

**Mémoire présenté devant l'ENSAE ParisTech  
pour l'obtention du diplôme de la filière Actuariat  
et l'admission à l'Institut des Actuaires**

le \_\_\_\_\_

Par : Déborah SEROR et David NKIHOUABONGA YENGUE

Titre: Modélisation des comportements de rachats dans un cadre de risques  
compétitifs

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

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

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

*signature*

*Entreprise :*

*Nom :*

*Signature :*

*Membres présents du jury de la filière*

*Directeur de mémoire en entreprise :*

*Nom :*

*Signature :*

*Invité :*

*Nom :*

*Signature :*

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

Signature du responsable entreprise

*Secrétariat*

Signature du candidat

*Bibliothèque :*

## Abstract

*Key words: surrenders, competing risks, Fine&Gray approach, subdistribution, time-dependent covariate.*

---

Insurers have been much concerned about surrenders for a long time because they strongly impact their reserves and their ALM strategy. As a result, the European directive Solvency II promotes the development of internal risk model (among which a whole unit is dedicated to the lapse risk). It strengthens the necessity to deeply study and understand this risk.

In our study we perform a survival analysis so as to model the contract lifetime while taking into account the main risk factors impacting the policyholder's behaviors, in the competing risk framework. We perform a Fine & Gray model (Fine and Gray 1999) taking into account structural and cyclical risk factors, on an insurance portfolio embedding Whole Life contracts. The comparison between our model and the Kaplan-Meier estimator is satisfying. However, our model fails to recover the empirical lapse rate trajectory. The model could be improved by specifying a baseline hazard function.

---

## Résumé

*Mots clés : rachats, risques compétitifs, approche de Fine and Gray, sous-distribution, covariable dépendant du temps*

---

Les rachats représentent un risque considérable pour l'assureur : une mauvaise estimation du risque de rachat engendre un provisionnement inadéquat pour faire face à celui-ci, et pourrait conduire à des problèmes de gestion actif/passif.

Face à ces enjeux, la directive Européenne Solvabilité II, qui encourage les assureurs à développer leur propre modèle interne de risque, dédie un module entier à l'évaluation du risque de rachat. Ainsi, les assureurs vont devoir modéliser au plus juste les comportements de rachat. C'est dans ce cadre que s'inscrit notre mémoire, dont l'objectif est de proposer une modélisation inédite du rachat par analyse de survie, dans un cadre de risques compétitifs. Nous mettons en œuvre un modèle de Fine&Gray (Fine and Gray 1999), intégrant des facteurs de risques structurels et conjoncturels sur un portefeuille américain de contrats « Whole Life ». La comparaison du modèle avec l'estimateur de Kaplan-Meier se révèle satisfaisante. Cependant, le taux de rachats prédits par le modèle sur le portefeuille présente des écarts avec celui observé. Une piste d'amélioration serait de considérer un modèle paramétrique en imposant une loi à la fonction de hasard de base.

---

## Remerciements

Nous tenons à remercier chaleureusement Xavier Milhaud pour la richesse du sujet proposé et la qualité de son encadrement tout au long du mémoire.

Nous remercions également l'équipe pédagogique de l'ENSAE pour la qualité des cours dispensés.

Enfin, nous exprimons notre gratitude envers nos proches pour leur soutien au quotidien.

## Sommaire

I.	Intuitions sur les facteurs de risque de rachat .....	7
I.1	Rachat conjoncturel et structurel .....	7
I.2	Etude bibliographique.....	8
I.3	Conclusion.....	9
II.	Présentation du produit « Whole Life ».....	11
II.1	Caractéristiques du « Whole Life » .....	11
II.2	Les primes du « Whole Life » .....	12
II.2.1	Les facteurs déterminant le montant de la prime.....	12
II.2.2	Périodicité de primes et frais .....	13
II.2.3	Calcul du montant de la prime.....	13
II.3	Le réseau de distribution du produit.....	15
II.4	La valeur de rachat .....	16
II.4.1	La valeur de rachat garantie.....	16
II.4.2	Les dividendes .....	17
II.4.3	Les dividendes de fin de contrat .....	18
II.4.4	Conclusion.....	18
II.5	La liquidité procurée par le Whole Life .....	19
II.5.1	Le rachat total du contrat .....	19
II.5.2	Les alternatives au rachat total proposées par l'assureur.....	19
II.5.3	Les couvertures d'assurance du contrat.....	19
II.6	Les options de clôture.....	20
II.7	Fiscalité du Whole Life .....	20
III.	Statistiques descriptives et sélection des variables explicatives.....	22
III.1	Description du portefeuille .....	22
III.2	Etude dynamique des clôtures de contrat .....	27
III.2.1	Le taux de rachat historique du portefeuille .....	27
III.2.2	Taux de rachat en fonction de la durée de vie des contrats .....	28
III.2.3	Evolution du nombre de rachats dans le temps en fonction de la date de souscription et de la durée du contrat .....	29
III.3	Sélection des variables explicatives caractéristiques de l'assuré.....	30
III.3.1	Etude de la durée des contrats sur tout le portefeuille .....	31
III.3.2	Etude de la durée des contrats en fonction des rachats totaux.....	32
III.3.3	Etude de la durée des contrats en fonction de la variable « acc.death.amount » .....	33
III.3.4	Etude de la durée des contrats en fonction du genre.....	34
III.3.5	Etude de la durée des contrats en fonction des classes d'âge .....	35
III.3.6	Etude de la durée des contrats en fonction des classes de risque.....	36
III.3.7	Etude de la durée des contrats en fonction du lieu de résidence.....	37
III.3.8	Etude de la durée des contrats en fonction de la fréquence de paiement des primes.....	38

III.3.9	Etude de la durée des contrats en fonction de la classe de la prime .....	39
III.4	Réalisation de tests statistiques.....	40
III.4.1	Test du chi 2 sur la variable rachat (« surrender.bit » dans la base de données » .....	40
III.4.2	Tests statistiques sur la variable durée du contrat.....	41
III.4.3	Tests de corrélation.....	42
III.4.4	Analyse des correspondances multiples .....	42
III.5	Prise en compte du contexte économique.....	44
III.5.1	Impact des marchés financiers sur le taux de rachat.....	44
IV.	Cadre théorique de l'analyse de survie avec risques compétitifs .....	48
IV.1	Cadre théorique.....	48
IV.1.1	Formalisme .....	48
IV.1.2	Approche par risques compétitifs .....	49
IV.1.3	Les grandeurs utilisées dans le cadre des risques compétitifs .....	50
IV.1.4	Modèle de régression dans le cadre de risques compétitifs .....	52
IV.2	Études empiriques.....	56
IV.2.1	Estimations empiriques.....	56
IV.2.2	Test de proportionnalité.....	58
V.	Modèle de Cox dans un cadre Fine&Gray avec une covariable dépendant du temps .....	65
V.1	Introduction d'une variable dépendant du temps dans le modèle de Cox .....	65
V.2	Prise en compte de l'approche Fine&Gray dans notre modèle .....	68
V.3	Résultats de la modélisation .....	68
V.4	Vérification de l'hypothèse proportionnalité.....	72
V.5	Comparaison de la fonction de survie du modèle avec l'estimateur de Kaplan-Meier .....	74
V.6	Reconstitution du nombre de rachats.....	78
V.7	Evolution vers un modèle paramétrique .....	80
VI.	Annexes .....	86
VI.1	Annexe statistiques descriptives.....	86
VI.1.1	Méthode de segmentation de la prime annuelle .....	86
VI.1.2	ACM : Tableaux de la construction des axes factoriels.....	87
VI.2	Exemples de fonctions de hasard empiriques.....	89
VI.3	Prise en compte de la covariable dépendant du temps et de l'approche Fine&Gray dans la programmation sous R .....	93
VI.3.1	Constitution initiale de la base de données.....	93
VI.3.2	Introduction de la performance de la covariable traduisant le contexte économique.....	93
VI.3.3	Prise en compte des risques compétitifs par l'approche Fine&Gray.....	95
VI.4	Test de proportionnalité des covariables du modèle de Cox approche Fine&Gray .....	97
VII.	Bibliographie .....	100

## INTRODUCTION

Un contrat d'assurance vie permet à son souscripteur de transmettre un capital, de se constituer un complément de retraite ou de financer un projet tout en bénéficiant de conditions fiscales avantageuses.

L'assureur prend l'engagement irrévocable de verser des prestations au(x) bénéficiaire(s) du contrat en fonction de la réalisation d'événements viagers en échange d'un engagement révocable du souscripteur du contrat de verser des primes.

Le caractère révocable de l'engagement du souscripteur a d'importantes conséquences pour l'assureur. En effet, cela implique que le souscripteur peut à tout moment cesser de verser ses cotisations et récupérer son épargne de manière partielle ou totale. On parlera alors respectivement de rachat partiel ou de rachat total du contrat d'assurance. L'assureur doit alors être en mesure de faire face à ses engagements.

Les rachats totaux représentent donc un risque considérable pour l'assureur. Une mauvaise estimation du risque de rachat, un mauvais provisionnement pour faire face à celui-ci pourraient conduire à des problèmes de gestion actif/passif. En effet, dans le cas d'un rachat massif par exemple, l'assureur pourrait être confronté à des problèmes de liquidité. Il devrait alors liquider des actifs adossés aux engagements et pourrait subir des pertes dues à un décalage entre les taux garantis aux assurés et les rendements de ses propres actifs.

De par son enjeu considérable pour l'assureur, le risque de rachat a été intégré au dispositif prudentiel Solvabilité 2 et devra ainsi faire l'objet d'une modélisation la plus précise possible. C'est dans ce contexte que s'inscrit notre mémoire dont l'objectif est de proposer une modélisation du taux de rachat sur une ligne de produits d'assurance-vie.

# I. Intuitions sur les facteurs de risque de rachat

## I.1 Rachat conjoncturel et structurel

Le risque de rachat est avant tout un risque de comportement humain : c'est l'aboutissement d'un processus de décision de l'assuré. Il convient de distinguer deux grandes catégories de rachat : les rachats conjoncturels et les rachats structurels.

Le rachat conjoncturel est lié aux conditions de marché et à la volonté de l'assuré d'investir dans des placements plus rémunérateurs. Cette possibilité s'apparente à une option américaine dont l'exercice se produit au moment le plus favorable pour l'assuré, et souvent le plus défavorable pour l'assureur.

En effet, lors d'une crise financière les assurés détenteurs d'un contrat sans taux garanti percevront un faible rendement, ceci étant du au fait que les performances financières des actifs de l'assureur soient moins bonnes. Ils seront alors tentés de racheter leur contrat. Au contraire, les détenteurs de contrat à taux garanti ont intérêt à conserver leur contrat. De même, le lancement de nouveaux produits par la concurrence peut inciter l'assuré à effectuer une sortie anticipée de son contrat.

Ainsi, au sein des facteurs de risque de rachat d'origine conjoncturelle nous pouvons distinguer les catégories suivantes :

- le contexte économique et financier : état des marchés financiers, taux de chômage, inflation, croissance etc.
- l'évolution de la législation : apparition de nouvelles taxes
- la concurrence : lancement de nouveaux produits, augmentation du spread entre le taux de rendement proposé par le contrat avec celui de la concurrence ...
- l'image et le rating de la compagnie.

Un second type de rachat, le rachat structurel, correspond à un désir ou besoin de liquidité immédiat de la part de l'assuré. Lors d'un rachat structurel l'assuré rachète pour des raisons personnelles généralement inconnues de l'assureur (achat immobilier, etc.) et cela même si les conditions de marché lui sont favorables.

De même, nous pouvons classifier les potentiels facteurs de risques structurels de la façon suivante :

- les caractéristiques de l'assuré : l'âge, la richesse de l'assuré, son état de santé etc.

(Un jeune actif sera plus enclin à racheter son contrat afin de financer des projets, tel qu'un achat immobilier, qu'un retraité.)

- le type de contrat : contrat d'épargne, de rentes ...
- les caractéristiques du contrat : le montant et le nivellement des primes, la fréquence de versement, les options proposées par le contrat, les pénalités de rachat ...

## 1.2 Etude bibliographique

Le risque de rachat a suscité de nombreux travaux de recherche dans la littérature. Cependant, rares sont les auteurs qui ont étudié le rachat dans un contexte à la fois conjoncturel et structurel. Les travaux concernent le rachat structurel ou le rachat conjoncturel mais rarement les deux en même temps.

La première étude portant sur la modélisation du comportement de rachat dans un cadre structurel est celle de S. Haberman et A. Renshaw (1986). Les deux chercheurs ont étudié par une approche statistique le comportement de rachats de contrats d'assurance vie en Ecosse en 1976. Ils ont établi que les facteurs majeurs de risques sont l'ancienneté et le type de contrat ainsi que la compagnie d'assurance concernée. Ils mettent en œuvre des modèles GLM et modélisent l'hétérogénéité du portefeuille via des termes d'interactions entre les facteurs de risque.

Le mémoire « Analyse des rachats d'un portefeuille vie individuelle : Approche théorique et Application pratique » Stéphane Fauvel et Maryse le Pévédic (2007) s'inscrit également dans un contexte structurel. Les deux auteurs développent un modèle basé sur l'espérance d'utilité de Von Neumann et Morgenstern afin de déterminer si l'assuré préfère racheter son contrat ou contracter un emprunt.

Les travaux portant sur le rachat conjoncturel sont plus nombreux. Certaines études concernent la valorisation financière de l'option de rachat des contrats d'Assurance-vie. Nous pouvons citer l'étude de Weixi Shen et Xu Huiping (2004) dans laquelle le rachat est modélisé par une option américaine. Il s'agit alors pour les deux auteurs de déterminer la valeur du contrat et la date optimale de rachat. Notre objectif étant l'étude du comportement de rachat et non sa valorisation financière, nous ne développerons pas de telles méthodes dans notre mémoire.

L'étude Changki Kim (2005) est très intéressante pour notre sujet. L'auteur étudie en effet l'impact de variables économiques sur le taux de rachat d'un portefeuille coréen. L'observation d'un pic de rachat lors de la forte hausse des taux d'intérêt qui s'est produite lors de la crise financière de Corée en 1997, lui suggère que l'évolution des taux d'intérêt et l'état des marchés financiers influencent de manière significative le rachat. Afin de vérifier ses intuitions, il effectue une régression logistique GLM (logit) en utilisant comme variables explicatives le spread entre le taux d'intérêt garanti par l'assureur et la référence du marché, des variables muettes pour les périodes de crises financières, le taux de chômage et le taux de croissance économique ainsi que l'âge du contrat. Il prend en compte d'éventuels effets de retard entre l'écart taux garanti-taux de marché et le taux de rachat (retards de 2 à 12 mois pris en compte). A l'issue de la régression logistique, il établit que les variables introduites sont toutes significatives sur le taux de rachat. De plus, il semblerait que l'assuré observe le spread entre les taux garantis-taux de marché durant une période de 2 à 6 mois avant de prendre la décision de racheter. Les résultats de son étude démontrent également que la modélisation du rachat est très sensible au type de produit considéré. En effet, selon qu'il s'agisse d'un contrat d'épargne pure, de rentes ou autre, l'effet des variables explicatives est plus ou moins accentué.

Dans son étude Xavier Milhaud (2011), constate l'existence d'un clivage dans la littérature entre rachat conjoncturel et structurel. Il propose alors un modèle GLM avec des variables explicatives caractéristiques de l'assuré et d'autres traduisant le contexte économique. La réalisation d'entretiens avec les

équipes marketing d'un grand assureur, lui permet d'identifier des facteurs de risques tels que la fiscalité, les pénalités de rachat ainsi que le commissionnement des agents de vente.

Dans un premier temps, l'auteur ne prend pas en compte le contexte économique et se concentre sur la modélisation du rachat structurel. Pour ce faire, il utilise des modèles de segmentation CART et LR et étudie l'impact de variables telles que l'âge, le sexe de l'assuré ou le type de contrat, le montant et la fréquence des primes, la richesse de l'assuré et l'ancienneté du contrat. Il en ressort que le sexe et la prime de risque de l'assuré ne sont pas des variables significatives. L'ancienneté du contrat, le type de contrat et l'option de participation aux bénéfices ont, au contraire, un impact important sur le rachat. Cette segmentation permet de dresser quelques types de profils risqués pour le rachat :

- les personnes jeunes
- les assurés qui versent une prime périodiquement (particulièrement les versements annuels et bimensuels)
- les assurés les moins fortunés
- les contrats avec une clause de participation aux bénéfices

Dans un second temps Xavier Milhaud (2011) prend en compte, via une modélisation mélange, l'environnement économique et la forte corrélation qui existe alors entre les comportements des assurés en période de crise. Le modèle développé donne de bons résultats sur certaines lignes de produits et tend à prouver qu'il n'est pas nécessaire de prendre en compte trop de facteurs de risque : l'ancienneté du contrat, le contexte économique et un troisième facteur discriminant suffisent dans la majorité des cas.

### 1.3 Conclusion

Dans la littérature, les modèles développés sur le risque de rachat peuvent se scinder en quatre grandes catégories :

- des modélisations financières avec valorisations d'options,
- des modélisations statistiques avec des séries temporelles : le rachat est étudié de manière collective (on agrège les décisions individuelles et on étudie le taux de rachat),
- des approches microéconomiques avec des critères d'espérance d'utilité (modélisation à l'échelle individuelle),
- des modélisations probabilistes avec des modèles GLM où le rachat est également étudié à l'échelle individuelle.

Nous allons privilégier une modélisation probabiliste car elle nous permet de modéliser la décision de rachat à l'échelle individuelle sans formuler d'hypothèse sur la rationalité de l'individu. Nous proposons une modélisation inédite du risque de rachat via des modèles de durée. Il s'agit d'un sous-ensemble de modèles probabilistes spécifiques à l'étude des durées de vie. Ils permettent de modéliser la loi d'occurrence d'un événement en fonction d'éventuels facteurs explicatifs.

Les modèles de survie ont parfois été utilisés avec succès afin de modéliser le risque de crédit Djaidja, Jan Beran et Abdel Yazid Karim(2006), ou encore le remboursement anticipé de crédits immobiliers Kim Hin Ho et Huiyong Su(2006).

Dans les modélisations du risque de rachat dans la littérature, l'ancienneté du contrat était une variable fortement significative. L'approche par modèle de durée nous permet de traiter cette variable non comme une variable explicative mais comme étant la variable d'intérêt.

Concernant les facteurs de risque de rachat, l'étude bibliographique menée nous amène à accorder la même importance aux dimensions conjoncturelles et structurelles du rachat. Ainsi, nous allons inclure des variables explicatives structurelles et conjoncturelles dans notre modèle afin de traduire le comportement de l'assuré dans son contexte économique.

Notre étude bibliographique nous a permis de développer des intuitions sur les facteurs de risques significatifs. L'âge, la richesse de l'assuré mais également la fréquence de versement des primes semblent être des variables explicatives pertinentes pour expliquer les rachats structurels. Enfin, des variables traduisant le contexte économique et financier tel qu'un indice boursier ou l'évolution des taux d'intérêt seront à considérer lors de notre modélisation.

## II. Présentation du produit « Whole Life »

Nous avons à notre disposition, afin d'effectuer notre étude, un portefeuille simulé à partir d'un portefeuille réel sur un produit d'assurance vie américain de type « Whole Life », commercialisé par une grande compagnie d'assurance internationale.

### II.1 Caractéristiques du « Whole Life »

Sur le marché de l'assurance-vie aux Etats-Unis deux grands types de produits existent : les produits « Whole Life » et les « Term Life ». La différence entre les deux réside dans la date d'échéance du contrat. Les « Term Life » présentent une date d'échéance du contrat fixée lors de la souscription : si le décès de l'assuré intervient après cette date aucune somme n'est versée au(x) bénéficiaire(s).

Les produits « Whole Life », quant à eux, permettent d'être couvert contre le décès pour une durée « illimitée » c'est-à-dire jusqu'au décès effectif de l'assuré. Ils garantissent le versement d'une indemnité suite au décès de l'assuré et possèdent également une valeur de rachat.

Nous allons détailler ci-dessous les caractéristiques et garanties spécifiques que présente le produit « Whole Life » sur lequel nous avons réalisé notre étude.

Le produit présente les garanties suivantes :

- un taux garanti qui n'est pas précisé pour des raisons de confidentialité,
- le souscripteur verse des primes périodiques dont le montant est fixé et reste constant au cours de la vie du contrat,
- la valeur de rachat croît lors de la vie du contrat,
- l'indemnité versée suite au décès de l'assuré est au minimum égale à la valeur nominale initiale du contrat (sous la condition que toutes les primes aient été payées et que l'assuré n'ait contracté aucun emprunt),
- une possibilité de contracter un emprunt sur la valeur de rachat durant la vie du contrat (choix entre un taux d'intérêt fixe ou un taux indexé sur l'indice Moody's).

L'assuré peut à tout moment racheter partiellement ou totalement son contrat à sa valeur de rachat. Lors d'un rachat total, l'assuré récupère la valeur de rachat et met fin au contrat. Dans le cadre de notre étude, nous nous intéresserons uniquement aux rachats totaux.

Concernant la fiscalité, le produit présente les caractéristiques suivantes :

- l'épargne constituée n'est pas taxée durant sa phase d'accumulation
- l'indemnité décès versée au bénéficiaire n'est pas taxée

- les prêts réalisés sur le contrat ne sont pas soumis à l'impôt sur le revenu tant que le contrat n'est pas clôturé.

## 11.2 Les primes du « Whole Life »

La prime est un élément central du produit « Whole Life » car, contrairement à d'autres produits avec versement unique ou versements libres, les versements de primes sont périodiques et rythment la vie du contrat. Ainsi, la structure des primes (montant, fréquence) est certainement significative pour le rachat du contrat.

Une caractéristique majeure du produit est le nivellement des primes. Le montant est fixé lors de la souscription et n'augmente pas au cours de la vie du contrat. Cette caractéristique est très importante pour notre étude car elle permet déjà d'écarter un potentiel facteur de risque de rachat : celui de la variation du montant de la prime au cours du contrat.

Cependant, le montant de la prime reste un potentiel facteur de risque de première importance car si l'assuré a l'impression de payer « trop » cher son contrat, celui-ci sera beaucoup plus tenté de le racheter. Par conséquent, nous allons nous intéresser à la tarification de la prime afin de dégager quels profils d'assurés paient une prime élevée.

### 11.2.1 Les facteurs déterminant le montant de la prime

Le montant des primes est fixé à la date de souscription et dépend des facteurs suivants :

- Valeur nominale initiale du contrat :

Il s'agit du capital décès initial garanti. Ce dernier est au minimum de 10 000\$ (sauf pour les contrats collectifs minimum de 1000\$). Il est réparti en cinq tranches de tarification :

Tranche 1 : \$ 1 000–\$ 24 999

Tranche 2 : \$ 25 000–\$ 99 999

Tranche 3 : \$ 100 000–\$ 249 999

Tranche 4 : \$ 250 000–\$ 999 999

Tranche 5 : \$ 1 000 000 +

Ce facteur sera par la suite assimilé à la richesse de l'assuré. En effet, il est légitime de penser que plus l'assuré est fortuné, plus son contrat d'assurance a une valeur élevée.

- Age à la date de la souscription
- Sexe du souscripteur
- Consommation de tabac : fumeur ou non fumeur
- Classe de risque de l'assuré : elle est établie sur critères médicaux. Les assurés sont répartis en trois classes :
  - classe « Substandard » présente un risque supérieur à la moyenne.

- classe « Standard » présente le niveau de risque moyen de la population.
- classe « Preferred » : regroupe les assurés considérés comme étant moins risqués que la moyenne.

### II.2.2 Périodicité de primes et frais

Le souscripteur peut choisir la fréquence de versement des primes. Des frais sont ajoutés à la prime selon la fréquence de versement choisie :

- annuellement (50 \$ à chaque versement)
- semestriellement (26 \$ à chaque versement)
- trimestriellement (15 \$ à chaque versement)
- mensuellement (6 \$ à chaque versement)

Ainsi plus le versement de primes est fréquent, plus les frais sont importants.

### II.2.3 Calcul du montant de la prime

Le montant de la prime périodique à verser est calculé de la façon suivante :

$$\mathbf{Prime = A \times L \times Mo + Frais}$$

Avec

<b>A</b>	Taux par tranche de 1000 \$ de capital décès. Elle dépend des caractéristiques de l'assuré cité au (II.2.1).
<b>L</b>	facteur de périodicité (exemple : L=0,5 pour une prime versée semestriellement)
<b>Mo</b>	capital décès initial garanti
<b>Frais</b>	frais du contrat (cf II 2.2)

**Tableau 1 variables utilisées pour le calcul de la prime**

Application : Calcul de la prime trimestrielle versée par un homme de 35 ans de classe de risque « preferred », non-fumeur dont le contrat présentent les caractéristiques suivantes :

- un capital décès de 100 000 \$

- les options « Disability Premium Waiver (DPW) » and « Accidental Death Benefit (ADB) » d'une valeur de 100 000 \$.

	Tarif en base annuelle	Facteur de périodicité	Résultat
Prime trimestrielle pour un capital décès de 100 000 \$	12,10 \$	× 0,2596	= 3,14 \$
Supplément de prime pour l'option « Disability Premium Waiver »	0,84 \$	× 0,2596	= 0,22 \$
Supplément de prime pour l'option « Accidental Death Benefit »	0,85 \$	× 0,2596	= 0,22 \$
Total			= 3,58 \$
Multiplication par le montant du capital garanti par 1000 \$ (3,58 x 100)			= 358 \$
Ajout frais de police			+ 15 \$
Prime trimestrielle totale			= 373 \$

Tableau 2 Exemple de calcul de la prime

Conclusion sur le montant de la prime :

Le montant de la prime est directement proportionnel au capital décès initial garanti *MO* qui peut être assimilable à la richesse de l'assuré. Ainsi, plus un assuré est fortuné, plus il acquiert un contrat d'assurance de valeur élevée et plus il paie des primes élevées.

Le montant de la prime dépend également, via le coefficient A, de l'âge, du sexe, de la consommation de tabac et du niveau de risque. Il semble alors qu'être jeune, fumeur ou « Substandard » engendre le paiement d'une prime plus élevée pour un même capital décès, qu'un individu plus âgé, non fumeur ou « Preferred ». Par conséquent, il est possible qu'un individu jeune, fumeur et/ ou « Substandard » développe un sentiment de lassitude de payer plus que son homologue pour un même capital décès garanti.

Enfin, la fréquence de versement intervient au niveau des frais ajoutés à la prime. Plus la fréquence de versement est élevée, plus l'assuré paiera des frais importants. De plus, verser fréquemment des primes peut entraîner un sentiment de lassitude chez l'assuré qui pourrait l'inciter à racheter son contrat.

### 11.3 Le réseau de distribution du produit

Nous avons souligné lors de notre étude bibliographique le rôle important joué par le réseau de distribution sur le rachat car les agents de vente conseillent leurs clients tout au long de la vie de leur contrat.

Les agents de vente reçoivent des commissions suite à de nouvelles souscriptions de contrats. Suite à l'apparition d'un nouveau produit sur le marché, il existe souvent des politiques commerciales chez les agents de vente visant à réaliser le maximum de souscriptions sur ces nouveaux produits. Par conséquent, l'agent de vente peut très bien avoir intérêt à inciter un assuré à racheter son contrat afin de souscrire à un nouveau produit.

Il est ainsi intéressant d'étudier comment s'effectue la rémunération du réseau de distribution sur notre produit Whole Life.

Frais pour les contrats dont la valeur nominale est supérieure à 10 000\$	1ère année du contrat	Années 2-10	Années 11 +
commission	50% <sup>1</sup>	4%	0%
Frais de gestion	0%	2%	2%
Frais de gestion supplémentaire	0%	0%	1%
Frais maximaux	50%	6%	3%

**Tableau 3 Grille des frais appliqués au contrat**

Le taux de commission est le pourcentage prélevé par le réseau de distribution (courtier, agence ...).

Concernant notre produit, les agents de vente se rémunèrent essentiellement sur les primes versées lors des deux premières années qui suivent la souscription. Au delà de la seconde année, ils perçoivent une commission beaucoup plus faible sur les contrats souscrits.

Par conséquent, il est probable que les agents de vente incitent leurs clients à racheter assez tôt, dès la deuxième année par exemple.

Ce facteur de risque est impossible à intégrer dans notre modélisation car il nécessite des informations très difficiles à avoir pour l'assureur : l'assuré qui a racheté son contrat a-t-il souscrit à un autre contrat peu de temps après ?

---

<sup>1</sup> Le taux de commission de 50% la première année s'applique jusqu'à l'âge de 60 ans. Pour la tranche d'âge 61-65 ans, il est de 47,5%. Au-delà de cette tranche, le taux de commission pour la première année décroît avec l'âge à la souscription (1% par année au delà de 65 ans).

Un élément primordial du produit Whole Life dans le cadre de notre étude est le montant auquel peut prétendre l'assuré lors du rachat. Ce montant, appelé valeur de rachat, évolue de manière dynamique et impacte de manière évidente la décision de racheter à une date donnée.

## 11.4 La valeur de rachat

Les primes versées sur le contrat Whole Life, diminuées des frais de gestion, s'accumulent telle une épargne au cours de la vie du contrat. Le montant accumulé est appelé la valeur de rachat du contrat (en pratique, le montant réel récupéré suite au rachat correspond à la valeur de rachat diminuée des pénalités de rachat).

La valeur totale de rachat du contrat est la somme des trois quantités suivantes :

- la valeur de rachat garantie
- la valeur des dividendes
- la valeur des dividendes de fin de contrat

Nous allons détailler ci-dessous chaque composante de la valeur totale de rachat.

### 11.4.1 La valeur de rachat garantie

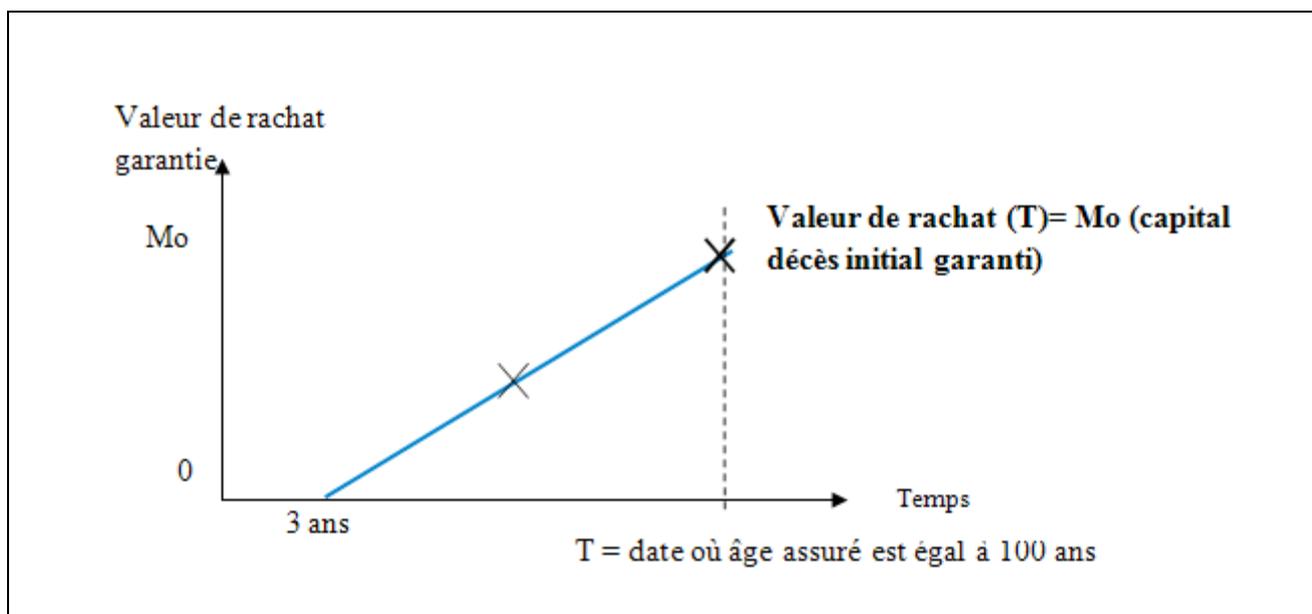


Figure 1 Evolution de la valeur de rachat garantie au cours de la vie du contrat

*Pour  $t < T$  (Age 100 ans), Valeur de rachat garantie(t) < Capital Décès Initial Garanti*

La valeur de rachat garantie croît chaque année à partir la troisième année du contrat. Elle est théoriquement égale au capital décès initial si l'assuré atteint l'âge de 100 ans.

La valeur de rachat garantie peut être empruntée mais elle ne peut pas faire l'objet de rachat partiel.

De plus, elle est nulle pendant les trois premières années de la vie du contrat. Cette caractéristique de la valeur de rachat est imposée par l'assureur afin de dissuader fortement l'assuré de racheter avant la troisième année. Ceci s'explique par le fait que les frais de gestion ne sont prélevés sur les primes du contrat qu'à partir de la deuxième année (cf tableau II.3). Ainsi, l'assureur ne prend pas de frais de gestion lors d'un rachat antérieur à la deuxième année.

Cette incitation est très forte car un assuré qui rachète son contrat avant trois ans ne récupère rien.

A partir de la troisième année, le profil de croissance de la valeur garantie (égalité avec le capital décès initial à l'âge de 100 ans) a pour objectif d'encourager l'assuré à conserver son « Whole life » afin de bénéficier d'une accumulation de son épargne.

#### 11.4.2 Les dividendes

Ils constituent la seconde composante de la valeur de rachat total. Contrairement à la valeur de rachat garantie, les dividendes ne sont pas garantis. Les dividendes sont généralement versés à partir de la fin de la deuxième année. Ils dépendent de la valeur du capital garanti et du sexe de l'assuré.

L'assureur effectue des provisions pour les risques financiers et techniques (mortalité). Si les résultats sont meilleurs que prévus, l'assureur verse l'excédent aux assurés par le biais des dividendes.

L'assuré peut choisir sous quelle forme il désire recevoir ses dividendes :

##### ➤ **Achat d'options de paid-up**

Il s'agit du mode le plus fréquemment choisi. Les dividendes sont utilisés pour acheter des extensions au contrat. Ces dernières permettent d'augmenter la valeur de rachat du contrat et de l'indemnité décès. Avec cette option, les dividendes peuvent également être rachetés afin de payer les primes du contrat. Ceci est très intéressant pour l'assureur car cela permet d'éviter les défauts de paiements.

##### ➤ **Réduction du montant des primes**

Les primes sont diminuées d'une somme égale au montant des dividendes. Le coût que représente le contrat pour l'assuré est ainsi réduit. Cette option va inciter l'assuré à ne pas racheter.

##### ➤ **Versement des dividendes en cash**

Cette option procure de la liquidité immédiate à l'assuré.

##### ➤ **Laisser en dépôt les dividendes**

Le souscripteur ne retire pas ses dividendes et accumule ainsi des intérêts sur ces dernières. Le taux d'intérêt est garanti comme supérieur à 3%. Les intérêts perçus sont soumis à des taxes.

##### ➤ **Réduire un prêt en cours**

##### ➤ **Achat de « Term Insurance » d'une échéance d'un an**

Pendant une année, l'indemnité décès du contrat est augmentée du montant du contrat « Term Insurance ».

Toutes ces options proposées par l'assureur ont l'objectif d'inciter les assurés à ne pas racheter leur contrat. En effet, elles répondent soit à un besoin de liquidité immédiate (versement en cash) ou bien elles visent à éviter un effet de lassitude (réduction des primes, d'un prêt en cours etc.).

Lors de notre étude bibliographique, nous avons mentionné le fait que les assurés possédant une option de participation aux bénéfices ont tendance à racheter plus que les autres. Cette option est présente pour tous les assurés sur notre produit, ainsi il ne s'agira pas d'un facteur de risque pertinent à inclure dans notre modèle.

### 11.4.3 Les dividendes de fin de contrat

Il s'agit de la troisième composante de la valeur de rachat.

Les dividendes de fin de contrat sont versés quand :

- le contrat arrive à maturité,
- le contrat est racheté,
- l'assuré décède.

Ces dividendes ne peuvent être empruntés ou rachetés pour payer des primes et ils ne sont pas garantis. De plus, ils ne sont versés qu'à partir de la 12<sup>ème</sup> année et sont égaux au pourcentage suivant

$$Dividende_{fin\_contrat}(t) = \frac{Valeur\ Rachat\ Garantie(t)}{Capital\ Décès\ Initial\ Garanti}$$

Cependant, ils ne peuvent excéder 25 \$ par tranche de 1 000 \$ de capital décès initial garanti.

Issue Age	End of Policy Year										Maximum per 1 000 \$
	11 or less	12	13	14	15	16	17	18	19	20 & more	
0-39	—	—	—	1%	1,5%	2%	2,5%	3%	3,5%	4%	25 \$
40-44	—	—	1%	1,5	2	2,5	3	3,5	4	4,5	25 \$
45 & over	—	1%	1,5	2	2,5	3	3,5	4	4,5	5,0	25 \$

**Tableau 4 Grille des taux appliqués pour les dividendes**

Les dividendes de fin de contrat représentent pour l'assureur une façon d'inciter l'assuré à conserver son contrat plus de douze ans.

### 11.4.4 Conclusion

La dynamique de la valeur de rachat a pour objectif d'inciter l'assuré à conserver le plus longtemps possible son contrat. Des options sont proposées à l'assuré dans le but d'éviter qu'il ne rachète par besoin de liquidité ou par effet de lassitude.

Nous avons souligné le fait que racheter avant la troisième année est totalement contraire à l'intérêt de l'assuré car il ne récupère rien de ses primes investies.

## 11.5 La liquidité procurée par le Whole Life

Dans la description commerciale du produit, l'assureur vante le fait que le Whole Life n'est pas uniquement dédié à la transmission de capital. Le produit peut servir à financer des projets au cours de la vie de l'assuré.

Cet argument publicitaire de l'assureur peut étonner car n'incite-t-il pas l'assuré au rachat en lui soulignant que l'épargne accumulée via son contrat est une source de liquidité ?

Au contraire, l'assureur anticipe ce besoin de liquidité de l'assuré en proposant des alternatives au rachat total.

