Master Actuariat de Dauphine

Mémoire présenté devant l'Université Paris Dauphine pour l'obtention du diplôme du Master Actuariat et l'admission à l'Institut des Actuaires

le _____

Par : LIMBOURG Jonathan

Titre: Modélisation des rachats des produits épargne en environnement Solvabilité II

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

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

Membre présent du jury de l'Institut des Actuaires :

Signature :

Entreprise :

Nom : KPMG S.A.

Signature :

Directeur de mémoire en entreprise :

Nom : BONNIN François

Signature :

Membres présents du jury du Master Actuariat de Dauphine :

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

Signature du responsable entreprise :

Secrétariat :

Bibliothèque :

Signature du candidat :

Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 PARIS Cedex

Modélisation des rachats des produits épargne en environnement Solvabilité II

Jonathan Limbourg
Encadré par François Bonnin et Loïc Seitz

Mémoire d'Actuariat
Année scolaire 2015/2016

Remerciements

Je tiens tout d'abord à remercier François Bonnin et Loïc Seitz pour leur implication dans ce mémoire, leurs conseils et le professionnalisme qu'ils m'ont transmis.

Je souhaite remercier Médéric Borde de m'avoir accueilli dans la belle structure qu'est l'équipe Actuariat Assurance de KPMG et pour son implication dans l'obtention de données pour la mise en pratique de l'étude.

Je remercie également Salma Jamal pour ses conseils avisés tant sur l'exercice que le sujet. J'en retiens des échanges très intéressants et formateurs.

Je remercie l'équipe Actuariat Assurance de KPMG pour leur bonne humeur et leur soutien.

Merci à Alexandra pour son soutien, ses relectures minutieuses et sa compréhension.

Je dédie ce mémoire à ma famille et à ma grand-mère sans qui rien n'aurait été possible. Merci pour votre engagement et votre confiance.

Résumé

L'opération d'assurance repose fondamentalement sur la compréhension, la valorisation et la gestion des risques couverts. Si les risques de marché sont de plus en plus maîtrisés par les assureurs, les risques comportementaux ne le sont pas encore. Le risque de rachat s'inscrit en effet dans cette catégorie et pourtant il est l'un des risques techniques les plus importants en épargne. De plus, les passifs étant représentés en juste valeur sous Solvabilité 2, la modélisation fine de ce risque aura de multiples impacts quantitatifs sur les principaux indicateurs : BE, SCR, fonds propres économiques et ratio de solvabilité.

Dans un premier temps, l'étude porte sur la compréhension et la modélisation du risque de rachat à l'aide d'un modèle mélange logit/non-paramétrique couplé à une approche multi-états type séparateurs à vaste-marge. L'accent est porté sur la construction théorique de ce type de modèle afin d'en comprendre les enjeux.

Dans un second temps le point de vu adopté est l'explication de la variable rachat par une liste de variables explicatives qui peuvent avoir une incidence sur le comportement de l'assuré. On distingue donc trois catégories de variables : les variables assurés (âge, niveau d'étude, ancienneté, etc.), les variables assureurs (taux servis, taux de rendement comptable) et les variables environnementales (financières, etc.).

Dans un troisième et dernier temps une mesure d'impact sur le BE de la modélisation proposée est réalisée. L'application sera faite sur le portefeuille d'un assureur français. L'analyse des résultats se fera par comparaison avec la modélisation actuelle de l'assureur. Tous les résultats seront réalisés et analysés à l'aide de l'outil de production Solvabilité II interne DEvent.

Mots-clés : Assurance-vie, épargne, Solvabilité II, rachats, régression logistique, régression non paramétrique par projection, SVM.

Abstract

Insurance operations rely mostly on different points such as comprehension, appreciation and risk management. If market risks are well handled by insurance providers on the other hand behavioral risk still remain hard to manage. The lapse risk enters in this category and yet it is one of the most important technical risk in savings. Moreover, liabilities are represented in fair-value under Solvency 2, the thorough modeling of this risk will have multiple quantitative impacts on main indicators such as : BE, SCR, economic own funds and solvency ratio.

In a first part, the study develops mainly the comprehension and the modeling of the lapse risk thanks to a logistic/non-parametrical model gathered with a multi-state such as support vector machine. The emphasis relies mainly on the theoretical construction of this type of model in order to understand its stakes.

In a second part, the point of view chosen is to explain the lapse variable with a list of explicative variables that may have an incidence on the behavior of the insured party. We can see three different categories of variables : policyholder's variables (age, level of education, seniority, etc.), insurer's variables (credited rate, accounting rate of return) and environmental variables (financial, etc.).

In a last and third part, it is important to measure the impact of the modeling which has been made and suggested on the BE. The application will be made on a French insurer portfolio. The analysis of these results will be made with a comparison with the actual modeling of the insurer. All the results will be made and analysed thanks to an internal Sovability 2 tool, DEvent.

Mots-clés : Savings, with-profit, Solvency II, lapse, logistic regression, nonparametric estimation by projection, SVM.

Table des matières

Introduction **1**

I Mise en contexte **3**

1 Assurance vie **4**
1.1 Présentation générale . . . 4
1.2 Produits d'épargne . . . 5
1.2.1 Produits sur fonds euro . . . 5
1.2.2 Produits en UC . . . 6

2 Solvabilité II **7**
2.1 Réglementation . . . 7
2.2 Bilan économique . . . 7
2.2.1 Valeur de marché des actifs . . . 7
2.2.2 Best Estimate . . . 7
2.2.3 Fonds propres . . . 10
2.3 Solvency Capital Requirement . . . 10
2.3.1 Une métrique de risque . . . 10
2.3.2 Module marché . . . 11
2.3.3 Module de souscription vie . . . 14
2.4 Marge de risque . . . 16
2.5 Bilan schématique et ratio de solvabilité . . . 17

3 Risque de rachats **18**
3.1 Quid ? . . . 18
3.1.1 Définition . . . 18
3.1.2 Risque et enjeux . . . 18
3.1.3 Rachats structurels et conjoncturels . . . 19
3.2 QIS 5 . . . 20
3.2.1 Définition et rôle des QIS et ONC . . . 20
3.2.2 Modélisation des rachats structurels . . . 20
3.2.3 Modélisation des rachats conjoncturels . . . 20
3.3 Impacts du choix de la loi . . . 22
3.3.1 Bilan économique . . . 22
3.3.2 SCR cessation . . . 22
3.3.3 SCR marché . . . 22
3.3.4 SCR global . . . 22
3.4 Problématique . . . 23

II Cadre de l'analyse **25**

4 Régression logistique **26**
4.1 Modèle linéaire . . . 26
4.2 Modèles linéaires généralisés . . . 28
4.3 Modèle logit . . . 29
4.4 Calibrage . . . 31
4.5 Pertinence dans le cadre du rachat . . . 32
4.6 Limites a priori . . . 33

5 Régression non paramétrique par projection **34**
5.1 Apports du non paramétrique . . . 34
5.2 Présentation de la régression par projection . . . 36
5.3 Calibrage . . . 37
5.3.1 Modèle approché . . . 37
5.3.2 Modèle général . . . 40
5.4 Intégration au modèle logit . . . 45
5.5 Limites a priori . . . 46

6 Modèle à états et SVM **47**
6.1 Besoin d'un modèle à état . . . 47
6.2 Présentation des SVM . . . 48
6.3 Etude théorique . . . 49
6.3.1 Cas linéaire séparable . . . 49
6.3.2 Cas linéaire non séparable . . . 53
6.3.3 Cas non linéaire . . . 55
6.4 Calibrage par algorithme SMO . . . 57
6.4.1 Heuristique de choix de deux multiplicateurs . . . 58
6.4.2 Optimisation pour deux multiplicateurs . . . 59
6.4.3 Mise à jour du SVM . . . 60
6.5 Application dans le cadre des rachats . . . 61
6.6 Limites a priori . . . 61

7 Outil ALM DEvent **62**
7.1 Projection de flux de trésorerie . . . 62
7.1.1 Outil de $3^{\text{ème}}$ génération . . . 62
7.1.2 Actif . . . 63
7.1.3 Passif . . . 65
7.1.4 Interactions Actif-Passif . . . 66
7.2 Calcul du BE . . . 66
7.3 Calcul du SCR . . . 67

8 Utilisation du GSE **68**

III Mise en œuvre de l'analyse **70**

9 Etude de données **71**
9.1 Caractéristiques de l'échantillon . . . 71
9.1.1 Données de rachats . . . 71

9.1.2 Données économiques . . . 72
9.2 Classification des données manquantes . . . 77
9.3 Algorithme des k plus proches voisins . . . 78

10 Application de la régression logistique 79
10.1 Résultats . . . 79
10.2 Qualité du modèle . . . 80
10.2.1 Taux d'erreur et Cross-Validation . . . 80
10.2.2 Courbe ROC . . . 81

11 Mise en place de la proposition de modèle hybride 84
11.1 Difficultés pratiques . . . 84
11.2 Résultats . . . 85
11.3 Taux d'erreur et Cross-Validation . . . 85
11.4 Courbe ROC et coefficient AUC . . . 86

12 Mise en place du modèle à états 87
12.1 Calibrage du SVM . . . 87
12.2 Intégration au modèle hybride . . . 89
12.2.1 Taux d'erreur et Cross-Validation . . . 89
12.2.2 Courbe ROC et coefficient AUC . . . 90

IV Etude d'impacts 92

13 Résultats de la modélisation type QIS 5 93
13.1 Best Estimate . . . 93

14 Impacts sur le Best Estimate 94
14.1 Modèle logit . . . 94
14.2 Modèle hybride . . . 96
14.3 Modèle à états . . . 97
14.4 Synthèse . . . 98

Conclusion 99

Glossaire 102

Bibliographie 104

A Etude de la vraisemblance du modèle logit 106
A.1 Calcul de la Hessienne . . . 106
A.2 Démonstration du caractère convexe . . . 106

B Démonstrations relatives au modèle non-paramétrique 108
B.1 Démonstration du caractère projecteur orthogonal . . . 108
B.2 Majoration de la variance de l'estimateur des moments . . . 109
B.3 Calcul du biais de l'estimateur de la variance . . . 110

