taking into account the rate volatilities observed.

An alternative dynamic modeling approach, taking into account the combined effects of changing economic conditions and policyholder behavior, is proposed with the aim of improving the adjustment of model laws.

2 The construction of termination laws

The definition of the new termination laws requires in a first step to construct the study bases with the available relevant data, to carry out a statistical analysis predictive of the bases and then to use the results of the study (termination scores) in the modeling of the resiliations that will be implemented in a Solvency 2 framework.

2.1 Stages of building the bases

The bases are based both on the monthly premium head-to-head data of the Caisses régionales du Crédit Agricole, extracted on a quarterly basis, and on the individual claims (death, PTIA³, ITD⁴) of Crédit Agricole Assurances (CAA).

Terminations are deducted differentially between 2 consecutive quarterly photos (T and $T + 1$) : they correspond to the observations that disappear (redemptions of the crédit) between T and $T + 1$ without this disappearance being linked to the contractual end of the ADE, or loan, or in the event of a loss (death, PTIA or ITD) and observations with a premium between T and $T + 1$ (direct ADE terminations).

Two aggregate annual bases are constructed that are used for modeling and one last annual base that is used to back-test modeling.

The choice of study period meets two constraints : the offers and distribution methods must be stabilized and the loss of life, PTIA and ITD must be known exhaustively in order to avoid assimilating resiliations to claims (the case of IBNRs...).

The analysis of the distribution of staff by seniority leads to the grouping of seniority over or equal to 15 years.

2.2 Modeling

A predictive statistical study by seniority of contracts is carried out. The results of the segmentations of the study are then used to construct the laws of termination.

2.2.1 The predictive statistical study

Taking into account the volumes at stake, the study is carried out on a sampling at 10% of the annual bases built previously, stratified by regional fund and seniority of contract.

or predictive variables ($\{X_k\}_{1 \leq k \leq p}$) both quantitative and qualitative (crédit insurance data or socio-demographic data INSEE...).

The termination risk in the year is quantified by a S risk score, or score that is calculated as a linear combination of the ($\{X_k\}_{1 \leq k \leq p}$) explanatory variables.

$$S_a = \sum_{1 \leq k \leq p} a_k X_k$$

the coefficients $\{a_k\}_{1 \leq k \leq p}$ being optimized for the prediction of Y .

3. total and irreversible loss of autonomy

4. total and irreversible incapacity

The supervised classification methods chosen for the study are DISQUAL⁵ and LOGIT.

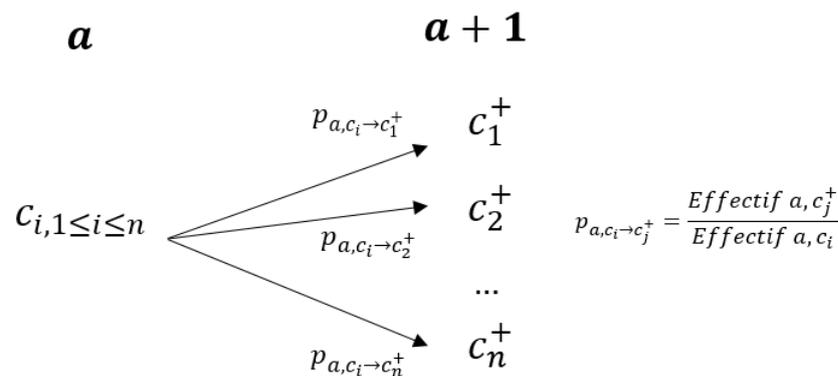
The performance of the models derived from these methods is assessed by analysis of AUC and ROC curves. For each seniority, the model of the best performing method is used to define the termination scores.

2.2.2 Trajectories and clusters

The segmentation of the previously constructed age scores makes it possible to define the (c_1, c_2, \dots, c_n) risk classes of the observations of the model. The study only deals with 3 and 4 class segments.

Gross termination rates t_{a,c_i} associated with each class (c_i) of and seniority a are estimated from the segmentation level set above.

Risk classes $(c_1^+, c_2^+, \dots, c_n^+)$ of one-year old a -age observations (o_{a+1}) are calculated from $a+1$ -age observation scores with adjusted inducers $(S_{a+1}(o_{a+1}))$. The association of the risk classes at the beginning of the year a with the classes of observations aged for one year $a + 1$ makes it possible to define a state change matrix :

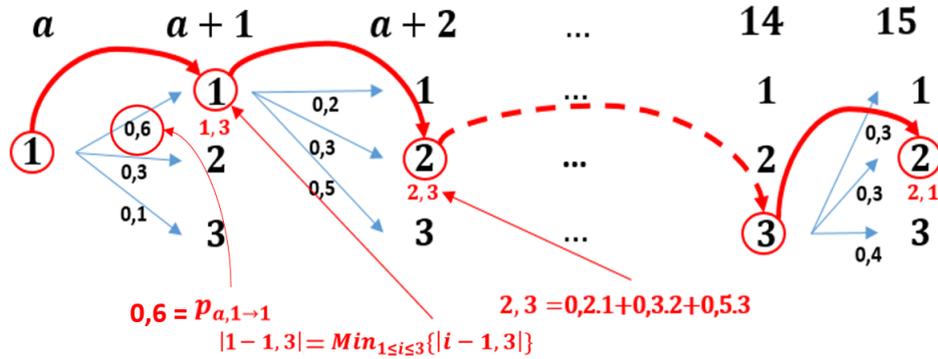


La classe de transition entre a et $a + 1$ $(c_{a+1,i})$, partant de la classe c_i est définie comme la classe de risque la plus « proche » de l'espérance de changement d'état entre a et $a + 1$ $(\sum_{1 \leq k \leq n} p_{a,c_i \rightarrow c_k^+} c_k^+)$.

The trajectory between a and 15 starting from the c_i class is iteratively defined as the $(15 - a)$ -tuple $(c_i, c_{a+1,i}, c_{a+2,i}, \dots, c_{15,i})$.

5. Discrimination on qualitative variables

The following example, with a 3-class granularity, summarizes the construction of a path (1.1.2...3.2) from a and class 1 :



Finally, clusters correspond to groupings of identical trajectories, as illustrated by the following example in the case of a 3 risk class granularity :

	a	...	13	14	15	
<i>Traject₁</i>	3	...	3	3	2	}
<i>Traject₂</i>			3	3	2	
<i>Traject₃</i>				3	2	
<i>Traject₄</i>					2	
						Cluster

The termination laws are obtained for each cluster after a step of smoothing the gross rates associated with the risk classes calculated previously.

The economic scenarios are integrated into the construction of the trajectories by adjusting the thresholds of the quantitative inducers which define the scores.

3 Implementation

3.1 The results

Dynamic termination laws are analyzed in an Solvency 2 framework, on the back-test model-point constructed previously, and are compared with historical laws as well as with actual terminations. The study looks at projections of contract headcount, financial flows, and best estimates (BELs) and risk costs.

The segmented law with 3 risk classes is implemented in a first stage without including a scenario of evolution of the economic context (central scenario); the induced redemption flows are compared to those resulting from historical redemption laws, as well as to the real redemption flow observed at one year : the numbers resulting from the dynamic law are closer to the real than those obtained with the historical law. Refining the granularity of segmentation (from 3 to 4 risk classes) has no significant impact.

The law segmented to 3 risk classes is then studied by integrating scenarios of single or multifactorial economic developments, involving crédit indicators (crédit rate) or ADE (age of the insured or ADE rate) : the tests highlight the dynamic potential and sensitivity of the segmented law to the modeled contexts.

Finally, the Best Estimate (BEL) and the Capital Requirement (SCR) - for death and work stoppage guarantees, resulting from the segmented law are compared with those resulting from the historical law : the BELs are generally quite similar ; on the other hand, the dynamic segmented law is slightly more demanding in terms of equity than the historical law with higher SCRs.

3.2 Limitations and optimization paths

The central alternative modeling has a moderate fit quality to the real, which can be optimized.

This may be related to the quality of the predictions defining the scores : one answer to this problem could be to review the sampling on which the modeling is carried out, by rebalancing the number of people. This could be done, for example, by sub-sampling unrepurchased contracts, leaving for each contract repurchased - characterized by its date and agency of subscription - only contracts signed in the same agency and at the same time.

Another reason may be the very principle of trajectory construction. The construction is based on adjustments to the driver thresholds of the scores that should be back-tested and optimized over time.

The limits can finally be explained by the levels of redemption probabilities inherited from the modeling base : in our example the model base is characterized by a particularly high volume of redemptions compared to the back-testing base which tends to increase the redemption probabilities overall, despite the segmentation. Dynamic parameterization adapted from the above scenarios can however correct and improve the performance of the model.

Remerciements

Cette étude a été réalisée au sein de la direction de l'Actuariat de Predica, société d'assurance vie du Groupe Crédit Agricole.

Mes remerciements vont tout d'abord à Omblin de Baudinière et Thibault Müller, de l'unité d'Actuariat produit prévoyance-emprunteur, pour leurs conseils, leur disponibilité et leurs encouragements.

Je remercie Laure Battistelli responsable du service modélisation et analyses prospectives pour ses remarques et Junior Fonkoua pour son accompagnement sur l'outil Prophet.

J'exprime ma gratitude à Anaïd Chahinian directrice de l'Actuariat de Predica, Fabrice Diakhate responsable du service inventaire et Stéphanie Gautier-Mavel pour leur soutien au cours de ce travail.

Je remercie l'équipe pédagogique du CNAM pour la qualité des enseignements dispensés.

Enfin, je tiens à témoigner de ma reconnaissance à mes collègues et à toutes les personnes qui ont contribué d'une manière ou d'une autre à la réalisation de ce mémoire.

Mémoire

Table des matières

1	Introduction générale	4
2	Le marché de l’ADE des crédits immobiliers	5
2.1	Le contrat d’assurance emprunteur	5
2.2	Un marché ADE habitat à maturité	7
2.3	Un contexte favorable pour le consommateur	8
2.3.1	Un marché des crédits immobiliers dynamique	8
2.3.2	Un cadre réglementaire qui favorise l’ouverture du marché de l’ADE	9
2.4	L’exposition des bancassureurs au risque de résiliation ADE	10
2.4.1	Les résiliations d’ordre structurel	10
2.4.2	Les résiliations d’ordre conjoncturel	10
2.5	L’exposition des bancassureurs au risque de résiliation ADE	11
2.5.1	Le Crédit Agricole un acteur majeur de l’ADE en France	11
2.5.2	Les crédits immobiliers couverts au Crédit Agricole	12
2.5.3	Les produits ADE des Caisses régionales et la distribution	13
2.5.3.1	Les produits ADE des Caisses régionales et la distribution	13
2.5.4	La modélisation actuelle des rachats	15
2.5.4.1	La composante conjoncturelle	15
2.5.4.2	La composante assurantielle	15
2.5.5	Les limites et recherche d’une méthode alternative	16
2.6	Conclusion	18
3	Les bases d’étude	19
3.1	Le choix de la période d’étude	19
3.2	Analyse structurelle préliminaire des crédits distribués par des Caisses régionales	20
3.3	Principes généraux de construction des bases d’études	20
3.3.1	Reconstitutions des flux et enrichissements	20
3.3.2	Traitement des valeurs manquantes ou aberrantes	21
3.3.3	Agrégation et échantillonnage	22
3.3.4	Contrôle et calage à la comptabilité	23
3.3.5	Contrôle de la qualité de l’ajustement de l’échantillon	23
3.3.6	Description de la structure des données	24
3.3.7	Analyse descriptive de quelques variables	25
3.3.8	Etude de la dynamique des indicateurs crédit et ADE mesurée dans les fichiers sources entre juillet 2015 et juillet 2019	25
3.3.8.1	Evolution des taux de crédits immobiliers effectifs – crédits amortissables	25
3.3.8.2	Evolution des taux ADE immobilier annualisés effectifs – couvrant les prêts amortissables	26
3.4	Conclusion	26

4	Modélisation des rachats en assurance emprunteur	27
4.1	Le cadre technique	27
4.1.1	L'existant	27
4.1.2	L'approche alternative	28
4.2	L'approche par des méthodes statistiques prédictives	29
4.3	Le scoring	30
4.3.1	Mise en œuvre	30
4.3.1.1	Etudes statistiques univariées	30
4.3.1.2	Etudes statistiques bivariées	31
4.3.1.3	Analyse des Correspondances Multiples (ACM)	32
4.3.1.4	Mise en œuvre des méthodes	32
4.4	Calcul des taux bruts moyens par ancienneté	38
4.5	Clustering des observations	39
4.5.1	Le calage des seuils	39
4.5.1.1	L'ajustement des seuils du modèle	40
4.5.1.2	L'ajustement des seuils de test	44
4.5.1.3	L'ajustement des seuils d'âges	45
4.5.2	Les transitions annuelles	45
4.5.3	Les matrices de transition	45
4.5.4	Les trajectoires	46
4.5.5	Les taux de rachats bruts et les effectifs moyens associés à chaque classe de risque $C_{a_0,p}$	49
4.5.6	Clustering	50
4.5.7	Le lissage des taux	51
4.6	Construction des Model-Points emprunteur (MP)	51
4.7	Conclusion	52
5	Mise en œuvre du modèle	54
5.1	La directive solvabilité 2	54
5.1.1	La Best Estimate Liability ou provision Best Estimate (BEL)	54
5.1.2	Le capital de solvabilité requis (SCR)	55
5.1.2.1	Le SCR global	56
5.1.2.2	Le SCR module	56
5.1.2.3	Le SCR sous-module	57
5.1.3	Zoom sur l'assurance emprunteur	57
5.2	Le modèle prophète	58
5.2.1	Cinématique et prise en compte des rachats dans le modèle	58
5.2.2	Chocs sur les rachats	60
5.2.3	Principe de calcul des effectifs	61
5.3	Etude de la sensibilité des rachats projetés à la loi de modélisation utilisée	62
5.3.1	Prérequis	62
5.3.2	Analyse des projections des effectifs de contrats rachetés	62
5.3.2.1	Impact de la granularité de la segmentation	62
5.3.2.2	Lois historiques et loi segmentée centrale à 3 modalités	63
5.3.2.3	Loi segmentée « centrale » et lois segmentées « dynamiques »	64
5.3.3	Impacts de l'approche alternative sur les effectifs de contrats et les flux financiers	67
5.3.4	Impacts de l'approche alternative sur les calculs de BEL et SCR de rachat des garanties décès et arrêt de travail	69
5.4	Conclusion	73
6	Conclusion générale	74

7	Annexes	76
7.1	Le marché de l’ADE des crédits immobiliers	76
7.1.1	Evolution de la réglementation	76
7.1.2	Etude de la dynamique des indicateurs crédit et ADE mesurée dans les fichiers sources entre juillet 2015 et juillet 2019	77
7.1.2.1	Evolution des taux de crédits immobiliers effectifs – crédits amortissables	77
7.1.2.2	Evolution des taux ADE immobilier annualisés effectifs – couvrant les prêts amortissables	78
7.2	Les bases d’étude	80
7.2.1	Sélection des agences les plus contributrices de la CR 802 pour l’ancienneté 5 au seuil de 10%	80
7.2.2	Proportion des agences (en encours crédit) retenues dans l’échantillonnage par ancienneté	80
7.2.3	Proportion des contrats filtrés dans l’étude comportementale	81
7.2.4	Décomposition en % de chiffre d’affaires et source d’exposition partielle	81
7.2.5	Distribution des effectifs (en nombre) par ancienneté	82
7.2.6	Réconciliation des primes décès de la base d’étude avec la comptabilité	82
7.2.7	Réconciliation des primes ITT de la base d’étude avec la comptabilité	83
7.2.8	Contribution du stock et des affaires nouvelles au chiffre d’affaires annuel dans la base d’étude	83
7.2.9	Ajustement global entre la base échantillonnée et la base totale	84
7.2.10	Piste d’optimisation de l’ajustement	84
7.2.11	Présentation des données	85
7.2.11.1	Structure des données	85
7.2.11.2	La définition des spreads de taux de crédit ou d’ADE	85
7.2.11.3	La description de la typologie des communes INSEE	86
7.2.12	Stratégie de gestion des valeurs manquantes ou aberrantes	87
7.2.13	Analyse descriptive de quelques variables	88
7.3	Modélisation des rachats en assurance emprunteur	92
7.3.1	Eléments théoriques	92
7.3.1.1	La discriminante qualitative	92
7.3.1.2	La regression logistique (LOGIT)	94
7.3.2	La mise en œuvre pour l’ancienneté 6	96
7.3.2.1	L’analyse des correspondances multiples (ACM) pour l’ancienneté 6	96
7.3.2.2	Méthode DISQUAL	97
7.3.2.3	La régression logistique	98
7.3.3	V de Cramer	100
7.3.4	Lambda de Wilks	100
7.3.5	Courbe ROC	101
7.3.6	AUC et interprétation	101
7.3.7	Synthèse des courbes ROC par anciennetés des méthodes DISQUAL et LOGIT	102
7.3.8	La grille de score	104
7.3.9	Le lien entre les taux de rachats et les scores	104
7.3.10	Calcul des facteurs d’ajustement des seuils des inducteurs	105
7.3.10.1	L’ajustement des seuils du modèle	105
7.3.10.2	L’ajustement des seuils de tests	105
7.3.11	Principe de calcul des taux bruts : l’estimateur de Hoem	106
7.3.12	Evolution des drivers d’ajustement des seuils de juillet 2015 à juillet 2019	107
7.3.13	Trajectoires modélisées par anciennetés pour une segmentation à 3 classes de risques	109

7.3.14	La matrice de transition	111
7.3.15	La table des trajectoires	111
7.3.16	Principe du Lissage des lois par la méthode whittaker-Henderson	112
7.3.16.1	Les résultats du lissage pour les segmentations à 3 et 4 classes de risques	113
7.4	Le back-testing	114
7.4.1	L'application de la grille de score en 1 ^{re} année	114
7.4.2	Modélisation des rachats	116

Chapitre 1

Introduction générale

L'assurance des emprunteurs (ADE) intervient lors de la souscription d'un prêt, le plus souvent pour un prêt immobilier. Il s'agit d'une assurance permettant à l'emprunteur de se protéger contre les différents risques susceptibles d'empêcher le remboursement de ses mensualités. Parmi les risques, sont principalement couverts le risque décès systématiquement, l'incapacité de travail et la perte d'emploi. Elle protège à la fois l'emprunteur et le prêteur en se substituant à l'emprunteur pour le remboursement des mensualités ou du solde du crédit en cas de sinistre. Bien que non obligatoire, cette assurance est exigée par les établissements de crédit dans la quasi totalité des prêts immobiliers. Le Best Estimate Liability (BEL) lié à la mise en place de la directive-cadre Solvabilité 2 se doit, comme son nom l'indique, de refléter la meilleure estimation de l'engagement de l'assureur. Dans le cadre de la garantie des contrats d'assurance emprunteur, l'évaluation du BEL dépend de deux grandes familles de paramètres :

1. d'ordre technique : taux de décès, d'invalidité, d'incapacité et d'actualisation
2. d'ordre conjoncturel et structurel :
 - lié à l'objet de l'assurance : le taux de remboursement anticipé ou de rachat du crédit
 - lié directement au contrat d'assurance : le taux de substitution

Du fait du contexte d'évolution rapide des taux d'emprunt et de l'ouverture récente et continue du marché de l'assurance emprunteur favorisée par les évolutions législatives, nous nous intéresserons dans ce qui suit à l'étude de la deuxième famille de risque limitée à l'assurance des prêts immobiliers distribués par les Caisses régionales du Crédit Agricole qui représente la plus grosse partie du marché de l'assurance emprunteur et qui est également la plus exposée aux évolutions ; nous regrouperons indifféremment sous le qualificatif de rachat ou de résiliation, les sorties de portefeuille relevant de ce risque conjoncturel.

L'objectif du mémoire est de proposer une optimisation de la modélisation actuelle des lois de résiliations des contrats emprunteurs, en utilisant des méthodes prédictives supervisées. L'étude se décompose en quatre parties :

- Une première partie présentera le contexte actuel du marché crédit et assurance emprunteur sur le domaine immobilier en général et le dispositif du Crédit Agricole en particulier.
- Une deuxième partie présentera la construction des bases d'étude et les données disponibles.
- Une troisième partie sera consacrée à la présentation la modélisation alternative des rachats à partir des bases d'études construites précédemment.
- Une quatrième partie sera consacrée à la mise en œuvre des nouvelles lois dans l'outil prophet, et à leur évaluation.
- Enfin une conclusion proposera des pistes d'optimisations.

Chapitre 2

Le marché de l'ADE des crédits immobiliers

2.1 Le contrat d'assurance emprunteur

Lorsqu'un individu souhaite obtenir un crédit, l'organisme prêteur exige de l'individu la souscription d'une assurance emprunteur. Cette souscription peut se faire auprès de la compagnie d'assurance du prêteur s'il s'agit d'un bancassureur ou d'un assureur externe sous réserve de respecter certaines conditions comme l'équivalence des garanties.

Les éléments définis à la souscription, indispensables à la modélisation des risques, sont les suivants :

- L'emprunteur doit remplir un questionnaire de santé ou une déclaration d'état de santé. Ce n'est toutefois plus le cas, pour certains crédits, à partir du 1^{er} Juin 2022, grâce à l'entrée en vigueur de la loi n° 2022-270¹ ou loi Lemoine (crédits <200 k€ et arrivant à terme avant les 60 ans de l'assuré).
- L'âge de l'individu à la souscription doit être inférieur à des seuils définis selon le risque couvert : c'est ce qu'on appelle les limites de garanties à l'entrée dans l'assurance. Au Crédit Agricole, les limites d'âge sont de 70 ans pour le décès, 60 ans pour l'arrêt de travail et 50 ans pour la perte d'emploi.
- L'individu peut décider ou se voir imposer de se couvrir contre les risques décès, la Perte Totale et Irréversible d'Autonomie, l'Incapacité Temporaire de travail Partiel ou Totale, l'Invalidité Permanente Partielle ou Totale D'autres garanties, moins fréquemment souscrites, sont également possibles : Perte d'Emploi, Hospitalisation, Maladies critiques, etc...
- L'individu définit la quotité de son assurance emprunteur, qui représente le pourcentage du capital emprunté couvert par l'assureur : celle-ci peut aller de 1 à 100% par co-emprunteur. A noter enfin que certains produits ne proposent pas de couverture pour des crédits.

1. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000045268729>

Ainsi, une fois l'individu assuré, les modalités de paiement des cotisations sont définies :

- Il existe deux tarifications possibles, selon que les cotisations sont définies sur le Capital Initial emprunté (CI), ou sur le Capital Restant Dû (CRD), montant du capital que l'assuré doit encore rembourser à une certaine date d'évaluation. Lors de la tarification, l'assureur prend en compte le questionnaire de santé ou la déclaration d'état de santé, afin d'évaluer si oui ou non, l'assuré fera l'objet d'une tarification de risque aggravé.
- L'assuré peut décider ou se voir imposer la fréquence de paiement de ses cotisations. L'échéancier des cotisations ADE peut être adossé ou non à celui du remboursement du financement couvert qui peut être à amortissement mensuel, trimestriel, semestriel ou annuel. Il est également possible pour l'assuré de choisir entre les deux options suivantes : l'option « tarif garanti » qui oblige l'assuré à payer le niveau de cotisations convenu au moment de la signature du contrat, et donc ne permet aucun ajustement, même en cas d'aggravation de l'état de santé de l'assuré ; et l'option « tarif révisable » par suite d'une baisse ou d'une hausse du niveau du risque assuré, ce qui permet entre autres à l'assureur de réviser de façon unilatérale le niveau de cotisations. Au Crédit Agricole, le tarif à l'entrée est garanti durant la durée de couverture. L'échéancier de paiement des cotisations est mensuel, indépendamment de la périodicité du financement...
- Il est possible ou non de mettre en place une exonération des cotisations lorsque l'assuré se trouve en Incapacité Temporaire de travail ou en Invalidité Permanente. Au Crédit Agricole la cotisation est maintenue.

Durant la période de couverture qui démarre dès l'octroi du prêt et se termine à l'arrivée à échéance des remboursements, l'assuré peut rester en bonne santé jusqu'au terme du prêt, ou peut être sinistré suivant les quatre catégories de risques couverts par l'assureur :

- La garantie Décès qui assure en cas de décès toute cause, sauf exclusions légales et contractuelles (Suicide, fraude, fait de guerre, etc...), le remboursement du CRD à la date du sinistre augmenté des intérêts éventuellement dus. Il existe généralement une période contractuelle qui débute à la date d'effet du contrat pendant laquelle la garantie ne s'applique pas. Cette période, appelée délai de carence, permet d'éviter l'antisélection.
- La Perte Totale et Irréversible d'Autonomie (PTIA), qui représente un état grave et irréversible d'invalidité, nécessitant l'assistance d'une tierce personne pour réaliser les actes ordinaires de la vie. Dans ce cas le CRD est pris en charge par l'assureur.
- En cas d'arrêt de travail, soit donc en cas d'Incapacité Temporaire de travail Partielle ou Totale (ITT) ou d'Invalidité Permanente Partielle ou Totale (ITD), l'assureur prend en charge les échéances du crédit. L'indemnisation s'exerce au-delà d'une période, appelée délai de franchise. Contrairement au délai de carence qui s'applique systématiquement en début de contrat et est spécifique au risque Décès, la franchise débute à la date de survenance du sinistre et ne s'applique qu'aux risques arrêt de travail (ITT / ITD). Les prestations cessent dès la reprise du travail, le passage à la retraite de l'assuré.

Le schéma suivant résume les différences entre le délai de carence pour le Décès et le délai de franchise pour l'Arrêt de travail, figure 2.1 :

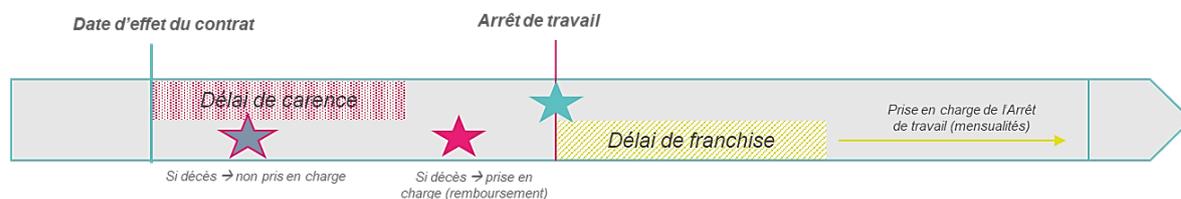


Fig. 2.1 – Vie du contrat ADE : mise en œuvre des prestations

2.2 Un marché ADE habitat à maturité

Le chiffre d'affaire de l'assurance emprunteur s'élève à 10 milliards d'euros pour l'année 2019 ([FA, 2021]), soit une hausse de près de 55% par rapport à 2007.

Cette évolution a principalement été portée par celle de l'habitat qui augmente de près de 64% et qui contribue à environ 70% des primes totales d'assurance emprunteur en 2019, figure 2.2.

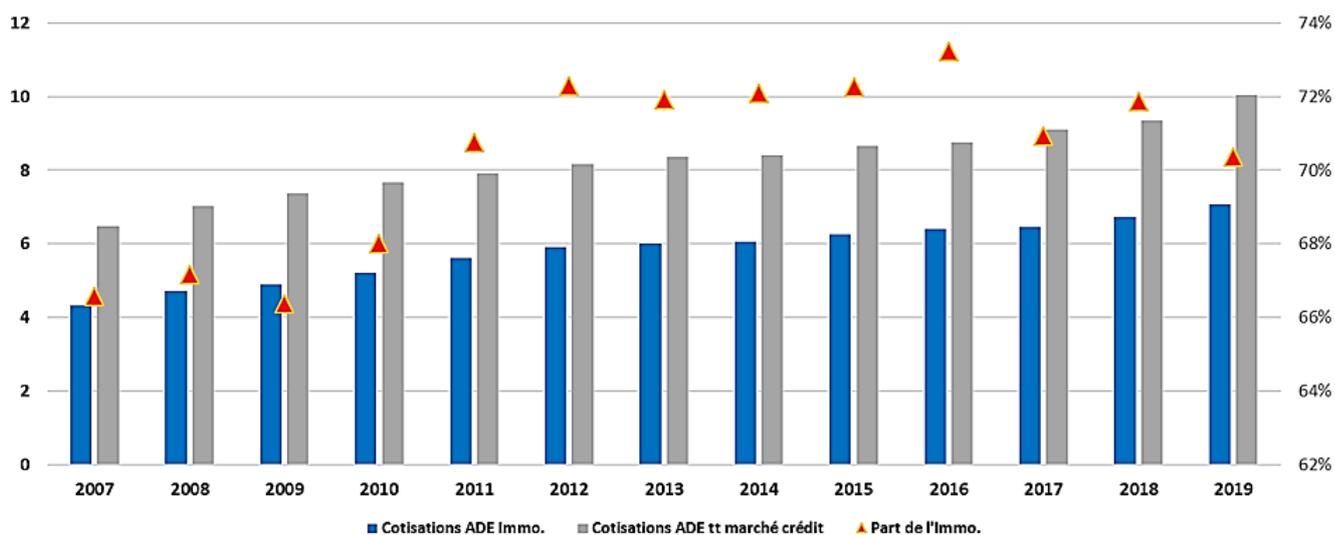


Fig. 2.2 – Evolution des montants de cotisations d'assurance emprunteur (Md€) en France de 2007 à 2019

2.3 Un contexte favorable pour le consommateur

2.3.1 Un marché des crédits immobiliers dynamique

L'augmentation du chiffre d'affaire ADE est en lien avec celle des encours crédit (Figure 2.3) sur la même période.



Fig. 2.3 – Encours de crédits à l'habitat aux particuliers (en Mds€ d'euros)

La croissance des encours de crédits immobiliers est notamment favorisée par un contexte très favorable de taux, figure 2.4, caractérisé par une baisse forte de 2010 à 2016, puis plus légère jusqu'au début 2022 (après une augmentation fin 2016), mais également par une augmentation modérée des prix de l'immobilier (hors Ile de France) et enfin par des incitations fiscales d'investissement sur le marché locatif, avec successivement les lois Scellier (2009-12), Duflot (2013-14) et le dispositif Pinel à partir de 2015.

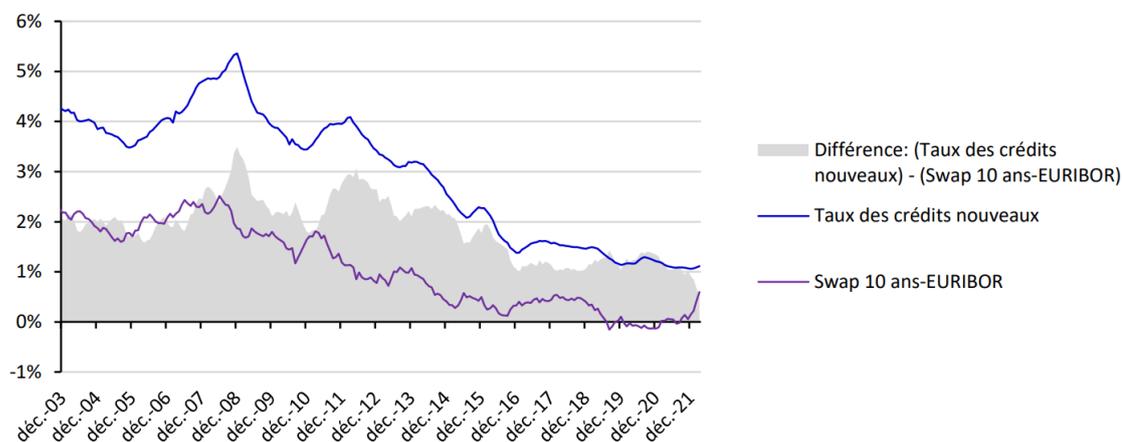


Fig. 2.4 – Taux des crédits nouveaux à l'habitat

L'analyse de la composition des flux mensuels, figure 2.5), permet de constater le poids important entre autres des rachats de crédit sur la période 2015-2018, qui ont eu un impact significatif sur les résiliations ADE, mais également sur les nouvelles entrées en portefeuille.

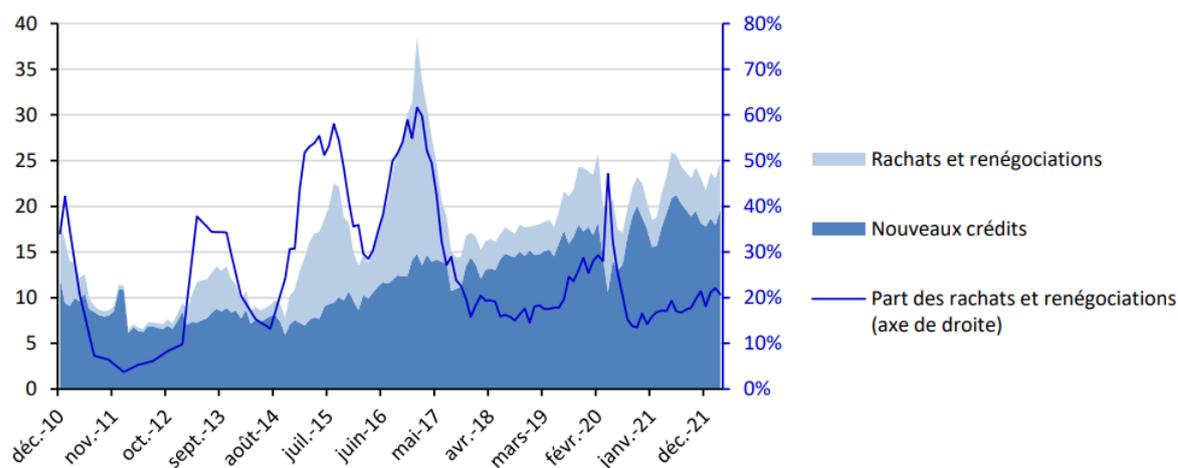


Fig. 2.5 – Production mensuelle de crédits à l'habitat et rachats et renégociations (en Mds d'euros)

2.3.2 Un cadre réglementaire qui favorise l'ouverture du marché de l'ADE

Depuis 2011, le marché de l'assurance emprunteur a connu de nombreuses évolutions techniques et réglementaires, visant à ouvrir le marché à la concurrence pour permettre aux assurés de bénéficier de tarifs plus attractifs :

- La loi Lagarde (2009) et la loi bancaire (2013) ont pour objectif :
 - Une meilleure information des emprunteurs : Depuis le 1^{er} juillet 2009, la Loi Lagarde introduit une obligation faite aux banques de distribuer une notice informative sur le thème de l'assurance emprunteur. Cette fiche d'information apporte aux emprunteurs une meilleure compréhension des assurances de prêts, elle doit être remplie conjointement par l'emprunteur et l'organisme prêteur.
 - Le libre choix de l'assurance de prêt immobilier : La Loi Lagarde, complétée par les mesures de la loi de séparation et de régulation des activités bancaires - SRAB (Loi n°2013-672 du 26 juillet 2013) ont supprimé une disposition législative qui autorisait les organismes financiers à imposer leur propre assurance emprunteur aux consommateurs. Les emprunteurs peuvent donc présenter une assurance tierce lors de la souscription de leur prêt immobilier dans les conditions suivantes : Le contrat présenté par l'emprunteur doit avoir un niveau de garantie au moins égal au contrat d'assurance dispensé par l'établissement prêteur. Si tel est le cas, le prêteur ne peut refuser sans justification le contrat d'assurance proposé. Il ne peut pas non plus modifier les conditions financières du prêt en contrepartie de son acceptation d'un contrat autre que celui qu'il distribue.
 - Le coût de l'assurance emprunteur doit être comparable au taux du crédit. Il sera donc exprimé en taux annuel effectif de l'assurance TAEA L'assureur est tenu d'informer l'établissement prêteur de toute modification du contrat ou de non paiement des échéances par l'emprunteur.
- La loi sur la consommation, dite loi Hamon (2014) vient compléter la loi Lagarde et permet à l'emprunteur ayant souscrit son assurance de prêt auprès de sa banque, de la résilier pendant 1 an à compter de la signature de l'offre de prêt. La loi Hamon concerne tous les nouveaux contrats d'assurance emprunteur relatifs à des offres de prêt signées auprès des banques à partir du 26 juillet 2014, en couverture des prêts immobiliers à usage d'habitation ou mixte (habitation et professionnel).

- Le CCSF (Comité Consultatif du Secteur Financier) (2015) arrête une liste de critères permettant d'apprécier l'équivalence des garanties et structure les demandes de substitutions. Cette liste est utilisée par les banques depuis le 1^{er} mai 2015. Depuis le 1^{er} octobre 2015, les exigences de la banque doivent figurer dans la Fiche Standardisée d'Information (FSI) remise à chaque futur emprunteur, le plus en amont possible.
- En 2016, la loi Sapin 2 - ou loi Bourquin, (loi n° 2016-1691, du 9 décembre 2016) vient renforcer le dispositif précédent dont les effets sont jugés insuffisants ; elle donne la possibilité à l'emprunteur de résilier son contrat tous les ans et de procéder à sa substitution, en respectant un préavis de 2 mois précédant la date anniversaire de la signature du contrat de prêt. Elle est entrée en vigueur en 2 temps, depuis le 1^{er} mars 2017 pour tous les nouveaux contrats signés à partir de cette date et depuis le 1^{er} janvier 2018 pour tous les contrats signés antérieurement.
- Enfin la loi Lemoine (Loi n°2022-270 du 28 février 2022), en œuvre dès 2022, donne un droit de résiliation sans frais à tout moment de l'assurance emprunteur. Elle s'est appliquée également en 2 temps, à partir du 1^{er} juin 2022 pour les nouvelles offres de prêts et, à partir du 1^{er} septembre 2022, pour les contrats d'assurance en cours. Outre cette mesure phare, la loi Lemoine vise également à améliorer l'accès au marché de l'assurance, en :
 - Supprimant le questionnaire médical pour les emprunteurs dont le montant de financement est inférieur à 200 000€ pour une personne seule, ou 400 000€ pour un couple
 - Reduisant de 10 à 5 ans le droit à l'oubli pour les pathologies les plus lourdes.

Les différentes étapes législatives destinées à faciliter la substitution des contrats d'assurance emprunteur avant 2022 sont résumées en annexe figure 7.1. En 2022, avec l'entrée en vigueur de la loi Lemoine, l'assuré peut finalement substituer son contrat ADE à tout moment (figure 7.2 en annexe).

2.4 L'exposition des bancassureurs au risque de résiliation ADE

Il existe deux types de situations qui mettent fin au contrat d'assurance.

2.4.1 Les résiliations d'ordre structurel

Ces résiliations sont soit liées au terme normal du crédit (sous-jacent de l'assurance) et au remboursement de sa dernière échéance soit liées aux aux sorties d'âge des garanties d'assurance prévues au contrat (précisés dans les Condition générales et notices) ou enfin, à la mise en œuvre des garanties décès, PTIA ou ITD avec le remboursement au client du capital restant dû de son financement couvert.

Les consolidations des ITT entraînent également des sorties de portefeuille des assurés. Toutefois ces situations étant difficilement identifiables et peu nombreuses, elles ne seront pas prises en compte dans cette étude.

2.4.2 Les résiliations d'ordre conjoncturel

Ces résiliations sont liées soit au rachat du crédit immobilier, et conséquemment à la résiliation du contrat d'assurance, ou à la possibilité de résiliation ou de substitution du contrat d'assurance, à la faveur d'un contexte de marché favorable ADE ou crédit.

2.5 L'exposition des bancassureurs au risque de résiliation ADE

2.5.1 Le Crédit Agricole un acteur majeur de l'ADE en France

L'ADE des Caisses régionales de Crédit Agricole couvre environ 8 millions d'emprunteurs et indemnise annuellement 10 000 décès, 40 000 arrêts de travail et 1 500 pertes d'emploi. Le Crédit Agricole par son chiffre d'affaires est le deuxième assureur emprunteurs en France, figure 2.6 :

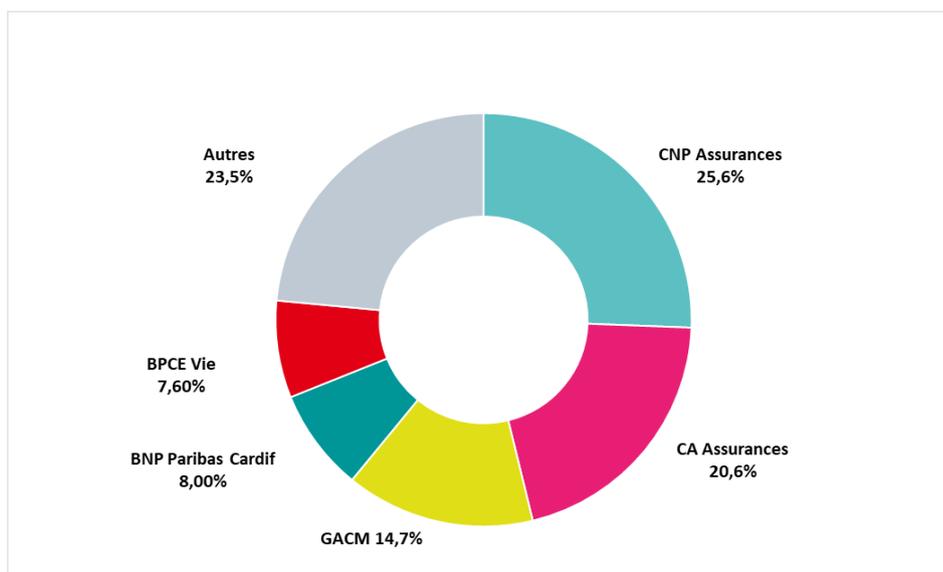


Fig. 2.6 – Parts de marché des principaux acteurs en 2020 (% des cotisations cumulées des 24 principaux acteurs)

L'activité est en croissance soutenue en 2020 par rapport au marché, figure 2.7 :

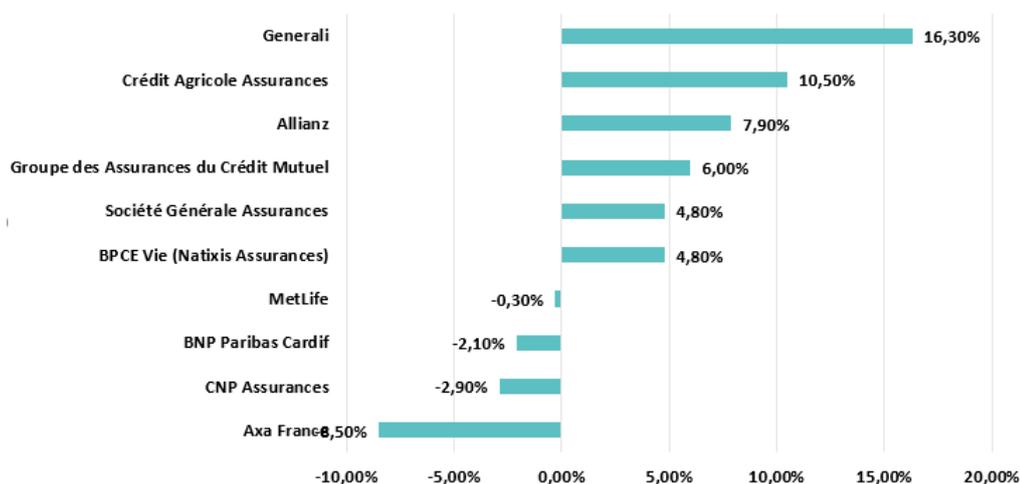


Fig. 2.7 – Variation 2020/2019 (%) du chiffre d'affaires 2019/2020

Le marché dominant est le marché des prêts immobiliers qui représente près de 70% du chiffre d'affaires ADE, figure 2.8 :

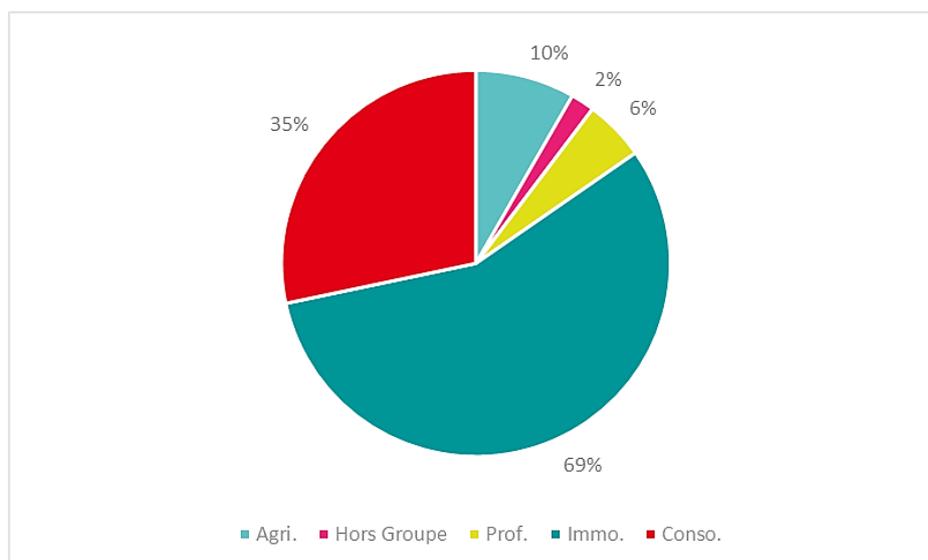


Fig. 2.8 – Poids de des contrats habitat en nombre dans l'ADE des Caisses régionales

2.5.2 Les crédits immobiliers couverts au Crédit Agricole

Les principales familles structurantes de crédit habitats distribués sont les suivants :

- Les crédits à taux fixe, parmi lesquels sont distingués :
 - Les crédits amortissables.

Il s'agit du crédit immobilier classique utilisé pour financer l'achat d'un logement. Une mensualité, contenant à la fois une partie du capital emprunté et les intérêts définis en avance, doit être versée par l'emprunteur chaque mois pendant toute la durée du crédit. Par exemple pour un emprunt sur 15 ans à mensualités fixes composées d'une charge d'intérêt et du remboursement du capital emprunté, figure 2.9.

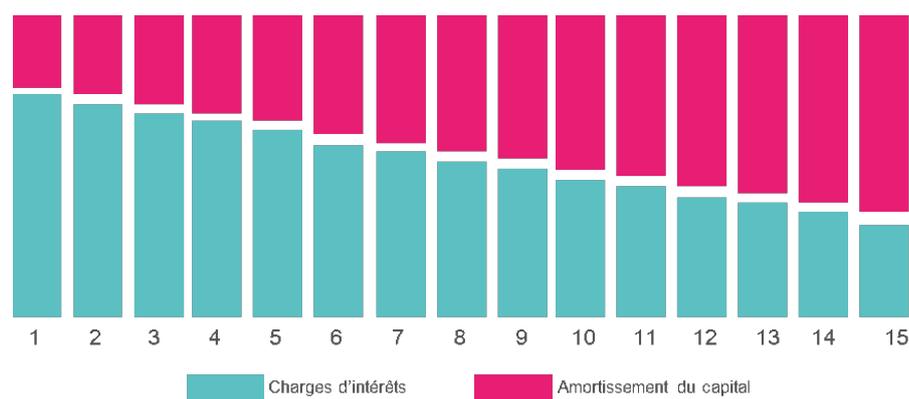


Fig. 2.9 – Echéancier prêt à taux fixe

Cette famille comprend entre autres les prêts à taux 0 (PTZ) et les prêts amortissables « standard » avec ou sans différé partiel, avec ou sans période d'anticipation.

- Les prêts in fine en capital ou en capital et intérêts.

La définition est similaire à celle d'un crédit amortissable, mais le capital emprunté n'est remboursé qu'à l'échéance de l'emprunt, ainsi les mensualités ne contiennent que les intérêts, figure 2.10 :

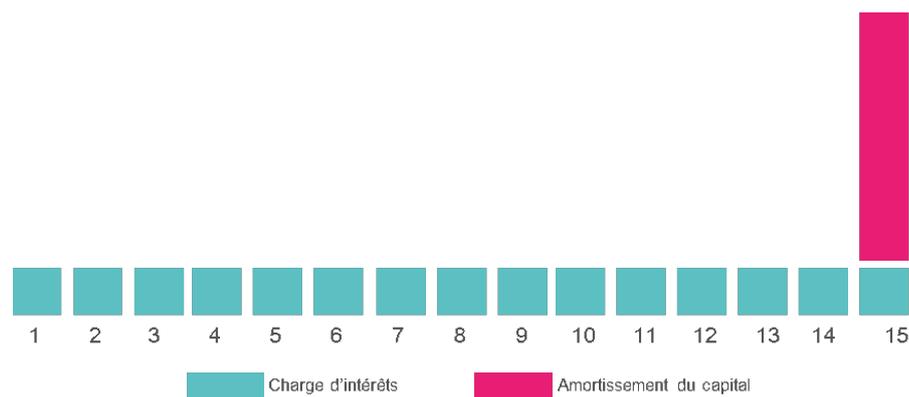


Fig. 2.10 – Echéancier prêt in fine

- Les crédits à taux variables.

Ce sont des prêts dont la mensualité peut évoluer à la hausse comme à la baisse en fonction du taux initialement prévu. Le taux de ces crédits est fixé initialement sur un indice de marché monétaire de référence, l'indice Euribor, qui régit les taux immobiliers et leur évolution. Le taux appliqué est révisé annuellement en fonction de l'évolution à la hausse ou à la baisse du taux de référence. Il peut être capé, c'est-à-dire limité à la hausse.

Par la suite, l'étude sera restreinte aux financements à taux fixe, amortissables qui constituent l'essentiel des financements couverts en ADE du portefeuille.

2.5.3 Les produits ADE des Caisses régionales et la distribution

2.5.3.1 Les produits ADE des Caisses régionales et la distribution

L'offre ADE proposée (Figure 2.11) a pour objectif de satisfaire l'ensemble des besoins des clients emprunteurs. Les crédits habitats sont couverts par le même socle contractuel Groupe d'assurance groupe indépendamment des projets habitat qu'ils financent. L'offre est complétée par une offre spécifique (« hors norme ») pour couvrir :

- Les gros capitaux au-delà-du plafond du contrat Groupe.
- Les emprunteurs ayant dépassé l'âge limite à l'entrée dans le contrat groupe.

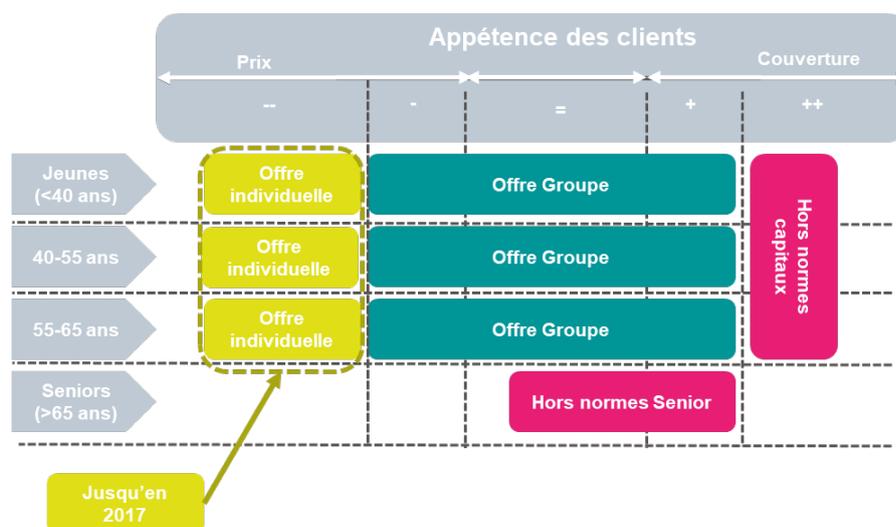


Fig. 2.11 – Caractéristiques des garanties de l'offre habitat

Les offres groupe et hors norme étaient complétées jusqu'en 2017 par une offre individuelle pour les clients privilégiant le tarif à la couverture.

L'offre ADE habitat groupe qui nous intéresse dans l'étude est également caractérisée par une tarification à taux fixe par tranche d'âge sur capital initial (CI). Elle couvre en outre la profession du client. Les caractéristiques détaillées du contrat habitat sont les suivantes (Figure 2.12). L'engagement de l'assureur n'est pas le même selon le risque couvert : soit il verse le capital restant dû, soit il se substitue à l'assuré et rembourse temporairement les échéances du prêt ou une partie des échéances.

	Décès	PTIA	ITT/INV	INV AERAS	P. E.
Type de garantie	Capital	capital	Rente temporaire	Garantie proposée par l'assureur si les garanties ITT/IPT sont refusées pour raisons médicales ou accordées accordées mais avec réserves - conformément à la convention Aéras révisée	Rente temporaire
Engagement de l'assureur	Versement d'un capital	Versement d'un capital	Rente temporaire		Rente temporaire
Age limite adhésion	70 ans	70 ans	60 ans		50 ans
Age limite prestation	70 ans	70 ans	65 ans		55 ans
Franchise			90 jours		90 jours
Plafond	1,2M€ puis 1,5M€		8000€ par mois et assuré		50% ou 100% de l'échéance - max. 3500€ par mois par assuré
Décès accidentel temporaire	Décès accidentel limité à 200k€ - 90j				
Définition			sa profession		
Couverture des inactifs ITT			oui		
Exclusion MNO			non		non

Fig. 2.12 – Cartographie de l'offre habitat du Crédit Agricole

1. Définition de la Perte Totale et Irréversible d'Autonomie (PTIA).

L'assuré est en état de Perte Totale et Irréversible d'Autonomie lorsqu'en cours d'assurance les trois conditions suivantes sont cumulativement réunies :

- l'invalidité dont est atteint l'assuré le place dans l'impossibilité totale et définitive de se livrer à toute occupation ou à toute activité rémunérée ou pouvant lui procurer gain ou profit. elle met définitivement et de façon permanente l'assuré dans l'obligation de recourir à l'assistance d'une tierce personne pour accomplir l'ensemble des actes ordinaires de la vie (se laver, s'habiller, se nourrir, se déplacer).
- la date de la PTIA telle que reconnue par l'assureur doit intervenir avant la fin du mois de son 70ème anniversaire.

2. Définition de l'Incapacité Temporaire Totale.

L'assuré est en état d'ITT lorsque, en cours d'assurance, les trois conditions suivantes sont cumulativement réunies :

- Si l'assuré exerce une activité professionnelle à la veille du sinistre :
 - L'assuré se trouve, à la suite d'un Accident ou d'une maladie, dans l'incapacité totale, reconnue médicalement, d'exercer votre activité professionnelle (définie comme l'activité professionnelle que Vous exerciez au jour du sinistre) même à temps partiel.
 - Cette incapacité est continue et persiste au-delà d'une période de franchise de 90 jours, période pendant laquelle aucune prestation n'est due par l'assureur.
 - Cette incapacité doit être justifiée par la production des pièces prévues contractuellement « En cas d'Incapacité Temporaire Totale »).
- Si l'assuré n'exerce pas d'activité professionnelle ou s'il est demandeur d'emploi à la veille du sinistre :
 - L'assuré se trouve, à la suite d'un Accident ou d'une maladie dans l'incapacité totale, reconnue médicalement, d'exercer vos activités habituelles non professionnelles même à temps partiel.
 - Cette incapacité est continue et persiste au-delà d'une période de franchise de 90 jours, période pendant laquelle aucune prestation n'est due par l'assureur.

- Cette incapacité doit être justifiée par la production production des pièces prévues contractuellement en cas d’Incapacité Temporaire Totale.
3. Définition de l’Invalidité (INV) Permanente Totale (IPT) - ou totale et définitive (ITD)
L’assuré est en état d’Invalidité Permanente Totale lorsque, en cours d’assurance, les deux conditions suivantes sont cumulativement réunies :
- A l’issue d’un état d’Incapacité Temporaire Totale défini ci-dessus, l’assuré se trouve dans l’impossibilité reconnue médicalement, d’exercer, même à temps partiel, une quelconque activité professionnelle ou une activité habituelle non professionnelle.
 - Cette invalidité doit être justifiée par la production des pièces prévues contractuellement en cas d’Invalidité Permanente Totale.