### 11.5.1 Le rachat total du contrat

L'assuré peut décider de racheter son contrat à tout moment de la vie du contrat. C'est l'étude de ce comportement qui fait l'objet de notre mémoire. La valeur de rachat nette est établie de la façon suivante :

***valeur de rachat nette = valeur de rachat + dividendes – valeurs des prêts contractés.***

Une fois le contrat racheté, toute couverture d'assurance garantie par le contrat est rendue caduque et le contrat ne peut être rétabli.

### 11.5.2 Les alternatives au rachat total proposées par l'assureur

L'assuré peut accéder à tout moment à la valeur de rachat de son contrat par le biais d'emprunt ou de retrait de dividendes.

- Rachat de dividendes : l'assuré rachète uniquement les dividendes qu'il a perçus sur son contrat.
- Emprunt contracté sur le contrat : la valeur maximale qu'il peut emprunter au cours de l'année  $t$  est égale :

*Valeur de rachat (fin année  $t$ ) + dividendes perçues – primes non payées de l'année en cours  
– valeur prêt éventuel déjà contracté*

### 11.5.3 Les couvertures d'assurance du contrat

- L'indemnité décès garantie

Si l'assuré décède alors que le contrat Whole Life est encore en vigueur, une indemnité décès au moins égale au montant nominal initial est versée au bénéficiaire à condition que toutes les primes aient été payées et qu'il n'ait pas contracté d'emprunt.

- Extension de couverture

Il est possible d'acquérir une indemnité décès supérieure à celle garantie par le contrat si les dividendes ont été utilisés pour acheter des « paid-up additions » ou des contrats temporaires. Il est également possible d'acheter une option « accidental rider » qui permet d'obtenir une indemnité décès supérieure en cas de décès accidentel.

Il semblerait intuitif que l'achat d'extension au contrat fidélise le client et le rende moins susceptible de racheter son contrat.

## 11.6 Les options de clôture

Si la prime due n'est pas reçue à la fin d'une période, le contrat est clôturé. Selon l'option de fin de contrat choisie par le souscripteur ce dernier peut :

- transformer son Whole Life en un contrat temporaire « Paid-Up Extended term Insurance » : à la date de clôture la valeur de rachat du contrat est utilisée pour acheter un contrat « Term Life ».
- prolonger son contrat comme un « Reduced-Paid Up Insurance » : la valeur de rachat du Whole Life est utilisée afin d'acheter un nouveau Whole Life d'une valeur nominale plus faible.
- effectuer le rachat du contrat à sa valeur de rachat : le rachat du contrat peut être effectué à tout moment de la vie du contrat.
- rétablir le contrat sous certaines conditions dictées par la compagnie d'assurance.

## 11.7 Fiscalité du Whole Life

Les contrats Whole Life présentent un certain nombre d'avantages fiscaux qui sont réglementés par des lois fiscales américaines. L'une d'entre elle, la loi TAMRA, intervient dans le cadre de notre étude.

Cette loi a pour objectif d'empêcher des souscripteurs d'acheter un produit Whole Life en versant une unique prime d'un montant important pour emprunter ensuite la valeur de rachat sans être soumis à des taxes.

La législation TAMRA définit un seuil maximal de montant de primes versées durant les sept premières années de la vie du contrat. Le dépassement de ce seuil conduit le contrat Whole Life à être traité comme un MEC (Modified Endowment Contract). Ceci engendre une taxation des prêts ou des retraits effectués sur le contrat.

Tout changement sur le contrat, tel qu'une augmentation de la valeur nominale ou l'ajout de rider, entraîne le renouvellement de la période de 7 ans.

## Conclusions du chapitre II :

Nous avons établi que les primes rythment la vie du contrat et sont d'une importance majeure dans la problématique du rachat. En effet, leur montant est fixé lors de la souscription en fonction de caractéristiques de l'assuré. Nous avons l'intuition que le comportement de rachat des assurés sera différent selon les primes versées. De plus, la fréquence de versement impacte certainement leur décision : des versements fréquents majorent les frais et pourraient créer un sentiment de lassitude.

De plus, l'étude des caractéristiques du produit nous suggère que l'assureur veut dissuader l'assuré de racheter son contrat très tôt via une valeur de rachat nulle jusqu'à la troisième année d'ancienneté. Enfin l'assureur propose une série d'options qui ont pour but d'éviter que l'assuré ne rachète son contrat par besoin de liquidité immédiate.

Nous allons affiner nos intuitions sur les facteurs de risques développées jusqu'à présent en réalisant des statistiques descriptives sur notre portefeuille.

### III. Statistiques descriptives et sélection des variables explicatives

Dans un premier temps, nous nous concentrons sur la sélection de facteurs de risques caractéristiques de l'assuré. Pour ce faire, nous allons réaliser des statistiques descriptives sur notre portefeuille afin d'évaluer l'impact des caractéristiques de l'assuré sur la durée de vie de son contrat. Dans un second temps, nous analyserons le taux de rachat historique observé sur le portefeuille par rapport au contexte financier. L'objectif étant de sélectionner une variable conjoncturelle à intégrer dans notre modélisation.

#### III.1 Description du portefeuille

Le portefeuille est composé de 29 321 contrats, souscrits entre janvier 1995 et juillet 2010. Le portefeuille a été arrêté en juillet 2010. Les données à notre disposition pour chaque contrat sont les suivantes :

Variable	Modalités et répartition au sein du portefeuille	Remarques
Date de souscription du contrat	Dates comprises entre 1995 et 2010	-Notre base de données commence en janvier 1995, et nous n'avons dans notre portefeuille que des contrats souscrits à partir de janvier 1995. - La base de données s'achève en juillet 2010. Pour les contrats encore en vigueur à cette date nous n'avons pas l'information sur leur éventuel rachat après cette date. Nous avons donc une censure administrative fixe à droite.
Date de fin du contrat	Relative à chaque contrat.	Date non renseignée si le contrat est encore en vigueur à la date d'arrêt du portefeuille (juillet 2010).
Genre de l'assuré	Female 50%  Male 50%	La répartition du portefeuille par genre est proche de 50/50 (fréquences arrondies)

Fréquence de paiement de la prime	Monthly 36% Quaterly 19% Semi-annual 7% Annual 23% Other 15%	La modalité "Other" correspond aux versements libres.
Classe de risque de l'assuré (Risk State)	Substandard Smoker 2% Substandard nonSmoker 2% Standard Smoker 34% Standard nonSmoker 49% Preferred nonsmoker 11% Preferred Smoker 2%	Pour tarifier le contrat, l'assureur prend en compte ces classes de risques. Elles prennent en compte l'état de santé de l'assuré et de son statut fumeur/non fumeur. L'assuré est rattaché à l'une de ses classes après un examen médical. Plus l'assuré est considéré comme risqué, plus sa prime est élevée. De même, un individu fumeur paie une prime plus importante.
Age de souscription (Underwriting age)	Young 47% Middle 34% Old 19%	Le découpage en tranches d'âge est le suivant : Young [0-35 ans] Middle ]35 ans -55 ans] Old ]55 ans -85 ans]
Living Place	EastCoast 20% WestCoast 5% Other 75%	
Prime Annualisée	prime annuelle moyenne = 561 \$ prime maximale = 6 950 \$ prime médiane = 403 \$	Variable continue qui correspond au montant de la prime versée annuellement par l'assuré.
Accident Death Amount	Rider 16% NoRider 84%	Cette variable indique la souscription ou non d'une option permettant d'augmenter l'indemnité de décès en cas d'un décès accidentel.
Surrender Bit	Variable binaire qui vaut : - 1 si le contrat a fait l'objet d'un rachat total (38%) -0 sinon (62%).	Notons que si le contrat n'est pas racheté (Surrender Bit=0), à la date d'arrêt du portefeuille (Juillet 2010) ; celui-ci est soit en cours (In Force) soit clos pour une raison autre que le rachat.

Etat du contrat	<p>In Force 49%</p> <p>Surrender 38%</p> <p>Death 4%</p> <p>Other Lapses 9%</p>	<p>Il s'agit de préciser l'état du contrat en juillet 2010. Ce dernier peut être toujours en vigueur (In Force) ou clôturé. Trois types de fin de contrat sont alors possibles :</p> <ul style="list-style-type: none"> <li>- Soit l'assuré est décédé (Death)</li> <li>- Soit le contrat a été racheté (Surrender)</li> <li>- Soit le contrat est clos pour une autre raison (Other Lapses).</li> </ul>
-----------------	---	--

**Tableau 5 Les différentes variables au sein du portefeuille**

La répartition des assurés selon les variables du portefeuille peut être visualisée à l'aide de diagrammes et boxplot. (Cf. annexes)

➤ Distribution du montant de la prime annuelle

Le boxplot du montant des primes annuelles nous apporte une information supplémentaire concernant la distribution des primes. En effet, nous pouvons constater une asymétrie : la moyenne (560,9 \$) est supérieure à la médiane (402,6 \$).

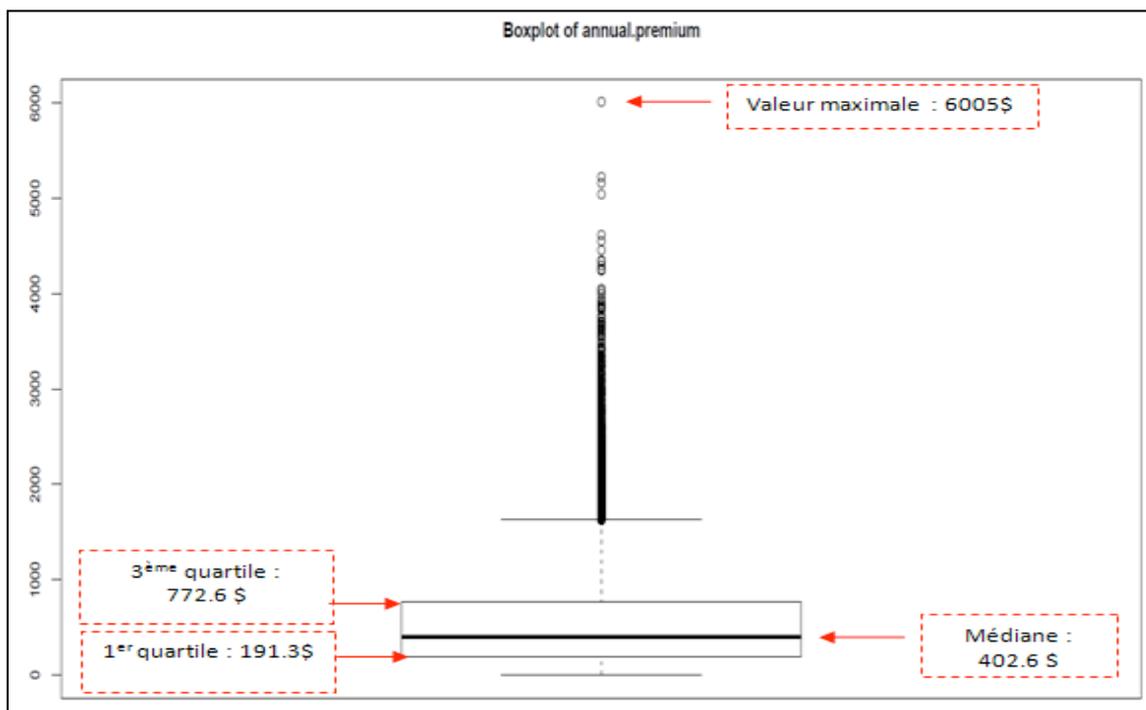


Figure 2 Boîte à moustache de la distribution de la prime

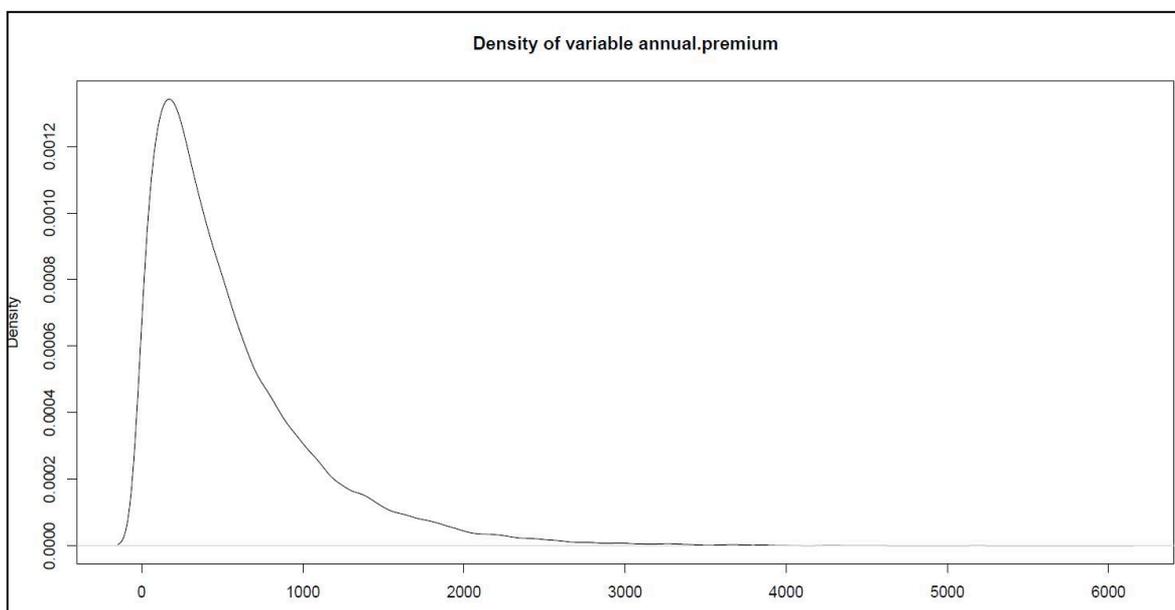


Figure 3 densité de probabilité de la prime

Ainsi, 75% des primes annuelles du portefeuille sont inférieures à 772,6 \$. Près de la moitié des primes du portefeuille sont inférieures à 402,6 \$. Dans la suite de notre étude nous allons segmenter cette variable en classes.

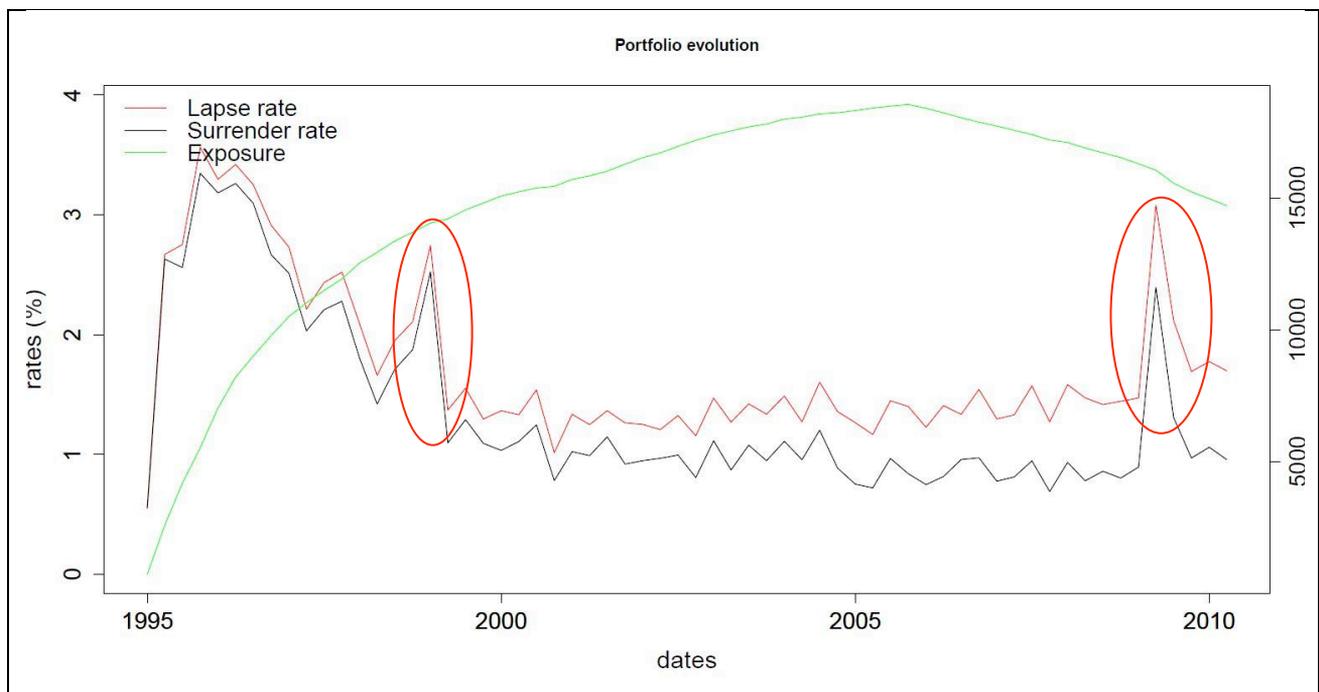
Nous avons ainsi décrit les différentes variables présentes dans le portefeuille. Notre objectif est maintenant d'établir quelles sont les variables qui influencent le plus le rachat. Dans un premier temps, nous allons étudier la dynamique du taux de rachat sur notre portefeuille, puis étudier l'influence des différentes variables sur la durée de vie des contrats.

## III.2 Etude dynamique des clôtures de contrat

### III.2.1 Le taux de rachat historique du portefeuille

Nous observons sur ce graphique le nombre de contrats en cours (courbe verte) par trimestre. Les contrats peuvent être clôturés suite à trois causes :

- Décès de l'assuré
- Rachat du contrat
- Autres causes (maturité notamment, défaut de paiements ...) regroupées au sein de la modalité « Other Lapses »



**Figure 4 évolution du portefeuille entre le lancement du produit et la censure administrative**

La compagnie d'assurance a lancé le produit en 1995. La courbe verte « Exposure » représente le nombre de contrats en vigueur dans le portefeuille. Elle présente une forte croissance jusqu'en 2005, due au lancement du produit.

Le taux de clôture est calculé pour chaque trimestre (i.e. nombre de contrats clôturés divisé par le nombre de contrats en cours lors du trimestre). Le taux de clôture est tracé en rouge. On observe deux pics majeurs de clôture de contrats en janvier 1998 et en avril 2009. Cependant le nombre de contrats clôturés comprend le nombre d'assurés ayant racheté, le nombre d'assurés décédés et le nombre de contrats clos pour d'autres raisons. Nous allons nous intéresser plus spécifiquement aux rachats.

De même le taux de rachat correspond au nombre de contrats clôturés suite à un rachat divisé par le nombre de contrats en cours lors du trimestre. Nous observons que les taux de clôture et les taux de rachat ont des évolutions similaires. Ceci laisse penser que c'est le comportement de rachat qui domine l'évolution du taux de clôture. En effet, il semble légitime de penser que le taux de décès et de contrats

arrivant à maturité demeurent à peu près constant pour chaque trimestre. Ce sont par conséquent les pics de rachats qui conditionnent les pics de clôtures.

Nous avons visualisé l'évolution du taux de rachat du portefeuille entre 1995 et 2010. Cependant, ce graphique ne nous donne pas d'informations sur l'évolution du taux de rachat en fonction de l'ancienneté du contrat. Or c'est ce que nous allons modéliser avec les modèles de survie.

### III.2.2 Taux de rachat en fonction de la durée de vie des contrats

Le taux de rachat représenté sur le graphique ci-dessous a été calculé de la façon suivante :

$$\text{Taux de rachat}(\text{trimestre } i) = \frac{\text{Nombre de rachats effectués au } i\text{ème trimestre}}{\text{Nombre de rachats totaux}}$$

Ce graphique nous permet donc de visualiser la distribution des rachats en fonction de l'ancienneté du contrat. On observe que le taux de rachat décroît avec l'ancienneté du contrat. Il se stabilise à partir du 20ème trimestre soit à partir de la 5ème année de la vie du contrat. Les variations observées à partir du 57ème trimestre (soit de la 14ème année) doivent être considérées comme des effets de bord causés par un faible nombre de contrats concernés (seuls 30 contrats ont une durée de vie supérieure à 57 trimestres).

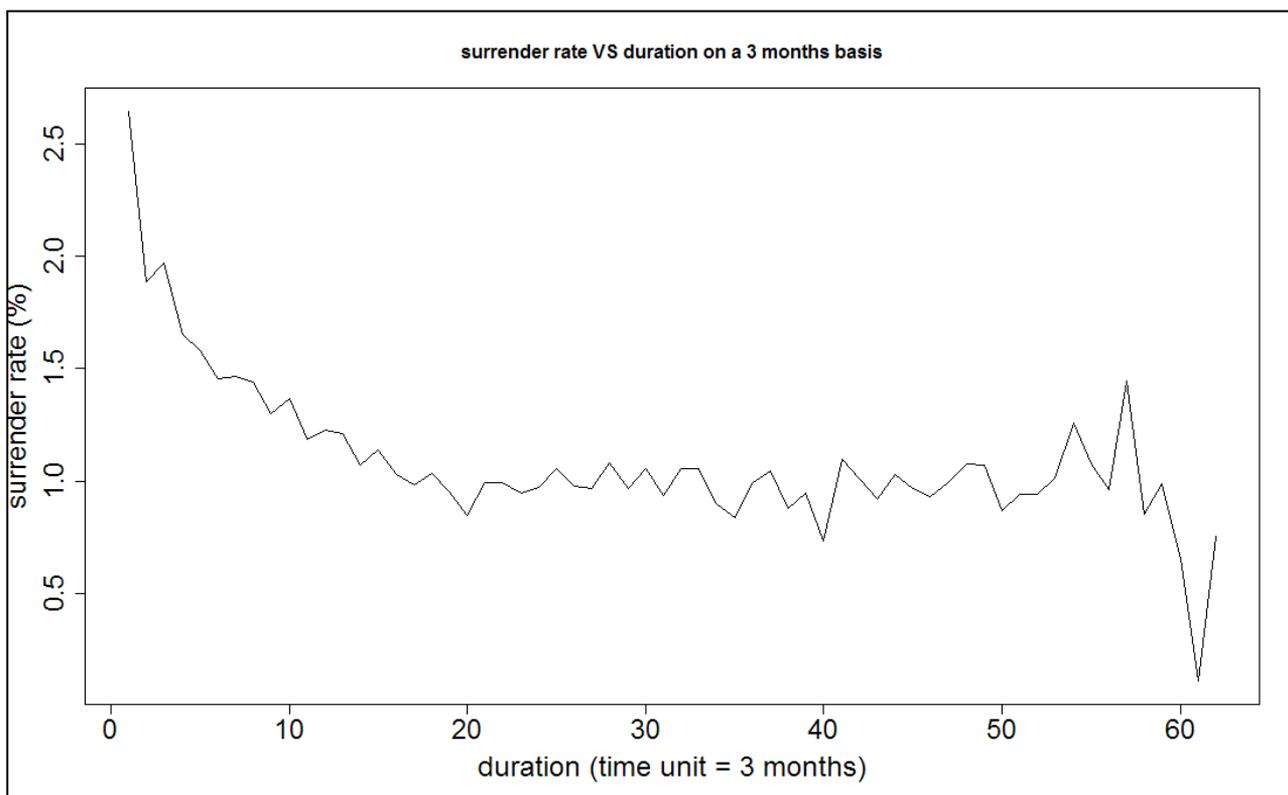


Figure 5 évolution du taux de rachat par trimestre

### III.2.3 Evolution du nombre de rachats dans le temps en fonction de la date de souscription et de la durée du contrat

Nous allons croiser les informations acquises via les visualisations du taux historique et du taux en fonction de l'ancienneté du contrat. Pour cela, nous allons visualiser sur un même graphique l'effet cohorte et l'effet de l'ancienneté du contrat.

Une cohorte désigne l'ensemble des assurés qui ont souscrit durant une même période. Dans notre cas, le pas est trimestriel : il s'agit donc de l'ensemble des contrats souscrits au cours d'un même trimestre.

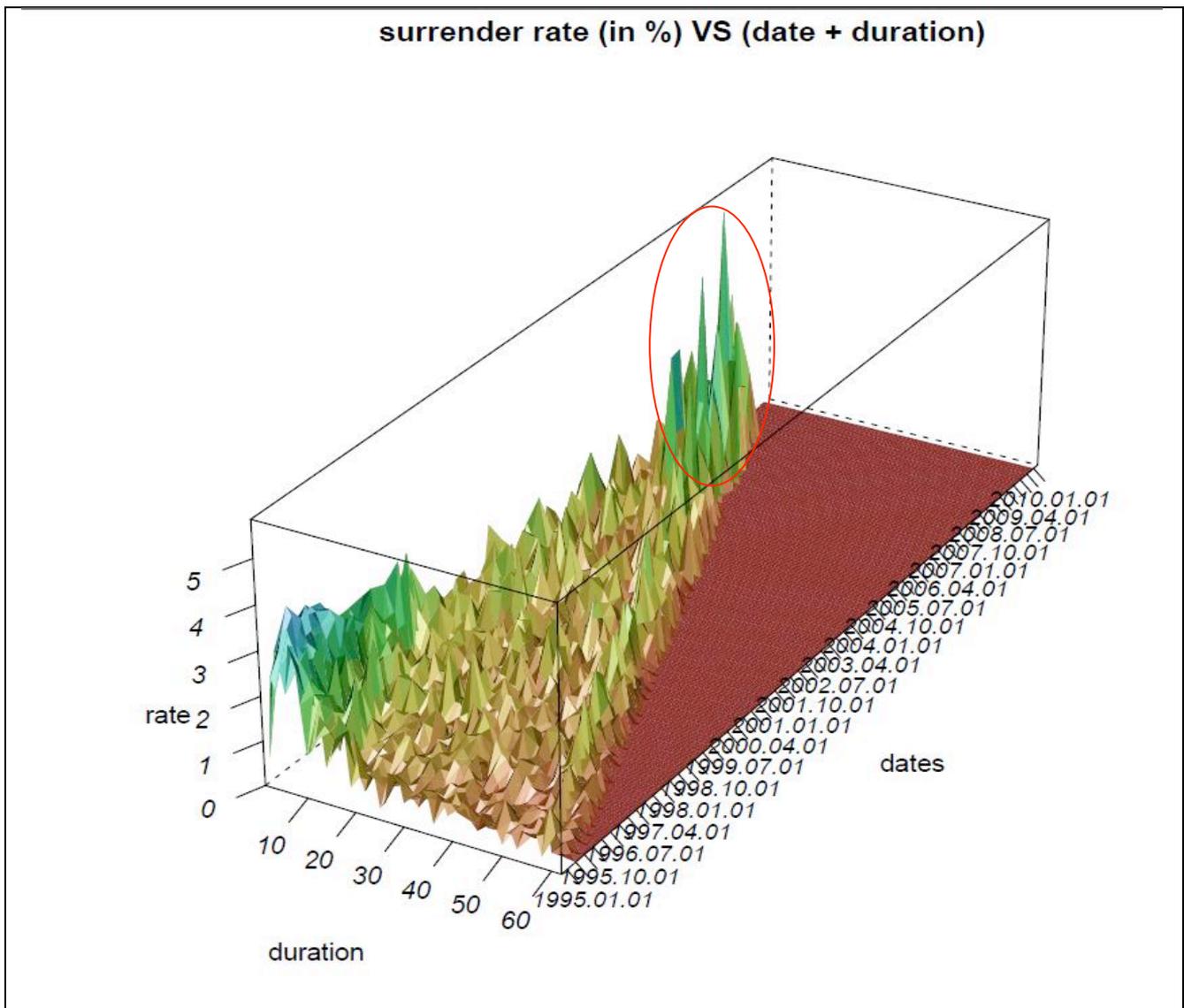


Figure 6 Taux de rachat par cohorte

*Remarque : la durée des contrats (variable « duration » est en trimestre).*

Nous pouvons tout d'abord retrouver les analyses effectuées dans le III.2.1 et le III.2.2 Nous avons en effet souligné dans le III.2.2, la décroissance du taux de rachat en fonction de l'ancienneté du contrat : pic de rachats sur les 3 premières années de la vie du contrat puis stabilisation à partir de la 5ème année.

Dans le III.2.1, nous avons établi l'existence de pics de rachats en 1998 et en 2009. La visualisation du taux par cohorte nous permet de constater que l'on observe une évolution similaire du taux de rachat en fonction de l'ancienneté (décroissance puis stabilisation) du rachat mais avec des niveaux plus ou moins élevés selon l'année de souscription. En effet, pour les cohortes entre 2008 et 2010, le pic de rachats des premières années est particulièrement élevé par rapport aux autres cohortes.

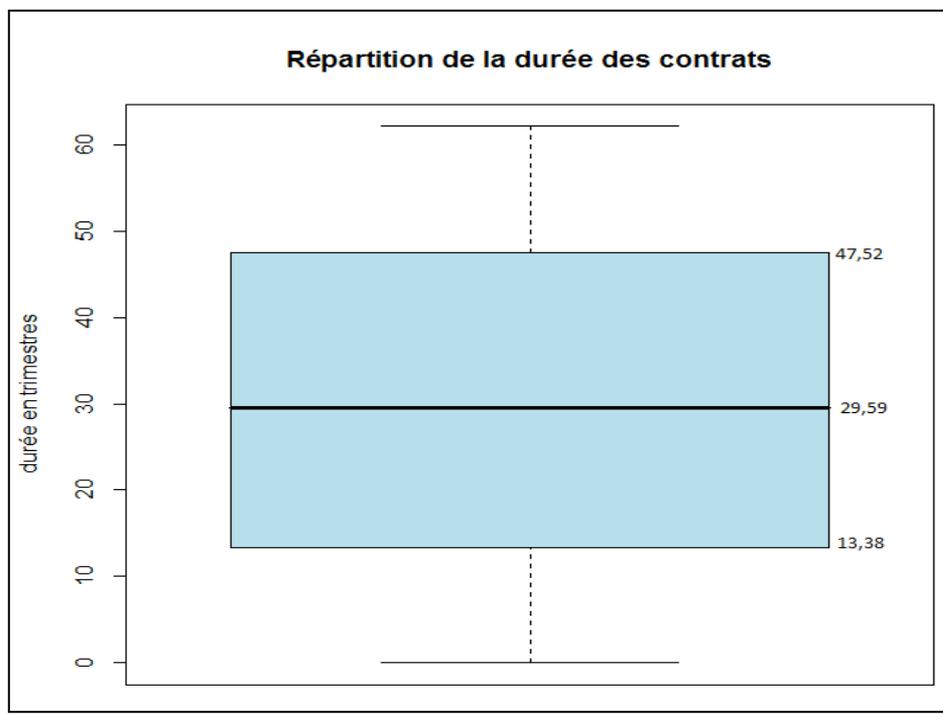
Cette étude nous permet de formuler une hypothèse très importante : la probabilité de racheter son contrat n'est pas constante dans le temps. Elle décroît fortement durant les trois premières années du contrat, puis se stabilise à partir de la 5<sup>ème</sup>.

De plus, il existe un effet cohorte. En effet, selon la date d'entrée dans le portefeuille, le phénomène observé est plus ou moins accentué.

### **III.3 Sélection des variables explicatives caractéristiques de l'assuré**

Nous allons étudier l'influence des différentes variables à notre disposition (âge, classe de risque ...) sur la durée du contrat d'assurance. Cette étude est effectuée dans un premier temps de façon qualitative à l'aide de box plots.

### III.3.1 Etude de la durée des contrats sur tout le portefeuille



**Figure 7** boîte à moustache de la durée des contrats

La durée médiane des contrats du portefeuille est de 30 trimestres à peu près, soit 8,75 années. La durée maximale est de 62 trimestres. 75% des contrats sont clôturés après 13,38 trimestres, soit un peu plus de 3 ans. Cela est cohérent avec les caractéristiques du produit et la volonté de l'assureur de conserver pendant au moins 3 ans les assurés en portefeuille.

### III.3.2 Etude de la durée des contrats en fonction des rachats totaux

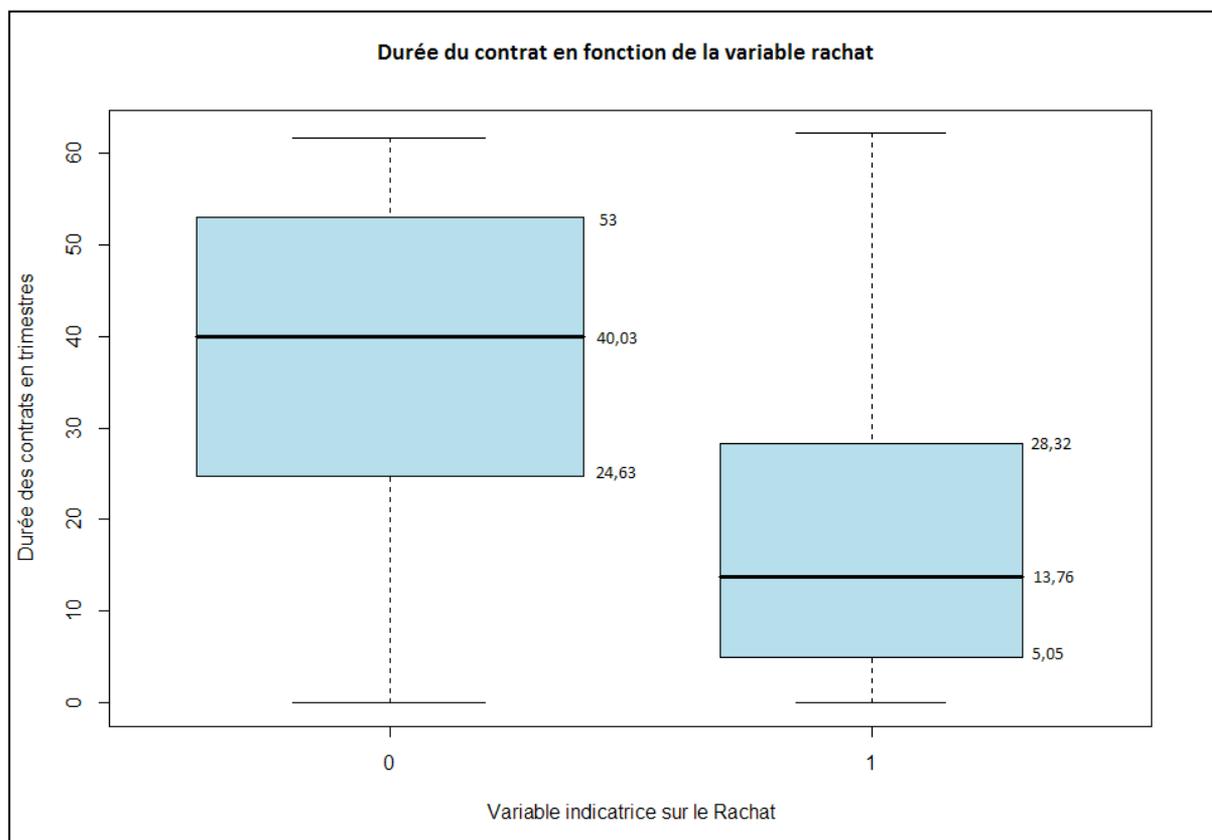


Figure 8 boîte à moustache de la durée des contrats en fonction de la variable rachat

De ce graphique nous remarquons que pour ceux qui n'ont pas fait de rachats totaux, la durée médiane des contrats est d'environ 10 ans (40 trimestres), tandis que pour ceux qui ont fait des rachats totaux celle-ci est d'environ 3 ans et demi (14 trimestres). Ceci confirme l'observation effectuée au III.2 : presque la moitié des rachats totaux s'effectue durant les trois premières années de la vie du contrat.

### III.3.3 Etude de la durée des contrats en fonction de la variable « acc.death.amount »

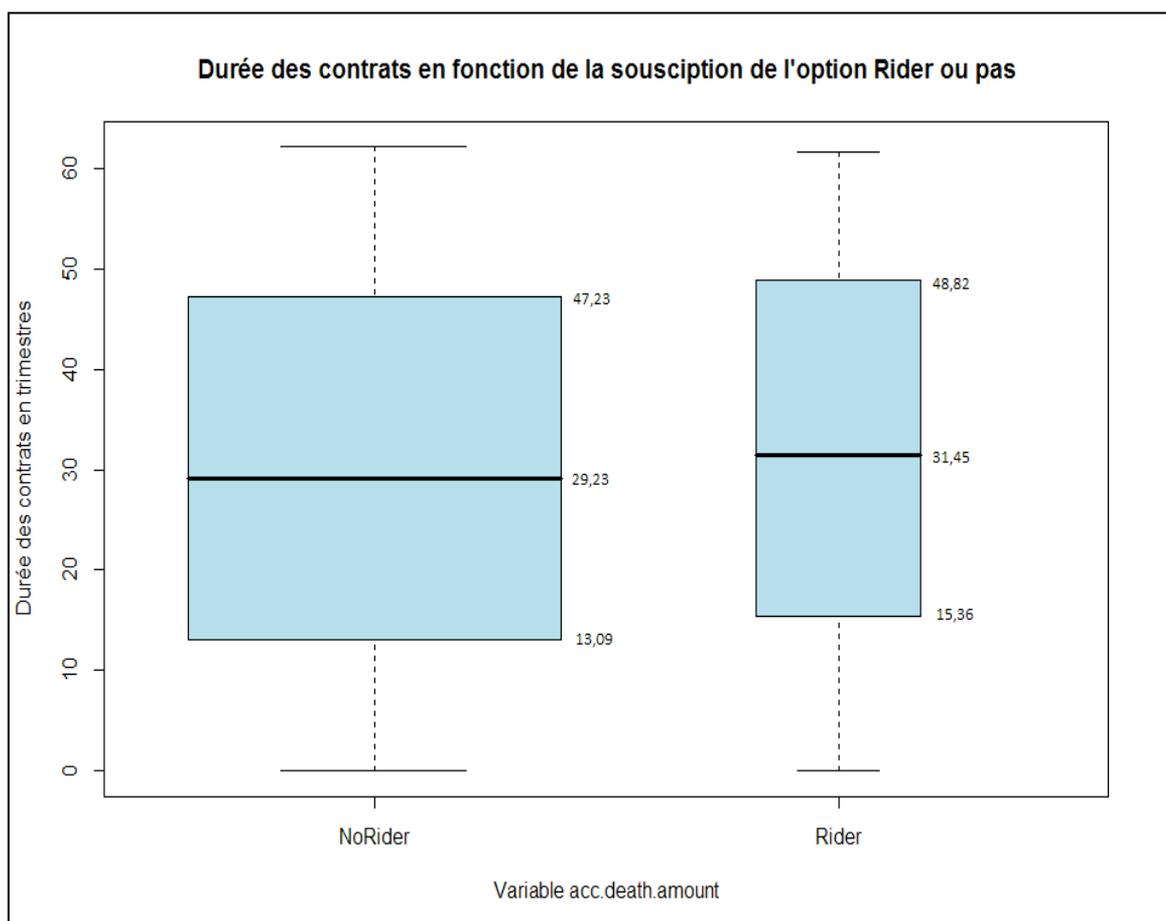


Figure 9 boîte à moustache de la durée des contrats en fonction de l'option Rider

L'option rider, présentée au II.6, correspond à une extension de couverture d'assurance. Elle permet de toucher un capital décès plus important en cas de décès accidentel.