Table des figures

1.1 Répartition des montants de l'épargne des ménages français 4
1.2 Répartition en parts de marché 5
1.3 Fiscalité des contrats d'assurance vie en France 6

2.1 Décomposition des flux du Best Estimate déterministe 8
2.2 Schéma des flux du Best Estimate déterministe 8
2.3 Décomposition des flux du Best Estimate sur le scénario 1 9
2.4 Décomposition des flux du Best Estimate sur le scénario 2 9
2.5 Exemple d'ANR 9
2.6 Eclatement schématique du SCR 11
2.7 Décomposition des flux du Best Estimate sur le scénario 1 choqué 12
2.8 Décomposition des flux du Best Estimate sur le scénario 2 choqué 12
2.9 Exemple de SCR Taux 13
2.10 Bilan économique sous Solvabilité II 17

3.1 Paramètres des lois minimum et maximum des ONC 21
3.2 Graphique des lois minimum et maximum des ONC 21

4.1 Exemple de régression linéaire 26
4.2 Résumé des MLG et des fonctions de lien associées 29
4.3 Régression linéaire sur une variable binaire 29

5.1 Données et comportement du modèle logit en fonction de l'ancienneté 35
5.2 Exemple d'utilisation de la vitesse de convergence 45

6.1 Exemple de classification linéaire par SVM 48
6.2 Exemple de données non linéairement séparables 53
6.3 Exemple de classification non linéaire par SVM 55

7.1 Organisation des outils internes 62
7.2 Architecture de l'outil DEvent 63

9.1 Liste des variables de rachats observées 71
9.2 Evolution des indices CAC 40 et EUROSTOXX 50 72
9.3 Evolution de l'indice MSCI Real Estate 73
9.4 Evolution du taux actuariel de l'OAT 10 ans française 74
9.5 Evolution de l'indice CPI français 75
9.6 Evolution des variables sous forme d'indice et représentation des crises 76
9.7 Evolution des variables représentatives des crises 77

10.1 Taux d'erreur estimés pour le modèle logit 81
10.2 Courbe ROC estimée pour le modèle logit 83

11.1 Taux d'erreur estimés pour le modèle hybride 85
11.2 Courbe ROC estimée pour le modèle hybride 86

12.1 Taux d'erreur estimés pour le SVM linéaire 87
12.2 Tracé du nuage de points et de la frontière linéaire 88
12.3 Taux d'erreur estimés pour le SVM gaussien 88
12.4 Tracé du nuage de points et de la frontière gaussienne 89
12.5 Taux d'erreur estimés pour le modèle à états 90
12.6 Courbe ROC estimée pour le modèle à états 90

13.1 BE calculé avec la modélisation type QIS 5 93
13.2 Graphique des flux de BE - Modélisation QIS 5 93

14.1 BE calculé avec le modèle logit 94
14.2 Graphique des flux de BE - Modèle logit 95
14.3 BE calculé avec le modèle hybride 96
14.4 Graphique des flux de BE - Modèle hybride 96
14.5 BE calculé avec le modèle à états 97
14.6 Graphique des flux de BE - Modèle à états 97
14.7 Graphique des flux de BE des différents modèles 98

Introduction

L'assurance vie reste le placement préféré des français à fin 2015 ($37,7\%$ de l'épargne des ménages) et les montants qu'elle représente sont colossaux (1650 Mds €)[Banque De France, 2016]. Ce montant explique à lui seul les enjeux de gestion actif passif que représente ces produits pour les assureurs.

La gestion des produits d'assurance repose sur la compréhension et la maîtrise des risques sous-jacents de la part des assureurs. Les risques liés à l'opération d'assurance vie sont de natures diverses et variées. Les risques de marchés sont souvent mieux maîtrisés grâce à la quantité de données disponibles et le fort développement des modèles financiers. L'entrée en vigueur de la directive Solvabilité II renforce ce besoin de compréhension et de maîtrise des risques.

Au mois de mai 2016 les assureurs français ont observé $9,4$ Mds € de prestations pour les produits d'épargne[FSA, 2016]. Ces prestations sont conséquence de la liquidité offerte par ce type de contrat en cas de décès ou de simple décision de rachat. Ainsi, à tout moment, l'assuré peut récupérer à sa seule demande son épargne. Cette option de liquidité est un risque que l'assureur doit intégrer dans la gestion de son business.

Depuis le 1^{er} janvier 2016, la directive Solvabilité II impose la représentation des passifs en juste valeur au passif du bilan économique. Cette représentation se traduit majoritairement par le Best Estimate. L'évaluation du Best Estimate repose sur la meilleure estimation des hypothèses et des paramètres de la projection. L'estimation des prestations est donc un point central du calcul du Best Estimate des produits épargne. Si la modélisation du décès est aujourd'hui maîtrisé (tables réglementaires, tables d'expérience, etc.), la modélisation du rachat demande encore à être approfondie.

Les modélisations existantes ne répondent souvent pas aux critères de compréhension et de meilleure estimation requis par la directive Solvabilité II. Un besoin de mis-à-jour des modèles est donc créé. Dans le compte-rendu de l'Institut des Actuaires de la commission Solvabilité II du 10 mars 2016[Institut des Actuaires - ACPR, 2016], l'ACPR demande aux assureurs de ne pas se limiter aux lois ONC largement utilisées et de saisir pleinement les limites de la modélisation retenue.

L'environnement économique mouvant et oppressant actuel pousse également les assureurs à revoir l'importance qu'ils accordent au risque de rachat. Le contexte de taux durablement bas nourri par la politique monétaire de la Banque Centrale Européenne (Quantitative Easing depuis mars 2015)[M. Charrel, 2016] diminue peu à peu le rendement des produits d'assurance vie et augmente le risque de rachat massif.

C'est dans ce cadre économique et réglementaire que se place l'étude. L'objectif est de com-

prendre le risque afin de proposer une modélisation adaptée.

Dans un premier temps, un focus sur le contexte économique et réglementaire est réalisé. Le risque de rachat est développé et les lois ONC sont présentées. Ceci permet de comprendre le risque et de mieux l'appréhender afin de proposer une modélisation adaptée.

Le besoin de prendre en compte les facteurs comportementaux des assurés mènent à une modélisation de type régression logistique. Un modèle de ce type est donc construit étape après étape dans un second temps de l'étude. Des modèles hybrides et à états sont construits afin de répondre aux besoins a priori de la modélisation du rachat.

Dans un troisième temps le modèle proposée est mis en pratique sur le portefeuille d'un assureur français : il est calibré et testé. L'étude de la qualité du modèle est faite à travers diverses méthodes mathématiques usuelles afin de s'assurer du pouvoir de prédiction du modèle proposé.

Dans un dernier temps, une sensibilité du BE aux lois de rachats retenues est effectuée. Les différentes lois proposées sont testées afin de mesurer leur impact sur le BE.

Première partie

Mise en contexte

Chapitre 1

Assurance vie

1.1 Présentation générale

Il est important de mesurer le périmètre des produits d'assurance vie pour en comprendre les enjeux et en affiner la modélisation. Une bonne compréhension du concept d'assurance sur la vie et une exposition des chiffres clés de ce business sont donc nécessaires pour appréhender l'étude.

L'assurance vie est une opération par laquelle un assureur s'engage à verser une prestation à un bénéficiaire en cas de réalisation d'un évènement aléatoire afférent à la durée de vie de l'assuré. Cet engagement nait de primes versées par l'assuré. Ces opérations représentent des montants colossaux en France : près de 1650 Mds €.

A fin 2015, l'étude Autorité de Contrôle Prudentiel et de Résolution (ACPR) sur l'épargne des ménages français[Banque De France, 2016] montre que l'assurance vie reste le placement préféré des Français (37, 7% contre 13, 7% en livrets divers). Voici un récapitulatif de cette étude :

Répartition des encours épargnes		
Principaux placements financiers (Mds €)	**4 377**	**100%**
Actifs liquides et non risqués	1 056,8	24,1%
Numéraire	67,4	1,5%
Dépôts à vue	374,6	8,6%
Livrets	600,0	13,7%
OPC monétaires	14,8	0,3%
Autres actifs non risqués	1 702,0	38,9%
Comptes à terme	74,6	1,7%
PEL / PEP	252,9	5,8%
Assurance-vie en supports euros	1 374,5	31,4%
Actifs liquides et risqués	582,3	13,3%
Autres actifs risqués	1 035,9	23,7%
Actions non cotées	760,7	17,4%
Assurance-vie en supports UC	275,2	6,3%

FIGURE 1.1 – Répartition des montants de l'épargne des ménages français

Source : Banque de France - Epargne des ménages 2016

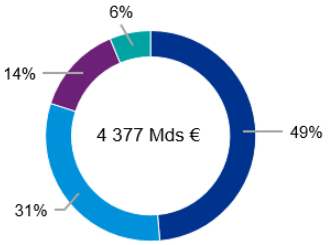

FIGURE 1.2 – Répartition en parts de marché

Source : Banque de France - Epargne des ménages 2016

L'intérêt des français pour l'assurance vie, face au livret A qui est historiquement le plus populaire, s'explique par un taux de rendement du livret A trop bas ($0,75\%$ à octobre 2016). Cependant les contrats en Unités de Compte (UC) rencontrent un succès grandissant ces dernières années car, même si supérieur au livret A, le rendement moyen des produits d'assurance vie reste relativement bas ($2,2\%$[Banque De France, 2016]).

Selon une étude de la Fédération Française des Assurances[FSA, 2016], l'assurance vie (fonds euros comme UC) a récolté $10,8$ Mds € de cotisations au mois de mai 2016. La moyenne des cotisations mensuelles observées depuis janvier 2015 est de 10 Mds €. La même étude mentionne que les assureurs ont observé $9,4$ Mds € de prestations au mois de mai 2016.

Ces chiffres montrent à la fois l'attrait que représente l'assurance vie pour les ménages français, mais également le niveau des montants engagés et l'importance des enjeux que ces montants représentent.