La garantie IPT ne peut se cumuler avec la garantie ITT. La perception de la garantie IPT fait cesser la prise en charge au titre de la garantie ITT.

2.5.4 La modélisation actuelle des rachats

Les lois de rachats utilisé dans l’outil de projection de portefeuille prophète, sont basées sur des études retrospectives actualisées dès lors que les indicateurs de suivi du risque mettent en avant une défaillance de ces dernières.

La modélisation sous-tend que les rachats des contrats emprunteur adossés à un prêt immobilier peuvent avoir deux origines différentes :

- Une composante, liée au rachat du prêt sous-jacent, et fortement en lien avec l’environnement économique.
- Une composante spécifique à l’assurance, en lien avec les évolutions réglementaires autorisant le changement d’assurance emprunteur en cours de vie du prêt bancaire.

2.5.4.1 La composante conjoncturelle

Pour la composante crédit, il est émis l’hypothèse que les taux de rachats sont fonction de l’ancienneté du contrat, avec un taux moyen sur la durée de projection des engagements (30 ans) ajusté sur des hypothèses de volumes de crédits futurs (estimés au niveau «Groupe », à partir de remontées détaillées des Caisses régionales). La forme de la courbe des rachats historiques est établie sur la base des données 2015-2017. Son niveau est ensuite ajusté en fonction des taux de rachats moyens attendus par le Crédit Agricole sur la période 2019-2021. L’effet Hamon historique constaté sur la période d’observation est neutralisé ($H\%$) pour calculer le coefficient correcteur. Un même coefficient correcteur ϕ est ainsi appliqué à l’ensemble de la courbe d’historique (excepté l’ancienneté 0).

2.5.4.2 La composante assurantielle

Le taux de rachat final de 1^{re} année est exprimé comme la somme d’un taux de rachat total ajusté historique (neutralisé de l’effet Hamon $H\%$) et de l’effet Hamon actualisé ($H^l\%$) :

$$Tx_{final,anc_1} = \frac{Tx_{final,anc_1} - H\%}{\phi} + H\%$$

Pour les années suivantes, le taux « Bourquin » estimé à partir des premières données ($B\%$) est augmenté d’un facteur de sécurité de 25%.

$$Tx_{final,anc_i} = \frac{Tx_{étape_1,anc_i} - H\%}{\phi} + B\%(1 + 25\%)$$

Pour mieux comprendre la loi finale, la nouvelle loi Immo est décomposée étape par étape (Figure 2.13). Les taux bruts sur la période d'étude sont d'abord isolés, puis les taux retraités de l'effet conjoncturel uniquement sont ensuite pris en compte pour obtenir enfin la loi finale retraitée de l'effet conjoncturel ainsi que des effets Bourquin et Hamon.

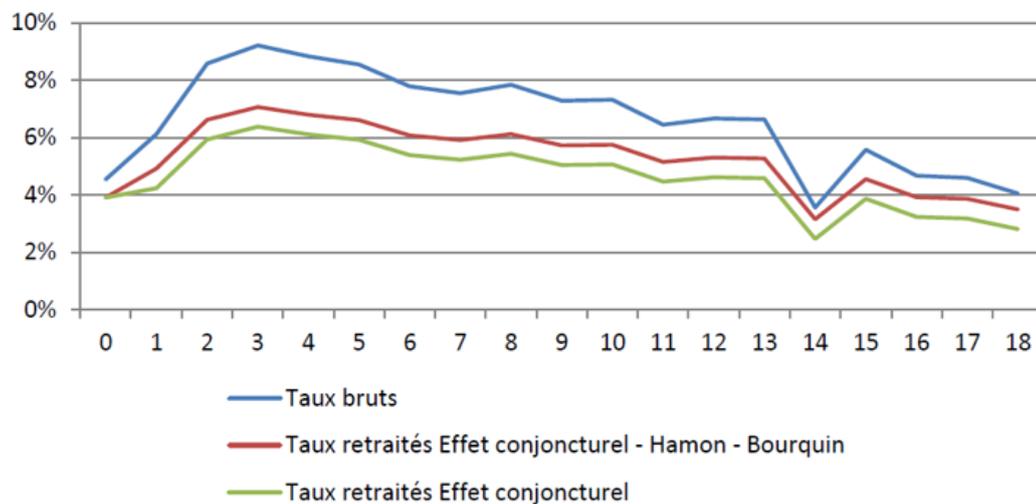


Fig. 2.13 – Lois de résiliations immo par ancienneté

2.5.5 Les limites et recherche d'une méthode alternative

La méthode actuelle se révèle insuffisante pour capter la volatilité des rachats au cours du temps, liée à celles des deux composantes identifiées ci-dessus. Cette volatilité est illustrée sur le graphique ci-dessous figure 2.14 avec les taux de rachats moyens évalués sur les périodes 2015-2016, 2016-2017 et 2018-2019.

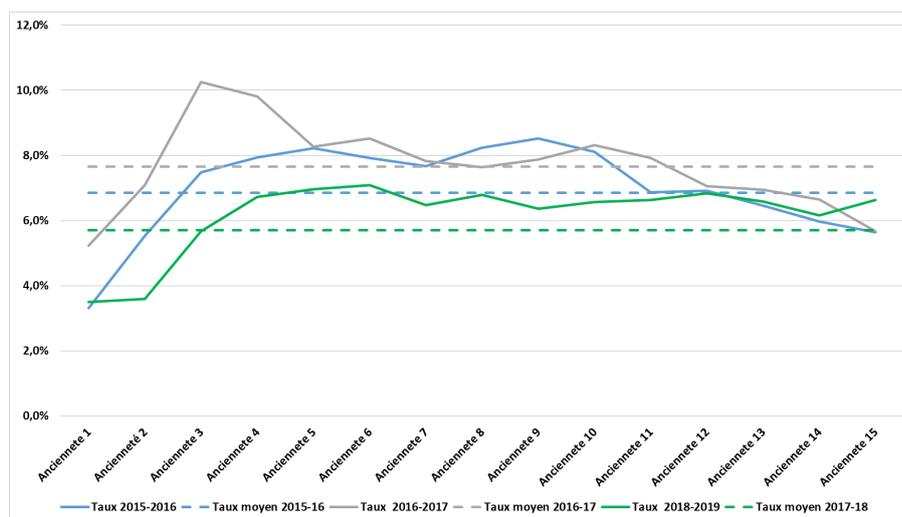


Fig. 2.14 – Evolution des taux de rachats moyens constatés – périodes [2015,2016], [2016,2017] et [2018,2019]

L'approche actuelle reproduit les comportements observés et induit des écarts d'expérience, d'autant plus importants que la volatilité est concentrée sur les contrats de moins de 10 ans qui pèsent le plus dans le portefeuille - 80% du portefeuille, figure 2.15 :

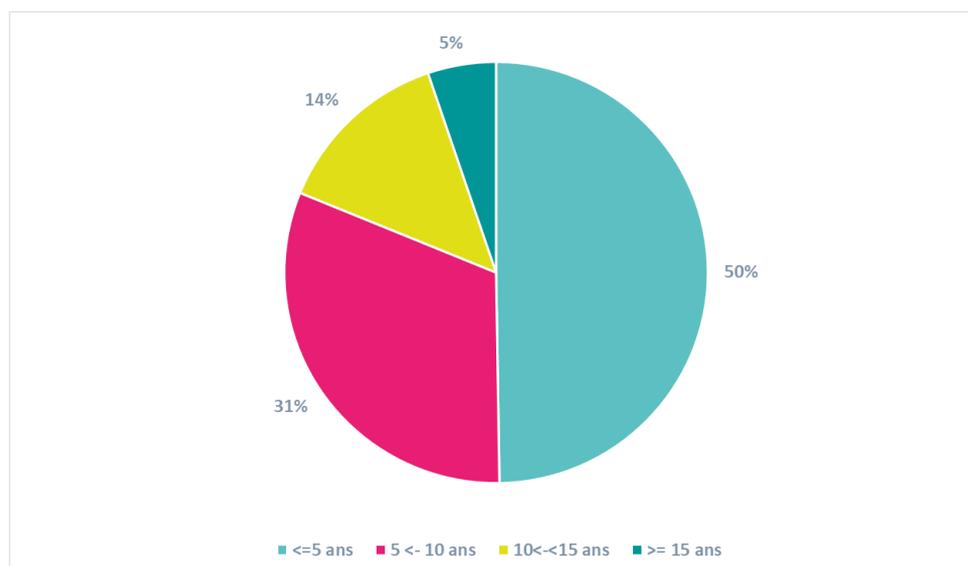


Fig. 2.15 – Structure en nombre du marché ADE habitat par tranches d'ancienneté

L'approche actuelle est caractérisée par des lois moyennes de rachats par ancienneté, représentées ci-dessous, figure 2.16 :

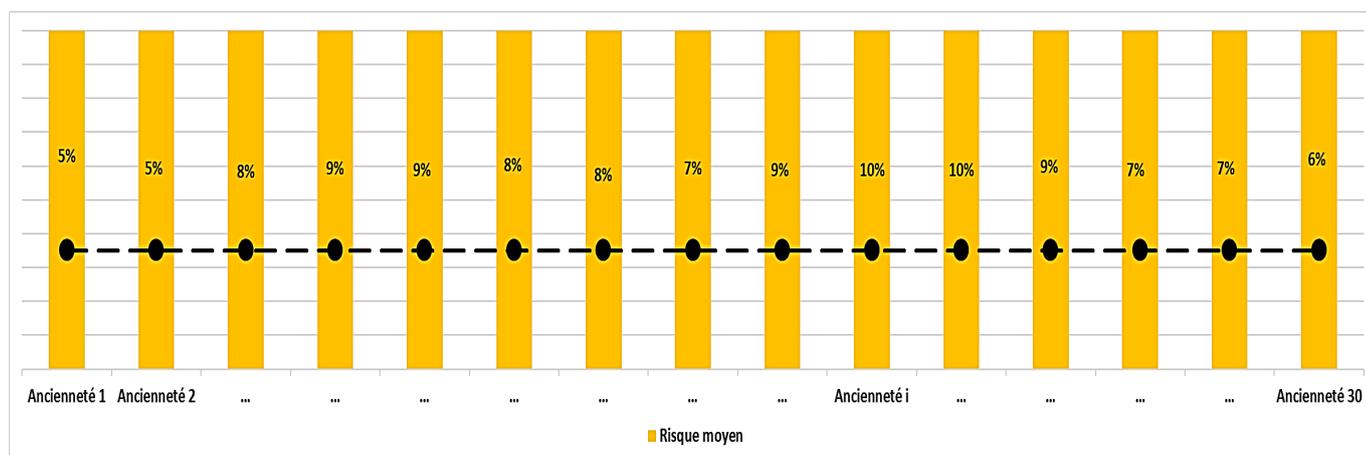


Fig. 2.16 – Lois historiques moyennes par ancienneté

Chaque contrat projeté se voit affecter la même loi de rachat, dépendant uniquement de l'ancienneté du contrat. Une approche alternative de modélisation dynamique, prenant en compte les effets combinés de l'évolution du contexte économique et du comportement des assurés va être l'objet de l'étude qui suit, avec comme objectif d'améliorer l'ajustement des lois du modèle. Cette approche ne nécessite pas d'ajustement calculé sur des données externes à celles utilisées pour les projections.

Le principe (Figure 2.17) est, pour chaque ancienneté, d'affecter les contrats dynamiquement à une classe de risque, à l'aide de scores d'attrition dépendant uniquement de l'ancienneté du contrat. Le contrat se voit affecter la loi de rachat dépendant de son ancienneté et de son score, ci-dessous, pour une segmentation à 3 classes (risque faible, moyen ou élevé).

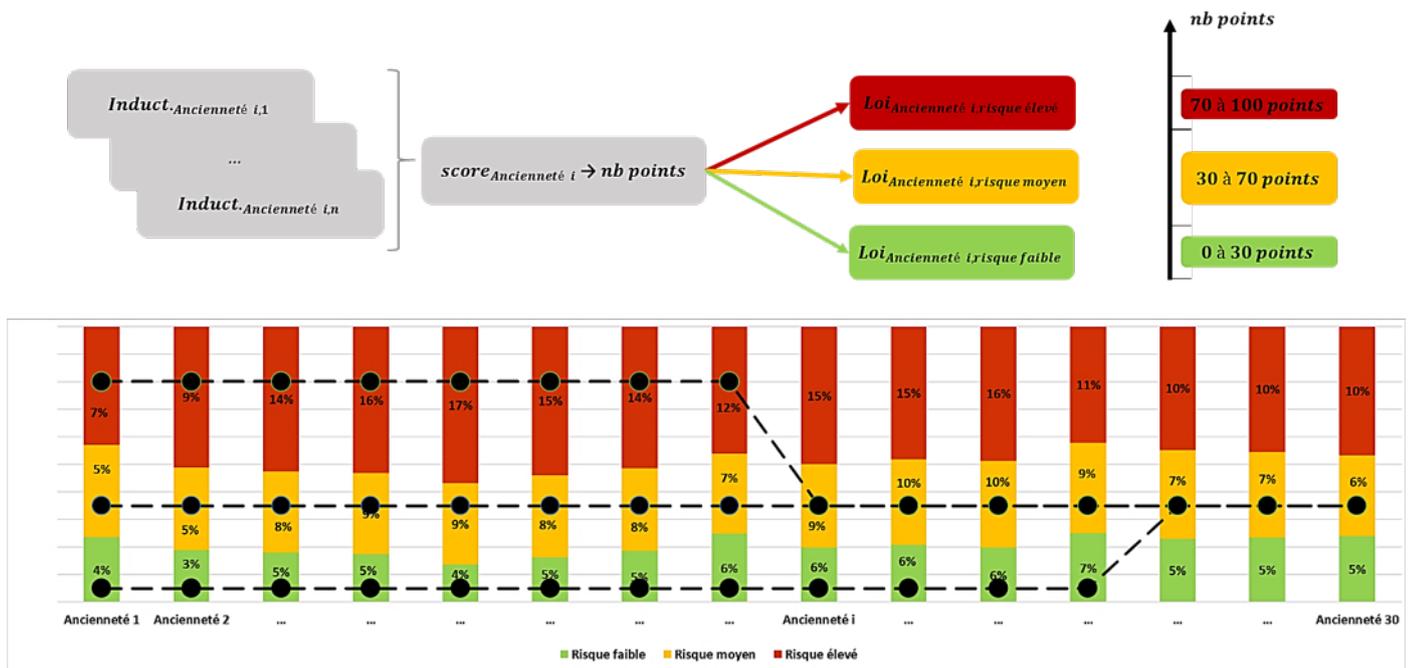


Fig. 2.17 – Approche segmentée dynamique

2.6 Conclusion

Depuis une dizaine d'année le contexte économique et législatif entourant l'assurance des emprunteurs a beaucoup évolué. Les taux de crédit immobilier après avoir significativement baissé jusqu'en 2022 ont commencé à augmenter depuis. Cette baisse a facilité l'accès au marché du crédit, et a été renforcé par la volonté du législateur d'ouvrir le marché de l'ADE qui s'est traduite par une baisse tarifaire de l'assurance. En outre l'assuré a aujourd'hui la faculté de résilier à tout moment.

Cette situation conduit à une forte variabilité des rachats à modéliser et a un impact sur les engagements futurs.

La modélisation statique des rachats par ancienneté actuellement mise en œuvre n'étant pas pleinement satisfaisante, une approche dynamique alternative doit être recherchée afin de mieux s'ajuster au réel. La mise en place de cette approche requiert la construction de bases annuelles de modélisation, qui va être présentée dans la partie suivante.

Chapitre 3

Les bases d'étude

La construction des bases d'étude nécessite de travailler sur des fichiers issus de deux systèmes d'informations :

- Les données ADE tête par tête et de gestion des sinistres provenant des systèmes d'information des Caisses régionales.
- Les données comptables de primes et de sinistres de l'assureur utilisées pour les arrêtés techniques afin de valider la base d'étude construite.

Avant de construire les bases, il est nécessaire de choisir et justifier la période d'étude retenue.

3.1 Le choix de la période d'étude

Le choix de la période d'étude est piloté par les deux hypothèses structurantes suivantes :

1. Stabilité dans le temps des offres et des processus de distribution. L'internalisation de l'ADE des Caisses régionales au sein de Crédit Agricole Assurances en 2016, a eu pour conséquence la distribution à partir de 2019 d'une nouvelle offre (dite « H2 ») accompagnée d'évolutions des outils informatiques et des modes distributifs.
2. Une qualification optimale des rachats et de ses inducteurs comportementaux. Les sorties pour cause de rachat, ne sont pas en lecture directe ; elles se déduisent du reste des sorties après avoir écarté les origines suivantes :
 - La fin de financement dont l'échéance est disponible dans le fichier source.
 - Le dépassement d'âge limite contractuel des garanties qui est également connu
 - L'occurrence d'un sinistre décès ou PTIA dont la date de survenance n'est pas immédiatement connue. En effet, les sinistres survenus lors d'un exercice (N) ne sont exhaustivement identifiés qu'à la fin du deuxième exercice suivant la survenance : 95% sont identifiés en année N , 98% en année $N + 1$ et 100% en année $N + 2$.

La période d'étude (2015-2019) qui répond aux contraintes ci-dessus sera donc retenue pour à la fois construire et back-tester la modélisation dynamique des rachats. Ce choix permettra entre autres de vérifier la capacité d'ajustement conjoncturel du modèle dynamique, construit dans un contexte réglementaire « Hamon » et back-testé dans un contexte « Bourquin ». Une analyse préliminaire des crédits distribués dans le réseau des Caisses régionales est conduite pour préciser le périmètre de contrats pris en compte dans l'étude.

3.2 Analyse structurelle préliminaire des crédits distribués par des Caisses régionales

L'étude des fichiers têtes par têtes ADE donne la structure des crédits immobiliers couverts qui est stable dans le temps, ci-dessous, figure 3.1 :

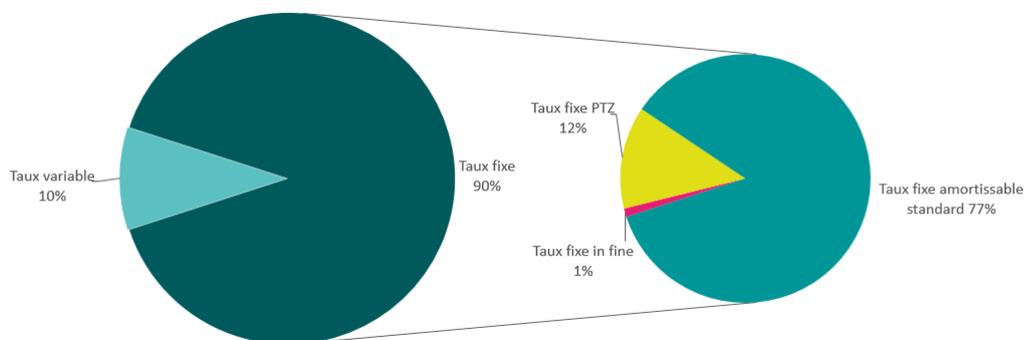


Fig. 3.1 – Distribution des familles de crédits en nombre %

Le périmètre de l'étude qui suit sera restreint aux crédits à taux fixe standard hors prêts in fine et PTZ, qui représentent près de 80% du portefeuille des crédits immobiliers couverts en ADE.

Une fois précisés la période et le périmètre de l'étude, la construction des bases de données, à partir des fichiers sources est lancée. Cette construction se fait concrètement en plusieurs étapes :

Dans une 1^{re} étape, il s'agira de reconstituer les flux à partir de photos de stocks trimestriels, de les enrichir avec des données calculées à partir des fichiers bruts initiaux ou de données exogènes, de retraiter les valeurs incohérentes ou manquantes.

Dans une 2^e étape, les observations seront agrégées et échantillonnées compte tenu des volumétries en jeu.

Enfin, à l'issue de la construction la cohérence du contenu des bases sera contrôlée à l'aide de sources adverses comptables.

3.3 Principes généraux de construction des bases d'études

3.3.1 Reconstitutions des flux et enrichissements

Les bases d'étude annuelles sont construites de manière différentielle à partir.

- Des fichiers des assurés qui contiennent les caractéristiques des contrats externes ou internes en stock, produits tous les trimestres (04/*N*, 07/*N*, 10/*N* et 01/*N* + 1) par les traitements ADE de l'ensemble des 39 Caisses régionales du Groupe Crédit Agricole. Elle est complétée par des données provenant des bases INSEE ([Bases INSEE]).
- De l'historique des sinistres ADE (à fin 2021) de l'ensemble des Caisses régionales du Groupe Crédit Agricole. Seuls sont retenus les sinistres décès, PTIA et ITD qui vont permettre d'extraire de la population étudiée les sorties de portefeuille résultant de ces sinistres durant les périodes d'étude.
- De la prise en compte des sorties d'âge ADE contractuelles et des crédits arrivant à leur terme au cours de la période étudiée.

Les contrats présents en début de période sont extraits et leur évolution est suivie jusqu'au terme de la période d'étude (1 an), pour obtenir une base contenant les contrats toujours en portefeuille ou sortis consécutivement à une substitution du contrat ADE, ou à un rachat du crédit, sous-jacent au contrat d'assurance. Le traitement se décompose en 3 grandes étapes, figure 3.2 :

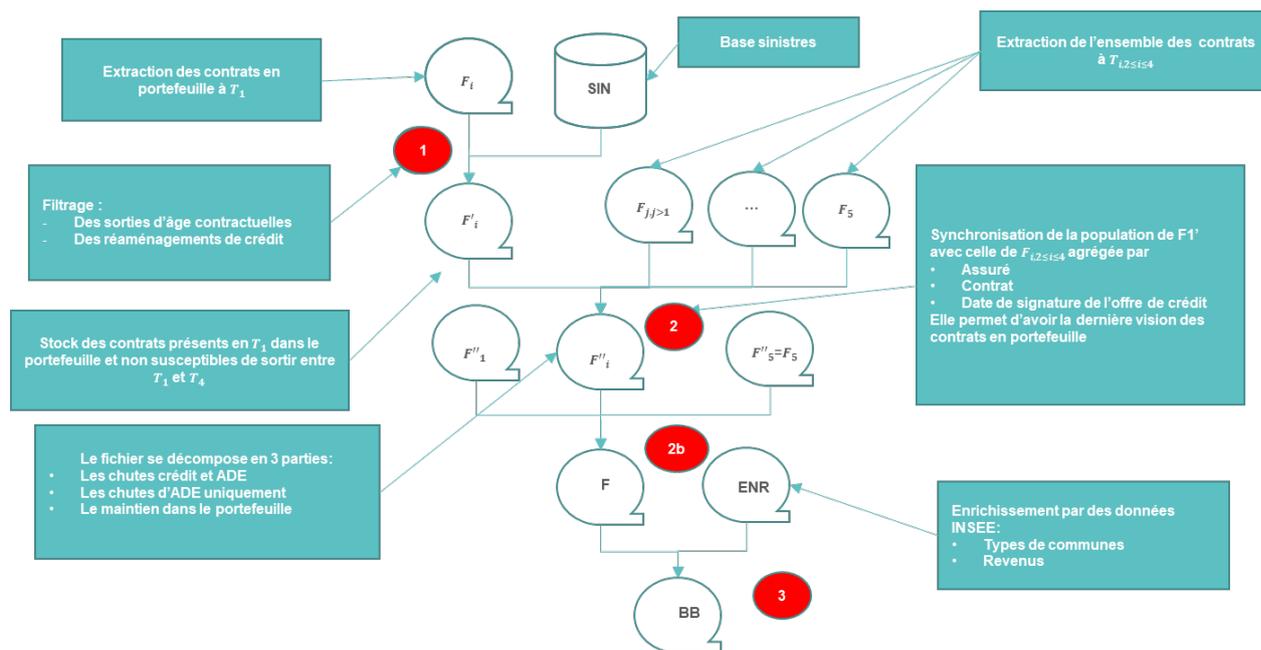


Fig. 3.2 – Dessin de chaîne simplifiée du traitement de constitution des bases d'études

Trois bases brutes d'observation annuelle $07/N \rightarrow 07/N + 1$ sont ainsi construites :

- Deux bases :
 - Une base [BB1] qui sert à la construction du modèle et à la segmentation des classes de risques.
 - Une seconde [BB2] est utilisée avec la 1^{re} pour le calcul des taux de rachats sur chaque classe de risques.
- Enfin, une base qui est utilisée pour évaluer la robustesse du modèle construit précédemment [BB3].

Les bases brutes ci-dessus sont restreintes au périmètre d'étude.

3.3.2 Traitement des valeurs manquantes ou aberrantes

La stratégie de gestion des anomalies est synthétisée en annexe 7.2.12 - figure 7.21. L'analyse de la qualité des données des fichiers sources préalable à la construction des bases donne les éléments statistiques d'anomalies suivants :

- Pour la base [B1] : un taux d'anomalie de 0,13% (Figure 3.3) :

	Nb lignes en anomalie	Nb lignes total
Hors_Scope_Etude	2 622	1 891 070
Scope_Etude (base [B1])	7 723	5 673 210
Total	10 345	7 564 280

Fig. 3.3 – Qualité de données - Base [B1]

- Pour la base [B2] : un taux d'anomalie de 0,15% (Figure 3.4) :

	Nb lignes en anomalie	Nb lignes total
Hors_Scope_Etude	2 303	1 459 381
Scope_Etude (base [B2])	8 968	5 683 182
Total	11 271	7 142 563

Fig. 3.4 – Qualité de données - Base [B2]

L'anomalie la plus fréquente dans les deux bases concerne la durée initiale du prêt qui représente plus de 90% des anomalies. Compte tenu des volumétries en jeu, les lignes en anomalie sont supprimées.

3.3.3 Agrégation et échantillonnage

Pour les besoins de l'étude, une agrégation des bases [B1], [B2] et [B3] est réalisée à la maille Caisse régionale \times Assuré \times contrat (identifié par son code et sa date de signature). Cette agrégation conduit respectivement aux bases [BA1], [BA2], et [BA3], avec les effectifs suivants (Figure 3.5) :

	BA1	BA2	BA3
Rachats	265 805	227 760	214 775
Maintien en portefeuille	3 141 019	3 069 210	3 525 512

Fig. 3.5 – Effectifs bases agrégées [BA_i]

Les volumétries des bases agrégées construites ci-dessus précédemment sont d'environ 3 millions de lignes. Pour faciliter la mise en œuvre des méthodes statistiques dans l'étude comportementale, il est procédé à un échantillonnage représentatif de la base brute d'étude agrégée [BA1]. La base obtenue est la base [BEA1]. Une cible d'environ 300 000 lignes conduit au choix à un échantillonnage stratifié par Caisse régionale et ancienneté de 10% des bases brutes, en considérant les agences des réseaux les plus « contributrices » et représentatives des politiques de distribution des Caisses régionales.

Pratiquement, par Caisse régionale \times ancienneté :

- Les agences sont triées par ordre décroissant d'encours crédit habitat (CRD).
- Seules les agences $CRD_1, CRD_2, \dots, CRD_{i_{CR, Anc}}$ vérifiant $\sum_{i < i_{CR, Anc}} CRD_i < 10\%$ et $\sum_{i=i_{CR, Anc}} CRD_i \geq 10\%$ sont sélectionnées.

Par exemple, pour la Caisse régionale 802 et l'ancienneté 5, seules les 5 agences les plus contributrices sont retenues comme l'illustre la figure 7.9 en annexe.

On ne retiendra ainsi qu'environ 3 à 4% des agences des Caisses régionales par ancienneté, Figure 7.10 en annexe.

L'échantillonnage se traduit par les effectifs suivants (Figure 3.6) :

	BEA1
Rachats	27 610
Maintien en portefeuille	318 643

Fig. 3.6 – Effectifs base échantillonnée agrégée

Certains contrats apparaissent exposés partiellement (environ 5% des contrats de la base) au risque de rachat sur la période d'étude choisie (figures 7.11 et 7.12 en annexe), du fait :

- De l'atteinte des limites d'âge du contrat d'assurance.
- Du terme du contrat de prêt.
- De l'occurrence d'un sinistre (DC, PTIA, ITD).

Avant l'agrégation destinée à l'étude comportementale, ces contrats sont filtrés. Ils seront naturellement pris en compte avec leur exposition pour les calculs des taux bruts.

Par ailleurs, les anciennetés supérieures à 15 ans, qui sont unitairement inférieures en nombre au seuil de 2%, sont regroupées sur l'ancienneté 15 ans, figure 7.13 en annexe.

Il sera appliqué aux anciennetés supérieures à 15 ans, les lois de l'ancienneté 15 ans résultant du regroupement.

3.3.4 Contrôle et calage à la comptabilité

La comparaison du chiffre d'affaire décès et ITT sur un an calculé à partir de la base brute constituée à mi-année avec celui des arrêtés de compte annuels ne fait pas ressortir d'incohérence, figure 7.14 et Figure 7.15 en annexe.

Par ailleurs, la figure 7.16 en annexe illustre la distribution du chiffre d'affaires annuel entre le flux net d'affaires nouvelles de l'année (8%) et le stock (92%). Cette distribution est également conforme aux données comptables.

3.3.5 Contrôle de la qualité de l'ajustement de l'échantillon

Les segmentations étant réalisées sur chaque ancienneté, la qualité de l'échantillonnage est évaluée pour chaque ancienneté de 1 à 15 ans. L'ajustement est globalement satisfaisant pour les besoins de l'étude, figure 3.7 :

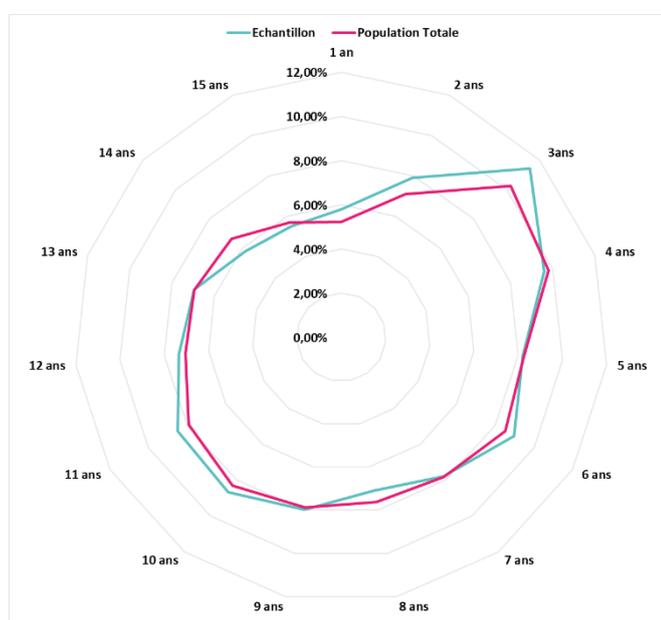


Fig. 3.7 – Ajustement [BE1] échantillon/pop. Totale par Ancienneté

L'ajustement peut être optimisé sur les anciennetés pour lesquelles les écarts sont jugés significatifs, en identifiant les Caisses régionales qui contribuent le plus aux écarts constatés sur ces anciennetés, et en les excluant de la sélection. Ainsi, sur les anciennetés ≤ 3 ans, les Caisses régionales 812, 836, 839, 868, 887 et 894 des anciennetés ≤ 3 peuvent être exclues, figure 7.17 en annexe.

L'amélioration recherchée peut être constatée graphiquement - (Figure 7.18 en annexe) (ou par un χ^2 d'ajustement)

Afin de conserver un effectif échantillonné de 10% de la population totale, il convient de procéder, pour les anciennetés traitées à une nouvelle extraction de la population résiduelle en adaptant le seuil comme suit :

$$Seuil_{cible} = \frac{10\% \text{ population totale}}{\text{population résiduelle}}$$

Cependant, cette optimisation permet de gagner en ajustement mais entraîne une perte en représentativité des Caisses dans la base.

Dans le cadre de cette étude, l'ensemble des Caisses régionales sera conservé et le retraitement ci-dessus ne sera pas mis en œuvre.

3.3.6 Description de la structure des données

Les données utilisées pour cette étude sont à la fois qualitatives et quantitatives (continues ou discrètes). Les données brutes ont été privilégiées pour préserver la qualité des modèles et également maîtriser le coût des actualisations futures :

1. variables numériques continues :
 - Le taux du prêt client (taux nominal du crédit).
 - Le spread par rapport à un taux de prêt de "marché" (voir au 7.2.11.2 en annexe).
 - Le spread par rapport à un taux ADE de "marché" (voir au 7.2.11.2 en annexe).
 - Le taux de cotisation ADE client (Le taux du tarif moyen effectif).
 - Le taux de décote commerciale ADE.
 - L'âge de l'assuré en années.
 - La durée du crédit courant en mois.
 - Le montant de capital initial emprunté.
 - Le montant de revenu médian («Revenu_Median»).
 - Le taux (quotité) de couverture DC.
 - Le taux (quotité) de couverture ITT.
2. variables qualitatives :
 - La catégorie socio-professionnelle de l'assuré.
 - La civilité de l'assuré.
 - Le code secteur de distribution - cf. identifiant du réseau commercial de la Caisse régionale.
 - Les types de communes INSEE (voir annexe - figure 7.20) :
 - Type Commune1.
 - Type Commune2.
 - Type Commune3.
 - L'indicateur de présence de co-emprunteur(s) sur le crédit courant.
 - L'indicateur de présence de restrictions médicales (totales ou partielles) sur un des risques couverts (décès et/ou ITT) du financement courant.

On y trouve ainsi les trois grandes catégories de données suivantes qui sont :

- Liées directement au profil de l'assuré : âge, civilité...
- Liées à son opération de crédit : montants, durées...
- Liées à son assurance de prêts : niveau de couverture (risque et quotité), tarif...

Auxquelles il est ajouté quelques variables géodémographiques (extraites des bases de l'INSEE - [Bases INSEE]) caractérisant l'environnement de l'assuré qui sont :

- Liées à son lieu de résidence : nombre d'habitants, typologie de commune (rurales ou urbaines). La description est donnée en annexe au 7.2.11.3
- Liées au niveau de vie dans son lieu de résidence : revenu de salaire médian.

Des nouvelles variables susceptibles d'enrichir le modèle seront créées, à partir des variables existantes :

- Les taux de marché ADE et crédit moyens calculés en début de période sur l'ensemble du portefeuille ADE, sur les productions nouvelles et sur des familles de crédit et ADE comparables.
- Les écarts (spreads) crédit (resp. ADE) pour chaque observation entre les taux crédit (resp. ADE) du contrat considéré et les taux de crédit (resp. ADE) moyens de marché calculés ci-dessus. Le principe du calcul des spreads est présenté au 7.2.11.2 en annexe.
- Les marqueurs suivants, caractérisant l'assuré :
 - « urbain » et de « CSP plus » : L'indicateur vaut 1 si l'assuré réside dans une commune de plus de 10 000 habitants et a une CSP élevée ; vaut 0 sinon ($\mathbb{1}_{(TypeCommune3 \geq 3)} \mathbb{1}_{(CSP = 1)}$)

- Jeune et de « CSP plus ». L'indicateur vaut 1 si l'assuré a moins de 38 ans et a une CSP élevée ($\mathbb{1}_{(Age \leq 38 \text{ ans})} \mathbb{1}_{(CSP = 1)}$).

Une analyse descriptive statique peut être conduite à partir des bases étant précédemment construites et validées.

3.3.7 Analyse descriptive de quelques variables

L'étude des variables suivantes permet de mieux caractériser les contrats étudiés.

- 90% des financements empruntés sont <200k€. Il est à noter le poids significatif des montants inférieurs à 100k€, dû aux renégociations de crédits sur les financements d'ancienneté élevée (7.2.13 - figure 7.22 en annexe).
- 60% des durées des financements sont comprises entre 10 et 20 ans (7.2.13 - figure 7.23 en annexe).
- 50% des financements immobiliers ont pour objet l'achat d'un logement (7.2.13 - figure 7.24 en annexe). Près de 15% des financements ont pour objet un investissement locatif.
- Près de 80% des contrats sont couverts totalement ou partiellement en ITT (7.2.13 - figure 7.25 en annexe). Les contrats non couverts en ITT le sont soit parce qu'ils ne sont pas éligibles (par exemple pour les financements in fine totaux ou les prêts courts) ou parce que les assurés peuvent y renoncer (dans le cadre de financements locatifs par exemple), ou encore parce que les assurés ont atteint la limite d'âge pour cette garantie.
- 60% des assurés couverts ont moins de 45 ans - dont près de 30% ont moins de 35 ans (7.2.13 - figure 7.26 en annexe). En outre, 46% sont des femmes, 76% sont multi-emprunteurs (Financement sur plusieurs têtes) , et 10% des assurés ont des restrictions de couverture.
- Les assurés sont des urbains à 73% et résident à 50% dans des villes de moins de 200 000 habitants (7.2.13 - figure 7.27 en annexe).
- Ils sont à 60% des artisans, commerçants, employés ou chefs d'entreprise (7.2.13 - figure 7.28 en annexe). L'analyse des nouvelles variables créées ci-dessus montre que les jeunes CSP+ représentent 17% des contrats et que les urbains CSP+ représentent 5% des contrats. Elle montre également que 45% des assurés ont un autre encours assuré dans le groupe (immobilier ou professionnel).

3.3.8 Etude de la dynamique des indicateurs crédit et ADE mesurée dans les fichiers sources entre juillet 2015 et juillet 2019

3.3.8.1 Evolution des taux de crédits immobiliers effectifs – crédits amortissables

La période [2015, 2019] est marquée par une baisse globale des taux de crédit pour l'ensemble des tranches de durée de crédit, avec une baisse particulièrement marquée sur la période [Avril 2016, Janvier 2017], voir annexe 7.1.2.1 - figure 7.3. Le comportement des taux crédits évalués sur les bases d'étude est en ligne avec ceux disponibles dans les rapports institutionnels.

La mise en perspective de la situation de juillet 2015 avec celle de juillet 2019 montre que la baisse des taux dans les deux cas concerne les contrats les plus récents , voir annexe 7.1.2.1 - figures 7.4 et 7.5 . Les taux ont baissé de manière homogène pour l'ensemble des anciennetés et des tranches d'âge entre juillet 2015 et juillet 2019 ; les deux courbes sont assez similaires.

3.3.8.2 Evolution des taux ADE immobilier annualisés effectifs – couvrant les prêts amortissables

La période [2015, 2019] est marquée par une baisse globale des taux ADE pour les tranches d'âge les plus basses jusqu'en avril 2019 et une remontée depuis octobre 2017 pour les tranches d'âge les plus hautes et globalement depuis avril 2019 pour l'ensemble des tranches d'âge, voir annexe 7.1.2.2 - figure 7.6. La période [2015,2016] s'inscrit dans une baisse des taux pour l'ensemble des tranches. La baisse résulte de l'ouverture du marché de l'ADE.

La mise en perspective de la situation de juillet 2015 avec celle de juillet 2019 montre logiquement une accentuation de la segmentation tarifaire par tranche d'âge qui se traduit par une augmentation tarifaire sur les tranches d'âge à l'entrée les plus élevées et une diminution sur les tranches d'âge les plus basses, voir annexe 7.1.2.2 - figures 7.7 et 7.8. Structurellement, l'effet de segmentation sur les contrats les plus jeunes l'emporte sur la stabilité des contrats les plus anciens.

Compte tenu de la faculté de négociation du réseau des Caisses régionales, l'évolution des indicateurs de crédit ou d'assurance par ancienneté est fortement corrélé à la fois aux volumes de rachats constés ou de négociations clients acceptées.

3.4 Conclusion

Cette partie a été consacrée à la construction des bases d'études, qui sont destinées :

- A la modélisation et au calcul des taux bruts de rachat par ancienneté et segment de risque.
- Au back-testing de la modélisation sur des données plus récentes.

L'étude comportementale se faisant à la maille contrat d'assurance, la base d'étude est agrégée par identifiant assuré par date de signature des contrats. Compte tenu de sa volumétrie la base est échantillonnée de manière stratifiée par ancienneté et Caisse régionale à 10

Les données des Caisses régionales sont enrichies par des données socio-démographiques INSEE.

Le périmètre de l'étude est restreint aux contrats ADE couvrant les prêts immobiliers amortissables qui constitue la plus grande du portefeuille ADE du Crédit Agricole et qui sont également les plus exposés aux risques de rachats.

Le choix des périodes a été guidé par des contraintes de disponibilité, de fiabilité et de stabilité des données. En outre, base de back-testing présente l'intérêt de s'inscrire dans un cadre économique et législatif légèrement différent celle de modélisation :

- La base de modélisation se situe dans un contexte Hamon tandis que de celle de back-test se situe dans un contexte « Bourquin ».
- Les taux de crédit et d'ADE moyens mesurés en début de période sur la base de back-test sont significativement plus faibles que ceux mesurés sur la base de modélisation.

Cette caractéristique devrait permettre de mesurer la capacité d'ajustement du modèle au contexte économique et tester également sa robustesse face à un changement législatif.

Chapitre 4

Modélisation des rachats en assurance emprunteur

Il a été vu précédemment que l'approche actuelle de la modélisation des rachats basée sur l'historique nécessitait d'être challengée dans le contexte actuel caractérisé par une évolution rapide des indicateurs économiques. Cette partie a pour objectif de présenter l'approche alternative dynamique, en partant de la segmentation basée sur des scores d'attrition de chaque ancienneté du portefeuille en classe de risques, pour aboutir à la définition de loi de rachats qui seront mises en œuvre par la suite.

L'étude étant structurée par la façon d'utiliser les lois ultérieurement, une étape préalable consiste à s'intéresser au cadre technique de mise en œuvre.

4.1 Le cadre technique

4.1.1 L'existant

Actuellement les lois de rachats emprunteur du périmètre immobilier sont déterminées uniquement en fonction de l'ancienneté des lignes de Model-Points (MP) projetées, figure 4.1 :

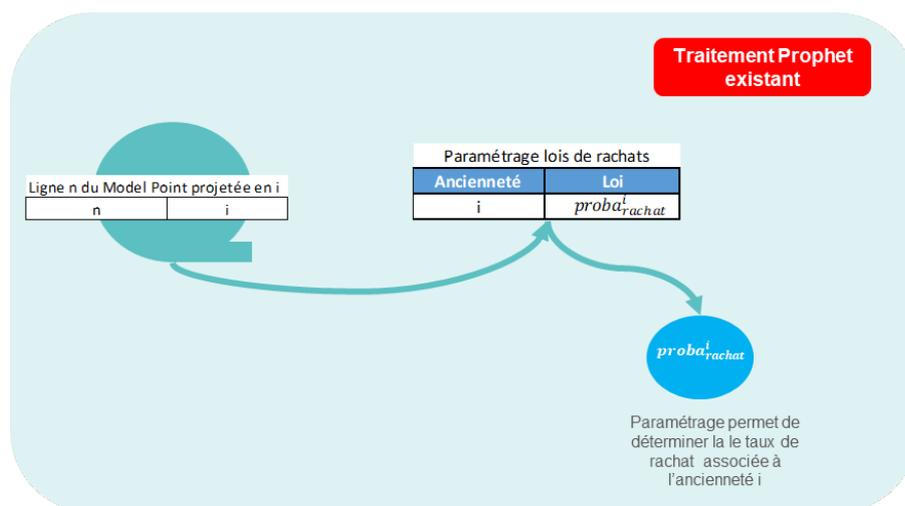


Fig. 4.1 – Approche actuelle

4.1.2 L'approche alternative

L'approche proposée consiste en la mise en œuvre de scores d'attrition par ancienneté permettant le calcul pour chaque ligne de Model-Point, des lois de rachat, dynamiquement au cours de la projection dans prophet. Une première option serait d'intégrer la gestion des scores dans prophet (Figure 4.2) :

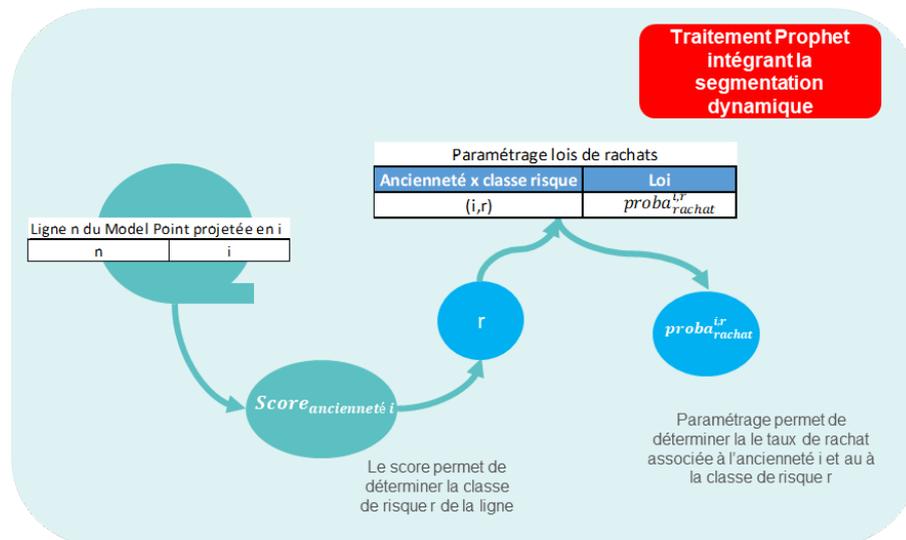


Fig. 4.2 – Approche alternative : mise en œuvre des scores dans prophet

Ce choix n'a pas été retenu car outre son intérêt limité, il impliquerait en termes de paramétrages, des mises à jour lourdes et régulières de l'outil de mise en œuvre.

Une deuxième option – qui a été préférée, consiste à externaliser de l'outil de projection le calcul des scores et des lois associées, qui conduit à introduire entre autres la notion de trajectoire (présentée en détail dans la section suivante) qui permet, pour chaque contrat, à partir de son ancienneté et sa classe de risque initiale de caractériser ses classes de risques au cours du temps et donc ses lois de rachats pour chaque ancienneté, jusqu'au terme de la projection (30 ans). L'optimisation du modèle consiste à minimiser le nombre de trajectoires en les regroupant en clusters. Cette solution permet donc d'affecter à un cluster chaque contrat à partir de son ancienneté et sa classe de risque initiale.

Le paramétrage consiste à implémenter dans prophet les lois par ancienneté pour chaque cluster modélisé, figure 4.3 :

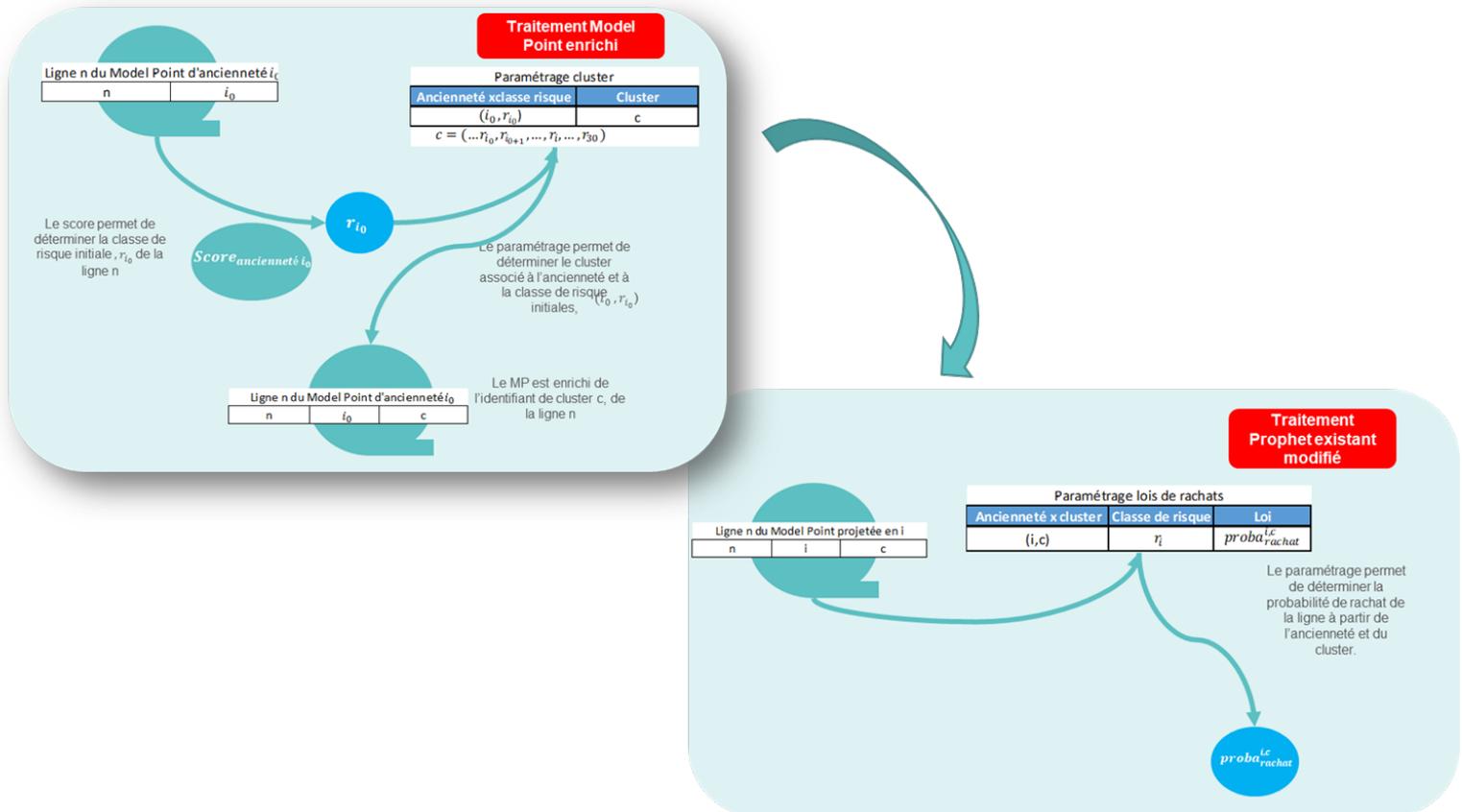


Fig. 4.3 – Approche alternative : mise en œuvre des scores dans les traitements MP

Le prérequis technique étant posé, la partie qui suit décrira d'une part la construction des clusters et les lois associées et d'autre la préparation des données permettant la mise en œuvre du modèle.

4.2 L'approche par des méthodes statistiques prédictives

La mise en place se décompose en plusieurs étapes, pour aboutir aux lois de rachats qui seront finalement utilisées dans l'outil prophet présenté ultérieurement à l'occasion de la mise en œuvre. Dans cette section, les noms des bases définies en partie 3 sont repris.

Etape 1 - Le scoring :

A partir de la base agrégée échantillonnée [BAE1], sur laquelle est appliquée deux méthodes supervisées de classement, un score de rachat par ancienneté de contrat est construit qui permet d'affecter chaque assuré-contrat à une classe de risque. Le nombre de classes de risques optimal sera discuté dans la partie 5.

Etape 2 - Le calcul des taux bruts moyens par ancienneté :

Les scores par ancienneté sont appliqués sur les bases agrégées complètes [B1] et [B2], en s'assurant que le risque de rachat est bien fonction croissante du score. Chaque ligne des bases détaillées [B1] et [B2] est affectée à une classe de risque. Les taux bruts de rachats par ancienneté sont ensuite calculés pour chaque base, desquels sont déduits les taux bruts moyen par ancienneté et classe de risque.

Etape 3 – Le clustering des assurés – prêts :

L'objectif est de construire pour chaque contrat, une trajectoire de classe de risque sur une durée de contrat maximum de 30 ans, à partir de la classe de risque et l'ancienneté initiale. La modélisation se limitera à 15 ans compte tenu du regroupement à 15 ans des anciennetés ultérieures. Dans le paramétrage, les anciennetés supérieures à 15 ans auront la loi de l'ancienneté 15 ans. Cette étape de modélisation permet de définir le référentiel associant à chaque ancienneté et classe de risque initial, un identifiant de cluster.

Etape 4 - L'enrichissement des Model-Points emprunteur :

Il s'agit ici, dans les traitements de production des Model-Points, de valoriser pour chaque contrat, une nouvelle variable « cluster », calculée à partir de son ancienneté de sa classe de risque initiale et du référentiel construit dans l'étape 3 ci-dessus. Cette variable permettra ensuite d'activer la loi de rachat paramétrée pour le cluster associé au contrat, pour chaque ancienneté.

L'étape 3 est réalisée à chaque mise à jour du modèle. Toutefois, il est envisageable de la mettre en œuvre entre deux mises à jour, à la demande ou de façon régulière (exemple en début d'exercice annuel). Cette possibilité n'est pas cependant pas traitée dans cette étude.

4.3 Le scoring

L'objectif (voir [SAPORTA, 2011], [TUFFERY, 2017] et [STA201, 2016]) est d'expliquer la variable Y qualitative « cible » de rachat (présence ou départ du client dans durant la période d'étude) à partir d'un certain nombre de variables explicatives ou prédictives $\{X_k\}_{1 \leq k \leq p}$ à la fois quantitatives et qualitatives (données d'assurance ou socio-démographiques). Le risque de « départ » dans la période d'étude est quantifié par une note de risque S , ou score qui est calculé comme combinaison linéaire des variables explicatives $\{X_k\}_{1 \leq k \leq p}$. $S = \sum_{1 \leq i \leq p} a_i X_i$, les coefficients $\{a_k\}_{1 \leq k \leq p}$ étant optimisés pour la prédiction de Y .

Le choix des méthodes se porte donc assez naturellement vers deux méthodes supervisées de classement :

- La méthode d'analyse factorielle discriminante – plus exactement une déclinaison de cette méthode : la méthode DIScrimination sur variables QUALitatives (DISQUAL).
- La méthode de régression logistique binaire (LOGIT).

Les principes sous-jacents des 2 méthodes sont rappelés succinctement en annexe au 7.3.1. La partie qui suit présente leur mise en œuvre sur la base d'étude agrégée [BEA1]

4.3.1 Mise en œuvre

Les deux méthodes présentées précédemment vont être mises en œuvre sur la base échantillonnée agrégée des contrats [BEA1], pour chaque ancienneté a de contrats ($1 \leq a \leq 15$), après avoir sélectionné les variables d'étude les plus pertinentes.

4.3.1.1 Etudes statistiques univariées

Pour chaque ancienneté, les variables à la fois les plus discriminantes (analyses univariées) et les moins liées deux à deux (analyses bivariées) sont sélectionnées.

Distribution des variables explicatives par rapport à la variable à expliquer¹

Les analyses univariées pour chaque ancienneté conduisent à découper les variables continues et à regrouper les modalités des variables explicatives. Ce travail a été mené avec comme objectif :

- D'éviter les modalités avec des effectifs trop faibles.
- D'augmenter les différences de distributions des variables relativement à la variable à expliquer.

Liaisons entre les variables explicatives la variable à expliquer

L'analyse fait appel au coefficient de Cramer (V de Cramer) qui varie de 0 (liaison nulle) à 1 (liaison maximale) - voir annexe 7.3.3.

L'analyse des V de Cramer des variables explicatives par rapport à la variable à expliquer, par anciennetés, donne les listes des variables les plus discriminantes, triées par ordre décroissant. Seules sont sélectionnées les variables au-dessus d'un seuil de 5%.

Pour les anciennetés 14 et 15, il n'y a pas de variable qui satisfasse au seuil des 5%. En conséquence, pour ces anciennetés il ne sera pas possible de définir de score de risque pertinent. En conséquence, la modélisation des rachats pour ces anciennetés se limitera à l'estimation d'un taux théorique moyen.