Nous pouvons observer que les contrats avec une option « rider » présentent une durée médiane supérieure d'un peu plus de 2 trimestres. Ainsi, cela confirme une intuition formulée lors de l'étude du produit Whole Life au II.6 : une option de type « rider » **fidélise le client et l'incite à conserver son contrat.**

### III.3.4 Etude de la durée des contrats en fonction du genre

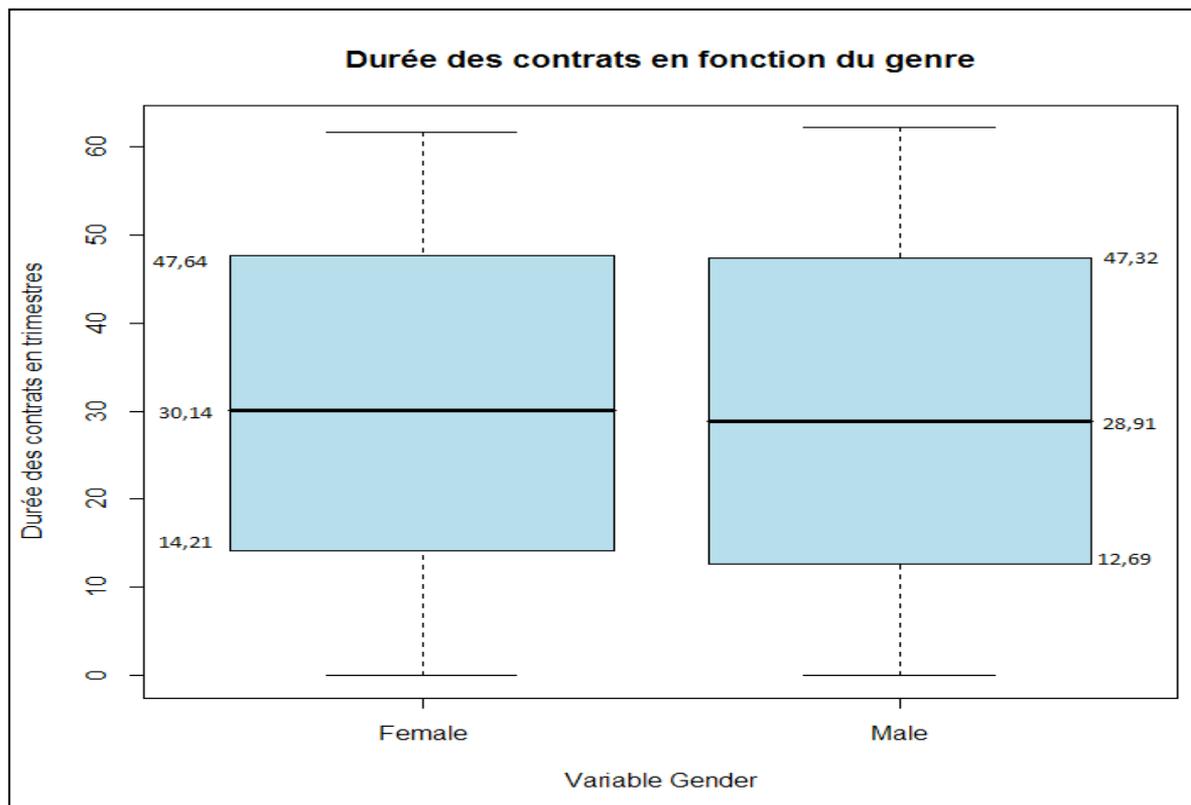


Figure 10 boîte à moustache de la durée des contrats en fonction du genre

Nous remarquons que la répartition de la durée des contrats est quasiment la même par genre. Nous avons une durée médiane des contrats de 7ans et demi pour les femmes, et de 7 ans et un trimestre pour les hommes. Ceci vient confirmer une intuition développée lors de notre étude bibliographique qui est que le **sexe n'influence pas le rachat**. Nous vérifierons ultérieurement, par des tests statistiques, si nous pouvons écarter la variable sexe de notre modélisation.

### III.3.5 Etude de la durée des contrats en fonction des classes d'âge

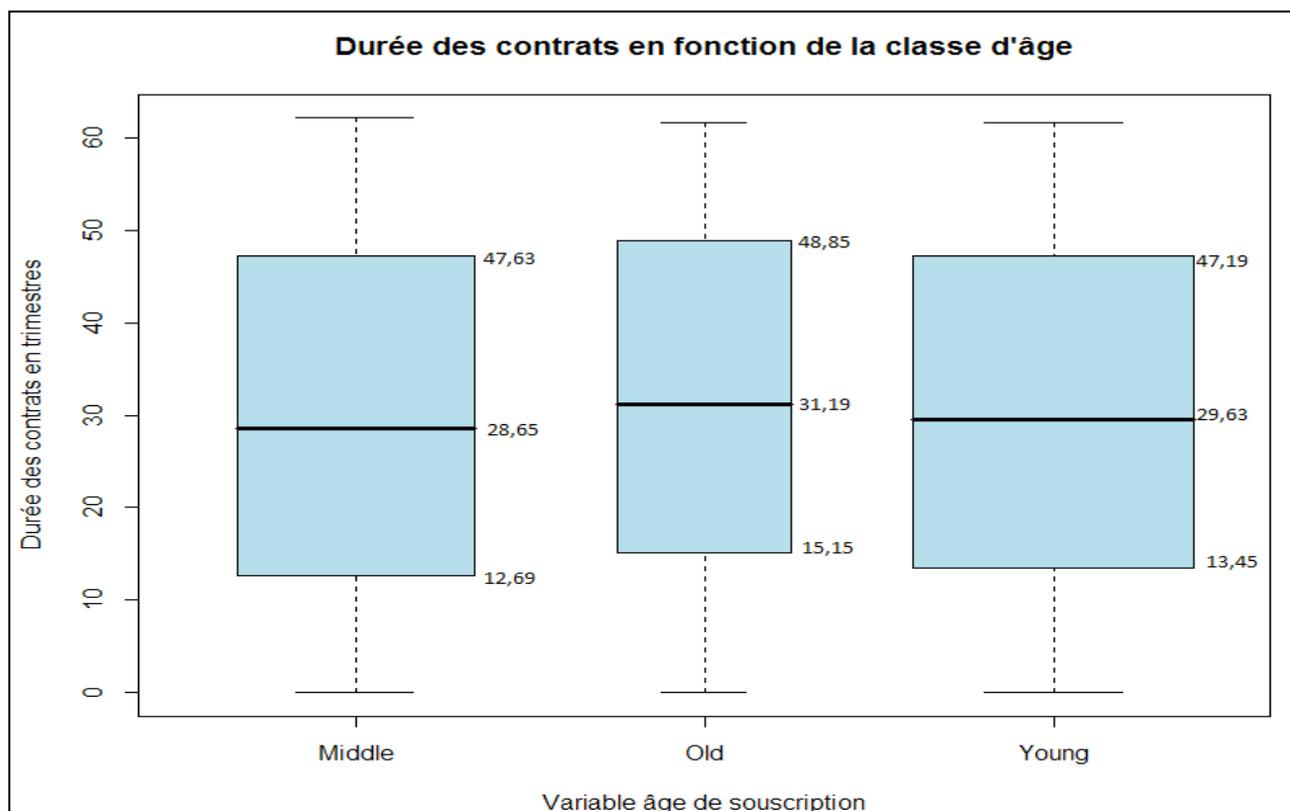


Figure 11 boîte à moustache de la durée des contrats en fonction de la classe d'âge

Nous avons constaté précédemment que près de 47% des assurés du portefeuille sont jeunes [0,35[, 34% sont d'âge moyen ([35,55[), et une proportion plus faible (19%) est âgée ([55-85[). Nous observons une légère différence de durée de vie des contrats. Les assurés appartenant à la classe la plus âgée (« Old ») semblent conserver leur contrat plus longtemps que les deux autres classes d'âge (âge inférieur à 55 ans). Cette observation est cohérente avec les réflexions menées lors de notre étude bibliographique. Les personnes jeunes présentent plus de risque de racheter que les personnes plus âgées car elles sont généralement plus en quête de liquidité afin de financer des projets que les personnes retraités.

### III.3.6 Etude de la durée des contrats en fonction des classes de risque

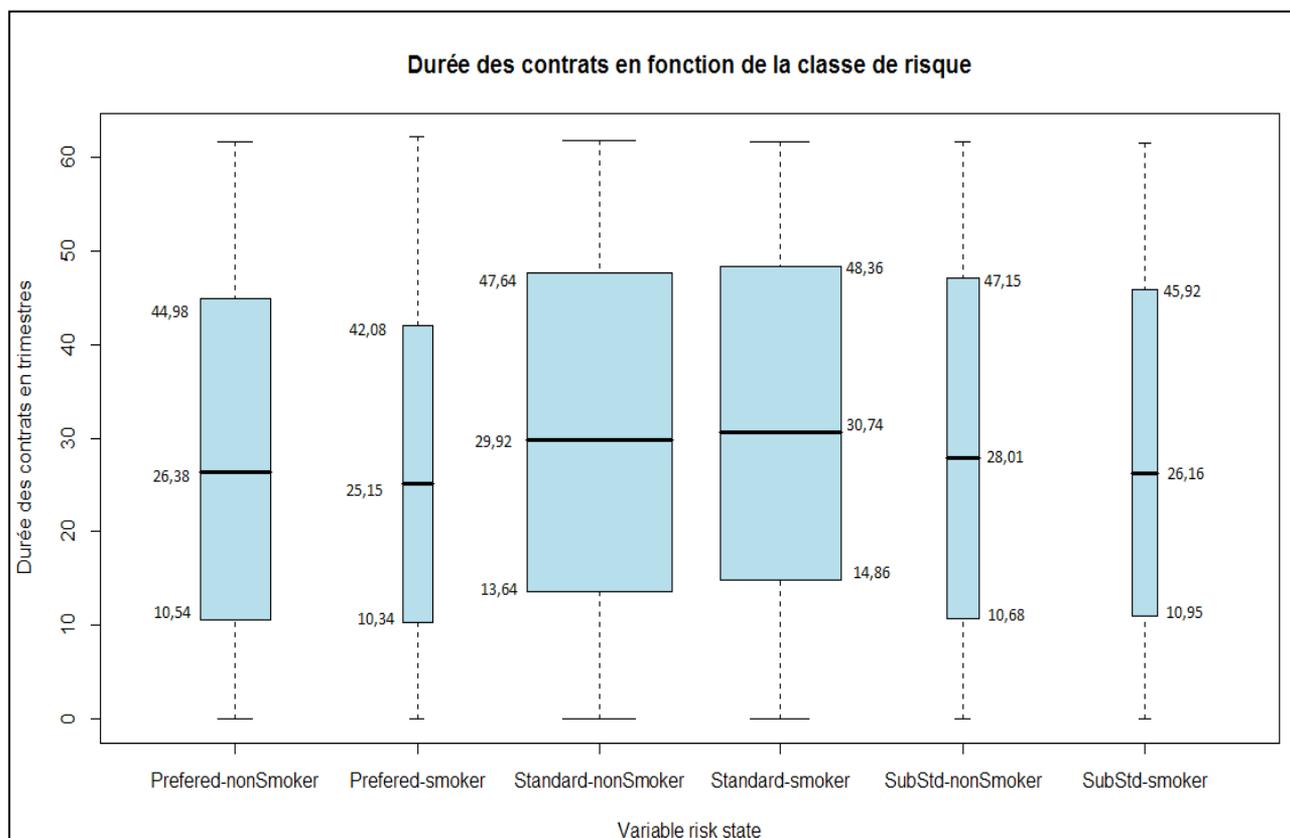


Figure 12 boîte à moustache de la durée des contrats en fonction de la classe de risque

De ces tracés de boîtes à moustaches, nous déduisons qu'il serait possible de grouper les modalités de la variable classes de risque en deux catégories :

- la classe de risque « Standard » : risques dans la moyenne.
- les classes de risques « extrêmes » : composé de la classe « Preferred » (meilleurs risques) et de la classe « Substandard » (risques les plus grands).

En effet, les contrats des assurés de la classe de risque « Standard », présente des durées de vie médianes plus élevées que ceux des assurés de risques « extrêmes ».

Concernant les risques « Preferred », ces assurés sont les plus choyés par les assureurs car ils présentent un risque plus faible que la moyenne. Par conséquent, il est possible qu'ils soient sollicités par d'autres assureurs et fassent jouer la concurrence plus facilement. Ils rachètent alors leur contrat afin de bénéficier de tarifs plus intéressants ailleurs.

Concernant les risques « Substandard », c'est une explication inverse qui s'impose. En effet, ces derniers paient des primes plus élevées que leurs homologues à cause de leur niveau de risque considéré comme supérieur. Par conséquent, ils peuvent se lasser de payer cher leur contrat d'assurance et décident alors de racheter. Pour les classes de risque « extrêmes », les fumeurs semblent racheter plus rapidement que les non-fumeurs. Une explication pourrait être que, à profil de santé comparable, un fumeur paie une prime plus élevée qu'un non-fumeur.

### III.3.7 Etude de la durée des contrats en fonction du lieu de résidence

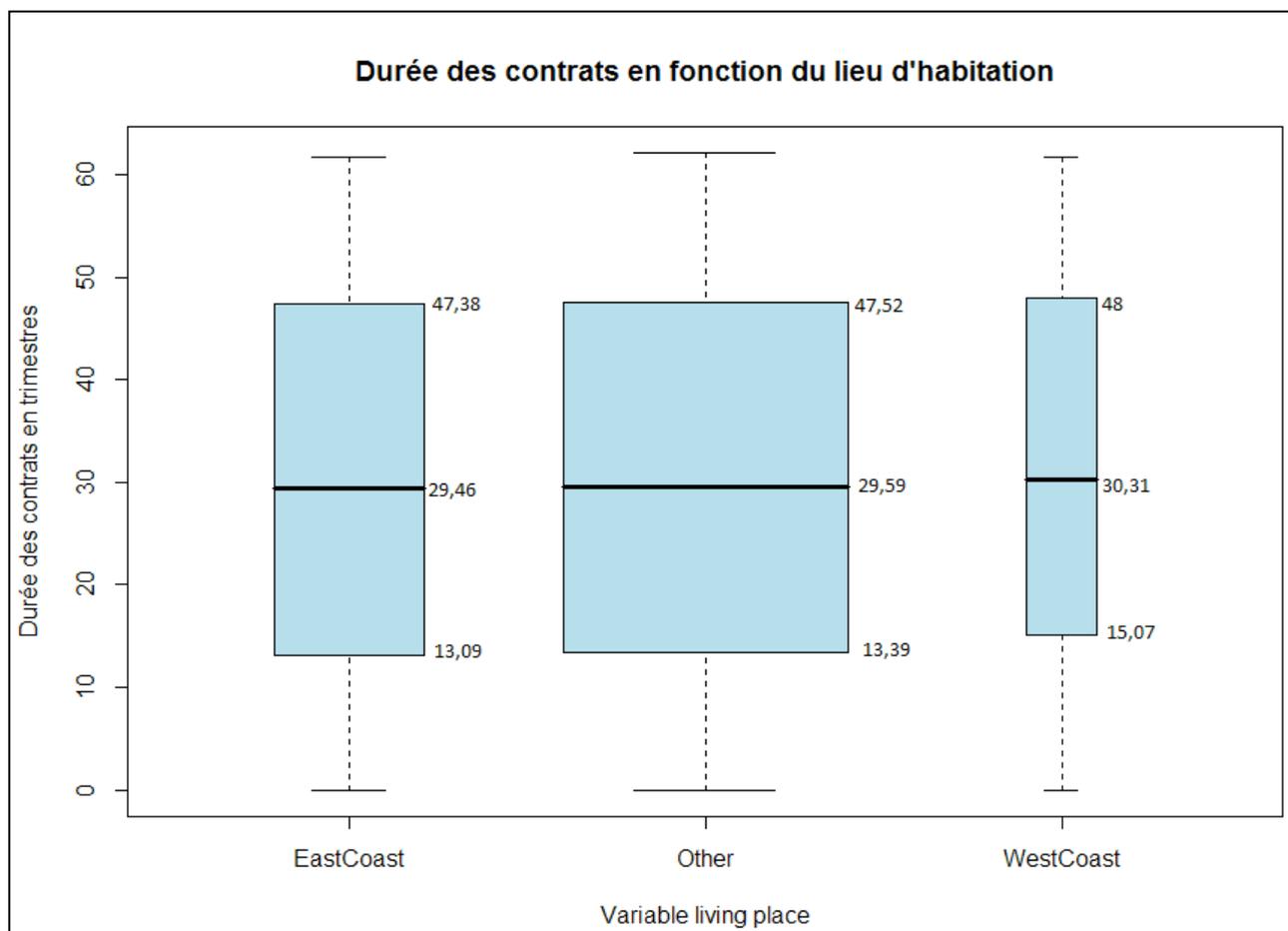


Figure 13 boîte à moustache de la durée des contrats en fonction du lieu d'habitation

Notre portefeuille de clients est majoritairement composé de clients résidants en dehors des côtes Est et Ouest des Etats-Unis (modalité « Other »), puis de clients résidants sur la côte Est, et enfin des clients résidants sur la côte Ouest. La répartition des durées suivant les lieux de résidence est quasiment la même.

Cette variable ne semble pas spécialement impacter les comportements de rachat, une hypothèse que nous tenterons de confirmer au moment de la sélection de variable lors de la modélisation de type analyse de survie.

### III.3.8 Etude de la durée des contrats en fonction de la fréquence de paiement des primes

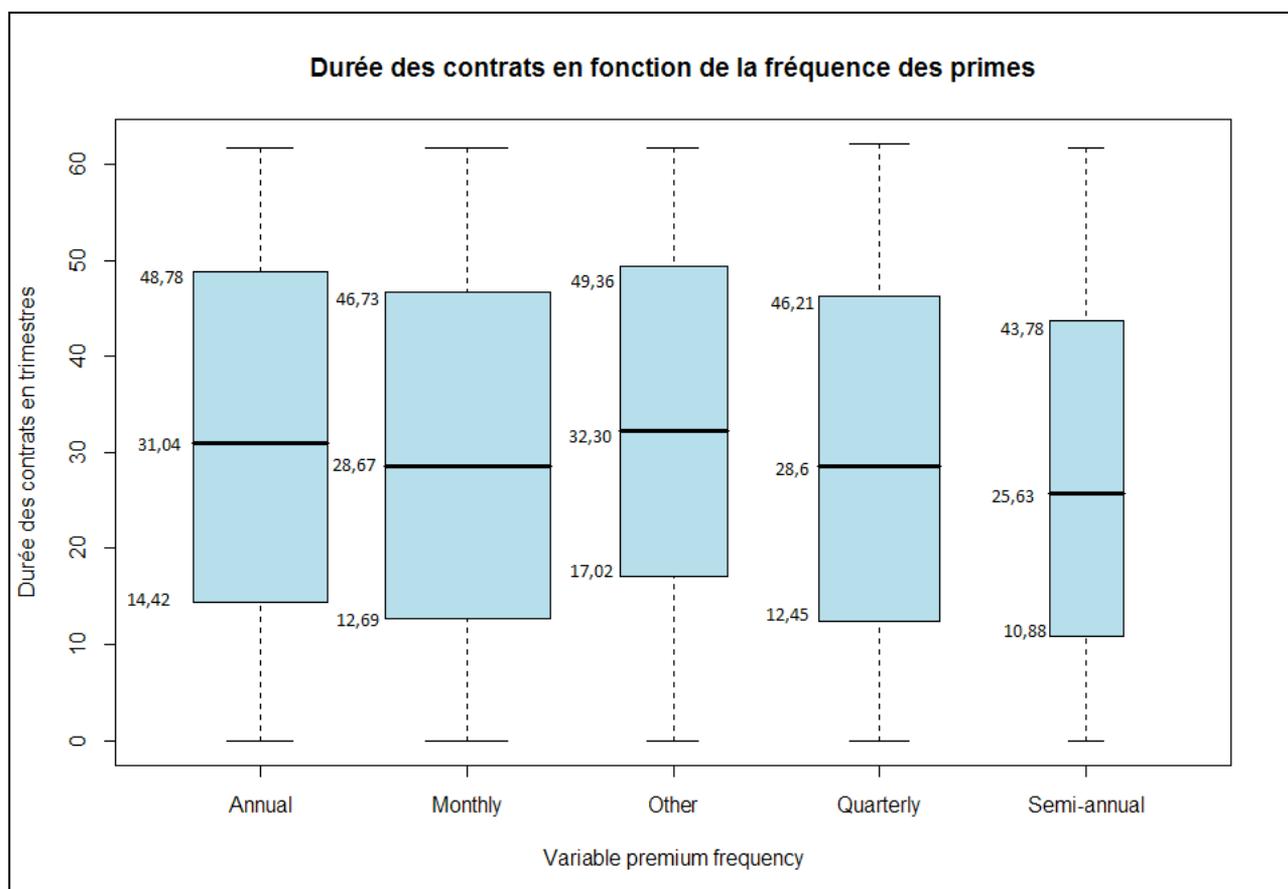


Figure 14 boîte à moustache de la durée des contrats en fonction de la fréquence de paiement des primes

De ces boîtes à moustache on remarque que les modalités de fréquence de paiement des primes qui ont des durées de contrat les plus importantes sont les modalités versement libre (« Other »), et versement annuel. Cela confirme les intuitions que nous avons, plus le client paye sa prime fréquemment, moins il a tendance à conserver le contrat.

### III.3.9 Etude de la durée des contrats en fonction de la classe de la prime

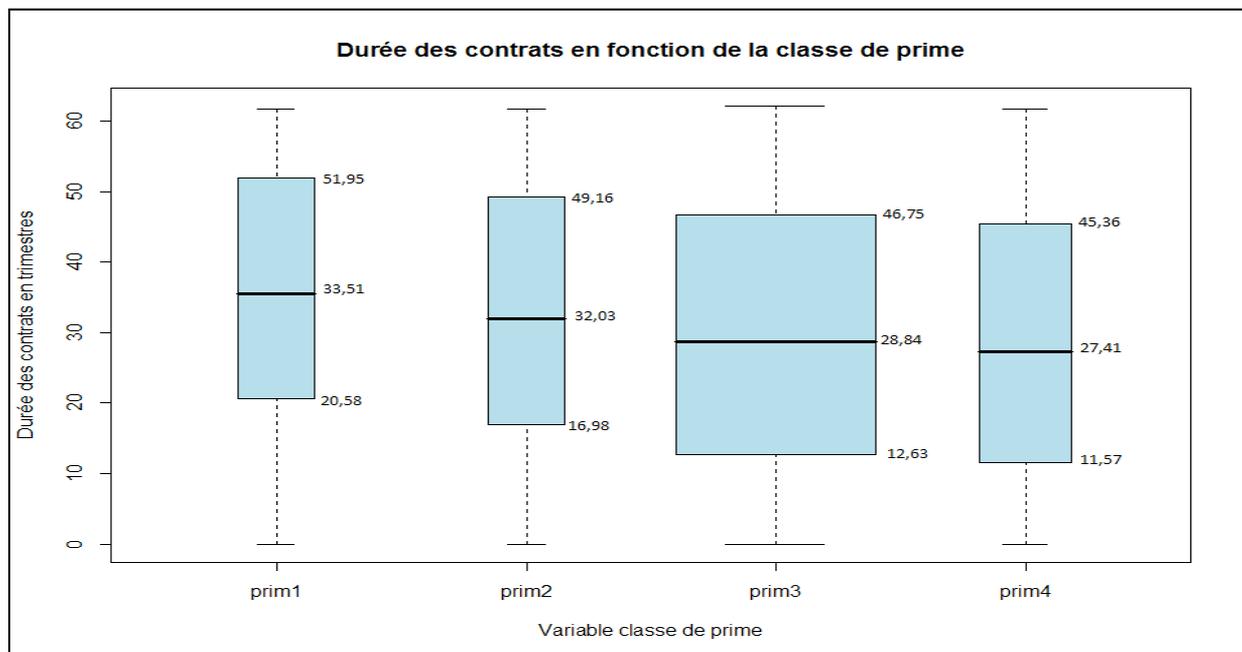


Figure 15 boîte à moustache de la durée des contrats en fonction de la classe de prime

Après avoir segmenté la prime en 4 classes (les classes sont nommées par ordre croissant) à l'aide d'une méthode Anova (voir détails en annexe), nous traçons le boxplot de la durée des contrats en fonction des classes de paiement des primes. On déduit que plus la prime est élevée, plus le client aura tendance à ne pas rester longtemps. La majorité des clients 66% à peu près sont dans la classe prim3, et payent une prime annualisée comprise entre 78 et 152 \$.

#### Conclusion :

Cette première approche qualitative nous permet tout d'abord de constater que les rachats totaux s'effectuent en moyenne au bout de 3 ans et demi (i.e 14 trimestres). Pour les contrats qui ne font pas l'objet d'un rachat, la durée de vie est en moyenne de 10 ans (i.e 40 trimestres).

Ensuite, les différents boxplots réalisés nous ont permis de visualiser l'influence des différentes variables sur la durée de vie du contrat et de confirmer des intuitions formulées à l'issue de notre étude bibliographique.

Il semble en effet que les variables « acc.death.amount », « underwriting.age », « risk.state », « premium », soient significatives sur le rachat. Les variables « gender » et « living place » semblent, quant à elles, avoir peu d'impact. Nous allons par la suite vérifier ces observations par des tests statistiques.

### III.4 Réalisation de tests statistiques

Dans cette rubrique nous allons réaliser des tests statistiques afin d'appuyer ou éventuellement contredire nos premières analyses sur les variables clefs influant sur le rachat. De même nous allons tenter de dégager des profils types, autrement dit des profils plus à même de racheter leur contrat.

#### III.4.1 Test du chi 2 sur la variable rachat (« surrender.bit » dans la base de données »

Premièrement, nous effectuons un test du chi 2 entre la variable rachat (variable qui vaut 1 s'il y a eu rachat total de l'épargne du contrat, et 0 sinon), et toutes les autres variables qualitatives. Les autres variables qualitatives étant la tranche d'âge, le profil de risque, le sexe, la région et l'année de souscription (pour essayer de voir s'il y a éventuellement des effets rachats par cohorte).

La p-value pour le test est fixée à 5%. A titre de rappel, le test du chi 2 entre deux variables s'effectue sur des variables à modalités. Soient  $X$  et  $Y$  deux variables à modalités et  $n$  le nombre total d'observations. Supposons que la variable  $X$  ait  $p$  modalités et que la variable  $Y$  ait  $r$  modalités, le tableau de contingence est de la forme suivante :

$X \backslash Y$	$X_1$	...	$X_p$	Total
$Y_1$	$n_{11}$	$n_{1j}$	$n_{1p}$	$n_{1.}$
...	$n_{i1}$	$n_{ij}$	$n_{ip}$	$n_{i.}$
$Y_r$	$n_{r1}$	$n_{rj}$	$n_{rp}$	$n_{r.}$
Total	$n_{.1}$	$n_{.j}$	$n_{.p}$	$n$

Tableau 6 tableau illustratif du chi2

Avec  $\forall j \in \{1, \dots, p\} n_{.j} = \sum_{i=1}^r n_{ij}$  et  $\forall i \in \{1, \dots, r\} n_{i.} = \sum_{j=1}^p n_{ij}$

Les  $n_{i.}$  et  $n_{.j}$  étant les effectifs marginaux.

Les effectifs théoriques (effectifs en cas d'indépendance des variables) sont donnés par la formule suivante :

$$\overline{n_{ij}} = \frac{n_{i.} \times n_{.j}}{n}$$

La distance du chi 2  $D_{\chi}^2$  calcule l'écart entre les effectifs observés et les effectifs théoriques par la formule suivante :

$$D_{\chi}^2 = \sum_{i=1}^r \sum_{j=1}^p \frac{(\overline{n_{ij}} - n_{ij})^2}{\overline{n_{ij}}}$$

Sous l'hypothèse nulle ( $H_0$  étant l'indépendance des variables  $X$  et  $Y$ )  $D_{\chi}^2$  suit une loi du chi 2 à  $(r - 1) \times (p - 1)$  degrés de liberté ( $\chi_{(r-1) \times (p-1)}^2$ ).

Enfin pour un seuil  $\alpha$  fixé (5% dans notre cas), la valeur critique correspond au quantile de seuil  $1-\alpha$  de la variable  $\chi^2_{(r-1)\times(p-1)}$ . Si la statistique de test est plus grande que la valeur critique, l'hypothèse nulle  $H_0$  est rejetée ; sinon celle-ci ne l'est pas.

Test du chi2 : l'hypothèse nulle étant l'indépendance des variables					
	tranche d'âge	Profil de risque	Genre	Région	Année de souscription
Rachat total {0,1}	Rejet de l'hypothèse nulle	Rejet de l'hypothèse nulle	Rejet de l'hypothèse nulle	Non rejet de l'hypothèse nulle	Rejet de l'hypothèse nulle

Tableau 7 résultats des tests du chi2

Nous déduisons de ces tests que seule la caractéristique de la zone géographique n'influe pas de manière significative sur le fait de racheter ou non.

#### III.4.2 Tests statistiques sur la variable durée du contrat

Pour tester s'il y a un lien éventuel entre la variable donnant la durée des contrats (« duration ») et les variables qualitatives nous allons utiliser deux types de tests :

- le test de Wilcoxon si la variable qualitative concernée n'a que deux modalités
- le test de Kruskal Wallis si la variable concernée a au moins 3 modalités.

	tranche d'âge	Profil de risque	Genre	Région	Accident death amount
Test	Kruskal Wallis	Kruskal Wallis	Wilcoxon	Kruskal Wallis	Wilcoxon
Résultat	Rejet de l'hypothèse nulle				

Tableau 8 résultats tests de Kruskal Wallis

Nous déduisons de ces tests statistiques que toutes ces variables qualitatives impactent de manière significatives la durée de vie du contrat.

### III.4.3 Tests de corrélation

Nous allons maintenant tester la corrélation entre la variable duration et la variable montant annuel des primes (quantitative continue). En calculant le coefficient de corrélation linéaire de Pearson entre la durée des contrats et le montant des primes annuelles nous avons ( $\bar{x}$  et  $\bar{y}$  étant les moyennes empiriques de nos variables):

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = -0,07$$

Le coefficient de corrélation de Pearson n'est pas significatif (très proche de 0), un lien linéaire ne permettrait pas d'expliquer le lien entre la durée des contrats et le montant annuel des primes.

Afin d'essayer de savoir s'il existe un lien non linéaire entre la variable durée du contrat et montant annuel des primes, nous allons calculer la valeur du coefficient de Kendall pour nos deux variables. Le coefficient de Kendall est basé sur les rangs associés aux contrats lors du tri de nos variables aléatoires. Nous obtenons :

$$\rho_{Kendall} = -0,06$$

Le coefficient de corrélation de Kendall étant très proche de 0, ce qui nous incite à nous questionner afin de savoir si le montant de la prime influence la durée des contrats. Cependant nous avons vu qu'une segmentation judicieuse de la prime en classes amenait à de meilleures conclusions (les classes ayant des primes annualisées plus importantes, conservaient leur contrat moins longtemps).

### III.4.4 Analyse des correspondances multiples

Afin d'essayer d'analyser l'influence des autres variables du portefeuille sur la durée des contrats, nous allons effectuer une analyse des correspondances multiples. Ainsi, nous discrétisons les variables primes et durée des contrats en fonction de leurs fonctions de répartition (nous prenons les intervalles 0-1<sup>er</sup> quartile, puis 1<sup>er</sup> quartile-médiane, puis médiane-3<sup>e</sup> quartile, et enfin 3<sup>e</sup> quartile-1). Ainsi les durées sont regroupées en 4 groupes (des plus petites durées aux plus grandes dur1, dur2, dur3 et dur4), et les primes sont également regroupées en 4 groupes (des montants les plus petits aux plus grands prim1, prim2, prim3 et prim4). Pour éviter la confusion entre la modalité « Other » de la variable fréquence des primes, et la modalité « Other » de la variable lieu d'habitation, celle de la variable fréquence des primes a été renommée « Free ». Vu la faible fréquence des classes de risques « Substandard Smoker », « Substandard NonSmoker » et « Preferred Smoker », nous avons fusionné les 6 classes de risque en 3 classes de risque. Nous avons regroupé les « Substandard », les « Standard » et les « Preferred ».

Ensuite nous construisons nos axes factoriels sans prendre en compte l'état des contrats, puis nous projetons les modalités de la variable état du contrat sur le plan factoriel construit à l'aide des autres variables du portefeuille.

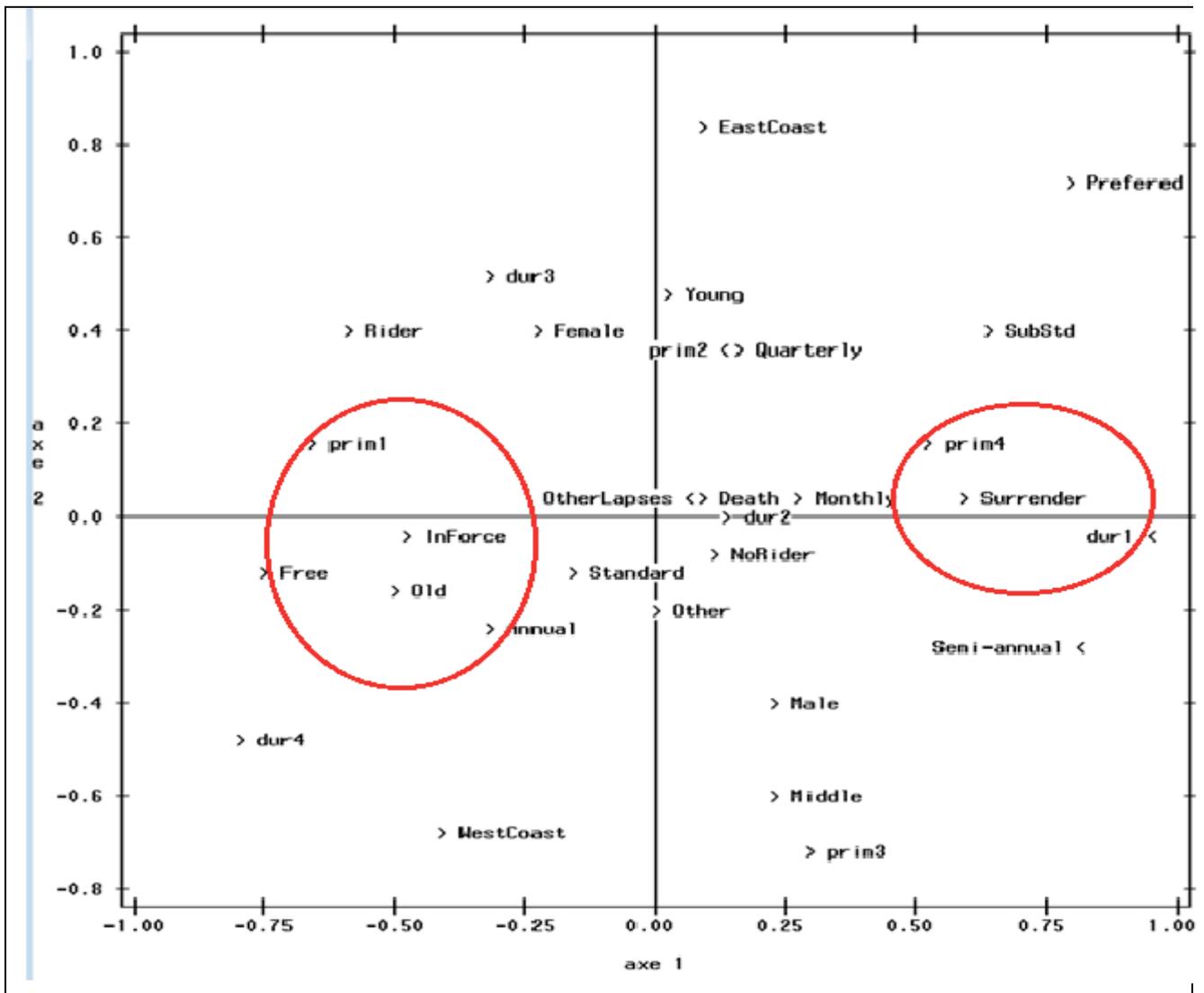


Figure 16 graphique plan factoriel de l'ACM

Nous allons essentiellement centrer notre analyse sur les modalités surrender (rachat), que nous opposerons à la modalité in force (en vigueur). Les modalités les plus contributives (voir annexe) pour l'axe 1 sont : {« dur1 », « dur4 », « prim1 », « prim4 », « rider », « free », « old », « preferred »}. Les modalités les plus contributives pour l'axe 2 sont : {« female », « male », « prim3 », « young », « middle », « preferred », « eastcoast »}. Lors de l'interprétation nous accorderons ainsi plus d'importance à ces modalités. Nous remarquons qu'au sein de notre portefeuille, les clients les plus propices au rachat sont ceux qui ont les primes les plus élevées. La durée de leur contrat est parmi les plus petites (on en déduit ainsi qu'il y a beaucoup de rachats dans les premiers mois voire les premières années après la souscription). Ce sont essentiellement des clients qui n'ont pas souscrits à l'option « Rider », et les classes de risques les plus proches sont les « Preferred » et les « Substandard ». Par opposition les clients ayant un contrat en vigueur sont essentiellement des personnes âgées, ceux ayant souscrits à l'option « Rider », ceux ayant des versements libres ou annuels pour leurs primes.