1.2 Produits d'épargne

Les contrats d'assurance vie les plus représentés sur le marché français sont les contrats dits d'épargne. Ces contrats s'apparentent à une gestion d'actif : en effet l'assuré verse des primes à l'assureur qui va les investir dans des actifs financiers tout en lui accordant certaines options et garanties. Le capital ainsi constitué sera disponible pour l'assuré.

Deux grandes sous-catégories des produits d'épargne se distinguent : les fonds euro et les fonds UC. Le principe de base reste commun, mais les engagements de l'assureur sont complètement différents selon le produit.

1.2.1 Produits sur fonds euro

Les produits sur fonds euro garantissent le capital investi et un taux minimum de revalorisation [1]. In fine, l'assuré récupère son capital majoré des intérêts garantis et des intérêts distribués par l'assureur.

1. Ce taux minimum peut être nul et son expression peut être brute ou nette

Ce type de contrat est donc avantageux car sécuritaire ; le risque engendré est porté par l'assureur.

L'épargne présente également un avantage fiscal. Les intérêts produits par le contrat d'assurance vie sont d'abord abattus des cotisations sociales (la Contribution Sociale Généralisée (CSG) de $8,2\%$, la Contribution au Remboursement de la Dette Sociale (CRDS) de $0,5\%$ et les prélèvements sociaux de $6,8\%$). Deux possibilités s'offrent ensuite à l'assuré : soit l'intégration de ces intérêts dans la déclaration d'impôts sur le revenu, soit le prélèvement de l'impôt lors de la sortie du produit (principe du Prélèvement Forfaitaire Libératoire (PFL)). Quel que soit le mode d'imposition choisi par l'assuré, le taux d'imposition est une fonction décroissante de l'ancienneté du contrat. Voici un tableau récapitulatif de l'imposition[Code Des Impôts, 2016] :

Fiscalité des contrats d'assurance vie (hors prélèvements sociaux)		
Date de souscription	**Abattement sur les produits**	**Taux d'imposition**
Avant 01/01/1983	Exonération	0%
Entre 02/01/1983 et 31/12/1989	Abattement annuel de 4 600 € pour une personne seule et 7 200 € pour un couple soumis à une imposition commune	45% si ancienneté < 2 ans 25% si 2 ans ≤ ancienneté < 4 ans 15% si 4 ans ≤ ancienneté < 6 ans 7,5% si 6 ans ≤ ancienneté
Après 01/01/1990		45% si ancienneté < 2 ans 25% si 2 ans ≤ ancienneté < 4 ans 15% si 4 ans ≤ ancienneté < 8 ans 7,5% si 8 ans ≤ ancienneté

FIGURE 1.3 – Fiscalité des contrats d'assurance vie en France

L'avantage le plus important est acquis à partir de 8 ans d'ancienneté. Ces avantages ont donc un impact sur la durée moyenne des contrats.

Le dernier avantage est la disponibilité du capital. Si l'assuré exprime un besoin de liquidité, il peut racheter son contrat à tout moment moyennant des pénalités de rachats ou bénéficier d'une avance. En cas de décès de l'assuré, le capital est reversé aux bénéficiaires.

1.2.2 Produits en UC

Les produits en UC différent des produits sur fonds euro de par la nature de l'engagement de l'assureur. En effet, dans la catégorie fonds euro, l'engagement de l'assureur est exprimé en euros et est associé à un taux minimum garanti. Dans les produits en UC, la garantie est exprimée en nombre de parts de l'UC. Le risque est porté par l'assuré : il n'y a pas de garantie du capital ni des intérêts.

Cette catégorie de produit n'est pas le cœur de cette étude, mais elle intervient dans la compréhension du rachat.

Il existe des produits hybrides : les contrats multi-supports. Dans ce cas le capital est scindé en une partie en fonds euro et une partie en UC.

Chapitre 2

Solvabilité II

2.1 Réglementation

Depuis le 1er janvier 2016, la directive Solvabilité II est entrée en vigueur en Europe. Cette dernière a amené une nouvelle vision aux assureurs européens : une vision économique. La juste valeur est l'idée directrice de cette nouvelle réglementation. Les spécifications techniques sont données par l'European Insurance and Occupational Pensions Authority (EIOPA) au niveau européen, et la mise en application est assurée par l'ACPR en France.

La réglementation est elle-même divisée en 3 piliers. Le premier pilier concerne les exigences quantitatives. Ce 1er pilier est constitué de 2 grands principes : le bilan économique et le Capital de Solvabilité Requis (SCR). Le second pilier est à mi-chemin entre exigences quantitatives et exigences qualitatives. Les deux éléments clés sont la gouvernance et l'Own Risk and Solvency Assessment (ORSA). Le troisième et dernier pilier concerne la transparisation et la communication. Deux rapports phares en font parties : le RSR et le SFCR.

Le but de l'étude est de mesurer les impacts quantitatifs de la modélisation des rachats en environnement Solvabilité II, un focus est donc réalisé sur le premier pilier de la norme prudentielle.

2.2 Bilan économique

2.2.1 Valeur de marché des actifs

En norme Solvabilité I, les actifs sont représentés en valeur d'achat retraitée d'amortissements et de dépréciations. Une vision rétrospective - mais mise à jour annuellement - des risques de marché est donc adoptée. Ceci ne permet donc pas d'avoir une représentation du bilan en accord avec le marché. C'est pourquoi le concept de bilan économique intègre une représentation de l'actif en valeur de marché. Le principe clé est la fidélité du bilan et de la solvabilité de la compagnie.

2.2.2 Best Estimate

Solvabilité II introduit également une vision économique du passif. Les provisions techniques prudentes en normes françaises sont remplacées par le Best Estimate (BE) [1]. La vision en juste

1. Dénommé meilleure estimation dans le règlement délégué

valeur du passif et donc des engagements est maintenant adoptée. Compréhension et maîtrise du risque sont maître-mots pour parvenir à cette évaluation.

Le BE est défini dans la Directive Solvabilité II[Parlement et Conseil Européens, 2009] comme *la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents.*

Les risques se décomposent en deux catégories : les risques couvrables (exemple : garantie plancher) et les risques non couvrables (exemple : mortalité). Pour les risques couvrables, la valorisation doit être cohérente avec le marché. Par exemple : le risque d'assurance associé à une garantie plancher repose sur l'exercice d'une option de vente américaine dont le sous-jacent est défini contractuellement, il doit être valorisé comme tel. Pour les risques non-couvrables, la valorisation repose sur le principe de la meilleure estimation des hypothèses.

Pour bien comprendre la méthode de calcul du BE, un exemple de produit épargne est présenté. Soit un contrat d'épargne dont une prime unique d'1 M € est versée à $t = 0$. Le contrat dispose d'un Taux Minimum Garanti (TMG) de 2%. L'horizon de projection est fixé à 5 ans. Le rendement comptable de l'assureur est supposé constant égal à 3%. La politique de participation aux bénéfices de l'assureur est supposée être de 85% du rendement comptable. Le taux de mortalité est supposé constant égal à 1% et le taux de rachat à $1,5\%$. A la fin de la projection, tous les assurés rachètent leur contrat. Soit une courbe de taux sans risque arbitraire.

Rendement scénario central : **3,0%**

Année	0	1	2	3	4	5
Prime	1 000 000 €	0 €	0 €	0 €	0 €	0 €
Rendement annuel de l'actif	0 €	30 000 €	30 015 €	30 030 €	30 045 €	30 060 €
Produits servis	0 €	25 500 €	25 513 €	25 526 €	25 538 €	25 551 €
Décès	0 €	10 000 €	10 005 €	10 010 €	10 015 €	10 020 €
Rachats	0 €	15 000 €	15 008 €	15 015 €	15 023 €	1 017 533 €
Montant de l'épargne	1 000 000 €	1 000 500 €	1 001 000 €	1 001 501 €	1 002 002 €	0 €
Flux total annuel	**1 000 000 €**	**25 000 €**	**25 013 €**	**25 025 €**	**25 038 €**	**1 027 553 €**

Taux sans risque	-	3,0%	3,0%	3,0%	3,0%	3,0%
Facteur d'actualisation	-	0,971	0,943	0,915	0,888	0,863

BE déterministe	**979 371 €**

FIGURE 2.1 – Décomposition des flux du Best Estimate déterministe

La prime est comptabilisée en $t = 0$; elle n'intervient donc plus dans le calcul du BE mais ouvre des droits futurs. La mise à jour de l'épargne est en fait un ajout d'intérêts crédités puis un retrait en cas de décès ou de rachats en fin d'année.

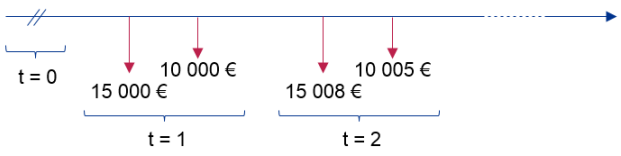

FIGURE 2.2 – Schéma des flux du Best Estimate déterministe

L'utilisation d'un Générateur de Scénarios Economiques (GSE) permet de calculer un BE stochastique. Ceci permet de prendre en compte les options asymétriques du contrat telles que le TMG et le rachat. En introduisant un deuxième scénario de rendement de l'actif dans l'exemple précédent, il est possible de valoriser les garanties du TMG de 2% et de la clause de participation aux bénéfices. De part la prise en compte de ces options en faveur de l'assuré, le BE stochastique est en général plus élevé que le BE déterministe. Il en résulte une meilleure estimation des engagements.