En synthèse, figure 4.4, il ressort que les variables les plus fréquemment sélectionnées sont :

- Des variables liées au crédit (taux, montants ou durée).
- Des variables directement liées à l'assurance (taux de cotisation, taux de remise tarifaire) ou indirectement (l'âge de l'assuré qui est fortement lié au niveau de tarif comme le montrera l'étude bivariée).

L'analyse montre également que les critères liés à l'assurance entrent en compte dans la décision de rachat au delà du délai « Hamon ».

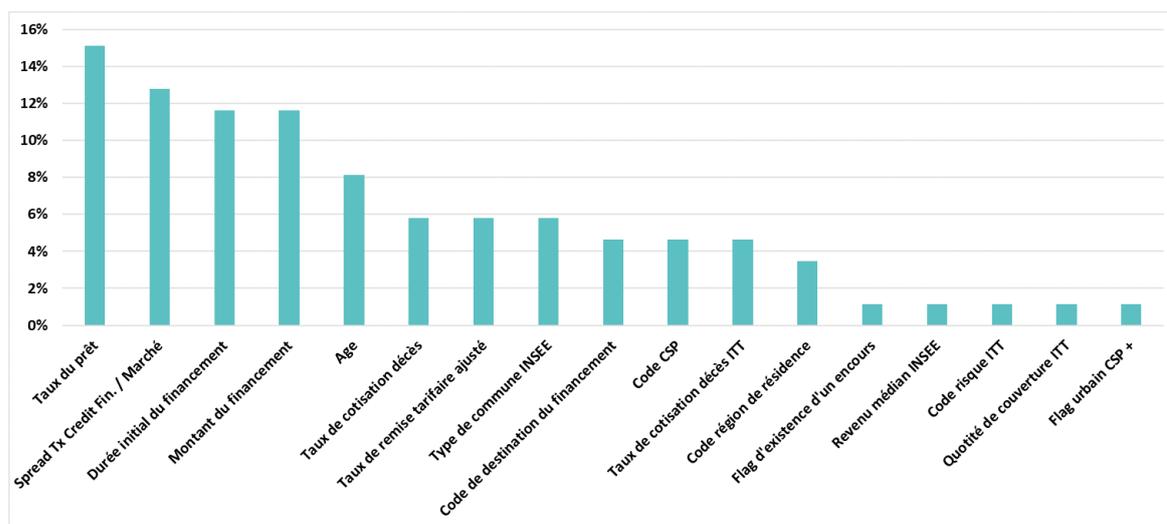


Fig. 4.4 – Fréquence des variables sélectionnées – Analyse univariée

4.3.1.2 Etudes statistiques bivariées

L'analyse des V de Cramer des variables explicatives entre elles par ancienneté, donne la liste (partielle) suivante des couples de variables les plus liées, triées par ordre décroissant. Seules les variables des couples dont le V de Cramer est au-dessus de 0,4 sont sélectionnées.

En synthèse, figure 4.5, il ressort que :

- Les variables les plus discriminantes sont globalement conservées.
- Les variables assurantielles (taux de cotisation, taux de remise tarifaires) fortement liées entre elles et à l'âge des assurés ne sont pas conservées.

1. = Flag_Assurance

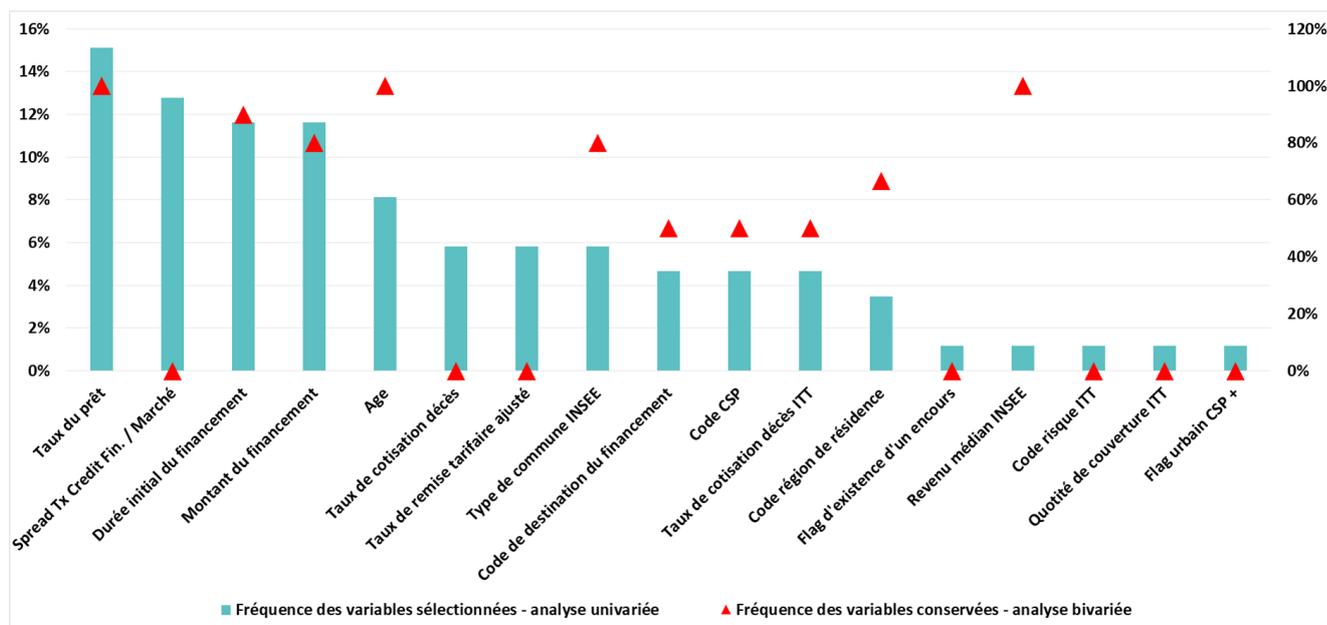


Fig. 4.5 – Fréquence des variables sélectionnées x fréquence des variables conservées – Analyse bivariée

4.3.1.3 Analyse des Correspondances Multiples (ACM)

Pour chaque ancienneté, l'ACM est effectuée sur le tableau disjonctif complet des variables sélectionnées précédemment.

Afin d'orienter notre étude, seuls les 3 premiers axes sont retenus. Ils définissent les 3 premiers plans factoriels de l'ACM. En toute rigueur, l'équivalent du critère de Kaiser pour l'ACM, nous amènerait à retenir les axes correspondant à des valeurs propres supérieures à $\frac{1}{nb\ variables}$. Seules les représentations des modalités dont le cos^2 sur au moins 1 des 2 axes est supérieure à 0,02 sont prises en compte pour l'analyse.

Pour l'ancienneté 6 par exemple (voir 7.3.2.1 en annexe), le 1^{er} plan factoriel montre une séparation par le 2^e axe des modalités "Dur._Fin. ≤ 237 mois" et "Tx_Cotis. < 0,293%" d'une part et des modalités "Dur._Fin. > 237 mois" et "Tx_Cotis. > 0,293%" d'autre part. Les résiliations concernent davantage les contrats de durée supérieurs 19,5 ans et dont le taux ADE est supérieur à 0,293% annuel. Le 2^e plan factoriel montre une séparation par le 1^{er} axe des modalités "CSP_1_2" et "CSP_Autres". Le 1^{er} axe sépare lui les modalités "Âge ≤ 43" et "Mt_Fin. > 100,8k€" d'une part et les modalités "Âge > 43" et "Mt_Fin. ≤ 100,8k€" d'autre part. Les résiliations concernent davantage les financements dépassant 100,8k€ couverts pour des assurés âgés de moins de 43 ans et à CSP favorable. Enfin le 3^e plan factoriel montre une séparation par le 2^e axe des modalités "Tx_Prêt > 3,15%" d'une part et des modalités "2,75 < Tx_Prêt ≤ 3,15%" et "Tx_Prêt ≤ 2,75%" d'autre part. Les rachats concernent les contrats couvrant les financements caractérisés par des taux de crédit supérieurs à 3,15%.

4.3.1.4 Mise en œuvre des méthodes

Une étape préalable consiste en un échantillonnage des données d'étude. Deux tiers des observations définissent l'échantillon d'apprentissage et un tiers celui de validation.

L'analyse discriminante qualitative :

Pour chaque ancienneté sont sélectionnés les axes satisfaisant au test de F à 15% (par défaut) appliqué au lambda de Wilks (voir annexe 7.3.4), avec un test du F inférieur à 5%, pour éviter d'aller chercher les axes de faible inertie différenciant trop peu les individus.

Pour chaque ancienneté le modèle est optimisé en le simplifiant progressivement par une suppression un à un les axes sélectionnés et en faisant converger de manière satisfaisante l'aire sous la courbe ROC de l'échantillon d'apprentissage avec celle de validation. L'objectif est d'éviter les situations de surapprentissage qui pénaliseraient le pouvoir de généralisation et donc la robustesse du modèle.

Par exemple pour l'ancienneté (figure 7.30 en annexe), 6 axes ont été sélectionnés initialement. L'optimisation conduit à n'en retenir 4, pour lesquels la différence d'AUC entre l'échantillon d'apprentissage et de validation est minimal.

Il est à noter ici que la dégradation du pouvoir de généralisation du modèle liée à sa complexité est assez limitée. L'optimisation des modèles conduit à la synthèse des AUC des échantillons d'apprentissage et de validation par ancienneté suivante, fFigure 4.6 :

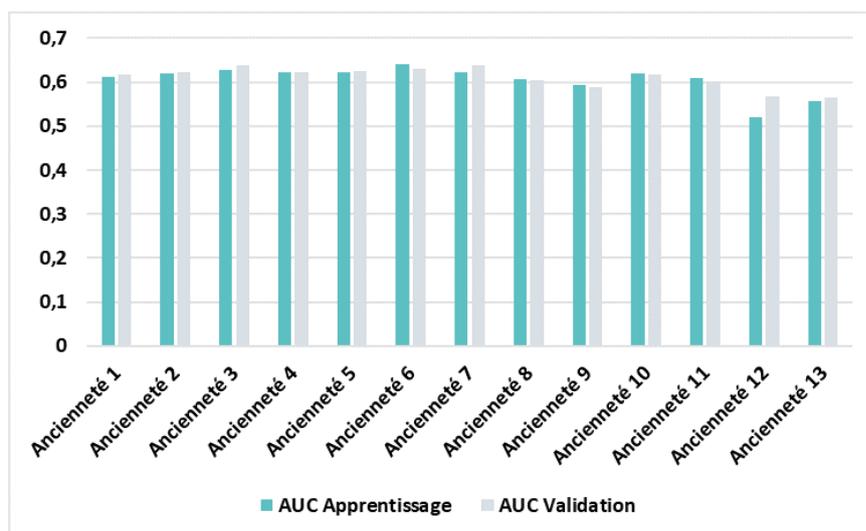


Fig. 4.6 – Synthèse de l'optimisation des modèles DISQUAL par ancienneté

Il est procédé aux tests de normalités et d'homoscédasticité, sous jacents à la méthode discriminante qualitative. Les tests de normalité des axes sélectionnés par le test de Kolmogorov-Smirnov (compte tenu de la volumétrie étudiée) font ressortir des p-values amenant à rejeter l'hypothèse nulle de normalité. Le résultat de ces tests pourra expliquer la qualité des modèles produits par cette méthode. La mise en œuvre d'une méthode non paramétrique que l'on est tenté de lui préférer améliore sensiblement les performances du modèle, mais reste la plupart du temps en deçà de la méthode logistique exposée plus loin). L'étude sera donc poursuivie avec la méthode paramétrique. Les tests d'homoscédasticité par le test de Levene mené sur les axes ne permettent pas de rejeter significativement l'hypothèse nulle d'homoscédasticité.

Le calcul des grilles de score pour chaque ancienneté :

Les résultats du modèle sont réappliqués² à l'ensemble des individus et non plus seulement au seul échantillon d'apprentissage ce qui permet d'obtenir des coefficients moins biaisés.

A partir des coefficients de la fonction discriminante linéaire pour les axes sélectionnés, la fonction de score sur les axes de l'ACM est calculée par différence des deux colonnes correspondant aux modalités de la variable à expliquer.

la valeur du score est ensuite calculée³ sur les modalités des variables initiales. Ces scores conduisent à la définition d'une grille de scores présentée au 7.3.8 en annexe.

La grille obtenue est appliquée à l'ensemble des individus.

2. procédure SAS DISCRIM

3. procédure SAS SCORE

Les modalités des variables retenues par exemple pour l'ancienneté 6 présentées au 7.3.2.2 (en annexe) par ordre de nombre de points décroissants sont ainsi les suivantes :

- La durée totale du financement > 237 mois à hauteur de 26 points.
- Le taux de prêt > 3,15%, à hauteur de 20 points.
- Les codes CSP 1 et 2 (les CSP+), à hauteur de 17 points.
- Le taux de prêt > 2,75% et ≤ 3,15%, à hauteur de 15 points.
- L'âge ≤ 43 ans, à hauteur de 15 points.
- Le taux de cotisation ADE total annualisé ≤ 0,293%, à hauteur de 12 points.
- Le montant de financement ≤ 100,8k€, à hauteur de 10 points.

L'analyse des grilles par ancienneté obtenue par la discriminante qualitative fait ressortir le poids significatif sur les scores de l'ensemble des anciennetés des modalités des variables crédit (taux de prêt, montant et durée de financement), mais également celles des âges sur les 7 premières années du contrat d'assurance (Figure 4.7).

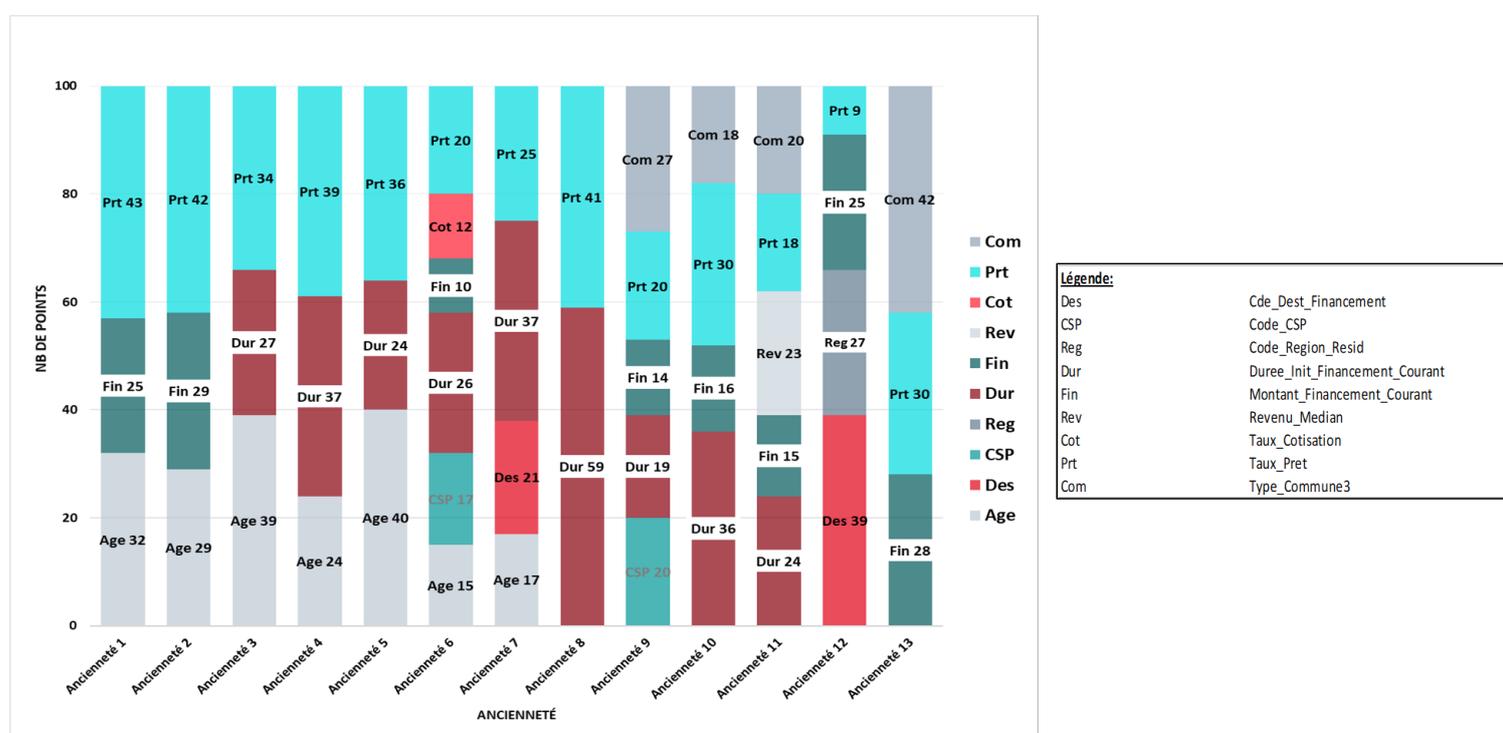


Fig. 4.7 – Cartographie des scores par variable et poids de la modalité maximum associée - DISQUAL

La régression logistique :

La régression logistique sous-tend une hypothèse d'absence de colinéarité (multicolinéarité) entre les variables. Nous procédons à une analyse du facteur d'inflation de variance (VIF) et des index de condition (CI)⁴ qui ne fait pas apparaître des $VIF > 10$ et des $CI > 100$. Il peut donc être considéré que la multicolinéarité est faible et que cette hypothèse d'application de la méthode de régression logistique est satisfaite.

Il est procédé à une étape de sélection pas à pas - de type stepwise, des variables utilisées dans le modèle, avec un seuil de significativité pour entrer et rester dans le modèle de 5%. On s'assure ensuite que la sélection des variables ne peut pas être simplifiée par une analyse de critères classiques d'ajustement du modèle, comme par exemple le critère BIC (Schwartz). Enfin il est vérifié que les modalités des variables sélectionnées ont un χ^2 de Wald supérieur au seuil critique de 3,84.

4. via les options VIF et COLLIN de la procédure REG de SAS

La procédure de construction itérative du modèle est lancée. Pour chaque ancienneté les étapes de création peuvent être visualisées graphiquement. Voir par exemple Figure 7.33) en annexe pour l'ancienneté 6.

Comme lors de la mise en œuvre de l'analyse discriminante, le modèle est réappliqué⁵ à l'ensemble des individus et plus seulement au seul échantillon d'apprentissage ce qui permet d'obtenir des coefficients moins biaisés.

A partir des coefficients de la régression logistique, la grille de score est calculée de la même façon que précédemment, avec l'analyse discriminante qualitative.

Les modalités des variables retenues par exemple pour l'ancienneté 6 présentées au 7.3.2.3 (en annexe) par ordre de nombre de points décroissants sont ainsi les suivantes :

- Le taux de prêt $> 3,15\%$, à hauteur de 32 points.
- Le taux de prêt $> 2,75\%$ et $\leq 3,15\%$, à hauteur de 25 points.
- Les codes CSP 1 et 2 (les CSP+), à hauteur de 19 points.
- L'âge ≤ 43 ans, à hauteur de 19 points.
- Le taux de cotisation ADE total annualisé $\leq 0,293\%$, à hauteur de 11 points.
- Le montant de financement $\leq 100,8k$ €, à hauteur de 10 points.
- La durée totale du financement > 237 mois à hauteur de 9 points.

L'analyse des grilles par ancienneté obtenue par la régression logistique fait logiquement ressortir des similitudes avec celle obtenue par la discriminante qualitative, à savoir le poids significatif (en nombre de points) des modalités des taux de prêt, durée et montant de financement sur les scores de l'ensemble des anciennetés, et celles des âges sur les 7 premières années, figure 4.8.

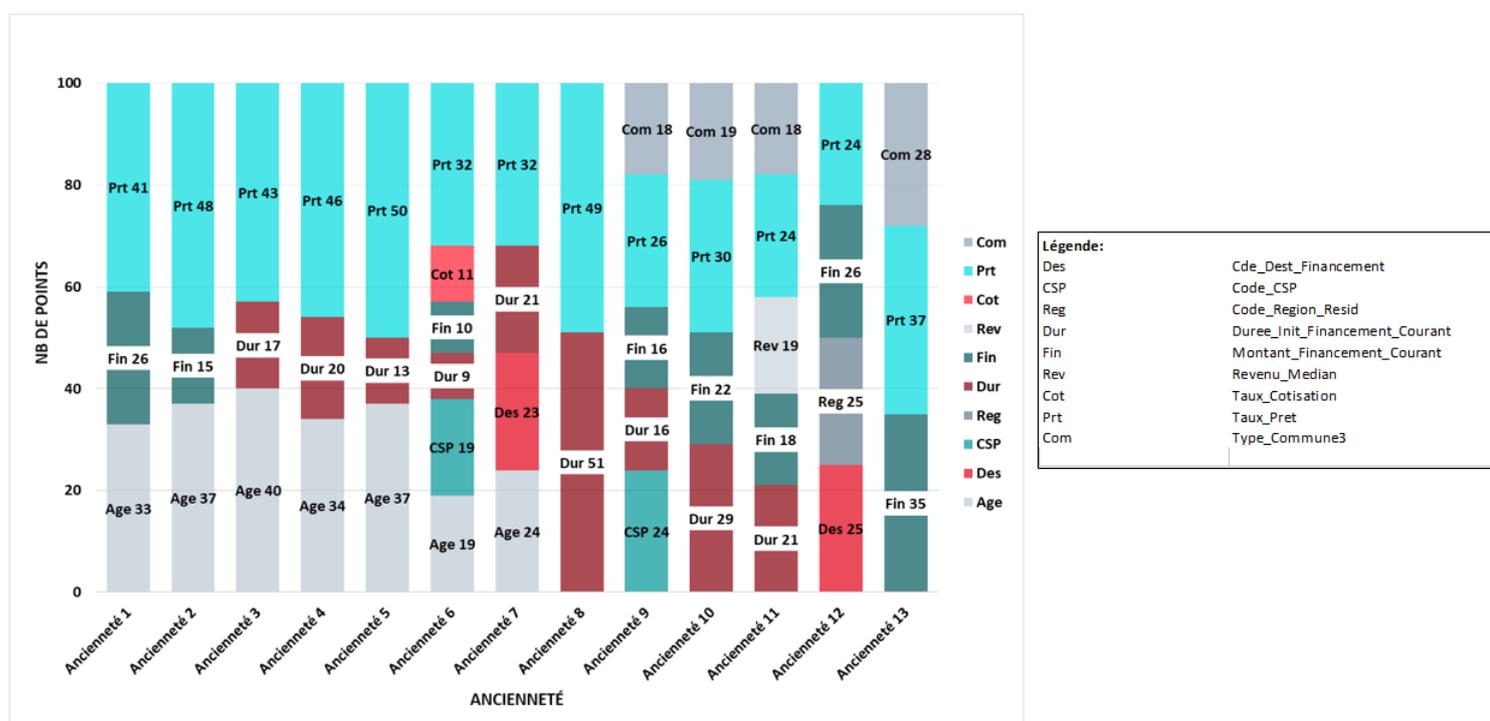


Fig. 4.8 – Cartographie des scores par variable et poids de la modalité maximum associée - LOGIT

Des différences de pondération des modalités des variables entre les scores des 2 méthodes sont constatées. Ce point est illustré dans ce qui précède, dans la décomposition des scores de l'ancienneté 6 par modalités.

5. procédure SAS LOGISTIC

L'analyse de la performance des modèles utilisés dans l'étude :

Le croisement des écarts relatifs (jauges dans le tableau qui suit (Figure 4.9), avec le poids des modalités des variables dans les scores (fond rouge dans le tableau ci-dessous) montre que les écarts les plus importants sont globalement concentrés sur les variables ayant une contribution moyenne ou faible. Des écarts relatifs significatifs (>14%) sont cependant constatés sur la durée des anciennetés 4, 6 et 7 et le taux de prêt de l'ancienneté 5 qui contribuent fortement aux scores de ces anciennetés. Ces écarts contribuent à expliquer les différences de performance entre les 2 méthodes.

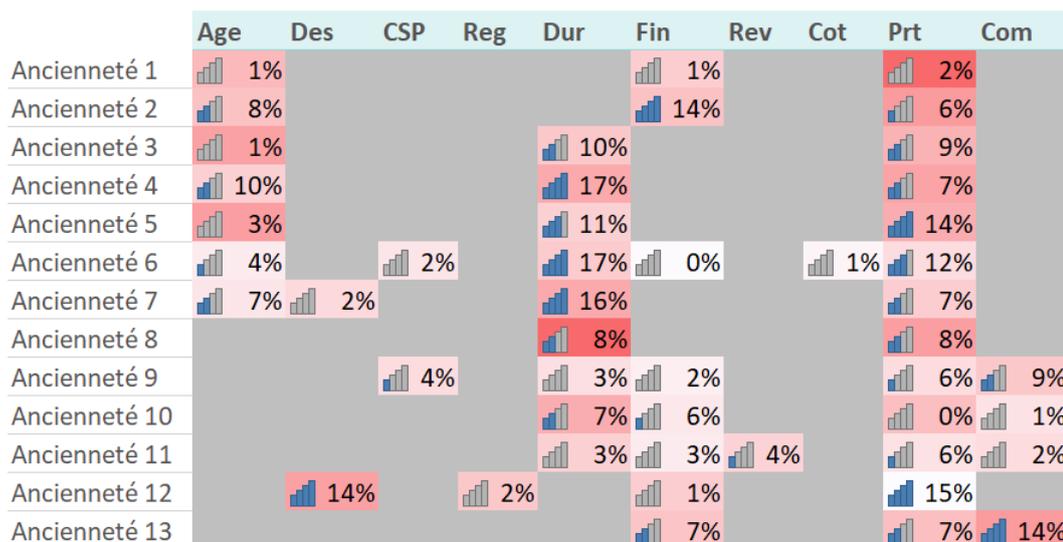


Fig. 4.9 – Cartographie des écarts relatifs x poids des modalités entre LOGIT et DISQUAL

Les performances des deux modèles sont analysées à la fois, en superposant par ancienneté les courbes ROC des régressions linéaires LOGIT et des analyses discriminantes DISQUAL (les courbes pour l'ensemble des anciennetés sont disponibles en annexe 7.3.7 - figure 7.38), et en comparant également les AUC des 2 modèles, figure 4.10, la synthèse des AUC par ancienneté.

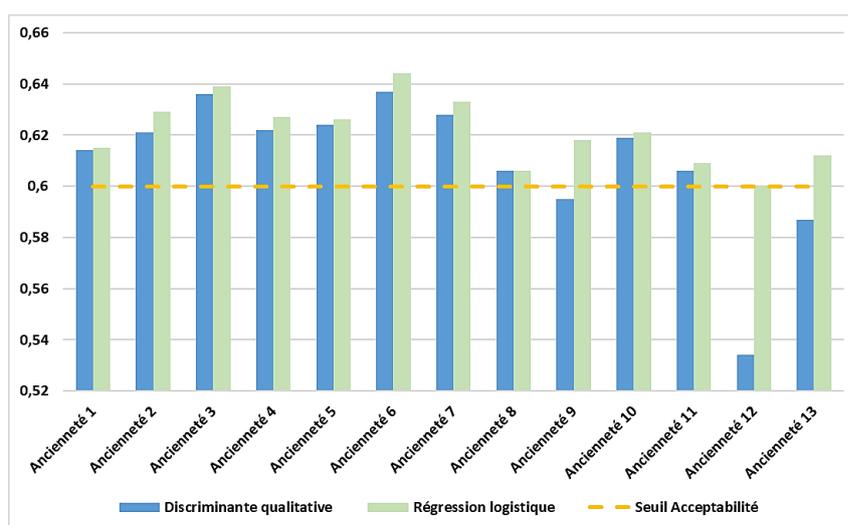


Fig. 4.10 – Par ancienneté et méthode prédictive

Les AUC de la régression logistique sont globalement supérieures à celles de l'analyse discriminante. En outre, les courbe ROC de la regression logistique sont systématiquement au dessus de celles de la régression logistique (voir 7.3.7 - figure 7.38 en annexe)

Les performances de la régression logistique apparaissent donc globalement supérieures à celles de l'analyse discriminante, même si le modèle est dans l'ensemble assez peu discriminant selon les critères habituels rappelés dans le tableau 7.37 en annexe.

Les deux raisons vraisemblables sont, sont liées aux données en entrée de l'analyse discriminante qui ne satisfont pas convenablement les deux hypothèses principales requises par cette méthode :

- La normalité des distributions.
- L'homoscédasticité des des matrices de variances/covariances.

En revanche l'hypothèse d'absence de multicolinéarité requise pour la méthode de régression logistique est davantage satisfaite. Les grilles de score par ancienneté, provenant de la méthode la plus performante seront retenues.

Le poids des variables sélectionnées dans les scores :

La modélisation a été réalisée un contexte économique baissier des taux ADE et de crédit, à l'origine de taux de résiliations significatifs. Elle est caractérisée également par le cadre réglementaire ADE "Hamon" dans laquelle elle a été réalisée.

L'analyse des scores construits à partir des 2 méthodes fait apparaître 2 groupes de scores : un 1^{er} groupe modélisant les anciennetés ≤ 8 ans qui est composé essentiellement d'indicateurs crédit ou ADE (Montant de CI, taux de crédit, taux de prêt, durée de financement âge de l'assuré) et un 2^e groupe modélisant les anciennetés > 8 ans qui intègre des indicateurs socio-démographiques (typologie de la commune, niveau de revenu médian, CSP). Ce point s'explique par la durée moyenne effective des crédits (autour de 8-10 ans), au delà de laquelle les résiliations (entraînées entre autres par les remboursements anticipés, les investissements locatifs) sont davantage structurelles et moins sensibles à la seule évolution du contexte économique global.

Dans le 1^{er} groupe de scores les variables les plus structurantes sont le taux de prêt et l'âge de l'assuré. Cette observation montre d'une part le poids significatif du crédit dans la résiliation du contrat d'assurance et ce dès la 1^{re} année de souscription du contrat d'assurance, pendant laquelle l'assuré peut librement résilier son contrat d'assurance au profit d'un contrat alternatif offrant le même niveau de garanties (délai Hamon). Elle montre d'autre part que l'âge de l'assuré fortement corrélé aux tarifs ADE, joue un rôle important dans la décision de résiliation et ce même au-delà de la 1^{re} année de souscription. Ce point permet de penser que le rachat du crédit au-delà de la 1^{re} année peut avoir été motivé par un souhait de résilier l'ADE dans un contexte réglementaire ne permettant pas de le faire directement (pour profiter, par exemple, d'un niveau tarifaire plus favorable ou d'une réduction d'assiette de primes). Cette hypothèse peut expliquer le taux réel de rachat à un an 2019 (sur la base de back-test) calculé dans un cadre réglementaire ADE "Bourquin", qui ressort plus faible que ce que l'évolution réglementaire conduisait à anticiper.

Le calcul des scores et des classes de risque sur la base échantillonnée agrégée et sur la base agrégée complète :

Les grilles de scores $score_i$ (normalisées sur une base de 100 points, voir 7.3.8 en annexe) sont établis pour chaque ancienneté i à partir de la méthode prédictive la plus performante. La classe de risque $classe_i$ à n modalités pour chaque ancienneté i est définie comme regroupements des $score_i$.

Pour $n = 3$:

- $[0, 30 \text{ points}]$: risque faible.
- $]30, 70 \text{ points}]$: risque moyen.
- $]70, 100 \text{ points}]$: risque élevé.

Pour $n = 4$:

- $[0, 30 \text{ points}]$: risque faible.
- $]30, 50 \text{ points}]$: risque moyennement faible.
- $]50, 70 \text{ points}]$: risque moyennement élevé.
- $]70, 100 \text{ points}]$: risque élevé.

Les scores et classes de risque de chaque observation de la base considérée sont définies comme suit :

$$\begin{aligned} \text{score}(\text{observation}) &= \sum_{1 \leq i \leq 15} \text{score}_i(\text{observation}) \mathbb{1}_{\{\text{ancienneté}(\text{observation})=i\}} \\ \text{classe}(\text{observation}) &= \sum_{1 \leq i \leq 15} \text{classe}_i(\text{observation}) \mathbb{1}_{\{\text{ancienneté}(\text{observation})=i\}} \end{aligned}$$

Les grilles sont construites sur la base échantillonnée agrégée des contrats du [BEA1]

La robustesse des grilles calculées est vérifiée en les appliquant aux bases [BA1] et [BA2] complètes.

L'analyse des taux de rachats par nombre de points et regroupements par classes de risques sur les bases échantillonnées et les bases complètes permet de constater que :

- Pour toutes bases, le taux de rachat est bien fonction croissante du nombre de points.
- L'ajustement du score appliqué sur les bases [BEA1] et [BEA2] est satisfaisant.
- Pour un même nombre de points, le taux de rachat sur la base [BA2] est inférieur à celui sur la base [BA1]. L'écart se réduit avec l'ancienneté.

Les probabilités de rachat en fonction scores (avec ou sans regroupements en classes de risque) peuvent être représentés graphiquement par ancienneté et pour chaque base ([BEA1], [BA1] et [BA2]). Le graphique pour l'ancienneté 5 est disponible en annexe, figure 7.39).

Et pour l'ensemble des anciennetés et des segmentations sur la base agrégée complète [BA1] et la base échantillonnée [BEA1], la synthèse ci-dessous, figure 4.11, permet de s'assurer de la cohérence des résultats :

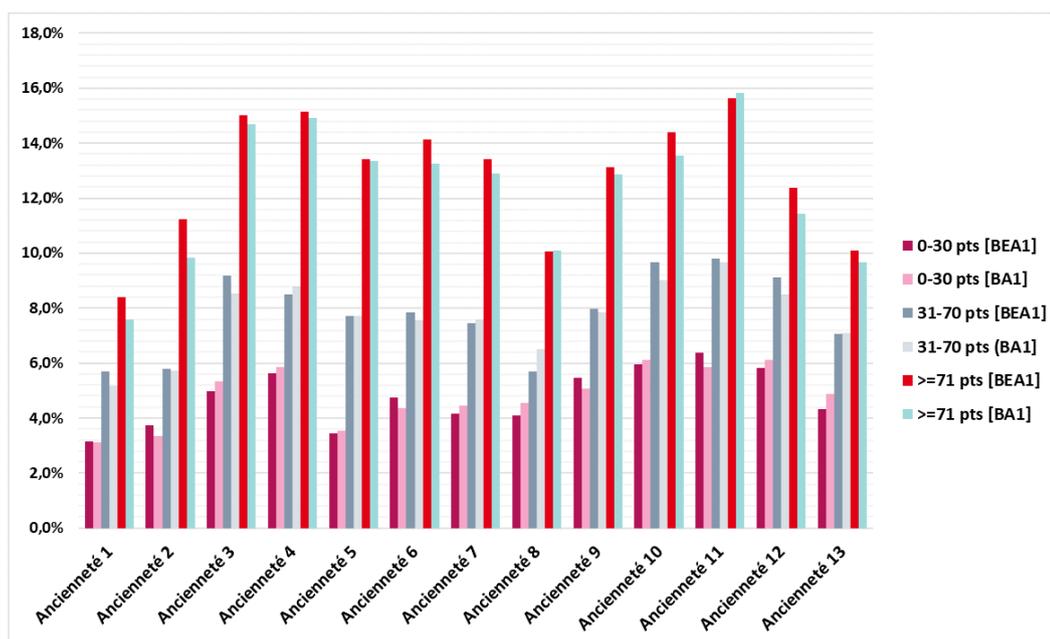


Fig. 4.11 – Synthèse des probabilités de rachats en fonction des classes de risques sur les bases d'étude agrégées

4.4 Calcul des taux bruts moyens par ancienneté

Les taux bruts par ancienneté et classes de risques sont calculés comme moyenne pondérée par les effectifs des taux bruts par classe de risque et ancienneté des bases [B1] et [B2] estimés par l'application de l'estimateur de Hoem présenté en annexe au 7.3.111.

Les taux moyens élevés pour les anciennetés 3 à 6 ans sont cohérents avec les rachats constatés sur la période. En même temps, une forte dispersion des taux de rachats par classe de risque est à noter, ci-dessous (Figure 4.12) :

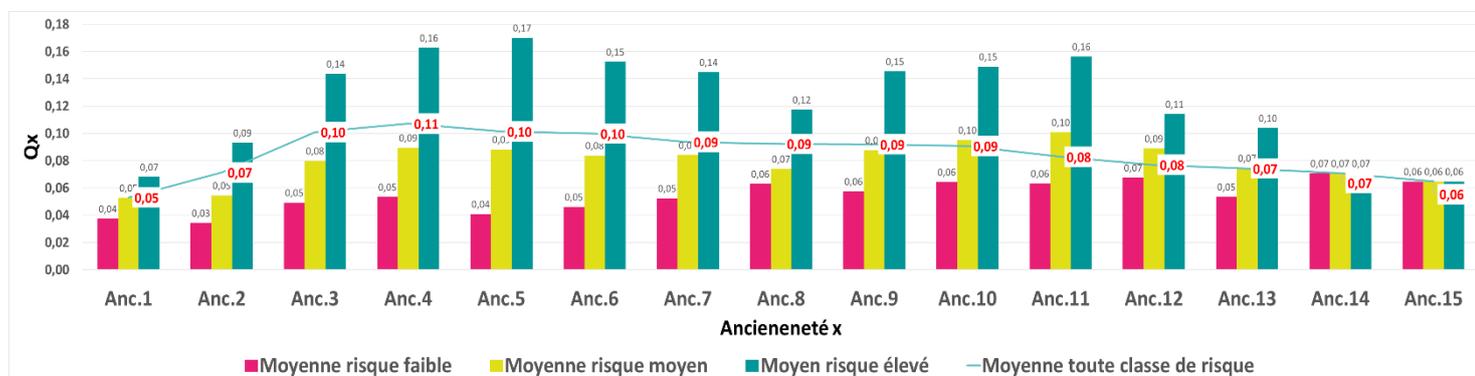


Fig. 4.12 – Taux bruts par classe de risque x Ancienneté

4.5 Clustering des observations

Dans ce qui suit le contexte de modélisation (« modèle ») sera distingué de celui du back-test (« test ») entre deux mises à jour du modèle. Dans cette présentation, la mise à jour intermédiaire est réalisée avec la base de back-test [B3].

Les termes d'indicateur (resp. seuil) et variable de score (resp. modalité de variable de score) seront également indistinctivement utilisés.

Par exemple, pour le score de l'ancienneté 6, ci-dessous, les taux 1,75% et 3,15% sont les seuils de l'inducteur taux de prêt et 43 le seuil de l'inducteur âge.

$$nb\ points_{anc\ 6} = 19.1_{\hat{Age} \leq 43} + 19.1_{CSP=1ou2} + 19.1_{Durée \geq 237\ mois} + 10.1_{Montant=100,8k\text{€}} \\ + 11.1_{Tx\ Cotisation\ ADE > 0,293\%} + 25.1_{2,75\% < Tx\ prêt \leq 3,15\%} + 32.1_{Tx\ prêt > 3,15\%}$$

4.5.1 Le calage des seuils

Avant de présenter le principe de l'ajustement, les inducteurs des seuils concernés sont les variables retenues dans le modèle et susceptibles d'évoluer dans le temps :

- Les indicateurs économiques :
 - Les taux de crédit
 - Les taux tarifaires ADE
- Les indicateurs structurels du portefeuille :
 - Les montants de CI
 - Les durées des financements

Les évolutions moyennes par ancienneté de ces variables considérées comme des drivers des seuils sont présentées au 7.3.12 - figures 7.43 à 7.46 en annexe.

Deux types de calages sont identifiés :

- Un calage des seuils du modèle, qui est mis en œuvre dans l'étape de construction du modèle (entre autres, pour la définition des trajectoires abordées un peu plus loin). L'idée est de définir dynamiquement un score pour l'ancienneté initiale a_0 « vieillie » à n ($1 \leq a_0 \leq 15, 1 \leq n \leq 14 - a_0$), à partir de celui de l'ancienneté $a_0 + n$ provenant de l'étude statistique qui précède. Ce score initial d'ancienneté a_0 projeté à n doit également intégrer les contextes baissiers des taux de crédit ou d'ADE pour la présente étude que des contextes haussiers taux ADE depuis fin 2019 ou crédit depuis 2022. Le principe est de conserver les indicateurs du score de l'ancienneté $a_0 + n$ et d'ajuster les seuils des inducteurs listés ci-dessus. Cette correction permet d'utiliser les scores par ancienneté construits dans l'étude statistique pour estimer les classes de risques prospectives (qui définiront les trajectoires que l'on verra plus loin).
- Un calage des seuils de test, qui est mis en œuvre au moment du back-testing, pour le calcul de la classe de risque initiale qui permettra de déterminer les lois de résiliations utilisées dans les projections de l'outil prophète. L'idée est d'utiliser les scores du modèle et leurs inducteurs en ajustant leurs seuils afin d'intégrer au mieux l'évolution du contexte entre le modèle et le test.

Remarque : cette utilisation repose sur l'hypothèse forte que les listes des inducteurs composant les scores par ancienneté demeurent stables dans le temps et que seuls les seuils évoluent, selon précisément l'ajustement présenté dans cette partie. Ceci est certainement vérifié si les inducteurs sont eux-mêmes assez stables dans le temps et l'est intuitivement moins s'ils évoluent. Dans ce cas la mise à jour des scores et de leur composition, par une nouvelle étude statistique s'avère nécessaire. la methode la plus performante.

4.5.1.1 L'ajustement des seuils du modèle

l'objectif est de définir ici le $Seuil(I)_{modèle}^{a_0, n}$ de l'inducteur I du score de l'ancienneté initiale a_0 « vieillie » à n .

Il est calculé à partir :

- Du seuil de l'inducteur I du score de l'ancienneté a_0+n , noté $Seuil(I)_{modèle}^{a_0+n}$,
- D'un estimateur de l'inducteur I calculé sur les données du modèle et d'ancienneté $a_0 + n$, noté $\overline{I}_{modèle, a_0+n}$ et
- D'un estimateur de I calculé sur les données du modèle (mod.) et d'ancienneté a_0 projeté en n , noté $\overline{I}_{modèle, a_0, n}$; cet estimateur nécessite de prendre des hypothèses sur l'évolution de I au cours des $14 - a_0$ années suivant le date de construction du modèle.

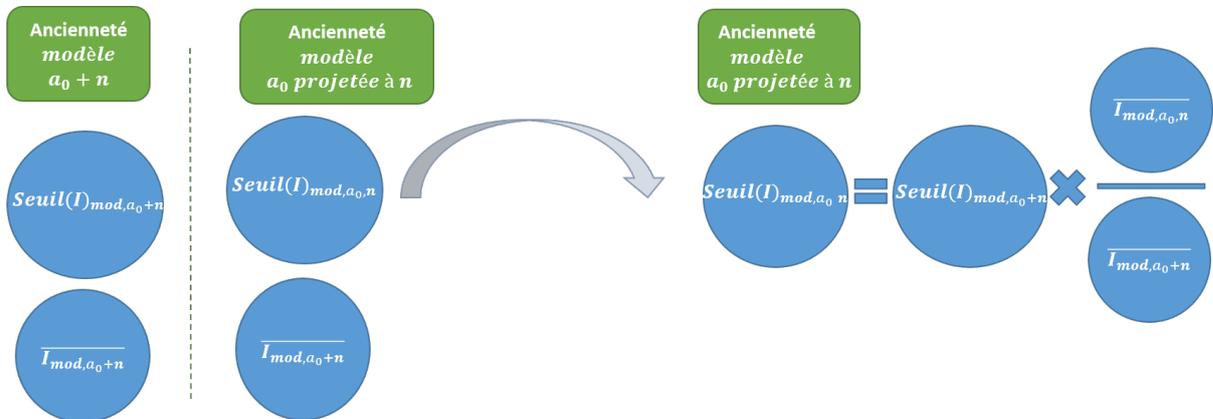


Fig. 4.13 – principe du calage des seuils du modèle

L'ajustement proposé consiste à faire évoluer les seuils des indicateurs du modèle comme les indicateurs moyens. Pour l'ancienneté initiale a_0 projetée en n il nécessite donc :

- De calculer pour $a_0 + n$ la moyenne arithmétique de I sur la base de modélisation [B1] : $\overline{I_{\text{modèle},a_0+n}}$
- De calculer pour a_0 la moyenne arithmétique de I sur la base de modélisation [B1] : $\overline{I_{\text{modèle},a_0}}$
- D'intégrer un scénario économique d'évolution en n de $\overline{I_{\text{modèle},a_0}}$, ci-dessus :

$$\overline{I_{\text{modèle},a_0,n}} = \overline{I_{\text{modèle},a_0}} (1 + EvO_{I,\text{modèle}}^{a_0,n})$$

Le seuil ajusté de l'indicateur I du score d'ancienneté a_0 projeté en n , $n \leq 14 - a_0$ est défini ci-dessous, à partir du seuil de l'inducteur I du score d'ancienneté $a_0 + n$ du modèle, tel que :

$$Seuil(I)_{\text{modèle}}^{a_0,n} = Seuil(I)_{\text{modèle}}^{a_0+n} \frac{\overline{I_{\text{modèle},a_0}} (1 + EvO_{I,\text{modèle}}^{a_0,n})}{\overline{I_{\text{modèle},a_0+n}}}$$

Par exemple, à partir du paramétrage (Figure 7.40 en annexe), le facteur d'ajustement des seuils du modèle est calculé comme suit pour le taux de prêt d'ancienneté 3 ans projeté à 3 ans (avec une hypothèse économique paramétrée d'une chute de 20% du taux).

A partir de $a_0 = 3$, $EvO_{Tx \text{ Prêt},\text{modèle}}^{a_0,n} = -20\%$, $\overline{I_{\text{modèle},3}} = 2,73\%$, $\overline{I_{\text{modèle},6}} = 3,08\%$ le facteur d'ajustement est tel que :

$$\frac{\overline{I_{\text{modèle},3}} (1 + EvO_{Tx \text{ Prêt},\text{modèle}}^{3,3})}{\overline{I_{\text{modèle},6}}} = \frac{2,73\% (1 - 20\%)}{3,08\%} = 71\%$$

De la même façon, l'ajustement du taux de prêt d'ancienneté 3 ans projeté à 4 ans (avec une hypothèse économique paramétrée d'une chute de 20% du taux) est calculé comme suit :

$$\frac{\overline{I_{\text{modèle},3}} (1 + EvO_{Tx \text{ Prêt},\text{modèle}}^{3,4})}{\overline{I_{\text{modèle},7}}} = \frac{2,73\% (1 - 20\%)}{3,25\%} = 67\%$$

Remarque : en l'absence du calage ci-dessus, le seuil projeté correspond au seuil de l'ancienneté du modèle, i.e. $Seuil(I)_{\text{modèle}}^{a_0,n} = Seuil(I)_{\text{modèle}}^{a_0+n}$

Avec les notations qui précèdent, en considérant par exemple le seuil de déclenchement de rachat de l'indicateur taux de crédit dans les 2 situations suivantes, il viendrait :

- En situation haussière des taux du modèle, entre les anciennetés 2 et 1 :
 - $Seuil(Tx \text{ prêt})_{\text{modèle}}^2 = Seuil(Tx \text{ crédit})_{\text{modèle}}^{1,1} = 1,5\%$
 - $\overline{Tx \text{ prêt}}_{\text{modèle}}^1 = 2\%$
 - $\overline{Tx \text{ prêt}}_{\text{modèle}}^2 = 1\%$
 - La situation d'un contrat d'un an avec un taux de crédit de 2% , projeté à 1 an, contribuerait positivement, toutes choses étant égales par ailleurs au score de rachat des contrats d'ancienneté 1 projetés à 1 an, figure 4.14 :

$$\mathbb{1}_{Tx \text{ crédit} \geq 1,5\%}(2\%) = 1$$

L'augmentation (resp. baisse) du taux résultant de la seule tendance du marché dans les situations ci-dessus induirait une augmentation (resp. diminution) du risque de rachat, ce qui apparaît incohérent. La solution proposée d'ajustement des seuils a pour objectif de « corriger » ces effets. Sans prise en compte d'évolution à 1 an, $Evo_{Tx\ prêt, modèle, 1}^1 = 0\%$, le retraitement des seuils $Seuil(Tx\ prêt)_{modèle, 1}^1$ dans chacune des situations précédentes - et illustré graphiquement ci-dessous, donne :

1. $Seuil(Tx\ prêt)_{modèle}^{1,1} = 1,5\% \frac{2\%(1+0\%)}{1\%} = 3\%$.

Ce qui aboutit cette fois à une contribution nulle au score des anciennetés 1 an projetées à un an (Figure 4.16, ci-dessous) :

$$\mathbb{1}_{Tx\ crédit \geq 3\%}(2\%) = 0$$

2. $Seuil(Tx\ prêt)_{modèle}^{1,1} = 2,5\% \frac{2\%(1+0\%)}{2\%} = 1,52\%$.

Ce qui aboutit cette fois à une contribution positive au score des anciennetés 1 an projetées à un an (Figure 4.17, ci-dessous) :

$$\mathbb{1}_{Tx\ crédit \geq 1,25\%}(2\%) = 1$$

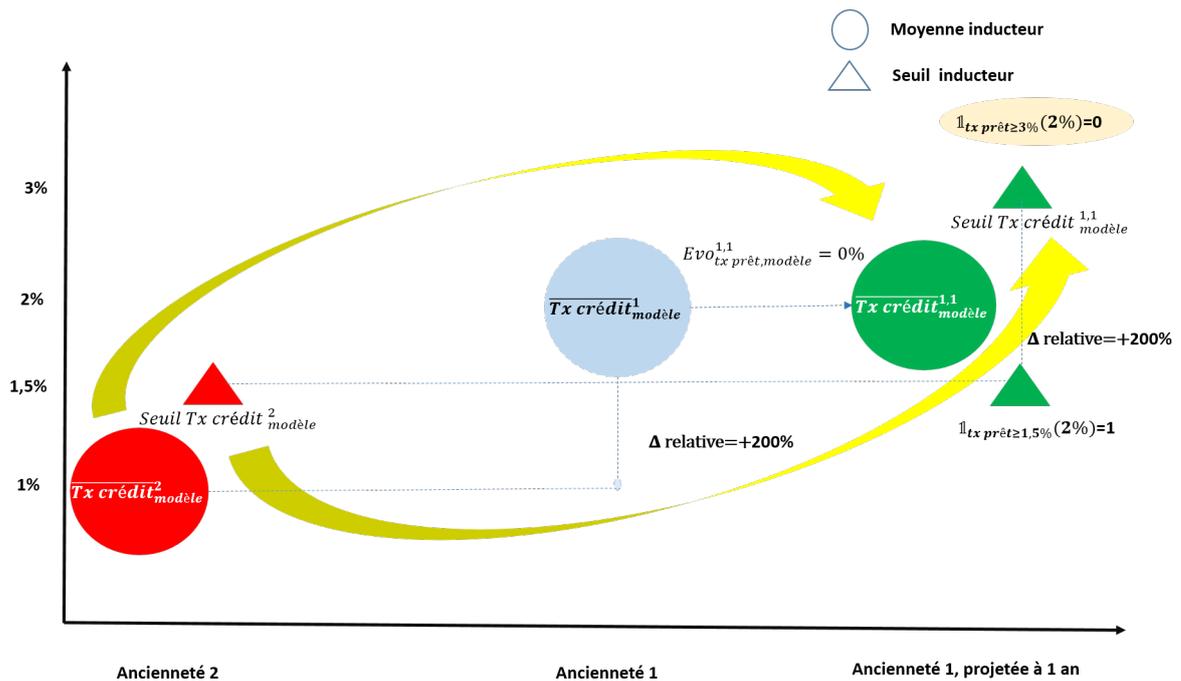


Fig. 4.16 – Ajustement seuil taux de crédit - situation haussière des taux (avec ajustement)

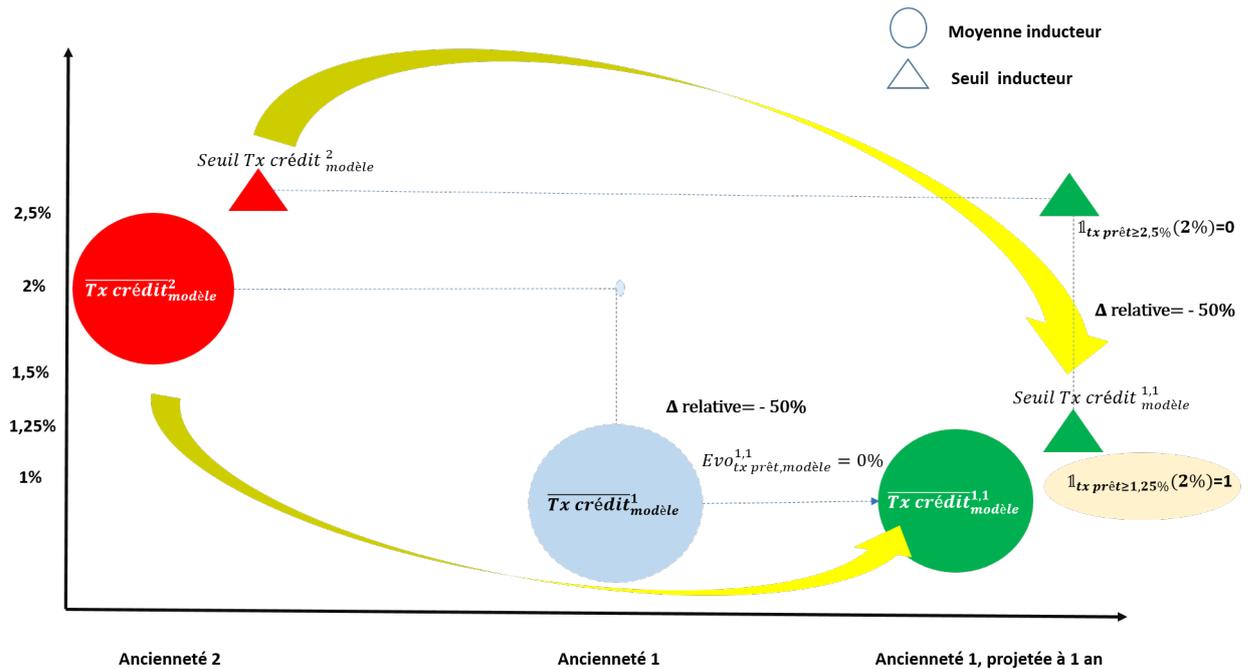


Fig. 4.17 – Ajustement seuil taux de crédit - situation baissière des taux (avec ajustement)

4.5.1.2 L’ajustement des seuils de test

L’objectif est ici de définir le seuil $Seuil(I)_{test,a_0}$ de l’inducteur I du score de l’ancienneté initiale a_0 , à partir du seuil $Seuil(I)_{modèle,a_0}$ de l’inducteur I du score de l’ancienneté a_0 , d’un estimateur $\overline{I}_{modèle,a_0}$ de l’inducteur I calculé sur les données du modèle et d’ancienneté a_0 et d’un estimateur de I, \overline{I}_{test,a_0} calculé sur les données de test et d’ancienneté a_0 .

L’ajustement consiste à faire évoluer les seuils des indicateurs de back-test comme les indicateurs moyens.