Conclusion: l'ACM réalisée nous permet de confirmer la pertinence des variables montant discrétisé de la prime, souscription de l'option rider, âge, classe de risque et fréquence de paiement des primes dans la modélisation du rachat. Nous allons maintenant sélectionner une variable externe pertinente afin de traduire l'influence du contexte financier sur notre modèle.

### III.5 Prise en compte du contexte économique

L'observation du taux de rachat du portefeuille, entre janvier 1995 et juillet 2010, (partie III.1.2) nous a permis de constater l'existence de deux pics majeurs de rachats en octobre 1998 et en avril 2009. Ces deux pics de rachats ne peuvent être expliqués par des variables caractéristiques de l'assuré car ces dernières sont fixes dans le temps. De plus, ces deux pics surviennent dans un contexte économique agité : crise financière en Asie en 1998 et crises des « subprimes » entre 2007 et 2009. Par conséquent, il semble pertinent d'inclure une variable conjoncturelle dans notre modèle afin de capter l'influence sur le rachat du contexte économique.

#### III.5.1 Impact des marchés financiers sur le taux de rachat

Dans son étude Changki Kim (2005), introduit comme variable explicative dans son modèle GLM le « spread » entre le taux garanti de l'assureur et la référence du marché.

Le produit sur lequel s'effectue notre étude possède un taux minimum garanti. Cependant, pour des raisons de confidentialité il n'est pas dévoilé. Ainsi, nous ne pouvons pas introduire comme variable le « spread » entre ce taux et une référence du marché.

Nous nous sommes intéressés dans un premier temps à l'influence des taux d'intérêt sur le rachat. Pour ce faire, nous avons pris comme référence les taux d'intérêt des obligations long-terme émises par l'état américain. Nous avons considéré les variations des taux d'intérêt des « T-Bonds » de maturité 10 ans.

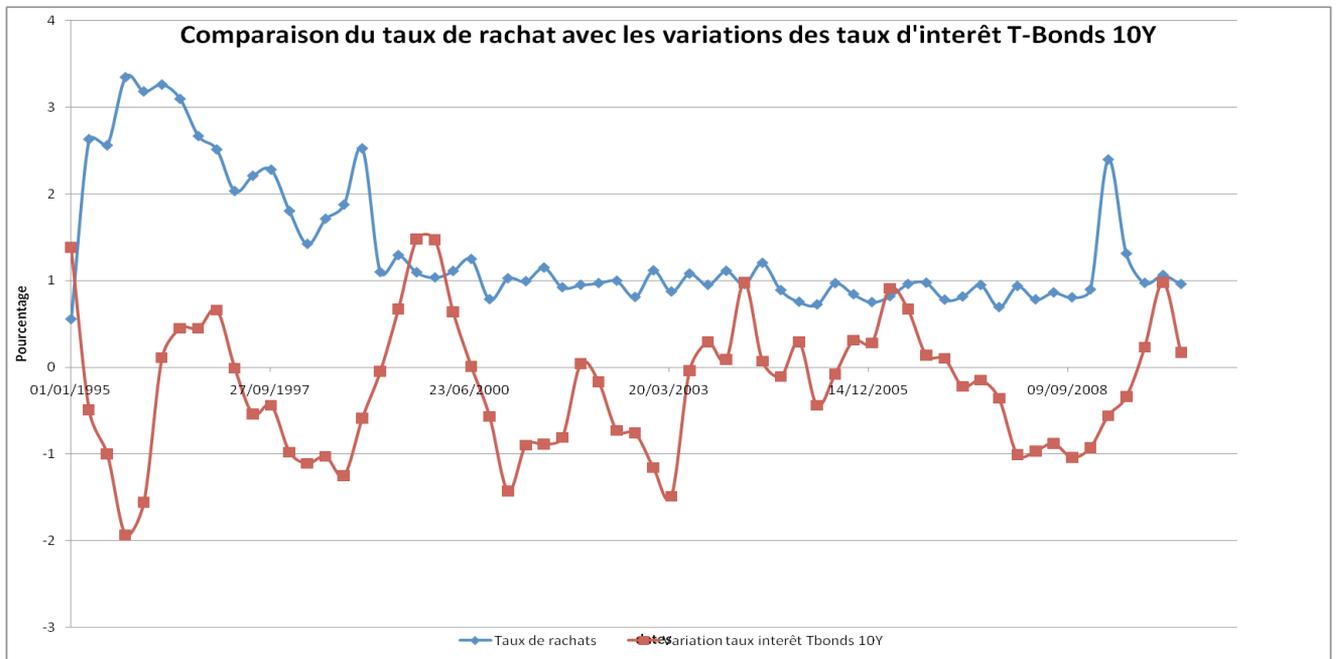


Figure 17 comparaison du taux de rachat avec les variations des taux d'intérêts des "treasury bond"

Cette comparaison graphique ne nous permet pas de détecter un éventuel lien entre l'évolution des taux d'intérêt long-terme et le taux de rachat.

Nous avons alors considéré l'introduction d'une variable conjoncturelle traduisant la survenance d'une crise financière. Nous avons comparé le taux de chômage, qui augmente fortement lors d'une crise ainsi que l'évolution d'un indice boursier américain avec le taux de rachat.

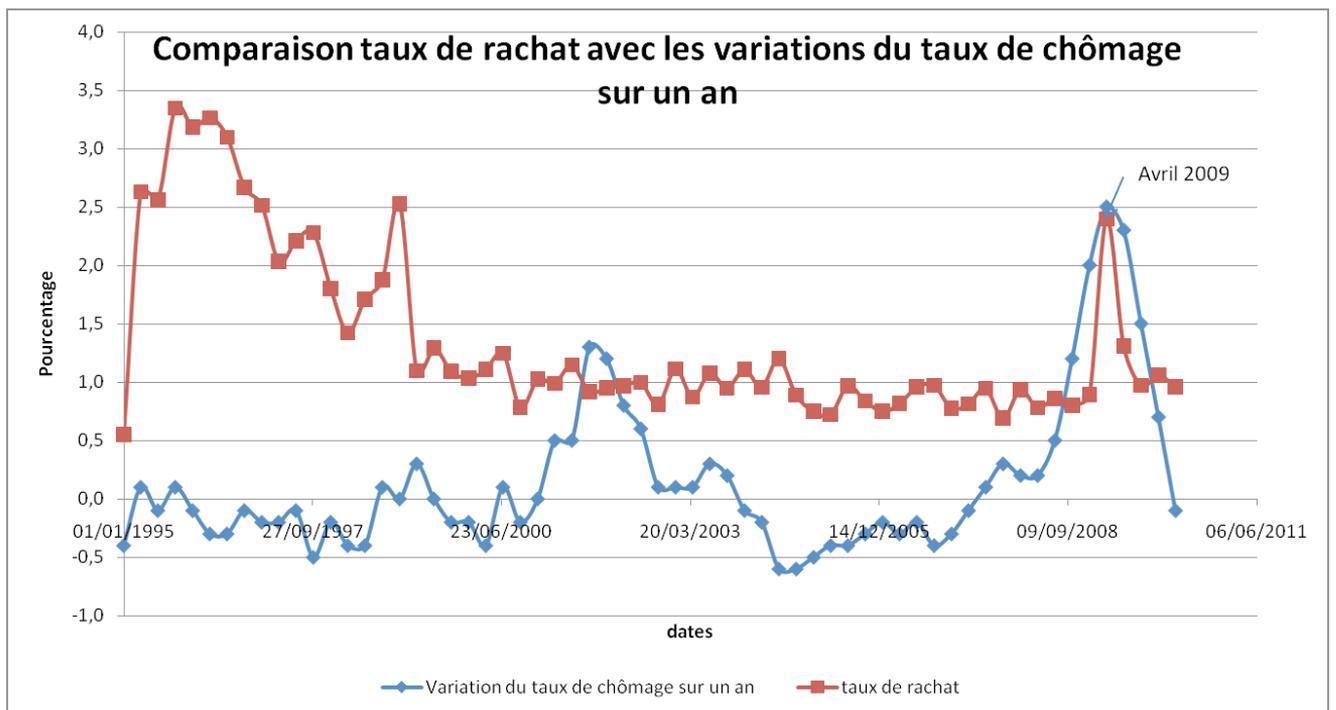


Figure 18 comparaison du taux de rachat avec le taux de chômage

Les variations du taux de chômage sont intéressantes car elles montrent que la crise financière des « subprimes » débutée en 2007, a eu un impact sur le taux de chômage lors de l'année 2009. En effet, entre

avril 2008 et avril 2009 le taux de chômage a varié de 2.5%. Ceci permet d'expliquer le taux de rachat de 2009. La crise des subprimes a débuté en 2007 cependant il y a un effet retard entre le début de la crise et le moment où elle affecte les assurés.

La variable du taux de chômage est certes satisfaisante pour capter le pic de 2009 cependant elle n'explique pas le pic de 1998. Ceci paraît normal puisque l'origine de la crise de 1998 est asiatique.

Nous avons alors étudié l'indice boursier américain du DowJones.

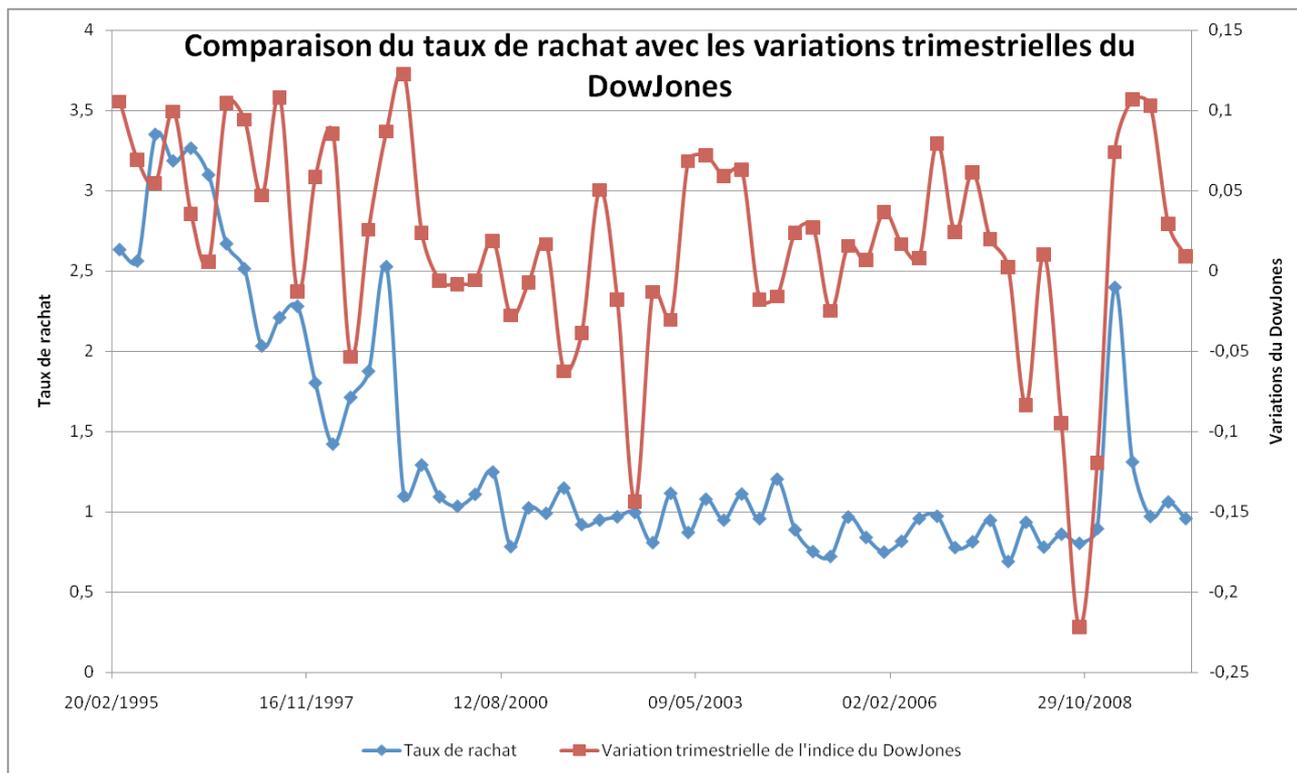


Figure 19 comparaison du taux de rachat avec l'indice DowJones

Nous observons sur le graphique que les variations positives du DowJones s'accompagnent de hausse du taux de rachat. Le coefficient de corrélation entre les variations de l'indice du DowJones et le taux de rachat est de 40%.

Il semble réaliste de prendre en compte un effet retard dans l'influence du DowJones sur le taux de rachat. Dans son étude Changki Kim (2005), établit que le temps de réaction des assurés est compris entre 3 et 6 mois. Notre pas étant trimestriel, le pic de rachat observé en avril 2009 est ainsi probablement causé par la chute des marchés financiers en octobre 2008.

L'existence d'un effet retard sur le rachat et le fait qu'une chute du DowJones soit généralement suivie d'une forte remontée de celui-ci, expliquent que le coefficient de corrélation entre les deux grandeurs étudiées soit positif.

Ainsi, il semble que les assurés aient tendance à racheter leur contrat lors d'une crise financière. Ce comportement de rachat s'explique par le fait que les assurés perçoivent des dividendes (cf. II.4.2), indexés sur la performance de la compagnie. Par conséquent, en période de crise financière, détenir un produit

offrant de potentiels dividendes perd de son attrait si les dividendes ne sont pas versés. Les assurés sont alors enclins à racheter leur contrat.

# IV. Cadre théorique de l'analyse de survie avec risques compétitifs

## IV.1 Cadre théorique

Dans cette partie nous présentons le cadre théorique dans lequel s'inscrit notre modélisation.

### IV.1.1 Formalisme

Nous notons :

$T$  la variable aléatoire recensant la durée des contrats,

$C$  la censure fixe, qui correspond à la durée entre la date d'arrêt du portefeuille (juillet 2010) et la date de souscription.

Au sein de notre base de données, nous observons la variable  $\bar{T} = \min(T, C)$  ou encore le couple  $(\bar{T}, \delta)$ , avec  $\delta = \mathbf{1}_{\{T \leq C\}}$ .

On peut donc définir le processus de comptage suivant :

$$N(t) = \sum_{i=1}^n N_i(t) \text{ avec } N_i(t) = \mathbf{1}_{\{T_i \leq t; T_i \leq C_i\}}$$

$N_i$  est le processus de comptage associé à l'assuré  $i$ .  $\{T_i \leq t; T_i \leq C_i\}$  représente la durée de contrat observable pour celui-ci. Sous l'hypothèse que la loi de la censure est indépendante de celle de la durée du contrat :

$$Y_i(t) = \mathbf{1}_{\{T_i \geq t\}} \mathbf{1}_{\{C_i \geq t\}} = \mathbf{1}_{\{\bar{T}_i \geq t\}}$$

$Y_i$  est prévisible car  $\{\bar{T}_i \geq t\}$  est le complémentaire d'un ensemble prévisible.

$$\text{On note } Y(t) = \sum_{i=1}^n Y_i(t)$$

On définit la fonction de hasard  $\lambda(t)$  de la façon suivante :

$$\lambda(t) = \frac{f(t)}{S(t)} = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta | T > t]}{\Delta}$$

Avec  $f(t)$  la densité de la variable aléatoire, et  $S(t)$  la fonction de survie ;  $S(t) = \mathbb{P}[T > t]$

La fonction de hasard représente la probabilité instantanée que le contrat soit clôturé durant l'intervalle de temps  $[t, t+\Delta]$ , sachant qu'il était encore en vigueur à la date  $t$ .

La filtration associée à notre processus est :

$$\mathcal{F}_t = \sigma(\{N_i(s), Y_i(s)\} \mid i \in \{1, \dots, n\}, s \leq t)$$

Elle correspond à l'information dont on dispose à l'instant  $t$ .

Notons  $T$  l'horizon de nos processus ( $0 \leq T \leq \infty$ ). Nous allons considérer la décomposition de Doob-Meyer sur les processus de comptage (sous-martingales positives) énoncée de la façon suivante.

Soit  $\{\mathcal{F}_t, t \leq T\}$  une filtration sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$

Soit  $N = \{N(t), t \geq 0\}$  un processus de comptage adapté à  $\mathcal{F}_t$  tel que  $\mathbb{E}[N(t)] < \infty \forall t$ .

Alors il existe un unique processus  $A(\cdot)$ ,  $\mathcal{F}_t$ -prévisible, croissant càdlàg et nul en zéro presque sûrement

vérifiant

$\mathbb{E}[A(t)] < \infty \forall t$ , tel que :

$$M(t) = N(t) - A(t), \forall t \leq T \text{ soit une martingale continue à droite.}$$

Dans le cas d'un processus de comptage censuré tel que nous l'avons défini, nous avons :

$$A_i(t) = \int_0^t Y_i(s) \lambda(s) ds$$

#### IV.1.2 Approche par risques compétitifs

Nous nous intéressons à la loi de rachat d'un contrat. Cependant, nous avons évoqué précédemment le fait que le rachat n'est pas la seule raison pour laquelle un contrat d'assurance peut être clos.

La clôture du contrat peut être causée par les raisons suivantes:

- décès de l'assuré
- rachat
- autres causes (maturité du contrat, défaut de paiement des primes etc.)

La durée des contrats au sein du portefeuille peut être schématisée par le graphe suivant :

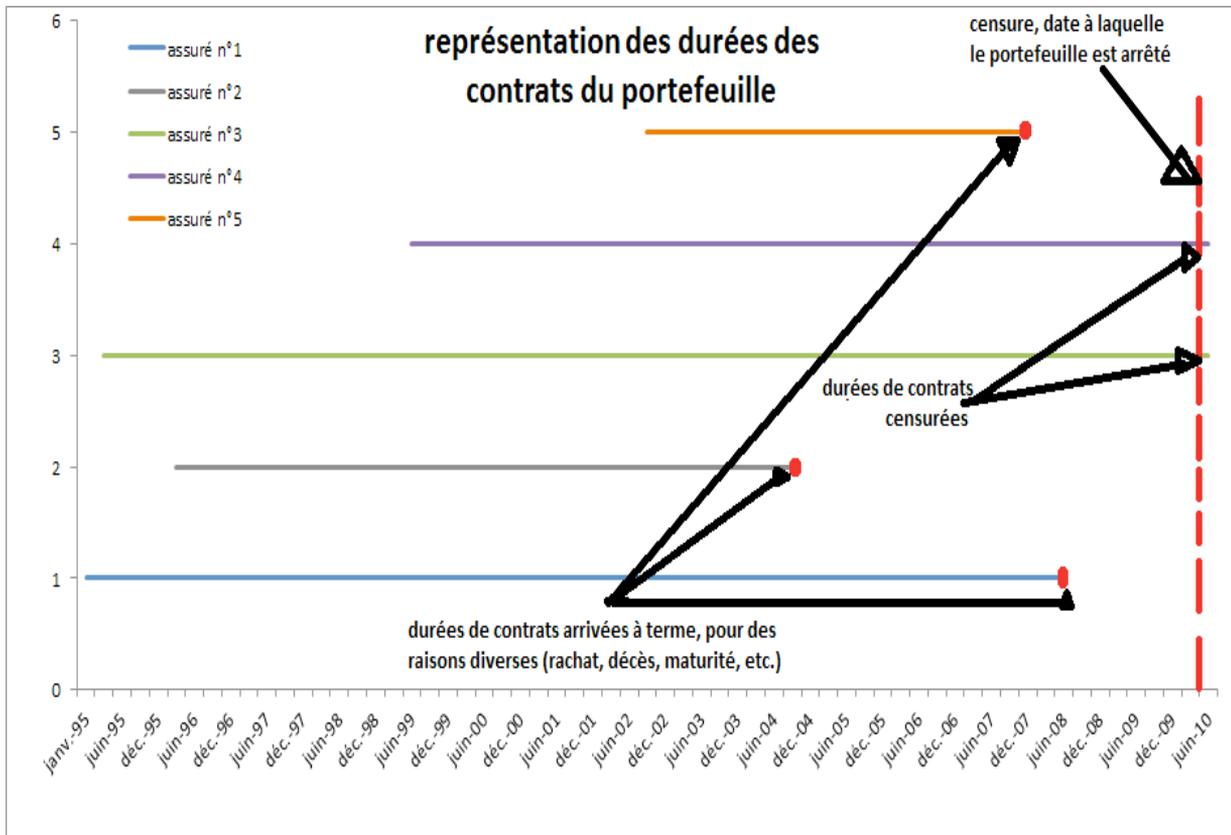


Figure 20 illustration des durées de contrats au sein d'un portefeuille

Par conséquent pour la durée du contrat des assurés, au sein du portefeuille on dispose des informations suivantes ( $\bar{T} = \min(T, C), J_T$ ) ;  $J$  étant une variable à valeurs dans  $\{0, \dots, p\}$ , spécifiant le type de risque réalisé (décès, rachat, autres causes de clôture), ou éventuellement le fait que le contrat soit en vigueur.

Lorsqu'il s'agit d'étudier l'occurrence d'un seul événement d'intérêt, les méthodes statistiques telles que les estimations Kaplan-Meier, les tests du Log-Rank et les modèles de Cox proportionnelles sont courantes. Nous avons dans un premier temps considéré la possibilité de se reporter au cadre évoqué précédemment avec un seul événement d'intérêt et une censure. Il s'agissait alors de considérer que l'événement d'intérêt est le rachat, et d'inclure les risques concurrents au rachat (le décès etc.) dans la censure. Nous avons décidé de rejeter cette approche car une donnée censurée est une donnée pour laquelle nous n'avons pas pu observer l'événement d'intérêt, mais ce dernier peut très bien se produire plus tard. Pour illustrer cela, nous pouvons prendre l'exemple d'une étude où l'on s'intéresse à la guérison d'un patient suite à un traitement. Si le patient déménage, l'information de son éventuelle guérison n'est pas observée. Cependant cette dernière peut très bien avoir lieu. Le fait de déménager durant l'étude est une censure. Au contraire, si l'individu décède d'un accident pendant l'étude, il ne peut évidemment plus guérir. Ainsi, considérer le décès accidentel comme une censure est inexact et conduit à surestimer la probabilité de guérison.

Il en est de même pour notre étude. Regrouper les autres causes de clôtures des contrats dans la loi de la censure nous conduirait à surestimer la probabilité de rachat. En effet, les risques rachat, décès et maturité sont exclusifs. Par conséquent, nous allons adopter un cadre d'étude de risques compétitifs en nous référant à l'ouvrage de Melania Pintilie « *Competing risks a practical perspective* »(2006).

#### IV.1.3 Les grandeurs utilisées dans le cadre des risques compétitifs

Pour la clarté de l'explication, on se limitera à un modèle avec un événement d'intérêt 1 et un événement compétitif 2.

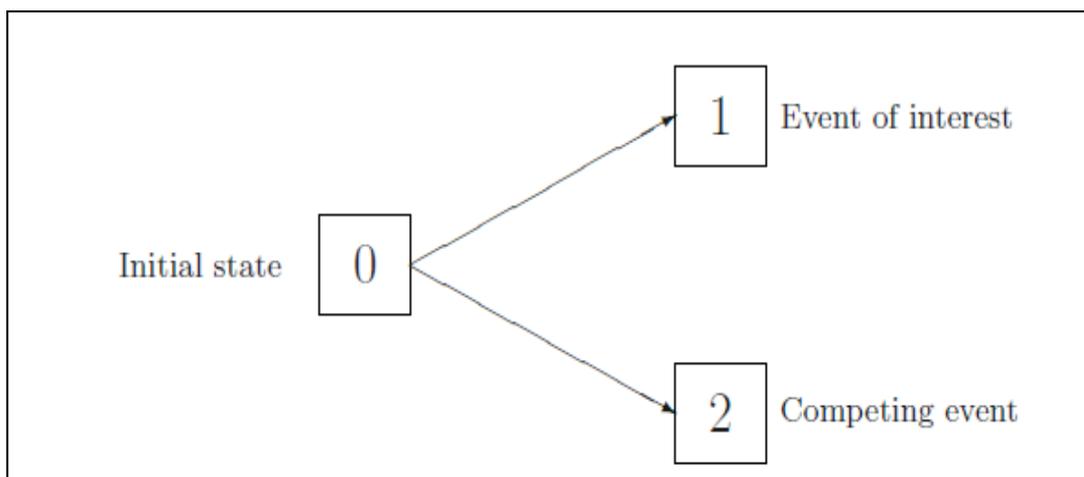


Figure 21 illustration graphique des risques compétitifs

Le processus de risques compétitifs  $(J_t)_{t>0}$  indique pour chaque date  $t$  l'état dans lequel se trouve un individu,  $J_t \in \{0,1,2\}$ . Un individu est initialement à l'état 0 et y demeure tant que, ni l'événement d'intérêt, ni le risque compétitif n'aient eu lieu.

L'individu peut :

- soit resté en 0

- soit migrer à l'état 1 si l'événement d'intérêt se produit.
- soit migrer à l'état 2 si l'événement compétitif se produit.

Les états 1 et 2 sont absorbants, au sens où une fois que l'individu y est, il ne les quitte plus. Introduisons le temps d'arrêt  $\tau$  :

$$\tau = \inf(t > 0 | J_t \neq \mathbf{0}) ; \tau \text{ est le temps d'occurrence d'un évènement.}$$

L'observation du processus  $(J_t)_{t>0}$  peut-être soumise à une censure à droite et/ou une troncature à gauche. Si l'observation du processus est soumise uniquement à une censure à droite  $C$ , on observe le couple  $(\bar{T} = \min(T, C), (J_t)_{t>0})$ . Dans le cas d'un évènement censuré, l'observation est stoppée quand l'individu est encore dans l'état initial 0 (ce qui correspond aux contrats en vigueur dans notre portefeuille). Dans le cas contraire, le temps d'occurrence d'un évènement et le type d'évènement ont été observés.

Dans la rubrique qui suit, nous allons ainsi nous intéresser à des grandeurs clés utilisées dans le cadre des risques compétitifs.

#### IV.1.3.1 La fonction de hasard spécifique

La fonction de hasard spécifique représente la probabilité instantanée que le contrat soit clôturé suite à la cause  $k$ , durant l'intervalle de temps  $[t, t+\Delta]$  sachant qu'il était encore en vigueur à la date  $t$ . Elle est donnée par la formule suivante :

$$\lambda_k(t) = \frac{f(t)}{S(t)} = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, J = k | T > t]}{\Delta}$$

Avec ici  $J$  la variable aléatoire représentant les différentes causes de clôture d'un contrat (décès, rachat et autres causes dans notre cas). On définit de plus la fonction de hasard spécifique cumulée qui correspond à un risque cumulé par :

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds$$

L'ensemble des fonctions de hasard spécifiques détermine totalement le comportement stochastique du processus car :

$$\lambda(t) = \sum_{k=1}^n \lambda_k(t) \quad \text{et} \quad S(t) = \mathbb{P}[T > t] = e^{-\int_0^t \sum_{k=1}^n \lambda_k(s) ds}$$

avec  $n$  le nombre de risques compétitifs.

#### IV.1.3.2 La fonction d'incidence cumulée (CIF)

Dans le cadre des risques compétitifs, l'étude de la fonction d'incidence cumulée est également pertinente, car cette grandeur représente la proportion attendue d'individus subissant un certain évènement compétitif au cours du temps. Elle est donnée par :

$$F_k(t) = \mathbb{P}[T < t, J = k] = \int_0^t f_k(s) ds \text{ et}$$

$$f_k(t) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, J = k]}{\Delta} = \lambda_k(t) \times S(t) \text{ d'où } F_k(t) = \int_0^t \lambda_k(s) \times S(s^-) ds$$

Ainsi la probabilité de subir un événement  $k$  avant la date  $t$  correspond à une « somme » sur chaque intervalle infinitésimal  $[s, s + ds]$  de la probabilité d'être encore à l'état 0 (dans notre cas correspond à la probabilité que le contrat soit en vigueur) juste avant  $s$ , multipliée par la probabilité instantanée de subir l'évènement  $k$  sur  $[s, s + ds]$ .

#### IV.1.4 Modèle de régression dans le cadre de risques compétitifs

Pour effectuer une analyse de survie dans le cas de risques compétitifs, il existe essentiellement deux approches :

- la modélisation des fonctions de hasard spécifiques.
- la modélisation des fonctions de hasard de la sous-distribution (notion qui sera expliquée dans la partie consacrée à cette approche).

##### IV.1.4.1 Modélisation des causes de hasard spécifiques

Comme nous avons vu précédemment, pour un risque donné  $k \in J$ , la fonction de hasard spécifique est donnée par :

$$\lambda_k(t) = \frac{f(t)}{S(t)} = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, J = k | T > t]}{\Delta}$$

Cette approche consiste à modéliser l'évènement d'intérêt comme si les risques compétitifs n'existaient pas. Elle est à privilégier si l'on s'intéresse à l'étude pure d'un processus (exemple : cas de traitement médical où on ignore les autres causes de décès). Elle permet en effet de connaître l'effet de covariables sur le processus. Cependant, pour prendre en compte les autres risques compétitifs, et pouvoir déterminer une proportion d'individus sur un portefeuille subissant le risque d'intérêt, il est nécessaire de modéliser toutes les fonctions de hasard spécifiques afin d'obtenir  $\lambda(t)$  et  $S(t)$  :

$$\lambda(t) = \sum_{k=1}^n \lambda_k(t) \quad \text{et} \quad S(t) = \mathbb{P}[T > t] = e^{-\int_0^t \sum_{k=1}^n \lambda_k(s) ds}$$

Nous avons décidé de rejeter cette approche car elle impliquerait de modéliser tous les processus des risques compétitifs (décès et autres causes).

##### IV.1.4.2 Modélisation du hasard de la sous distribution

La seconde approche consiste à étudier un processus dérivé du processus initial, la sous-distribution, obtenue en stoppant judicieusement le processus initial. Cette approche a été développée par Jan

Beyersmann, Arthur Allignol et Martin Schumacher dans leur ouvrage « Competing Risks and Multistate Model with R ».

Soit  $(J_t)_{t>0}$  le processus initial des risques compétitifs et  $\tau$  : le temps d'occurrence d'un évènement ( $\tau = \inf(t > 0 | J_t \neq 0)$ ).

Le temps  $\tau$  est le temps au bout duquel  $(J_t)_{t>0}$  quitte l'état initial (i.e. jusqu'à l'occurrence d'un évènement qui peut être 1 ou 2 dans le cas d'un modèle à deux risques compétitifs). L'idée de l'approche par la sous-distribution est de considérer la variable aléatoire  $\vartheta$ , et non plus  $\tau$ , qui représente la durée avant que l'évènement d'intérêt 1 ne se produise. L'objectif étant de prendre en compte le fait que le risque compétitif 2 puisse se produire en premier. Il s'agit alors d'arrêter le processus  $(J_t)_{t>0}$  juste avant le temps  $\tau$  (i.e. en  $\tau^-$ ) si le processus a bougé dans l'état du risque compétitif 2. Ainsi on introduit la variable suivante :

$$\xi_t = \mathbf{1}_{\{J_t \neq 2\}} \times J_t + \mathbf{1}_{\{J_t = 2\}} \times J_{\tau^-}$$

où  $J_t \in \{0, 1, 2\}$

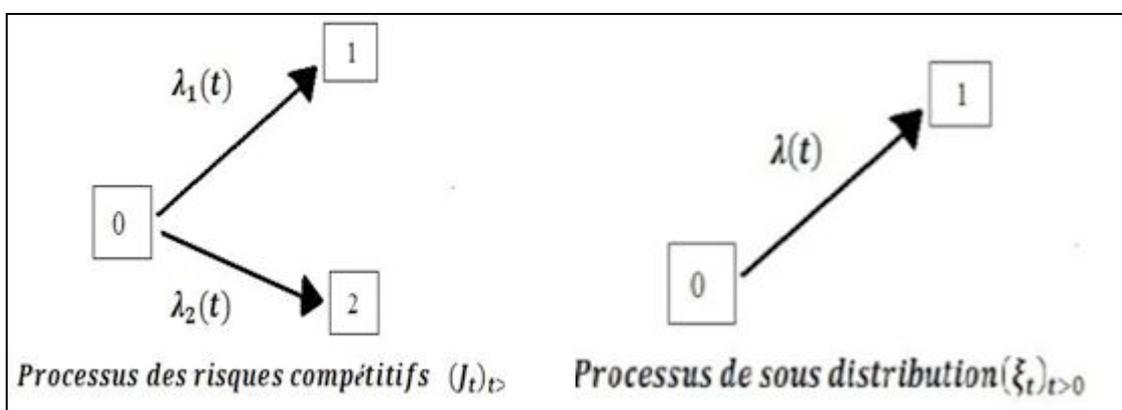


Figure 22 illustration graphique du concept de la sous distribution

Ainsi,  $(J_t)_{t>0}$  et  $(\xi_t)_{t>0}$  sont identiques tant que le processus  $(J_t)_{t>0}$  n'a pas sauté dans l'état de l'évènement compétitif 2.

Si  $(J_t)_{t>0}$  saute dans l'état 2, l'individu reste coincé dans l'état initial du processus de la sous distribution  $\xi_t = 0 \forall t \geq \tau$ .

Ainsi, l'interprétation de l'état initial évolue :

$\{J_t = 0\}$  signifie qu'il n'y a pas évènement à la date  $t$ .

$\{\xi_t = 0\}$  signifie qu'il n'y a pas d'évènement de type 1 à la date  $t$ .

Comme énoncé précédemment, nous nous intéressons donc à la variable  $\vartheta$  définie de la sorte :

$$\vartheta = \inf(t > 0 | \xi_t \neq 0)$$

$$\vartheta = \begin{cases} \tau & \text{si } J_\tau = 1 \\ \infty & \text{si } J_\tau = 2 \end{cases}$$

Par conséquent, la fonction de répartition associée à la variable  $\vartheta$  coïncide avec la fonction d'incidence cumulée de l'évènement 1 :

$$\mathbb{P}[\vartheta \leq t] = \mathbb{P}[\tau \leq t, J_\tau = 1] \quad \forall t \in [0, \infty)$$

La fonction de hasard de la sous-distribution du risque d'intérêt est la suivante :

$$\lambda_1(t)dt = \mathbb{P}[\vartheta \in [t, t + dt] | \vartheta \geq t]$$

ou de manière équivalente :  $\lambda_1(t) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, J_t = 1 | \{T > t\} \cup \{T \leq t, J_t \neq 1\}]}{\Delta}$

L'approche par sous-distribution a été introduite par Jason P. Fine et Robert J. Gray dans « A proportional hazard model for the subdistribution of a competing risk » (1999). Dans leur étude, Fine and Gray qualifient la variable aléatoire  $\vartheta$  rattachée à la sous distribution comme étant impropre car elle peut valoir l'infini.

L'ensemble à risque rattaché au hasard  $\lambda(t)$  est défini à tout instant  $t$  par :

$$\{T > t\} \cup \{T \leq t, J_t \neq 1\}.$$

L'ensemble à risque à une date  $t$  est donc constitué :

- des individus qui n'ont pas encore quitté l'état initial 0 à l'instant  $t$
- mais également des individus qui ont quitté l'état initial 0 avant l'instant  $t$  suite à un événement compétitif.

Cet ensemble à risque n'est pas naturel car il comporte des individus qui ont quitté l'état 0 pour un état compétitif et ne sont donc plus susceptibles de subir l'évènement 1. Il est difficile de conceptualiser un ensemble à risque où certains individus ont une probabilité nulle de vérifier  $\{J_\tau = 1\}$ .

Par conséquent, l'ensemble à risque est généralement inconnu. Il est cependant possible de lever cette difficulté dans le cas de données censurées à droite. Les solutions développées nécessitent une modélisation explicite de la distribution de la censure. En pratique, une censure aléatoire est fréquemment supposée.

➤ **Cas d'un individu  $i$  qui a subi l'évènement compétitif 2 à une date  $T_i$  et qui est soumis à une censure à droite aléatoire  $C_i$  :**



Pour ces individus, on connaît  $\tau_i$  et  $J_{\tau_i} = 2$ . Par conséquent  $\vartheta_i = \infty$ . Cependant dans le cas de données censurées, on a accès à la variable  $\vartheta_i \wedge C_i = C_i$  avec  $C_i$  la censure pour l'individu  $i$ . Ainsi, la censure permet de lever le problème de la variable  $\vartheta_i$  impropre.

Cependant, très souvent en pratique on ne connaît pas la censure  $C_i$ . L'observation peut avoir été arrêtée à la date d'occurrence du risque compétitif en  $\tau_i$  (par exemple si l'événement 2 est le décès : l'individu cesse d'exister). Même si l'événement compétitif 2 n'est pas fatal, la collecte de données peut être programmée pour s'achever en  $\tau_i$ . Ces difficultés d'ordre pratique causées par les censures aléatoires ont été résolues et implémentées sous R, voir l'article de Ruan & Fine « *Multiple imputation* ». La troncature à gauche, quant à elle, fait encore l'objet de recherches.

Dans le cadre de notre étude, nous sommes face à une censure fixe purement administrative (arrêt du portefeuille). Ceci nous simplifie la tâche car cette censure fixe nous permet de connaître à tout instant l'ensemble des individus à risque.

### ➤ Cas de la censure administrative

Le processus sous risque associé à la sous-distribution est :

$$Y_i(t) = \mathbf{1}(t \leq T_i \wedge C_i) + \mathbf{1}(T_i < t \leq C_i, J_{T_i} = 2)$$

Comme nous l'avons mentionné précédemment, le second terme de l'expression ci-dessus pose généralement problème car il est inconnu. La censure administrative représente une exception où le second ensemble est connu. Etant donné que la censure est uniquement due à une fin administrative d'étude, la durée potentielle de censure pour tout individu est connue en avance. Ainsi, les difficultés évoquées précédemment sur la définition de l'ensemble à risque ne se posent pas. Nous pouvons alors utiliser les modules statistiques standards implémentant le modèle de Cox à condition de définir correctement l'ensemble à risque.

Bilan sur les deux approches :

L'approche à privilégier, modélisation par hasard spécifique ou par hasard de la sous-distribution, dépend de la finalité de l'étude réalisée. En effet, pour étudier l'impact de covariables sur un processus, l'approche par cause-spécifique est satisfaisante. C'est souvent l'approche retenue dans les études médicales où l'objectif est de déterminer l'effet d'un médicament sur la durée de vie du patient. Cependant, une telle approche ne permet pas de prédire une proportion d'individu subissant l'événement d'intérêt. Or l'objectif de notre mémoire est de reconstituer le taux de rachat sur un portefeuille. Par conséquent, nous décidons d'adopter l'approche dite de Fine&Gray consistant à étudier la sous-distribution.