Voici les flux sur un premier scénario dont le rendement est revu à la baisse :

Rendement scénario 1 : **1,5%**

Année	0	1	2	3	4	5
Prime	1 000 000 €	0 €	0 €	0 €	0 €	0 €
Rendement annuel de l'actif	0 €	15 000 €	14 925 €	14 850 €	14 776 €	14 702 €
Produits servis	0 €	20 000 €	19 900 €	19 801 €	19 701 €	19 603 €
Décès	0 €	10 000 €	9 950 €	9 900 €	9 851 €	9 801 €
Rachats	0 €	15 000 €	14 925 €	14 850 €	14 776 €	989 951 €
Montant de l'épargne	1 000 000 €	995 000 €	990 025 €	985 075 €	980 150 €	0 €
Flux total annuel	**1 000 000 €**	**25 000 €**	**24 875 €**	**24 751 €**	**24 627 €**	**999 752 €**
Taux sans risque	-	3,0%	3,0%	3,0%	3,0%	3,0%
Facteur d'actualisation	-	0,971	0,943	0,915	0,888	0,863

BE scénario 1	**954 645 €**

FIGURE 2.3 – Décomposition des flux du Best Estimate sur le scénario 1

Et sur un second scénario dont le rendement est revu à la hausse :

Rendement scénario 2 : **4,5%**

Année	0	1	2	3	4	5
Prime	1 000 000 €	0 €	0 €	0 €	0 €	0 €
Rendement annuel de l'actif	0 €	45 000 €	45 596 €	46 200 €	46 813 €	47 433 €
Produits servis	0 €	38 250 €	38 757 €	39 270 €	39 791 €	40 318 €
Décès	0 €	10 000 €	10 133 €	10 267 €	10 403 €	10 541 €
Rachats	0 €	15 000 €	15 199 €	15 400 €	15 604 €	1 083 840 €
Montant de l'épargne	1 000 000 €	1 013 250 €	1 026 676 €	1 040 279 €	1 054 063 €	0 €
Flux total annuel	**1 000 000 €**	**25 000 €**	**25 331 €**	**25 667 €**	**26 007 €**	**1 094 381 €**
Taux sans risque	-	3,0%	3,0%	3,0%	3,0%	3,0%
Facteur d'actualisation	-	0,971	0,943	0,915	0,888	0,863

BE scénario 2	**1 038 767 €**

FIGURE 2.4 – Décomposition des flux du Best Estimate sur le scénario 2

Finalement le BE stochastique et l'Actif Net Réévalué (ANR) [2] sont obtenus :

BE stochastique	**996 706 €**
ANR	**3 294 €**

FIGURE 2.5 – Exemple d'ANR

2. ANR = Valeur de marché de l'actif − BE

2.2.3 Fonds propres

La solvabilité de l'entreprise se mesure à travers deux éléments : les Fonds Propres (FP) et le SCR. La détermination des FP représente donc une étape essentielle sous Solvabilité II.

Ceux-ci sont divisés en deux catégories : les fonds propres de base et les fonds propres auxiliaires. Les FP admis en représentation du SCR sont déterminés par le biais de règles de Tiering : il existe 3 Tiers qui représentent la qualité de ces derniers. Le but étant de détenir des FP de haute qualité pour faire face à la déviation maximum prévue par le SCR. La qualité des fonds propres est mesurée par rapport à la disponibilité de ces fonds, et à leur subordination. Si l'élément de FP a une durée déterminée, il est nécessaire que cette durée soit supérieure à celle des engagements d'assurance pour être éligible[Parlement et Conseil Européens, 2009].

2.3 Solvency Capital Requirement

Le SCR est le capital de solvabilité requis. Il représente un niveau minimum de fonds propres qu'une compagnie d'assurance doit détenir pour être solvable et donc pouvoir faire face aux variations de ses engagements. La méthodologie d'évaluation du SCR est ici présentée.

2.3.1 Une métrique de risque

La directive Solvabilité II introduit également une nouvelle métrique de risque. Cette métrique repose sur la notion de Value-at-Risk (VaR). La VaR correspond à une déviance maximum de la valeur sous risque sur une période donnée et à un niveau de confiance donné.

Définition : Soit $(X_k)_{k\in\mathbb{R}_+}$ un processus réel. Soit $t \in \mathbb{R}_+$. La VaR de X_t à horizon δ au seuil α est le réel que l'on note $\mathrm{VaR}_{1-\alpha}(X_t, \delta)$ tel que :

$$\mathrm{VaR}_\alpha(X_t, \delta) = \underset{Y\in\mathbb{R}}{\operatorname{argmin}}\{P(X_{t+\delta} - X_t \leqslant -Y)) \leqslant 1 - \alpha\}$$

Sous Solvabilité II, la déviance maximum de l'ANR à horizon 1 an et au niveau de confiance de $99,5\%$ est mesurée. Cette déviance est appelée SCR.

Les articles 103 à 108 de la directive[Parlement et Conseil Européens, 2009] définissent le SCR. Ce dernier est décomposé en plusieurs modules et sous-modules comme l'illustre le schéma suivant :

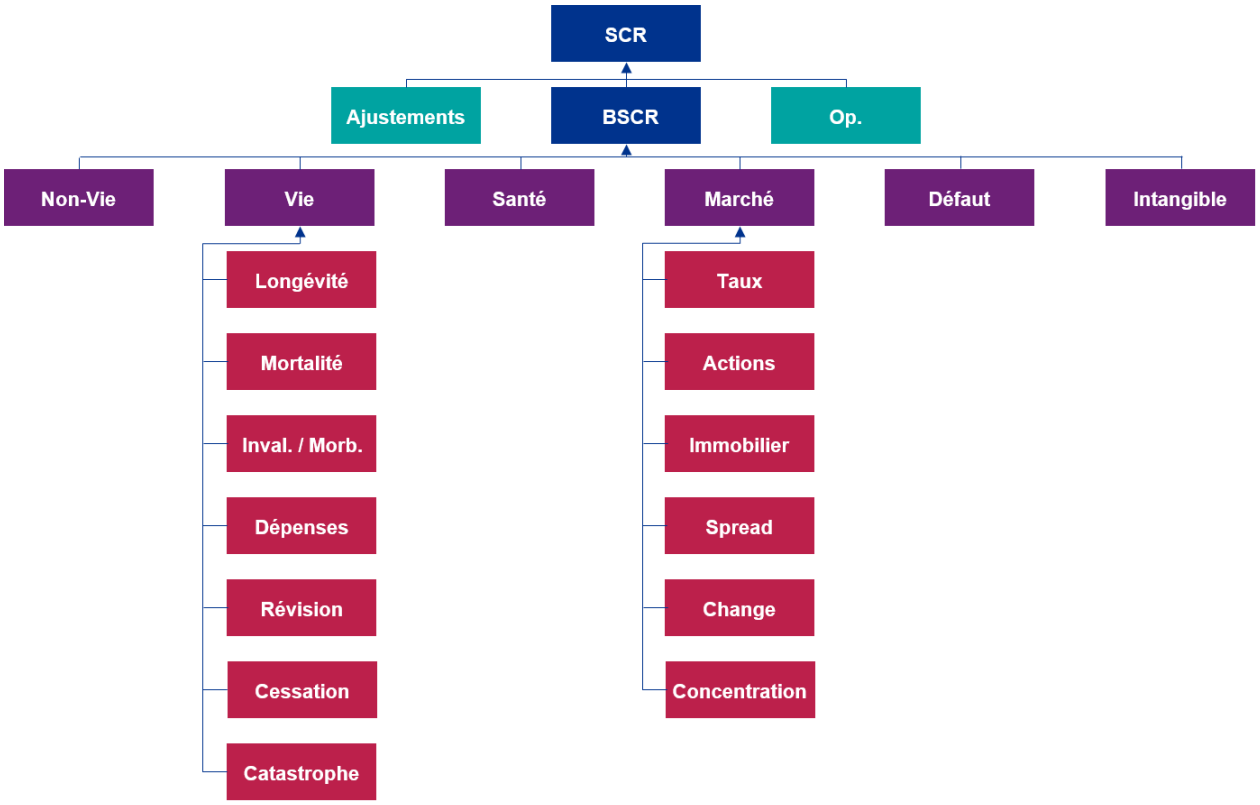

FIGURE 2.6 – Eclatement schématique du SCR

Seuls les modules de souscription vie et de marché sont présentés compte tenu du périmètre de l'étude, mais il est utile de noter que les modules de souscription santé et de souscription non-vie se développent aussi en sous-modules.

2.3.2 Module marché

Le module marché représente la déviance maximale de l'ANR incombant uniquement aux risques de marché. Chaque sous-module est calculé comme une différence d'ANR choqué et non choqué. Un exemple de calcul du sous-module taux est présenté.

Taux

Le risque de taux est le risque de voir l'ANR baisser consécutivement à une hausse ou baisse de la courbe de taux sans-risque de référence.

L'exemple de calcul de BE dans la figure 2.5 à la page 9 est ici repris. L'actif de départ est, par hypothèse, composé uniquement de la prime. Un ANR non-choqué a déjà été calculé.

Le BE du produit choisi est sensible aux taux. Les flux d'actif sont supposés insensibles car le rendement comptable est constant et le BE est quant à lui sensible au mouvements des taux par l'effet d'actualisation.

Le SCR étant une VaR, le principe est de calculer un bilan économique choqué après application d'un choc de taux. Ces chocs sont calibrés, revus et fournis par l'EIOPA dans les spécifications techniques[EIOPA, 2014].

Pour le sous-module taux, deux courbes choquées sont fournies : un choc haussier et un choc baissier. Les deux chocs sont appliqués et le choc le plus défavorable en matière d'impacts sur l'ANR est conservé.

Ici seule la baisse des taux sera défavorable. En effet une baisse des taux ne changera pas l'actif initial (qui est égal à la prime) et va mécaniquement augmenter le BE par augmentation des facteurs d'actualisation. L'ANR sera donc réduit. De la même manière, il est déductible qu'une hausse des taux augmentera l'ANR.