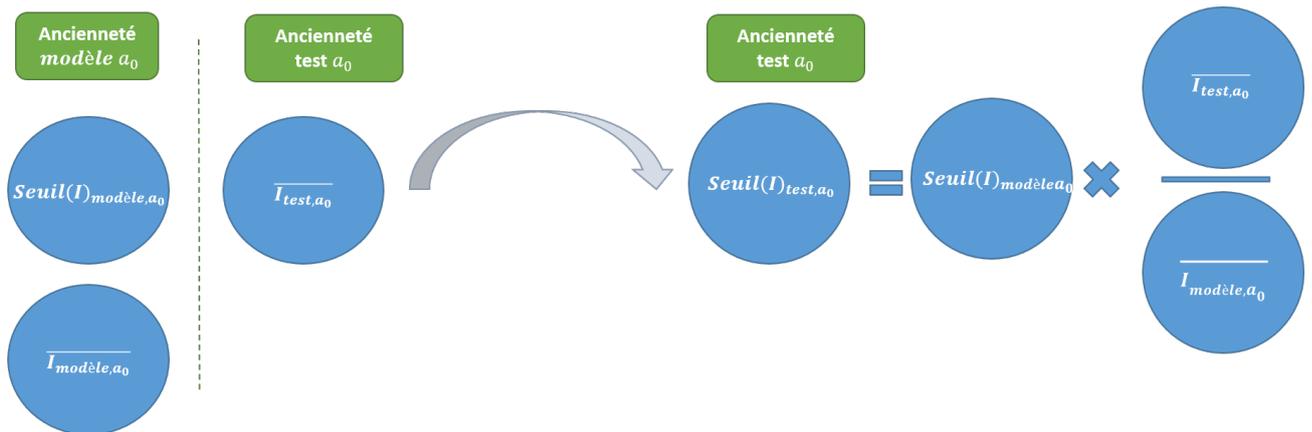


Fig. 4.18 – Principe du calage des seuils de test

Pour l’ancienneté initiale il nécessite donc :

- De calculer pour a_0 la moyenne arithmétique de l’inducteur I sur les données de la base de modélisation [B1] : $\overline{I}_{modèle,a_0}$
- De calculer pour a_0 la moyenne arithmétique de l’inducteur I sur les données de la base de back-test [B3] : \overline{I}_{test,a_0}

Le seuil ajusté de l'indicateur I du score d'ancienneté a_0 de test est défini ci-dessous, à partir du seuil de l'inducteur I du score d'ancienneté a_0 du modèle, tel que :

$$Seuil(I)_{test}^{a_0} = Seuil(I)_{modèle}^{a_0} \frac{\overline{I_{test,a_0}}}{\overline{I_{modèle,a_0}}}$$

Par exemple, à partir du paramétrage (Figure 7.41 en annexe), le facteur d'ajustement des seuils de back-test pour le taux de prêt d'ancienneté 3 ans est calculé comme suit : avec $a_0 = 3$, $\overline{I_{modèle,3}} = 2,73\%$, $\overline{I_{test,3}} = 1,84\%$, le facteur d'ajustement est tel que :

$$\frac{\overline{I_{test,3}}}{\overline{I_{modèle,3}}} = \frac{1,84\%}{2,73\%} = 67\%$$

De la même façon, l'ajustement des seuils de test pour le taux de prêts d'ancienneté 6 ans est obtenu comme suit :

$$\frac{\overline{I_{test,6}}}{\overline{I_{modèle,6}}} = \frac{2,44\%}{3,08\%} = 79\%$$

4.5.1.3 L'ajustement des seuils d'âges

Avec les notations qui précèdent, l'ajustement spécifique des seuils d'âge du modèle d'ancienneté a_0 projetés en n , $Seuil(\hat{Age})_{modèle}^{a_0,n}$ consiste à appliquer multiplicativement au seuil du modèle d'ancienneté $(a_0 + n)$ le rapport de l'âge moyen de l'ancienneté a_0 vieilli à n , $(\hat{Age}_{modèle,a_0} + n)$ sur l'âge moyen de l'ancienneté $(a_0 + n)$, $\hat{Age}_{modèle,a_0+n}$. Une déformation multiplicative de $+/-$ $Evo_{\hat{Age},modèle}^{a_0,n} \%$ peut être ajoutée pour moduler à travers l'âge le risque de résiliation :

$$Seuil(\hat{Age})_{modèle}^{a_0,n} = Seuil(\hat{Age})_{modèle}^{a_0+n} \frac{(\hat{Age}_{modèle,a_0} + n)(1 + Evo_{\hat{Age},modèle}^{a_0,n})}{\hat{Age}_{modèle,a_0+n}}$$

Les seuils d'âge de test peuvent être ajustés selon l'approche présentée au 4.5.1.2, ci-dessus.

4.5.2 Les transitions annuelles

Les transitions annuelles $T_{i \rightarrow j}^{a_0,n}$ de l'observation d'ancienneté initiale a_0 projetée en n , $n \leq 14 - a_0$ correspond au passage de l'état i à l'état j entre l'année $a_0 + n$ et $a_0 + n + 1$.

L'état i correspond à la classe de risque de l'observation résultant de l'application du score ajusté (défini précédemment) d'ancienneté $a_0 + n$. L'état j correspond à la classe de risque de l'observation « vieillie » résultant de l'application du score ajusté d'ancienneté $a_0 + n + 1$. Le vieillissement de l'observation consiste à incrémenter d'un an l'ancienneté du contrat et l'âge de l'assuré.

4.5.3 Les matrices de transition

Pour chaque ancienneté initiale a_0 projetée en n , $n \leq 14 - a_0$ et chaque état initial i défini par sa classe de risque, l'objectif est de définir un état de transition moyen \overline{trans}_i et une probabilité de rachat moyenne associée $Pr_{\overline{trans}_i}$.

Pour alléger le formalisme et sans perdre en généralité, la notation $a = a_0 + n$ est adoptée dans ce qui suit.

Avec les notations suivantes :

- Pour les anciennetés a , $1 \leq a \leq 15$.
- Pour r états de transitions possibles, correspondant aux r classes de risques que l'on se fixe.

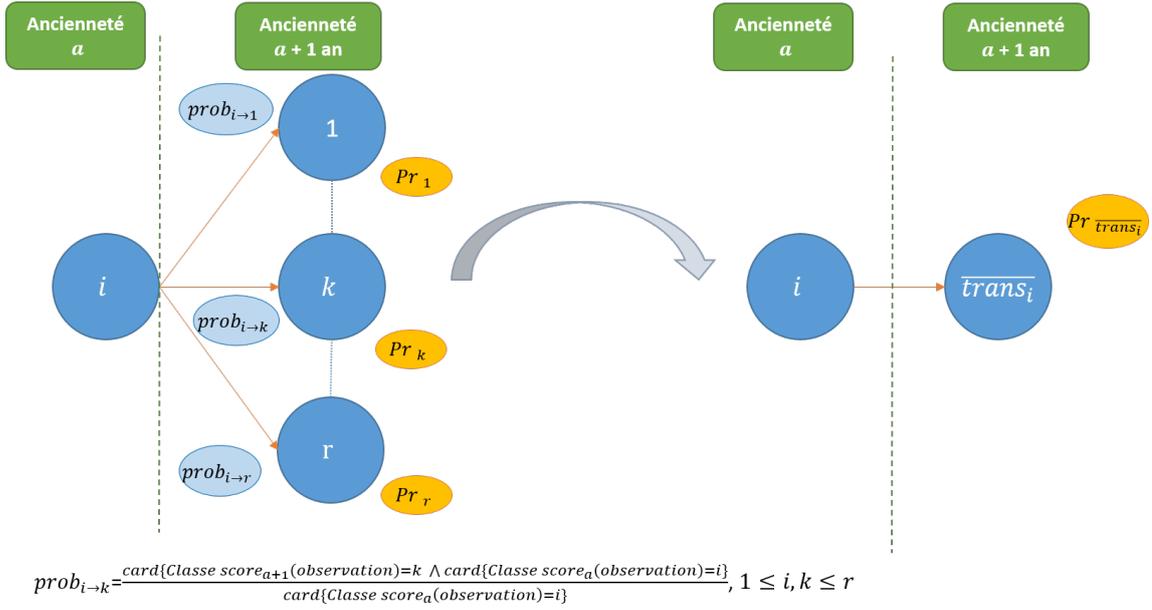


Fig. 4.19 – Présentation de la notion de transition moyenne

La matrice de transitions $M_{a \times r, 10}$, détaillée en annexe au 7.3.14, permet de définir, pour chaque ancienneté a et chaque classe de risque initial i :

- L'état en fin de période $a + 1$, ef_i^a :

$$|ef_i^a - \mathbb{E}[T_{i \rightarrow \cdot}^{a, a+1}]| = \text{Min}_{1 \leq k \leq r} |k - \mathbb{E}[T_{i \rightarrow \cdot}^{a, a+1}]|, 1 \leq ef_i^a \leq r$$

où $T_{i \rightarrow \cdot}^{a, a+1}$ est la variable aléatoire de transition partant de la classe de risque i en a .

- Le taux de rachat en fin de période $a + 1$, $txmf_i^{a+1}$:

$$txmf_i^{a+1} = \mathbb{E}[Txr_{i \rightarrow \cdot}^{a+1}]$$

où $Txr_{i \rightarrow \cdot}^{a+1}$ est la variable aléatoire de taux de rachat en $a + 1$ associée à la transition partant de la classe de risque i en a .

- L'effectif en fin de période $a + 1$, $Lxmf_i^{a+1}$:

$$Lxmf_i^{a+1} = \mathbb{E}[Lx_{i \rightarrow \cdot}^{a+1}]$$

où $Lx_{i \rightarrow \cdot}^{a+1}$ est la variable aléatoire d'effectif en $a + 1$ associée à la transition partant de la classe de risque i en a .

Une extraction de la matrice de transition pour l'ancienneté 1 projetée à 5 ans, pour $r=3$ classes de risque est donnée en annexe, figure 7.50.

4.5.4 Les trajectoires

Pour chaque ancienneté initiale a_0 et chaque état initial i défini par sa classe de risque, l'objectif est ici d'associer de manière itérative les transitions, pour finalement obtenir les projections des classes de risques ($a_0 + n$), jusqu'à l'ancienneté maximale de 15 ans. Pour $a = a_0 + n$, ci-dessous, figure 4.20 :

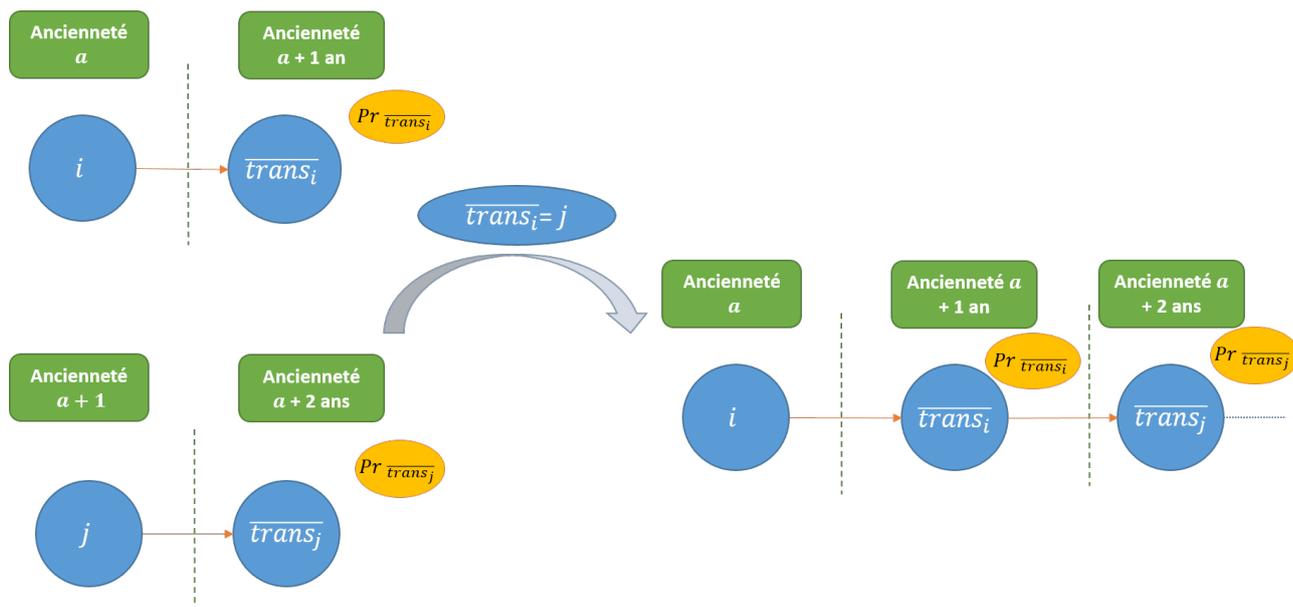


Fig. 4.20 – Présentation de la notion de trajectoire

En reprenant les notations ci-dessus, les trajectoires de changement d'états

$T_a = (C_{a_0,1}, \dots, C_{a_0,k}, \dots, C_{a_0,15})$ sont construites de manière itérative, comme suit, avec :

- Les anciennetés initiales $a_0, 1 \leq a_0 \leq 15$
- La classe de risque $i, 1 \leq i \leq r$
- La projection à l'année $p, a_0 \leq p \leq 15$

		Année de projection (p)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ancienneté (a)	1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$	$C_{1,6}$	$C_{1,7}$	$C_{1,8}$	$C_{1,9}$	$C_{1,10}$	$C_{1,11}$	$C_{1,12}$	$C_{1,13}$	$C_{1,14}$	$C_{1,15}$
	2		$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	$C_{2,5}$	$C_{2,6}$	$C_{2,7}$	$C_{2,8}$	$C_{2,9}$	$C_{2,10}$	$C_{2,11}$	$C_{2,12}$	$C_{2,13}$	$C_{2,14}$	$C_{2,15}$
	3			$C_{3,3}$	$C_{3,4}$	$C_{3,5}$	$C_{3,6}$	$C_{3,7}$	$C_{3,8}$	$C_{3,9}$	$C_{3,10}$	$C_{3,11}$	$C_{3,12}$	$C_{3,13}$	$C_{3,14}$	$C_{3,15}$
	4				$C_{4,4}$	$C_{4,5}$	$C_{4,6}$	$C_{4,7}$	$C_{4,8}$	$C_{4,9}$	$C_{4,10}$	$C_{4,11}$	$C_{4,12}$	$C_{4,13}$	$C_{4,14}$	$C_{4,15}$
	5					$C_{5,5}$	$C_{5,6}$	$C_{5,7}$	$C_{5,8}$	$C_{5,9}$	$C_{5,10}$	$C_{5,11}$	$C_{5,12}$	$C_{5,13}$	$C_{5,14}$	$C_{5,15}$
	6						$C_{6,6}$	$C_{6,7}$	$C_{6,8}$	$C_{6,9}$	$C_{6,10}$	$C_{6,11}$	$C_{6,12}$	$C_{6,13}$	$C_{6,14}$	$C_{6,15}$
	7							$C_{7,7}$	$C_{7,8}$	$C_{7,9}$	$C_{7,10}$	$C_{7,11}$	$C_{7,12}$	$C_{7,13}$	$C_{7,14}$	$C_{7,15}$
	8								$C_{8,8}$	$C_{8,9}$	$C_{8,10}$	$C_{8,11}$	$C_{8,12}$	$C_{8,13}$	$C_{8,14}$	$C_{8,15}$
	9									$C_{9,9}$	$C_{9,10}$	$C_{9,11}$	$C_{9,12}$	$C_{9,13}$	$C_{9,14}$	$C_{9,15}$
	10										$C_{10,10}$	$C_{10,11}$	$C_{10,12}$	$C_{10,13}$	$C_{10,14}$	$C_{10,15}$
	11											$C_{11,11}$	$C_{11,12}$	$C_{11,13}$	$C_{11,14}$	$C_{11,15}$
	12												$C_{12,12}$	$C_{12,13}$	$C_{12,14}$	$C_{12,15}$
	13													$C_{13,13}$	$C_{13,14}$	$C_{13,15}$
	14														$C_{14,14}$	$C_{14,15}$
	15															$C_{15,15}$

Fig. 4.21 – Matrice des classes de risques projetées

Avec :

- La classe de risque d'ancienneté a et d'année de projection p , $C_{a_0,p}$ définie telle que :

$$C_{a_0,p} = \begin{cases} 0 & \text{si } p < a_0 \\ i, 1 \leq i \leq r, & \text{si } p = a_0 \\ e f_{C_{a_0,p-1}}^{a_0+p} & \text{sinon} \end{cases}$$

- Le taux brut de rachat associé à la classe de risque $C_{a_0,p}$ défini tel que :

$$tx_{C_{a_0,p}}^p = \begin{cases} 0 & \text{si } p < a_0 \\ txr_i^{a_0}, 1 \leq i \leq r, & \text{si } p = a_0 \\ txmf_{C_{a_0,p-1}}^{a_0+p} & \text{sinon} \end{cases}$$

- L'effectif associé à la classe de risque $C_{a_0,p}$ défini tel que :

$$Lx_{C_{a_0,p}}^p = \begin{cases} 0 & \text{si } p < a_0 \\ Lx_i^{a_0}, 1 \leq i \leq r, & \text{si } p = a_0 \\ Lxmf_{C_{a_0,p-1}}^{a_0+p} & \text{sinon} \end{cases}$$

Chaque trajectoire est ainsi caractérisée par son ancienneté et la classe de risque initiale.

La mise en œuvre du procédé aboutit à $13 \times r + 2$ trajectoires au total, i.e. 41 trajectoires dans notre cas $r = 3$.

Par exemple, les 3 trajectoires d'ancienneté initiale et $a_0 = 9$ et $r = 3$ peuvent être illustrées comme suit, figure 4.22 :

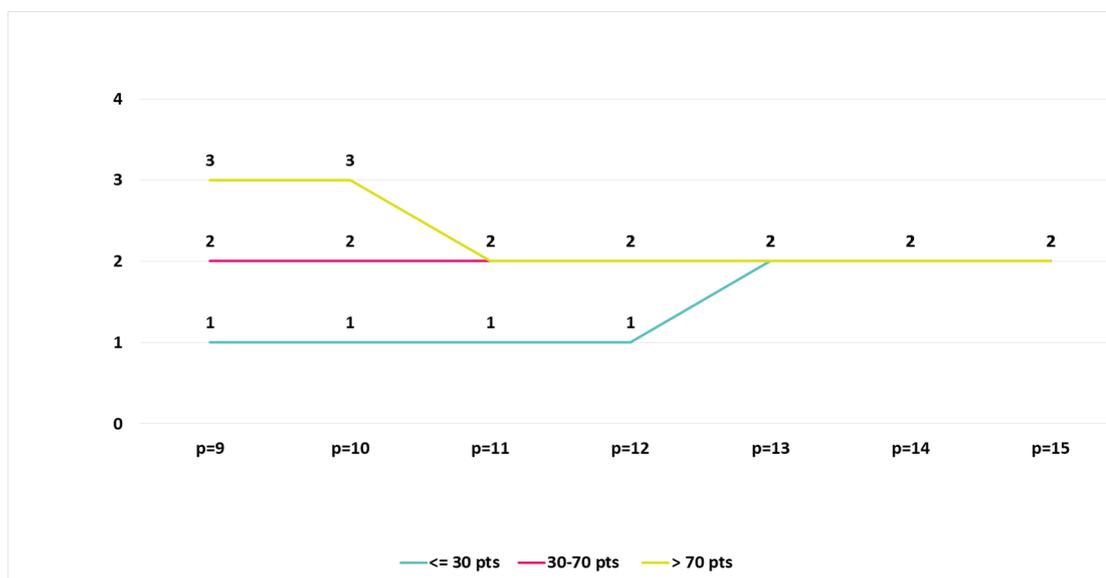


Fig. 4.22 – Les trajectoires pour $r = 3$ et $a_0 = 9$

- La classe de risque initiale 1 reste inchangée jusqu'à la 3^e année de projection, puis passe en classe 2 jusqu'au terme (15^e année)
- La classe de risque initiale 2 reste inchangée jusqu'au terme (la 15^e année)
- La classe de risque initiale 3 reste inchangée jusqu'à la 7^e année de projection, puis passe en classe 2 jusqu'au terme (15^e année)

Le détail des trajectoires par anciennetés pour $r = 3$ est disponible en annexe 7.3.13 - figure 7.48. Il permet de constater, pour l'ensemble des anciennetés, une convergence au cours des projections vers la classe de risque « médiane », la classe 2.

Pour optimiser le nombre de lois de rachats à paramétrer des groupes de trajectoires (voir 4.5.6) partageant le même paramétrage sont constitués. Ce regroupement nécessite pour une même classe de risque et une même ancienneté projetée, de partager la même probabilité de résiliation et le même effectif. Cette contrainte nous conduit à définir (voir 4.5.5) pour chaque ancienneté projetée et chaque classe de risque, à partir des probabilités de résiliation et effectifs des trajectoires :

- Une probabilité de résiliation moyenne.
- Un effectif moyen.

4.5.5 Les taux de rachats bruts et les effectifs moyens associés à chaque classe de risque $C_{a_0,p}$

En reprenant les notations ci-dessus, l'objectif est de définir les taux bruts de rachat et les effectifs moyens associés à chaque état de la trajectoire $(C_{a_0,a_0}, \dots, C_{a_0,k}, \dots, C_{a_0,15})$ ci-dessus, avec :

- Pour l'ancienneté $a_0 = p = 1$:

Ancienneté	Etat début période	Taux de rachat moyen de la classe i de l'ancienneté 1 et de l'année de projection 1
$a_0, a_0 = 1$	$i, 1 \leq i \leq r$	$txm_i^1 = tx_i^1$

Ancienneté	Etat début période	Effectif moyen de la classe i de l'ancienneté 1 et de l'année de projection 1
$a_0, a_0 = 1$	$i, 1 \leq i \leq r$	$Lxm_i^1 = Lx_i^1$

- Pour une ancienneté $a_0 > 1$ et $p \leq a_0 \leq 15$:

Ancienneté	Etat début période	Taux de rachat moyen de la classe $C_{a_0,p}$ de l'ancienneté a_0 et de l'année de projection p pondérée des effectifs
$a_0, a_0 > 1$	$C_{a_0,p}$	$txm_{C_{a_0,p}}^p = \frac{\sum_{k \leq p} Lx_{C_{k,p}}^p tx_{C_{k,p}}^p \mathbb{1}_{\{C_{k,p}=C_{a_0,p}\}}}{\sum_{k \leq p} Lx_{C_{k,p}}^p \mathbb{1}_{\{C_{k,p}=C_{a_0,p}\}}}$

Ancienneté	Etat début période	Effectif moyen de la classe $C_{a_0,p}$ de l'ancienneté a_0 et de l'année de projection p
$a_0, a_0 > 1$	$C_{a_0,p}$	$Lxm_{C_{a_0,p}}^p = \frac{\sum_{k \leq p} Lx_{C_{k,p}}^p \mathbb{1}_{\{C_{k,p}=C_{a_0,p}\}}}{\sum_{k \leq p} \mathbb{1}_{\{C_{k,p}=C_{a_0,p}\}}}$

Un extrait de la table des trajectoires, pour $r = 3$ et $p = 5$ et a_0 , est donnée en annexe, Figure 7.51.

4.5.6 Clustering

L'objectif de cette étape est de regrouper en clusters les trajectoires qui ont le même comportement et qui partageront le même paramétrage, figure 4.23.

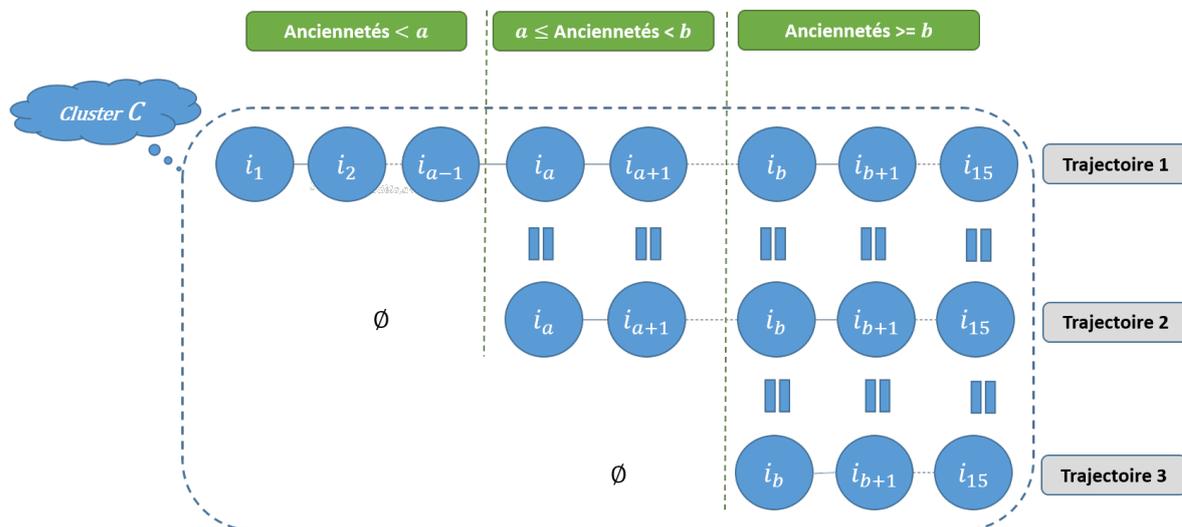


Fig. 4.23 – Présentation de la notion cluster de trajectoires

Les trajectoires de classes et taux associés sont regroupés comme suit

$T_a = (C_{a,1}, \dots, C_{a,p}, \dots, C_{a,15})_{1 \leq a \leq 15}$: Les trajectoires sont classées par anciennetés croissantes.

Par récurrence :

- Pour $a = 1$, $Cluster(T_1) = 1$
- Pour $a > 1$, $Cluster(T_k)_{1 \leq k \leq a} = clust_k$, $clust_k \leq k$
- Pour $a + 1$:

$$Cluster(T_{a+1}) = clust_{a+1} = \begin{cases} clust_k & \text{si } \exists k, 1 \leq k \leq a/C_{a+1} = R_{a+1}C_k \\ Max_{1 \leq k \leq a} \{clust_k\} + 1 & \text{sinon} \end{cases}$$

avec :

- $C_n = \sum_{1 \leq i < 15} C_{n,i} 10^i$
- $R_k C_n$ reste euclidien de C_n par 10^k , noté $C_n \equiv R_k C_n [10^k]$

Ainsi pour $r = 3$, 11 clusters (i.e. 11 lois) sont identifiés à partir de 41 trajectoires initiales. Les trajectoires regroupées dans le cluster 1 sont les suivantes, figure 4.24 :

AncADEGrp	Id_Trajectoire	trajectoire	cluster
1	1	111111111111222	1
6	16	000011111111222	1
7	19	000001111111222	1
8	22	000000111111222	1
9	25	000000011111222	1
10	28	000000001111222	1
11	31	000000000111222	1
12	34	000000000011222	1
13	38	00000000000222	1
14	40	00000000000022	1
15	41	00000000000002	1

Fig. 4.24 – Clusterisation des trajectoires cluster 1 ($r = 3$)

4.5.7 Le lissage des taux

Les taux bruts de rachats par cluster présentent certaines irrégularités. Il est indispensable de lisser les taux bruts afin d'absorber les variations dues aux fluctuations d'échantillonnage. 4 méthodes sont disponibles :

1. Les modélisations paramétriques.
2. Les lissages paramétriques.
3. Les lissages non-paramétriques.
4. Les modèles relationnels.

La méthode de Whittaker-Henderson (W-H) qui appartient à la troisième famille est privilégiée. Le principe de cette méthode est rappelé succinctement au 7.3.16 en annexe. Elle consiste à associer des valeurs lissées régulières et fidèles aux données brutes sans aucun paramètre à estimer, ni hypothèse sur la forme de la loi de rachat.

Dans les applications, $k = 3$ est retenu comme le degré de différenciation optimal, indépendamment de la courbe à lisser.

En outre, plus a est grand et plus le lissage est favorisé aux dépens de la fidélité aux données brutes, et inversement. Nous prendrons $k = 3$ qui est une valeur intermédiaire à notre objectif de lissage souhaité.

Les graphiques 7.52 et 7.53 en annexe restituent les taux lissés pour chacun des clusters identifiés dans l'étape précédente dans le cas d'une segmentation à 3 classes ($r = 3$) et à 4 classes de risques ($r = 4$).

Les taux par cluster et ancienneté qui seront paramétrés dans l'outil prophet sont dorénavant définis.

4.6 Construction des Model-Points emprunteur (MP)

L'application de la grille de score sur la base [BA3], agrégée à la maille $CR \times assuré \times contrat$ permet de disposer pour chaque observation de la base de :

- Son ancienneté.
- Son score initial.

L'objectif est d'affecter à chaque observation, son cluster qui permettra de connaître son taux de résiliation dans le paramétrage de l'outil prophet :

$$\text{Ancienneté} \times \text{cluster} \longrightarrow \text{taux de résiliation}$$

Le tableau de clusterisation des trajectoires (Figure 4.24), permet d'établir une correspondance entre l'ancienneté, la classe de risque initiale la trajectoire associée et le cluster correspondant :

Ancienneté	Score initial	Trajectoire	Cluster
$a_0, a_0 > 1$	C_{a_0, a_0}	$(C_{a_0, p})_{1 \leq p \leq 15}$	$Clust_{a_0}$

De cette correspondance est déduite une table qui à chaque ancienneté et score initial associe son cluster :

Ancienneté	Score initial	Cluster
$a_0, a_0 > 1$	C_{a_0, a_0}	$Clust_{a_0}$

Par exemple, pour $r = 3$, le paramétrage du cluster 1 sera utilisé par les lignes de Model-Points ayant les anciennetés initiales et les scores (nbpoints_Deb) suivants, figure 4.25 :

AncienneteADEGrp	nbpoints_Deb	cluster
1	1	1
4	2	1
6	2	1
8	2	1
9	2	1
11	2	1
12	2	1
13	2	1
14	2	1
15	2	1

Fig. 4.25 – Table de correspondance entre l’ancienneté x score et le cluster

Le cluster est affecté simplement à l’aide d’un format d’association ($\$FScore$) sous SAS⁶. Chaque observation du MP ADE utilisé par l’outil prophet est enrichie pour finir avec l’identifiant de cluster. Ainsi 2 jeux de MP ADE sont disponibles : un jeu correspondant à une segmentation en 3 classes de risques et un jeu correspondant à 4 classes de risques.

Le score du modèle (transition en 1^{re} année) est appliqué à la base de modélisation et au Model-Point de tests. L’analyse par classe de risque de la structure et de l’ancienneté moyenne résultant de l’application de ce score, permet de constater :

- Un transfert de la contribution des classes les plus risquées vers les classes les moins risquées, ci-dessous, figures 7.54 et 7.55, pour les segmentations à 3 et 4 classes de risques.
- Une diminution des anciennetés moyennes sur les classes de risques les plus faibles et les plus élevées, ci-dessous, figures 7.56 et 7.57, pour les segmentations à 3 et 4 classes de risques.

L’affectation des lois étant fonction de l’ancienneté initiale et du score en 1^{re} année, les clusters qui seront mis en œuvre dans les projections seront ceux correspondant aux trajectoires les moins risquées. Cependant la diminution des anciennetés sur les classes les plus élevées dans les mêmes proportions que celle constatée sur les classes les moins risquées devrait limiter la diminution attendue des rachats. Les résultats de la mise en œuvre du modèle seront étudiées dans la partie suivante.

4.7 Conclusion

Dans cette partie, une 1^{re} sous-partie a été consacrée à la réalisation d’une étude comportementale à 1 an des assurés emprunteurs des Caisses régionales. Cette étude a eu pour objectif d’identifier les facteurs déterminants des rachats et de construire des scores d’attrition par ancienneté. Ces scores ont conduit à définir des classes de risques par ancienneté.

6. $Cluster = put(cats(Anciennete, Score), \$FScore)$

Une 2^e sous-partie a été consacrée pour chaque ancienneté (a) à construire de manière itérative les projections de chaque classe de risque sur $(15 - a)$ années. Cette construction a nécessité de définir la notion de transition et à l'étendre à celle de trajectoire, la trajectoire matérialisant le résultat de la projection des classes de risques de chaque ancienneté. L'optimisation du nombre de trajectoires (et donc du nombre de lignes de paramétrage des taux de résiliations dans l'outil prophet) a conduit à introduire la notion de cluster comme regroupement de trajectoires.

Cette construction a nécessité de définir la notion d'ajustement des scores via les seuils des inducteurs les définissant. Le but de cet ajustement a été double :

- Mettre en cohérence les anciennetés projetées dans la construction avec celles de l'étude comportementale.
- Introduire des scénarii économiques dans la modélisation prospective, et en particulier de pouvoir mettre en œuvre le modèle dans un contexte différent de celui de la modélisation (e.g. dans un contexte haussier de taux de crédit actuel alors que la modélisation s'est faite dans un contexte baissier).

Pour chaque cluster, un taux brut a été associé à chaque classe de risque qui le définit. Un lissage des taux bruts de chaque cluster par la méthode whittaker-Henderson a permis de définir le paramétrage à paramétrer dans l'outil prophet.

Le score de 1^{re} année a été calculé sur la base agrégée de back-testing. L'identifiant du cluster associé à ce score a été ensuite injecté dans le Model-Point (MP) de back-test utilisé en entrée de prophet dans la partie suivante. Pour chaque ligne de MP, la connaissance de son ancienneté et de l'identifiant de cluster permet de récupérer le paramétrage des taux de rachat dans prophet tout au long de la projection.

La connaissance du risque futur de chaque ligne de MP projeté est ainsi intégralement contenue dans le cluster et est donc externalisée de prophet qui contient uniquement le paramétrage des taux par ancienneté et par cluster.

Chapitre 5

Mise en œuvre du modèle

L'objectif de cette partie est de mettre en œuvre l'approche vue précédemment dans un cadre Solvabilité 2. Après avoir rappelé le cadre de cette directive précisant la notion de Best Estimate Liability (BEL) et de Capital de Solvabilité Requis (SCR), le modèle de projection des passifs emprunteur incluant les rachats sera présenté. Enfin, les impacts des différentes approches de modélisation des rachats sur l'écoulement des effectifs, sur le BEL central et le SCR de rachat emprunteur sera analysé.

5.1 La directive solvabilité 2

Solvabilité 2 (S2) est une réforme européenne de la réglementation prudentielle s'appliquant au secteur de l'assurance dérivée de Bâle II pour les banques. La directive repose sur trois piliers ayant chacun un objectif :

- Le Pilier I : Exigences quantitatives en fonds propres et provisions techniques. A pour objectif de définir les exigences quantitatives, notamment en matière de fonds propres et de calculs des provisions techniques. Deux niveaux de fonds propres sont définis : le MCR (Minimum Capital Requirement ou Capital Minimum Requis) et le SCR (Solvency Capital Requirement ou Capital cible).
- Le Pilier II : Gouvernance des risques et exigences qualitatives Ce pilier définit les exigences qualitatives de suivi des risques interne aux sociétés en matière d'organisation et de gouvernance des organismes.
- Le Pilier III : Discipline de marché Ce pilier définit l'ensemble des exigences en matière d'informations prudentielles et de publication à destination des autorités de contrôle d'une part et au public d'autre part.

5.1.1 La Best Estimate Liability ou provision Best Estimate (BEL)

Le concept de Best Estimate, dit « meilleure estimation », se traduit par le passage d'une évaluation prudente des provisions techniques à une évaluation économique plus réaliste. La provision Best Estimate doit refléter au plus juste le profil de risque du portefeuille, elle doit être fiable et objective sans prise en compte d'une marge de sécurité.

Plus précisément, l'article 77 de la directive 2009/138/CE définit la provision Best Estimate comme « la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents ».

Pour calculer la provision Best Estimate, le portefeuille est supposé en run-off à partir de la date d'évaluation, c'est-à-dire qu'il est considéré comme un groupe fermé (aucun nouveau contrat ne peut s'ajouter) et les flux futurs sont projetés jusqu'à l'extinction des contrats du portefeuille. La cinquième étude quantitative d'impact (QIS5) précise que le BEL peut être calculé selon deux méthodes équivalentes :

- Méthode directe : calcul du BEL selon la définition de l'EIOPA (valeur actualisée des flux probables futurs liés au Run-off, ie sans prise en compte d'affaires nouvelles, des engagements au passif). Le Best Estimate est :

$$BE = \sum_t \frac{E_{\mathbb{Q}}(CF_t^{out} - CF_t^{in})}{(1 + r_t)^t}$$

Avec :

- \mathbb{Q} : Probabilité risque neutre
- CF_t^{out} : les cash-flows sortant à la date de projection t
- CF_t^{in} : les cash-flows entrant à la date de projection t
- Les flux à prendre en compte sont les flux associés aux contrats existants en portefeuille sur la période de projection retenue.
- La définition des flux à l'intérieur des frontières du contrat sont les suivants :
 - (+) Prestations à payer : Décès, rachats, maturités
 - (-) Primes à recevoir des assurés au titre des contrats en portefeuille à la date de calcul
 - (+) Frais sur primes et prestations
 - (+) Commissions des partenaires distributeurs
 - (+) Autres (prélèvement sociaux...)
- Méthode indirecte : calculer les autres éléments du passif et en déduire le BE à partir de l'égalité Actif-passif

5.1.2 Le capital de solvabilité requis (SCR)

Le SCR (ou Capital de Solvabilité Requis) est le capital requis aux organismes d'assurance pour garantir leur solvabilité. Il permet à l'organisme d'absorber les sinistres imprévus et correspond au montant des fonds propres à détenir pour limiter la probabilité de ruine à un an à 0,5%.

Le SCR peut être calculé de plusieurs façons différentes : soit par une formule standard calibrée uniformément sur le marché européen, soit par un modèle interne développé par l'assureur et après autorisation du superviseur, soit par une combinaison de ces deux méthodes. L'approche retenue pour le calcul est l'approche standard.

Pour déterminer le capital de solvabilité requis, la formule standard propose une architecture par modules et par sous-modules, ci-dessous - figure 5.1. Les modules reprennent la segmentation par LoB (Line of Business) des risques. Les risques de Marché, Santé, Vie et Non-Vie sont dissociés et forment chacun un module de risque, eux-mêmes subdivisés en sous-modules. Le calcul du SCR suit une approche dite bottom-up (de bas en haut) qui commence par le calcul des $SCR_{sous-module}$, se poursuit par celui des SCR_{module} pour aboutir à celui du BSCR (SCR de base ou Basic SCR). Le SCR global est déduit du BSCR.

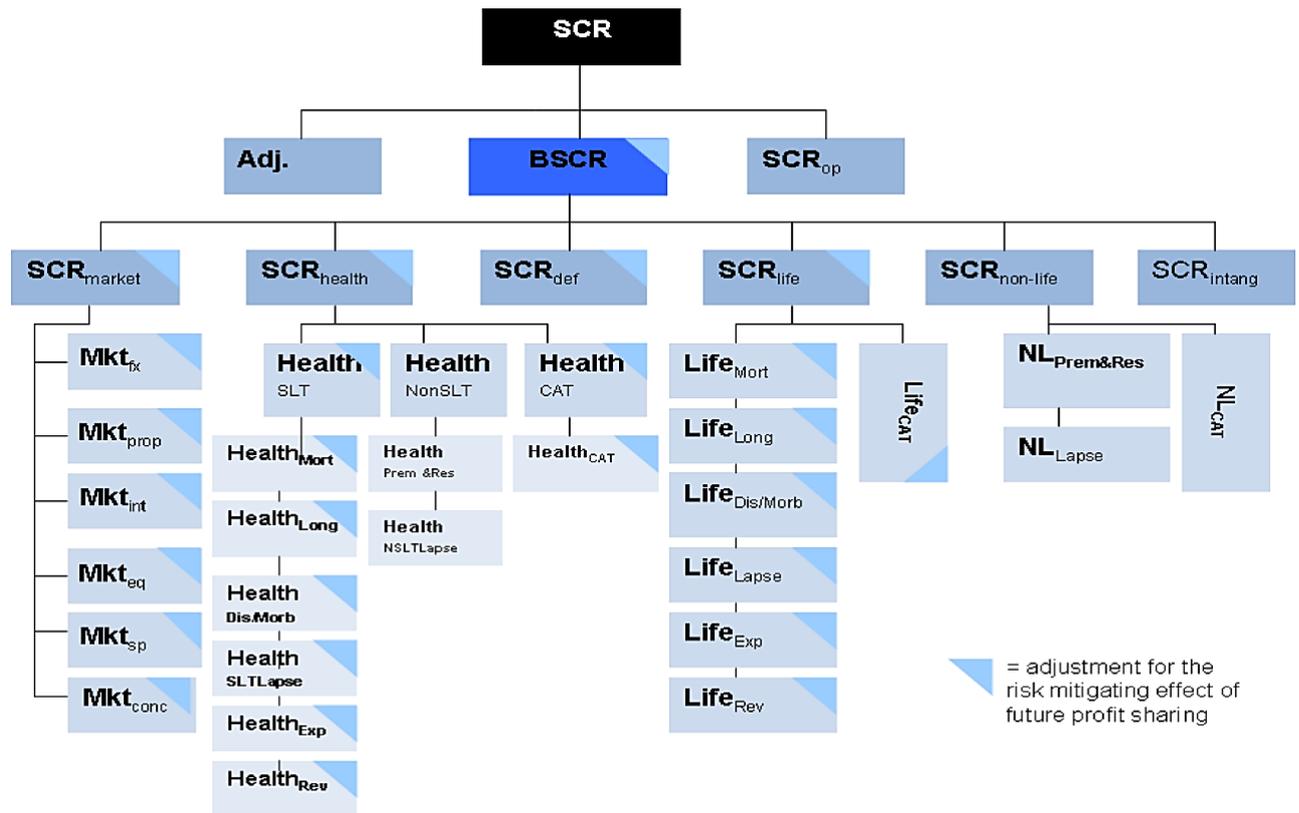


Fig. 5.1 – SCR - structure de la formule standard

5.1.2.1 Le SCR global

Selon l'architecture qui précède, le SCR est calculé selon la façon suivante :

$$SCR_{global} = BSCR + SCR_{opérationnel} - Adj$$

Le BSCR (Basic SCR) est le capital requis avant tout ajustement, il est déterminé par l'agrégation des besoins en capital pour les modules de risques suivants (les risques liés au capital exigé pour chaque module sont donnés dans la cartographie 7.58 en annexe :

- Le module Market (marché) : il intègre l'ensemble des risques de marchés.
- Le module Default.
- Le module Life (vie).
- Le module Health (santé) : il intègre les risques liés aux produits d'assurance santé.
- Le module Non life (non vie) : il est associé au risque de souscription des produits non vie.
- Le risque intang (intangible) mesure le risque lié aux actifs incorporels et intangibles.

5.1.2.2 Le SCR module

Le niveau de capital requis pour un module de risque est égal à l'agrégation des niveaux de capital requis pour chacun des chocs :

$$SCR_{module} = \sqrt{\sum_{i,j} corr_{i,j}^{module} SCR_{sous_module_i} SCR_{sous_module_j}}$$

Avec :

- $corr_{i,j}^{module}$: le coefficient de corrélation entre le sous-module i et le sous-module j , issu de la matrice de corrélation du module de risque
- $SCR_{sous_module_i}$: le capital requis pour le sous-module i
- $SCR_{sous_module_j}$: le capital requis pour le sous-module j

5.1.2.3 Le SCR sous-module

Le SCR de ces sous-modules se calcule alors par une approche dite delta-NAV, où la NAV correspond à l'actif net réévalué (Net Asset Value) :

$$SCR_{sous-module} = \Delta NAV$$

où :

$$\begin{aligned} \Delta NAV &= NAV^{Avant choc} - NAV^{Après choc} \\ &= (Actif^{Avant choc} - BE^{Avant choc}) - (Actif^{Après choc} - BE^{Après choc}) \end{aligned}$$

soit :

$$SCR_{sous-module} = \Delta NAV = \Delta Actif - \Delta BE$$

5.1.3 Zoom sur l'assurance emprunteur

L'assurance emprunteur, l'objet de cette étude, couvre plusieurs risques, elle se répartit sur différentes LoB, figure 5.2 :

- la branche Vie (Life) pour la garantie décès ;
- la branche Santé, assimilable à de la Vie (SLT Health) pour les garanties incapacité et invalidité ;
- la branche Non-Vie (Non-Life) pour la garantie perte d'emploi.

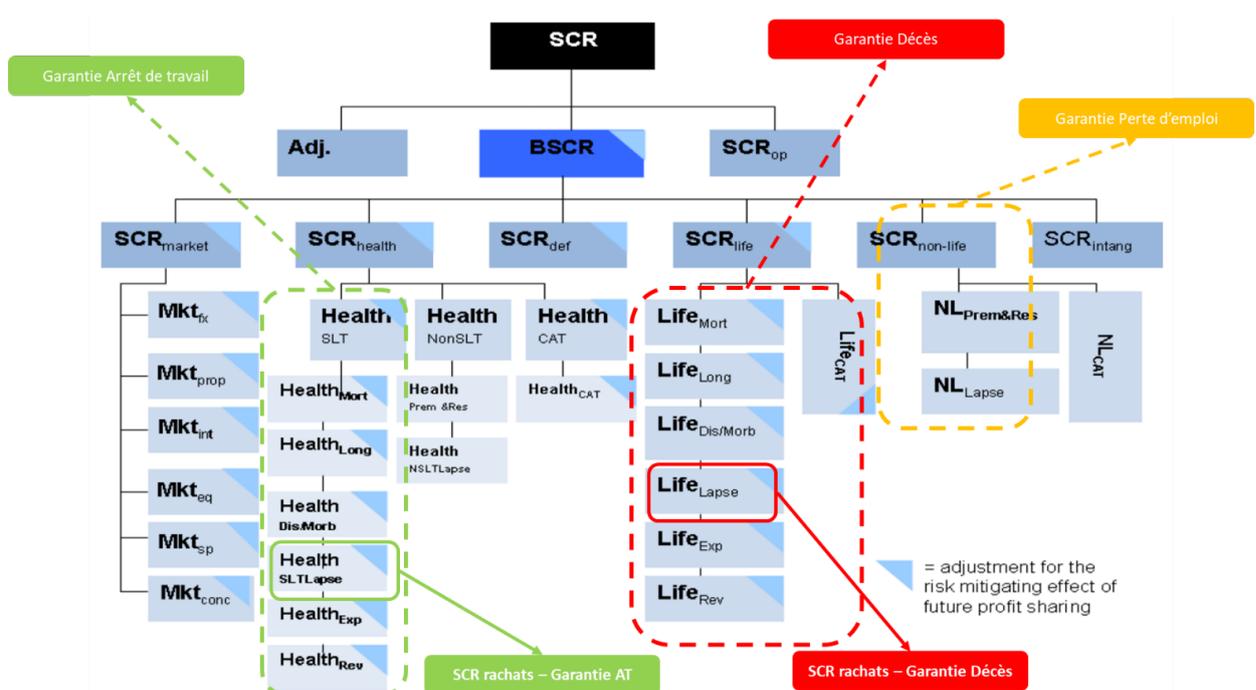


Fig. 5.2 – SCR - structure de la formule standard : branches d'activités concernées par l'assurance emprunteur

La section qui suit, après avoir présenté les principes de projection du modèle, analysera les différentes lois de rachats selon les 3 axes suivants :

- Les effectifs rachetés au cours du temps
- Le BE central pour la garantie Décès et pour la garantie Arrêt de travail
- Les SCR du sous-module rachat (coût du risque) de la garantie décès et de la garantie arrêt de travail

Les calculs de SCR de rachats se feront comme vu précédemment selon l'approche delta NAV ; les scénarios de choc ne concernant que le passif, la variation de l'actif suite à un choc étant nulle. Le SCR de rachats est tel que :

$$SCR_{Rachat}^{Garantie} = -\Delta BE_{Rachat}^{Garantie} = BE_{Garantie}^{Aprèschocrachat} - BE_{Garantie}^{Avantchocrachat}$$

Avec :

Garantie : garantie décès ou arrêt de travail

5.2 Le modèle prophète

Cette section présente la dynamique « par état » du modèle de projection, qui forme le socle des calculs implémentés à chaque pas de temps.

5.2.1 Cinématique et prise en compte des rachats dans le modèle

La modélisation du passif d'un produit d'assurance Emprunteur dans le modèle de projection est réalisée par une projection tête par tête des lignes de Model-Points de passif. Cette projection tête par tête fait appel à un modèle à états, qui se définit par :

- Une structure d'états :
 - Irréversibles :
 - Décès de l'assuré (état absorbant ¹)
 - Résiliation du contrat (état absorbant)
 - Arrivée à terme du contrat (état absorbant)
 - PTIA, étant assimilé à un décès, et donc regroupé avec le Décès (cf. ci-dessous : état absorbant)
 - Invalidité (ITD : non absorbant car peut déboucher sur deux autres états, l'arrivée à terme ou le décès)
 - Réversibles :
 - Valide
 - Incapacité (ITT)
- Une structure de transitions, qui correspondent à des probabilités de transition d'un état à l'autre, calculées en tenant compte des caractéristiques de l'assuré et de lois d'expérience.

1. Dans le cadre d'un modèle à état, un état i est dit absorbant si $p_{i,j} = \delta_{i,j}$ (i.e. $p_{i,j} = 0$ pour tout état j différent de i : une fois l'état atteint, la seule transition possible est un maintien dans l'état)

Les différents états possibles pour un assuré sont les suivants :

- Etats possibles pour un assuré « en bonne santé » en début de période, figure 5.3 :

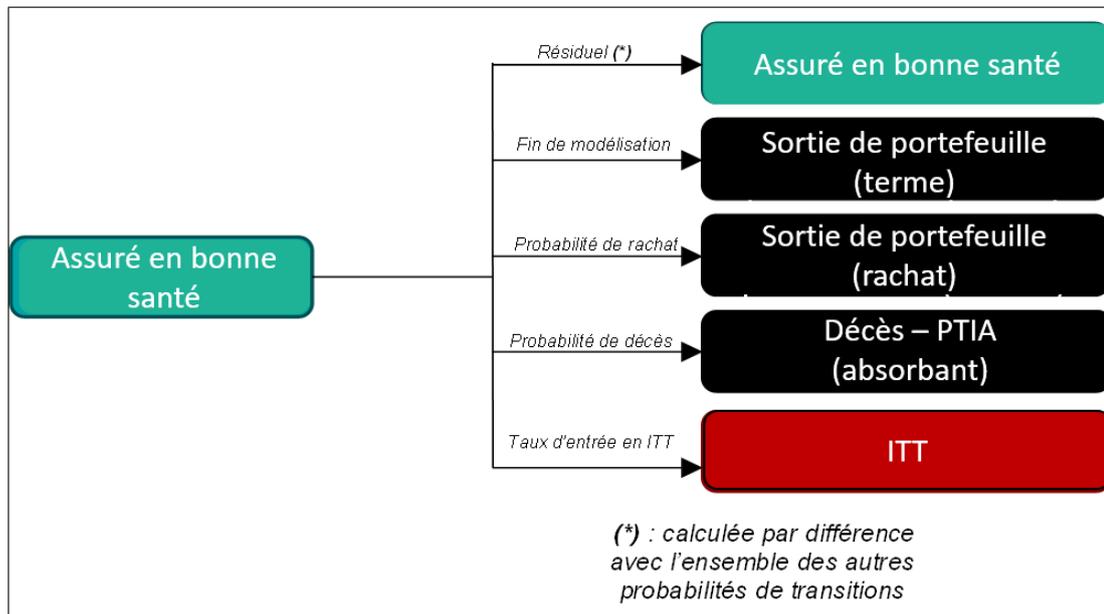


Fig. 5.3 – SCR - Transitions à partir de l'état valide

- Etats possibles pour un assuré en état d'Incapacité (ITT) en début de période (Figure 5.4) :

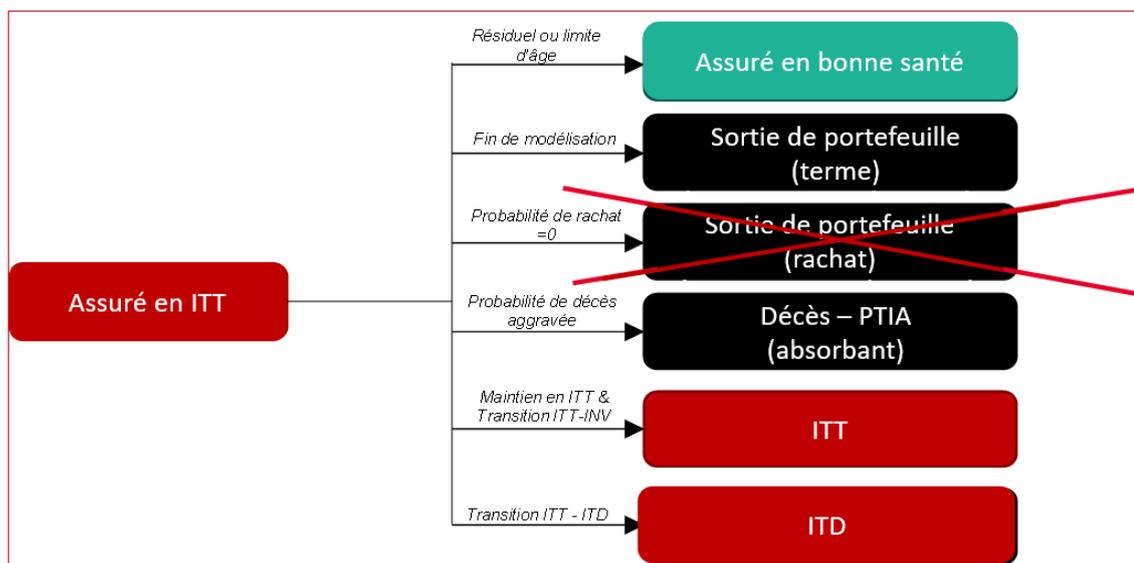


Fig. 5.4 – SCR - Transitions à partir de l'état ITT

- Etats possibles pour un assuré en état d'Incapacité (ITD) en début de période (Figure 5.5) :

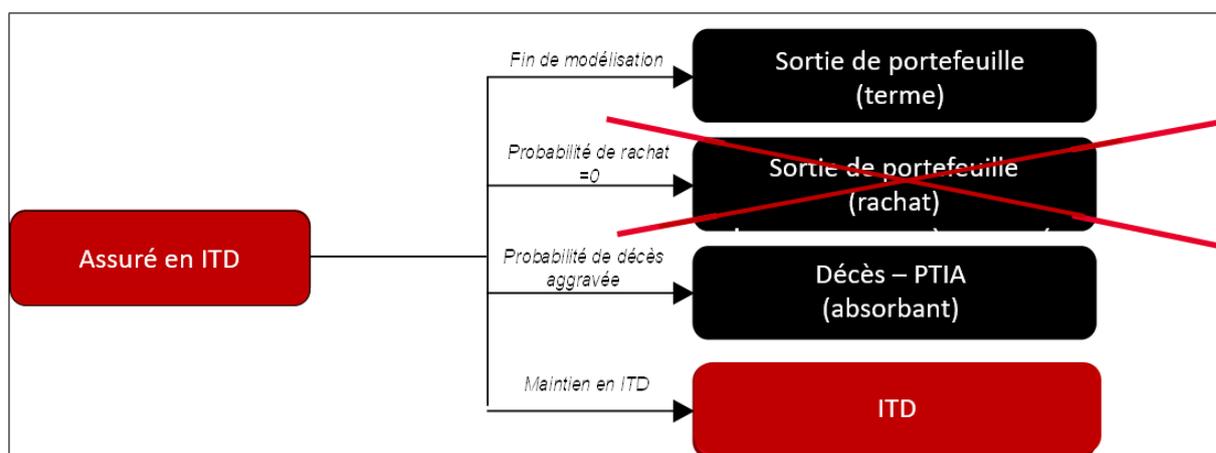


Fig. 5.5 – SCR - Transitions à partir de l'état ITD

Les résiliations ne sont pas autorisées pour les assurés en états d'incapacité et d'invalidité, tel qu'illustré ci-dessus ; Ce qui peut se résumer par le schéma suivant, figure 5.6 :

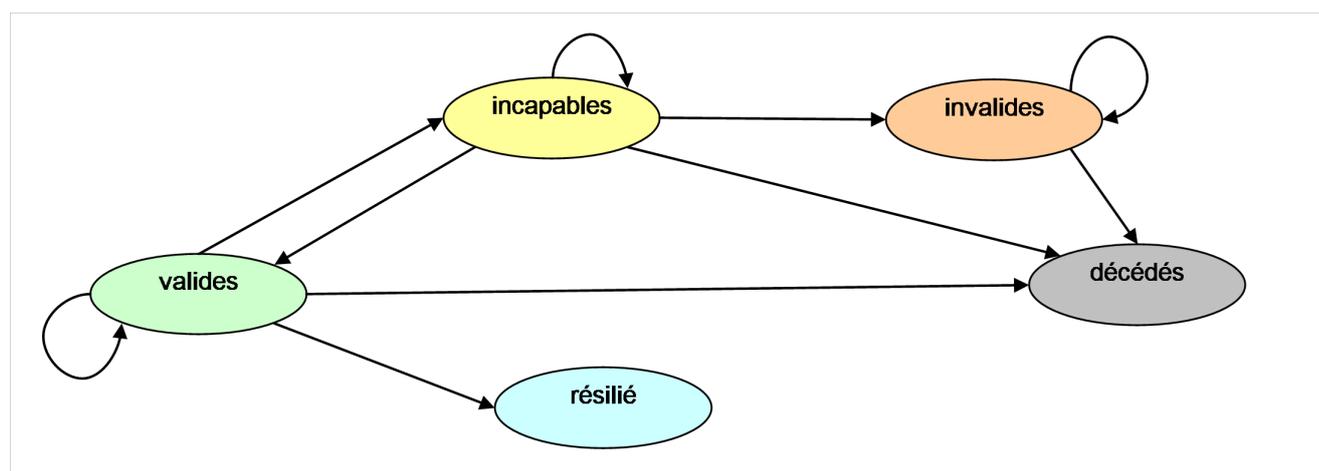


Fig. 5.6 – SCR - modèle prophète : graphe des changements d'état

Les projections sont ainsi réalisées sur deux types de population initiale, les adhésions en vigueur « non sinistrées » qui correspondent aux assurés en bonne santé, ou aux assurés « valides », et les adhésions en cours qui correspondent aux assurés en Incapacité (ITT) ou invalidité (ITD).