## IV.2 Études empiriques

Dans cette rubrique nous allons estimer de façon non paramétrique les fonctions de hasard des différents risques de notre portefeuille, puis nous réaliserons des tests graphiques qui nous permettront de savoir si un modèle à hasards proportionnels est pertinent.

### IV.2.1 Estimations empiriques

Afin de modéliser la loi de rachat de notre portefeuille, nous nous intéresserons à la modélisation de la fonction de hasard de la sous-distribution. Comme nous l'avons énoncé précédemment, étant dans un modèle à risques compétitifs, nous allons nous intéresser à :

$$\lambda_1(t) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta, J_t = 1 \mid \{T > t\} \cup \{T \leq t, J_t \neq 1\}]}{\Delta}, \text{ la cause 1 étant le rachat.}$$

Dans le cadre général, un estimateur de la fonction de hasard est donné par l'estimateur de Nelson-Aalen. Celui-ci est donné par la formule suivante :

$$\widehat{\lambda}(t) = \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{\{T_i \leq t\}}}{\sum_{j=1}^n \mathbf{1}_{\{T_j \geq T_i\}}}$$

avec  $\bar{T} = \min(T, C)$   $\delta = \mathbf{1}_{\{T \leq C\}}$ .

Notons que cet estimateur est beaucoup plus robuste sur les premières années de l'estimation car plus on avance dans le temps, moins on a d'observations (contrats clôturés). Ainsi il faut interpréter avec prudence l'estimation obtenue pour des grandes durées (l'estimation du hasard pour les durées supérieures à 12 ans est effectuée sur 80 contrats pour un portefeuille d'environ 30 000 contrats à l'origine). Concernant l'estimateur de la fonction de hasard de la sous-distribution  $\widehat{\lambda}_1(t)$ , celui-ci s'effectue de la même façon mais en modifiant l'ensemble à risque (terme au dénominateur) comme expliqué en partie VI.1.4.2.

Ci-dessous l'estimation de la fonction de hasard de la sous distribution pour les différents risques du portefeuille :

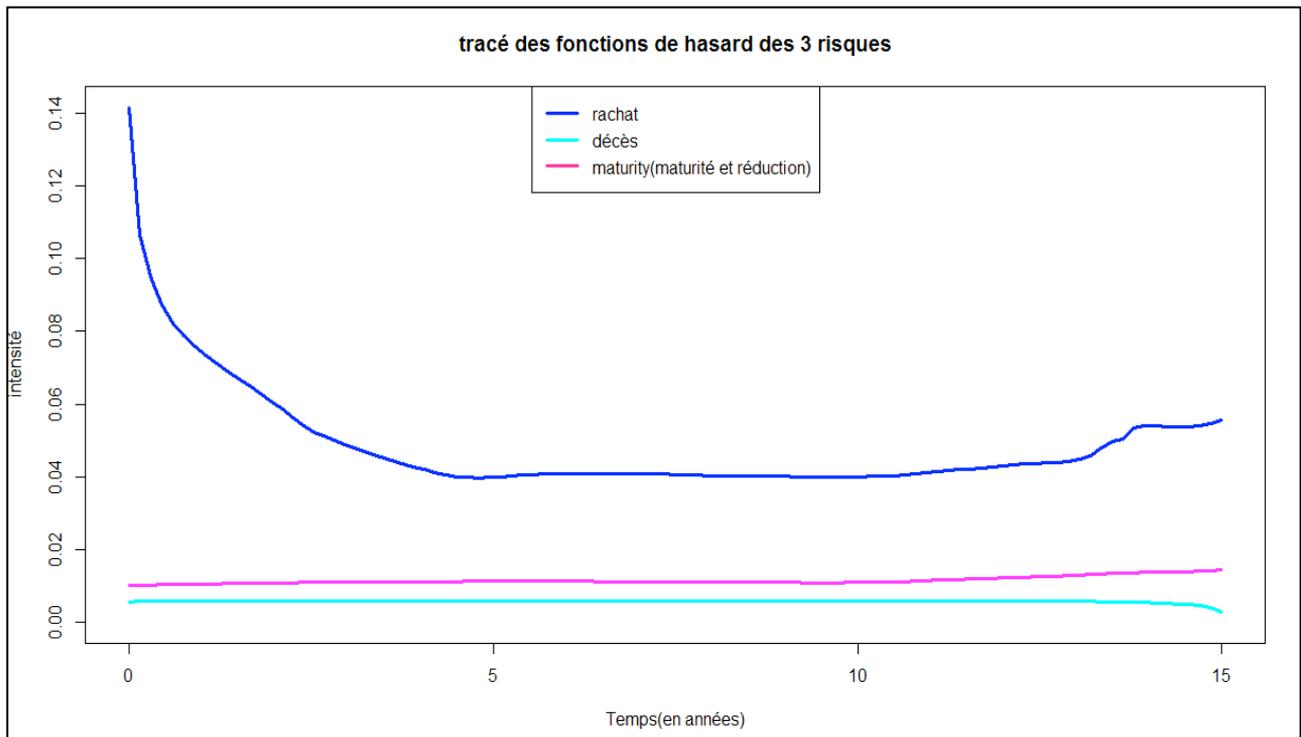


Figure 23 tracé des fonctions de hasard empiriques

Nous remarquons que pour les risques décès et les autres causes de clôture regroupées dans la modalité « OtherLapses », la fonction de hasard est quasi constante. Ceci s'apparente à des fonctions hasard de variables aléatoires suivant des lois exponentielles.

Quant à la fonction de hasard du rachat, celle-ci est très forte au début, elle a une pente très raide durant la 1ere année, puis elle décroît progressivement jusqu'à la 5e année. Entre la 5e et la 10e année elle est approximativement constante.

Nous pouvons tracer les fonctions de hasard empiriques par covariable afin de visualiser leur effet. Ci-dessous, la variable « acc.death.amount », qui correspond à la présence ou non d'une option de capital supplémentaire en cas de décès, sert à stratifier le portefeuille. Cela signifie qu'elle permet de constituer deux sous-populations (chaque modalité de la variable définit une sous-population) sur lesquelles l'estimation de la fonction de hasard est effectuée.



Figure 24 exemple de tracé de fonction de hasard pour l'option Rider

Nous présentons dans l'annexe B les autres tracés de la fonction de hasard rachat stratifiée par covariable.

#### IV.2.2 Test de proportionnalité

Dans les modèles de survie, la classe de modèles à hasards proportionnels semi-paramétrique est très répandue. Nous allons tester de façon empirique si un tel modèle serait valide sur nos données.

Un modèle à hasard proportionnel consiste à supposer que la fonction de hasard de tout individu s'écrit sous la forme :  $h_{\theta}(t) = \theta h(t)$  avec  $\theta$  un paramètre positif inconnu. Le modèle de Cox appartient à cette famille et permet d'intégrer des variables explicatives pour définir le paramètre  $\theta$  :  $\theta = \exp(x^t\beta)$  avec  $x$  : vecteur de  $p$  variables explicatives, et  $\beta$  vecteur de  $p$  paramètres.

Dans le cadre d'un modèle de Cox, les fonctions de hasard de deux individus  $i$  et  $j$  vérifient :

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp(X_i\beta)}{\lambda_0(t)\exp(X_j\beta)} = \frac{\exp(X_i\beta)}{\exp(X_j\beta)}$$

Pour des covariables constantes dans le temps, avec peu de modalités, l'hypothèse de proportionnalité peut être vérifiée graphiquement en traçant  $\text{Log}(\text{Log}(-S_{KM}(t)))$  en fonction de la durée  $t$ , où  $S_{KM}$  est l'estimateur de Kaplan-Meier.

Ceci se démontre de la façon suivante :

Sous l'hypothèse de hasards proportionnels, nous avons :

$$\lambda_i(t) = \lambda_0(t)\exp(X_i\beta) \text{ d'où } \Lambda_i(t) = \int_0^t \lambda_0(u)\exp(X_i\beta) du = \Lambda_0(t)\exp(X_i\beta)$$

$$\text{Or } S(t) = \exp(-\Lambda(t)) \text{ Ainsi } \text{Log}(-\text{Log}(S(t))) = \text{Log}(\Lambda_0(t)) - X_i\beta$$

Par conséquent, si le modèle de Cox est correct, les courbes de l'estimateur de Kaplan-Meier à l'échelle log-log, obtenues pour les différentes modalités d'une même covariable doivent être approximativement parallèles.

Nous pouvons réaliser une telle vérification graphique pour toutes nos covariables fixes. Concernant la variable externe variant dans le temps, une telle vérification graphique n'est pas possible. Cependant, dans le V.4, nous pourrions procéder à une vérification par test statistique.

- **Test de l'hypothèse de proportionnalité pour la variable Gender :**

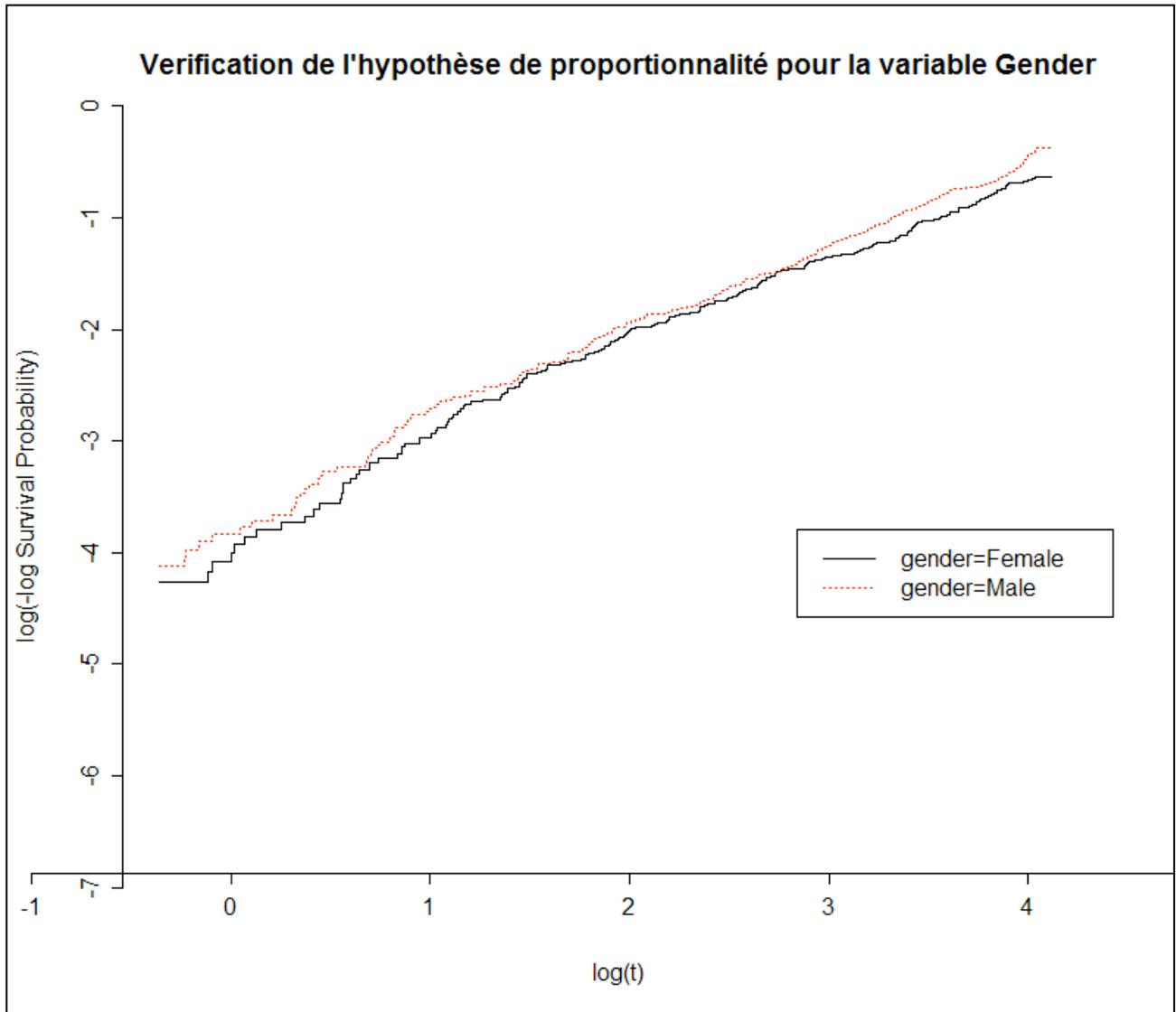


Figure 25 vérification graphique de l'hypothèse de proportionnalité pour la variable genre

Le défaut majeur de toute procédure de vérification graphique est sa subjectivité. En effet, sur quels critères décider du parallélisme des courbes. Ce problème a été soulevé par David G.Kleinbaum et Mitchel Klein (2005) qui recommandent une approche conservatrice consistant à supposer l'hypothèse de proportionnalité tant qu'il n'y a pas de fortes preuves de non-parallélisme des courbes log-log. Une fois un modèle calibré, nous pourrions effectuer des tests statistiques plus rigoureux afin de vérifier la proportionnalité des hasards. Nous considérons l'hypothèse de proportionnalité vérifiée pour la variable « gender ».

- Test de l'hypothèse de proportionnalité pour la variable « acc.death.amount » :

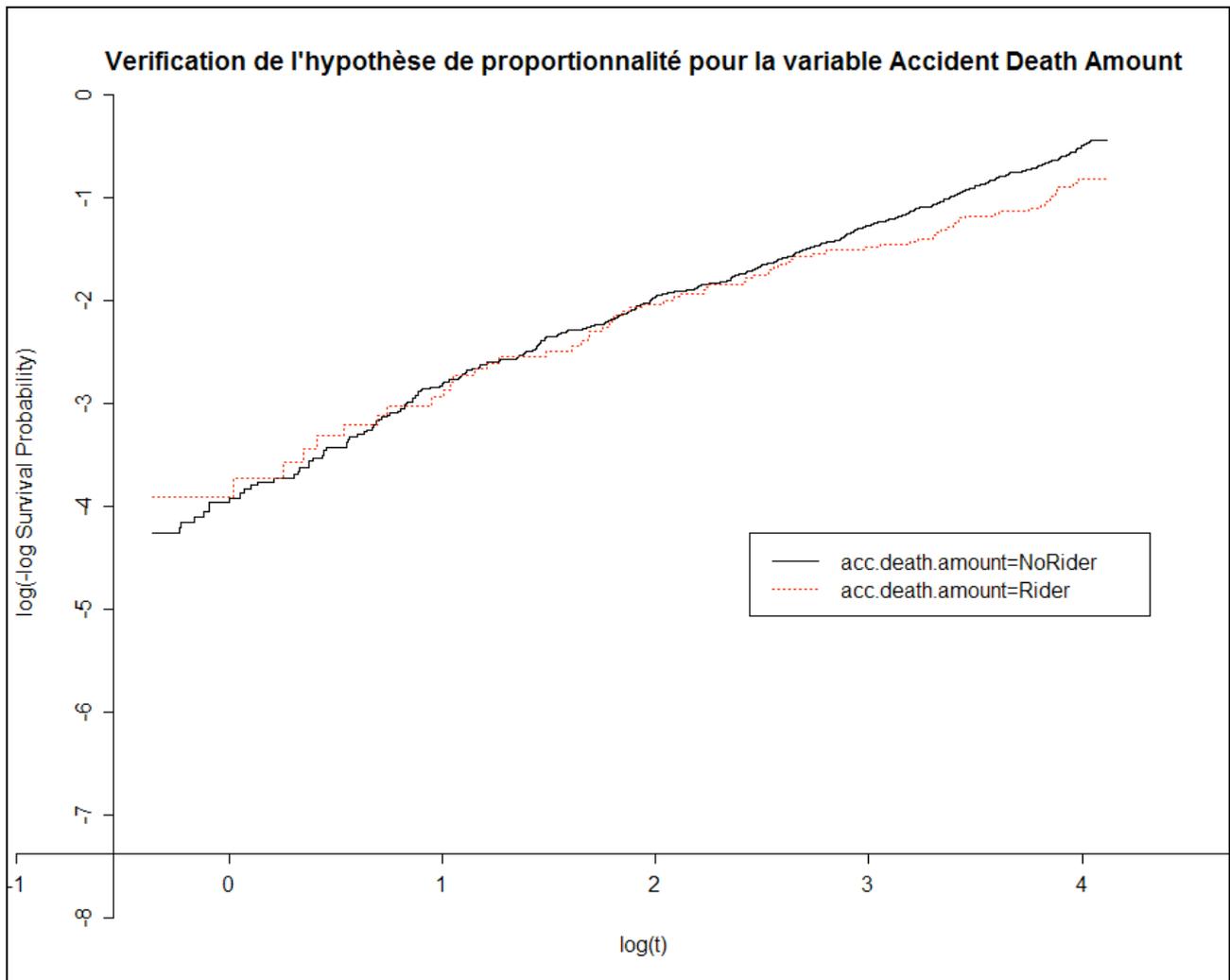


Figure 26 vérification graphique de l'hypothèse de proportionnalité pour la variable "acc.death.amount"

Les courbes suggèrent un non-parallélisme pour les faibles durées de vie des contrats. Cependant, au-delà de 20 trimestres, les courbes sont parallèles. Nous pouvons considérer que l'hypothèse est vérifiée.

- Test de l'hypothèse de proportionnalité pour la variable « risk.state » :

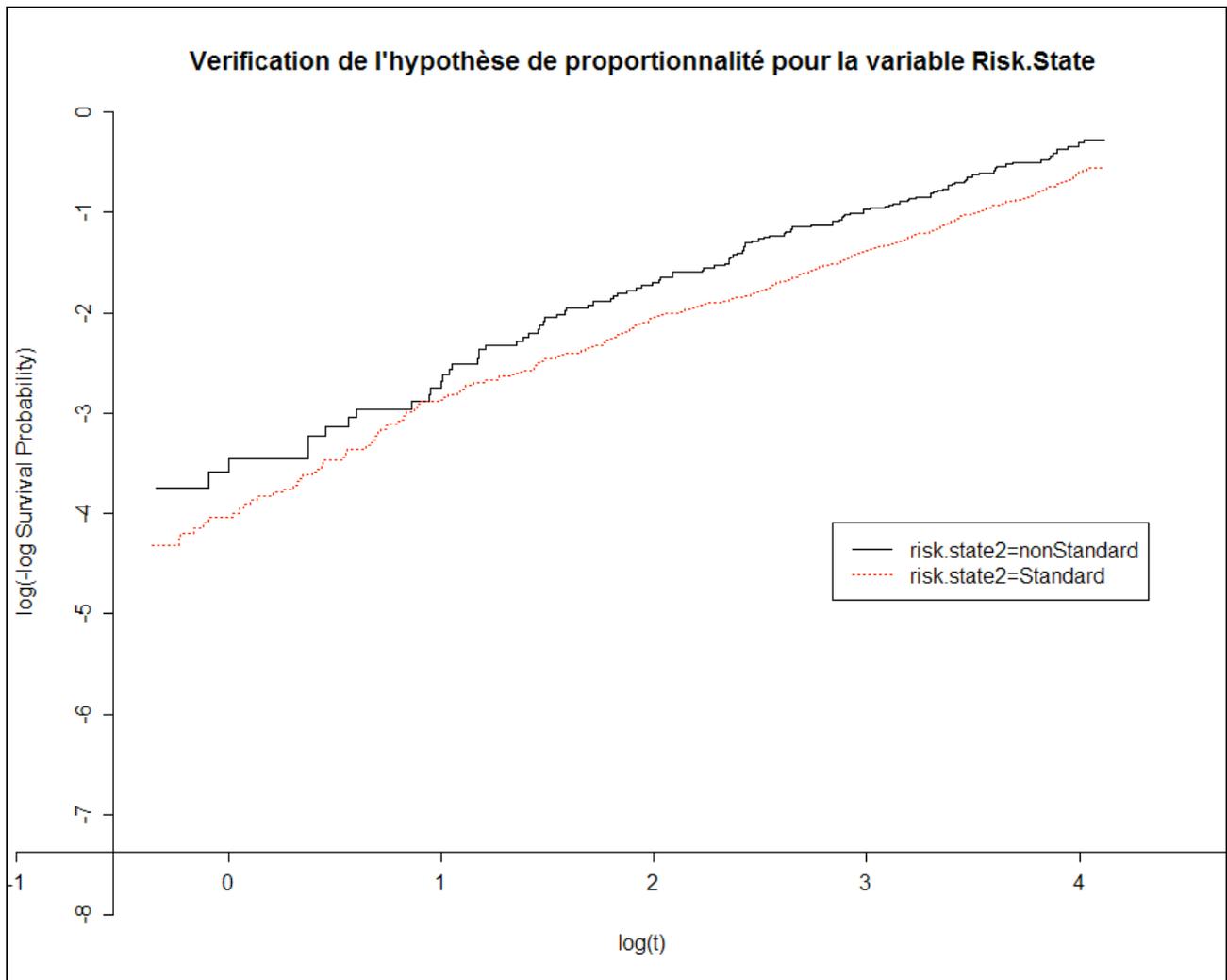


Figure 27 vérification graphique de l'hypothèse de proportionnalité pour la variable "risk.state"

De même, l'hypothèse de proportionnalité pour la variable « risk.state » est vérifiée.

- Test de l'hypothèse de proportionnalité pour la variable « premium.frequency » :

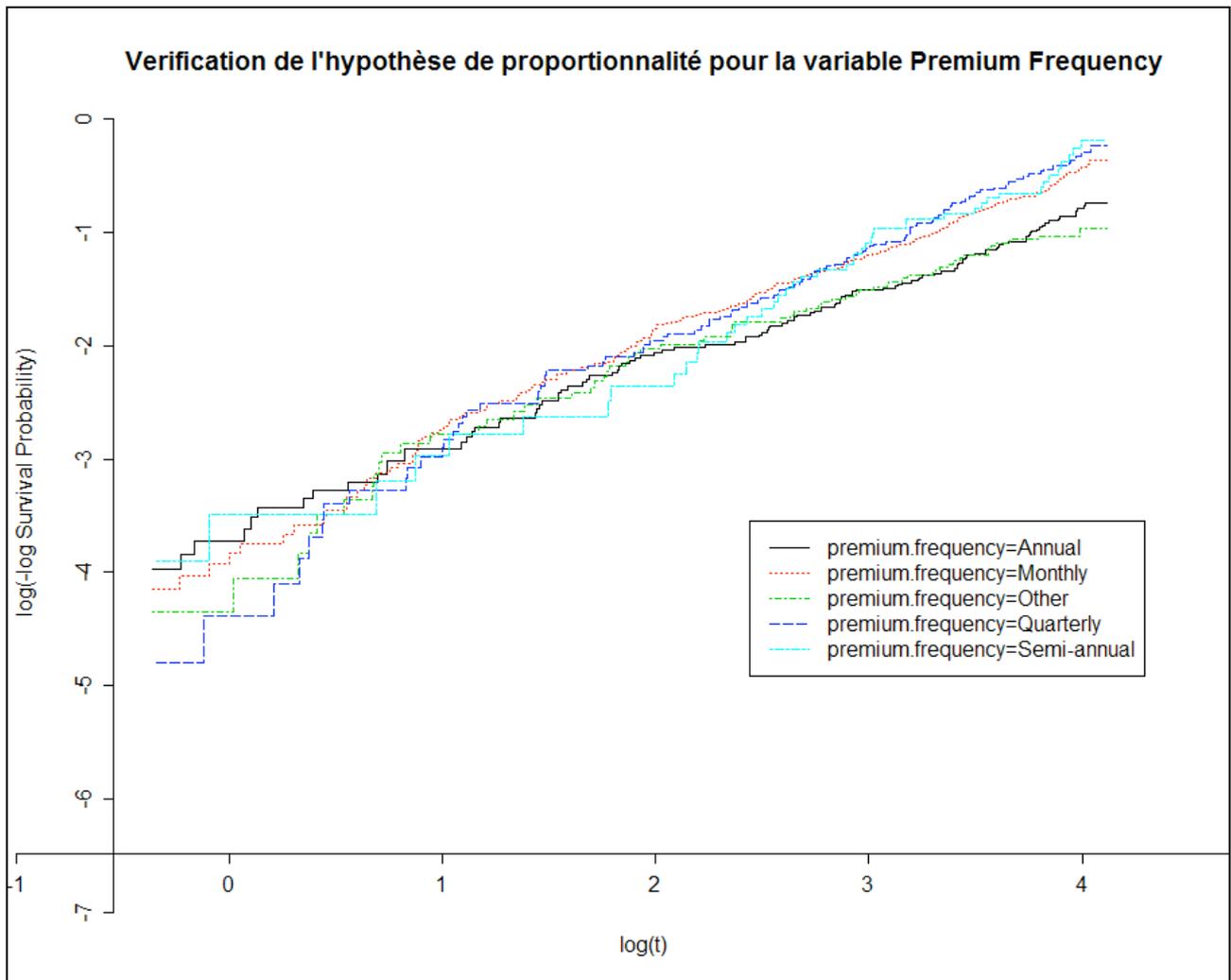


Figure 28 vérification graphique de l'hypothèse de proportionnalité pour la variable "premium.frequency"

L'hypothèse de proportionnalité semble contestable pour les modalités semestrielle, trimestrielle et annuelle. Elle nécessite d'être confirmée par un test statistique plus rigoureux que nous effectuerons dans le V.4.

- Test de l'hypothèse de proportionnalité pour la variable « premium » :

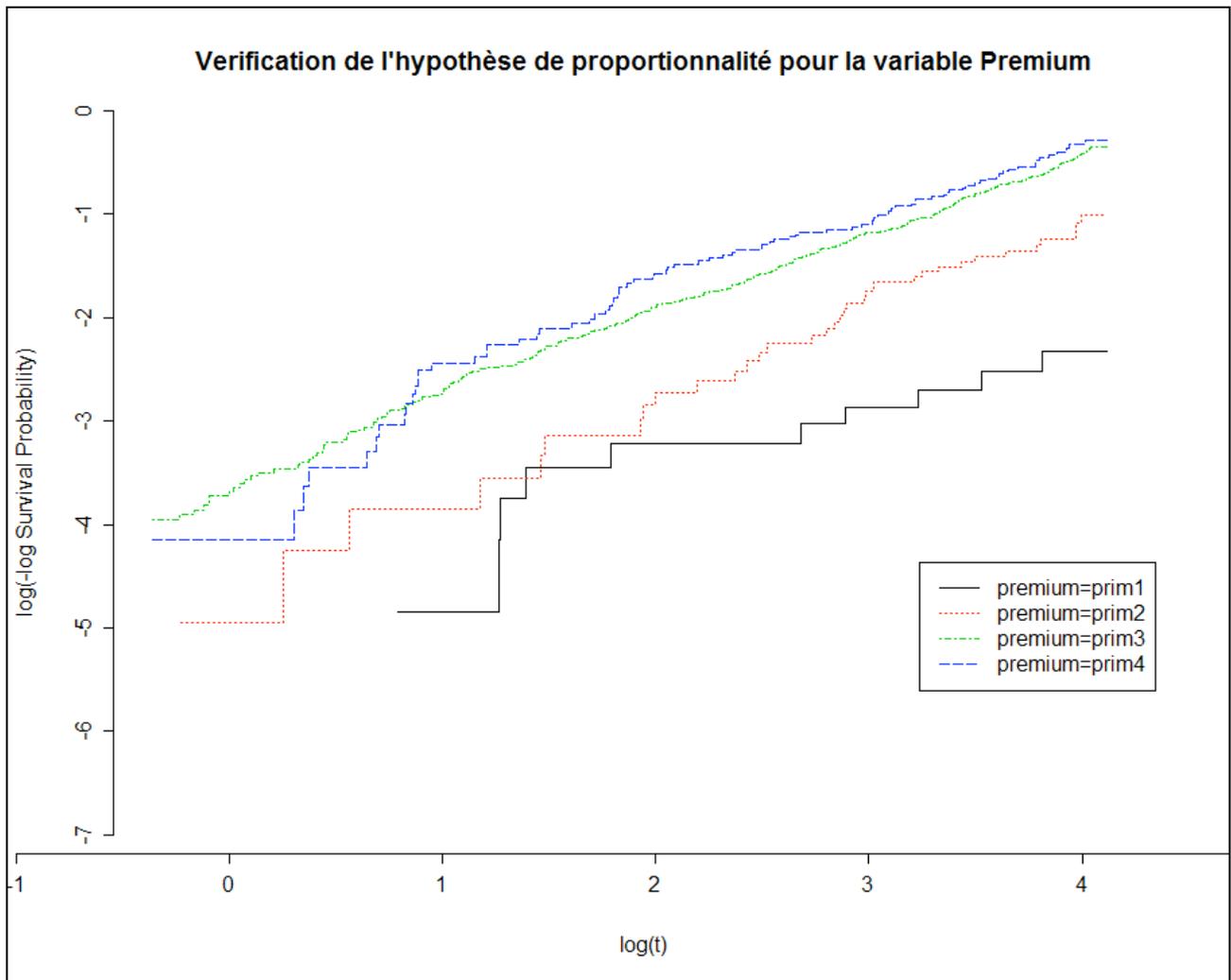


Figure 29 vérification graphique de l'hypothèse de proportionnalité pour la variable classe de prime

L'hypothèse semble vérifiée pour les modalités prim1, prim2 et prim3. La prim4 présente une indication de non-parallélisme avec la prim3 pour les faibles durées (i.e. durées inférieures à sept trimestres). Nous acceptons tout de même l'hypothèse de proportionnalité.

- Test de l'hypothèse de proportionnalité pour la variable « underwriting.age » :

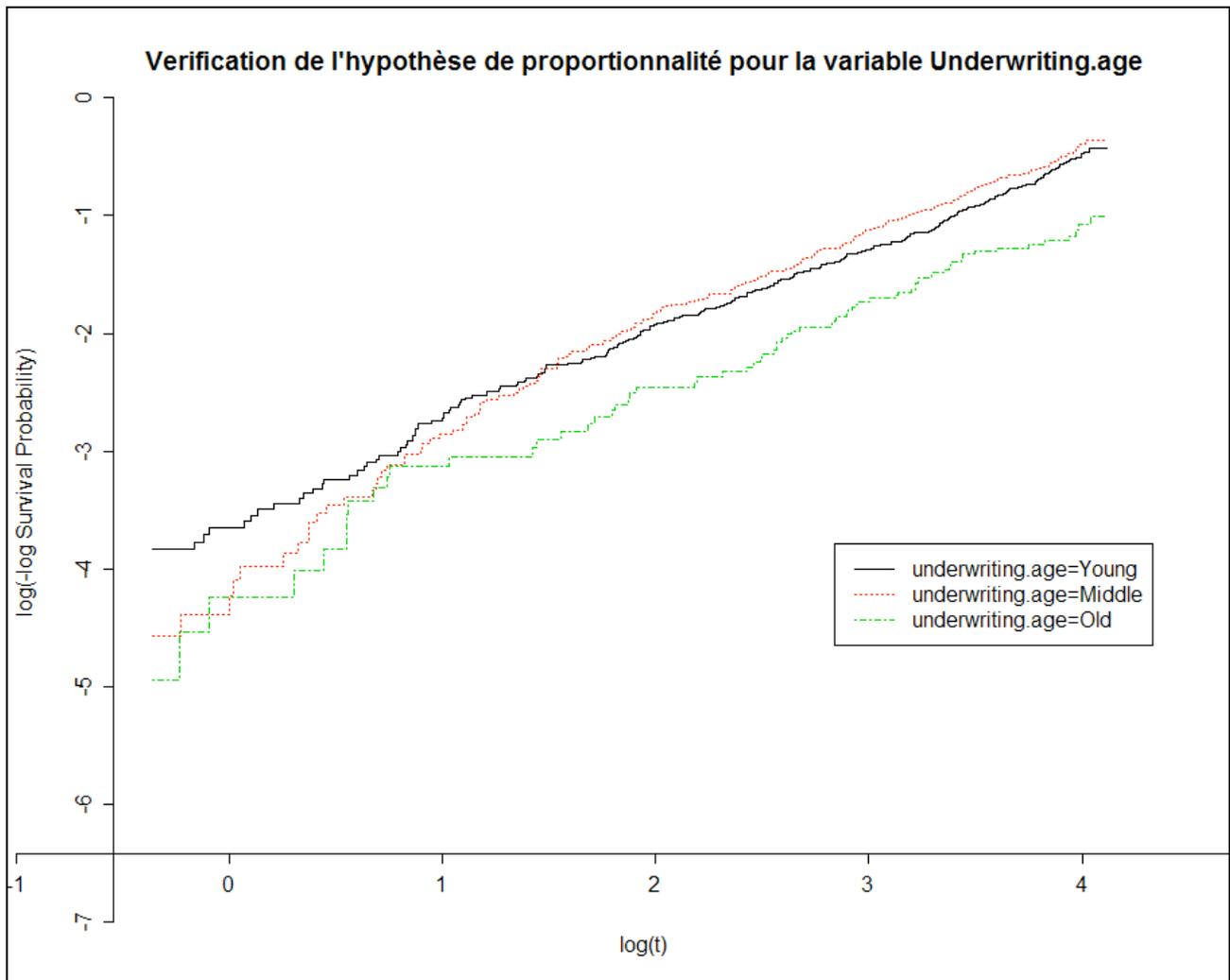


Figure 30 vérification graphique de l'hypothèse de proportionnalité pour la variable "underwriting.age"

De même que précédemment, à partir de cinq trimestres l'hypothèse semble valide.

Conclusion :

D'après les tests graphiques effectués dans le chapitre 4, l'hypothèse de proportionnalité semble vérifiée pour les covariables présentes dans notre portefeuille. Par conséquent, nous allons utiliser un modèle de Cox à hasards proportionnels afin de modéliser le taux de rachat.

# V. Modèle de Cox dans un cadre Fine & Gray avec une covariable dépendant du temps

## V.1 Introduction d'une variable dépendant du temps dans le modèle de Cox

Dans un modèle à hasards proportionnels de Cox, la fonction de hasard de l'individu  $i$  vérifie :

$$\lambda_i(t) = \lambda_0(t) \exp(\mathbf{X}_i(t)' \boldsymbol{\beta})$$

avec :

- $\mathbf{X}_i$ : vecteur ( $p \times 1$ ) des covariables de l'individu  $i$ .
- $\lambda_0$  : fonction non négative et non spécifiée appelée « baseline ». Il s'agit de la fonction de hasard de base.
- $\boldsymbol{\beta}$ : vecteur de  $p \times 1$  colonnes correspondants aux coefficients pour chaque covariable.

Dans le modèle de Cox, la « baseline » n'est pas spécifiée. Aucune forme ne lui est imposée : il s'agit d'un modèle semi-paramétrique. Dans cette approche, on s'attache essentiellement à définir le positionnement d'une population par rapport à une autre, sans chercher à déterminer le niveau absolu de risque.

Le ratio des hasards entre deux individus avec des vecteurs de covariables  $X_i$  et  $X_j$  **fixes** dans le temps est donné par :

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp(\mathbf{X}_i \boldsymbol{\beta})}{\lambda_0(t) \exp(\mathbf{X}_j \boldsymbol{\beta})} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{\exp(\mathbf{X}_j \boldsymbol{\beta})}$$

Les covariables peuvent être de deux types :

- constantes dans le temps : caractéristiques fixes de l'assuré (sexe, âge à la date de souscription etc.)  $\mathbf{X}_i$ .
- dépendantes du temps  $\mathbf{X}_i(t)$ .

A l'issue de chapitre 2, nous avons conclu sur la pertinence d'inclure dans le modèle :

- **des variables caractéristiques de l'assuré et de son contrat** : âge à la souscription, fréquence des primes, classe de risque etc. Ces variables sont fixes, c'est-à-dire constantes dans le temps.
- **une variable externe traduisant le contexte économique** : performance de l'indice du Dow Jones. Cette variable est stochastique et varie dans le temps.

Les covariables fixes dans le temps ne posent pas de problème conceptuel car nous avons dans notre portefeuille pour chaque individu un vecteur  $X_i$ , constitué de ses caractéristiques fixes, clairement définies.

row.names	id.contract	issue.date	end.date	acc.death.amount	gender	premium.frequency	risk.state	underwriting.age	living.place
39	39	1995-03-30	1999-12-07	NoRider	Male	Monthly	Standard-smoker	Middle	WestCoast
846	846	1995-06-19	2000-04-20	NoRider	Male	Monthly	Standard-nonSmoker	Middle	Other
1034	1034	1995-06-10	1998-01-20	NoRider	Female	Other	Standard-nonSmoker	Young	Other
1063	1063	1995-06-07	1996-09-03	NoRider	Male	Annual	Standard-nonSmoker	Middle	Other
1491	1491	1995-05-15	2009-04-07	NoRider	Male	Monthly	Prefered-nonSmoker	Middle	Other
1723	1723	1995-06-25		NoRider	Female	Monthly	Standard-nonSmoker	Young	Other
1942	1942	1995-05-16		NoRider	Male	Annual	Standard-smoker	Middle	EastCoast

X1 vecteur de covariables fixes de l'individu 1

La prise en compte d'une variable externe traduisant le contexte économique est plus problématique car elle est variée dans le temps et nécessite une réflexion approfondie. Plusieurs solutions se sont présentées à nous.

➤ Approche 1 : se ramener à une covariable fixe dans le temps

Il s'agit d'inclure dans le modèle, pour chaque individu, la valeur du DowJones à une période pertinente. Nous avons envisagé d'introduire la performance de l'indice au cours du dernier trimestre de la vie du contrat. Nous aurions ainsi pour chaque individu, une variable Xi traduisant le contexte économique au moment de la clôture du contrat, si clôture, ou à la date d'arrêt du portefeuille, si le contrat est toujours en vigueur. Cette approche présente deux défauts majeurs.

Tout d'abord, la pertinence de cette covariable pour les contrats encore en vigueur est discutable : la censure de fin de portefeuille est purement administrative, ainsi connaître la variation du DowJones sur le trimestre Mars-Juillet 2010 ne présente pas de pouvoir explicatif.

De plus, cette approche nous empêcherait d'utiliser notre modèle dans un but prédictif car elle nécessite de connaître la date de fin du contrat. Nous l'avons donc rejetée.