Un choc baissier arbitraire[3] a été appliqué et donne les résultats suivants :

Rendement scénario 1 : **1,5%**

Année	0	1	2	3	4	5
Prime	1 000 000 €	0 €	0 €	0 €	0 €	0 €
Rendement annuel de l'actif	0 €	15 000 €	14 925 €	14 850 €	14 776 €	14 702 €
Produits servis	0 €	20 000 €	19 900 €	19 801 €	19 701 €	19 603 €
Décès	0 €	10 000 €	9 950 €	9 900 €	9 851 €	9 801 €
Rachats	0 €	15 000 €	14 925 €	14 850 €	14 776 €	989 951 €
Montant de l'épargne	1 000 000 €	995 000 €	990 025 €	985 075 €	980 150 €	0 €
Flux total annuel	**1 000 000 €**	**25 000 €**	**24 875 €**	**24 751 €**	**24 627 €**	**999 752 €**
Taux sans risque	-	2,9%	2,9%	2,9%	2,9%	2,9%
Facteur d'actualisation	-	0,972	0,944	0,918	0,892	0,867
BE scénario 1	**959 064 €**					

FIGURE 2.7 – Décomposition des flux du Best Estimate sur le scénario 1 choqué

Rendement scénario 2 : **4,5%**

Année	0	1	2	3	4	5
Prime	1 000 000 €	0 €	0 €	0 €	0 €	0 €
Rendement annuel de l'actif	0 €	45 000 €	45 596 €	46 200 €	46 813 €	47 433 €
Produits servis	0 €	38 250 €	38 757 €	39 270 €	39 791 €	40 318 €
Décès	0 €	10 000 €	10 133 €	10 267 €	10 403 €	10 541 €
Rachats	0 €	15 000 €	15 199 €	15 400 €	15 604 €	1 083 840 €
Montant de l'épargne	1 000 000 €	1 013 250 €	1 026 676 €	1 040 279 €	1 054 063 €	0 €
Flux total annuel	**1 000 000 €**	**25 000 €**	**25 331 €**	**25 667 €**	**26 007 €**	**1 094 381 €**
Taux sans risque	-	2,9%	2,9%	2,9%	2,9%	2,9%
Facteur d'actualisation	-	0,972	0,944	0,918	0,892	0,867
BE scénario 2	**1 043 592 €**					

FIGURE 2.8 – Décomposition des flux du Best Estimate sur le scénario 2 choqué

L'ANR choqué et le SCR taux sont :

3. Courbe de taux plate égale à 2,9%

BE stochastique choqué	1 001 328 €
ANR choqué	-1 328 €
SCR taux	4 622 €

FIGURE 2.9 – Exemple de SCR Taux

Action

Le risque action est le risque de voir son actif se déprécier en raison d'une baisse du niveau des actions détenues.

Les spécifications techniques[EIOPA, 2014] décomposent le choc à appliquer en deux parties. Les actions sont décomposées en deux catégories : les actions cotées dans l'Organisation de Coopération et de Développement Économiques (OCDE) (Type 1) et les actions non cotées dans l'OCDE (Type 2). Les premières sont choquées à la baisse de 39% de leur valeur de marché et les secondes à 49%. A ce choc s'ajoute un effet Dampener qui permet d'adapter le choc en fonction du niveau moyen de l'indice sur les trois dernières années[4].

Un régime différent est appliqué aux actions détenues pour participations stratégiques. Compte tenu des objectifs particuliers de la détention de ces titres, le choc appliquée est de 22%.

Le SCR action final est obtenu par agrégation à l'aide d'une matrice de corrélation.

Immobilier

Le risque immobilier est le risque de voir son actif déprécier en raison de l'explosion d'une bulle immobilière qui ferait chuter la valeur de marché des titres immobiliers détenus.

Le choc à appliquer fourni par les spécifications techniques[EIOPA, 2014] est de 25%.

Ce sous-module concerne tous les types de titres : exploitation, investissement et parts de sociétés immobilières.

Spread

Le risque de spread est le risque de voir son actif se déprécier en raison d'évènements de crédit. L'évènement crédit peut se manifester de deux manières : un changement de la qualité de crédit de l'émetteur (changement de rating) ou une incapacité de l'émetteur à rembourser ses dettes (perte de coupons ou de nominaux).

La formule standard ne prévoit pas de choc pour les obligations d'états européens. Pour les autres titres de crédit, l'intensité du choc est fonction du rating[5] et de la duration de Macaulay du titre.

4. Avec comme contrainte : $|\text{Dampener}| \leqslant 10\%$
5. Le rating retenu est le deuxième meilleur fourni par les agences reconnues

Le SCR spread se décompose comme la somme de trois éléments : la partie obligations, la partie produits dérivés [6] et la partie produits structurés [7].

Change

Le risque de change est le risque de voir son actif déprécier en raison d'un mouvement du taux de change entre différentes monnaies.

Pour chaque monnaie étrangère [8], un choc haussier du change de 25% et un choc baissier du change de 25% doivent être appliqués. Comme pour le sous-module taux, le scénario le plus défavorable est conservé. Ce procédé fournit donc un SCR par monnaie étrangère. Le SCR change est finalement obtenu par somme de tous ces SCR.

Concentration

Le risque de concentration mesure la surexposition de l'actif à un émetteur donné. Détenir une proportion trop importante de son actif en titres d'un même émetteur est risqué car une situation de défaut implique une perte massive.

Ce sous-module concerne tous les produits de crédit, d'action et d'immobilier. Les actifs liés aux opérations en UC ne sont pas concernés.

Un SCR concentration est calculé pour chaque émetteur. Il est égal à un pourcentage [9] de l'exposition à cet émetteur (méthode déterministe sans calcul d'un ANR choqué). Le SCR change final est ensuite obtenu par agrégation à l'aide d'une matrice de corrélation.

Agrégation

Une fois que les sous-modules sont calculés, le SCR marché est obtenu par agrégation à l'aide d'une matrice de corrélation. Les spécifications techniques en fournissent deux selon le sens le plus défavorable du choc de taux.

2.3.3 Module de souscription vie

Le module de souscription vie correspond à une déviance maximale de l'ANR inhérente à l'opération d'assurance sur la vie. Comme le module marché, il est décomposé en sous-modules qui sont eux-mêmes calculés comme un delta d'ANR choqué et non choqué.

6. CDS, TRS, CLN, etc.
7. Repackaged loans, ABS, CDO, etc.
8. La monnaie locale est la monnaie dans laquelle la compagnie prépare ses états financiers
9. Ce pourcentage est fonction de la part d'exposition de l'actif et du rating de l'émetteur

Longévité

Le risque de longévité est le risque d'une sous-tarification ou d'un sous-provisionnement en raison d'un allongement de la durée de vie. Les produits de garanties en cas de vie type rentes viagères sont sensibles à ce risque.

Un choc relatif baissier de la mortalité de 20% est appliqué pour obtenir un ANR choqué. Le module longévité est finalement égal à la différence entre les deux ANR.

Mortalité

Le risque de mortalité est le pendant à la hausse du risque de longévité. Les produits de type capitaux décès y sont sensibles.

Un choc relatif haussier de la mortalité de 15% est appliqué.

Invalidité / Morbidité

Le risque d'invalidité et de morbidité s'inscrit dans la suite des risque de mortalité et de longévité mais il concerne les risques de maladies particulières.

A l'image du risque de taux du module marché, deux chocs sont fournis par les spécifications techniques : un choc haussier et un choc baissier du taux de morbidité et d'invalidité. Le plus défavorable est conservé.

Dépenses

Le risque de dépenses mesure l'exposition de l'ANR par rapport aux hypothèses de frais dans la projection des contrats. Le montant du module est le résultat de la combinaison de deux chocs : une hausse de 10% des frais et une hausse de 1% du taux d'inflation des coûts.

Révision

Le risque de révision a pour but d'évaluer l'impact d'une évolution du montant des annuités due à l'inflation, à des changements législatifs ou à la dégradation de l'état des sinistrés.

Le choc appliqué est une hausse de 4% des taux de passage de dépendance partielle à dépendance totale.

Cessation

Le sous-module cessation évalue le risque d'une évolution défavorable du taux de rachat des contrats.

Trois chocs sont donnés par les spécifications techniques : un choc relatif haussier de 50% du taux de rachat, un choc relatif baissier de 50% du même taux et un rachat total de 40% de portefeuille la première année.

Seul le plus défavorable des trois chocs est conservé.

Catastrophe

Le module catastrophe correspond à une hausse instantané de la mortalité sur les 12 prochains mois.

Le choc appliqué est un choc absolu haussier de 15% des taux de mortalité. Ce choc n'est appliqué que sur les produits dont l'impact sur les provisions techniques est défavorable.

Agrégation

La technique d'agrégation des sous-modules est comparable à celle du module marché. Elle repose sur l'utilisation d'une matrice de corrélation.

2.4 Marge de risque

Un nouveau poste est crée au passif du bilan économique : la Marge de Risque (RM). La directive[Parlement et Conseil Européens, 2009] précise que la calcul de cette marge doit être effectué *de manière à garantir que la valeur des provisions techniques est équivalente au montant que les entreprises d'assurance et de réassurance demanderaient pour reprendre et honorer les engagements d'assurance et de réassurance.*

La RM correspond au coût d'opportunité de la détention du SCR jusqu'à écoulement total des engagements. Le taux de ce coût d'opportunité est identique pour toutes les compagnies et est réévalué périodiquement [10]. Finalement le coût en capital s'exprime comme :

$$CoC = 6\%.\sum_{t\geq 0}\frac{SCR(t)}{(1+r_t)^t}$$

En pratique, l'évaluation de la RM est difficile car la projection du SCR est très coûteuse en temps de calcul. Les assureurs utilisent donc des méthodes d'approximation pour la déterminer. Les plus courantes sont :

i) Calcul exact par des méthodes déterministes
ii) Proxys pour les SCR de modules/sous-modules
iii) Proxys pour le SCR global directement
iv) Formule fermée pour la RM
v) Calcul comme un pourcentage conservatif du BE actuel (approche Solvabilité I)

10. Ce taux est de 6% actuellement et est réévalué par l'Autorité de Contrôle

2.5 Bilan schématique et ratio de solvabilité

Voici un schéma illustrant les différents éléments se rapportant au bilan économique :

