



**HAL**  
open science

## **Facteurs explicatifs du rachat en Assurance-Vie : classification et prévisions du risque de rachat**

Xavier Milhaud, Stéphane Loisel, Véronique Maume-Deschamps

### ► **To cite this version:**

Xavier Milhaud, Stéphane Loisel, Véronique Maume-Deschamps. Facteurs explicatifs du rachat en Assurance-Vie : classification et prévisions du risque de rachat. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494798

**HAL Id: inria-00494798**

**<https://inria.hal.science/inria-00494798>**

Submitted on 24 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# FACTEURS EXPLICATIFS DU RACHAT EN ASSURANCE-VIE : CLASSIFICATION ET PRÉVISIONS DE RISQUE DE RACHAT

*Xavier Milhaud*

*Adresse : 5 Passage D'enfer, 75014 - Paris*

*Rattaché à AXA Cessions et Université Lyon 1, ISFA, Laboratoire SAF*

*Stéphane Loisel*

*Affiliation : Université de Lyon, UCBL Lyon 1, ISFA, Laboratoire SAF*

*Véronique Maume-Deschamps*

*Affiliation : Université de Lyon, UCBL Lyon 1, ISFA, Laboratoire SAF*

## Résumé en français

La compréhension de la dynamique des taux de rachat est un point crucial pour une compagnie d'assurance. En effet, elle intervient à différents niveaux, lors du design de nouveaux produits ou dans le calcul de l'European Embedded Value (EEV) par exemple. L'assureur est amené à faire des hypothèses sur les taux de rachat par produit, ce qui peut avoir un impact majeur si ces hypothèses ne sont pas vérifiées. Ainsi, l'assureur est exposé à de nombreux risques parmi lesquels le risque d'anti-sélection et de hasard moral, ou encore le risque de taux lorsqu'il doit emprunter de l'argent pour rembourser la valeur de rachat à l'assuré. Le risque d'anti-sélection ([1]) est important dans la mesure où l'assureur peut finalement n'avoir que des "mauvais" risques en portefeuille : c'est le cas notamment si les "bons" risques rachètent leur contrat au même moment.

Les premières recherches académiques sur les taux de rachat remontent aux années 1990 ([2]) et deux perspectives y sont développées : soit l'assuré a un besoin urgent d'argent ce qui le pousse à racheter son contrat (de manière générale, une récession économique ferait augmenter les taux de rachat), soit l'assuré a un raisonnement rationnel et rachète son contrat si les taux du marché montent (il pourra ainsi profiter de l'opportunité de souscrire un nouveau contrat ayant les mêmes garanties à un moindre coût). [3] propose de séparer la relation long-terme des taux de rachat, taux d'intérêt et taux de chômage et court-terme (dû à des mécanismes d'ajustement) grâce à une technique de co-intégration. Dans [4], les rachats sont étudiés dans le contexte de la crise économique de Corée en 2001 et on note l'importance du contexte économique et financier dans les décisions des assurés ; d'autres modèles comme le modèle binomial négatif et des processus de comptage ont aussi été testés par [5] et [6].

En Assurance Vie, certaines caractéristiques de contrats jouent un rôle majeur dans la décision de l'assuré de racheter son contrat. Les conditions de souscription, l'âge, la profession ainsi que d'autres facteurs propres à la situation de l'assuré lors de la souscription influencent ses décisions. Il est également possible que l'environnement économique et financier soit important. Deux modèles de segmentation nous ont permis de développer

ces idées : les arbres de classification et de régression ([7], [8]), et la régression logistique ([9]). Les contrats de type Prévoyance ainsi que ceux d'Épargne sont impactés, et les résultats montrent clairement que la garantie de participation au bénéfice est très discriminante. Nous nous focalisons dans cette étude sur des produits de type Mixte. Nous présentons dans un premier temps les fondamentaux de chacun des modèles ainsi que leurs hypothèses et limites. Nous nous concentrons principalement sur la théorie des arbres de classification, bien moins connue que celle liée à la régression logistique. La notion de gain d'homogénéité et d'instabilité de l'algorithme CART y sont particulièrement développés, ainsi que la méthode qui permet de palier à ce problème d'instabilité : les forêts aléatoires. Puis nous testons différents facteurs comme possibles déclencheurs de la décision de rachat dans le but de segmenter le portefeuille en classe de risque : nous disposons de l'âge de l'assuré à la souscription, la durée du contrat, la fréquence de la prime, la prime de risque, la prime d'épargne, le sexe, la clause de participation au bénéfice et un indicateur de la richesse de l'assuré. La durée du contrat et l'option de participation au bénéfice apparaissent comme des éléments essentiels parmi les facteurs explicatifs de la décision de rachat. L'intérêt de l'utilisation de ces deux méthodes réside dans leur complémentarité : l'algorithme CART, méthode non-paramétrique, permet de mettre en valeur l'importance des variables en les classant dans un ordre décroissant. Le résultat de cette procédure peut être directement utilisé comme input dans la procédure de régression logistique (comme variables explicatives). Cela limite le nombre de paramètres à estimer dans la régression logistique tout en conservant l'essentiel de l'information contenue dans les données et nécessaire à la segmentation. Toutefois et à des fins de comparaison, ce n'est pas ce raisonnement qui est suivi et les deux modèles sont présentés en concurrence d'un point de vue de l'erreur de classification (mêmes variables explicatives dans le modèle non-paramétrique et le modèle paramétrique en input). Les résultats d'erreur sont quasiment similaires mais chaque méthode apporte une information différente : CART rend mieux compte de l'importance de chaque facteur tandis que la régression logistique donne un résultat plus facile à appréhender (une probabilité de rachat) et fournit un outil de comparaison numérique (les odd-ratios) lorsque les valeurs de variables explicatives diffèrent entre deux assurés.