L'entrée en vigueur de la loi Bourquin donne la possibilité aux assurés de résilier leur contrat chaque année. Cette loi génère une augmentation des demandes (de la part des assurés) de renégociation du taux du contrat à la baisse. Ainsi, le modèle tient compte des taux de dérogations concédés lors de la projection des primes.

Le modèle est paramétré à l'aide de lois et hypothèses permettant de passer d'un état à l'autre. A ces lois sont appliqués des chocs permettant de calculer les Best Estimate avant et après chocs.

5.2.2 Chocs sur les rachats

Comme il a été vu précédemment, selon la formule standard du pilier 1 de Solvabilité 2, le risque de rachat est inclus dans le module de risque Life ainsi que dans le module de risque Health SLT. Pour les produits d'assurance emprunteur, il concerne ainsi la garantie décès et la garantie arrêt de travail.

- Le choc hausse des rachats consiste en une hausse permanente immédiate de 50% des taux de rachats par rapport au cas central. Le taux de rachat obtenu après choc ne peut toutefois pas dépasser 100
- Le choc baisse des rachats consiste en une baisse permanente immédiate de 50% des taux de rachats par rapport au cas central. La baisse du taux de rachat ne peut toutefois pas dépasser 20 points de pourcentage.
- Le choc rachats massifs consiste en une interruption immédiate de 40% des polices d'assurance pour lesquelles une interruption engendrerait une augmentation des provisions techniques hors marge de risque.

Le terme « immédiat » peut être interprété de plusieurs manières : sur la première période de projection ou sur la première année de projection. Il est ici supposé que le choc est appliqué sur la première période (i.e. le premier mois) et que les taux de rachats sont « normaux » et non-ajustés sur tous les mois suivants, même ceux de la première année de projection.

5.2.3 Principe de calcul des effectifs

Les taux de rachats sont paramétrés dans une table prophet par ancienneté de souscription et par identifiant de cluster (ID_CLUSTER), défini précédemment et permettant d'associer à chaque ligne de Model-Point sa loi de rachat, en fonction de l'ancienneté et du risque initial du contrat. La variable ID_CLUSTER a été ajoutée aux Model-Points en entrée de prophet. Le pas de projection de prophet étant mensuel, les taux annuels paramétrés doivent être mensualisés. Le taux mensuel de rachat pour l'ancienneté m en mois est calculé comme suit, par l'estimateur de Kaplan-Meier, à partir du taux annuel pour l'ancienneté a en années et le cluster c :

$$tx_rachat_mensuel_m^c = 1 - (1 - tx_rachat_annuel_a^c)^{1/12}$$

Pour $m, 12a - 1 \leq m < 12a$ et $1 \leq m \leq 360$ La prise en compte dans le modèle des chocs de rachats vus précédemment est présentée en annexe - 7.4.2.

L'effectif de rachat associé à la ligne n du Model-Point projeté à k mois est défini tel que :

$$Rachats_n^k = Valides_n^k \cdot tx_rachat_mensuel_{m_n+k}^{c_n}$$

où :

- c_n : le cluster associé à la ligne n
- m_n : l'ancienneté initiale de la ligne n
- $Valides_n^k$: l'effectif de la population valide associé à la ligne n , projeté en k . Cet effectif satisfait à la récurrence suivante :

$$\begin{aligned} Valides_n^k &= Valides_n^{k-1} - Rachats_n^{k-1} - Décès_n^{k-1} - Incap_n^{k-1} - Inval_n^{k-1} - Terme_n^{k-1} \\ &\quad + RétabIncap_n^{k-1} \end{aligned}$$

Avec :

- $Décès_n^{k-1}$: l'effectif des décès dans la période $[k-1, k[$
- $Incap_n^{k-1}$: l'effectif des entrées en incapacité dans la période $[k-1, k[$
- $Inval_n^{k-1}$: l'effectif entrée en invalidité dans la période $[k-1, k[$
- $Terme_n^{k-1}$: l'effectif din de contrats la période $[k-1, k[$
- $RétabIncap_n^{k-1}$: l'effectif des sorties d'incapacité dans la période $[k-1, k[$

On en déduit l'effectif de rachat total projeté à k mois est défini tel que tel que :

$$Rachats^k = \sum_{n \in \mathcal{N}} Rachats_n^k$$

Où \mathcal{N} est l'ensemble des lignes du Model Point projeté.

Après avoir analysé l'impact de la granularité de la segmentation sur les projections, l'étude présentera l'intérêt de la nouvelle approche par rapports aux lois statistiques classiques et s'intéressera enfin à la sensibilité des différentes lois de rachats aux scénariis de chocs de rachats S2.

5.3 Etude de la sensibilité des rachats projetés à la loi de modélisation utilisée

L'objectif de cette partie est d'évaluer la capacité des lois de rachat construites précédemment à s'ajuster automatiquement à l'évolution rapide du contexte et à intégrer dynamiquement des scénariis prospectifs. La mise en œuvre s'appuie sur le Model-Point du T4 2018 qui correspond à une vision du stock de contrats à décembre 2018 et qui est projeté sur 30 ans dans l'outil prophet. Les familles de contrats retenues dans ce stock sont celles des contrats utilisées dans la modélisation. Le Model-Point a été enrichi des identifiants de cluster de rachat qui permettent tout au long de la projection, d'associer chaque ligne de modèle point à la loi de rachat paramétrée.

5.3.1 Prérequis

Dans ce qui suit, la loi de rachats construite à partir de l'historique des rachats réels sur la période 2015-2017 sera dénommée loi historique ou historique brute.

Les lois segmentées à k ($k = 3$ ou 4) classes ou modalités (m) de risque et sans prise en compte de scénariis d'évolution économique seront dénommées lois segmentées centrales à k classes ou encore lois « centrales km ».

Les lois segmentées à k ($k = 3$ ou 4) classes ou modalités (m) de risque et avec prise en compte de scénariis d'évolution économique seront dénommées loi segmentées dynamiques à k classes ou encore lois « dynamiques km ».

Le qualificatif de loi segmentée recouvre aussi bien les lois segmentées dynamiques que les lois segmentées centrales ; enfin, les qualificatifs de loi segmentée centrale ou dynamique sous entendent une segmentation à 3 classes de risque. Le taux de rachats moyen pondéré par les effectifs sur 30 ans est défini comme suit :

$$\text{Taux rachat moyen pondéré} = \frac{\sum_{i=1}^{30} \text{Taux de rachat}_i \text{Effectif de contrats}_{i-1}}{\sum_{i=0}^{29} \text{Effectif de contrats}_i}$$

5.3.2 Analyse des projections des effectifs de contrats rachetés

L'objectif de cette partie est d'analyser les flux de rachats sur 30 ans résultant des projections des stocks, selon les lois historiques et segmentées. L'évolution est calculée sur une base 100, 100 étant l'effectif du stock initial. Aucun scénario économique prospectif n'est paramétré ici dans la construction des lois segmentées.

5.3.2.1 Impact de la granularité de la segmentation

Les projections considérées dans cette partie sont celles des effectifs de contrats rachetés sur 30 ans résultant d'une loi segmentée centrale à 4 classes de risques.

Il s'agit ici de vérifier si l'affinement de la granularité des classes de risques, en subdivisant la classe moyenne [30,70 points[en 2 sous-classes [40,50 points[et [50,70 points[a un impact sur modélisation des rachats. Le graphique ci-dessous, figure 5.7, illustre les résultats obtenus pour les lois centrales à 3 modalités (« centrale 3m ») et à 4 modalités (« centrale 4m »).

Il n'apparaît pas de différence significative induite par la granularité des classes de risques dans la construction des lois de rachats dynamique.

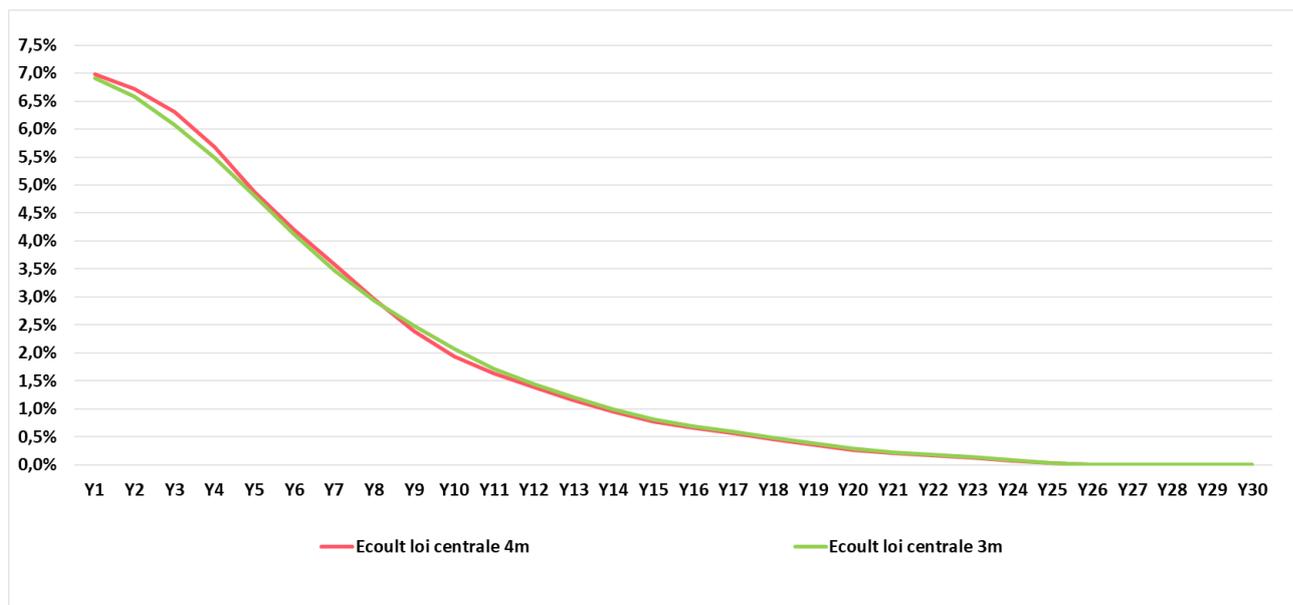


Fig. 5.7 – Scénario central - écoulement effectif rachats - approche segmentée « centrale 3m » vs « centrale 4m »

Le calcul du taux de rachats moyen pondéré par les effectifs pour les 2 lois confirme l'absence d'impact significatif de la granularité sur les taux de rachats, ci-dessous, figure 5.8 :

Taux rachat moyen pondéré loi centrale 3m	7,1%
Taux rachat moyen pondéré loi centrale 4m	7,2%

Fig. 5.8 – Taux de rachats moyen pondéré - loi segmentée « centrale 3m » vs « 4m »

Compte tenu du coût de paramétrage des lois basées sur une segmentation à 4 classes qui se traduit par le doublement du nombre de clusters et donc du nombre de lignes de paramétrage par rapport à une segmentation à 3 classes, la suite de l'étude se fera sur la base d'une segmentation à 3 classes.

5.3.2.2 Lois historiques et loi segmentée centrale à 3 modalités

Cette partie considère les projections 2019 sur 30 ans résultant de la mise en œuvre de la loi de rachats historiques brute (« histo. brute ») et de la loi segmentée centrale à 3 modalités (« centrale 3m »).

Les contrats présents dans le stock du T4 2018 sont suivis sur la période 2018-2021. Les effectifs de contrats réellement rachetés annuellement sur la période sont déduits des sinistres survenus et des termes des contrats.

Les effectifs sont rapportés à l'effectif de contrats initial pour travailler sur une base 100. L'effectif de contrats présents dans les model points fin 2018 et réellement rachetés à fin 2019 représente 5,9% de l'effectif initial.

L'évolution réelle constatée en 2020 et 2021 paraissant atypique (effectifs en hausse significative par rapport à 2019) en lien probable avec le contexte de COVID, seul l'effectif de rachats réel à 1 an est pris en compte dans l'analyse.

L'évolution du contexte économique entre la période de modélisation (2016-2017) et celle de la projection, (à partir de 2019), est marquée par un ralentissement de la baisse des taux de crédit, voir figure 7.4, et une stabilité, voire une légère hausse des tarifs ADE, voir figure 7.7. L'évolution de ces inducteurs structurants du modèle se traduit logiquement par une baisse des rachats, liée à une moins grande incitation des emprunteurs à résilier leur contrat. Le résultat attendu de la mise en œuvre de la loi segmentée doit refléter ce comportement des emprunteurs.

Les résultats sont présentés figure 5.9, ci-dessous. L'effectif réel à 1 an des contrats rachetés est superposé aux projections précédentes et est identifié par la marque (▲) sur le graphique

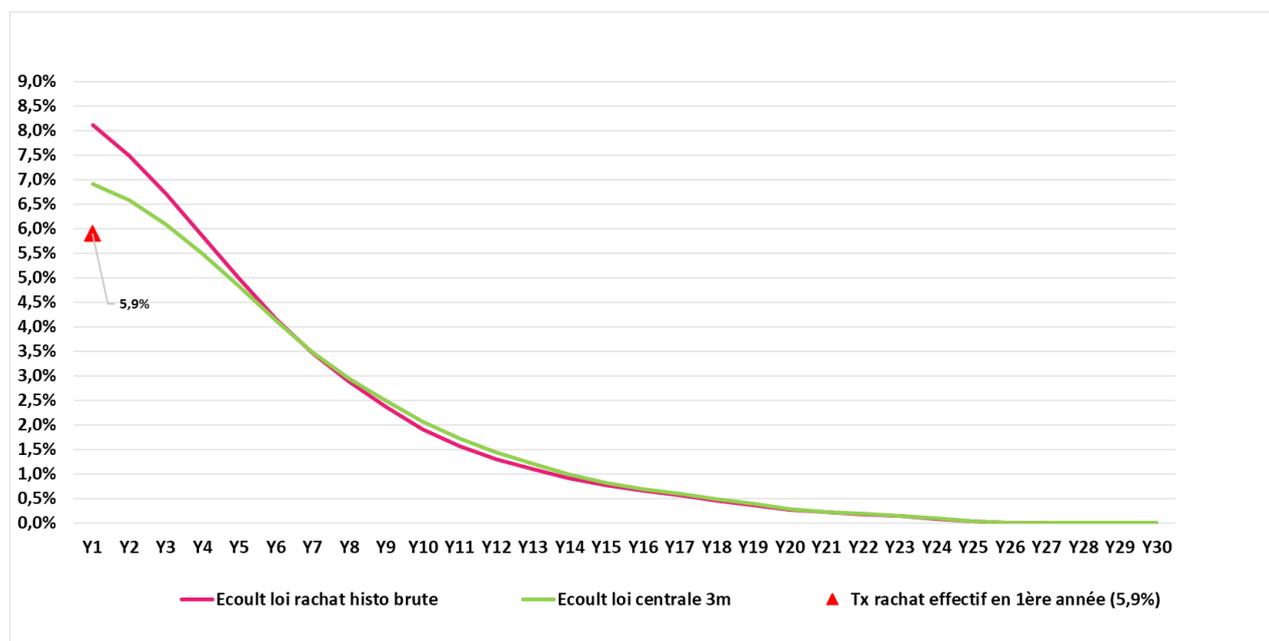


Fig. 5.9 – Scénario central - écoulement effectif rachat lois histo. vs approche segmentée « centrale »

Pour les 2 lois, les effectifs de contrat rachetés sont fonctions décroissantes strictes de la durée de projection. Ce point est expliqué par les sorties du portefeuille (en run off) qui ne sont pas amorties par d'éventuels mouvements à la hausse des taux de rachats.

La loi segmentée à 3 classes (« centrale 3m ») donne des niveaux de rachats globalement plus faibles que la loi historique brute et tend vers le niveau de rachat réel à 1 an. Ce constat est confirmé par le calcul du taux moyen pondéré, figure 5.10, ci-dessous :

Taux rachat moyen pondéré central 3m	7,1%
Taux rachat moyen pondéré histo. brute	7,7%

Fig. 5.10 – Taux de rachats moyen pondéré – loi segmenté « centrale 3m ». vs loi historique

Ce résultat est conforme à ce qui était attendu et reflète la capacité de la loi segmentée à s'ajuster au contexte et à limiter les écarts les écarts d'expérience en 1^{re} année. Cette propriété est intéressante pour une utilisation dans le cadre de la norme IFRS17 entrée en vigueur récemment.

5.3.2.3 Loi segmentée « centrale » et lois segmentées « dynamiques »

Cette partie s'intéresse aux projections 2019 sur 30 ans résultant de la mise en œuvre des lois segmentées centrale et dynamique de rachats à 3 modalités, en y intégrant différents scénarii économiques prospectifs.

Dans la pratique, l'intégration de scénarii économiques consiste à paramétrer la déformation souhaitée des seuils d'inducteurs qui entrent dans la modélisation des rachats (par exemple le taux de crédit, le taux tarifaire ADE ou l'âge de l'emprunteur dans ce qui suit), à lancer le calcul des clusters à partir du modèle ainsi que les lois de rachats associées et à affecter les nouveaux identifiants de cluster sur les lignes du Model-Point projeté en entrée de prophet.

Scénario 1 : scénario de hausse des taux de crédit (« Crédit Up »)

Le paramétrage de ce scénario consiste à augmenter les seuils des taux de crédit de 10% par an sur les 5 premières années de projection et à laisser inchangés les seuils des années suivantes.

Cette situation s'interprète comme une dégradation de contexte de marché du crédit pour les assurés qui ne parviennent pas à trouver des taux de crédit plus favorables que celui qui leur est appliqué et de fait ne sont pas incités à résilier leur contrat de prêt et donc ADE (du fait du crédit). Logiquement les niveaux de rachats attendus résultant de la loi dynamique devraient donc être inférieurs à ceux résultant de la loi centrale

Les résultats des projections sont présentés (Figure 5.11), ci-dessous :

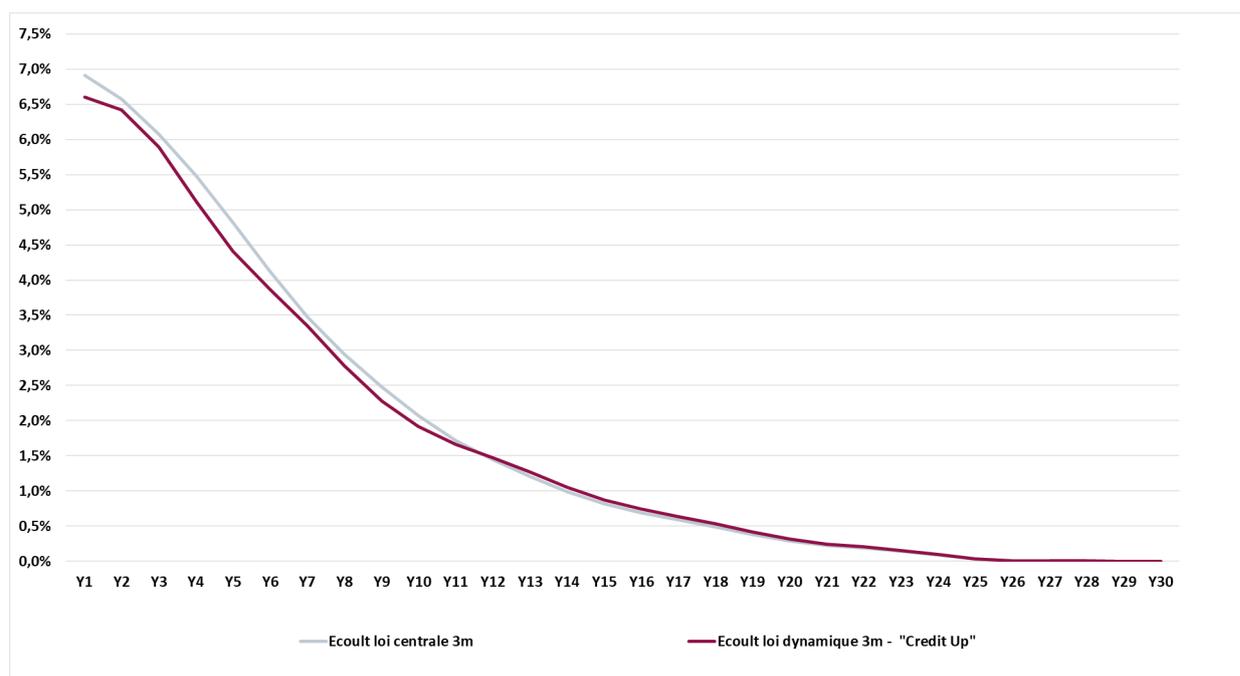


Fig. 5.11 – Scénario dynamique «Crédit Up» : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »

Il apparaît que la loi segmentée dynamique donne des niveaux de rachats globalement plus faibles que la loi segmentée centrale. Ce constat est confirmé par le calcul du taux moyen pondéré, figure 5.12 ci-dessous :

Taux rachat moyen pondéré loi "centrale 3m"	7,1%
Taux rachat moyen pondéré loi "dynamique 3m - Crédit Up"	6,6%

Fig. 5.12 – Taux de rachats moyen pondéré lois segmentées : « centrale 3m », vs « dynamique 3m Crédit Up»

Ce résultat est conforme à ce qui était attendu et reflète la sensibilité de la loi segmentée dynamique à l'intégration d'un scénario économique de hausse des taux.

Scénario 2 : scénario de baisse des taux de crédit (« Crédit Down »)

Le paramétrage de ce scénario consiste à diminuer les seuils des taux de crédit de 10% par an sur les 5 premières années de projection et à laisser inchangés les seuils des années suivantes.

Cette situation s'interprète comme une amélioration de contexte du marché du crédit caractérisé par une baisse des taux qui incite les assurés à résilier leur contrat de prêt et donc ADE pour aller les renégocier à la concurrence. Logiquement les niveaux de rachats attendus résultant de la loi dynamique devraient être inférieurs à ceux résultant de la loi centrale.

Les résultats des projections sont présentés figure 5.13, ci-dessous :

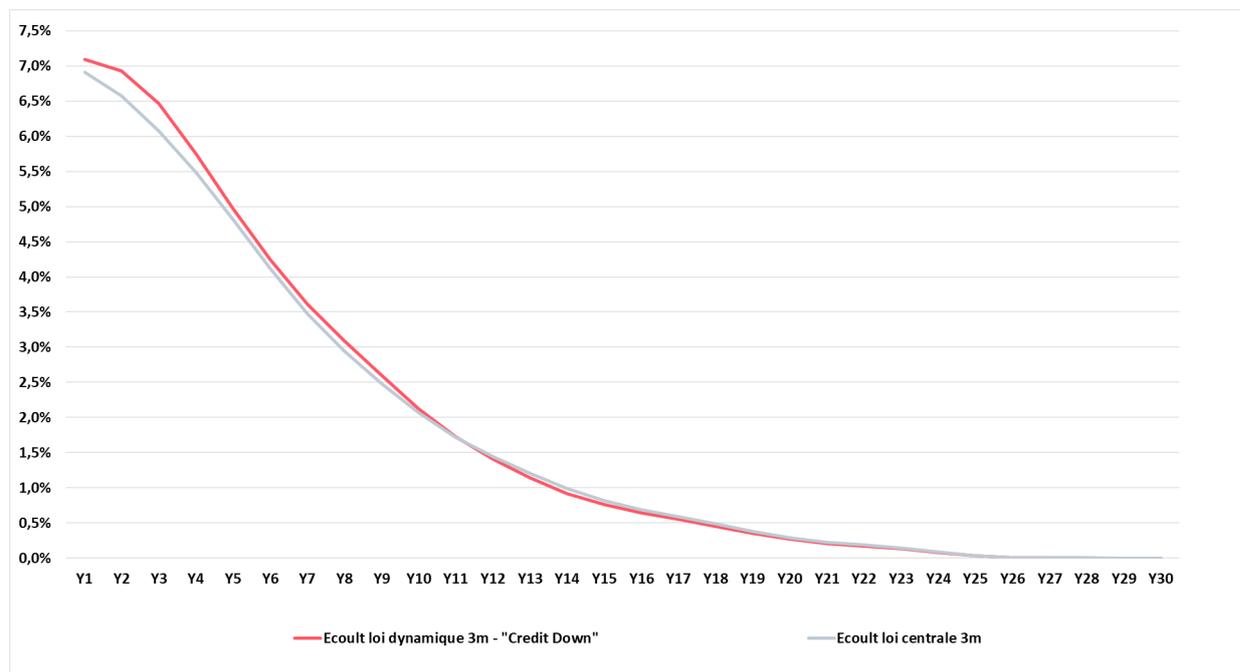


Fig. 5.13 – Scénario dynamique «Crédit Down » : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »

La loi segmentée dynamique donne des niveaux de rachats globalement plus élevés que la loi segmentée centrale. Ce constat est confirmé par le calcul du taux moyen pondéré, figure 5.14 ci-dessous :

Taux rachat moyen pondéré loi "centrale 3m"	7,1%
Taux rachat moyen pondéré loi "dynamique 3m - Crédit Down"	7,5%

Fig. 5.14 – Taux de rachats moyen pondéré lois segmentées : « centrale 3m ». vs « dynamique 3m Crédit Down»

Ce résultat est conforme à ce qui était attendu et reflète la sensibilité de la loi segmentée dynamique à l'intégration d'un scénario économique de baisse des taux.

Scénario 3 : scénario mix ADE/Crédit (« Mix ADE/Crédit »)

Le paramétrage de ce scénario consiste à augmenter les seuils des taux de crédit, à baisser les taux ADE et les âges des assurés de 10% par an sur les 5 premières années de projections, et à laisser inchangés les seuils des années suivantes.

Cette situation s'interprète à la fois comme une situation de marché crédit défavorable pour les assurés, caractérisée par des taux à la hausse et à une situation de marché ADE favorable avec des tarifs ADE plus faibles et un élargissement des segments de clientèles concernées (ciblage de clientèles plus âgées). Les inducteurs ADE (tarif et âge) incitent les assurés à résilier leur contrat ADE et entraînent une hausse des rachats. Inversement la hausse des taux de crédit entraîne une baisse des rachats en dissuadant les assurés de résilier leur contrat de prêt et donc ADE. Ce scénario correspond au contexte actuel, caractérisé par une volonté de compenser l'augmentation des taux de crédit liée à l'inflation par une facilitation d'accès au marché de l'ADE aux assurés et à une ouverture plus large à la concurrence.

Selon que le poids du crédit l'emporte sur l'ADE ou non, le résultat obtenu sera soit une augmentation soit une baisse globale du niveau des rachats par rapport à la loi segmentée centrale.

Les résultats des projections sont présentés figure 5.15, ci-dessous :

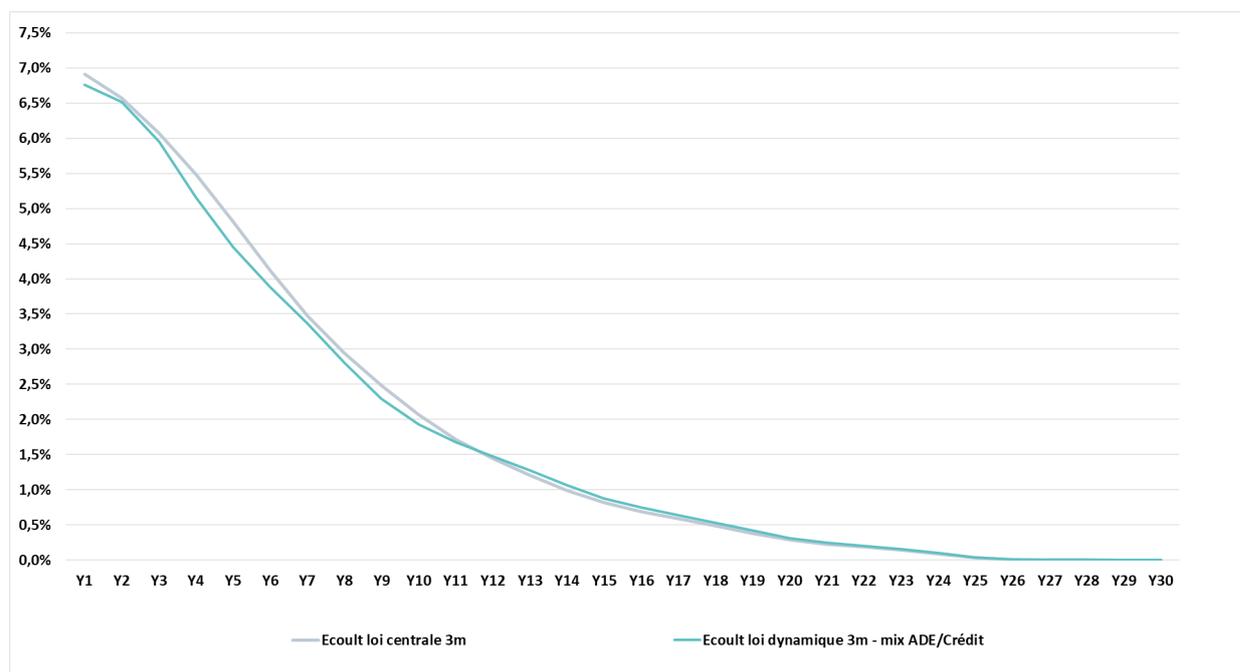


Fig. 5.15 – Scénario dynamique «mix ADE/Crédit » : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »

La loi segmentée dynamique donne des niveaux de rachats plus faibles que la loi segmentée centrale. Ce constat est confirmé par le calcul du taux moyen pondéré, figure 5.16 ci-dessous :

Taux rachat moyen pondéré loi "centrale 3m"	7,1%
Taux rachat moyen pondéré loi "dynamique 3m - mix ADE/Crédit"	6,7%

Fig. 5.16 – Taux de rachats moyen pondéré lois segmentées : « centrale 3m ». vs « dynamique 3m mix ADE/Crédit »

D'autre part, les niveaux de rachats obtenus sont légèrement inférieurs à ceux résultant de la mise en œuvre du scénario 1, « Crédit Up » présenté ci-dessus : 6,7% vs 6,6%.

Le résultat obtenu permet de conclure que l'impact de l'évolution des taux de crédit sur les rachats est supérieur à celui des inducteurs ADE sélectionnés dans ce scénario.

5.3.3 Impacts de l'approche alternative sur les effectifs de contrats et les flux financiers

Ce qui précède montre que la mise en œuvre de la loi segmentée centrale induit moins de rachats que la loi historique. L'objectif ici est d'étudier son impact sur l'écoulement des effectifs de contrats et en particulier sur ceux des sorties selon leur nature (décès ou terme). Comme précédemment, les effectifs sont exprimés en base 100 (100 étant l'effectif de contrats initial).

La loi segmentée centrale entraîne un écoulement sensiblement plus rapide des effectifs de polices, figure 5.17 ci-dessous :

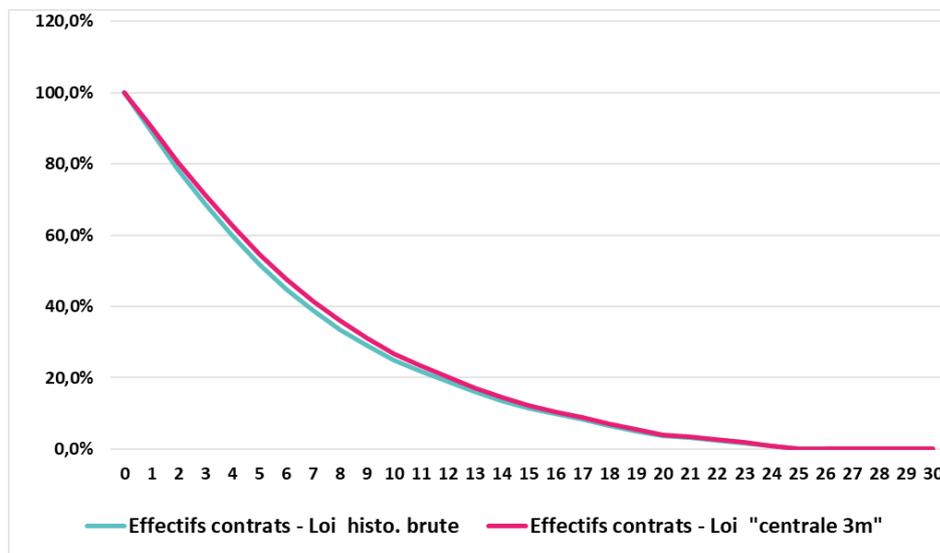


Fig. 5.17 – Ecoulement effectifs : flux de contrats

Concernant les sorties, il est à noter tout d’abord le poids marginal des décès par rapport aux rachats et aux termes figure 5.18, ci-dessous :

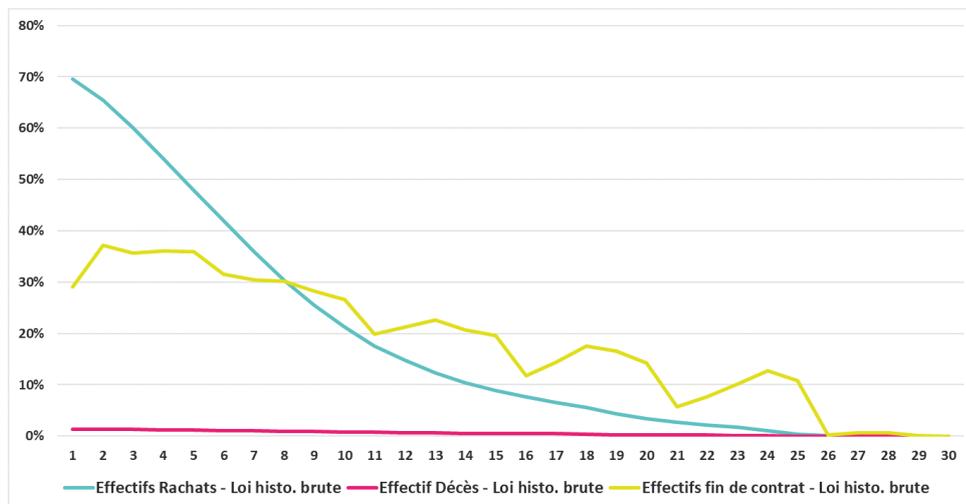


Fig. 5.18 – Ecoulement effectifs : poids des sorties par nature (%)

Il apparaît également de nombreuses arrivées à termes de crédits ou sorties d’âge en décès, aux années 11, 16 et 21, accentuées par un phénomène d’agrégations inhérent à la construction des Model-Points.

La comparaison des écoulements par nature de sorties résultant de la mise en œuvre des deux lois de rachats, permet de constater logiquement que la loi dynamique segmentée génère des écoulements moins rapides que la loi historique, ces derniers dépendant uniquement du nombre de contrats présents, figures 5.19 et 5.20, ci-dessous :

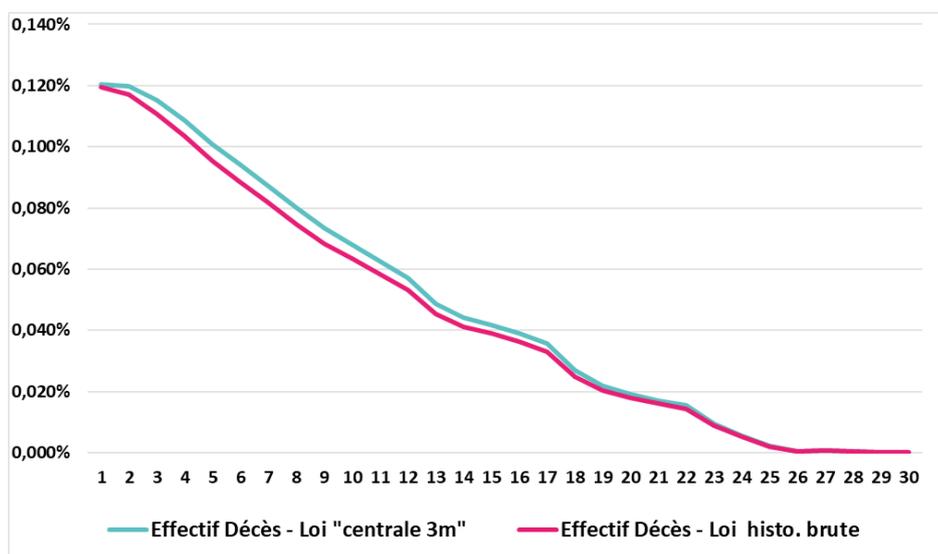


Fig. 5.19 – Ecoulement effectifs : flux de décès

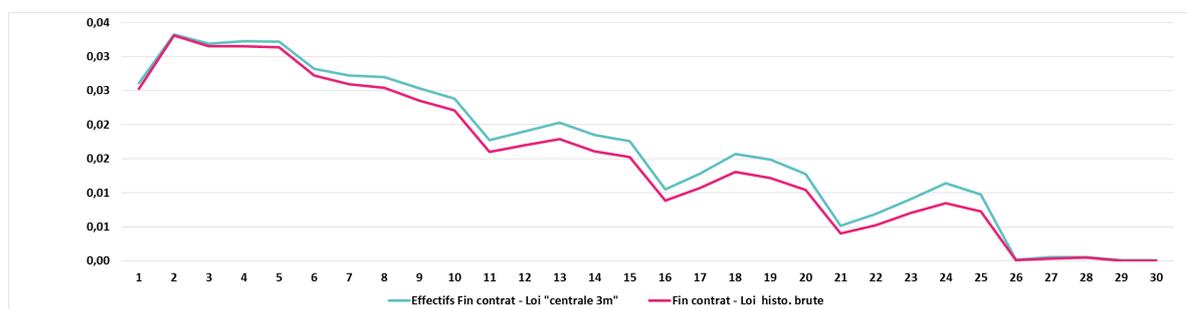


Fig. 5.20 – Ecoulement effectifs : flux des termes

5.3.4 Impacts de l'approche alternative sur les calculs de BEL et SCR de rachat des garanties décès et arrêt de travail

Le BE s'obtient comme suit :

$$BE = -P + S + fr + Com$$

A partir flux suivants actualisés au taux sans risques :

- P : le montant des primes (hors Taxes).
- S : le montant des prestations versées.
- fr : le montant des frais.
- Com : le montant de la rémunération sur flux du réseau et apporteurs d'affaires.

L'objectif est ici de comparer pour les scénarii de rachats centraux, les projections des composantes des comptes de résultats des garanties décès et arrêt de travail obtenues par l'approche historique et l'approche segmentée à 3 classes. L'étude est réalisée en base 100, 100 étant le montant de primes en 1^{re} année de projection du scénario central de rachat pour la garantie étudié (décès ou arrêt de travail)

Comme cela a été constaté dans ce qui précède, la loi dynamique génère moins de rachats que la loi historique. En conséquence les montants de primes décès résultant de la loi historique sont globalement inférieurs à ceux résultant de la loi dynamique, ci-dessous, figures 5.21 et 5.22 :

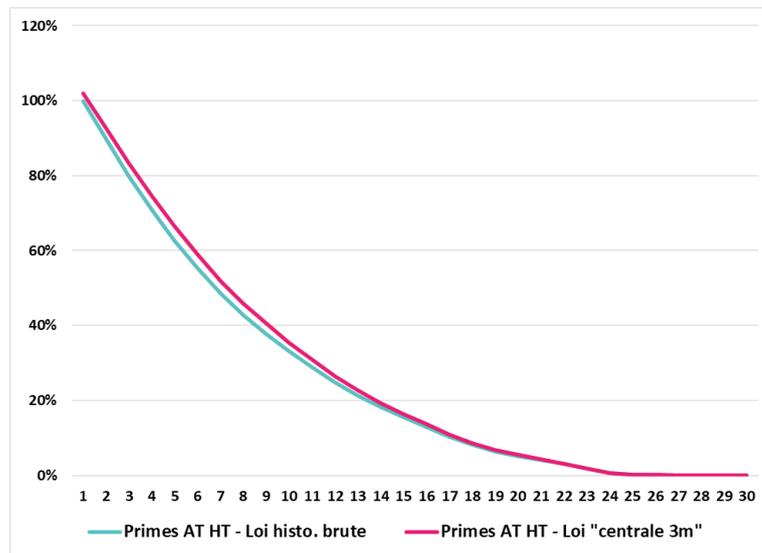


Fig. 5.21 – BEL/SCR Garantie décès : flux de primes - loi segmentée vs histo.

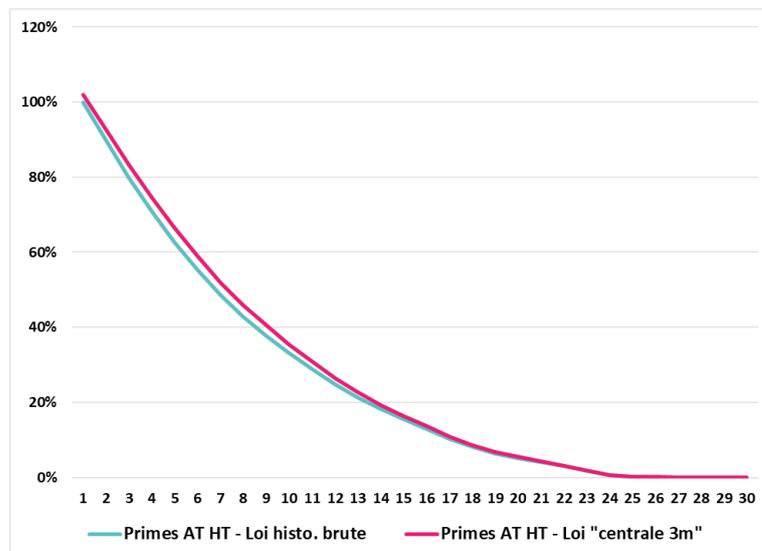


Fig. 5.22 – BEL/SCR Garantie arrêt de travail : flux de primes - loi segmentée vs histo.

Pour la même raison, les montants de prestations résultant de la loi historique sont globalement inférieurs à ceux résultant de la loi dynamique, ci-dessous, figures 5.23 et 5.24 :

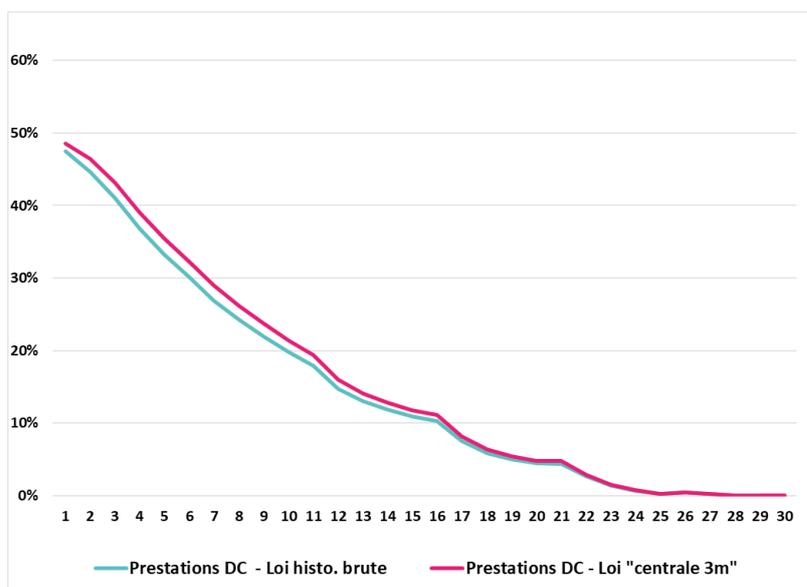


Fig. 5.23 – BEL/SCR Garantie décès : flux de prestations - loi segmentée vs histo.

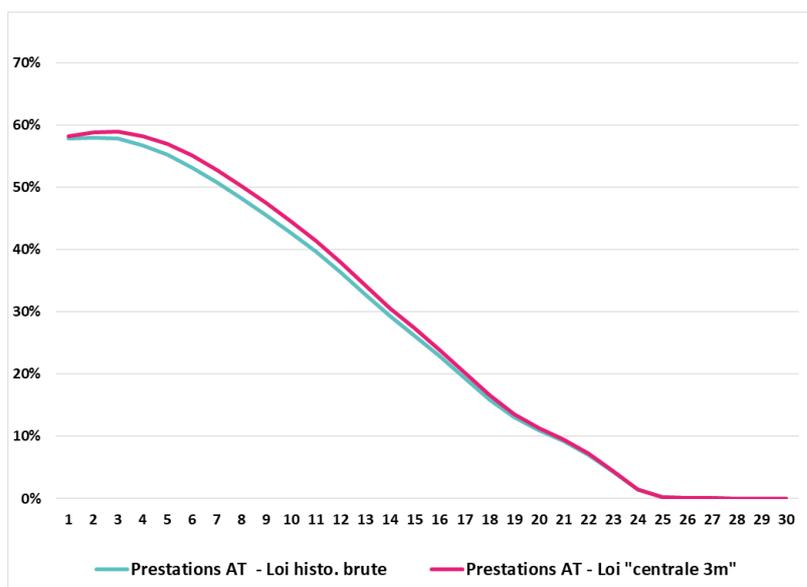


Fig. 5.24 – BEL/SCR Garantie arrêt de travail : flux de prestations - loi segmentée vs histo.

La prise en compte de cette nouvelle loi, entraîne uniquement une translation à la hausse de la sinistralité mais pas de déformation de cette dernière.

Pour la projection des prestations décès, il apparaît des diminutions de prestations aux années 12, 17 et 22. Ces diminutions sont liées à la structure des financements couverts dans le portefeuille, en lien avec les volumes de tombées de crédits aux années 11, 16 et 21, constatés précédemment. Le phénomène est atténué pour les prestations arrêt de travail, par les sorties d'âges en invalidité/incapacité.

Deux postes du BE sont directement liés aux primes : les frais et les commissions fixes. Par conséquent ces derniers évoluent comme ce que nous avons observés sur le chiffre d'affaire. Le produit Emprunteur crédit agricole est également basé sur un commissionnement variable dépendant du résultat technique global des risques décès et arrêt de travail.

L'analyse des flux, ci-dessous figure 5.25, montre un impact non significatif de la loi de rachat segmentée :

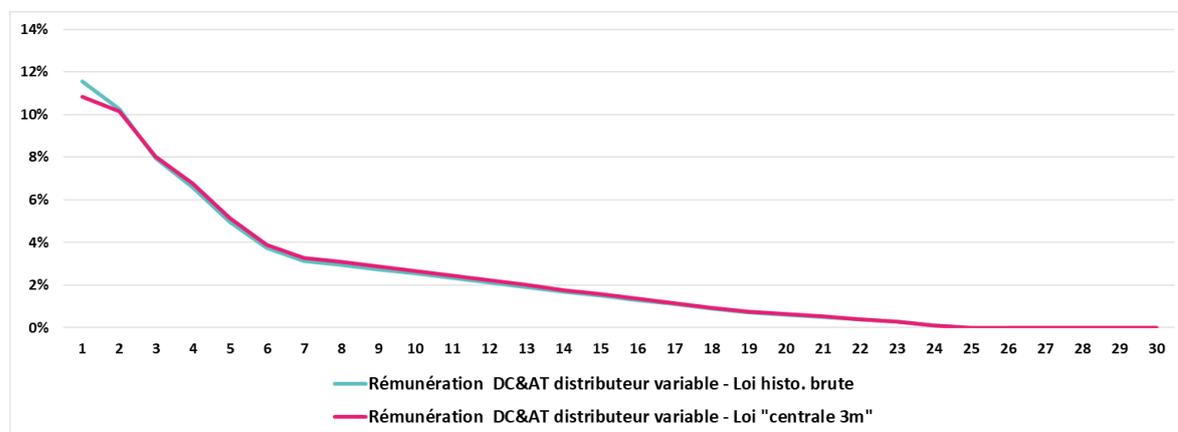


Fig. 5.25 – BEL/SCR Garantie décès : flux de commissions variables DC&AT loi segmentée vs histo.

En synthèse, l'analyse comparée dans le temps des composantes du BEL pour les lois segmentée centrale et historique ne fait pas apparaître d'écart significatif et laisse penser que BEL des deux lois devraient être assez proches. Les observations sont confirmées par les calculs des BEL centraux pour les 2 lois en base 100, synthétisés dans les tableaux 5.26 pour la garantie décès et 5.27 pour la garantie Arrêt de travail ci-dessous :

Base 100	central histo	central modele	Variation(%)
Primes HT	100,00	103,20	3%
Prestations	62,66	65,00	4%
Rémunération du réseau distributeur (fixe et variable)	41,03	41,81	2%
Frais	1,49	1,54	3%
BEL	5,18	5,15	-1%

Fig. 5.26 – Impact BEL/SCR - approche alternative - garantie décès

Base 100	central histo	central modele	Variation(%)
Primes HT	100,00	104,77	5%
Prestations	88,08	90,93	3%
Rémunération du réseau distributeur (fixe et variable)	45,43	47,03	4%
Frais	1,51	1,57	4%
BEL	35,02	34,76	-1%

Fig. 5.27 – Impact BEL/SCR - approche alternative - garantie arrêt de travail

Les tableaux font bien ressortir des niveaux de primes, prestations, commissions et frais moins importants pour la loi historique que pour la loi segmentée et des des BEL centraux équivalents pour les 2 lois (autour de 1% d'écart).

Afin d'avoir une vision SCR, les comptes de résultats obtenus par l'approche historique et segmentée centrale sont comparés pour les scénarii de rachats centraux et choqués. Cette analyse conduit à calculer les BEL de chaque loi dans les différentes situations, desquels seront déduits les SCR correspondants. L'étude est réalisée en base 100, 100 étant le montant de primes total du scénario central de rachat pour la garantie étudié (décès ou arrêt de travail)

L'impact pour la garantie décès de la mise en œuvre de la loi segmentée centrale (modèle) par rapport à la loi historique est synthétisé dans le tableau 5.28 :

	central histo	rachat massif histo	choc hausse rachat histo	choc baisse rachat histo	central modele	rachat massif modele	choc hausse rachat modele	choc baisse modele
Base 100								
Primes HT	100,00	60,63	78,08	131,21	103,20	63,42	83,00	134,77
Prestations	62,66	41,35	50,01	80,99	65,00	43,28	53,42	83,49
Rémunération du réseau distributeur (fixe et variable)	41,03	23,97	32,48	52,93	41,81	24,80	33,92	54,01
Frais	1,49	0,89	1,16	1,96	1,54	0,93	1,23	2,01
BEL	5,18	5,58	5,57	4,68	5,15	5,60	5,58	4,74
coût du choc		0,40	0,39	0,00		0,45	0,43	0,00

Fig. 5.28 – Impact BEL/SCR - approche alternative - garantie décès

Pour la loi historique comme pour la loi segmentée, le coût de risque maximum est obtenu pour le choc de rachat massif.

Enfin, le tableau ci-dessus permet également de constater que la loi dynamique segmentée est plus sensible aux chocs de rachats que la loi historique et contribue là encore une mobilisation plus importante de fonds propres. Le SCR résultant de la mise en œuvre de la loi dynamique segmentée est plus élevé de 9% que celui obtenu à partir de la loi historique.

5.4 Conclusion

Dans cette partie, la loi de rachats dynamiques a été analysée dans un cadre solvabilité 2, en la comparant aux lois historiques construites comme moyennes sur les années 2015-2017 des rachats observés. Pour ce travail, un stock de contrats immobiliers de décembre 2018, a été projeté dans l'outil prophet en paramétrant les lois de rachats souhaitées. L'étude a consisté en une analyse des projections des effectifs des contrats, puis des flux financiers et enfin des BEL et coûts du risque.

La loi segmentée à 3 classes de risques a été mise en œuvre dans un 1^{er} temps sans y intégrer de scénario d'évolution de contexte économique (scénario central). Les flux de rachats induits ont été comparés à ceux résultant des lois de rachats historiques, ainsi qu'au flux réel de rachats constatés à un an (contrats du stock de 2018 vus en 2019). Les effectifs résultant de la loi dynamique se rapprochent davantage du réel que ceux obtenus avec la loi historique brute. L'affinement de la granularité de la segmentation (passage de 3 à 4 classes de risque) est sans impact significatif.

La loi segmentée à 3 classes de risques a ensuite été mise en œuvre en intégrant des scénarii d'évolutions économiques mono ou multifactoriels, impliquant des indicateurs crédit (taux de crédit) ou ADE (âges des assurés ou taux ADE) : les tests ont mis en évidence le potentiel dynamique et la sensibilité de la loi segmentée aux contextes.

Enfin le Best Estimate (BEL) et le capital requis (SCR) – pour les garanties décès et arrêt de travail, résultant de la loi segmentée ont été comparés à ceux résultant de la loi historique. Les BEL sont globalement assez proches. En revanche, la loi segmentée dynamique est un peu plus exigeante en fonds propres que la loi historique avec des SCR plus élevés.

Chapitre 6

Conclusion générale

Depuis quelques années l'assurance emprunteur connaît de nombreux changements réglementaires et les assurés sont de plus en plus incités à comparer leur contrat ADE au marché. La modélisation retrospective actuelle des risques de rachats (présentée au 2.5.4) ne permet pas de capter et d'intégrer directement l'impact de l'évolution de contexte dans les projections futures du portefeuille emprunteur. Elle nécessite des mises à jours fréquentes sous peine d'occasionner des écarts d'expérience préjudiciables et requiert des coefficients d'ajustement prospectifs, calculés à partir de données externes.