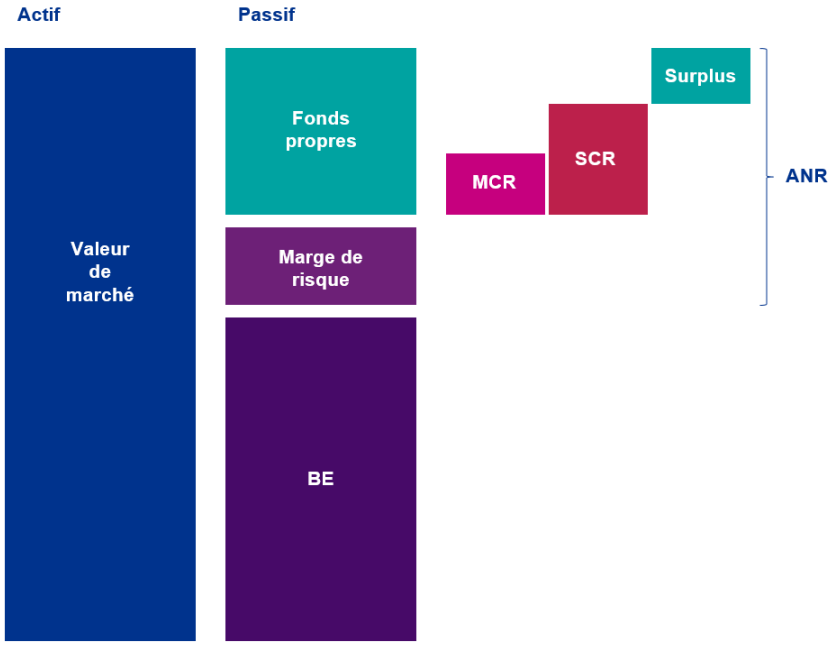

FIGURE 2.10 – Bilan économique sous Solvabilité II

Ce schéma permet également d'illustrer le ratio de solvabilité. Ce ratio est défini comme le rapport des fonds propres éligibles et du SCR. Il doit être supérieur à 100% pour que la compagnie soit solvable. Un ratio égal à 100% traduit la capacité à encaisser un choc bicentenaire et donc une seule ruine sur 200 ans.

Chapitre 3

Risque de rachats

3.1 Quid ?

3.1.1 Définition

Le rachat désigne l'action d'un assuré lorsqu'il décide, de son plein gré, de prélever son épargne avant le terme de son contrat. La valeur de rachat est définie[Code Des Assurances, 2016a] comme *la différence entre la valeur actuelle des engagements respectivement pris par l'assureur et par l'assuré.*

La valeur de rachat s'exprime comme :

$$\text{Valeur}_{\text{rachat}} = \sum_{i=0}^{N} \frac{\text{Flux assureur}_{t_i}}{(1+r_{t_i})^{t_i}} - \sum_{i=0}^{N} \frac{\text{Flux assuré}_{t_i}.P(\{\text{Assuré présent en t}_i\})}{(1+r_{t_i})^{t_i}}$$

Cette valeur de rachat est égale à la Provision Mathématique (PM) de l'assuré. En effet la PM est définie comme la différence entre les valeurs actuelles des engagements de l'assureur et de l'assuré.

Cependant, l'assureur peut pénaliser l'assuré qui souhaite prélever son épargne avant le terme du contrat en appliquant une pénalité de rachat. Le Code des Assurances stipule que cette pénalité ne peut excéder 5% de la PM et doit être nulle si l'ancienneté du contrat est supérieure ou égale à 10 ans.[Code Des Assurances, 2016b].

Le rachat apparait finalement comme une option pour l'assuré : il peut la déclencher quand il le souhaite et elle lui donne un pouvoir de liquidité sur son épargne. Mais ce pouvoir de liquidité à un coût pour l'assureur qui ne peut le contraindre que par le biais d'une pénalité encadrée légalement.

3.1.2 Risque et enjeux

Pour comprendre en quoi le rachat représente un risque pour l'assureur, il faut revenir aux fondamentaux des produits d'épargne. Comme présenté dans la partie 1.2 à la page 5, l'assureur investit les primes de l'assuré dans des actifs afin de lui garantir son capital majoré d'intérêts. Lors de l'exercice de l'option de rachat par l'assuré, l'assureur doit revendre ces actifs pour récupérer tout ou partie de la valeur de rachat. Deux situations défavorables sont alors possibles :

i) Les titres peuvent être illiquides et donc la revente forcée et rapide peut baisser leur valeur de marché

ii) Les titres peuvent être en situation de moins-value latente et le rachat peut pousser la réalisation de ces moins-values

Ces explications permettent d'illustrer en quoi ces situations sont défavorables à l'assureur et font du rachat un risque.

Certains titres en portefeuille ne sont pas toujours liquides et sont acquis par l'assureur dans une stratégie de Buy and Hold [1]. Une revente prématurée de ces titres peut donc altérer la stratégie de l'assureur en faisant à la fois baisser leur valeur de marché car illiquides et en faisant perdre des produits financiers futurs qui devaient permettre à l'assureur de servir les intérêts garantis.

D'autres titres se trouvent parfois en situation de moins-values latente tout en ayant un fort potentiel de remontée du cours. Les titres actions en sont un bon exemple en période de crise économique : ils subissent souvent de fortes chutes sur de courtes périodes mais une remontée du cours est régulièrement observée. La valeur d'exercice est alors très sensible à la date de cession du titre. La moins-value n'est donc pas toujours réellement défavorable à l'assureur, mais la contrainte de vente induite par le rachat peut le devenir.

Le rachat apparait donc comme un risque pour l'assureur, il peut être qualifié d'option asymétrique pour l'assuré.

Maintenant que le risque est identifié, il est nécessaire de le mesurer pour en comprendre les enjeux.

En régime de croisière, les rachats observés sur les portefeuilles d'assurance n'excèdent pas les 5%. Dans ce cas, le risque est maîtrisable par les assureurs : ils l'ont intégré dans leur gestion d'actif et de trésorerie pour répondre aux besoins des assurés. Mais un risque ne se mesure pas en espérance ; il se mesure sur les quantiles extrêmes. Il est difficile de discuter de la probabilité d'occurrence des évènements extrêmes avant même d'étudier la modélisation du risque, néanmoins il est possible d'appréhender la force de ces évènements a priori.

Et sur ce point, le risque de rachat n'est pas des moindres ; il peut mener à la faillite. L'exemple d'une crise économique [2] le montre aisément. Lors d'une crise économique affectant les marchés actions, les titres sont en moins-value latente franche et ne servent que peu de dividendes. Deux effets sont alors perceptibles : une baisse des rendements financiers qui affectent l'épargne des assurés et une perte de confiance [3] des assurés en tout système économique. Dès lors, les assureurs risquent d'essuyer une vague massive de rachats et donc de réaliser d'importantes moins-values. Ils ne peuvent plus faire face à l'intégralité de leurs engagements envers les assurés.

3.1.3 Rachats structurels et conjoncturels

La gestion du risque de rachat se révèle essentielle pour un assureur qui commercialise un produit d'épargne. Cette gestion passe d'abord par la compréhension des mécanismes sous-jacents au

1. Stratégie qui consiste en l'achat de titre en vue d'une détention à long-terme sans essayer d'encaisser une plus-value mais en se contentant des produits financiers dégagés tels que les coupons, les dividendes et les loyers
2. Evènement non négligeable et observé récemment (Crise des *subprime* de 2008)
3. Le comportement des assurés n'est pas toujours rationnel face aux situations extrêmes

risque. Comme abordé précédemment, le rachat est une décision prise par l'assuré : maîtriser ce risque revient donc à maîtriser le comportement de ce dernier. Le premier pas de cette démarche est d'isoler les facteurs influençant le comportement des assurés.

Les facteurs peuvent a priori être séparés en deux catégories : les facteurs propres à l'assuré (âge, sexe, niveau d'étude, revenus, nombre de personnes à charge, avantages fiscaux et facteurs dont ils dépendent tels que l'ancienneté en France, etc.) et les facteurs économiques (niveaux des taux, niveaux des indices immobilier et action, inflation, etc.).

En pratique, ces deux catégories donnent lieu à la scission du rachat en deux composantes : les rachats structurels et les rachats conjoncturels. Les rachats structurels correspondent aux rachats liés aux caractéristiques du contrat et des assurés tandis que les rachats conjoncturels sont liés à l'environnement économique et concurrentiel.

3.2 QIS 5

3.2.1 Définition et rôle des QIS et ONC

La mise en place de la réforme Solvabilité II est passée par plusieurs étapes. Pendant son élaboration, plusieurs études d'impacts ont été menées : les Quantitative Impact Studies (QIS). Le cinquième et dernier QIS porte sur les exigences quantitatives dans leur globalité à savoir les provisions techniques, le SCR et la classification en FP.

Des Orientations Nationales Complémentaires (ONC) ont accompagné le QIS 5 en 2010. Ce sont de vraies lignes directrices fournies par l'ACPR en vue des premiers états quantitatifs. Des directives ont donc été données pour la modélisation des rachats et bon nombre d'assureurs les utilisent encore aujourd'hui [4]. Les modélisations proposées pour ces lois ont été reprises dans les ONC de 2013[ACPR, 2013].

3.2.2 Modélisation des rachats structurels

L'utilisation de lois d'expérience été préconisée pour la partie structurelle des rachats. Les lois d'expérience se reposent sur l'historique observé des rachats. Ces lois peuvent dépendre de nombreux paramètres tels que l'âge de l'assuré, l'ancienneté fiscale, etc.

Les ONC proposent des simplifications en cas de manque de données avec, par exemple, l'utilisation d'un taux moyen unique applicable à tous les contrats d'un groupe homogène de risque [5]. La proposition pour ce type de rachat est finalement peu restrictive.

3.2.3 Modélisation des rachats conjoncturels

Les ONC donnent davantage de directions sur la partie conjoncturelle des rachats. Elles précisent que ces derniers sont *couramment modélisés par une fonction dépendant uniquement de l'écart entre le taux servi et un taux dépendant de l'environnement économique, souvent appelé taux de*

4. Benchmark parmi les clients KPMG
5. Les Model Point (MP) désignent souvent ces groupes homogènes de risque

rendement espéré par l'assuré.