En dernière partie, nous discutons des différences entre les deux modélisations en termes de résultats numériques et d'un point de vue opérationnel. Enfin nous proposons des pistes d'amélioration de ces modèles, notamment l'utilisation d'algorithme de *reéchantillonnage* pour équilibrer les proportions de la variable binaire réponse (rachat) et stabiliser les résultats, mais aussi la prise en compte d'un côté fonctionnel de l'étude par la dépendance importante au contexte économique.

**Mots-clés :** classification, segmentation, CART, régression logistique, rachat.

## Abstract

Understanding the dynamics of the surrender (lapse) rates is a crucial point for insurance companies. For example, the calculation of the European Embedded Value (EEV) or the design of a new product is partially based on assumptions on the surrender rates. The insurer can be faced to many problems such as : first, policy lapse might make her unable to fully recover her initial expenses (cost of procuring, underwriting, and issuing new business). Second, policyholders who have adverse health or other insurability problems tend not to lapse their policies, causing the insurer to experience more claims than expected if the lapse rate is high : this is the so-called *moral hazard* and *adverse selection* where there remain only “bad risks” ([1]). Third, massive early surrender poses a liquidity threat to the insurer who is subjected to interest rate risk (the interest rate is likely to change over the period of the contract).

What causes lapses has attracted certain academic interest since 1980’s : originally two main hypotheses have been suggested to explain lapse behavior. The emergency fund hypothesis ([2]) contends that policyholders use cash surrender value as emergency fund when facing personal financial distress (so the lapse rate increases during economic recessions). On the other hand, the interest rate hypothesis conjectures that the surrender rate rises when the market interest rate increases, because the investor acts as the opportunity cost for owning insurance contracts. [3] suggested to separate the potential long-term relationship between the lapse rate, interest rate and unemployment rate from their short-term adjustment mechanisms thanks to the cointegrated vector autoregression approach. In the literature, [5] and [6] applied respectively the negative binomial and the zero-inflated models as counting processes, and [4] applied the logistic regression model with economic variables to explain the lapses on insurance policies during the economic crisis in Korea.

This study shows that some policy features and policyholder’s characteristics are crucial to explain the decision of the policyholder to surrender her contract. We point it out by applying two segmentation models to a life insurance portfolio : the Logistic Regression model ([9]) and the Classification And Regression Trees model ([7], [8]). We run the study with Endowment products but Protection as well as Savings lines of business are impacted. Results clearly explicit that the profit benefit option is highly discriminant. First we present the models and discuss their assumptions and limits. We mainly focus on the classification tree theory which is not as famous as the logistic regression one. The notions of purity and instability of the CART algorithm are addressed as well as the way to overcome this instability by the use of Random Forests. Then we test different policy features and policyholder’s characteristics to be lapse triggers in order to segment the portfolio in risk classes regarding the surrender choice : the premium frequency, the underwriting age, the duration of the contract, the risk premium, the saving premium, the gender, the profit benefit option and the face amount. The duration and the profit benefit option are essential. The complementarity of these two models is one of the key-points

developed in this paper : the non-parametric CART yields to the importance of each explanatory variable and allow us to select salient features of the data, which can then be input in the logistic regression model to limit the number of parameters to estimate and improve the results. In this paper, we do not follow this thought process because our goal is just to compare the classification error of the two methods (so the input are the same ones in the two models) and the model performances. Errors are similar although the information given by each modeling is quite different : CART provides us with the importance of each explanatory variable whereas the logistic regression output is both the probability for the policyholder to surrender her contract (easier to understand) and the odd-ratios, a very useful tool to compare the risks when considering two different policyholders.

Finally, we explore the main differences of both models in terms of numerical and operational results and discuss about it. Moreover, some ideas like functional methods and re-sampling techniques are suggested to deal with the dynamic environment (economic context) and improve the effectiveness and the accuracy of the results.

**Keywords :** classification, segmentation, CART, logistic regression, surrender.

**Les deux mots-clés à choisir parmi la liste proposée sont :**  
Apprentissage et classification, Modèles pour les assurances et les finances

## Références

- [1] William F. Bluhm. Cumulative antiselection theory. *Transactions of Society of actuaries*, 34, 1982.
- [2] Jean François Outreville. Whole-life insurance lapse rates and the emergency fund hypothesis. *Insurance : Mathematics and Economics*, 9 :249–255, 1990.
- [3] R.F. Engle and C.W.J. Granger. Cointegration and error-correction : Representation, estimation and testing. *Econometrica*, (55) :251–276, 1987.
- [4] Changki Kim. Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal*, pages 56–70, 2005.
- [5] Yusho Kagraoka. Modeling insurance surrenders by the negative binomial model. Working Paper 2005, 2005.

- [6] D. C. Atkins and R. J. Gallop. Re-thinking how family researchers model infrequent outcomes : A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 2007.
- [7] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Chapman and Hall, 1984.
- [8] B. Ghattas. Aggregation d'arbres de classification. *Revue de statistique appliquee*, 2(48) :85–98, 2000.
- [9] Joseph M. Hilbe. *Logistic regression models*. Chapman and Hall, 2009.