➤ Approche 2 : trajectoire de l'indice DowJones entre Janvier 1995 et Juillet 2010

Il s'agit de discrétiser la période Janvier 1995-Juillet 2010 en trimestres et de calculer la performance de l'indice DowJones sur chaque trimestre. Ainsi, 62 covariables sont ajoutées au portefeuille.

Id.contract	Issue.date	End.date	acc.death.amount	DJ Janv- Mars 95	DJ Mars- Juillet 95	[...]	DJ Mars- Juillet 2010
39	1995-03-30	1999-12-07	NoRider	0,11	0,07	...	0,07

Tableau 9 exemple d'insertion d'une variable dépendant du temps

Ceci nous permet d'inclure dans le modèle la trajectoire de la variation de l'indice DowJones de la date de début (Janvier 1995) à la date de fin du portefeuille (Juillet 2010).

Chaque assuré disposerait alors d'un vecteur  $X_i$ , fixe dans le temps, composé de 62 composantes où chaque composante  $X_{ij}$  représente la variation de l'indice DowJones durant le trimestre n° j.

Cette approche nous permettrait de prédire car il suffit de fixer un horizon de fin d'étude et de simuler l'indice DowJones entre la date de début de portefeuille et cet horizon.

Cependant, la visualisation de la fonction de hasard de l'individu i nous permet d'identifier un défaut majeur de cette approche :

$$\lambda_i(t) = \lambda_0(t) \exp(\sum_{j=1}^{60} X_j \beta_j + \sum_k X_{ik} \beta_{ik})$$

- $X_j$  avec  $j = 1..60$  : variation du DowJones sur l'intervalle n°j
- $\beta_j$ : les 60 coefficients calibrés par le modèle pour l'indice DowJones
- $X_{ik}$  : covariable fixe k de l'individu (âge, classe de risque etc.)

Nous pouvons en effet constater que les variations de l'indice sur la totalité des 62 trimestres sont prises en compte dans la fonction de hasard de l'individu i. Ainsi, cela revient à considérer que la performance de l'indice DowJones sur des trimestres où le contrat n'était pas ou plus en vigueur impacte la date de rachat (contrats où la date de souscription est postérieure à Janvier 1995 et/ou la date de rachat antérieure à Juillet 2010). Ceci n'a pas de sens, ainsi nous avons écarté cette approche.

➤ Approche 3 : segmenter la base en intervalle trimestriel

Il s'agit de segmenter chaque contrat en période de temps de pas trimestriel. Ainsi, chaque contrat n'est plus représenté par une ligne unique, mais par autant de lignes que de trimestres durant lesquels le contrat était en vigueur. Cette mise en forme est nécessaire pour l'implémentation du modèle sous R, cependant elle ne conduit pas à dupliquer l'individu lors de la modélisation. En effet, lors de la calibration du modèle, les individus seront identifiés par leur numéro de contrat et non par les lignes de la base de données.

	row.names	id.contract	issue.date	end.date	acc.death.amount	gender	premium.frequency	risk.state	underwriting.age	living.place
1	39	39	1995-03-30	1999-12-07	NoRider	Male	Monthly	Standard-smoker	Middle	WestCoast

	id.contract	issue.date	end.date	acc.death.amount	gender	premium.frequency	risk.state	underwriting.age	DJIAtend
1	39	1995-03-30	1995-04-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.11
2	39	1995-04-01	1995-07-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.07
3	39	1995-07-01	1995-10-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.05
4	39	1995-10-01	1996-01-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.1
5	39	1996-01-01	1996-04-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.04
6	39	1996-04-01	1996-07-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.01
7	39	1996-07-01	1996-10-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.1
8	39	1996-10-01	1997-01-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.09

Tableau 10 2eme exemple d'insertion d'une variable dépendant du temps

Pour chaque intervalle de temps (i.e pour chaque trimestre), nous calculons la variation de l'indice DowJones entre les dates de début et de fin de trimestre (variable DJIA.tend). Le détail du calcul est présenté en annexe C.

Ceci nous permet de nous affranchir du problème rencontré dans l'approche 2 car nous calculons la performance de l'indice DowJones uniquement sur les intervalles de temps où le contrat est en vigueur.

De plus, elle ne présente pas de problème pour la prédiction. En effet, quand on cherche à prédire, nous ne disposons pas de la date de rachat du contrat étant donné que c'est ce que nous cherchons à déterminer.

Il s'agit alors de fixer un horizon temporel, similaire à une date de censure administrative, et de segmenter chaque contrat en autant de lignes qu'il y a d'intervalles écoulés entre la date de souscription et l'horizon temporel. Il faudra ensuite simuler l'indice DowJones sur ces différents intervalles. Nous avons effectué une prédiction sur la partie du portefeuille que nous n'avons pas utilisé pour calibrer le modèle dans le V.5.

## V.2 Prise en compte de l'approche Fine& Gray dans notre modèle

A l'issue du chapitre 4, nous avons décidé d'adopter l'approche de Fine&Gray, consistant à étudier le processus de la sous distribution  $(\xi_t)_{t>0}$  et non plus le processus original des risques compétitifs  $(X_t)_{t>0}$ . La théorie a été présentée plus en détails dans le IV.1.4.2. La différence majeure de cette approche par rapport à une approche classique concerne le traitement des individus ayant subi un risque compétitif. Leur processus est stoppé en  $\tau^-$ , où  $\tau$  représente la date d'occurrence du risque compétitif. Les individus ayant subi un risque compétitif sont ainsi conservés dans l'ensemble à risque jusqu'à la date de censure administrative.

Il existe sous R des bibliothèques (bibliothèque « comprsk ») permettant de calibrer directement un modèle de Fine&Gray sans avoir à modifier soi-même le processus initial. Cependant, elles ne sont pas adaptées pour des modèles avec des variables dépendant du temps.

Par conséquent, ne pouvant utiliser ces bibliothèques, nous avons effectué les manipulations nécessaires sur nos données afin de respecter le cadre théorique de Fine&Gray. Ces manipulations sont décrites dans l'annexe C.

Le livre « Competing Risk and Multistate » de Jan Beyersmann, Arthur Allignol et Martin Schumacher (2012) présente une analyse assez similaire qui nous a guidés. Une fois l'ensemble à risque modifié et les processus stoppés pour les risques compétitifs, il suffit d'appliquer un modèle de Cox classique.

## V.3 Résultats de la modélisation

A l'issue des statistiques descriptives réalisées dans le II, nous avons déterminé que la variable « living.place » n'était pas significative comme facteur de risque de rachat. Nous avons également des

doutes sur la variable « genre ». Nous effectuons tout de même un premier modèle complet comportant toutes les covariables afin de pouvoir déterminer celles qui pourraient être exclues.

➤ Le modèle complet :

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_{DJIA} X_{DJIA} + \beta_{rider} X_{rider,i} + \beta_{gender} X_{gender,i} + \beta_{freq} X_{freq,i} + \beta_{risk.state} X_{risk.state,i} + \beta_{age} X_{age,i} + \beta_{premium} X_{premium,i} + \beta_{living} X_{living,i})$$

La modélisation réalisée sur 1/3 de notre portefeuille, soit 10 000 contrats, nous donne les résultats suivants :

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z )	
DJIAtend	1.82374	6.19496	0.29457	0.29952	6.089	1.14e-09	***
acc.death.amountRider	-0.23653	0.78936	0.04948	0.04933	-4.795	1.62e-06	***
premium.frequencyMonthly	0.21419	1.23886	0.04554	0.04562	4.695	2.66e-06	***
premium.frequencyOther	-0.28700	0.75051	0.06265	0.06298	-4.557	5.19e-06	***
premium.frequencyQuarterly	0.26182	1.29929	0.05163	0.05141	5.093	3.52e-07	***
premium.frequencySemi-annual	0.47755	1.61212	0.06445	0.06398	7.464	8.39e-14	***
underwriting.ageMiddle	0.08573	1.08952	0.03659	0.03648	2.350	0.0188	*
underwriting.ageOld	-0.28296	0.75355	0.04979	0.04997	-5.662	1.49e-08	***
premiumprim2	1.13345	3.10635	0.12836	0.12868	8.808	< 2e-16	***
premiumprim3	1.67819	5.35587	0.11216	0.11232	14.941	< 2e-16	***
premiumprim4	1.80442	6.07645	0.11765	0.11775	15.324	< 2e-16	***
risk.statePrefered-smoker	-0.08298	0.92037	0.11698	0.11394	-0.728	0.4664	
risk.stateStandard-nonSmoker	-0.26595	0.76647	0.05106	0.05064	-5.252	1.50e-07	***
risk.stateStandard-smoker	-0.37674	0.68609	0.05463	0.05440	-6.925	4.35e-12	***
risk.stateSubStd-nonSmoker	0.09076	1.09501	0.10381	0.10705	0.848	0.3965	
risk.stateSubStd-smoker	0.09804	1.10301	0.12333	0.12314	0.796	0.4260	
living.placeEastCoast	0.08051	1.08384	0.08797	0.08875	0.907	0.3643	
living.placeOther	0.07331	1.07606	0.08217	0.08302	0.883	0.3772	
genderMale	0.13898	1.14911	0.03351	0.03341	4.160	3.18e-05	***
---							
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							
		exp(coef)	exp(-coef)	lower .95	upper .95		
DJIAtend		6.1950	0.1614	3.4442	11.1426		
acc.death.amountRider		0.7894	1.2669	0.7166	0.8695		
premium.frequencyMonthly		1.2389	0.8072	1.1329	1.3547		
premium.frequencyOther		0.7505	1.3324	0.6634	0.8491		
premium.frequencyQuarterly		1.2993	0.7696	1.1748	1.4370		
premium.frequencySemi-annual		1.6121	0.6203	1.4221	1.8275		
underwriting.ageMiddle		1.0895	0.9178	1.0143	1.1703		
underwriting.ageOld		0.7536	1.3270	0.6832	0.8311		
premiumprim2		3.1064	0.3219	2.4139	3.9974		
premiumprim3		5.3559	0.1867	4.2975	6.6748		
premiumprim4		6.0764	0.1646	4.8242	7.6538		
risk.statePrefered-smoker		0.9204	1.0865	0.7362	1.1507		
risk.stateStandard-nonSmoker		0.7665	1.3047	0.6941	0.8464		
risk.stateStandard-smoker		0.6861	1.4575	0.6167	0.7633		
risk.stateSubStd-nonSmoker		1.0950	0.9132	0.8878	1.3506		
risk.stateSubStd-smoker		1.1030	0.9066	0.8665	1.4041		
living.placeEastCoast		1.0838	0.9226	0.9108	1.2898		
living.placeOther		1.0761	0.9293	0.9145	1.2662		
genderMale		1.1491	0.8702	1.0763	1.2269		
Concordance= 0.627 (se = 0.005 )							

Figure 31 résultats de la calibration du modèle

➤ Qualité d'ajustement globale du modèle :

Elle est donnée par les p-values des tests du ratio de vraisemblance, du test de Wald ou du test du Logrank :

Si la p-value < 5% pour les tests de LRT, Wald et Log-Rank, alors il existe au moins une covariable influant sur le taux de rachat.

Nous obtenons des p-values inférieures à 5%, ainsi notre modèle est pertinent au seuil de 5%.

Remarque : la significativité globale du modèle n'a de sens que si l'hypothèse de proportionnalité est vérifiée. Ainsi, pour pouvoir interpréter les statistiques des tests ci-dessus il faut supposer que l'hypothèse de proportionnalité est vérifiée. Nous avons effectué une vérification graphique de l'hypothèse pour toutes les covariables à l'exception du DowJones dans le IV.2.2. Nous effectuerons une vérification par test statistique de l'hypothèse de proportionnalité dans le prochain paragraphe.

➤ Significativité de chaque covariable sur la durée de vie du contrat :

Pour chaque covariable  $j$ , nous disposons de :

-  $\hat{\beta}_j$  le coefficient de la covariable dans le modèle, désigné par « coef » dans la sortie R

-  $\exp(\hat{\beta}_j)$  désigné par « exp(coef) »

-  $\sqrt{V(\hat{\beta}_j)}$  désigné par « se(coef) »

Pour chaque covariable, un test de Wald est effectué afin de tester sa significativité :

Test de  $H_0 : \beta_j = 0$

$$\text{Statistique de Wald : } z_j = \frac{\sqrt{n}\hat{\beta}_j}{\sqrt{V(\hat{\beta}_j)}} ; p_j = P(|U| > z_j) \text{ avec } U \sim N(0, 1)$$

La covariable « living.place » n'est pas significative et peut donc être exclue du modèle. La covariable « genre » quant à elle, est significative dans le modèle, donc nous la conservons. Concernant la covariable « risk.state », seule la modalité « Standard » de « risk.state » est significative. Ceci est vraisemblablement causé par la faible exposition de certaines classes : « Substandard smoker » (2% du portefeuille) et « Preferred Smoker » (2%). Il semblerait ainsi pertinent de fusionner certaines modalités de la covariable.

Nous avons en effet remarqué lors des statistiques descriptives que l'on pouvait distinguer le comportement des individus des classes de risque standard avec ceux des classes de risques « extrêmes ». Nous allons ainsi fusionner la variable en créant uniquement deux classes de risque : une classe « Standard », regroupant les fumeurs et non fumeurs, et une classe « Substandard and Preferred », regroupant également fumeurs et non-fumeurs.

La modalité « Middle » de la covariable « underwriting.age » n'est pas significative alors que la modalité « Old » est très significative. Il ne s'agit pas d'un problème d'exposition car la classe Middle concerne près de 34% du portefeuille. Ceci est à mettre en relief avec une observation formulée lors des statistiques descriptives. Nous avons en effet remarqué que les « Middle » et les « Young » avaient un comportement de rachat assez similaire, par opposition à la classe « Old ». Ces observations nous suggèrent une fusion des classes « Middle » et « Young ».

Les autres covariables du modèle sont significatives à 5%.

Un nouveau modèle incluant les améliorations précédentes peut alors être calibré :

```

n= 302508, number of events= 3584

              coef exp(coef) se(coef) robust se      z Pr(>|z|)
DJIAtend      1.83029   6.23571  0.29456   0.29946  6.112 9.84e-10 ***
acc.death.amountRider -0.23535   0.79030  0.04947   0.04937 -4.767 1.87e-06 ***
premium.frequencyMonthly 0.21651   1.24173  0.04553   0.04559  4.748 2.05e-06 ***
premium.frequencyOther -0.28789   0.74985  0.06264   0.06298 -4.571 4.86e-06 ***
premium.frequencyQuarterly 0.26376   1.30181  0.05160   0.05143  5.128 2.92e-07 ***
premium.frequencySemi-annual 0.47925   1.61487  0.06445   0.06396  7.493 6.74e-14 ***
ageYoung      0.31802   1.37440  0.04720   0.04745  6.703 2.05e-11 ***
premiumprim2   1.13071   3.09786  0.12834   0.12862  8.791 < 2e-16 ***
premiumprim3   1.67849   5.35747  0.11215   0.11229 14.948 < 2e-16 ***
premiumprim4   1.80629   6.08783  0.11763   0.11768 15.349 < 2e-16 ***
risk.state2Standard -0.31817   0.72748  0.04072   0.04063 -7.832 4.77e-15 ***
genderMale     0.14041   1.15075  0.03350   0.03340  4.204 2.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
DJIAtend      6.2357  0.1604  3.4673 11.2147
acc.death.amountRider 0.7903  1.2653  0.7174  0.8706
premium.frequencyMonthly 1.2417  0.8053  1.1356  1.3578
premium.frequencyOther 0.7498  1.3336  0.6628  0.8484
premium.frequencyQuarterly 1.3018  0.7682  1.1770  1.4399
premium.frequencySemi-annual 1.6149  0.6192  1.4246  1.8306
ageYoung      1.3744  0.7276  1.2524  1.5083
premiumprim2   3.0979  0.3228  2.4076  3.9861
premiumprim3   5.3575  0.1867  4.2991  6.6764
premiumprim4   6.0878  0.1643  4.8338  7.6671
risk.state2Standard 0.7275  1.3746  0.6718  0.7878
genderMale     1.1507  0.8690  1.0778  1.2286

Concordance= 0.626 (se = 0.005 )
Rsquare= 0.003 (max possible= 0.188 )
Likelihood ratio test= 851 on 12 df, p=0
Wald test = 672.9 on 12 df, p=0
Score (logrank) test = 721.4 on 12 df, p=0, Robust = 834.9 p=0

(Note: the likelihood ratio and score tests assume independence of
observations within a cluster, the Wald and robust score tests do not).

```

Figure 32 résultats de la calibration du modèle

➤ Effet des covariables :

L'exponentielle des coefficients mesure l'effet multiplicatif sur le hasard d'une augmentation d'une unité de la covariable, toutes choses égales par ailleurs.

- La modalité de référence pour la fréquence des primes est la modalité « Annual ». Les exponentielles des coefficients des autres modalités traduisent l'effet de la modalité par rapport à la référence « Annual ». Seule la fréquence « Other » (versement libre) diminue la probabilité de racheter par rapport à la fréquence annuelle. Le classement des fréquences de paiement des primes en termes de risque de rachat est le suivant : semestrielle > trimestrielle > mensuelle > annuelle > versements libres. Ce classement est bien celui qui a été observé lors des statistiques descriptives sur le portefeuille et semble conforme à l'intuition des praticiens.

- La présence d'une option en cas de décès accidentel « acc.death.amount.rider =1 » multiplie par 0.79 la fonction de hasard. Par conséquent, détenir sur son contrat cette option diminue la probabilité instantanée de rachat du contrat. Ceci est conforme aux statistiques descriptives, où nous avons remarqué que la durée de vie des contrats avec rider était supérieure à celle des contrats dépourvus de l'option. Cette option semble dissuader quelque peu l'assuré de racheter son contrat.

- Appartenir à la classe de risque « Standard » diminue la probabilité de racheter son contrat car cela multiplie par 0.73 la fonction de hasard. Cela est en adéquation avec les statistiques descriptives où les

assurés des classes de risques « extrêmes » présentaient des durées de vie de contrats inférieures aux classes « Standard ».

- Un assuré « jeune » (i.e âgé de moins de 55 ans) multiplie par 1.37 la probabilité de racheter son contrat. Nous avons bien formulé cette intuition lors de notre étude bibliographique : les jeunes rachètent plus car ils ont des besoins de financement plus importants.

- Concernant le montant des primes, nous avons observé sur les statistiques descriptives que plus le montant de la prime était élevé, plus la durée de vie du contrat était faible. Nous retrouvons cela dans les résultats du modèle. En effet, les coefficients suggèrent également que plus le montant des primes est important, plus la probabilité instantanée de racheter est forte.

- Concernant la covariable du DowJones, le coefficient est positif et vaut 1,83. Ceci signifie qu'une performance positive du DowJones augmente la probabilité de rachat. Ainsi, durant les périodes de crise financière, le taux de rachat diminue et durant les périodes haussières il augmente. Cette constatation est plutôt logique dans la mesure où le produit offre une garantie de rendement mais est à nuancer à cause de la baisse des dividendes perçus par l'assuré.

#### **V.4 Vérification de l'hypothèse de proportionnalité**

Nous avons évoqué précédemment que la significativité des coefficients et du modèle suppose que l'hypothèse de proportionnalité des covariables soit vérifiée. Les coefficients peuvent très bien être significatifs mais le modèle mal spécifié.

Nous avons dans la partie IV.2.2 vérifié graphiquement l'hypothèse de proportionnalité pour les covariables fixes. Nous n'étions pas en mesure de vérifier de façon empirique si l'hypothèse est valide pour l'indice DowJones qui dépend du temps. Nous allons maintenant procéder à des tests statistiques pour vérifier l'hypothèse de proportionnalité pour toutes les covariables y compris le DowJones.

Le principe du test réalisé est le suivant : une alternative au modèle à hasards proportionnels est un modèle avec des coefficients qui varient dans le temps :

$$\lambda_i(\mathbf{t}) = \lambda_0(\mathbf{t}) \exp(\mathbf{X}_i' \boldsymbol{\beta}(\mathbf{t}))$$

Si le coefficient varie dans le temps, l'impact de chaque covariable sur le hasard du rachat varie dans le temps. Le cas particulier  $\beta(t) = \beta$  implique la proportionnalité des hasards. Ainsi, si le graphique de  $\beta_j(t)$  en fonction du temps est une droite alors l'hypothèse de proportionnalité est validée. La procédure `cox.zph` sous R fournit les graphiques des résidus normalisés de Schoenfeld ainsi que les p.value du test statistique de l'hypothèse de proportionnalité :

$$H_0 \beta(t) = \beta$$

	rho	chisq	p
DJIAtend	-0.00283	0.0306	0.8611
acc.death.amountRider	-0.02143	1.6547	0.1983
premium.frequencyMonthly	0.01772	1.1377	0.2861
premium.frequencyOther	-0.01099	0.4390	0.5076
premium.frequencyQuarterly	0.02643	2.4980	0.1140
premium.frequencySemi-annual	0.02295	1.8686	0.1716
ageYoung	0.04143	6.2546	0.0124
premiumprim2	0.00529	0.1010	0.7506
premiumprim3	0.01151	0.4793	0.4887
premiumprim4	0.01879	1.2781	0.2583
risk.state2Standard	-0.01476	0.7826	0.3764
genderMale	-0.00654	0.1529	0.6958
GLOBAL	NA	18.2953	0.1070

Figure 33 tests de la dépendance des coefficients dans le temps

On ne rejette pas  $H_0$  au seuil  $\alpha$  si p-value  $> \alpha$

Ainsi pour toutes les covariables (sauf pour la variable âge), l'hypothèse de proportionnalité est acceptée au seuil de 5%. Elle est également acceptée pour toutes les covariables au seuil de 10%.

➤ Graphiques  $\beta_{DJIA}(t)$  en fonction du temps :

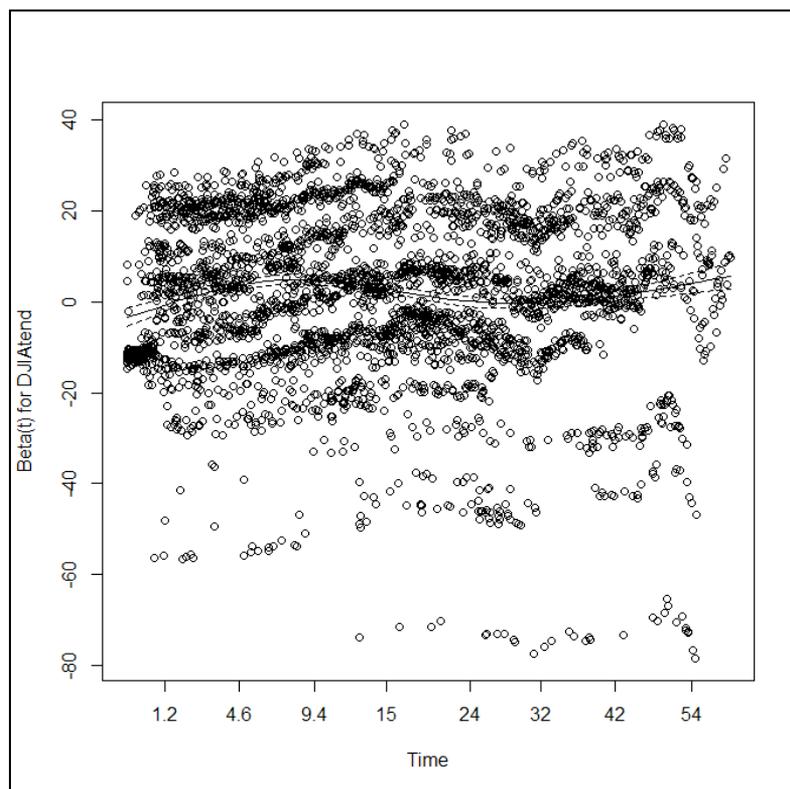


Figure 34 test de la dépendance du coefficient de l'indice DowJones dans le temps

➤ Graphiques  $\beta_{rider}(t)$  en fonction du temps :

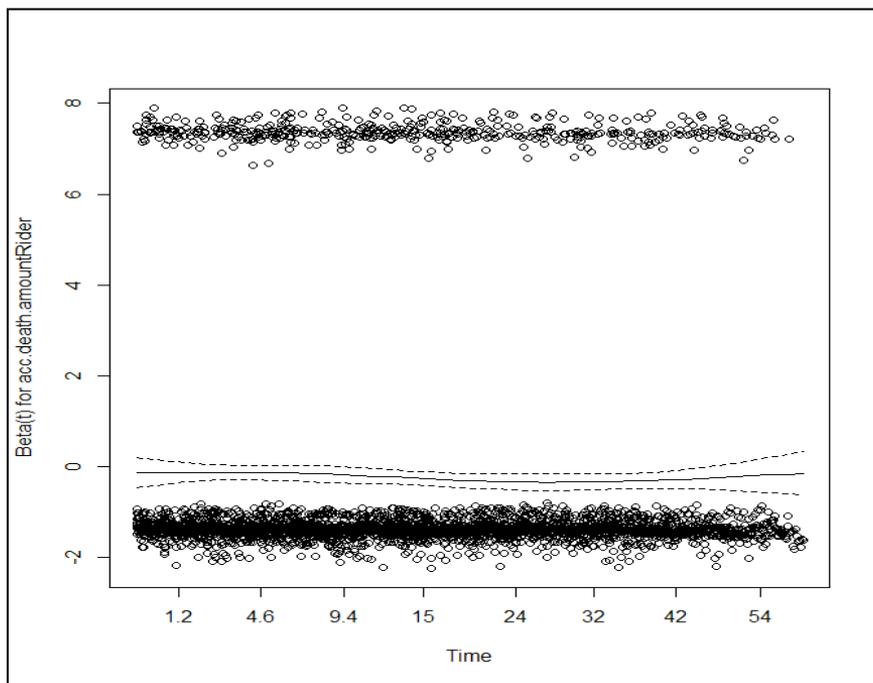


Figure 35 test de la dépendance du coefficient de l'option Rider dans le temps

Pour les autres covariables, les graphiques sont présentés en annexe D.

Nous avons ainsi validé le modèle par les tests de significativité et par les tests de validation de l'hypothèse de proportionnalité.

Il s'agit maintenant de valider le pouvoir prédictif de notre modèle en l'appliquant sur la partie du portefeuille que nous n'avons pas utilisée pour effectuer la calibration. Nous allons ainsi pouvoir comparer, sur cette partie du portefeuille, les prédictions fournies par le modèle avec la réalité.

Afin de valider la prédiction du modèle : deux approches sont possibles. La première consiste à comparer l'estimateur de Kaplan-Meier non paramétrique avec la fonction de survie estimée par le modèle sur une même sous-population.

La deuxième approche consiste à reconstituer le taux de rachat à partir de la prédiction du modèle et de le comparer avec le taux historique observé sur le portefeuille.

## V.5 Comparaison de la fonction de survie du modèle avec l'estimateur de Kaplan-Meier

Nous allons dans un premier temps comparer l'estimateur de Kaplan-Meier réalisé sur la population totale du portefeuille avec la fonction de survie de la « baseline » du modèle. La « baseline » du modèle est supposée représenter un individu moyen.

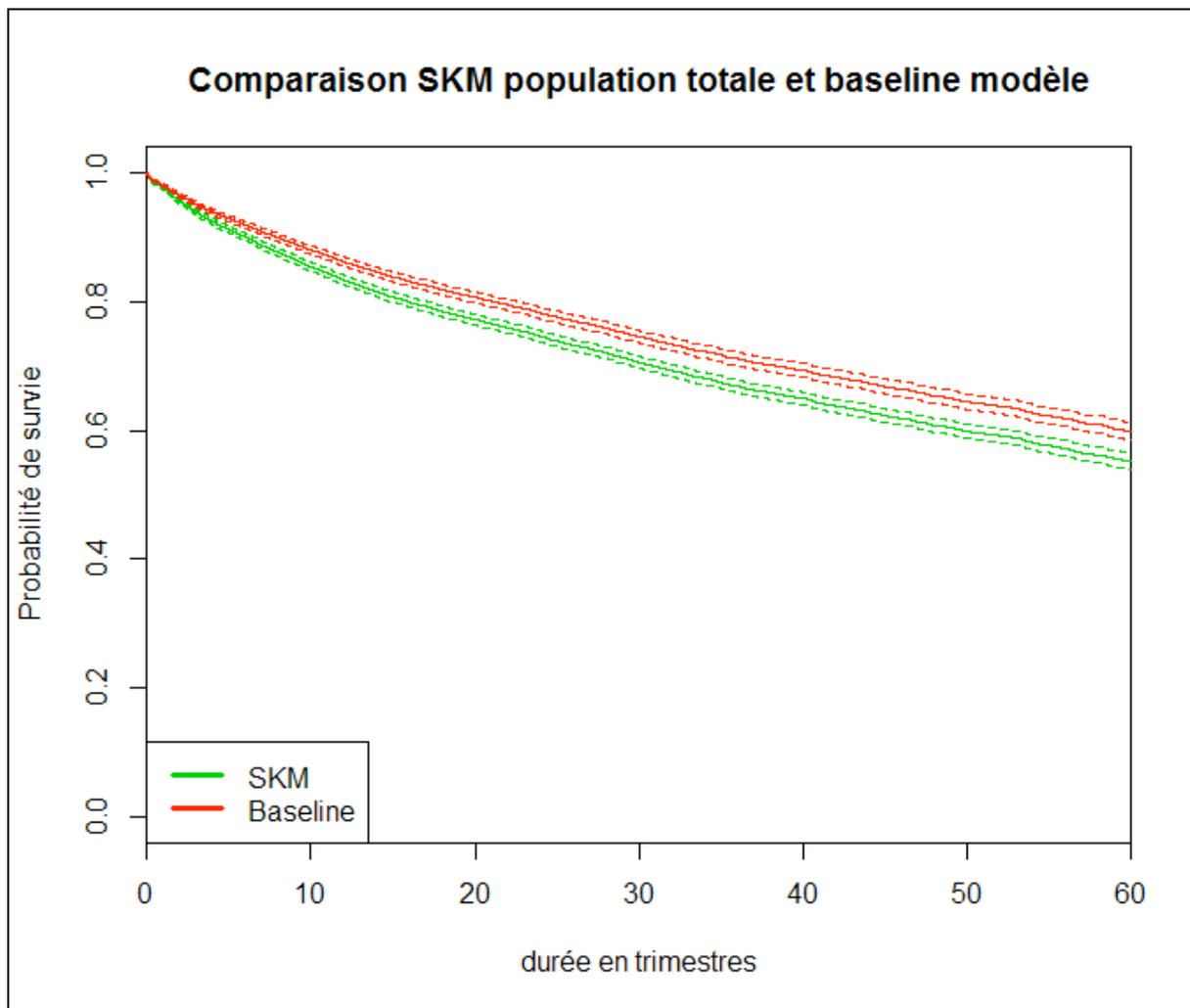


Figure 36 comparaison de la « Baseline » avec l'estimateur de Kaplan Meier

Nous pouvons être étonnés par le fait que la fonction de survie de la « baseline », censée représenter un individu moyen soit si loin de l'estimateur de Kaplan-Meier. Ceci s'explique par le fait que les populations concernées dans les deux cas ne sont pas les mêmes. L'estimateur de Kaplan-Meier s'applique sur la population totale alors que la « baseline » s'applique sur un individu moyen. En effet, la « baseline » correspond à la fonction de survie d'un individu fictif dont les covariables sont constantes dans le temps et fixées à la valeur moyenne des covariables du portefeuille. Ceci ne pose pas de problème conceptuel pour les covariates caractéristiques de l'assuré qui sont fixes dans le temps (sexe etc.) mais c'est en revanche problématique pour l'indice du DowJones. En effet, cela revient à considérer un assuré qui observe durant toute la vie de son contrat une performance du DowJones constante. Or, nous avons observé lors des statistiques descriptives sur notre portefeuille, un pic de rachat en 2009 lié à la crise financière. Ceci n'est pas capté par la « baseline », donc elle sous-estime la probabilité de racheter son contrat.

Sa fonction de survie est par conséquent au dessus de celle de l'estimateur de Kaplan-Meier.

Afin, de comparer l'estimateur de Kaplan-Meier à une population réelle et non à un individu fictif, nous allons définir une population significative sur notre portefeuille. Ceci consiste à fixer des modalités

pour les différentes covariables de façon à définir une sous-population de taille suffisante. On applique alors le modèle puis l'estimateur de Kaplan-Meier sur la sous-population.

La sous-population A est la suivante :

Covariable	Modalité
« Gender »	Female
« Acc.death.amount »	NoRider
« Risk.state »	Standard
« Premium.fréquence »	Monthly
« Premium »	Prim3
« Age »	Old

Tableau 11 exemple de sous population

Nous obtenons le graphique suivant :

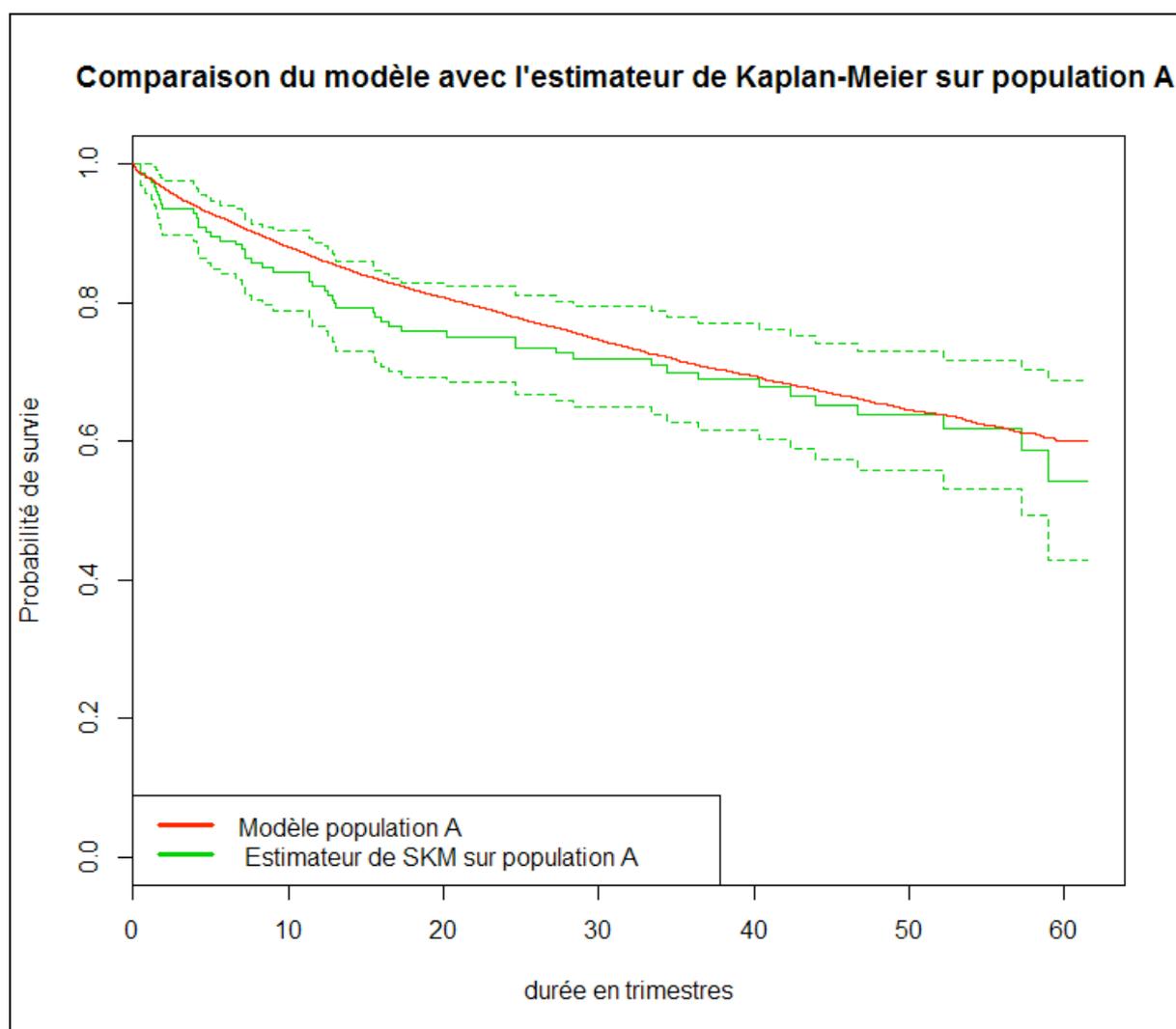
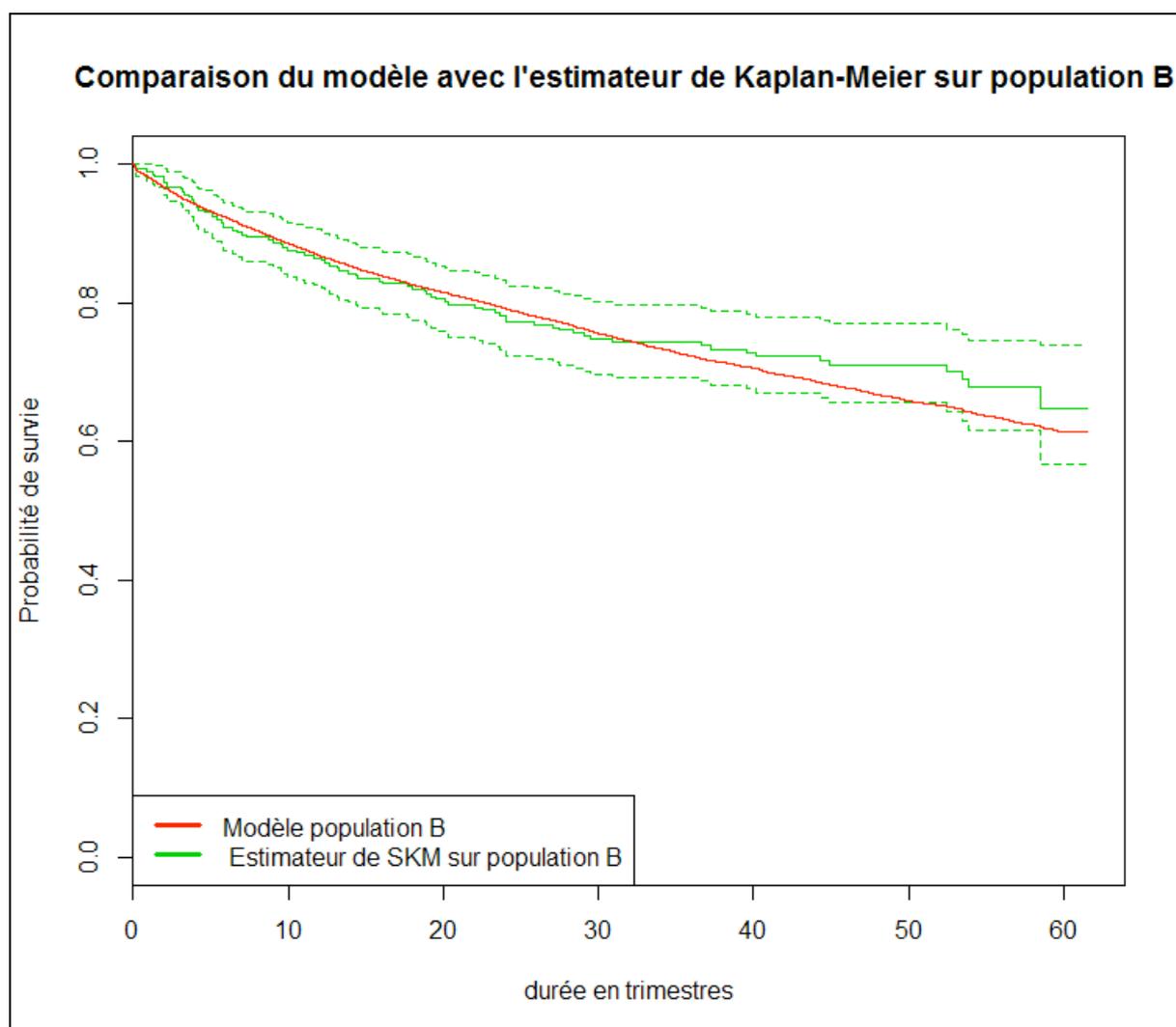


Figure 37 validation du modèle sur une sous population

Nous pouvons remarquer que la fonction de survie donnée par le modèle est située légèrement au dessus de l'estimateur mais toujours dans son intervalle de confiance à 95%. Cette comparaison du modèle avec l'empirique est satisfaisante.