L'objectif de l'étude présentée ici, a été de proposer une méthode alternative automatisée de modélisation des rachats qui s'appuie sur les données emprunteurs disponibles, et qui est caractérisée par une adaptabilité entre autres à l'environnement économique, affectant les comportements des assurés. Elle constitue une réponse permettant de quantifier l'effet possible des évolutions réglementaires les plus récentes comme les lois Bourquin ou Lemoine sur le portefeuille emprunteur et censées favoriser la concurrence, même si les premières études d'impact de l'entrée en vigueur de la loi Lemoine montrent une tendance haussière des tarifs ADE chez certains assureurs (voir[ARGUS, 2022]). Mais au delà du cadre réglementaire du marché de l'assurance, elle permet également de prendre en compte l'évolution récente du contexte économique marqué par l'inflation, une remontée rapide et significative des taux de crédit et une baisse de la collecte crédit.

Cependant la modélisation alternative centrale présente une qualité d'ajustement modérée au réel, qui peut avoir différentes causes. Elle peut s'expliquer par une qualité moyenne des prédictions définissant les scores : une piste de réponse à ce problème pourrait être de revoir l'échantillonnage sur lequel est réalisé la modélisation, en procédant à un rééquilibrage des effectifs. Ce rééquilibrage pourrait se faire, par exemple en sous-échantillonnant les contrats non rachetés en ne gardant pour chaque contrat racheté - caractérisé par sa date et son agence de souscription, que les contrats souscrits dans la même agence et au même moment. Une autre cause pourrait être liée à l'ajustement des seuils des inducteurs des scores qu'il conviendrait de suivre et d'optimiser dans le temps. Les limites constatées peuvent enfin s'expliquer par des niveaux de probabilités de rachats, héritées de la base de modélisation. Dans notre exemple la base du modèle est caractérisée par une volumétrie de rachats particulièrement élevée par rapport à la base de back-test qui tend à augmenter globalement les probabilités de rachats, malgré la segmentation. Néanmoins, un paramétrage dynamique adapté des scénarii, ci-dessus, peut corriger et améliorer les performances du modèle.

Par ailleurs, une optimisation du process consisterait à internaliser dans prophet le calcul dynamique du risque au cours du temps (via les scores) en mutualisant lorsque cela est possible le paramétrage de prophet avec celui des scénarii économiques utilisé dans la modélisation des résiliations.

Si le contexte économique le permet, l'ouverture du marché ADE visé par la réglementation devrait également se traduire par une augmentation significative des renégociations du coût global des montants empruntés, impliquant à la fois les conditions du contrat de prêt et celles de son assurance. Dans ce cadre, une modélisation dynamique des renégociations venant compléter celle des résiliations pourrait être pertinente.

Chapitre 7

Annexes

7.1 Le marché de l'ADE des crédits immobiliers

7.1.1 Evolution de la réglementation

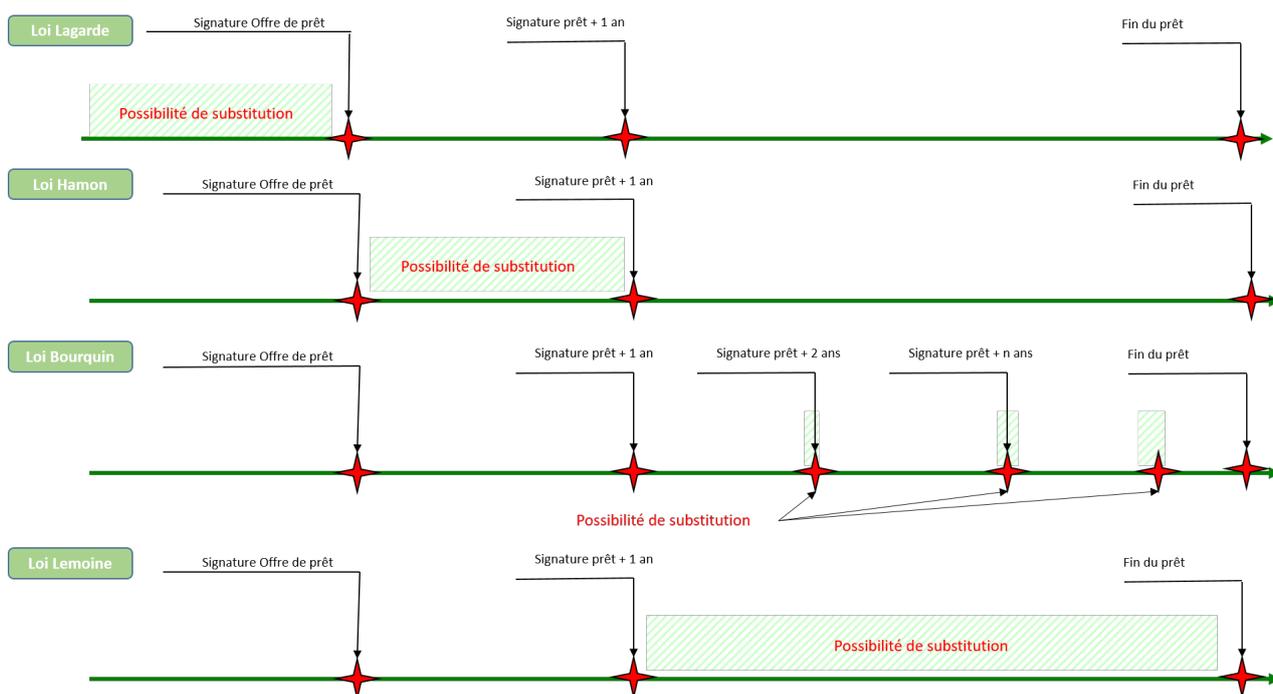


Fig. 7.1 – Impact cumulé des évolutions réglementaires sur les possibilités de substitutions

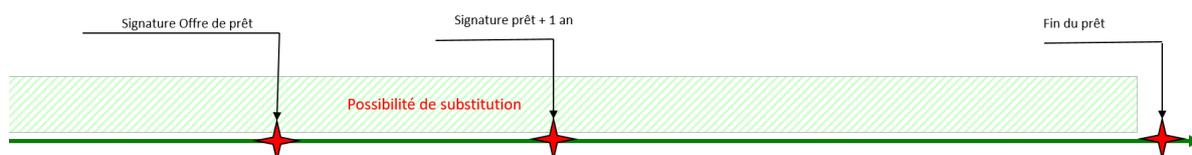


Fig. 7.2 – Impact des évolutions réglementaires sur les possibilités de substitutions

7.1.2 Etude de la dynamique des indicateurs crédit et ADE mesurée dans les fichiers sources entre juillet 2015 et juillet 2019

7.1.2.1 Evolution des taux de crédits immobiliers effectifs – crédits amortissables

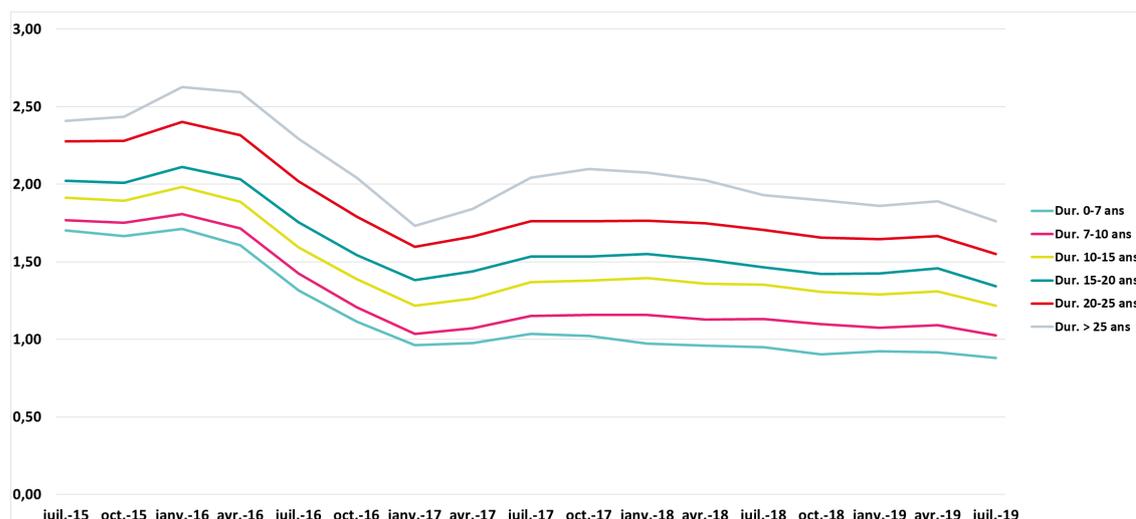


Fig. 7.3 – Evolution trimestrielle des taux de crédits immo. amortissables, par tranches d’âge – période [07/2015,07/2019]

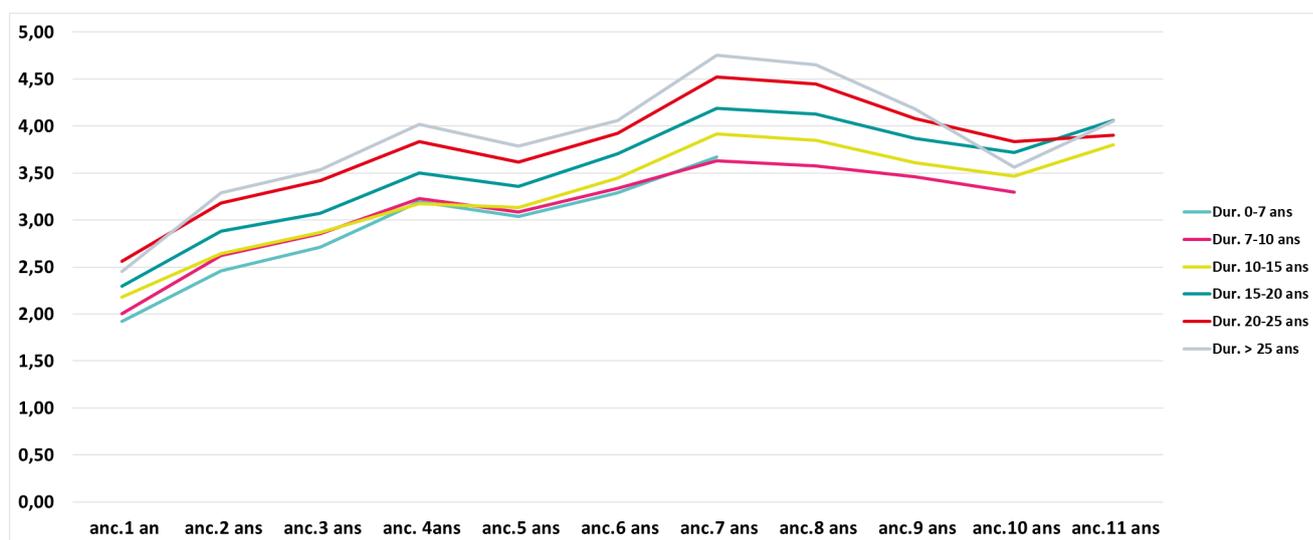


Fig. 7.4 – Structure des taux de prêts immo. Amortissables (en %) par ancienneté et tranches de durées - juillet 2015

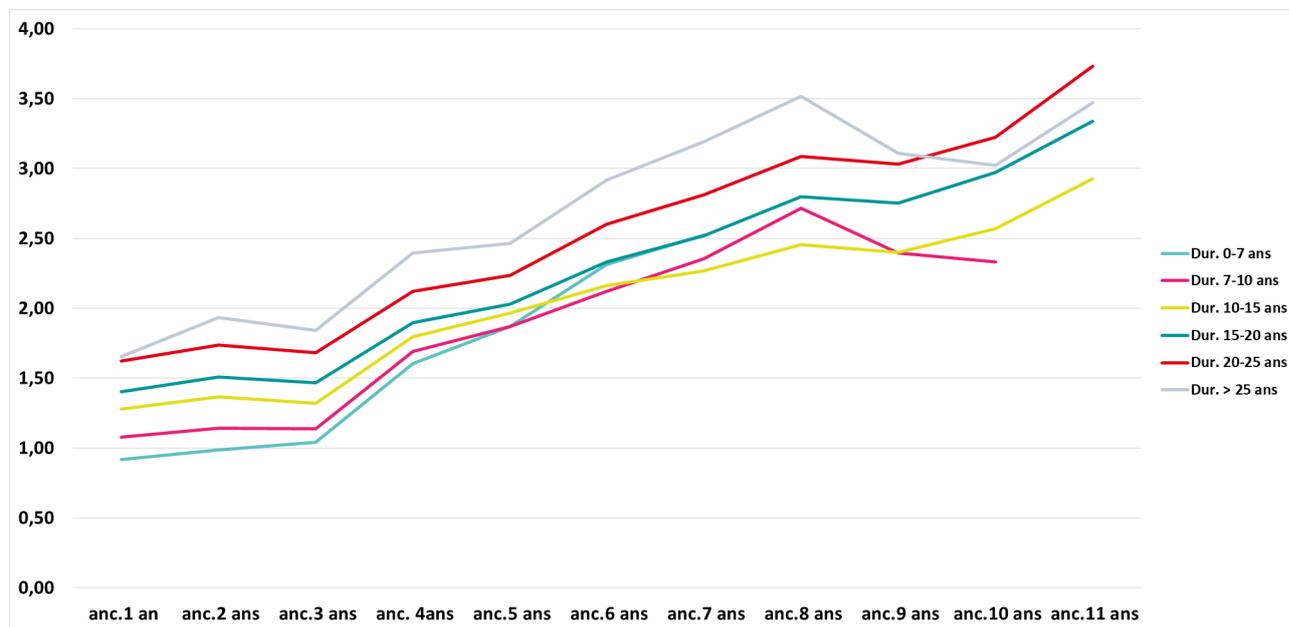


Fig. 7.5 – Structure des taux de prêts immo. Amortissables (en %) par ancienneté et tranches de durées - juillet 2019

7.1.2.2 Evolution des taux ADE immobilier annualisés effectifs – couvrant les prêts amortissables

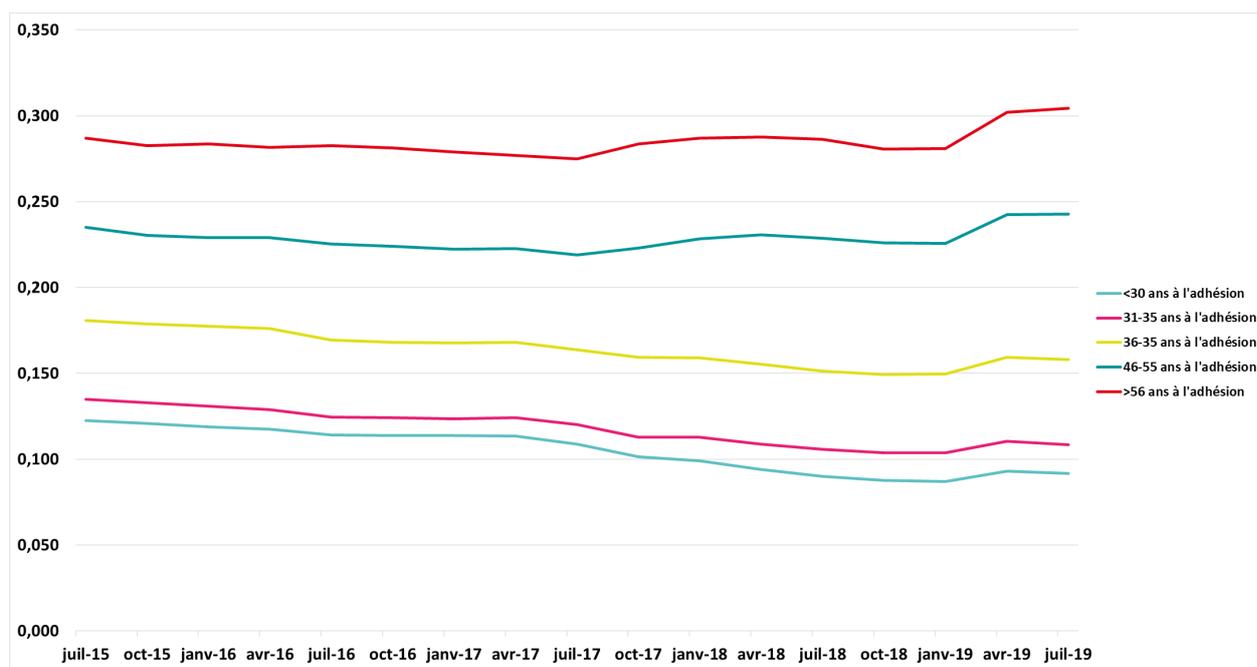


Fig. 7.6 – Evolution trimestrielle des taux ADE immo. annualisés, par tranches d'âge – période [07/2015, 07/2019]

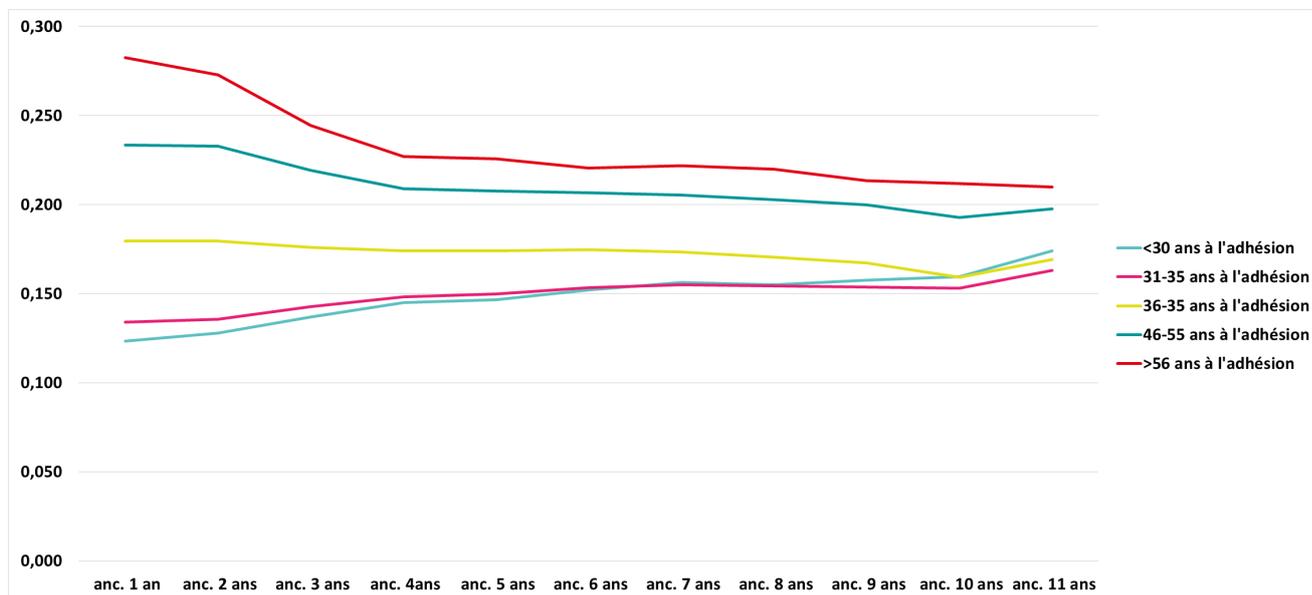


Fig. 7.7 – ADE - Structure des taux ADE immo. annualisés (en %) par tranche d'âge et ancienneté - juillet 2015

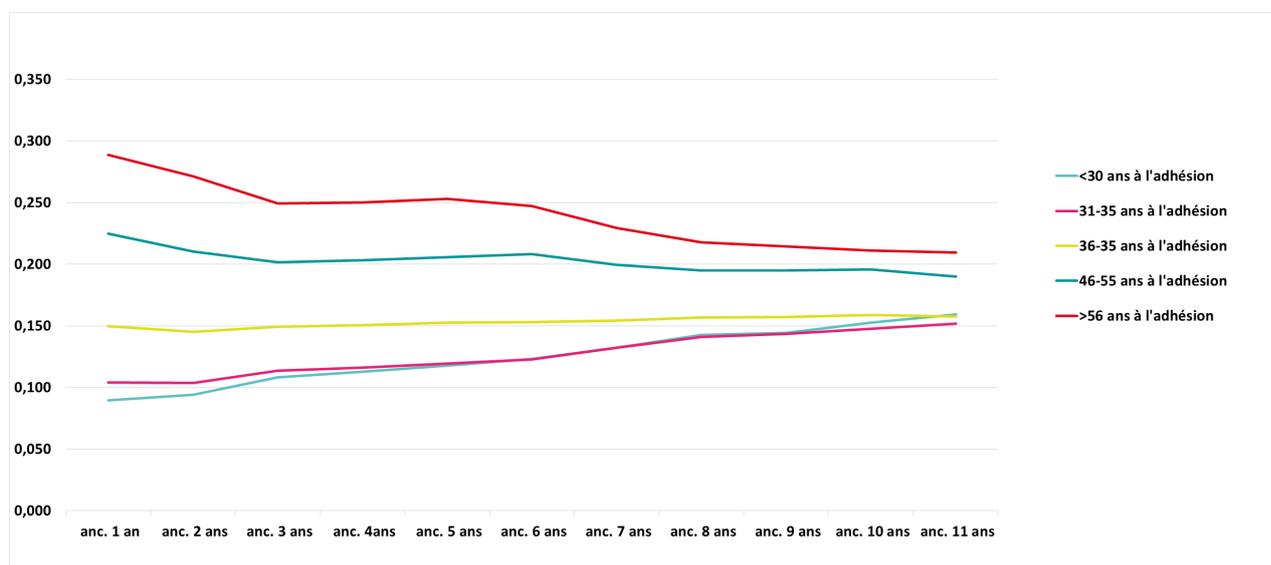


Fig. 7.8 – ADE - Structure des taux ADE immo. annualisés (en %) par tranche d'âge et ancienneté - juillet 2019

7.2 Les bases d'étude

7.2.1 Sélection des agences les plus contributrices de la CR 802 pour l'ancienneté 5 au seuil de 10%

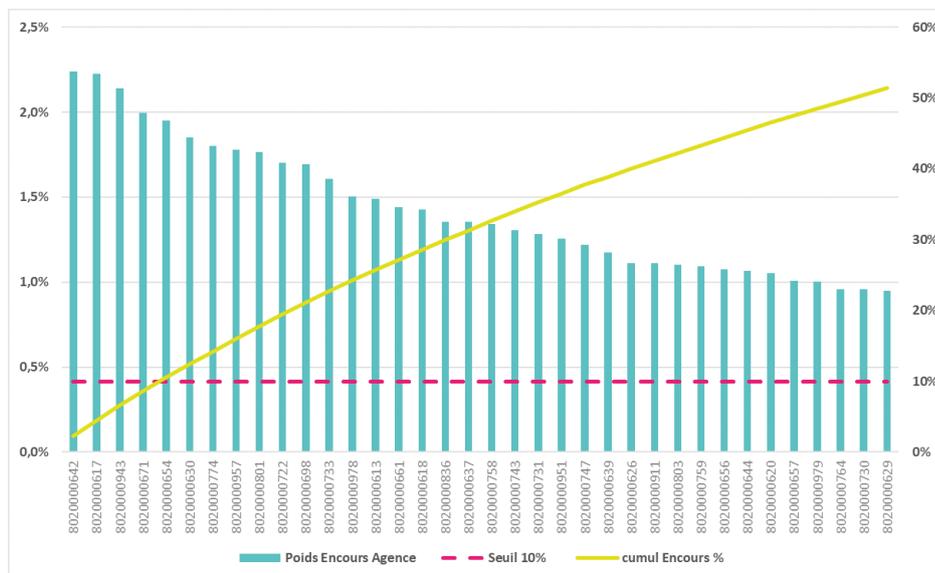


Fig. 7.9 – Echantillonnage stratifié – exemple de la CR 802

7.2.2 Proportion des agences (en encours crédit) retenues dans l'échantillonnage par ancienneté

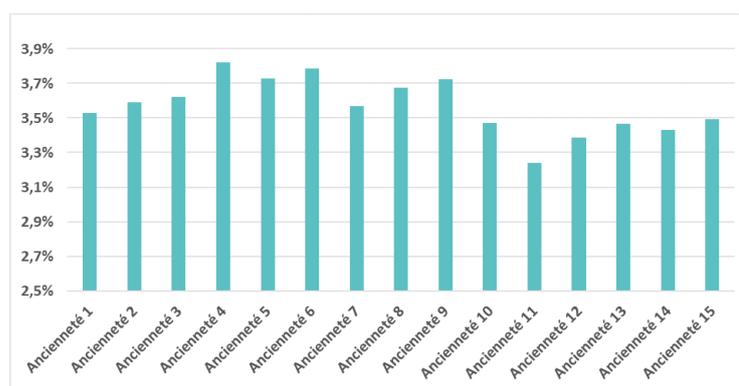


Fig. 7.10 – proportion des agences retenues dans l'échantillonnage par ancienneté avec un seuil de 10%

7.2.3 Proportion des contrats filtrés dans l'étude comportementale

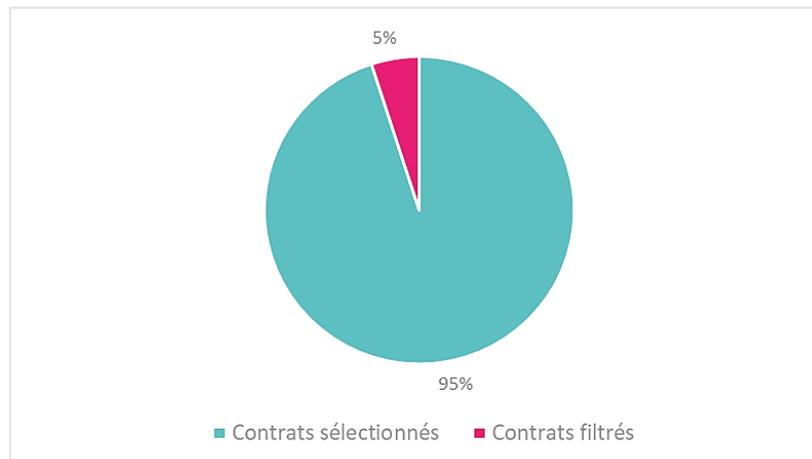


Fig. 7.11 – Périmètre de l'étude comportementale - poids des expositions partielles (contrats filtrés) - % des primes décès

7.2.4 Décomposition en % de chiffre d'affaires et source d'exposition partielle

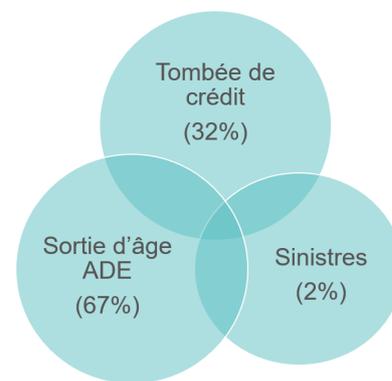


Fig. 7.12 – décomposition en % de CA décès des expositions partielles

7.2.5 Distribution des effectifs (en nombre) par ancienneté

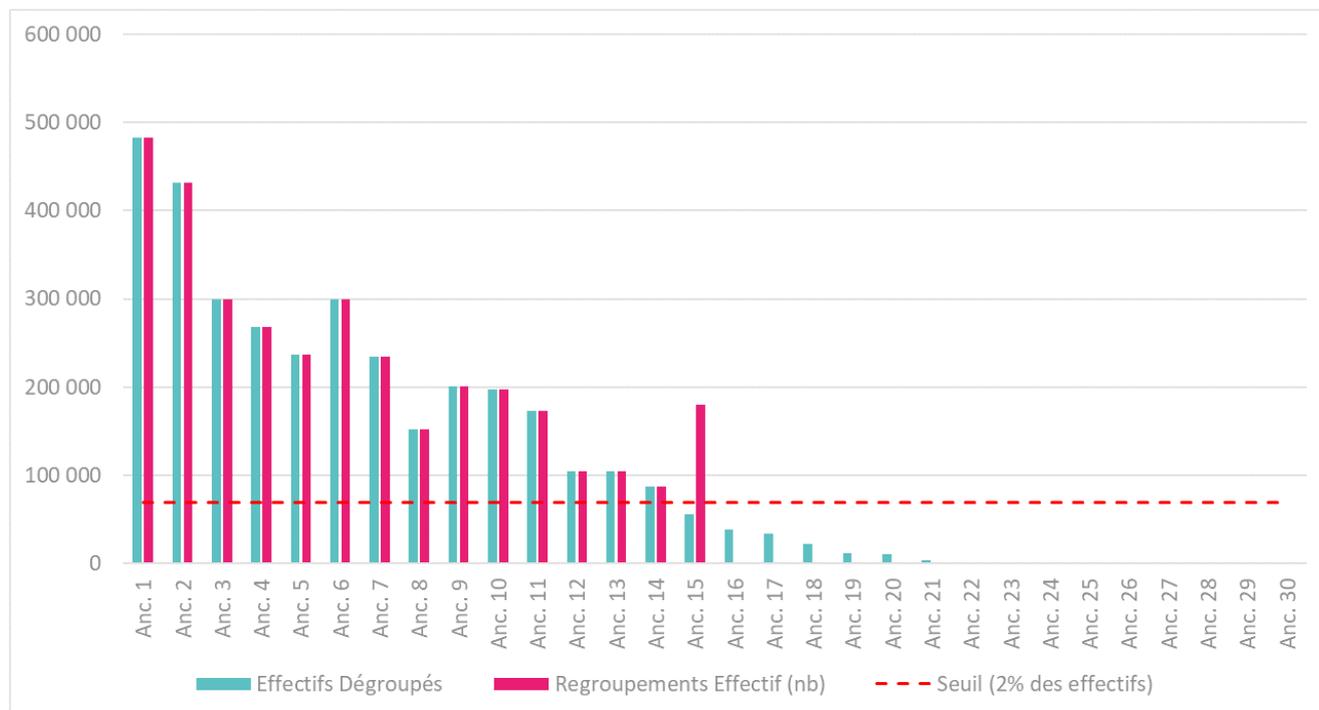


Fig. 7.13 – Distribution des effectifs en nombre par ancienneté – Base [B1]

7.2.6 Réconciliation des primes décès de la base d'étude avec la comptabilité

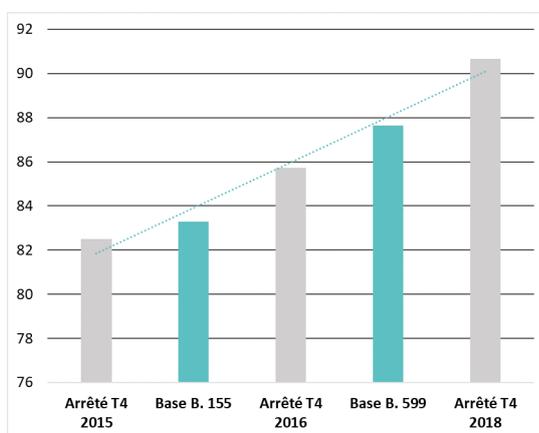


Fig. 7.14 – Primes DC HT – base 100

7.2.7 Réconciliation des primes ITT de la base d'étude avec la comptabilité

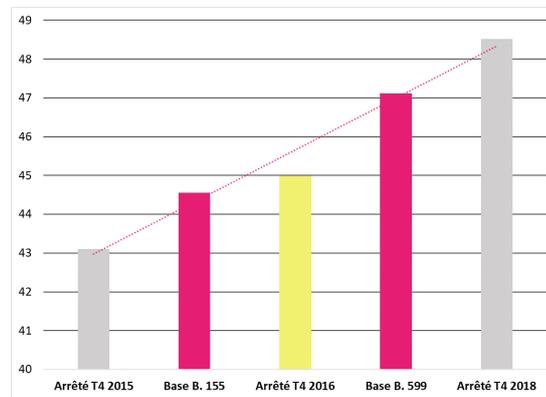


Fig. 7.15 – Primes ITT – base 100

7.2.8 Contribution du stock et des affaires nouvelles au chiffre d'affaires annuel dans la base d'étude

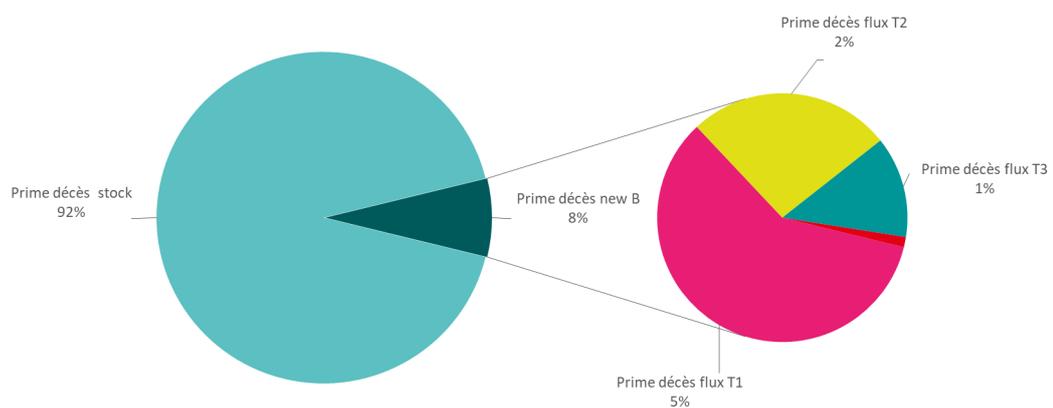


Fig. 7.16 – Contribution des générations de contrats au chiffre d'affaire (% primes décès)

7.2.9 Ajustement global entre la base échantillonnée et la base totale

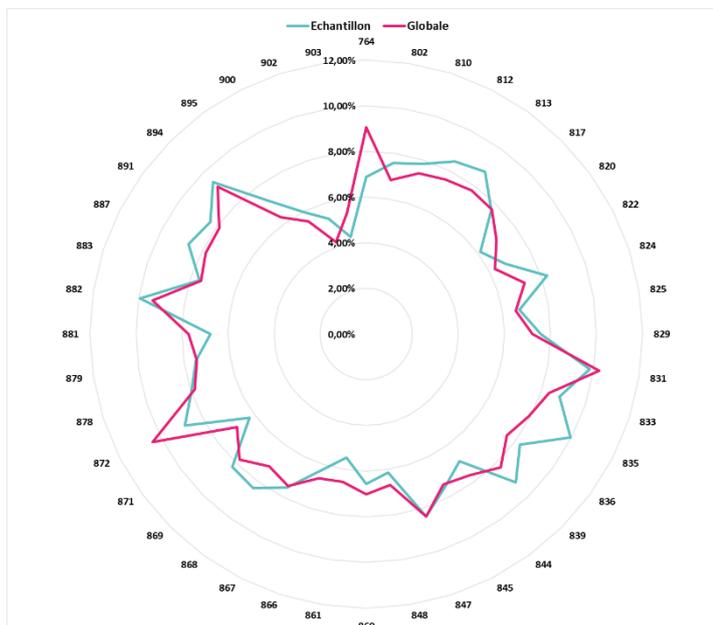


Fig. 7.17 – Ajustement [BE1] échantillon/pop. Totale par Caisse régionales

7.2.10 Piste d'optimisation de l'ajustement

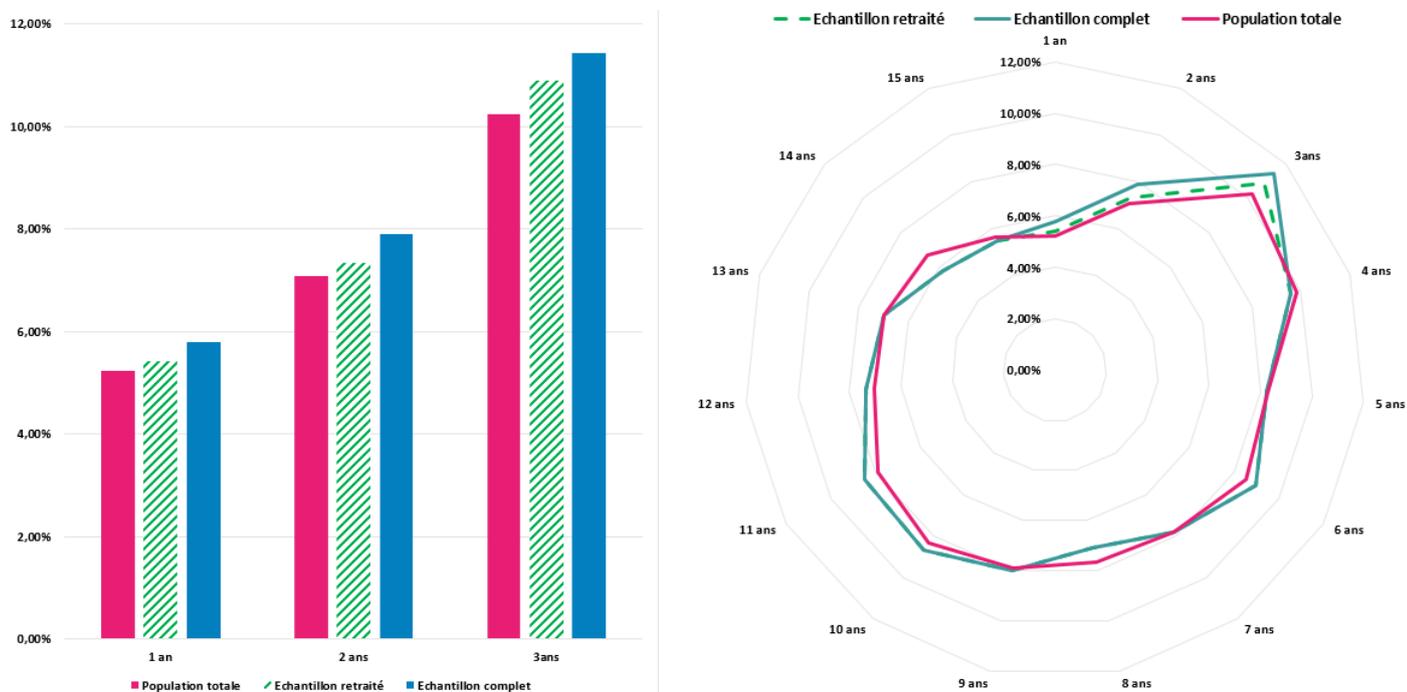


Fig. 7.18 – Ajustement [BE1] échantillon/pop. Totale par Caisse régionales

7.2.11 Présentation des données

7.2.11.1 Structure des données

Variable	Définition	Type Donnée
Montant_Financement_Courant	Capital initial de l'opération de crédit couverte	Source
Duree_Init_Financement_Courant	Durée initiale du contrat	Source
Taux_Remise_Tarifaire_Ref	Ecart (%) du taux de tarif avec un taux de référence	calculé
Taux_Pret	Taux nominal du prêt	Calculé
Taux_Couverture_DC	Quotité décès	Source
Taux_Couverture_ITT	Quotité ITT	Source
Taux_Cotisation	Tarif effectif total (hors perte d'emploi) = Prime totale/CI couvert en décès	Calculé
Taux_Cotisation_DC	Tarif effectif décès = Prime décès/CI couvert en décès	Calculé
Age	Âge début de période = date de début de période - date de naissance	Calculé
Revenu_Median	Source INSEE	Enrichissement
SpreadTxCredMarche	Ecart (%) du taux nominal de l'opération de crédit avec la taux nominal moyen des crédits récents (< 3 mois) en portefeuille de même durée résiduelle	Calculé
SpreadTxADEMarche	Ecart (%) du taux de tarif ADE du contrat avec la taux ADE moyens des contrats ADE récents (< 3 mois) en portefeuille de même durée résiduelle	Calculé
Anciennete_ADE	Ancienneté ADE = Date de début de période - date de signature du contrat ADE	calculée
AncienneteADEGrp	AncienneteADEGrp = min(Anciennete_ADE, 14)	calculée
Code_CSP	Code catégorie socio professionnelle	Source
Code_Sexe	Code sexe	Source
Type_Commune1	Type commune 1 nomenclature INSEE	Enrichissement
Type_Commune2	Type commune 2 nomenclature INSEE	Enrichissement
Type_Commune3	Type commune 3 nomenclature INSEE	Enrichissement
Type_CR	Type de CR nomenclature crédit Agricole SA	Source
FlagAssExt	Marqueur d'existence d'un encours crédit (tout marché crédit) couvert avec une ADE externe	Calculé
FlagExistEnc	Marqueur d'existence d'un encours crédit (tout marché crédit)	Calculé
Flag_Locatif	Marqueur d'investissement locatif de l'encours couvert	Calculé
Flag_coemprunt_Fin_Courant	Marqueur de co-emprunt de l'encours couvert	Calculé
Cde_Dom_Intervention	Code domaine d'intervention crédit (noyau crédit)	Source
Cde_Dest_Financement	Code destination du financement (noyau crédit)	Source
Code_Region_Resid	Code région résidence	Source

Fig. 7.19 – Cartographie des variables d'étude

7.2.11.2 La définition des spreads de taux de crédit ou d'ADE

Le calcul des spreads de taux de crédit ou ADE nécessite de calculer des moyennes et donc d'effectuer dans un 1^{er} des regroupements de contrats par familles crédit et ADE homogènes. Les familles identifiées sont concernent :

- Les financements à taux fixe
 1. Les financements in fine (y compris les financements court terme relais dont la durée est comprise au maximum entre 12 et 24 mois)
 2. Les prêts à taux zéro (PTZ)
 3. Les prêts amortissables standard
- les financements à taux variable

L'étude présentée étant limitée à la 3^e sous-catégorie de financements ci-dessus, il s'agit dans un 2^e temps de la segmenter selon des dimensions disponibles en base qui pilotent les taux. Pour le crédit les dimensions structurantes retenues sont la durée initiale pour le financement et l'âge de l'assuré pour l'ADE. Les tranches de segmentations sont les suivantes :

- La durée du financement :
 - 0-7 ans
 - 7-10 ans
 - 10-15 ans
 - 15-20 ans
 - 20-25 ans
 - > 25 ans
- L'âge de l'assuré
 - <30 ans
 - 31-35 ans
 - 36-35 ans
 - 46-55 ans
 - >56 ans

Les taux moyens sur les productions récentes (≤ 3 mois) sont calculés sur les segmentations ci-dessus (l'âge des assurés sont les âges à la souscription). Ils sont ensuite affectés à chaque assuré en fonction de son âge atteint et de la durée résiduelle de son contrat (= durée initiale- ancienneté). Ce taux correspond à un proxy de taux de marché auquel l'assuré pourrait prétendre s'il décidait de racheter son crédit ou de résilier son contrat ADE.

Le spread de taux (ADE ou crédit) est défini comme la différence entre le taux (ADE ou crédit) propre à l'assuré et le taux moyen de marché (ADE ou crédit) qui lui a été affecté précédemment (= Taux_ assuré - Taux_ marché_ assuré).

7.2.11.3 La description de la typologie des communes INSEE

Type_Commune1	Source INSEE ¹ : <ul style="list-style-type: none"> • RURAL 0 • URBAIN 1
Type_Commune2	source INSEE ² (à partir du code résidence emprunteur) <ul style="list-style-type: none"> • R -Rural 0 • C -Ville – centre 1 • B –Banlieue 2 • I - Ville isolée 3
Type_Commune3	source INSEE (à partir du code résidence emprunteur) <ul style="list-style-type: none"> • 0 : Autres • 1 : Unité urbaine de 2 000 à 4 999 habitants • 2 : Unité urbaine de 5 000 à 9 999 habitants • 3 : Unité urbaine de 10 000 à 19 999 habitants • 4 : Unité urbaine de 20 000 à 49 999 habitants • 5 : Unité urbaine de 50 000 à 99 999 habitants • 6 : Unité urbaine de 100 000 à 199 999 habitants • 7 : Unité urbaine de 200 000 à 1 999 999 habitants • 8 : Paris

Fig. 7.20 – Les codes communes INSEE

7.2.12 Stratégie de gestion des valeurs manquantes ou aberrantes

Variable	Problème détecté	Traitement effectué
Type de contrats	Des identifiants mal renseignés	Données supprimées
Date de signature ADI	Valeur manquante (impossible car donnée obligatoire pour déclencher le prélèvement de primes)	= date de signature crédit si elle est présente Sinon suppression en fonction de la fréquence
Date de signature crédit	Valeur manquante (impossible car donnée obligatoire pour déclencher le prélèvement de primes)	= date de signature ADI si elle est présente Sinon suppression de l'observation suppression en fonction de la fréquence
Durée du prêt	Valeurs manquantes Valeurs aberrantes (=0 ou >360 mois : durée maximale des prêts)	Réévaluation à la moyenne suppression en fonction de la fréquence
Date de fin de prêt	Mise en cohérence de la date de fin de prêt avec la durée	Si la durée du prêt n'est pas retraitée = date de fin de prêt Sinon recalcul à partir de la durée initiale, en prenant en compte la phase du crédit (instruction gestion) et l'éventuelle présence d'une période d'anticipation
Quotité Décès	<ul style="list-style-type: none"> Valeurs nulles (impossible car garantie obligatoire) Valeurs supérieures à 100% pour un seul assuré sur un même contrat 	Si les montants de primes DC et de remises commerciales et taux tarifaires DC catalogues sont présents = prime DC mensuelle/(tarif mensuel DC x remise X CI prêt) Réévaluation à la moyenne suppression en fonction de la fréquence
Quotité d'Arrêt de travail	Valeurs supérieures à 100% pour un seul assuré sur un même contrat	Si les montants de primes ITT et de remises commerciales et taux tarifaires catalogues ITT sont présents = prime ITT mensuelle/(tarif mensuel ITT x remise x CI prêt) Réévaluation à la moyenne ou suppression en fonction de la fréquence
Date d'effet	Valeurs manquantes Valeurs aberrante	Si les dates ADI et/ou prêts sont présentes, = max (date signature ADI, signature crédit) sinon Sinon réévaluation à la moyenne ou suppression en fonction de la fréquence
Année de naissance de l'assuré	<ul style="list-style-type: none"> Valeurs manquantes Valeurs aberrante 	Si le jour est manquant = 15 du mois Si le mois est manquant = mi-année (juin) Sinon réévaluation à la moyenne ou suppression en fonction de la fréquence

Fig. 7.21 – Descriptif des retraitements des données aberrantes ou manquantes

7.2.13 Analyse descriptive de quelques variables

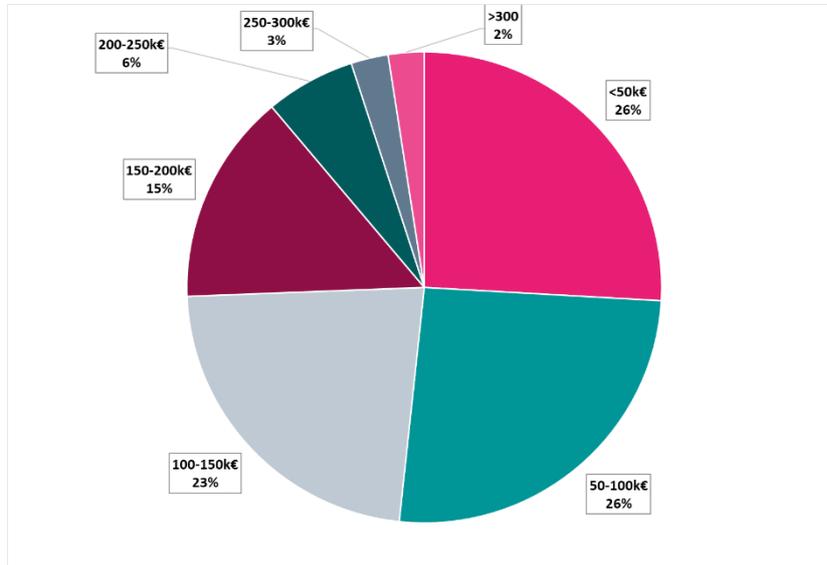


Fig. 7.22 – Analyse descriptive – distribution des capitaux empruntés

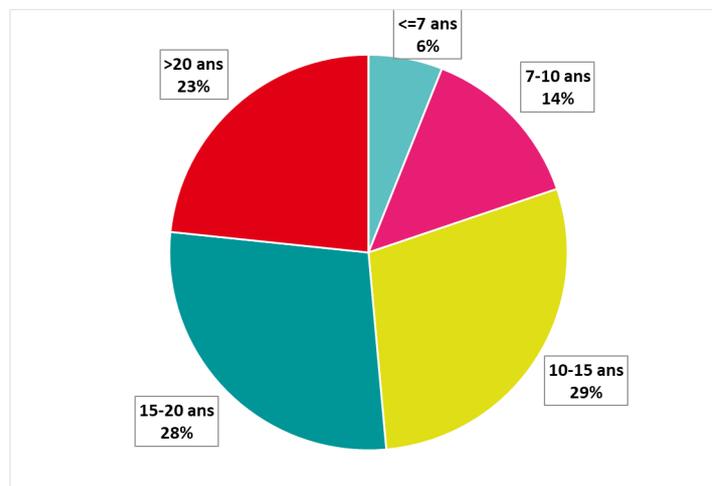


Fig. 7.23 – Analyse descriptive – durée des financements

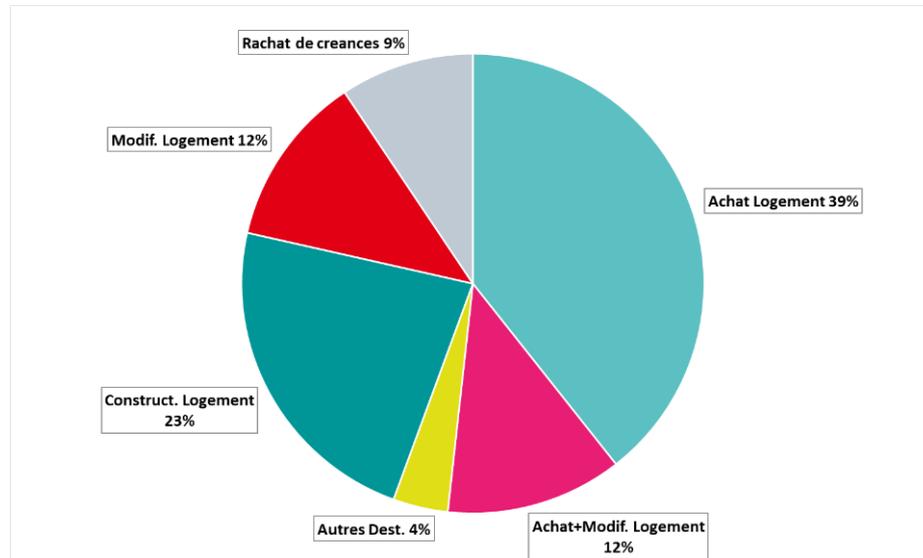


Fig. 7.24 – Analyse descriptive – destination des financements

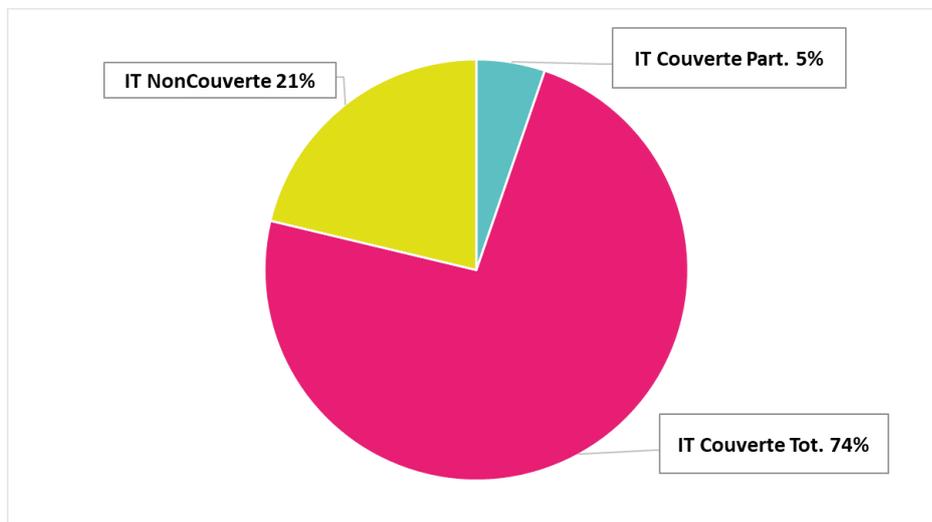


Fig. 7.25 – Analyse descriptive – couverture ITTT

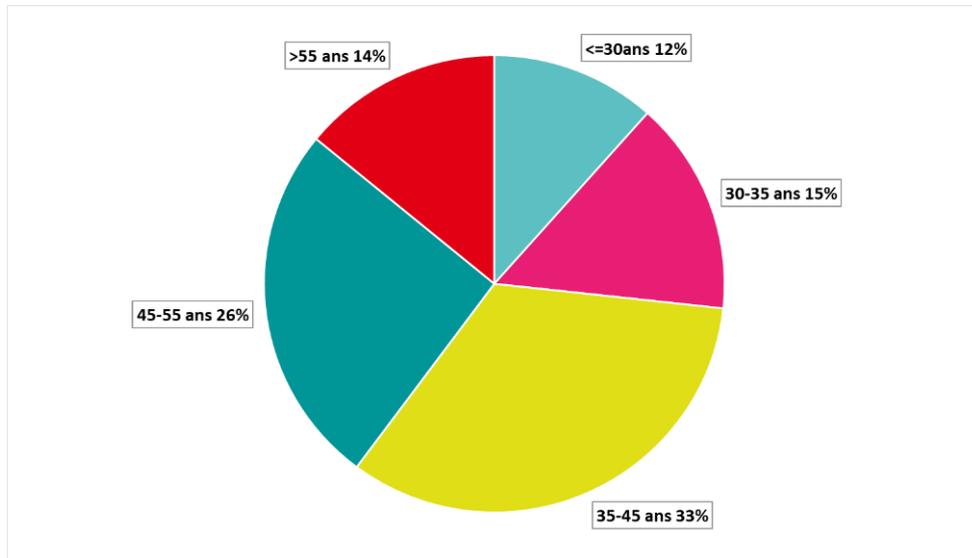


Fig. 7.26 – Analyse descriptive – âge des assurés

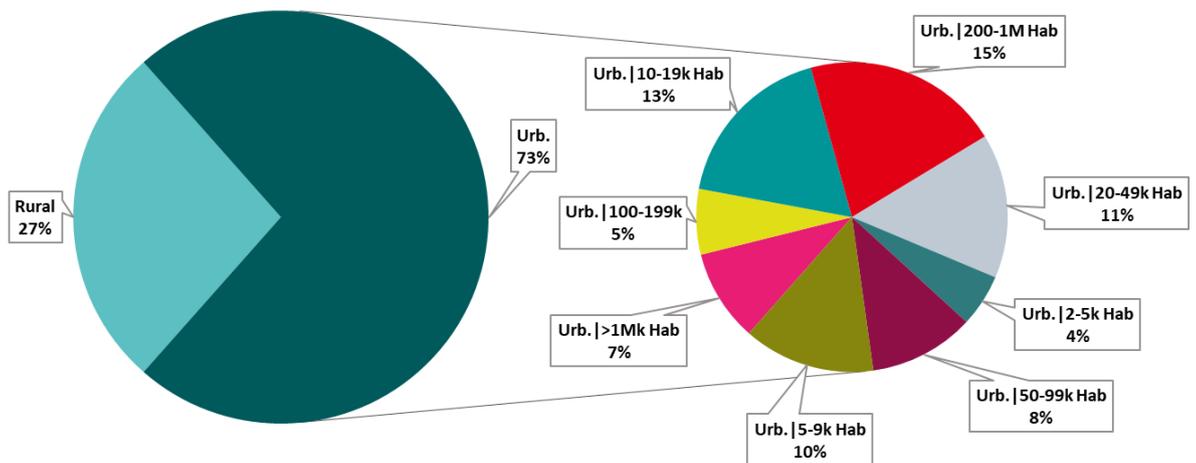


Fig. 7.27 – Analyse descriptive – typologie des lieux résidences des emprunteurs

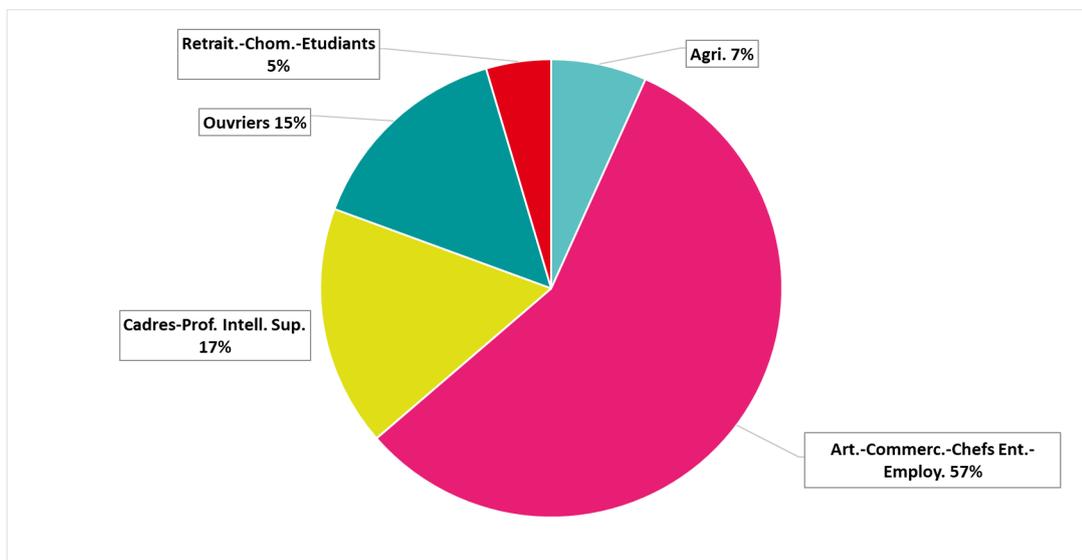


Fig. 7.28 – Analyse descriptive – CSP des emprunteurs

7.3 Modélisation des rachats en assurance emprunteur

7.3.1 Eléments théoriques

7.3.1.1 La discriminante qualitative

Cette section fait appel aux notations suivantes.

n : le nombre d'individus

Y variable cible qualitative à k modalités, correspondant à k groupes ou classes G_i

$\{X_k\}_{1 \leq k \leq p}$, p variables explicatives continues

Principe de l'analyse factorielle discriminante :

Il s'agit remplacer les X_j par des axes discriminants selon une méthode des moindres carrés : combinaisons linéaires des X_j prenant les valeurs les plus différentes (cf. principe de l'analyse en composantes principales). $k - 1$ axes sont obtenus (si le nombre d'individus $n > p > k$)

Les n individus forment un nuage de n points dans \mathbb{R}^p , formé des k sous-nuages G_i à différencier.