Elles fournissent deux lois de rachats conjoncturels dont l'une correspond à un plafond minimal et l'autre à un plafond maximal. Ces deux lois dépendent de plusieurs paramètres fournis par les ONC :

Paramètres des lois						
	Alpha	Beta	Gamma	Delta	RCm	RCM
Plafond M	-4%	0%	1%	4%	-4%	40%
Plafond m	-6%	-2%	1%	2%	-6%	20%

FIGURE 3.1 – Paramètres des lois minimum et maximum des ONC

Voici l'expression de chacune des lois, fonction des paramètres :

$$RC(R) = \begin{cases} RC_{max} & \text{si } R - TA < \alpha \\ RC_{max}.\dfrac{R - TA - \beta}{\alpha - \beta} & \text{si } \alpha < R - TA < \beta \\ 0 & \text{si } \beta < R - TA < \gamma \\ RC_{min}.\dfrac{R - TA - \gamma}{\delta - \gamma} & \text{si } \gamma < R - TA < \delta \\ RC_{min} & \text{si } \delta < R - TA \end{cases} \tag{3.1}$$

Les deux lois suivantes sont obtenues :

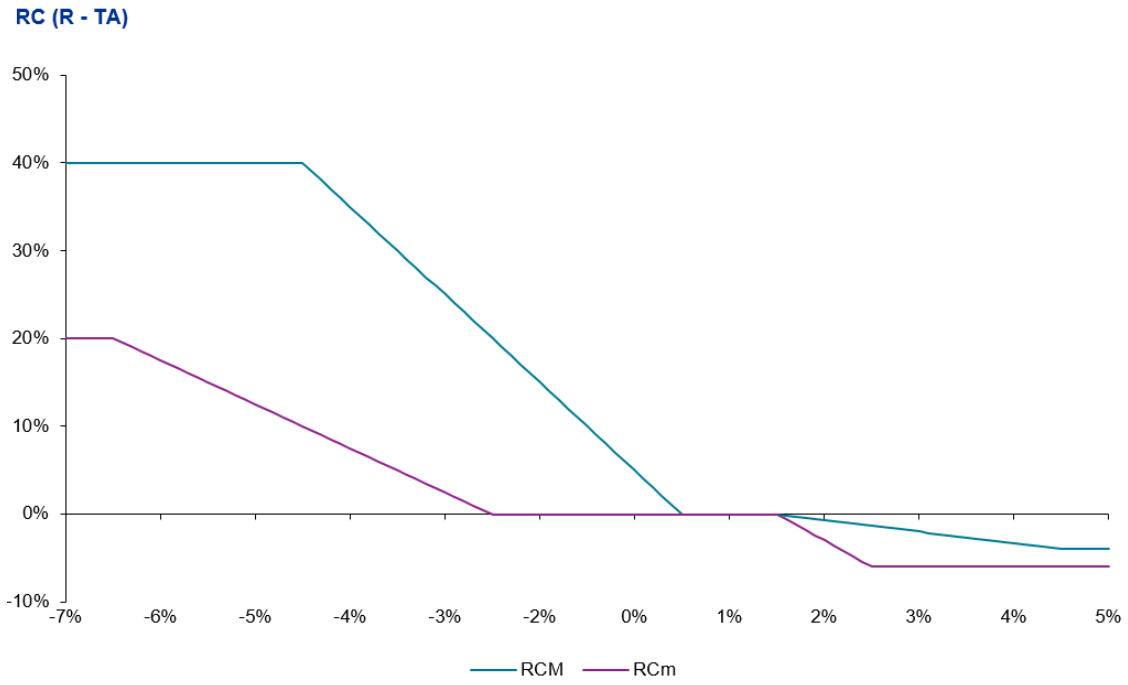

FIGURE 3.2 – Graphique des lois minimum et maximum des ONC

In fine, le rachat total s'exprime comme la somme des deux composantes structurelle et conjoncturelle bornées par 0 et 1. Les rachats conjoncturels peuvent être négatifs et signifient alors que le rachat total est inférieur au rachat structurel. Ce phénomène est envisageable lorsque le taux servi par l'assureur sur-performe le marché.

3.3 Impacts du choix de la loi

Maintenant que l'importance du risque que représente le rachat a été abordée, il est nécessaire de comprendre les impacts que peut avoir la modélisation retenue. Pour un produit d'épargne, tous les postes du bilan économique rattachés aux engagements sont impactés. De nombreux mécanismes se cachent derrière ces impacts et une bonne compréhension de ces derniers permet de comprendre l'intérêt d'une modélisation fine des rachats.

3.3.1 Bilan économique

En changeant la loi de rachat, les flux de sortie futurs associés au contrat sont directement modifiés. Mais les flux d'entrée le sont également par le biais d'une assiette de participation aux bénéfices différente et donc des intérêts crédités différents. Le sens de l'impact est difficilement prévisible, mais les flux composant le BE sont ainsi sensibles à la modélisation.

Une fois le BE recalculé, si la RM est égale à un pourcentage conservatif de ce dernier, il est évident qu'elle sera impactée de la même manière. Toutefois, si une autre méthode est utilisée, l'impact se mesure plutôt en fonction de l'impact sur le SCR. Celui-ci est discuté dans les parties 3.3.2,3.3.3 et 3.3.4.

Le dernier élément au passif du bilan économique est les FP ; ils sont obtenus par différence car la valeur de marché des actifs est insensible à la loi de rachat. L'impact est donc une somme des impacts observés sur le BE et la RM.

3.3.2 SCR cessation

Il est important de comprendre que la modélisation du rachat n'aura pas d'impact particulier sur le SCR cessation. En effet, le changement de modélisation du rachat ne change en rien l'application de la formule standard. C'est donc bien le taux de rachat qui est modifié est non le choc.

L'impact sur le SCR cessation est donc comparable à celui sur le BE.

3.3.3 SCR marché

Le SCR marché va être quant à lui très sensible à la partie conjoncturelle de la loi de rachat. L'idée est d'appliquer un choc initial sur la valeur de marché des actifs. Consécutivement à ce choc, les rendements futurs sont diminués et donc le taux servi est diminué. La sensibilité au taux servi de la loi retenue est donc déterminante dans le calcul du SCR marché.

Le comportement en situations extrêmes de la loi de rachat est capté dans l'évaluation du SCR marché. C'est la composante conjoncturelle qui est sensible dans ce calcul.

3.3.4 SCR global

Finalement tous les modules du SCR sont impactés et le SCR marché particulièrement donc, par agrégation, le SCR global l'est aussi. Le BE est impacté et donc par différence, les FP le sont.

Le ratio de solvabilité est ainsi sensible au choix de la loi.

Le sens des mouvements de tous les éléments précédents est très difficile à évaluer a priori, mais il est claire que ces mouvements ne sont pas négligeables et que le choix de la loi de rachat est un élément déterminant dans la production des états Solvabilité II.

3.4 Problématique

Comme vu dans la partie 3.1.3, les rachats traduisent le comportement des assurés. La proposition des ONC tente d'expliquer le rachat uniquement par l'ancienneté fiscale du contrat et la différence entre taux servi et taux compétiteur. La première faiblesse des ONC est donc de ne prendre en compte que très peu de variables explicatives du rachat.

Une deuxième faiblesse a priori est la scission du rachat en deux parties. Le rachat observé est un tout indissociable et il est difficile de capter toute l'explication du rachat en le scindant. Cette séparation est proposée pour des soucis de compréhension et d'interprétation, mais une modélisation du rachat comme un tout et une définition a posteriori des deux composantes paraît être une démarche plus prudente.

Comme vu dans la partie 3.3 à la page 22, les impacts sont certains et nombreux mais difficilement prévisibles. Il est important de comprendre qu'un assureur ne cherche pas seulement à baisser son BE et son SCR, ni à augmenter son ratio de solvabilité. Le premier objet de l'étude est de proposer une modélisation du rachat qui permette à l'assureur d'être à l'aise vis-à-vis de son risque, de le maîtriser. Ainsi, que la modélisation améliore le ratio de solvabilité ou le dégrade, l'issue de l'étude doit être une amélioration de la gestion du risque.

Cependant, de nombreuses modélisations ont déjà été proposées dans différents mémoires :

i) Logique floue par S. Detroulleau et S. Mouret[S. Detroulleau and S. Mouret, 2013]
ii) Vasicek, Black & Scholes saisonnier et option financière par A. Suru[A. Suru, 2011]
iii) Chain-Ladder par A. Suru
iv) Analyse macroéconomique par A. Suru

La proposition de la logique floue est intéressante par la prise en compte du comportement humain et par la quantification de ce dernier. Elle introduit des nuances intéressantes dans les comportements des assurés et des assureurs, néanmoins le modèle est délicat à calibrer et se calibre presque exclusivement sur du jugement d'expert.

Les propositions de modèles financiers tels que Vasicek et Black & Scholes reposent entièrement sur l'historique de rachat et ne prennent pas en compte les variables économiques lors de la projection. Il est également difficile d'affiner la granularité avec ce type de modélisation.

La modélisation par triangles de Chain-Ladder s'inscrit dans la même dynamique mais sans aléa dans la projection ; elle suppose que le passé se répétera dans le futur.

La dernière modélisation par analyse macroéconomique est en fait un proposition de modélisation par régression logistique, mais seuls les rachats conjoncturels sont modélisés. La forme du

modèle et la prise en compte détaillée des variables explicatives sont avantageux.

Bien d'autres mémoires se sont intéressés à la modélisation des rachats, mais ceux qui ne sont pas ici cités se sont intéressés à la seule partie structurelle des rachats.

Cette étude s'essaye donc à la proposition d'un modèle qui réconcilie les deux composantes du rachat, qui prenne en compte les facteurs comportementaux et qui soit adapté à une utilisation dans un modèle de projection.