Nous procédons de même sur une seconde population :

Covariable	Modalité
« Gender »	Male
« Acc.death.amount »	NoRider
« Risk.state »	Standard
« Premium.fréquence »	Other
« Premium »	Prim3
« Age »	Young



**Figure 38 validation du modèle sur une sous population**

La probabilité de survie du modèle est inférieure à celle de l'estimateur de Kaplan-Meier à partir du 30<sup>ème</sup> trimestre. Cependant, le modèle est toujours dans l'intervalle de confiance à 95% de l'estimateur de Kaplan-Meier, ce qui est satisfaisant.

Conclusion :

Les validations réalisées via la comparaison avec l'estimateur de Kaplan-Meier présentent des résultats encourageants, cependant elles ne sont pas exhaustives car elles concernent quelques sous-populations. Il faudrait procéder à de telles vérifications pour toutes les combinaisons possibles des modalités des différentes variables. Cependant, cette méthode de validation est fastidieuse. Une autre méthode de validation de notre modèle consiste à comparer, sur un échantillon assez important, le nombre de rachats observés avec celui prévu par le modèle. C'est cette approche que nous adoptons dans la rubrique qui suit.

## V.6 Reconstitution du nombre de rachats

L'approche permettant de valider notre modèle est la suivante : pour un échantillon (de taille assez importante, 6 000 contrats environ) n'ayant pas servi à la calibration, nous allons déterminer pour chaque trimestre le nombre de rachats estimés et comparer cette prévision avec le nombre de rachats observés par trimestre.

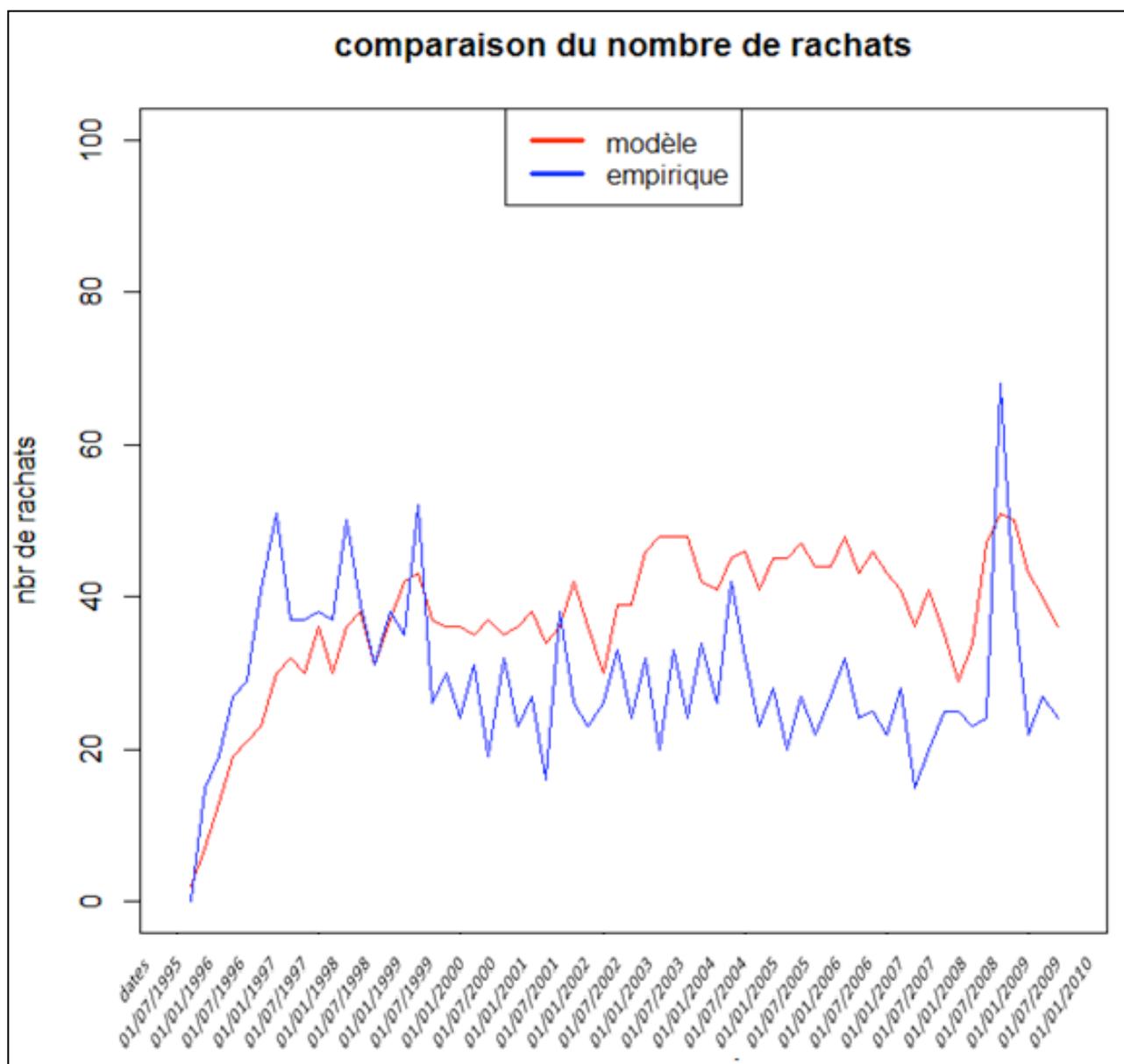


Figure 39 reconstitution du nombre de rachat sur un échantillon

Nous observons qu'au début de la période d'observation (janvier 1995) jusqu'à la fin de l'année 1998, le nombre de rachats observés sur le portefeuille croît fortement. Ce fort taux de rachat initial s'explique par le fait que le produit vient d'être lancé sur le marché. Il est probable que le produit fasse l'objet de politiques commerciales visant à réaliser un maximum de souscriptions. Certains clients peuvent avoir souscrit sous les conseils d'un agent de vente, mais réalisent quelques mois plus tard que le contrat n'est pas adapté à leurs attentes. De plus, le niveau du taux de rachat a tendance à diminuer à la fin de la période d'observation du portefeuille (hormis le pic de 2009). Cette diminution du taux de rachat est liée à la diminution de l'exposition du portefeuille : il n'y a plus de nouvelles souscriptions, que des clôtures de contrat.

Le modèle a du mal à capter cette dynamique de lancement du produit et sous-estime le nombre de rachat entre 1995 et 1998. Ceci n'a rien d'étonnant car nous n'avons intégré aucune covariable susceptible de traduire un tel effet de rodage. Il faudrait introduire dans le modèle une covariable traduisant les stratégies marketing du produit. Elles se décomposent généralement en deux phases :

- phase 1 : fortes incitations à la souscription lors des premières années suivant le lancement. Ceci se traduit par une période transitoire instable avec de nombreuses entrées dans le portefeuille (souscriptions) mais également de nombreuses sorties, notamment les rachats.

- phase 2 : délaissement du produit par les commerciaux suite à de nouvelles arrivées sur le marché. Le portefeuille connaît alors une phase de stabilité car le nombre de souscriptions diminue et les rachats causés par des « erreurs commerciales » (i.e. les clients ayant souscrit au produit alors qu'il n'est pas adapté à leurs attentes) peuvent être négligés.

Les stratégies marketing du réseau de distribution sont cependant difficiles à appréhender finement car elles sont confidentielles et propres à chaque acteur du réseau. Par conséquent, nous tentons d'appliquer notre modèle sur un portefeuille en régime de croisière, où les rachats faisant suite à des « erreurs marketing » peuvent être négligés.

La phase de stabilisation du portefeuille semble débuter en 2000. Notre modèle surestime alors le nombre de rachats, en moyenne 8 rachats de plus par trimestre, mais réussit à capter certaines tendances. En effet, il capte bien l'effet en 2009 de la crise financière des subprimes. Ainsi, la variable conjoncturelle introduite est satisfaisante mais l'impact de la crise est fortement sous-estimé.

Afin de tenter d'améliorer la performance de notre modèle, nous avons voulu le recalibrer uniquement sur un portefeuille stable. Nous avons enlevé les contrats souscrits entre 1995 et 1996, afin de tenter d'atténuer l'effet de rodage dû à la période lancement du produit. Nous obtenons les résultats suivants :

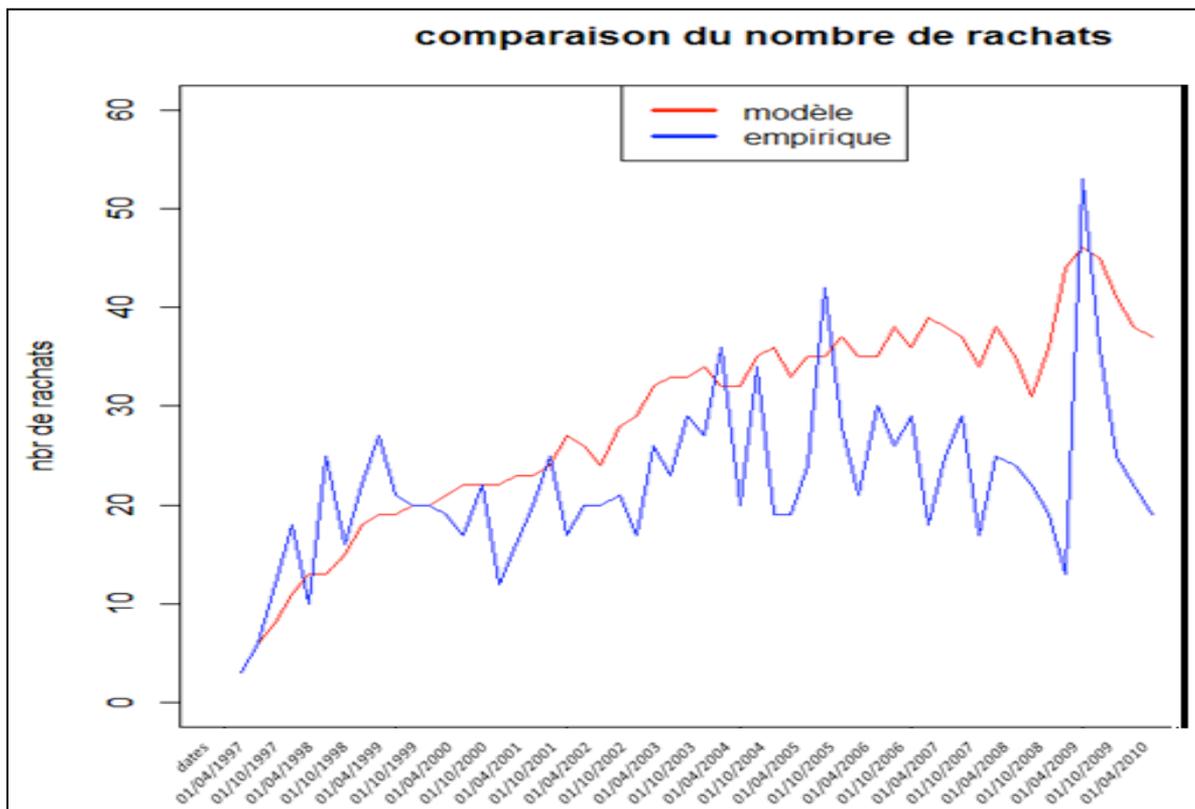


Figure 40 reconstitution du nombre de rachat sur un échantillon

Nous remarquons qu'en début de période d'observation, l'échantillon de validation présente moins de rachats que le précédent. Ceci confirme bien le fait que nous avons ainsi atténué l'effet de rodage. Cependant notre modèle sous-estime toujours, sur les premières années, le nombre de rachats puis les surestime. En calibrant notre modèle à partir de 1997, nous avons supposé que la période de rodage dure seulement deux ans. En réalité, elle dure certainement plutôt cinq ans, cependant nous n'aurions plus eu assez de données pour calibrer de manière satisfaisante le modèle.

## V.7 Evolution vers un modèle paramétrique

Outre l'effet de rodage existant sur notre portefeuille et qui n'est pas pris en compte par notre modèle, nous nous sommes interrogés sur sa capacité à bien prendre en compte la dimension structurelle du rachat. Cette dernière est dictée par la forme de la « baseline »  $\lambda_0(t)$ , car il s'agit de la fonction de hasard d'un individu moyen de la population dont toutes les covariables sont égales aux valeurs moyennes observées et sont fixes dans le temps. Par conséquent, il s'agit d'un individu fictif soumis à une variation du DowJones constante dans le temps. Cet individu fictif n'est pas vraiment soumis au contexte conjoncturel vu que ce dernier est stable dans le temps. Notre modèle étant semi-paramétrique, nous n'avons pas imposé de loi à la « baseline ». Elle a été estimée sur l'échantillon de calibration et peut être visualisée sur le graphique suivant :

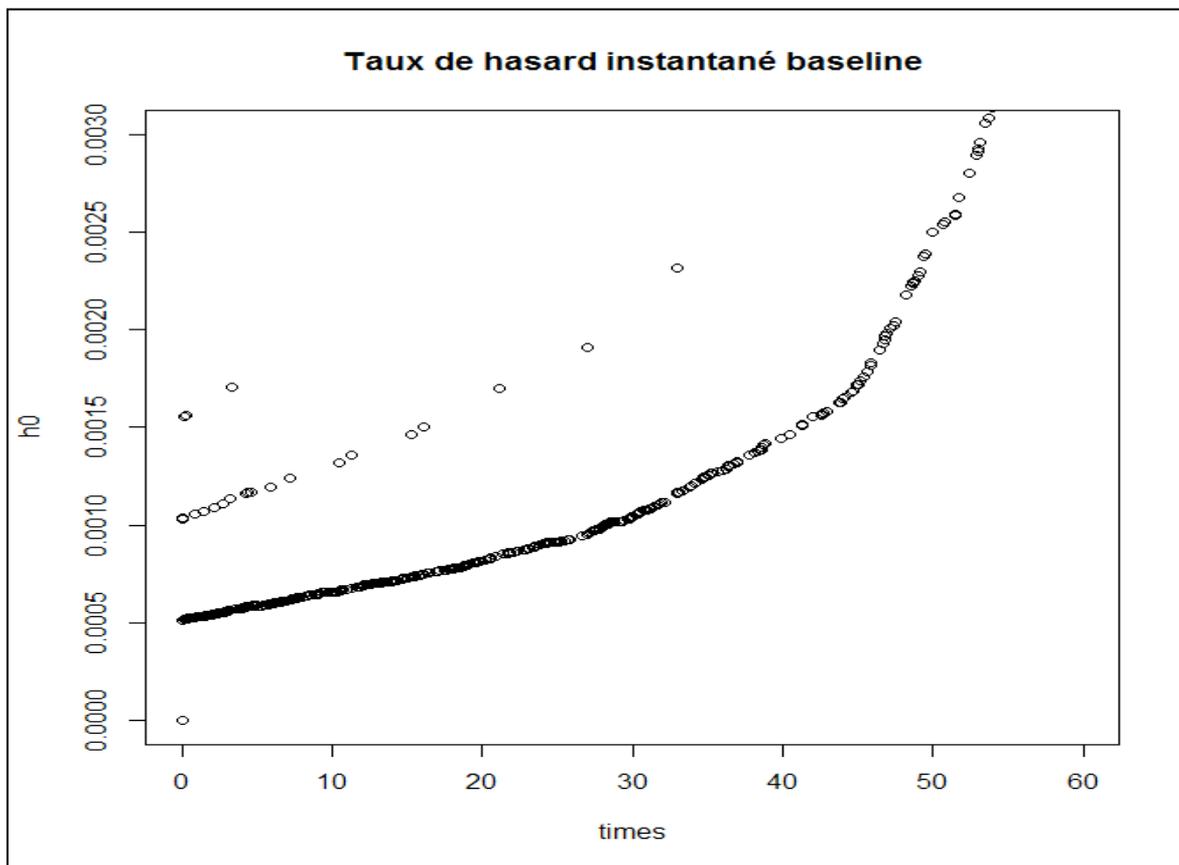


Figure 41 fonction « Baseline » du modèle

Nous remarquons que la « baseline » de notre modèle est croissante dans le temps. Pour rappel, la fonction de hasard empirique sur notre portefeuille est la suivante :

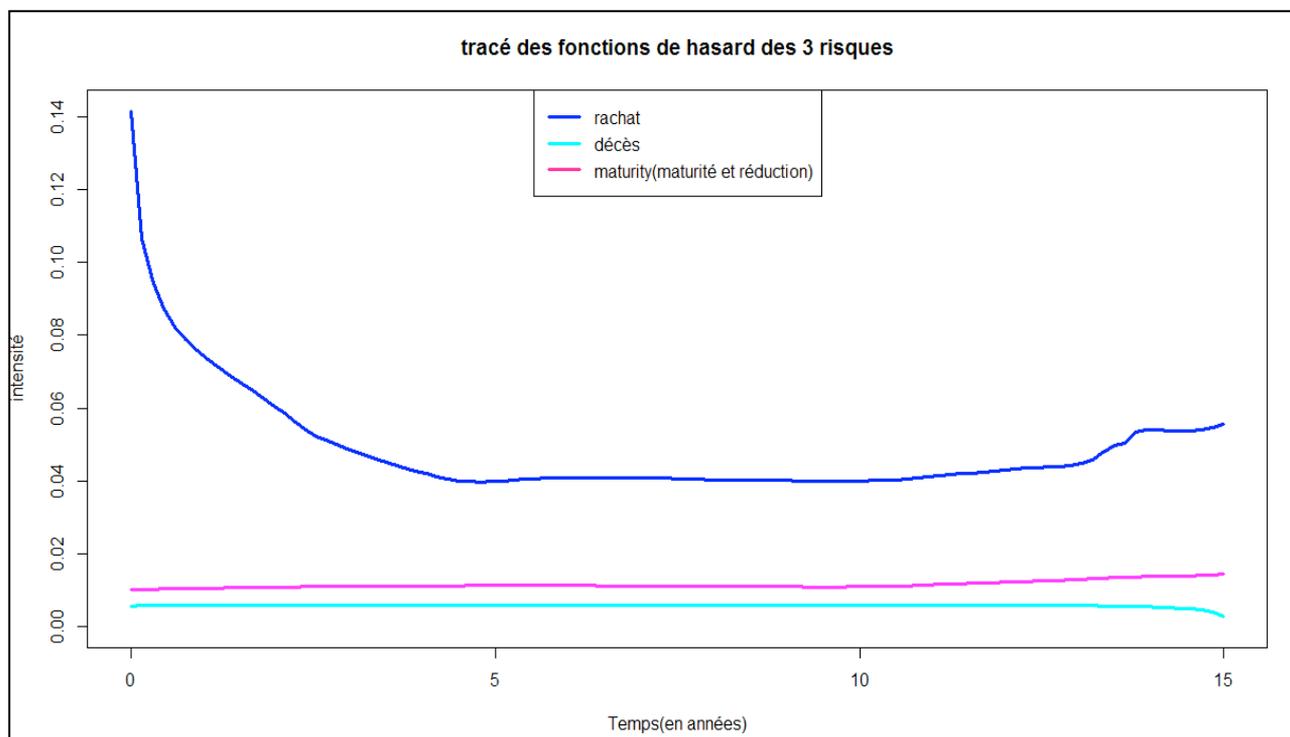


Figure 42 fonction de hasard des trois risques du portefeuille

Ainsi, contrairement à nos intuitions, un modèle semi-paramétrique ne calibre pas forcément une « baseline » conforme aux observations empiriques. Un modèle paramétrique, où la forme de la « baseline » est imposée, semble plus pertinent. Imposer à la « baseline » une distribution de Weibull, avec un paramètre de forme inférieur à 1, semble judicieux.

Il est possible de vérifier graphiquement si un modèle paramétrique de Weibull est adapté à nos données. En effet, dans une modélisation de Weibull, la fonction de survie vérifie :

$$S(t) = e^{-\lambda t^\alpha} \text{ d'où :}$$

$$\log(-\log(S(t))) = \log \lambda + \alpha \log(t)$$

On note  $\widehat{S}(t)$  : la fonction de survie empirique.

Si les points  $(\log(t), \log(-\log(\widehat{S}(t))))$  sont alignés, l'hypothèse d'une « baseline » suivant une loi de Weibull est adaptée.

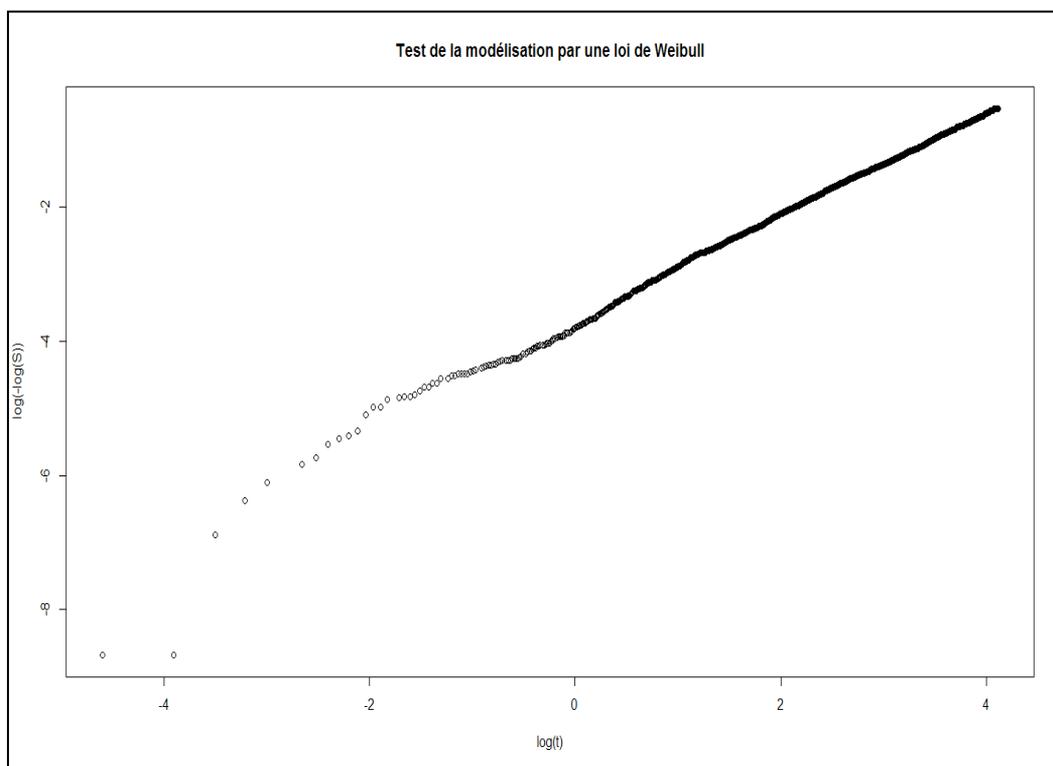


Figure 43 test graphique pour la modélisation par une loi de Weibull

Ainsi, l'hypothèse d'une « baseline » suivant une loi de Weibull est valide sur nos données.

Nous avons cherché de manière empirique le coefficient de forme qui correspondait le mieux à nos données. Il s'agit du coefficient  $\alpha=0,9$ .

Une fois le coefficient de forme fixé, nous avons calibré le modèle et obtenu les résultats suivants :

```

Call:
weibreg(formula = Surv(time1, sh.time, outcome == 2) ~ premium.frequency +
  DJIAtend + risk.state2 + acc.death.amount + premium, data = data_s1,
  shape = 0.9)

Covariate      Mean      Coef Exp(Coef)  se(Coef)    Wald p
premium.frequency
  Annual      0.250      0          1          (reference)
  Monthly     0.335     0.290     1.336     0.059     0.000
  Other       0.166    -0.237     0.789     0.081     0.004
  Quarterly   0.179     0.373     1.452     0.066     0.000
  Semi-annual 0.069     0.470     1.600     0.086     0.000
DJIAtend      0.012     2.292     9.897     0.372     0.000
risk.state2
  nonStandard 0.158      0          1          (reference)
  Standard    0.842    -0.378     0.685     0.052     0.000
acc.death.amount
  NoRider     0.830      0          1          (reference)
  Rider       0.170    -0.307     0.735     0.065     0.000
premium
  prim1       0.109      0          1          (reference)
  prim2       0.099     1.147     3.148     0.170     0.000
  prim3       0.660     1.698     5.465     0.148     0.000
  prim4       0.132     1.884     6.581     0.155     0.000
log(scale)                6.239  512.560  0.175     0.000

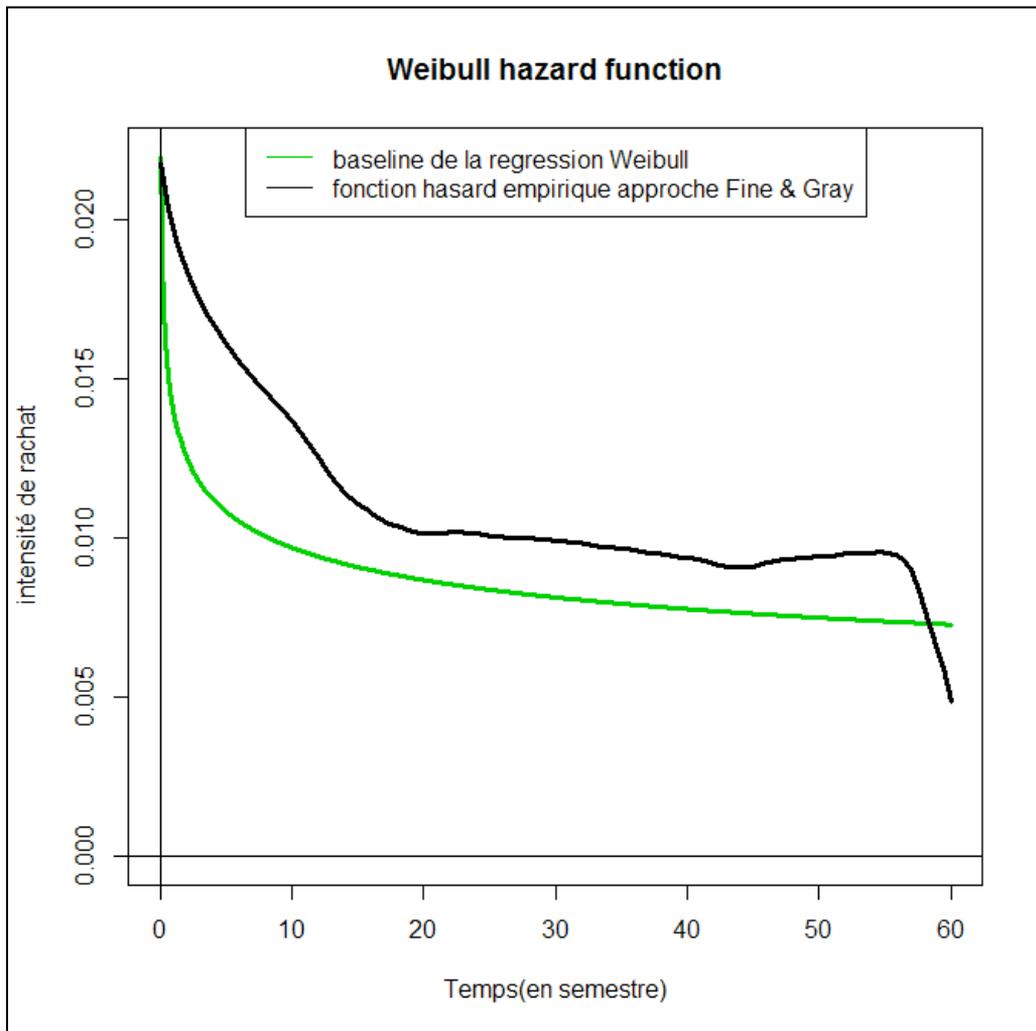
Shape is fixed at 0.9

Events                2141
Total time at risk    197467
Max. log. likelihood  -11502
LR test statistic      537
Degrees of freedom     10
Overall p-value        0

```

Figure 44 résultats du modèle de Weibull

La forme de notre « baseline » est la suivante:  $\lambda_0(t) = \Phi at^{\alpha-1}$  avec  $\Phi=512,56$  et  $\alpha=0,9$ .



**Figure 45** comparaison de la fonction de hazard de base du modèle de Weibull avec la fonction de hazard empirique du portefeuille

Il s'agirait maintenant, pour valider un tel modèle, de procéder à une comparaison avec l'estimateur de Kaplan-Meier et de reconstituer le taux historique de rachat. Par manque de temps, nous n'avons pu effectuer les mêmes vérifications que celles faites pour le modèle de Fine&Gray. Ainsi, cette modélisation paramétrique est donnée à titre d'ouverture et mériterait d'être davantage approfondie.

# Conclusion générale du mémoire

Notre mémoire propose une modélisation inédite des comportements de rachat par analyse de survie. Nous avons ainsi eu un long travail de définition et de mise en place du cadre théorique de notre étude. Nous avons remis en question et fait évoluer ce cadre à plusieurs reprises.

En effet, nous avons débuté notre étude dans un cadre classique d'analyse de survie puis nous avons rejeté cette approche afin d'évoluer vers un cadre de risques compétitifs.

Nous étions alors confrontés à plusieurs approches possibles pour notre modélisation : modéliser les fonctions de hasards de chaque risque compétitif ou une unique fonction de hasard, celle de la sous-distribution. Nous avons alors opté pour la seconde approche et avons implémenté un modèle à hasards proportionnels de type Fine&Gray.

Notre modèle se révèle satisfaisant concernant les deux points suivants :

- La calibration : les variables retenues pour la modélisation du taux de hasard du rachat sont significatives. Ces résultats sont confirmés par les études statistiques que nous avons faites en amont.
- La comparaison avec l'estimateur non-paramétrique de Kaplan-Meier sur des sous-populations de tailles significatives : les fonctions de survie issues du modèle sont contenues dans l'intervalle de confiance à 95% des estimateurs de Kaplan-Meier.

Cependant la reconstruction du nombre de rachats, à l'aide du modèle implémenté, n'est pas satisfaisante. Le modèle ne prend pas en compte le fort nombre de rachats qui suit le lancement du produit. Ensuite, il arrive à capter certaines tendances mais surestime le nombre de rachats.

Pour améliorer le modèle, nous nous sommes intéressés aux modèles de régression de fonctions de hasard paramétriques. Ces méthodes permettent de spécifier une loi pour la fonction de hasard de base (« baseline hazard»). Une première analyse nous suggère d'imposer une loi de Weibull. Nous aurions aimé approfondir davantage cette approche.

# VI. Annexes

## VI.1 Annexe statistiques descriptives

### VI.1.1 Méthode de segmentation de la prime annuelle

Nous avons décidé de segmenter la prime annuelle en classe vu qu'il était plus judicieux d'avoir une variable à modalité dans les modèles à hasards proportionnels que nous avons implémentés. De plus le fait d'avoir des modalités atténue l'effet des valeurs extrêmes lors de la calibration. Il y a par exemple dans notre portefeuille des contrats ayant une prime très faible.

La méthode utilisée est une méthode Anova expliquant la variable prime en fonction de la variable booléenne rachat, de sorte à créer des classes les plus homogènes possible. Le but de la méthode étant de maximiser la variance inter classes et de minimiser la variance intra classe. Le seuil a été fixé à 10<sup>-3</sup> après plusieurs seuils testés. Il a été choisi de sorte à ne pas avoir trop de classes fines, ou peu de classes avec beaucoup d'hétérogénéité. Notons que plus le seuil est bas, plus le modèle est contraignant et les classes sont nombreuses, sinon c'est l'effet inverse.

Les quatre classes obtenues (intervalles ouverts à gauche et fermés à droite) sont les suivantes :

- prim1 clients pour lesquels la prime annualisée est comprise entre 0 \$ et 78,85 \$
- prim2 clients pour lesquels la prime annualisée est comprise entre 78,85 \$ et 151,9 \$
- prim3 clients pour lesquels la prime annualisée est comprise entre 151,9 \$ et 1 068 \$
- prim4 clients pour lesquels la prime annualisée est supérieure à 1 068 \$.