Variance interclasse (« Between ») = variance des barycentres g_i (« centroïdes ») des classes G_i

$$B = \frac{1}{n} \sum_{1 \leq i \leq n} n_i (g_i - g)(g_i - g)^t$$

(matrice de covariance « Between »)

Variance intraclasse (« Within ») = moyenne des variances des classes G_i

$$W = \frac{1}{n} \sum_{1 \leq i \leq n} n_i V_i$$

(matrice de covariance « Within »)

Théorème de Huygens :

$$V = B + W$$

(Variance totale)

Il est impossible de trouver un axe u qui simultanément :

- maximise la variance interclasse sur u : $Max_u u^t B u$
- minimise la variance intraclasse sur u : $Min_u u^t W u$

Compromis : il est choisi de maximiser $\frac{u^t B u}{u^t W u}$, ou, ce qui est équivalent $\frac{u^t B u}{u^t V u}$ (Huygens), au lieu de maximiser $u^t B u$ ou minimiser $u^t W u$. Cette maximisation se fait sous hypothèse d'homoscédasticité (égalité des V_i , $1 \leq i \leq k$)

Il peut être montré que la solution u est le vecteur propre de $V^{-1}B$ associé à la plus grande valeur propre λ de $V^{-1}B$.

L'équivalence suivante est vérifiée : $\{u \text{ vecteur propre de } V^{-1}B\} \iff \{u \text{ vecteur propre de } W^{-1}B \text{ de valeur propre } \frac{\lambda}{1-\lambda}\}$

Les métriques induites par V^{-1} et W^{-1} sont équivalentes.

En particulier la distance d (distance de Mahalanobis) induite par W^{-1} entre deux points x et y est définie telle que :

$$d^2(x, y) = (x - y)^t W^{-1} (x - y)$$

L'analyse discriminante géométrique prédictive

x est classé dans le groupe G_i pour lequel la distance au centre g_i est minimale :

$$d^2(x, g_i) = (x - g_i)'W^{-1}(x - g_i) = x'W^{-1}x - 2g_i'W^{-1}x + g_i'W^{-1}g_i$$

Minimiser $d^2(x, g_i)$ équivaut à maximiser $2g_i'W^{-1}x + g_i'W^{-1}g_i$

$g_i'W^{-1}g_i = \alpha_i$ est une constante ne dépendant pas de x .

Pour chacun des k groupes G_i , la fonction discriminante de Fisher est obtenue après inversion de W telle que :

$$\alpha_i + \alpha_{i,1}X_1 + \alpha_{i,2}X_2 + \dots + \alpha_{i,p}X_p$$

Ensuite, x est classé dans le groupe pour lequel la fonction est maximale.

Le cas de deux groupes ($k = 2$) :

L'axe discriminant u , joint les deux centroïdes. Il peut être montré que $u = (a_1, a_2, \dots, a_p)$ est proportionnel à $W^{-1}(g_1 - g_2)$.

D'un point de vue prédictif, x est classé dans le groupe G_1 si :

$d^2(x, g_1) > d^2(x, g_2)$, soit

$2g_1'W^{-1}x - g_1'W^{-1}g_1 > 2g_2'W^{-1}x - g_2'W^{-1}g_2$, soit

$f(x) = (g_1 - g_2)'W^{-1}x - \frac{1}{2}(g_1'W^{-1}g_1 - g_2'W^{-1}g_2) > 0$, où f est la fonction de score de Fisher.

Pratiquement, la fonction f est obtenue par différence des fonctions discriminantes associées à chaque groupe.

L'analyse discriminante probabiliste (ou bayésienne)

Cette méthode s'appuie sur le théorème de Bayes et consiste (= règle bayésienne) à maximiser en x ,

$$\mathbb{P}(G_i|x) = \frac{\mathbb{P}(G_i)\mathbb{P}(x|G_i)}{\sum_{1 \leq j \leq k} \mathbb{P}(G_j)\mathbb{P}(x|G_j)}$$

($\mathbb{P}(G_i)$: proportion de G_i parmi les individus), ce qui équivaut à maximiser $\mathbb{P}(G_i)\mathbb{P}(x|G_i)$.

En faisant une hypothèse de multinormalité sur les distributions de $\mathbb{P}(x|G_i) = f_i(x)$, $1 \leq i \leq k$, et d'homoscédasticité, pour $k = 2$, la règle revient à classer x dans le groupe G_1 , si $f(x) > \text{Log} \frac{\mathbb{P}(G_2)}{\mathbb{P}(G_1)}$, où f est la fonction discriminante de Fisher définie précédemment. Il vient, en ajoutant une hypothèse d'équiprobabilité, que pour $k = 2$, la règle correspond à celle de l'analyse discriminante géométrique décrite plus haut.

Le cas de prédicateurs qualitatifs - la méthode DISQUAL

Elle permet de gérer la présence de variables qualitatives sans les traiter comme variables supplémentaires. La méthode se base sur la forme disjonctive de la variable qualitative X_j à m_j modalités ; $(\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_{m_j})$ définit les m_j variables indicatrices des modalités. Les m_j indicatrices sont équivalentes à la variable qualitative. Le score est une combinaison linéaire des indicatrices, ce qui revient à donner une note partielle à chaque modalité de chaque variable, le score final étant la somme des notes partielles. Les variables explicatives qui interviennent sont donc les indicatrices de toutes les variables.

Remarque : la matrice W n'est pas de plein rang, et n'est donc pas inversible car la somme des indicatrices des modalités de chaque variable vaut 1. Cela signifie qu'il y a une infinité de solutions équivalentes pour estimer les coefficients; une des solutions couramment utilisée consiste à ne prendre que les $(m_j - 1)$ indicatrices pour chaque variables qualitatives puisque la dernière est redondante.

Pratiquement, l'opération se fait en trois étapes :

- Etape 1 : analyse des correspondances multiples sur le tableau disjonctif complet des variables qualitatives et récupération des coordonnées des individus sur les axes factoriels les plus discriminants.
- Etape 2 : injection des coordonnées ci-dessus en entrée d'une analyse discriminante classique (cf. ci-dessus) qui conduit à une fonction de score de Fisher, combinaison linéaire des axes factoriels.
- Etape 3 : expression de la fonction de Fisher comme combinaison des indicatrices des modalités des variables initiales (qualitatives).

7.3.1.2 La régression logistique (LOGIT)

Cette section fait appel aux notations suivantes.

n : le nombre d'individus

Y : variable cible binaire ($Y = 0$ ou $Y = 1$)

$\{X_k\}_{1 \leq k \leq p}$, p variables explicatives continues, binaires ou qualitatives

$X = (X_1, X_2, \dots, X_p)$

$x = (x_1, x_2, \dots, x_p)$

Principe de la régression logistique binaire :

L'objectif visé est d'écrire l'espérance conditionnelle de Y comme combinaison linéaire des variables explicatives $\{X_k\}_{1 \leq k \leq p}$. L'objectif de la régression est de modéliser

$\mathbb{E}(Y \mid X = x) = \mathbb{P}(Y = 1 \mid X = x)$, (Y ne prenant que les valeurs 0 ou 1), non pas sous la forme :

$$\mathbb{P}(Y = 1 \mid X = x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (7.1)$$

mais sous la forme :

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

avec $\pi(x) = P(Y = 1 \mid X = x)$.

En effet, dans (7.1) $\mathbb{P}(Y = 1 \mid X = x)$ est borné alors que $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ ne l'est pas. La fonction $f : p \mapsto (\text{Log} \frac{p}{1-p})$ est appelée logit.

On a $\pi(x) = \frac{e^{\beta_0 + \sum_j \beta_j x_j}}{1 + e^{\beta_0 + \sum_j \beta_j x_j}}$

Les odds-ratios (OR) :

L'OR (i.e. «cote») d'une variable explicative mesure l'évolution du rapport des probabilités d'apparition de l'évènement $Y = 1$ contre $Y = 0$, lorsque X_i passe de x à $x + 1$. Dans ce cas, $\text{Logit}(\pi(x))$ augmente du coefficient β_i de X_i donc la cote $\frac{\pi(x)}{1 - \pi(x)}$ est multipliée par e^{β_i} . L'OR est défini comme suit :

$$OR = \frac{\frac{\pi(x+1)}{1 - \pi(x+1)}}{\frac{\pi(x)}{1 - \pi(x)}} = e^{\beta_i}$$

Remarque : pour les variables qualitatives il y a autant d'OR que de modalités moins une, l'une des modalités étant prise pour référence (généralement, son coefficient = 0).

Le cas de deux groupes ($k = 2$) :

Ce cas conduit au modèle pour $i, 1 \leq i \leq n$, $\pi(x^i) = \mathbb{P}(Y = 1 \mid X = x^i) = \frac{e^{\beta_0 + \sum_j \beta_j x_j^i}}{1 + e^{\beta_0 + \sum_j \beta_j x_j^i}}$

Les coefficients β_i recherches maximisent la fonction de vraisemblance :

$$\begin{aligned} L(\beta_1, \beta_2, \dots, \beta_p) &= \prod_{1 \leq i \leq n} \mathbb{P}(Y = y^i \mid X = x^i) = \prod_{1 \leq i \leq n} \pi(x^i)^{y^i} (1 - \pi(x^i))^{1-y^i} \\ &= \prod_{1 \leq i \leq n} \left(\frac{e^{\beta_0 + \sum_j \beta_j x_j^i}}{1 + e^{\beta_0 + \sum_j \beta_j x_j^i}} \right)^{y^i} \left(1 - \frac{e^{\beta_0 + \sum_j \beta_j x_j^i}}{1 + e^{\beta_0 + \sum_j \beta_j x_j^i}} \right)^{1-y^i} \end{aligned}$$

L'estimation des coefficients par maximisation de la vraisemblance est notée $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$. Elle conduit à une estimation $\pi(\hat{x}^i)$ de la probabilité conditionnelle $\pi(x^i)$ d'avoir $Y = 1$ sachant $X = x^i$.

Fiabilité des estimateurs $\hat{\beta}_j$: la statistique de Wald :

La significativité des estimateurs $\hat{\beta}_j$ et l'apport de la variable X_j sont testés à l'aide du χ^2 de Wald de X_j , définie comme le rapport $w = \frac{\hat{\beta}_j^2}{\sigma(\hat{\beta}_j)}$. Sous l'hypothèse H_0 que $\hat{\beta}_j = 0$, $\frac{\hat{\beta}_j}{\sigma(\hat{\beta}_j)}$ suit une loi $\mathcal{N}(0, 1)$. H_0 sera rejetée (ie $\hat{\beta}_j$ sera considéré significativement différent de 0) au seuil de 95% si $w \geq 4$ (exactement $3,84 = (1,96)^2$) dans ce cas l'IC de l'OR $e^{\hat{\beta}_j}$ à 95% ne contient pas 1. ($e^{\hat{\beta}_j} \geq 1$: influence positive de X_j ; $e^{\hat{\beta}_j} \leq 1$: influence négative de X_j).

Déviance et qualité d'un modèle ajusté :

La vraisemblance du modèle réduit à la constante est notée $L(\beta_0)$ et la vraisemblance du modèle saturé est notée $L(\beta_p)$ (i.e. avec toutes les variables explicatives). La déviance est définie telle que :

$$L(\beta_k) = -2[\text{Log}L(\beta_k) - \text{Log}L(\beta_p)] = -2 \sum_{1 \leq i \leq n} [y^i \text{Log}(\pi(x^i)) + (1 - y^i) \text{Log}(1 - \pi(x^i))]$$

Remarque : dans le cas où $Y = 0$ ou $Y = 1$, $D(\beta_k) = -2\text{Log}[L(\beta_k)]$, car $L(\beta_p) = 1$.

Optimisation du modèle :

L'objectif de la régression est logistique est de maximiser la vraisemblance du modèle ajusté $L(\beta_k)$, ce qui équivaut à minimiser $D(\beta_k)$. On procède pas à pas (méthode « stepwise ») en comparant les modèles emboîtés, à partir du modèle de base. Pour un ensemble donné de variables avec k degrés de libertés ($k = \text{nombre des variables quantitatives} + \text{nombre des modalités des variables qualitatives}$), on cherche les coefficients β_k qui maximisent $L(\beta_k)$. Une fois trouvés et s'ils sont différents de 0, il faut savoir si l'on peut améliorer le modèle en lui ajoutant q variables explicatives; la différence des déviances est calculée :

$$D(\beta_k) - D(\beta_{k+q}) = -2[\text{Log}L(\beta_k) - \text{Log}L(\beta_{k+q})]$$

Sous l'hypothèse H_0 de la nullité des q derniers coefficients, $D(\beta_k) - D(\beta_{k+q})$ suit une loi $\chi^2(q)$. Sous le seuil critique de la valeur du χ^2 (si la probabilité dépasse 5%), les q nouvelles variables sont rejetées.

En pratique, une variable est ajoutée à chaque étape (1 degré de liberté si la variable est quantitative ou m degrés si la variable est qualitative à m modalités), et la quantité $D(\beta_0) - D(\beta_k)$ est calculée telle que :

$$D(\beta_0) - D(\beta_k) = -2[\text{Log}L(\beta_0) - \text{Log}L(\beta_k)]$$

Plus $D(\beta_0) - D(\beta_k)$ est grande i.e. $-2\text{Log}L(\beta_k)$ petit, meilleur est le modèle.

Deux autres critères de sélection sont utilisés :

- Le critère d'Akaïké : $AIC = -2\text{Log}L(\beta_k) + 2k$.
- Le critère de Schwarz : $SC = -2\text{Log}L(\beta_k) + k\text{Log}(n)$ qui est souvent noté BIC (Bayesian information criterion).

7.3.2 La mise en œuvre pour l'ancienneté 6

7.3.2.1 L'analyse des correspondances multiples (ACM) pour l'ancienneté 6

L'ACM est réalisée¹ sur le tableau disjonctif complet des variables sélectionnées pour l'étude, soit 6 variables et 13 modalités. L'analyse est limitée aux trois 1^{ers} axes, qui définissent les trois 1^{ers} plans factoriels de l'ACM. Les sommes des \cos^2 des modalités dans chaque plan permet de sélectionner le plan factoriel le plus représentatif pour cette modalité. Il est défini comme le plan dans lequel cette somme est maximale. Par ailleurs les centroïdes des 2 classes de la variable cible ont été positionnés, en variables supplémentaires :

- "AssurGrpeFinCour" (assurance du financement courant dans le contrat groupe) correspond à la modalité 0 de la variable Flag_Assurance, i.e. les contrats non rachetés dans la période étudiée
- "AssurExtFinCour" (assurance du financement courant dans un contrat externe) correspond à la modalité 1 de la variable Flag_Assurance , i.e. les contrats rachetés dans la période étudiée

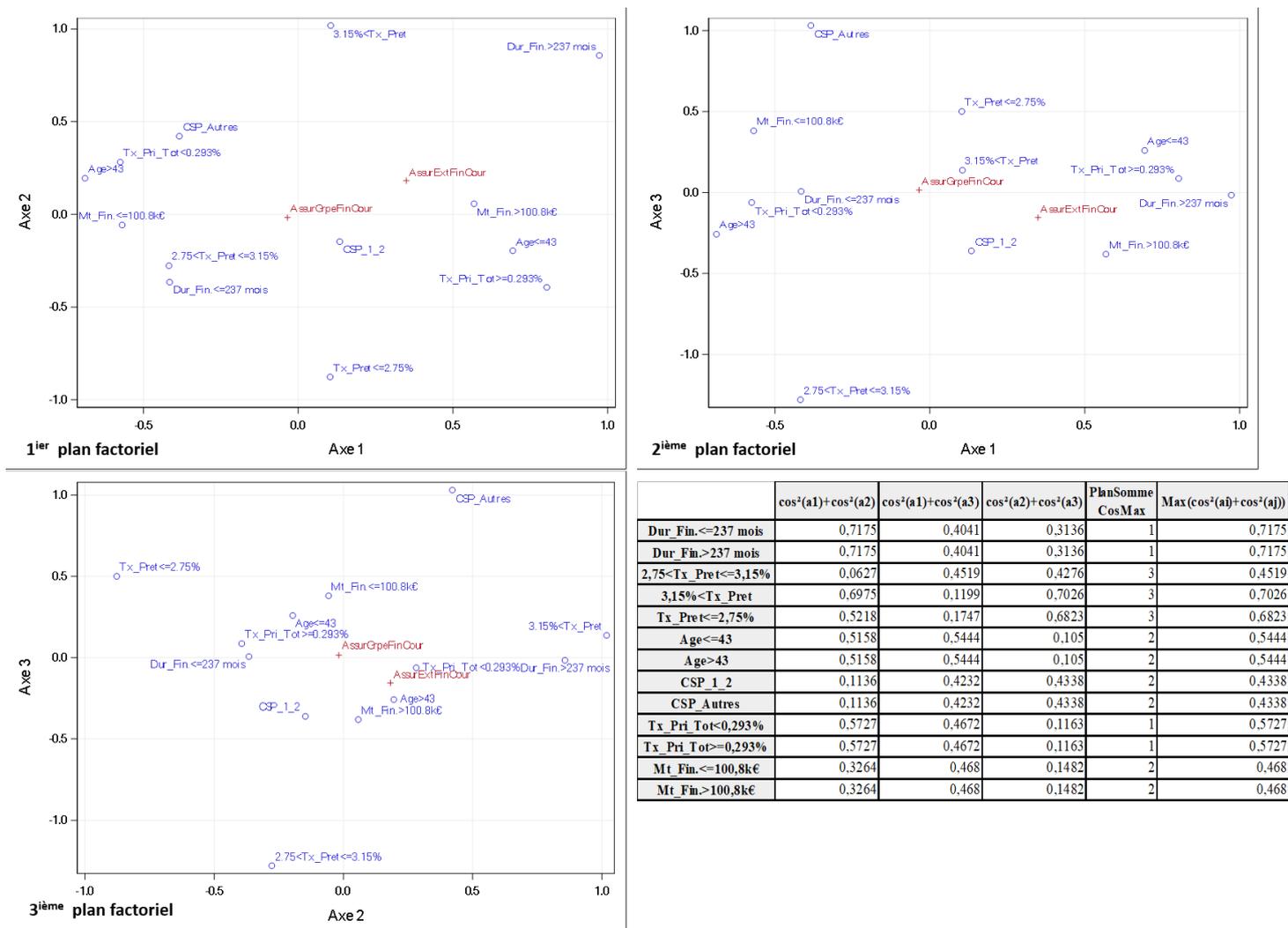


Fig. 7.29 – ACM - Les 3 1^{ers} plans factoriels pour l'ancienneté 6

1. proc SAS CORRESP

7.3.2.2 Méthode DISQUAL

Optimisation du modèle

Axes	AUC Apprentissage	AUC Validation
dim1 dim2 dim3 dim5 dim7 dim6	0.646	0.634
dim1 dim2 dim3 dim5 dim7	0.646	0.633
dim1 dim2 dim3 dim5	0.640	0.630
dim1 dim2 dim3	0.630	0.617
dim1 dim2	0.624	0.606
Dim1	0.617	0.599

Fig. 7.30 – Optimisation détaillée du modèle DISQUAL pour l'ancienneté 6

Calcul de la grille de score

Les scores déduits des coefficients sur les 4 axes sélectionnés sont les suivants (Figure 7.31) :

Obs	Variable	AssurExtFinCour	AssurGrpeFinCour	score
1	Constant	-2.51259	-0.09499	-2.41760
2	Dim1	0.65788	-0.06485	0.72273
3	Dim2	0.40312	-0.03973	0.44285
4	Dim3	-0.38140	0.03759	-0.41899
5	Dim5	-0.38814	0.03826	-0.42640

Fig. 7.31 – Fonctions de score par dimension, pour l'ancienneté 6 - DISQUAL

Puis les scores et nombres de points des modalités des variables initiales, (Figure 7.32) :

Obs	variable	_NAME_	Dim1	Dim2	Dim3	Dim5	Estimate	nbpoints
1	Age	Age>43	-0.68868	0.19488	-0.25757	0.35269	-0.45390	0
2	Age	Age<=43	0.69355	-0.19626	0.25939	-0.35519	0.45711	15
3	Code_CSP	CSP_Autres	-0.38385	0.42146	1.03087	0.62252	-0.78814	0
4	Code_CSP	CSP 1 2	0.13427	-0.14742	-0.36059	-0.21775	0.27569	17
5	Duree_Init_Financement_Courant	Dur_Fin.<=237 mois	-0.41520	-0.36572	0.00690	0.03279	-0.47891	0
6	Duree_Init_Financement_Courant	Dur_Fin.>237 mois	0.97307	0.85710	-0.01617	-0.07684	1.12237	26
7	Montant_Financement_Courant	Mt_Fin.<=100.8k€	-0.56848	-0.05733	0.38073	-0.68934	-0.30183	0
8	Montant_Financement_Courant	Mt_Fin.>100.8k€	0.56840	0.05732	-0.38068	0.68925	0.30179	10
9	Taux_Cotisation	Tx_Pri_Tot>0.293%	-0.57486	0.28168	-0.06191	0.07695	-0.29759	0
10	Taux_Cotisation	Tx_Pri_Tot<=0.293%	0.80330	-0.39362	0.08651	-0.10753	0.41586	12
11	Taux_Pret	Tx_Pret<=2.75%	0.10327	-0.87718	0.50055	0.36584	-0.67954	0
12	Taux_Pret	2.75<Tx_Pret<=3.15%	-0.41757	-0.27705	-1.27933	-0.29088	0.23558	15
13	Taux_Pret	3.15%<Tx_Pret	0.10512	1.01850	0.13734	-0.22161	0.56397	20

Fig. 7.32 – Nombre de points par modalité, pour l'ancienneté 6 - DISQUAL

7.3.2.3 La régression logistique

Sélection des variables et optimisation du modèle

La procédure SAS LOGISTIC associée à l'option STEPWISE conduit à sélectionner les 6 variables suivantes (Figure 7.33) :

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	Duree_Init_Financeme		1	1	154.6999		<.0001
2	Taux_Pret		2	2	78.4910		<.0001
3	Age		1	3	96.5616		<.0001
4	Code_CSP		1	4	56.6889		<.0001
5	Montant_Financement_		1	5	33.7119		<.0001
6	Taux_Cotisation		1	6	24.1078		<.0001

Fig. 7.33 – Selection stepwise des variables, pour l'ancienneté 6 - LOGIT

Les critères d'ajustement comme ceux d'Akaïke (AIC), de schwarz (SC) ou le -2Log-Vraisemblance (-2Log L) ci-dessous ne suggèrent pas de simplification du modèle (Figure 7.34) :

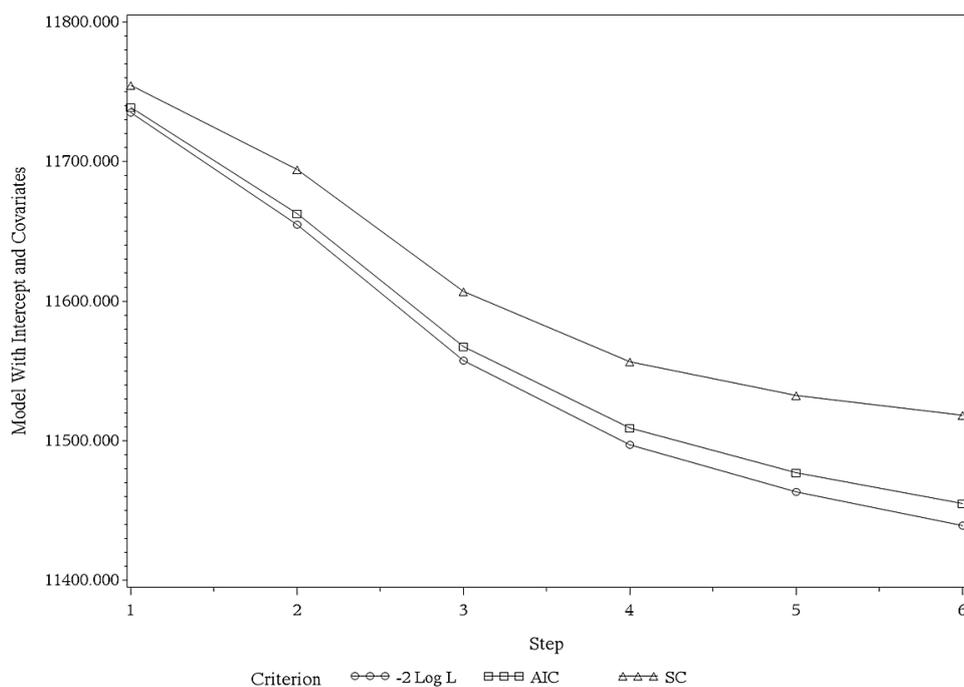


Fig. 7.34 – Statistiques d'ajustement par step du modèle – ancienneté 6

Les 6 étapes de la construction sont les suivantes (Figure 7.35) :

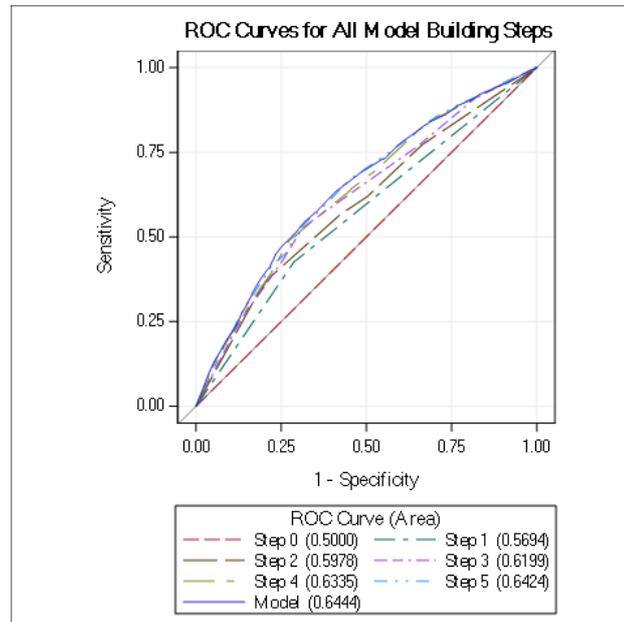


Fig. 7.35 – Etapes de construction du modèle pour l'ancienneté 6 - LOGIT

Calcul de la grille de score

La grille de score est la suivante (Figure 7.36) :

Obs	Variable	ClassVal0	Estimate	nbpoints
1	Age	Age>43	0	0
2	Age	Age<=43	0.4157	19
3	Code_CSP	CSP_Autres	0	0
4	Code_CSP	CSP_1_2	0.4156	19
5	Duree_Init_Financeme	Dur_Fin.<=237 mois	-0.2101	0
6	Duree_Init_Financeme	Dur_Fin.>237 mois	0	9
7	Intercept		-3.1794	0
8	Montant_Financement_	Mt_Fin.<=100.8k€	-0.2235	0
9	Montant_Financement_	Mt_Fin.>100.8k€	0	10
10	Taux_Cotisation	Tx_Pri_Tot>0.293%	0	0
11	Taux_Cotisation	Tx_Pri_Tot<=0.293%	0.2542	11
12	Taux_Pret	Tx_Pret<=2.75%	0	0
13	Taux_Pret	2.75<Tx_Pret<=3.15%	0.5505	25
14	Taux_Pret	3.15<Tx_Pret	0.7077	32

Fig. 7.36 – Nombre de points par modalité, pour l'ancienneté 6 - LOGIT

7.3.3 V de Cramer

L'objectif est de mesurer la liaison entre deux variables X et Y discrètes ou qualitatives. Une 1^{ère} étape consiste à déterminer les tableaux de contingence $X \times Y$ (à l lignes et c colonnes) sur la population (n individus) étudiée :

- Le tableau des effectifs observés ($T1$)
- Le tableau des effectifs théoriques ($T2$) (obtenu à partir des fréquences marginales $f_{i,j} = nf_i f_j$)

Avec les notations précédentes, χ_{max}^2 satisfait :

$$\chi_{max}^2 = n[Min(l, c) - 1]$$

Le V de Cramer est défini tel que :

$$V \stackrel{\text{déf}}{=} \sqrt{\frac{\chi^2}{\chi_{max}^2}}$$

V varie entre 0 (liaison nulle) et 1 (liaison maximale) et permet d'évaluer directement l'intensité de la liaison entre les deux variables. En outre, il ne dépend ni de la taille de la table, ni de la taille de l'échantillon.

7.3.4 Lambda de Wilks

Dans ce qui suit n désigne le nombre d'individus et p le nombre de variables du modèle considéré.

C'est le rapport

$$\Lambda \stackrel{\text{déf}}{=} \frac{\det(W)}{\det(V)}$$

Où W est la matrice des covariances intra-classe et V la matrice des covariances totale. Il varie entre 0 et 1. Plus il est bas meilleur est le modèle. A l'opposé si $\Lambda = 1$, tous les centroïdes sont égaux. Ce rapport permet de répondre à deux questions aux sujets des variables explicatives d'un modèle :

- Permettent-elles de discriminer les classes (Q1) ?
- Pourraient-elles être discriminées avec moins de variables (Q2) ?

Q1 :

dans le cas de deux classes, il peut être montré que $\frac{(1-\Lambda)}{\Lambda}$ suit une loi de *Fisher*($p, n-p-1$). A cette statistique de Fisher est associée une probabilité qui si elle est inférieure à un certain seuil (fixé entre 5% et 20%) permet de rejeter l'hypothèse d'égalité des centroïdes et caractérise la qualité du modèle

Q2 :

La réponse est apportée de manière itérative. En posant $\Lambda_0 = 1$, la variable donnant le plus petit lambda de wilks (noté Λ_1) est cherchée parmi tous les modèles réduits à une seule variable. A partir de l'hypothèse (étape p) que p variables minimisant le lambda de wilks Λ_p , parmi tous les modèles à p variables sont sélectionnées, le passage de l'étape p à $(p+1)$ s'effectue en sélectionnant la $(p+1)^e$ qui minimise le lambda de Wilks Λ_{p+1} parmi tous les modèles à $(p+1)$ variables.

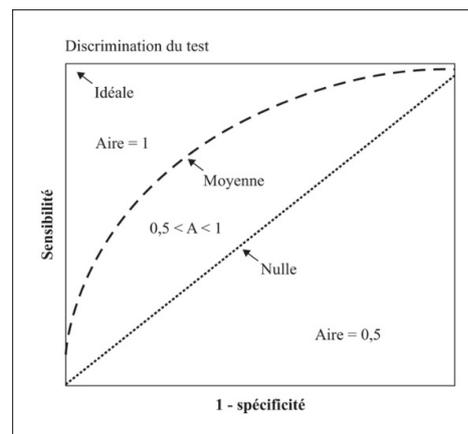
Critère d'arrêt de l'itération utilisé : le ratio (pour deux classes) $(n-2-p)(\frac{\Lambda_p}{\Lambda_{p+1}} - 1)$ suit une loi de *Fisher*($1, n-2-p$). Il est utilisé pour tester l'hypothèse H_0 selon laquelle la $(p+1)^e$ variable n'améliore pas le modèle. Si aucune variable ne fournit un Λ_{p+1} suffisamment faible pour que le ratio ci-dessus soit suffisamment grand pour que la probabilité associée soit inférieure à un certain seuil², l'itération est arrêtée.

2. 15% par défaut dans la procédure SAS STEPDISC mise en œuvre dans le choix des axes de l'ACM pour l'analyse discriminante

7.3.5 Courbe ROC

Le pouvoir discriminant d'un modèle de score est visualisé à l'aide d'une courbe appelée courbe ROC (Receiver operating characteristic). Elle représente - sur l'axe des ordonnées la probabilité de bien détecter un évènement au seuil s ($\alpha(s)$ - % de vrais positifs) en fonction - sur l'axe des abscisses, de la probabilité de mal détecter un non-évènement toujours au seuil s ($(1 - \beta(s))$ - % de faux positifs), lorsque l'on fait varier le seuil de séparation du score allant du maximum (tous les individus sont considérés comme non-évènements, $\alpha(s) = 0, \beta(s) = 1$) au minimum (tous les individus sont considérés comme évènements, $\alpha(s) = 1, \beta(s) = 0$). $\alpha(s)$ et $\beta(s)$ satisfont :

- $\alpha(s) = \mathbb{P}(\text{score}(x) \leq s | x) = \text{évènement}$
- $\beta(s) = \mathbb{P}(\text{score}(x) < s | x) = \text{non - évènement}$



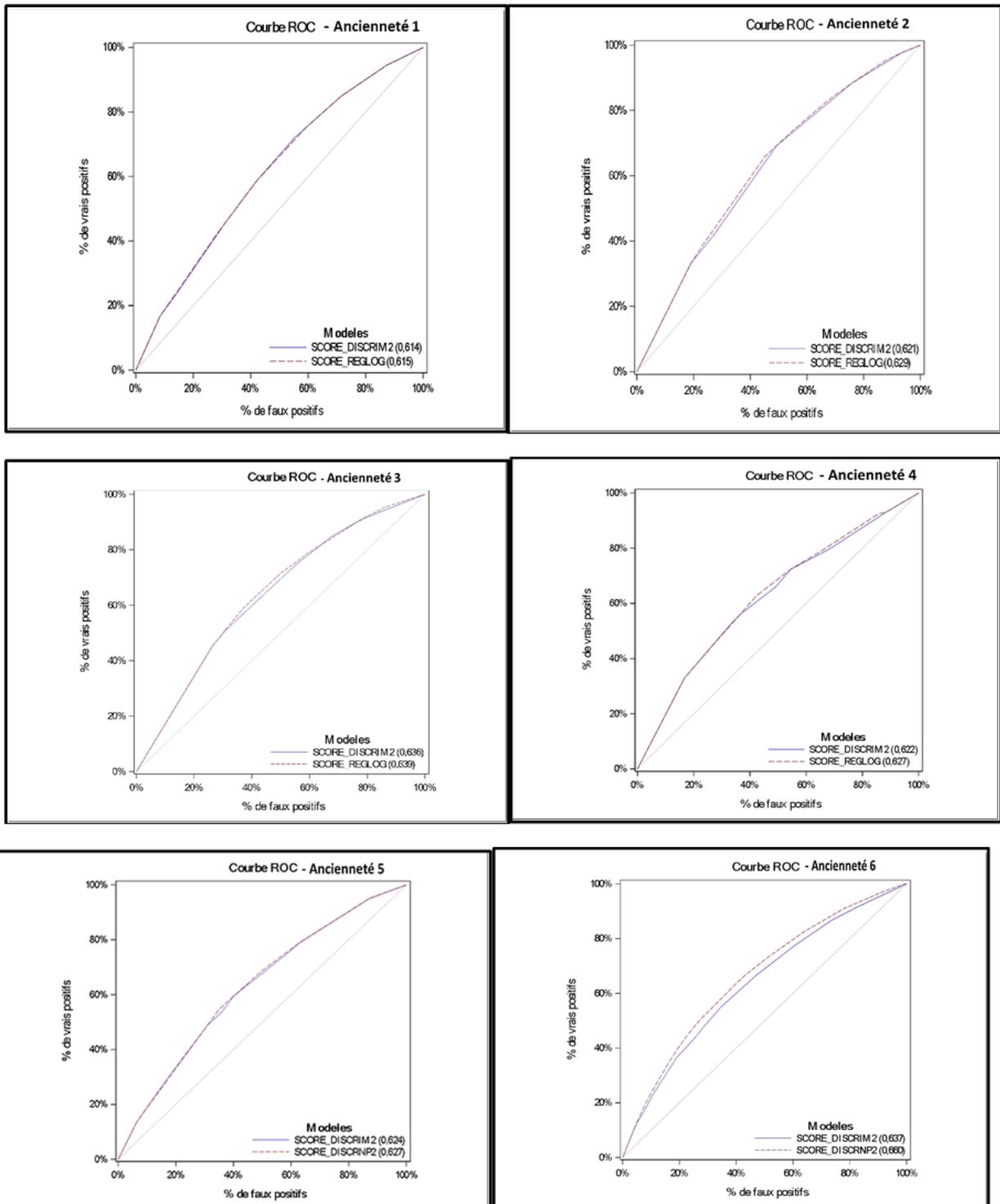
Plus la courbe s'approche du coin supérieur gauche de la figure ci-dessus meilleur est le modèle. Un modèle ayant une courbe correspondant à la diagonale ne classe pas mieux que le hasard.

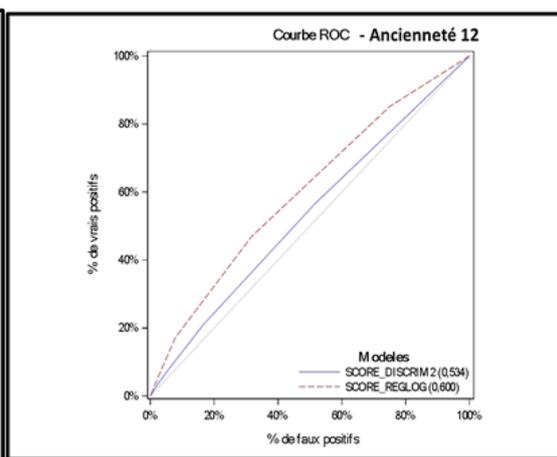
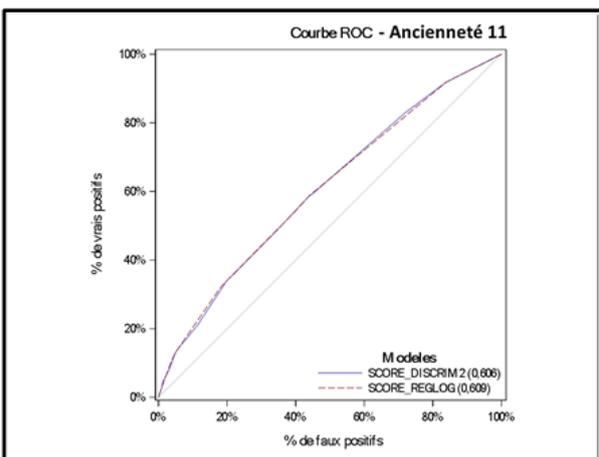
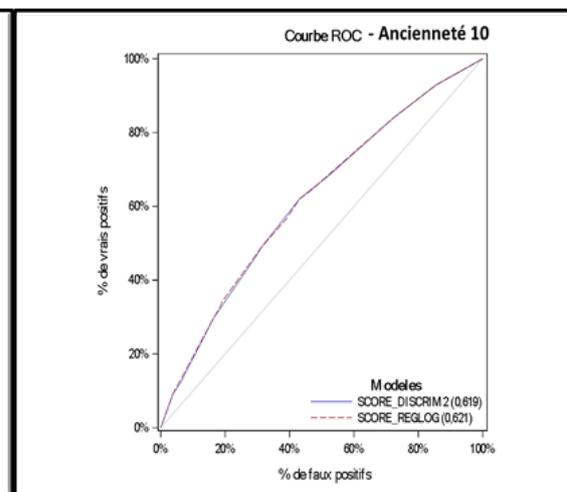
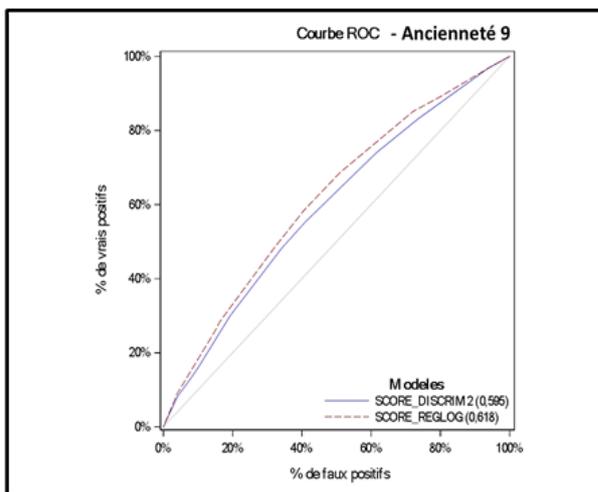
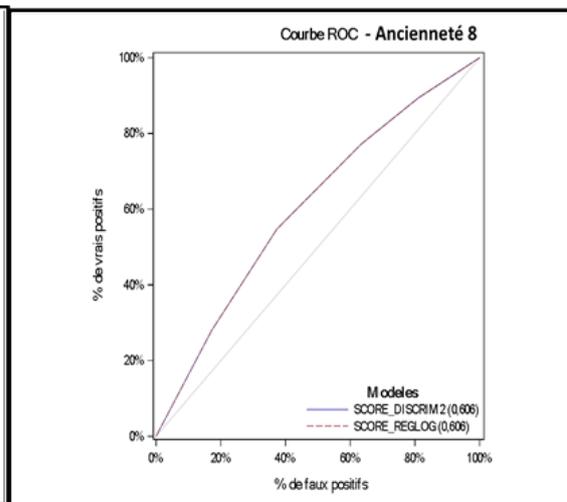
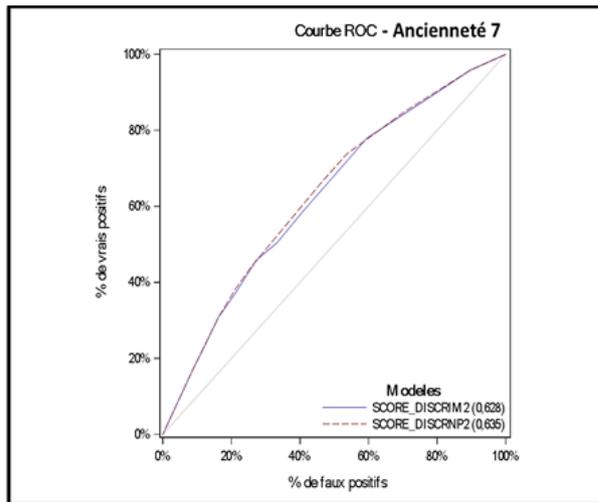
7.3.6 AUC et interprétation

AUC	Interpretation
$UC=0.5$	No discrimination, e.g., randomly flip a coin
$0.6 \geq AUC > 0.5$	Poor discrimination
$0.7 \geq AUC > 0.6$	Acceptable discrimination
$0.8 \geq AUC > 0.7$	Excellent discrimination
$AUC > 0.9$	Outstanding discrimination

Fig. 7.37 – AUC et interprétation

7.3.7 Synthèse des courbes ROC par anciennetés des méthodes DISQUAL et LOGIT





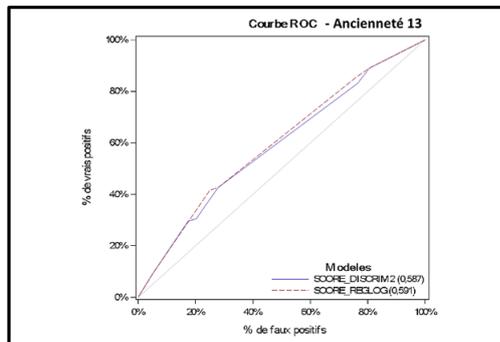


Fig. 7.38 – Courbes ROC DISQUAL et LOGIT par ancienneté

7.3.8 La grille de score

Un nombre de points $N_{i,j}$ compris entre 0 et 100 est défini comme suit, pour chaque modalité j de la variable i : A partir des coefficients $\{C_{i,j}\}_{i,1 \leq j \leq l(i)}$, associés aux variables $\{X_i\}_{1 \leq i \leq n}$ du modèle et où l_i est le nombre de modalités de X_i utilisées dans le modèle. Pour chaque modalité j de X_i , $N_{i,j}$ est défini comme suit :

$$N_{i,j} = 100 \frac{C_{i,j} - \text{Min}_{1 \leq k \leq l_i}(C_{i,k})}{\sum_n [\text{Max}_{1 \leq k \leq l_n}(C_{n,k}) - \text{Min}_{1 \leq k \leq l_n}(C_{n,k})]}$$

7.3.9 Le lien entre les taux de rachats et les scores

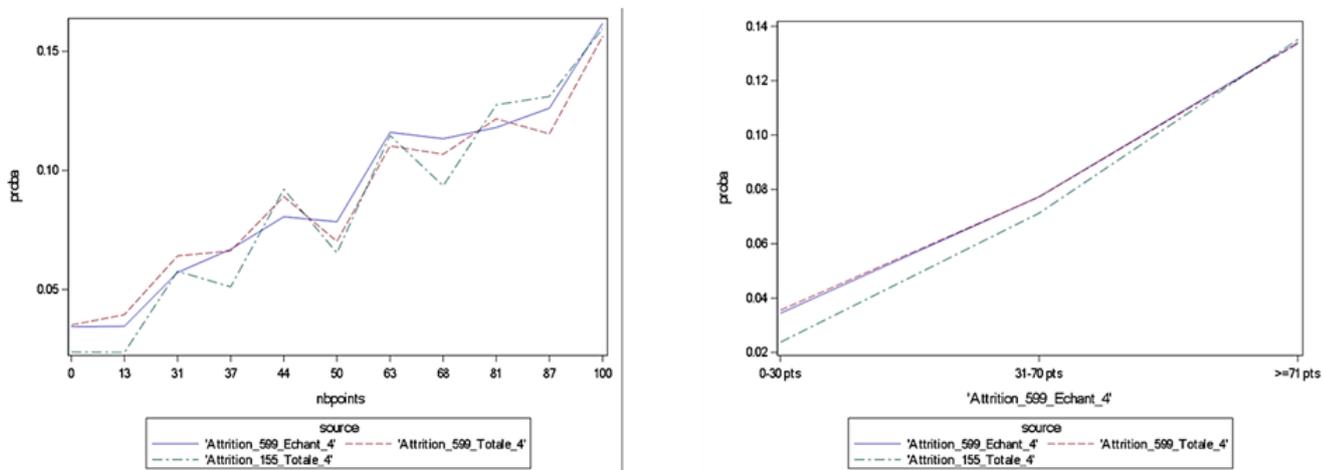


Fig. 7.39 – Représentation graphique des probabilités de rachats en fonction des scores pour l’ancienneté 5

7.3.10 Calcul des facteurs d'ajustement des seuils des inducteurs

7.3.10.1 L'ajustement des seuils du modèle

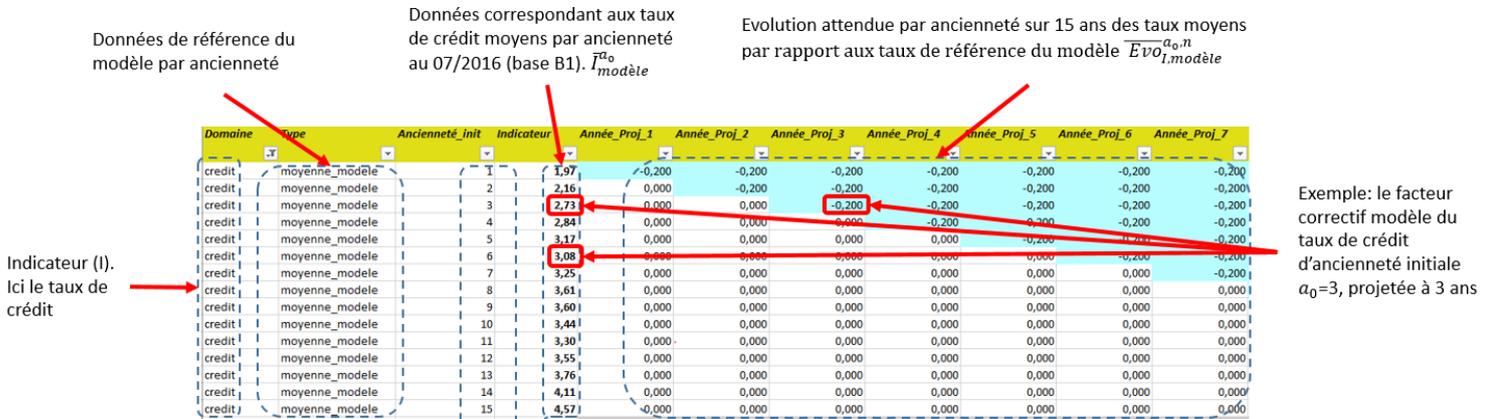


Fig. 7.40 – Exemple de calcul de facteur correctif modèle – cas $a_0 = n = 3$

7.3.10.2 L'ajustement des seuils de tests

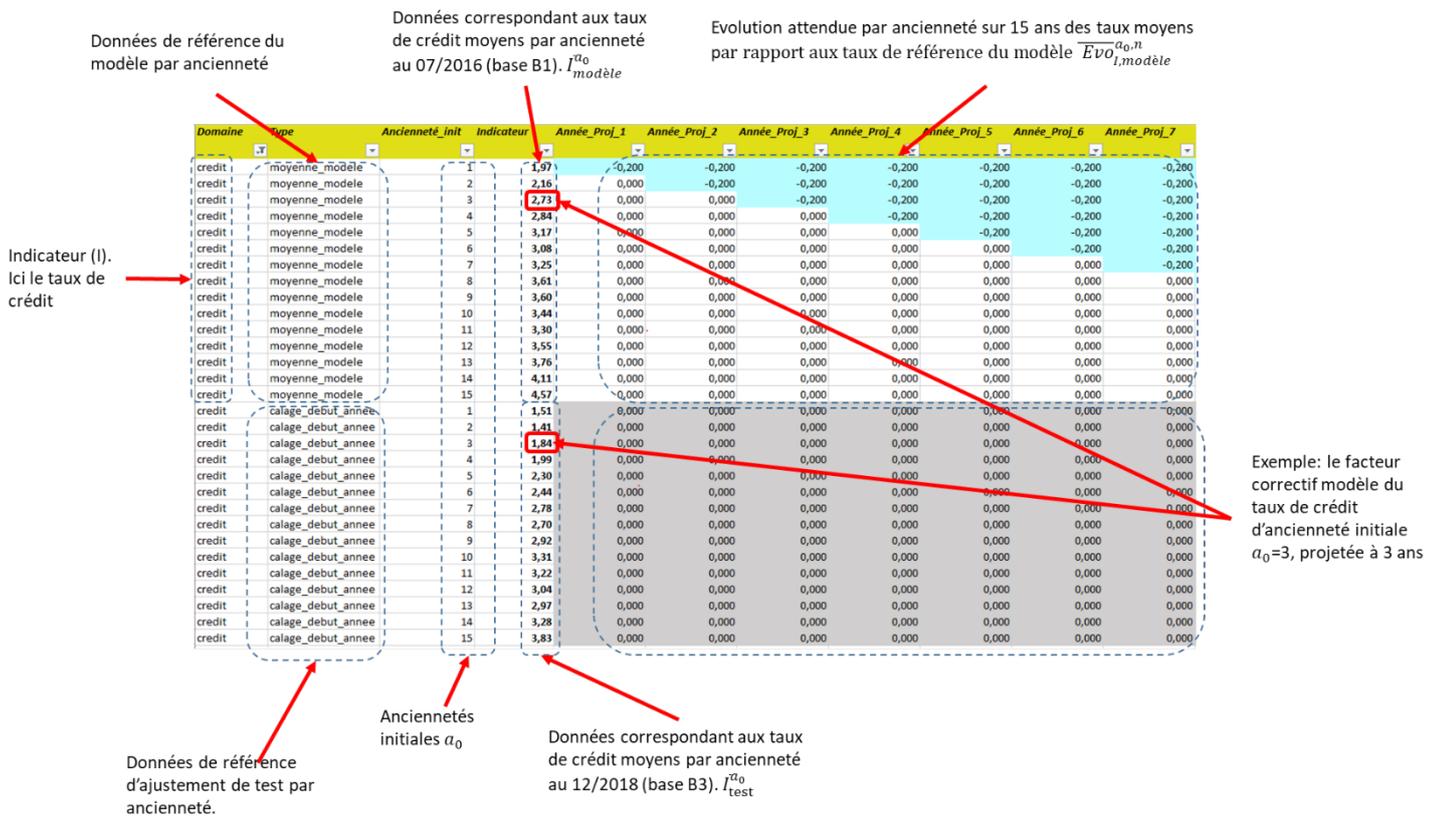


Fig. 7.41 – Exemple de calcul de facteur correctif test – cas $a_0 = n = 3$

7.3.11 Principe de calcul des taux bruts : l'estimateur de Hoem

L'estimation de la probabilité du rachat q_x d'un contrat entre l'ancienneté x et $x + 1$ à partir des données statistiques, est obtenue en divisant le nombre de rachats observés à cette ancienneté d_x par la population de contrats exposée au risque l_x :

$$q_x = \frac{d_x}{l_x}$$

Le diagramme de Lexis permet de représenter les différents cas possibles pour l'estimation des probabilités de rachat, et plus précisément de l'exposition au risque. Dans ce diagramme, chaque contrat sous observation est représenté par une diagonale et les axes représentent l'évolution du temps et de l'ancienneté. Dans la figure suivante (Figure 7.42), plusieurs cas, non exhaustifs, sont représentés :

- Le contrat est observé complètement entre les anciennetés x et $x + 1$.
- Le contrat est observé alors qu'il a une ancienneté supérieure à x , il quitte l'effectif avant d'avoir atteint l'ancienneté $x + 1$: le contrat est dit censuré sur la classe d'ancienneté $[x, x + 1[$.
- Le contrat entre dans l'effectif pendant la période d'observation alors que son ancienneté est déjà comprise entre x et $x + 1$: le contrat est dit tronqué sur la classe d'ancienneté $[x, x + 1[$.
- Même cas de figure que 3 et le contrat quitte l'effectif (rachat ou sinistre) avant d'avoir atteint l'ancienneté $x + 1$.

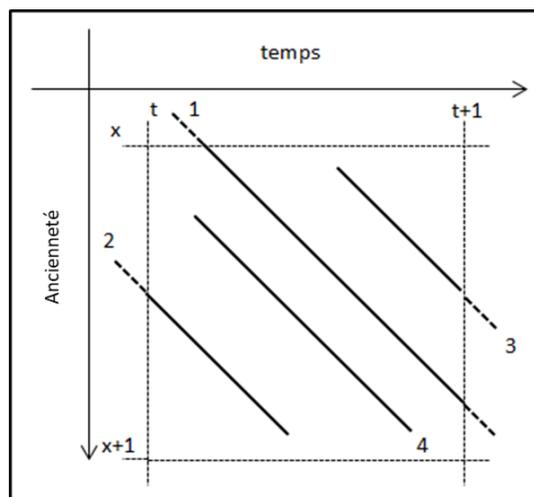


Fig. 7.42 – Diagramme de Lexis

Afin de prendre en compte ces contraintes et pour bien estimer la probabilité de rachat à l'ancienneté x , l'estimateur des moments de Hoem est utilisé. En effet, cet estimateur est le plus couramment utilisé avec l'estimateur de Kaplan Meier. Néanmoins, ce dernier peut présenter des traitements lourds lorsqu'il y a un nombre important de données car il demande de connaître la chronologie des rachats sur le portefeuille ainsi que le nombre de censures entre deux mouvements.

L'estimateur de Hoem est donc retenu :

$$\hat{q}_x = \frac{d_x}{er_x}$$

La quantité d_x correspond au nombre de contrats rachetés à l'ancienneté x .

Le terme er_x est calculé selon la formule suivante :

$$er_x = \sum_{j=1}^{n_x} er_x^{(j)}$$

où n_x désigne le nombre de contrats d'ancienneté x observés et $er_x^{(j)}$ correspond à l'exposition au risque du contrat j , elle est égale à la fraction d'année vécue par le contrat pendant l'observation.

Remarque : dans l'étude la durée d'observation est d'un an et tous les contrats d'ancienneté x observés sont présents en début de période d'observation (cas 1 et 2 du diagramme de Lexis 7.42, ci-dessus)

7.3.12 Evolution des drivers d'ajustement des seuils de juillet 2015 à juillet 2019

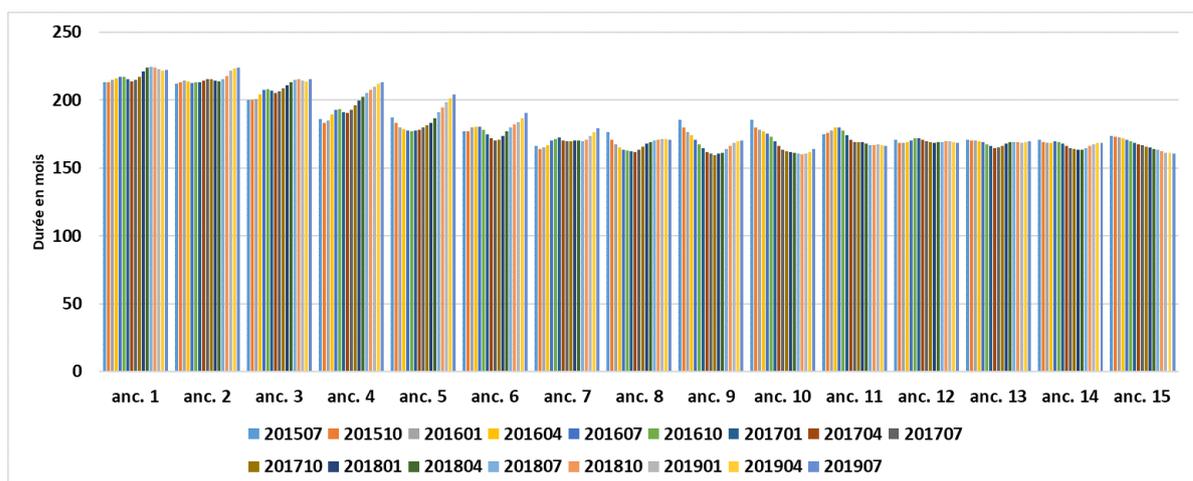


Fig. 7.43 – Moyenne des durées de crédits par ancienneté (yc sorties et réaménagements) [07/2015 ->07/2019]

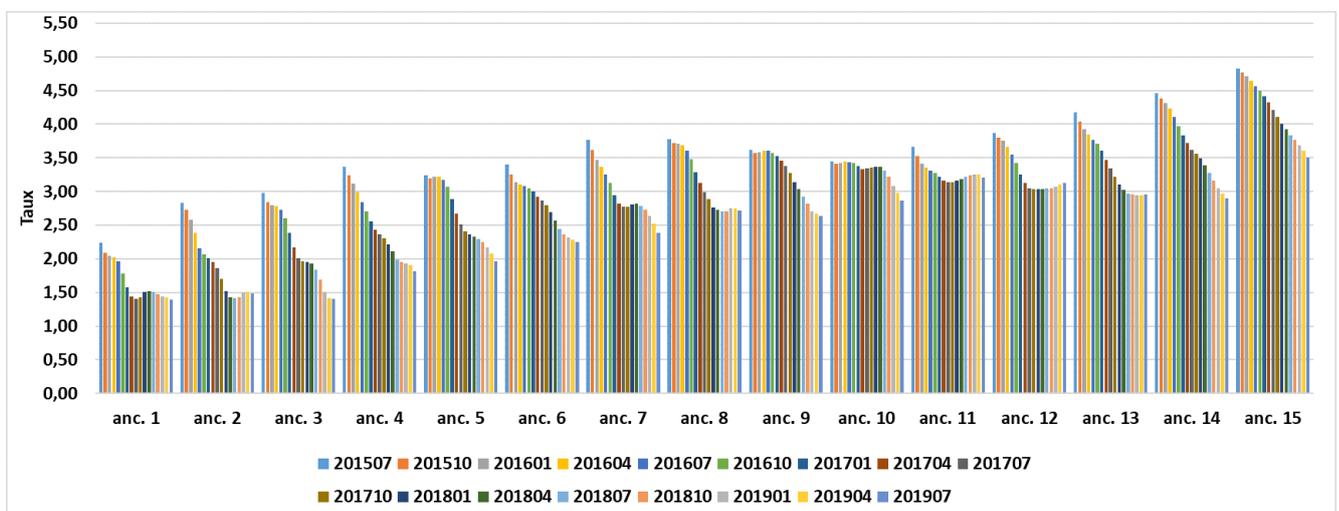


Fig. 7.44 – Moyenne des taux de crédits par ancienneté (yc sorties et réaménagements) [07/2015 -> 07/2019]

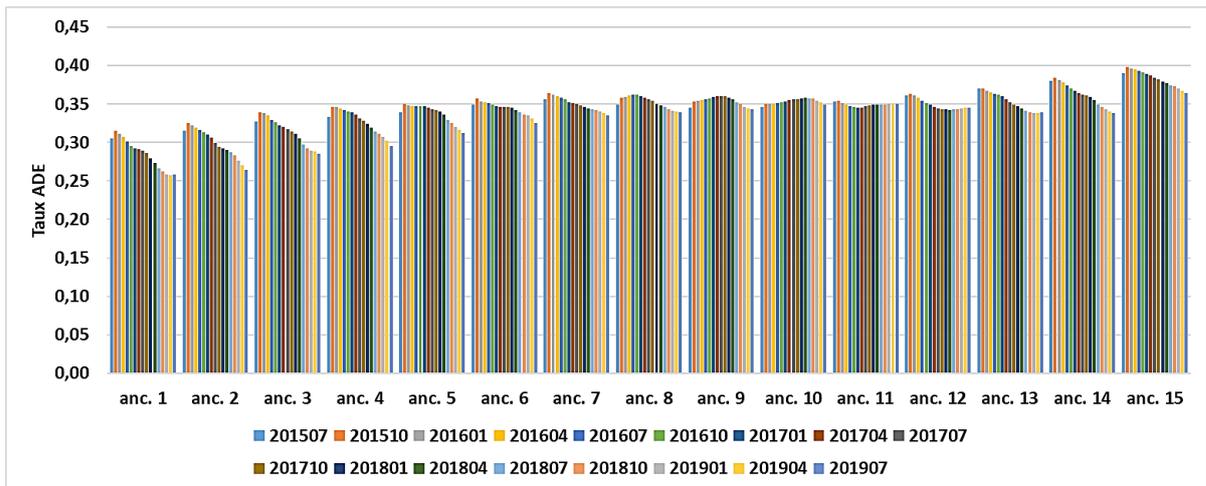


Fig. 7.45 – Moyenne des taux ADE par ancienneté (yc sorties et réaménagements) [07/2015 -> 07/2019]

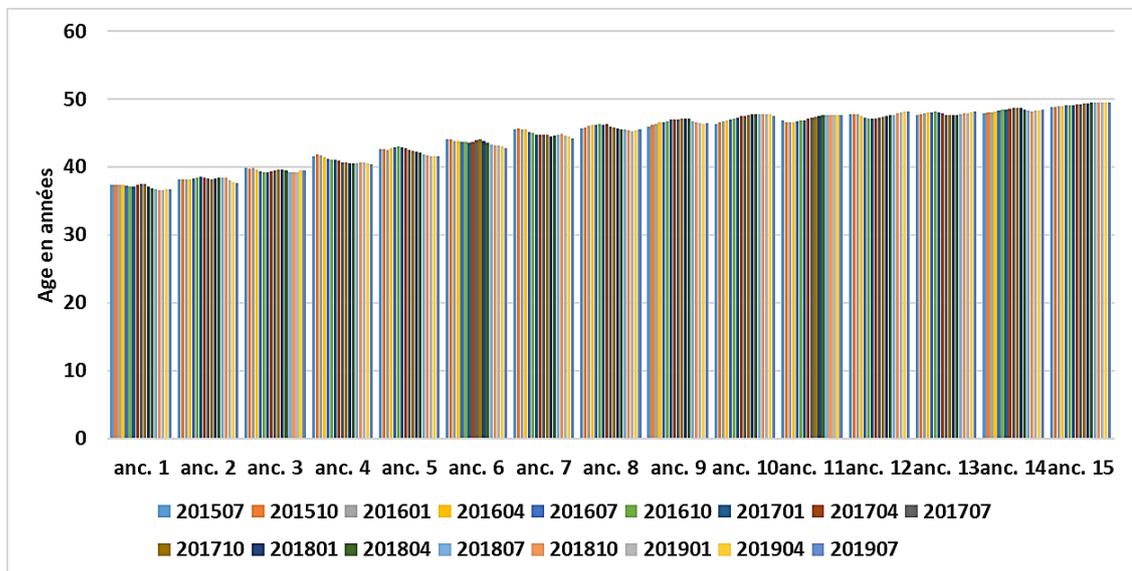


Fig. 7.46 – Moyenne des âges par ancienneté (yc sorties et réaménagements) [07/2015, 07/2019]

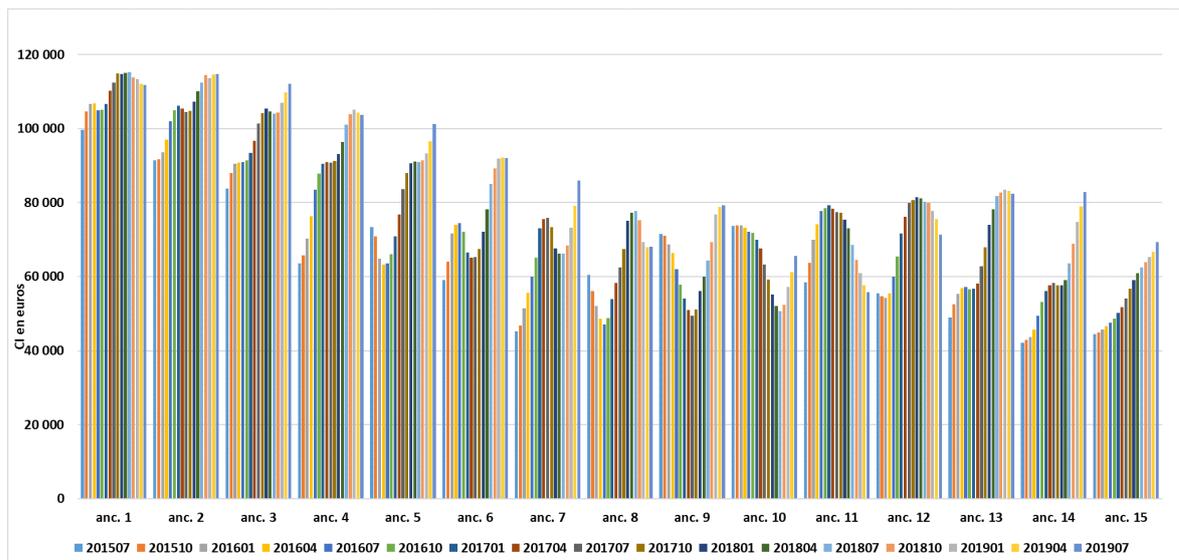
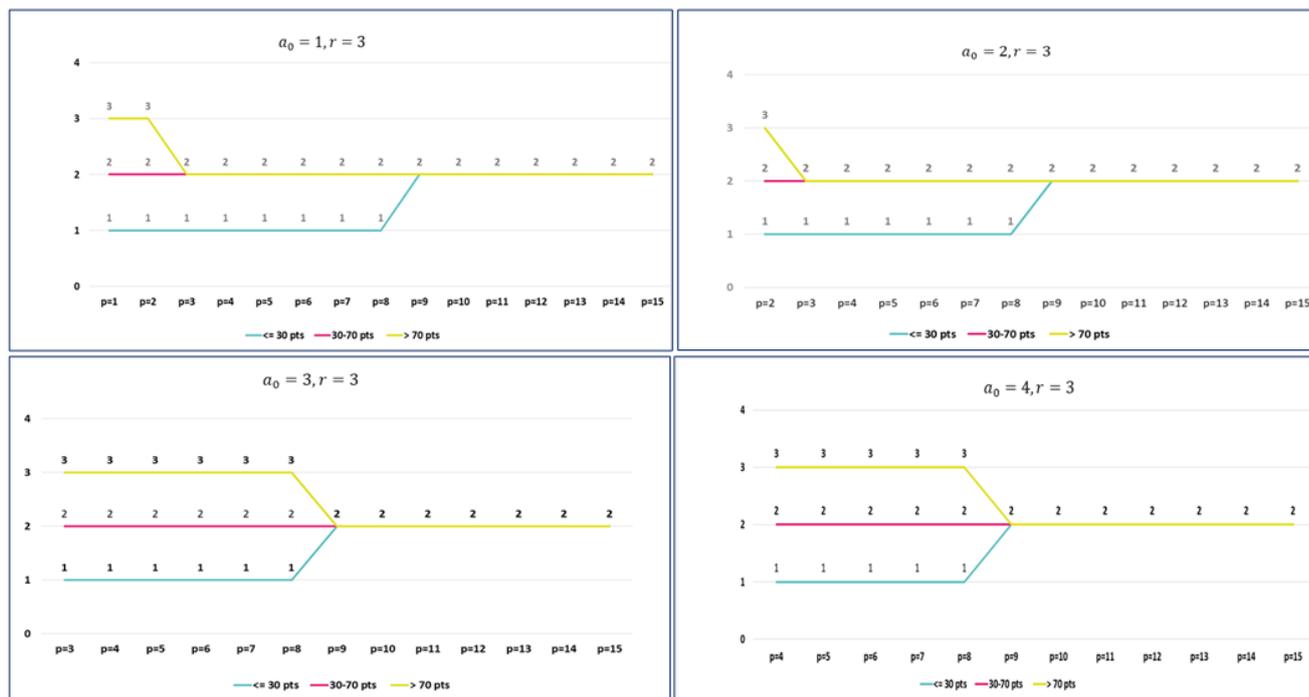


Fig. 7.47 – Moyenne des capitaux initiaux (CI) par ancienneté (yc sorties et réaménagements) [07/2015, 07/2019]

7.3.13 Trajectoires modélisées par anciennetés pour une segmentation à 3 classes de risques



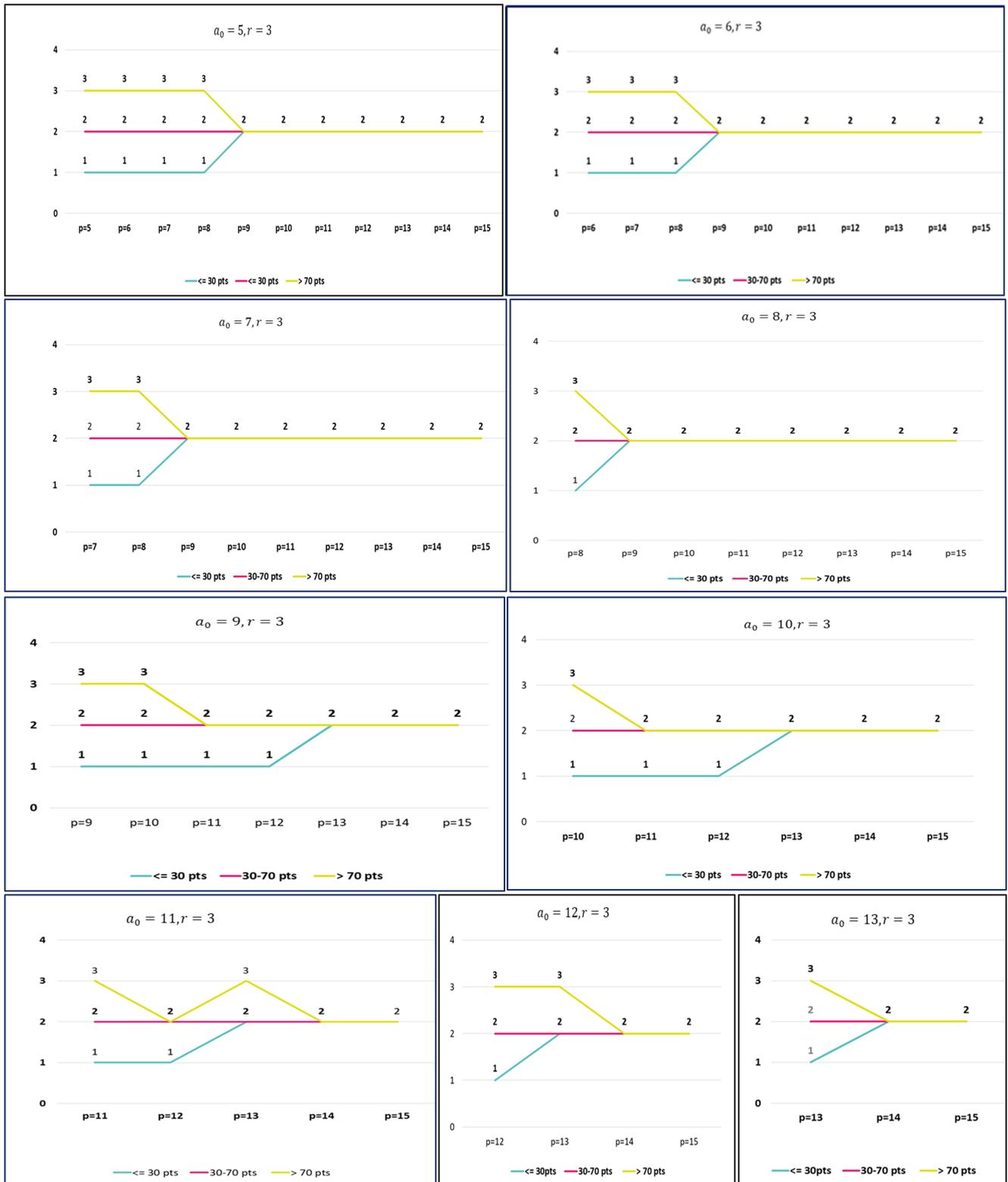


Fig. 7.48 – Trajectoires de risque par anciennetés, pour $r = 3$

7.3.14 La matrice de transition

Ancienneté	Etat début période	Transition	Proba. Transition	Taux brut Rachat (a + 1)	Etat fin période moyen	Etat fin période	Taux de rachat brut moyen fin période $txrmf_{i,1 \leq i \leq r}^{a+1}$	Effectif final (a + 1)	Effectif moyen Fin période $Lxmf_{i,1 \leq i \leq r}^a$
$a, 1 \leq a \leq 15$	$i, 1 \leq i \leq n$	$i \rightarrow 1$	$p_{i,1}^a$	txr_1^{a+1}	$E_i^a = \sum_{j=1}^r jp_{i,j}^a$	ef_i^a	$txmf_i^{a+1} = \sum_{j=1}^r txr_j^{a+1} p_{i,j}^a$	Lx_1^{a+1}	$Lxmf_i^{a+1} = \sum_{j=1}^r Lx_j^{a+1} p_{i,j}^a$
		$i \rightarrow j, 1 < j < r$	$p_{i,j}^a$	txr_j^{a+1}				Lx_j^{a+1}	
		$i \rightarrow r$	$p_{i,r}^a$	txr_r^{a+1}				Lx_r^{a+1}	

Avec $ef_i^a = Ent(E_i^a) \mathbb{1}_{E_i^a - Ent(E_i^a) \leq Ent(E_i^a) + 1 - Ent(E_i^a)} + (Ent(E_i^a) + 1) \mathbb{1}_{E_i^a - Ent(E_i^a) > Ent(E_i^a) + 1 - Ent(E_i^a)}, 1 \leq i \leq r$

Fig. 7.49 – Présentation des transitions

AncienneteADEGrp	AncienneteADEGrpj	Etat début période	nbpoints_Fin	Transition	Proba. Transition	Taux brut Rachat (a+1)	Etat fin période moyen	Etat final	Taux de rachat brut moyen fin période	Effectif final (a+1)	Effectif moyen Fin période
0	5	1	1	11	0,907	0,055	1,1	1	0,057	2613	2396
0	5	1	2	12	0,093	0,085	1,1	1	0,057	267	2396
0	5	2	1	21	0,025	0,055	2,2	2	0,098	3194	79362
0	5	2	2	22	0,746	0,085	2,2	2	0,098	97103	79362
0	5	2	3	23	0,230	0,145	2,2	2	0,098	29925	79362
0	5	3	1	31	0,001	0,055	2,7	3	0,126	275	195258
0	5	3	2	32	0,312	0,085	2,7	3	0,126	106989	195258
0	5	3	3	33	0,687	0,145	2,7	3	0,126	235574	195258

Fig. 7.50 – Exemple, pour $r = 3$ et $a = 6$ ($a_0 = 1$ et $n = 5$)

7.3.15 La table des trajectoires

AncienneteADEGrp	nbpoints_Deb	...	nbpoints_Deb_5	ProbaResil_Deb_5	Effectif_Deb_5	ProbaResilEffectif_Moy_5	Effectif_Deb_moy_5	...
1	1	...	2	0,108	130 677	0,099	122 562	...
1	2	...	2	0,108	130 677	0,099	122 562	...
1	3	...	3	0,146	278 380	0,141	144 342	...
2	1	...	3	0,126	159 842	0,141	144 342	...
2	2	...	3	0,126	159 842	0,141	144 342	...
2	3	...	3	0,149	83 765	0,141	144 342	...
3	1	...	2	0,098	137 940	0,099	122 562	...
3	2	...	2	0,098	137 940	0,099	122 562	...
3	3	...	3	0,132	39 971	0,141	144 342	...
4	1	...	1	0,060	17 959	0,052	27 645	...
4	2	...	2	0,092	97 294	0,099	122 562	...
4	3	...	3	0,126	22 377	0,141	144 342	...
5	1	...	1	0,048	37 331	0,052	27 645	...
5	2	...	2	0,084	130 497	0,099	122 562	...
5	3	...	3	0,153	49 285	0,141	144 342	...

Fig. 7.51 – Extrait de la table des trajectoires - $r = 3, p = 5$ et $1 \leq a_0 \leq 5$

7.3.16 Principe du Lissage des lois par la méthode whittaker-Henderson

- La plage de lissage retenue.

Un lissage est effectué sur un intervalle d'anciennetés $[x_1, x_n]$, où les données brutes sont considérées comme fiables, c'est-à-dire où il y a suffisamment d'effectifs sous risque

- La méthode whittaker-Henderson.

La méthode de Whittaker-Henderson consiste à minimiser la somme pondérée d'un critère de fidélité F et d'un critère de lissage S_k , $Min_g(F(g) + aS_k(g))$.

- Le critère de fidélité se définit comme suit :

$$F(g) = (\bar{g} - \bar{h})^t W (\bar{g} - \bar{h})$$

avec

- $\bar{h} = (h(x_i))_{1 \leq i \leq n}$ avec $h(x_i) = \hat{q}_{x_i}$, \hat{q}_{x_i} étant le taux brut de rachat du contrat d'ancienneté x_i , $1 \leq i \leq n$

- $\bar{g} = (g(x_i))_{1 \leq i \leq n}$, la solution du problème

$$\cdot W = \begin{bmatrix} w_{x_1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{x_n} \end{bmatrix} \text{ avec les poids } w_{x_i} = \frac{er_{x_i}}{\sum_{1 \leq j \leq n} er_{x_j}}, \text{ } er_{x_i} \text{ étant l'exposition}$$

au risque des contrats d'ancienneté x_i

- Le critère de régularité est calculé comme suit :

$$S_k(g) = \bar{g}^t (\mathcal{D}^{(k)t} \mathcal{D}^{(k)}) \bar{g}$$

avec :

- $\mathcal{D}^{(k)} \bar{g} = (\Delta^{(k)} g)(x_i)_{1 \leq i \leq k}$ et

- $(\Delta^{(k)} g)(x_i) \stackrel{\text{déf}}{=} \frac{((\Delta^{(k-1)} g)(x_{i+1}) - (\Delta^{(k-1)} g)(x_i))}{(x_{i+1} - x_i)}$ et $(\Delta^{(0)} g)(x_i) \stackrel{\text{déf}}{=} g(x_i)$

7.3.16.1 Les résultats du lissage pour les segmentations à 3 et 4 classes de risques

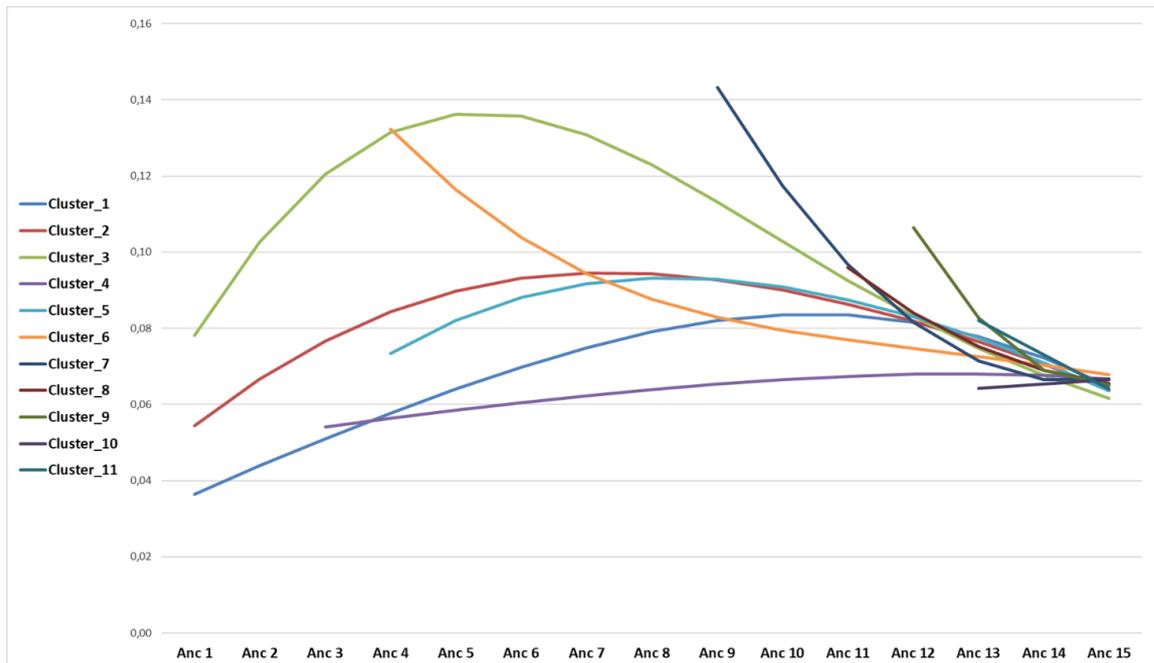


Fig. 7.52 – Taux lissés par clusters et anciennetés ($r = 3$)

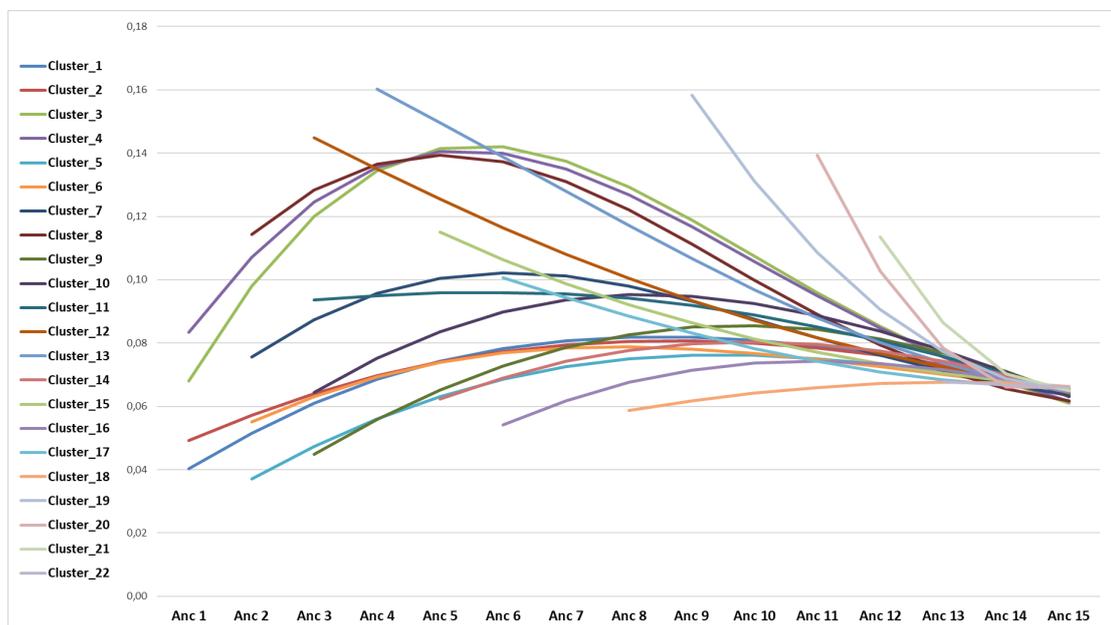


Fig. 7.53 – Taux lissés par clusters et anciennetés ($r = 4$)

7.4 Le back-testing

7.4.1 L'application de la grille de score en 1^{re} année

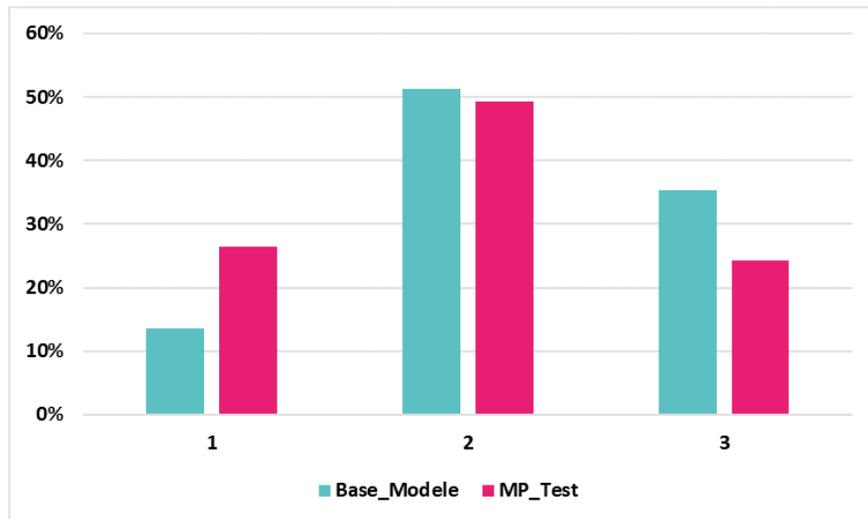


Fig. 7.54 – Score 1^{re} année – structure par classe de risque de la base modèle vs MP test – 3 classes

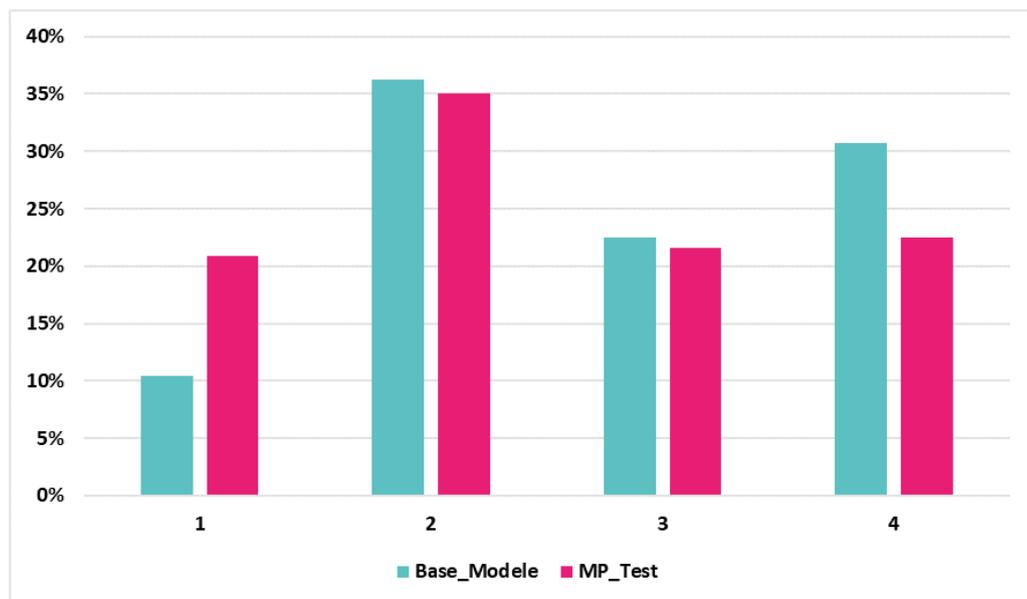


Fig. 7.55 – Score 1^{re} année – structure par classe de risque de la base modèle vs MP test – 4 classes

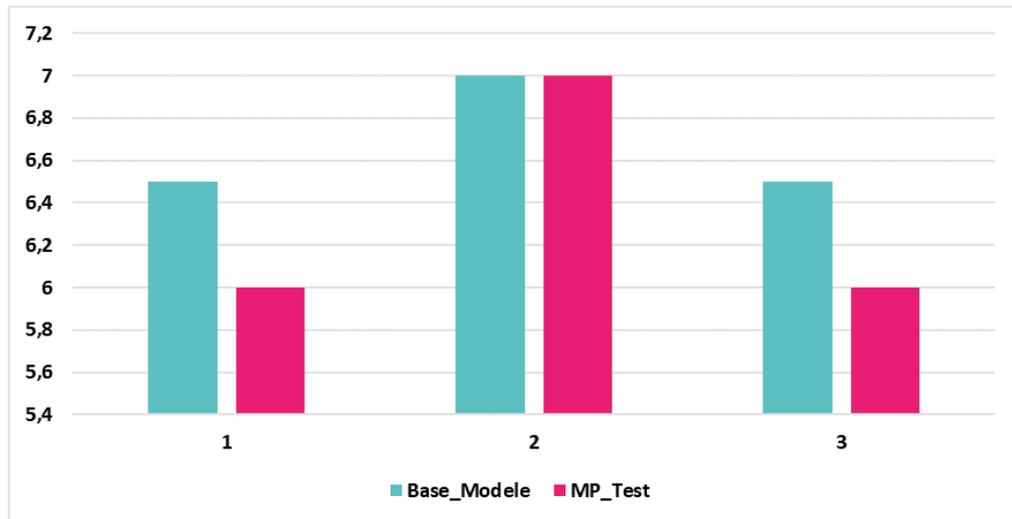


Fig. 7.56 – Score 1^{re} année – ancienneté moyenne par classe de risque de la base modèle vs MP test – 3 classes

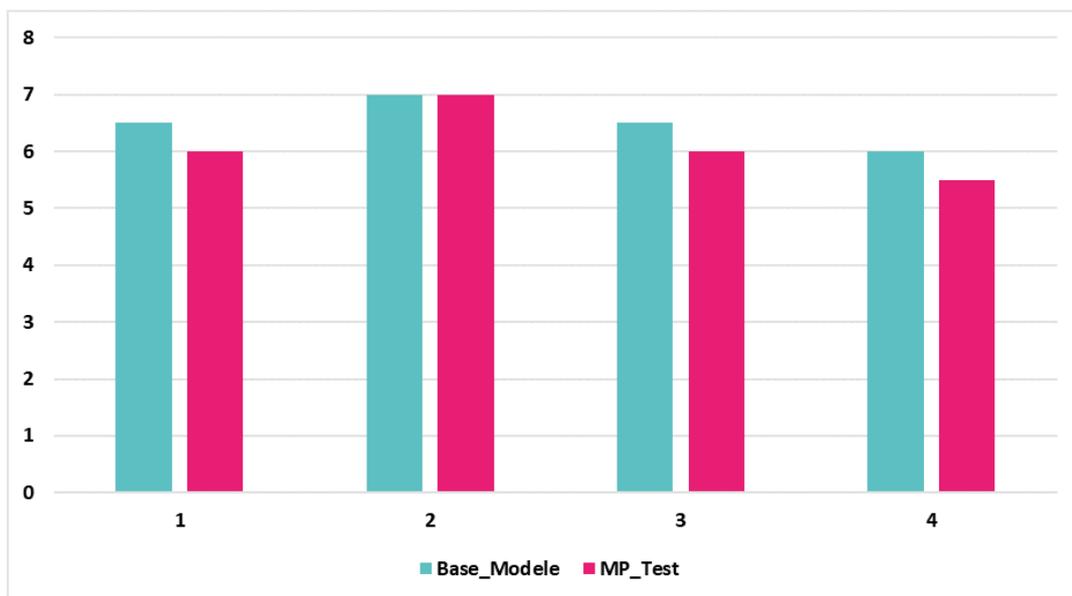


Fig. 7.57 – Score 1^{re} année – ancienneté moyenne par classe de risque de la base modèle vs MP test – 4 classes

7.4.2 Modélisation des rachats

Module	Sous_Module	Risques lié à l'exigence de capital du Module/Sous_module
Module Market (marché)		Evolution des taux d'intérêt
		Liquidité
		Concentration
		Marché d'action
		Marché immobilier
		Change
Module Default		Ecartement des spreads
		Défaut des contreparties Défaut financier
Module Life (vie)		Mortalité
		Longévité
		Incapacité/Invalidité
		Rachat
		Révision (concerne les rentes)
		Catastrophes exceptionnelles
Module Health (santé)	Health SLT (Similar to life)	Mortalité
		Longévité
		Incapacité/Invalidité
		De rachat
	Health non SLT	De frais
		De révision
		Primes et réserves
		Rachat
Module Non life (non vie)		prime et de réserve
		Rachat
		Catastrophe

Fig. 7.58 – SCR - cartographie des Modules/Sous-modules et risques liés au capital requis

- Le taux de rachat annuel

$$\begin{aligned}
 tx_rachat_ann_t = & \min(surr_pc \times (1 + shock_lap1) \times (1 + shock_SLTHlap1); 100) \\
 & + \max(shock_lap2 \times surr_pc; shock_lap2a \times 100) \\
 & + \max(shock_SLTHlap2 \times surr_pc; shock_SLTHlap2a \times 100)
 \end{aligned}$$

Grandeur	Variable Prophet	Description
$tx_rachat_ann_t$	SUR_RATE_ANN_PC	Taux de rachat annuel
$surr_pc$	IP_SURRENDER_PC	Taux de rachat
$shock_lap1$	shock_lap1	Choc additif de rachat, pour les hausses de rachats
$shock_SLTHlap1$	shock_SLTHlap1	Choc additif de rachat
$shock_lap2$	shock_lap2	Choc multiplicatif de rachat, pour les baisses de rachats
$shock_lap2a$	shock_lap2a	Plancher de l'impact du choc multiplicatif
$shock_SLTHlap2$	shock_SLTHlap2	Choc multiplicatif de rachat
$shock_SLTHlap2a$	shock_SLTHlap2a	Plancher de l'impact du choc multiplicatif

- Le taux de rachat mensuel

$$tx_rachat_t = \begin{cases} shock_mass_lapse & \text{si } t=1 \\ 1 - (1 - tx_rachat_ann_t)^{1/12} & \text{si } t > 1 \end{cases}$$

Grandeur	Variable Prophet	Description
tx_rachat_t	SUR_RATE	Taux de rachat mensuel
$tx_rachat_ann_t$	SUR_RATE_ANN_PC	Taux de rachat annuel
$shock_mass_lapse$	shock_mlap	Taux de rachat massif

- L'écoulement des effectifs

$$No_Incidence_t^{cancel} = tx_rachat_t \times No_Healthy_{t-1}$$

Grandeur	Variable Prophet	Description
$No_Incidence_t^{cancel}$	NO_INC_SUR_PP	Nombre de rachat unitaire
$No_Healthy_{t-1}$	NO_POLS_HEALTHY_PP	Nombre d'assurés « en bonne santé » au début de la période de projection,
tx_rachat_t	SUR_RATE	Taux de rachat mensuel au début de la période obtenu à partir de la loi de rachat pour une ancienneté de contrat t,

Bibliographie

- [ACPR, 2019] (2019). ACPR - Analyses et synthèses - Le financement de l'habitat en 2018.
- [ARGUS, 2022] (2022, novembre 16). Assurance emprunteur : la loi Lemoine, une réforme sous surveillance, S. Vié. Récupéré sur ARGUS de l'assurance : <https://www.argusdelassurance.com/>
- [BdF - WebStat] (s.d.). Banque de France - <http://webstat.banque-france.fr>.
- [BOCQUAIRE, 2016] Edith Bocquaire-Les grands principes de l'Actuariat - 2016.
- [CCSF, 2020] (2020). Rapport : Bilan de l'assurance - Actélior pour le Comité consultatif du secteur financier - CCSF.
- [ECHEVIN-SANSON-DHEM, 2019] (2019). D. Echevin O. Sanson, Marine D'HEM -Comprendre Et Conseiller L'Assurance Emprunteur - 2ème édition - L'ARGUS DE L'ASSURANCE.
- [FA, 2021] (2021). France assureurs- L'ASSURANCE FRANÇAISE, Données clés 2020.
- [HOSMER-LIMESHOW, 2000] (2000). Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd Ed.- John Wiley and Sons.
- [IGF - Rapport ADE, 2013] Assurance-emprunteur, Inspection Générale des Finances, Rapport N° 2013-M-086-02, novembre 2013, 47 pages.
- [Bases INSEE] Bases INSEE : typologie des communes et de distribution des revenus à partir de : <https://www.insee.fr>.
- [LAKKIS, 2013] (2013). M. Lakkis - Projection d'un portefeuille d'assurance emprunteur et analyse de la sensibilité des provisions Best Estimate en formule standard - ISUP.
- [Modélisation des chutes PREDICA, 2013] . Analyse des taux de chute du portefeuille ADI - PREDICA - janvier 2013.
- [MOREAU, 2012] (2012). C. Moreau - Assurance emprunteur : Analyse de l'impact de la nouvelle réglementation solvabilité II - UdS.
- [MULLER, 2020] (2020). T. Müller - Détermination d'un processus de construction et de suivi de loi Best Estimate et application au produit assurance Emprunteur.
- [PLANCHET-THEROND, 2006] (2006). F. Planchet et P. Thérond, Modèles de durée , Applications actuarielles, Economica éd.
- [SAPORTA, 2011] (2011). G. Saporta - Probabilités, analyse des données et statistique, 3ème édition - Technip.
- [STA201, 2016] (2016). Cours de l'UE STA201 (Analyse multivariée approfondie) du Master 2 de Statistiques du CNAM.
- [TOSETTI-BEHAR-FROMENTEAU, 2011] (2011). A.Tosetti, T. Behar et M. Fromenteau - Assurance, comptabilité, réglementation, actuariat, 2ème édition - Economica.
- [TUFFERY, 2017] (2017). S. Tufféry - Data mining et statistique décisionnelle - 4ème édition - Technip.

Table des figures

2.1	Vie du contrat ADE : mise en œuvre des prestations	7
2.2	Evolution des montants de cotisations d'assurance emprunteur (Md€) en France de 2007 à 2019	7
2.3	Encours de crédits à l'habitat aux particuliers (en Mds€ d'euros)	8
2.4	Taux des crédits nouveaux à l'habitat	8
2.5	Production mensuelle de crédits à l'habitat et rachats et renégociations (en Mds d'euros)	9
2.6	Parts de marché des principaux acteurs en 2020 (% des cotisations cumulées des 24 principaux acteurs)	11
2.7	Variation 2020/2019 (%) du chiffre d'affaires 2019/2020	11
2.8	Poids de des contrats habitat en nombre dans l'ADE des Caisses régionales	12
2.9	Echéancier prêt à taux fixe	12
2.10	Echéancier prêt in fine	13
2.11	Caractéristiques des garanties de l'offre habitat	13
2.12	Cartographie de l'offre habitat du Crédit Agricole	14
2.13	Lois de résiliations immo par ancienneté	16
2.14	Evolution des taux de rachats moyens constatés – périodes [2015,2016], [2016,2017] et [2018,2019]	16
2.15	Structure en nombre du marché ADE habitat par tranches ancienneté	17
2.16	Lois historiques moyennes par ancienneté	17
2.17	Approche segmentée dynamique	18
3.1	Distribution des familles de crédits en nombre %	20
3.2	Dessin de chaîne simplifiée du traitement de constitution des bases d'études	21
3.3	Qualité de données - Base [B1]	21
3.4	Qualité de données - Base [B2]	21
3.5	Effectifs bases agrégées [BA_i]	22
3.6	Effectifs base échantillonnée agrégée	22
3.7	Ajustement [BE1] échantillon/pop. Totale par Ancienneté	23
4.1	Approche actuelle	27
4.2	Approche alternative : mise en œuvre des scores dans prophet	28
4.3	Approche alternative : mise en œuvre des scores dans les traitements MP	29
4.4	Fréquence des variables sélectionnées – Analyse univariée	31
4.5	Fréquence des variables sélectionnées x fréquence des variables conservées – Analyse bivariée	32
4.6	Synthèse de l'optimisation des modèles DISQUAL par ancienneté	33
4.7	Cartographie des scores par variable et poids de la modalité maximum associée - DISQUAL	34
4.8	Cartographie des scores par variable et poids de la modalité maximum associée - LOGIT	35

4.9	Cartographie des écarts relatifs \times poids des modalités entre LOGIT et DISQUAL	36
4.10	Par ancienneté et méthode prédictive	36
4.11	Synthèse des probabilités de rachats en fonction des classes de risques sur les bases d'étude agrégées	38
4.12	Taux bruts par classe de risque \times Ancienneté	39
4.13	principe du calage des seuils du modèle	40
4.14	ajustement seuil taux de crédit - situation haussière des taux (sans ajustement)	42
4.15	Ajustement seuil taux de crédit - situation baissière des taux (sans ajustement)	42
4.16	Ajustement seuil taux de crédit - situation haussière des taux (avec ajustement)	43
4.17	Ajustement seuil taux de crédit - situation baissière des taux (avec ajustement)	44
4.18	Principe du calage des seuils de test	44
4.19	Présentation de la notion de transition moyenne	46
4.20	Présentation de la notion de trajectoire	47
4.21	Matrice des classes de risques projetées	47
4.22	Les trajectoires pour $r = 3$ et $a_0 = 9$	48
4.23	Présentation de la notion cluster de trajectoires	50
4.24	Clusterisation des trajectoires cluster 1 ($r = 3$)	50
4.25	Table de correspondance entre l'ancienneté \times score et le cluster	52
5.1	SCR - structure de la formule standard	56
5.2	SCR - structure de la formule standard : branches d'activités concernées par l'assurance emprunteur	57
5.3	SCR - Transitions à partir de l'état valide	59
5.4	SCR - Transitions à partir de l'état ITT	59
5.5	SCR - Transitions à partir de l'état ITD	60
5.6	SCR - modèle prophète : graphe des changements d'état	60
5.7	Scénario central - écoulement effectif rachats - approche segmentée « centrale 3m » vs « centrale 4m »	63
5.8	Taux de rachats moyen pondéré - loi segmentée « centrale 3m » vs « 4m »	63
5.9	Scénario central - écoulement effectif rachat lois histo. vs approche segmentée « centrale »	64
5.10	Taux de rachats moyen pondéré - loi segmenté « centrale 3m ». vs loi historique	64
5.11	Scénario dynamique «Crédit Up » : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »	65
5.12	Taux de rachats moyen pondéré lois segmentées : « centrale 3m ». vs « dynamique 3m Crédit Up »	65
5.13	Scénario dynamique «Crédit Down » : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »	66
5.14	Taux de rachats moyen pondéré lois segmentées : « centrale 3m ». vs « dynamique 3m Crédit Down »	66
5.15	Scénario dynamique «mix ADE/Crédit » : écoulement effectif de rachats lois segmentée « dynamique » vs « centrale »	67
5.16	Taux de rachats moyen pondéré lois segmentées : « centrale 3m ». vs « dynamique 3m mix ADE/Crédit »	67
5.17	Ecoulement effectifs : flux de contrats	68
5.18	Ecoulement effectifs : poids des sorties par nature (%)	68
5.19	Ecoulement effectifs : flux de décès	69
5.20	Ecoulement effectifs : flux des termes	69
5.21	BEL/SCR Garantie décès : flux de primes - loi segmentée vs histo.	70
5.22	BEL/SCR Garantie arrêt de travail : flux de primes - loi segmentée vs histo.	70
5.23	BEL/SCR Garantie décès : flux de prestations - loi segmentée vs histo.	71

5.24	BEL/SCR Garantie arrêt de travail : flux de prestations - loi segmentée vs histo.	71
5.25	BEL/SCR Garantie décès : flux de commissions variables DC&AT loi segmentée vs histo.	72
5.26	Impact BEL/SCR - approche alternative - garantie décès	72
5.27	Impact BEL/SCR - approche alternative - garantie arrêt de travail	72
5.28	Impact BEL/SCR - approche alternative - garantie décès	73
7.1	Impact cumulé des évolutions réglementaires sur les possibilités de substitutions	76
7.2	Impact des évolutions réglementaires sur les possibilités de substitutions	76
7.3	Evolution trimestrielle des taux de crédits immo. amortissables, par tranches d'âge - période [07/2015, 07/2019]	77
7.4	Structure des taux de prêts immo. Amortissables (en %) par ancienneté et tranches de durées - juillet 2015	77
7.5	Structure des taux de prêts immo. Amortissables (en %) par ancienneté et tranches de durées - juillet 2019	78
7.6	Evolution trimestrielle des taux ADE immo. annualisés, par tranches d'âge - période [07/2015, 07/2019]	78
7.7	ADE - Structure des taux ADE immo. annualisés (en %) par tranche d'âge et ancienneté - juillet 2015	79
7.8	ADE - Structure des taux ADE immo. annualisés (en %) par tranche d'âge et ancienneté - juillet 2019	79
7.9	Echantillonnage stratifié - exemple de la CR 802	80
7.10	proportion des agences retenues dans l'échantillonnage par ancienneté avec un seuil de 10%	80
7.11	Périmètre de l'étude comportementale - poids des expositions partielles (contrats filtrés) - % des primes décès	81
7.12	décomposition en % de CA décès des expositions partielles	81
7.13	Distribution des effectifs en nombre par ancienneté - Base [B1]	82
7.14	Primes DC HT - base 100	82
7.15	Primes ITT - base 100	83
7.16	Contribution des générations de contrats au chiffre d'affaire (% primes décès)	83
7.17	Ajustement [BE1] échantillon/pop. Totale par Caisse régionales	84
7.18	Ajustement [BE1] échantillon/pop. Totale par Caisse régionales	84
7.19	Cartographie des variables d'étude	85
7.20	Les codes communes INSEE	86
7.21	Descriptif des retraitements des données aberrantes ou manquantes	87
7.22	Analyse descriptive - distribution des capitaux empruntés	88
7.23	Analyse descriptive - durée des financements	88
7.24	Analyse descriptive - destination des financements	89
7.25	Analyse descriptive - couverture ITT	89
7.26	Analyse descriptive - âge des assurés	90
7.27	Analyse descriptive - typologie des lieux résidences des emprunteurs	90
7.28	Analyse descriptive - CSP des emprunteurs	91
7.29	ACM - Les 3 1 ^{ers} plans factoriels pour l'ancienneté 6	96
7.30	Optimisation détaillée du modèle DISQUAL pour l'ancienneté 6	97
7.31	Fonctions de score par dimension, pour l'ancienneté 6 - DISQUAL	97
7.32	Nombre de points par modalité, pour l'ancienneté 6 - DISQUAL	97
7.33	Selection stepwise des variables, pour l'ancienneté 6 - LOGIT	98
7.34	Statistiques d'ajustement par step du modèle - ancienneté 6	98
7.35	Etapas de construction du modèle pour l'ancienneté 6 - LOGIT	99
7.36	Nombre de points par modalité, pour l'ancienneté 6 - LOGIT	99

7.37	AUC et interprétation	101
7.38	Courbes ROC DISQUAL et LOGIT par ancienneté	104
7.39	Représentation graphique des probabilités de rachats en fonction des scores pour l'ancienneté 5	104
7.40	Exemple de calcul de facteur correctif modèle – cas $a_0 = n = 3$	105
7.41	Exemple de calcul de facteur correctif test – cas $a_0 = n = 3$	105
7.42	Diagramme de Lexis	106
7.43	Moyenne des durées de crédits par ancienneté (yc sorties et réaménagements) [07/2015 ->07/2019]	107
7.44	Moyenne des taux de crédits par ancienneté (yc sorties et réaménagements) [07/2015 -> 07/2019]	107
7.45	Moyenne des taux ADE par ancienneté (yc sorties et réaménagements) [07/2015 -> 07/2019]	108
7.46	Moyenne des âges par ancienneté (yc sorties et réaménagements) [07/2015, 07/2019]	108
7.47	Moyenne des capitaux initiaux (CI) par ancienneté (yc sorties et réaménagements) [07/2015, 07/2019]	109
7.48	Trajectoires de risque par anciennetés, pour $r = 3$	110
7.49	Présentation des transitions	111
7.50	Exemple, pour $r = 3$ et $a = 6$ ($a_0 = 1$ et $n = 5$)	111
7.51	Extrait de la table des trajectoires - $r = 3$, $p = 5$ et $1 \leq a_0 \leq 5$	111
7.52	Taux lissés par clusters et anciennetés ($r = 3$)	113
7.53	Taux lissés par clusters et anciennetés ($r = 4$)	113
7.54	Score 1 ^{re} année – structure par classe de risque de la base modèle vs MP test – 3 classes	114
7.55	Score 1 ^{re} année – structure par classe de risque de la base modèle vs MP test – 4 classes	114
7.56	Score 1 ^{re} année – ancienneté moyenne par classe de risque de la base modèle vs MP test – 3 classes	115
7.57	Score 1 ^{re} année – ancienneté moyenne par classe de risque de la base modèle vs MP test – 4 classes	115
7.58	SCR - cartographie des Modules/Sous-modules et risques liés au capital requis	116