## VI.1.2 ACM : Tableaux de la construction des axes factoriels

Aides à l'interprétation pour les modalités actives												
Variable	Modalités actives Modalité	!	CONTR POIDS		!	AXE1					!	
			COORD	CTR		RCTR	CO2	QLT	V.TEST			
duration2	dur1	!	4.11	3.24	!	0.95	20.4	1	31.7	31.7	104.55	!
	dur2	!	4.19	3.08	!	0.13	0.4	23	0.6	0.6	13.80	!
	dur3	!	4.19	3.08	!	-0.33	2.3	11	3.5	3.5	-34.82	!
	dur4	!	4.18	3.09	!	-0.80	13.7	2	21.0	21.0	-85.06	!
duration2			16.67	12.50			36.8					
gender	Female	!	2.78	6.25	!	-0.22	2.1	12	4.9	4.9	-43.06	!
	Male	!	2.78	6.25	!	0.22	2.1	13	4.9	4.9	43.07	!
gender			5.56	12.50			4.3					
annual_premium2	prim1	!	3.75	4.05	!	-0.66	12.1	3	20.7	20.7	-85.26	!
	prim2	!	4.30	2.82	!	0.14	0.4	22	0.6	0.6	14.12	!
	prim3	!	4.31	2.81	!	0.29	1.6	16	2.4	2.4	28.89	!
	prim4	!	4.30	2.82	!	0.51	5.2	6	7.7	7.7	51.34	!
annual_premium2			16.67	12.50			19.4					
acc_death_amount	NoRider	!	0.91	10.45	!	0.12	1.0	19	6.8	6.8	78.45	!
	Rider	!	4.64	2.05	!	-0.59	4.9	7	6.8	6.8	-47.91	!
acc_death_amount			5.56	12.50			5.9					
premium_frequency	Annual	!	4.25	2.93	!	-0.32	2.1	14	3.2	3.2	-33.00	!
	Free	!	4.71	1.90	!	-0.74	7.3	5	9.9	9.9	-57.78	!
	Monthly	!	3.59	4.42	!	0.28	2.5	10	4.4	4.4	39.45	!
	Quarterly	!	4.50	2.37	!	0.16	0.4	21	0.6	0.6	14.32	!
	Semi-annual	!	5.17	0.88	!	0.82	4.1	9	5.1	5.1	41.22	!
premium_frequency			22.22	12.50			16.4					
risk_state2	Prefered	!	4.81	1.67	!	0.80	7.5	4	10.0	10.0	57.86	!
	Standard	!	0.96	10.34	!	-0.16	1.8	15	12.2	12.2	-100.0	!
	SubStd	!	5.34	0.49	!	0.63	1.3	18	1.6	1.6	23.13	!
risk_state2			11.11	12.50			10.7					
underwriting_age	Middle	!	3.66	4.26	!	0.23	1.6	17	2.8	2.8	31.28	!
	Old	!	4.53	2.31	!	-0.51	4.2	8	5.9	5.9	-44.80	!
	Young	!	2.92	5.93	!	0.03	0.0	25	0.1	0.1	5.97	!
underwriting_age			11.11	12.50			5.8					
living_place	EastCoast	!	4.41	2.58	!	0.10	0.2	24	0.2	0.2	9.16	!
	Other	!	1.40	9.35	!	-0.00	0.0	26	0.0	0.0	-0.55	!
	WestCoast	!	5.30	0.58	!	-0.41	0.7	20	0.8	0.8	-16.55	!
living_place			11.11	12.50			0.9					

Tableau 12 annexe 1 ACM

Aides à l'interprétation pour les modalités actives

Variable	Modalités actives Modalité	! CONTR POIDS !		! AXE2 !						
		COORD	CTR	RCTR	CO2	QLT	V.TEST			
duration2	dur1	! 4.11	3.24 !	-0.05	0.1	25	0.1	31.8	-5.85 !	
	dur2	! 4.19	3.08 !	0.01	0.0	26	0.0	0.6	0.87 !	
	dur3	! 4.19	3.08 !	0.54	6.8	7	9.4	12.9	56.74 !	
	dur4	! 4.18	3.09 !	-0.49	5.6	9	7.7	28.8	-51.60 !	
duration2		16.67	12.50		12.5					
gender	Female	! 2.78	6.25 !	0.39	7.4	5	15.4	20.3	76.07 !	
	Male	! 2.78	6.25 !	-0.39	7.4	6	15.4	20.3	-76.09 !	
gender		5.56	12.50		14.8					
annual_premium2	prim1	! 3.75	4.05 !	0.15	0.7	17	1.1	21.8	19.37 !	
	prim2	! 4.30	2.82 !	0.35	2.6	11	3.5	4.1	34.75 !	
	prim3	! 4.31	2.81 !	-0.72	11.3	3	15.1	17.5	-71.85 !	
	prim4	! 4.30	2.82 !	0.16	0.5	19	0.7	8.4	15.58 !	
annual_premium2		16.67	12.50		15.1					
acc_death_amount	NoRider	! 0.91	10.45 !	-0.07	0.5	22	2.9	9.6	-50.89 !	
	Rider	! 4.64	2.05 !	0.38	2.3	13	2.9	9.6	31.08 !	
acc_death_amount		5.56	12.50		2.8					
premium_frequency	Annual	! 4.25	2.93 !	-0.22	1.1	16	1.5	4.7	-22.90 !	
	Free	! 4.71	1.90 !	-0.12	0.2	23	0.3	10.2	-9.50 !	
	Monthly	! 3.59	4.42 !	0.06	0.1	24	0.2	4.6	7.99 !	
	Quarterly	! 4.50	2.37 !	0.37	2.5	12	3.2	3.8	32.78 !	
	Semi-annual	! 5.17	0.88 !	-0.27	0.5	21	0.6	5.7	-13.65 !	
premium_frequency		22.22	12.50		4.4					
risk_state2	Prefered	! 4.81	1.67 !	0.70	6.3	8	7.5	17.5	50.39 !	
	Standard	! 0.96	10.34 !	-0.13	1.4	15	8.3	20.4	-82.30 !	
	SubStd	! 5.34	0.49 !	0.39	0.6	18	0.6	2.2	14.23 !	
risk_state2		11.11	12.50		8.2					
underwriting_age	Middle	! 3.66	4.26 !	-0.59	11.5	2	18.2	21.0	-80.06 !	
	Old	! 4.53	2.31 !	-0.17	0.5	20	0.7	6.6	-14.94 !	
	Young	! 2.92	5.93 !	0.49	11.1	4	21.9	22.0	90.18 !	
underwriting_age		11.11	12.50		23.1					
living_place	EastCoast	! 4.41	2.58 !	0.85	14.4	1	18.9	19.1	80.24 !	
	Other	! 1.40	9.35 !	-0.19	2.7	10	11.2	11.2	-76.49 !	
	WestCoast	! 5.30	0.58 !	-0.67	2.0	14	2.2	3.0	-26.72 !	
living_place		11.11	12.50		19.1					

Tableau 13 annexe 2 ACM

## VI.2 Exemples de fonctions de hasard empiriques

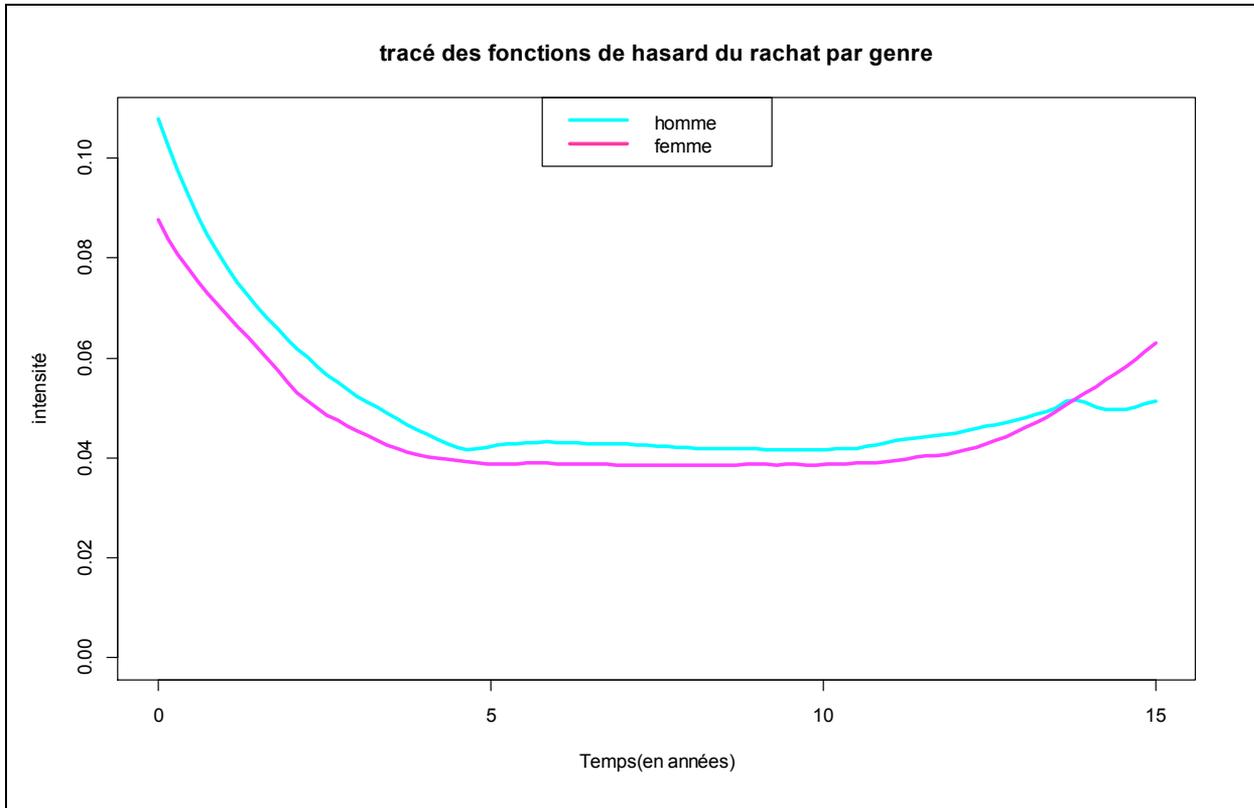
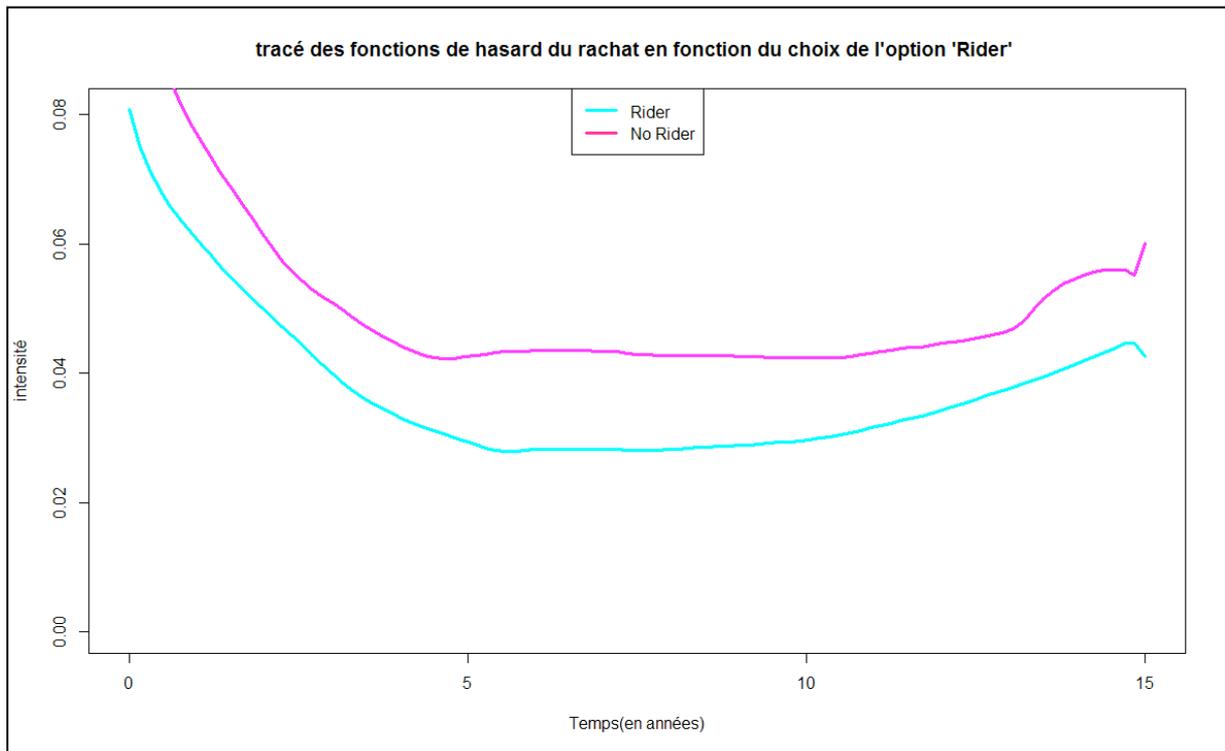


Figure 46 fonction de hasard empirique pour la variable genre

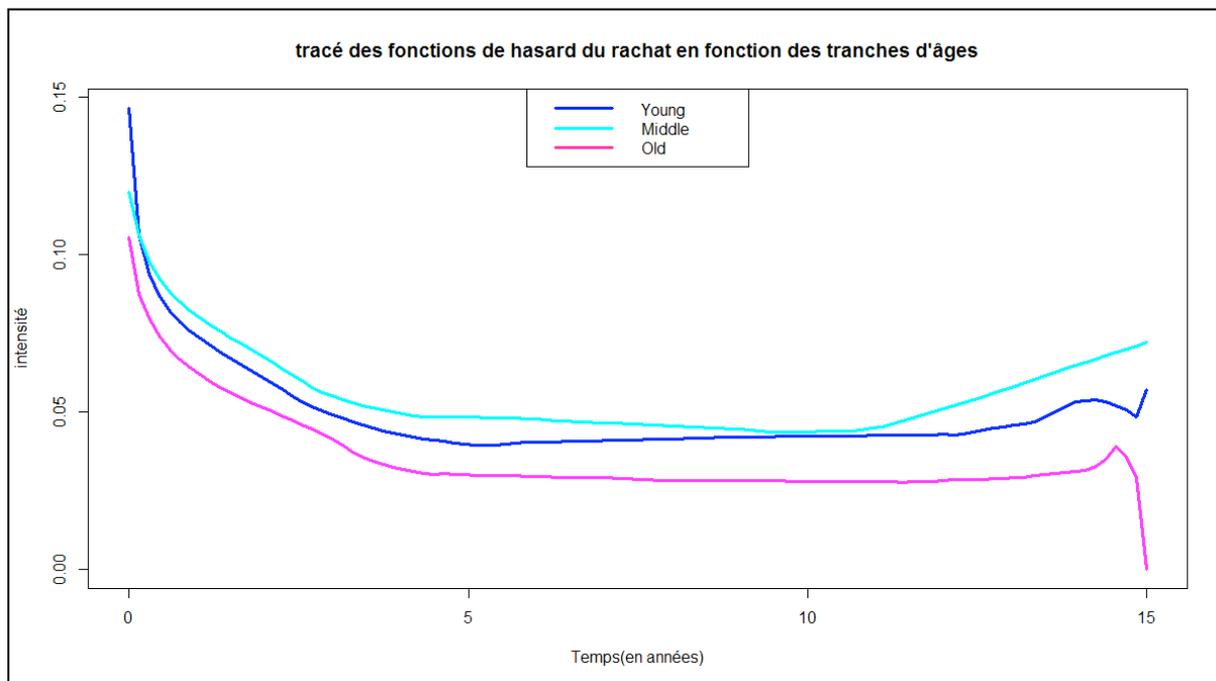
De ces tracés de fonctions de hasard empiriques, nous remarquons que les femmes ont durant quasiment les 15 ans de l'étude une probabilité instantanée de rachat inférieure à celle des hommes.

Cette tendance est inversée durant les deux dernières années de contrats. Notons que plus on avance dans le temps, moins l'estimateur empirique de la fonction de hasard est robuste. Ceci est dû au fait qu'il y a de moins en moins de données pour effectuer l'estimation. On a ainsi que 58 contrats dont la durée est supérieure ou égale à 15 ans (souscrits en 1995 et en vigueur en juillet 2010).



**Figure 47 fonction de hasard empirique pour la variable option Rider**

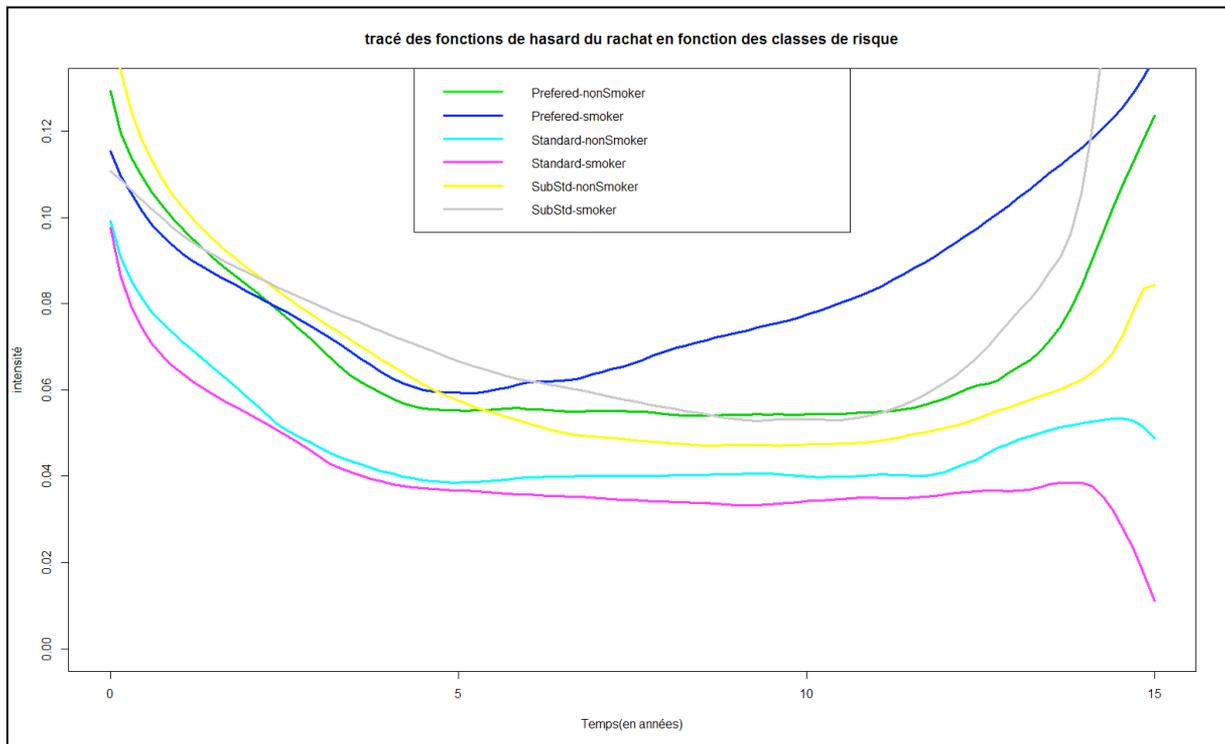
De ces tracés nous remarquons que d'une façon générale, les taux de hasard instantanés sont très forts au début du contrat, puis décroissent progressivement entre la 1ere et la 5eme année. Ce taux de hasard est quasiment constant entre la 5eme et la 10eme année. A partir de la 10eme année, il croit. Dès le début les clients ayant souscrits à l'option Rider (permettant d'augmenter le montant de l'indemnité en cas de décès accidentel) ont un taux de rachat instantané beaucoup plus faible, et ce durant toute la durée du contrat. Cet écart est beaucoup plus important entre la 5eme et la 10eme année du contrat, puis celui-ci tend à s'amoinrir entre la 10eme et la 12eme année à peu près (il y a un rapprochement des taux de rachat instantanés des deux catégories d'assurés). A partir de la 12eme année à peu près, cet écart s'agrandit progressivement.



**Figure 48 fonction de hasard empirique pour la variable underwriting age**

Nous remarquons de ce tracé de fonction de hasard que les personnes âgées (dont l'âge est supérieur à 55 ans), ont durant les quinze premières années, un taux de hasard instantané du rachat inférieur aux deux autres classes d'âges ( young et middle). Ceci est assez logique, car ceux-ci sont très souvent à la retraite, et ont en général moins de projets que les plus jeunes (achat d'un logement, d'un véhicule, etc.). Ainsi ce taux est élevé au début, mais reste toujours inférieur à celui des deux autres catégories, puis il décroît jusqu'à la 5eme année et après il est quasiment constant.

Le taux instantané de rachat des très jeunes (dont l'âge est inférieur à 35 ans) est quant à lui supérieur à celui des personnes âgées, et inférieur à celui des personnes un plus plus âgées (d'âge middle), sauf au tout début du contrat. Ainsi un client très jeune rachète probablement plus qu'une personne âgée et moins qu'une personne d'âge intermédiaire. En effet celui-ci a en général beaucoup plus de projets qu'une personne âgée (achat d'une voiture), et un peu moins qu'une personne d'âge intermédiaire (qui aura tendance à acheter un logement et voudra financer les études de son enfant, etc.).



**Figure 49 Figure 47 fonction de hasard empirique pour la variable risk state**

De ces tracés de fonctions de hasard nous remarquons que la classe standard (fumeur ou non) se démarque des deux autres classes. Elle est largement en dessous de celles-ci. Ainsi on peut déduire que les individus de cette classe ont moins tendance à racheter que les autres.

## VI.3 Prise en compte de la covariable dépendant du temps et de l'approche Fine & Gray dans la programmation sous R

### VI.3.1 Constitution initiale de la base de données

Pour chaque assuré :

- « issue.date »: date de souscription
- « end. date » : date de clôture du contrat (si clôture), si le contrat est censuré elle n'est pas renseignée
- les variables caractéristiques présentées dans le (III.1).
- « event » : variable indiquant l'état du contrat à la date de fin du portefeuille (Juillet 2010)
  - event=1 contrat toujours en vigueur en Juillet 2010
  - event=2 contrat ayant été racheté
  - event=3 contrat ayant été clôturé pour cause de décès de l'assuré (risque compétitif)
  - event=4 contrat ayant été clôturé pour causes diverses (risque compétitif)

### VI.3.2 Introduction de la performance de la covariable traduisant le contexte économique

Nous procédons en deux étapes pour adapter nos données à la prise en compte de la variable dépendant du temps.

- Dans un premier temps, nous discrétisons chaque contrat en intervalle trimestriel.

Sur chaque intervalle trimestriel :

1. Nous dupliquons les covariables caractéristiques de l'assuré (sexe, classe de risques etc .) car leurs valeurs ne varient pas dans le temps .

2. Nous calculons l'ancienneté du contrat aux bornes inférieures et supérieures de l'intervalle de temps.

Nous ajoutons ainsi deux nouvelles variables à notre base, nommées time1 et time2.

3. Nous créons une nouvelle variable status.time qui indique s'il s'agit d'un intervalle de temps intermédiaire du contrat ou s'il s'agit de l'intervalle final. La notion d'intervalle final désigne l'intervalle de temps sur lequel s'est produit un événement d'intérêt : rachat, risque compétitif ou censure administrative.

status.time=0 pour les intervalles de temps intermédiaires

status.time=1 pour l'intervalle final.

4. Le sens de la variable event évolue : elle correspond maintenant à l'état du contrat à la fin de l'intervalle considéré. Ainsi, pour un même assuré, la variable event prend des valeurs différentes selon l'intervalle de temps considéré.

- Sur les intervalles intermédiaires : event =1
- Sur l'intervalle final : event est défini comme au C.1

```
##NOTATIONS :
#   adj.beggin.dates[[i]] : date de souscription de l'assuré n° i
#   adj.end.dates[[i]] : date de fin du contrat = date de rachat (si rachat) OU = "2010-07-01" si risques compétitifs ou censure administrative

## ETAPE 1 : Pour chaque assuré, on génère la séquence de date entre sa date de souscription et sa date de fin de contrat pas trimestriel
for (i in 1:n.policyholders) {

  t1<-0
  dates.sequence[[i]] <- seq.Date(from = adj.beggin.dates[[i]],to=adj.end.dates[[i]], by = "3 months")

  # ETAPE 2 : on crée une nouvelle ligne par intervalle
  k=length(dates.sequence[[i]])

  if (k > 2) {
    data22<-data2[rep(i,(k-2)),] # on duplique les variables fixes de l'assuré sur chaque ligne
    data3[i,3]<-dates.sequence[[i]][2]
    data3[i,13]<-1 #on met event =1 pour la première ligne car k>2 donc contrat est en vigueur sur cette ligne
    data3[i,17]<-0 #status.time=0 pour la première ligne (sauf si k=2)

    for (j in 1:(k-2)){
      data22[j,2]<-dates.sequence[[i]][j+1]
      data22[j,3]<-dates.sequence[[i]][j+2] #pb : pour la dernière valeur on stocke la valeur arrondie
      data22[j,17]<-0 #car ligne intermédiaire
      data22[j,13]<-1 #on met status en Inforce
    }
    data22[k-2,3]<-data2[i,3]
    #valeur de status.time : dépend si un événement à eu lieu où non

    data22[k-2,17]<-1 # status.time=1 car dernier intervalle
    data22[j,13]<-data2[i,13] #on récupère la cause du competing risk

    # calcul des intervalles de durées
    data3[i,15]<-0
    data3[i,16]<-round(as.real((data3[i,3]-data3[i,2])/91),2)
  }
}
```

➤ Dans un second temps, nous calculons la performance du DowJones sur chaque intervalle.

Pour cela nous importons un fichier contenant les valeurs du DowJones à chaque date de début de trimestre entre Janvier 1995 et Juillet 2010.

```
eco.variables <- read.csv2("DJIA.csv",header=TRUE,colClasses=c("Date","numeric"))
eco.variables$VALUE<-as.numeric(eco.variables$VALUE)
```

Puis, pour chaque intervalle de temps, nous calculons la performance de l'indice de la façon suivante :

$$DJIA.tend[\text{intervalle } n^{\circ}i] = \frac{\text{Valeur de l'indice}[\text{date fin d'intervalle}] - \text{Valeur de l'indice}[\text{date début d'intervalle}]}{\text{Valeur de l'indice}[\text{fin d'intervalle}]}$$

```
#Calcul de la performance du DowJones sur chaque intervalle :|

for (i in 1:n.policyholders){

  d1<-as.numeric(eco.variables[which(eco.variables$DATE==adj.beggin.dates[[i]]),2]) #valeur de l'indice en début d'intervalle
  d2<-as.numeric(eco.variables[which(eco.variables$DATE==adj.end.dates[[i]]),2]) #valeur de l'indice en fin d'intervalle
  tend[i]<-round(as.numeric(((d2-d1)/d1),2)
  }

return(tend)
}

DJIAtend<-eco.variables.tendency(data_s$end.date,data_s$issue.date,eco.variables,delta,time.step,categ=T)
```

Nous obtenons alors la structure suivante :

	id.contract	issue.date	end.date	acc.death.amount	gender	premium.frequency	risk.state	underwriting.age	DJIAtend
1	39	1995-03-30	1995-04-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.11
2	39	1995-04-01	1995-07-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.07
3	39	1995-07-01	1995-10-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.05
4	39	1995-10-01	1996-01-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.1
5	39	1996-01-01	1996-04-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.04
6	39	1996-04-01	1996-07-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.01
7	39	1996-07-01	1996-10-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.1
8	39	1996-10-01	1997-01-01	NoRider	Male	Monthly	Standard-smoker	Middle	0.09

### VI.3.3 Prise en compte des risques compétitifs par l'approche Fine&Gray

Les objets de survie dans les modèles de durées correspondent aux variables aléatoires  $(X_i, C_i)$  où  $X_i$  représente pour l'individu  $i$  la durée avant l'occurrence de l'événement d'intérêt et  $C_i$  la censure.

Ces objets de survie sont créés sous R l'aide de la fonction Surv du package Survival. Dans un cadre standard où la durée est définie de façon continue, l'objet de survie est défini par :

Surv(duration,status) où status est une indicatrice qui vaut 1 ou 0 selon que l'observation soit censurée ou non.

Dans notre cas, les durées sont découpées en intervalles de temps ainsi l'objet de survie doit être défini sous la forme suivante :

Surv(time1,time2,status)

où time 1 (respectivement time 2) représente la durée du contrat en début (respectivement fin) d'intervalle  
status : variable indicatrice qui vaut 0 ou 1 selon que l'observation est censurée ou non.

La façon dont nous définissons la variable status est primordiale afin de bien prendre en compte les risques compétitifs. En effet, il faut veiller à ne pas les inclure dans la censure.

De plus, nous avons établi dans le 4.2 (« modélisation par approche Fine&Gray ») que pour étudier la subdistribution, il faut modifier l'ensemble à risque. En effet, pour les risques compétitifs il faut faire comme s'ils étaient encore susceptibles de racheter leur contrat. Il faut alors les conserver dans l'ensemble à risque et remplacer leur durée effective de vie de contrat par une durée fictive correspondant au cas où il n'aurait pas subi de risques compétitifs et seraient restés dans le portefeuille jusqu'à la date de censure administrative.

Pour définir un objet de survie conforme à notre approche théorique, nous avons apporté les modifications suivantes à notre base :

1. Nous avons créé une variable outcome qui résume les informations des variables

status.time et event :  $outcome = event \times status.time$

Pour chaque intervalle, elle peut prendre les valeurs suivantes :

- outcome = 0 : il ne se produit pas d'événement sur l'intervalle considéré (c'est un intervalle dit intermédiaire)
- outcome = 1 l'intervalle est censuré administrativement (i.e sa borne supérieure est Juillet 2010). Il s'agit de l'intervalle final.
- outcome = 2 un rachat a eu lieu sur l'intervalle.
- outcome = 3 ou 4 un risque compétitif à eu lieu sur l'intervalle.

2. Nous avons créé une variable cens.time. Cette variable est fixe pour chaque individu car

elle correspond à la durée entre la date de souscription de l'individu et la date de censure administrative. Cette variable représente la durée de vie du contrat en date de censure administrative.

Pour les contrats toujours en vigueur en Juillet 2010, les variables time 2 et cens.time coïncident sur l'intervalle final.

Pour les contrats ayant subi un risque compétitif, la variable cens.time viendra remplacer la variable time2 sur l'intervalle final. Ceci traduit le fait que l'on remplace pour ces contrats la durée réelle du contrat par la durée fictive du cas où ils seraient toujours dans l'ensemble à risque.

Pour les contrats rachetés, cette variable n'est pas utilisée.

Cette variable sera utilisée uniquement pour les contrats ayant subi un risque compétitif car elle remplace alors la durée time 2 dans l'intervalle final.

```
data_s1$outcome<-with(data_s1,event*status.time)
data_s1$sh.time<-with(data_s1,ifelse((data_s1$outcome==3)|(data_s1$outcome==4),data_s1$cens.time,data_s1$time2))
```

On définit alors la variable sh.time qui prend en compte les considérations précédentes.

## VI.4 Test de proportionnalité des covariables du modèle de Cox approche Fine& Gray

- Graphique  $\beta_{gender}(t)$  en fonction du temps

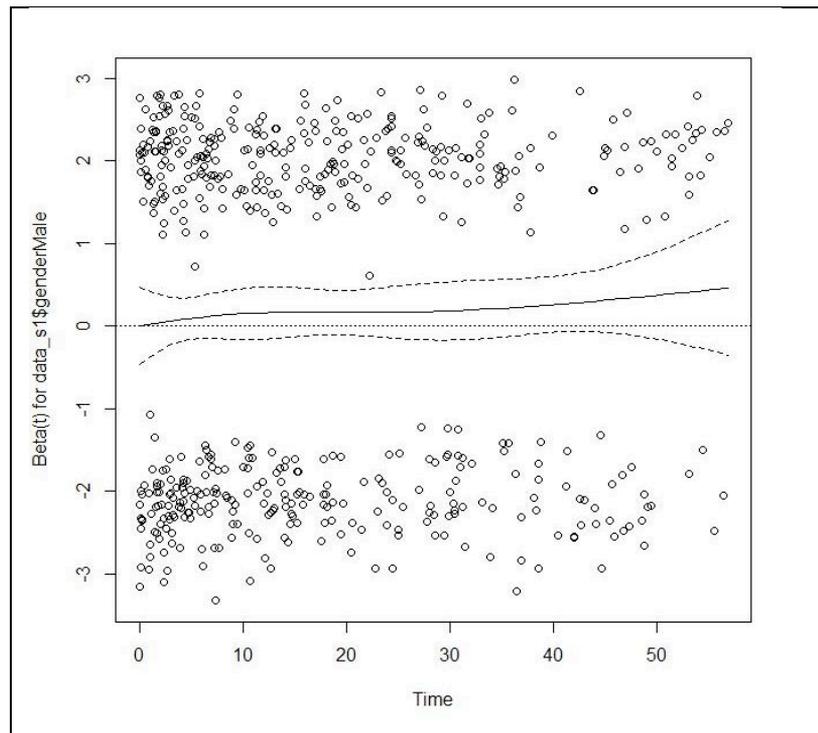


Figure 50 test de la dépendance du coefficient de la variable genre dans le temps

➤ Graphique  $\beta_{DJIA}(t)$  en fonction du temps

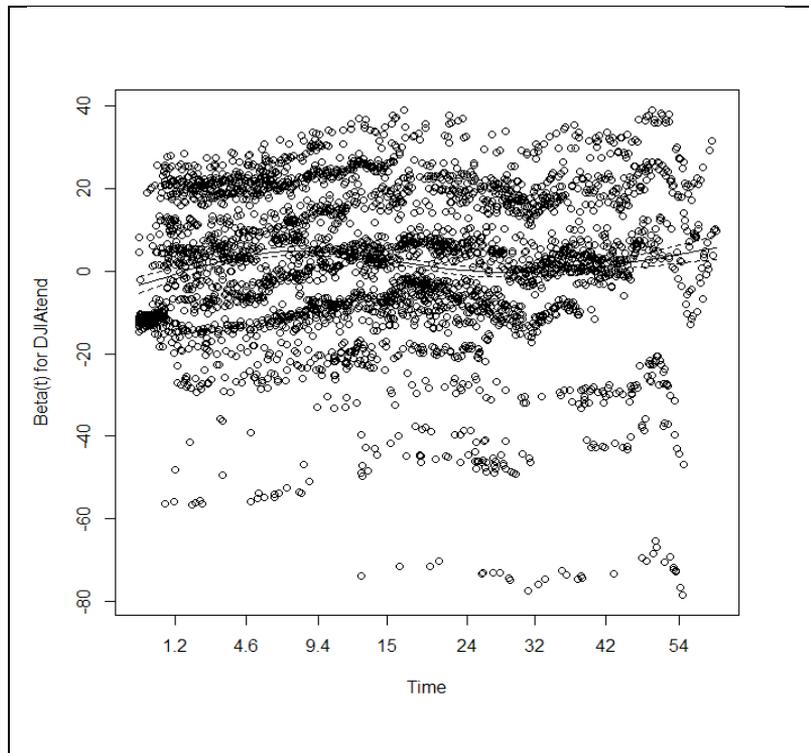


Figure 51 test de la dépendance du coefficient de l'indice DowJones dans le temps

➤ Graphique  $\beta_{frequency\ quarterly}(t)$  en fonction du temps

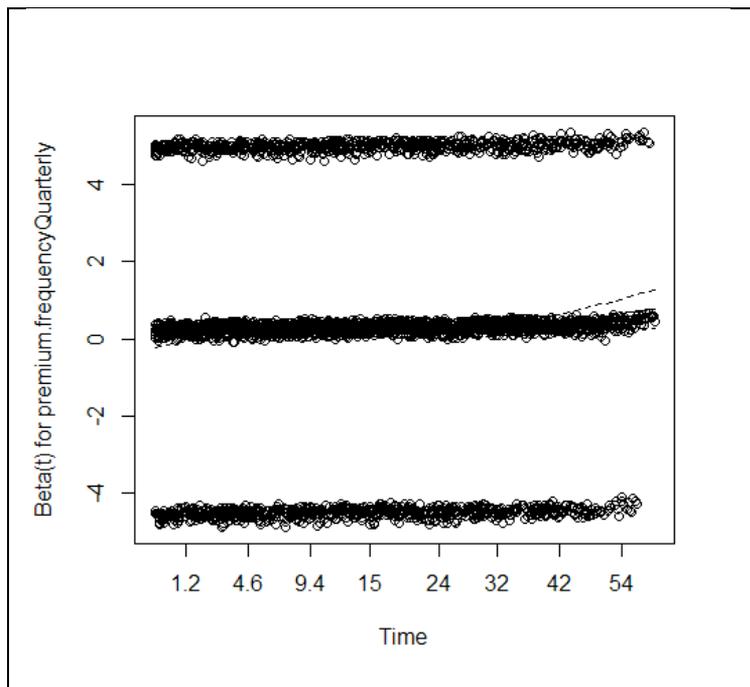


Figure 52 test de la dépendance du coefficient de la modalité "quarterly" dans le temps

➤ Graphique  $\beta_{frequency\ semi-annual}(t)$  en fonction du temps

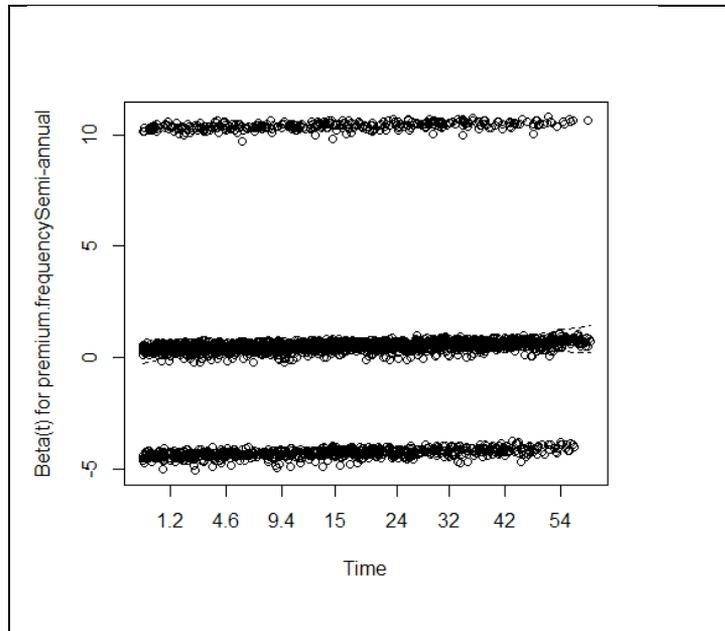


Figure 53 test de la dépendance du coefficient de la modalité "semi-annual" dans le temps

➤ Graphique  $\beta_{underwriting\ age\ middle}(t)$  en fonction du temps

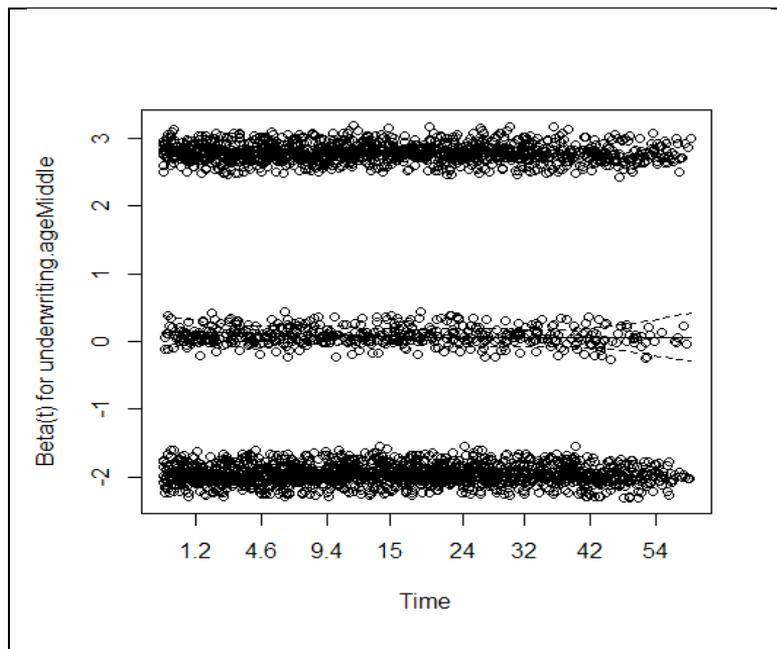


Figure 54 test de la dépendance du coefficient de la modalité "middle" dans le temps

## VII. Bibliographie

- Changki Kim. (2005). *Modeling surrender and lapses rates with economic variables*.
- Kim Hin Ho et Huiyong Su. (2006). *Structural prepayment risk behavior of the underlying mortgages for residential mortgage life insurance in a developing market*.
- S. Haberman et A.Renshaw. (1986). *statistical analysis of life insurance lapses*.
- Weixi Shen et Xu Huiping. (2004). *The valuation of unit-linked policies with or without surrender options*.
- David G. Kleinbaum et Mitchel Klein. (2005). *Survival Analysis*.
- Denis Fougère. (2007). *Modèles de durée*.
- Djaidja, J. B. (2006). *Credit risk modeling based on survival analysis with immunes*.
- Djaidja, Jan Beran et Abdel Yazid Karim. (2006). *Credit risk modeling based on survival analysis with immunes*.
- Frédéric Planchet et Pierre Therond. (2006). *Modèle De Durée - Applications Actuarielles*.
- Jan Beyersmann, Arthur Allignol et Martin Schumacher. (2012). *Competing Risks and Multistate Model with R*.
- Jason P. Fine et Robert J.Gray. (1999). *A proportionnal hazard model for the subdistribution of a competing risk*.
- Jean David Fermanian. (2009). *Modèles de durées*.
- Melania Pintilie. (2006). *Competing risks a practical perspective*.
- Philippe Saint Pierre. (2012). *Introduction à l'analyse des durées de survie*.
- Stéphane Fauvel et Maryse le Pévédic. (2007). *Analyse des rachats d'un portefeuille vie individuelle: approche théorique et application pratique*.
- Therry M. Therneau et Patricia M.Grambsch. (2000). *Modeling Survival Data - Extending the Cox Model*.
- Torben Martinussen et Thomas H. Scheike . (2006). *Dynamic Regression Models for Survival Data*.
- Xavier Milhaud. (2011). *Segmentation et modélisation des comportements de rachats en Assurance-vie*.