Deuxième partie

Cadre de l'analyse

Chapitre 4

Régression logistique

4.1 Modèle linéaire

Soit $n \in \mathbb{N}$. Soient X et Y deux variables aléatoires à support respectivement dans $\mathbb{R}^n$ et $\mathbb{R}$. Le problème de régression s'exprime comme un problème de prédiction de la variable aléatoire Y connaissant la variable aléatoire X. En voici un exemple :

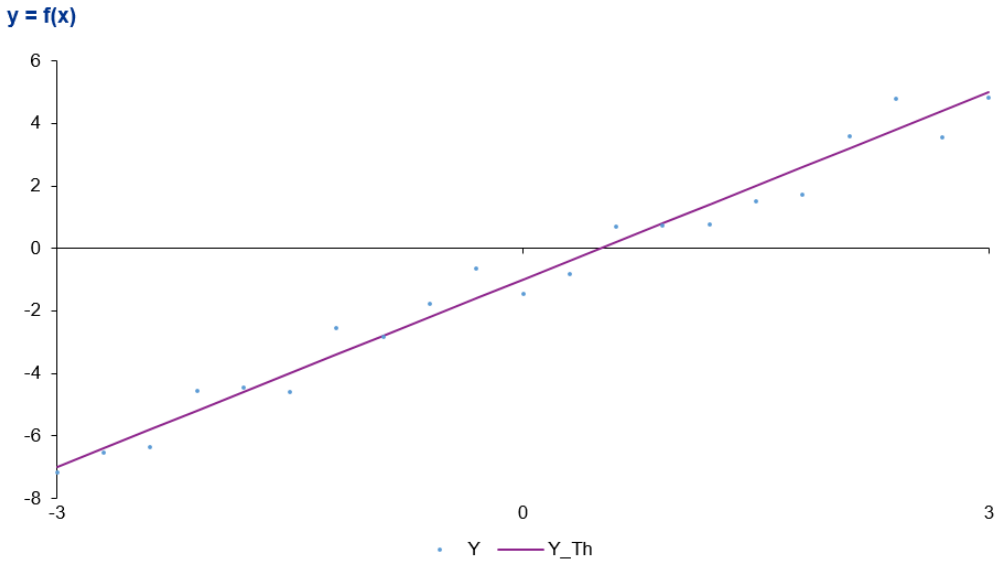

FIGURE 4.1 – Exemple de régression linéaire

Le principe, en dimension 1, est de trouver la droite qui passe au plus juste entre tous les points.

Certains éléments de la régression doivent être définis en amont de la résolution du problème.

Définition 4.1 (Régression).
Si Y est dans L^1 (i.e. $\mathbb{E}[|Y|] < \infty$), la régression de Y sur X est la fonction g définie par :

$$\begin{aligned} g : \quad \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \mathbb{E}[Y|X = x] \end{aligned}$$

Remarque 4.1.

i) Si X et Y sont indépendantes, alors la connaissance de X n'amène rien sur celle de Y et donc : $\mathbb{E}[Y|X = x] = \mathbb{E}[Y]$

ii) Si la fonction g est linéaire, la régression est dite linéaire.

Définition 4.2 (Résidus)**.**
Soit ξ la variable aléatoire à support dans $\mathbb{R}$ définie par : $\xi = Y - g(X)$. ξ représente l'erreur stochastique de l'approximation de Y par sa meilleure prévision sachant X, elle est appelée résidu de régression. Finalement :

$$Y = g(X) + \xi$$

Remarque 4.2.

$$\mathbb{E}[\xi|X] = \mathbb{E}[Y - g(X)|X] = \mathbb{E}[Y|X] - \mathbb{E}[g(X)|X] = g(X) - g(X) = 0 \text{ p.s.}$$

Le problème de régression consiste en la supposition d'une forme paramétrique de la fonction de régression g puis en la recherche des paramètres associés.

Soit $\beta \in \mathbb{R}^{n+1}$. La fonction g est supposée linéaire :

$$g(x) = \beta_0 + \sum_{i=1}^{n} \beta_i.x_i$$

Le problème de régression linéaire se résume donc au problème d'estimation des β. Le meilleur estimateur cherché est celui qui minimise l'erreur quadratique de prévision :

$$\hat{\beta} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}}\{\mathbb{E}[(Y - \beta_0 - \sum_{i=1}^{n} \beta_i.X_i)^2]\}$$

Le modèle peut être réécrit plus simplement en intégrant la constante au vecteur X. Soit $X = (1, X_1, ..., X_n)$. Le modèle s'écrit alors :

$$Y = \beta.X + \xi$$

Soient $(X^k; Y^k)_{k \in [\![1;N]\!]} \in (\mathbb{R}^{n+2})^N$ N observations indépendantes des variables aléatoires X et Y. Il est possible, par abus de notation, d'écrire Y le vecteur des Y^k et X la matrice dont les lignes sont les vecteurs X^k. Le vecteur $\hat{\beta}$ est donc tel que :

$$\hat{\beta} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}}\{f(\beta)\}$$

$$\text{où } f(\beta) = \sum_{k=1}^{N} (Y^k - \beta.X^k)^2$$

Il est bon de noter que si la matrice carrée $X'X$ est non dégénérée définie strictement positive, alors il existe une unique solution analytique à ce problème d'optimisation. Cette solution est appelée estimateur des moindres carrés et est donnée par :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

La preuve n'est pas rappelée ici puisque cet estimateur n'est pas utilisé dans l'étude. Cependant la présentation des modèles linéaires permet d'ouvrir sur une autre famille de modèles : les modèles linéaires généralisés.

4.2 Modèles linéaires généralisés

Les Modèles Linéaires Généralisés (MLG) sont définis à partir des fonctions de la famille exponentielle.

Définition 4.3 (Famille exponentielle)**.**
La famille exponentielle naturelle est l'ensemble $\mathcal{F}_\theta$ des fonctions $f_{\theta;\phi}$ telles que :

$$f_{\theta;\phi}(y) = exp\left[\frac{y\theta - b(\theta)}{\phi} + c(y;\phi)\right]$$

$$\text{où} \begin{cases} b \in \mathcal{D}^3 \\ b^{(1)} \text{ est inversible} \\ \theta \subset \mathbb{R} \\ \phi \subset \mathbb{R}_+^* \end{cases}$$

Remarque 4.3.
θ est appelé le paramètre naturel et ϕ le paramètre de dispersion.

Définition 4.4 (Fonction de lien)**.**
Soit $f_{\theta;\phi} \in \mathcal{F}_\theta$ une fonction de la famille exponentielle. La fonction de lien associée à $f_{\theta;\phi}$ est la fonction g définie par :

$$g(\mu) = (b^{(1)})^{-1}(\mu)$$

Définition 4.5 (MLG)**.**
Soit $\beta \in \mathbb{R}^{n+1}$. Soit g une fonction de lien. L'ensemble des modèles linéaires généralisés est l'ensemble des modèles de la forme suivante :

$$g(\mathbb{E}_\beta[Y|X = x]) = x'\beta$$

Remarque 4.4.
Les MLG généralisent le modèle linéaire dans le sens où ils forment un sous-ensemble des transformations bijectives du modèle linéaire.

Loi de $Y\|x$	Fonction de lien	Nom du lien
Bernoulli	$g(\mu) = logit(\mu) = log\left(\frac{\mu}{1-\mu}\right)$	Logit
Poisson	$g(\mu) = log(\mu)$	Logarithmique
Gamma	$g(\mu) = -\frac{1}{\mu}$	Réciproque
Gaussienne	$g(\mu) = \mu$	Identité
Bernoulli	$g(\mu) = F^{-1}_{\mathcal{N}(0;1)}(\mu)$	Probit
Log-normale	$g(\mu) = log(-log(1-\mu))$	Log-log

FIGURE 4.2 – Résumé des MLG et des fonctions de lien associées

4.3 Modèle logit

Cette section décrit la modélisation d'une variable aléatoire Y décrivant la réalisation ou non d'un évènement. De manière complètement équivalente, la variable Y est supposé à support $D_Y = \{0;1\}$: 1 correspond à la réalisation de l'évènement et 0 à la non réalisation. Le modèle linéaire n'est pas adapté au regard du support de Y. La principale critique du modèle linéaire est son ensemble d'arrivée. Pour modéliser une variable réponse sur $\{0;1\}$ il est incohérent d'utiliser un modèle dont l'ensemble d'arrivée est $\mathbb{R}$. Le graphique suivant montre bien la difficulté d'adaptation du modèle linéaire au cas binaire :

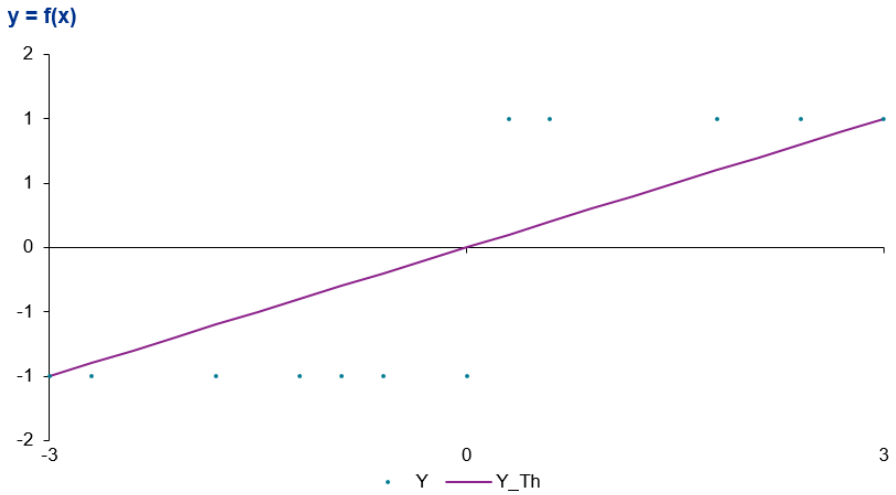

FIGURE 4.3 – Régression linéaire sur une variable binaire

Cependant le modèle linéaire peut être adapté à la situation par le biais d'une transformation appropriée. Finalement $Y|X = x$ est modélisé par une loi de Bernoulli dont le paramètre est une

transformation inversible d'un modèle linéaire.

La théorie des MLG propose plusieurs fonctions de lien de type sigmoïde particulièrement adaptées au cas présent car le lien est une fonction de $\mathbb{R}$ dans $[0;1]$.

Définition 4.6 (Modèle logit)**.**

Si la variable aléatoire $Y|X = x$ suit une loi de Bernoulli de paramètre $p_\beta(x) = \dfrac{e^{x'\beta}}{1+e^{x'\beta}}$, alors on dit que $Y|X$ suit un modèle logit.

Remarque 4.5.

i) Les liens probit et log-log sont tout aussi pertinents que lien logit cependant la symétrie et l'expression analytique de l'inverse du lien logit en font un candidat idéal

ii) Si X possède des composantes qualitatives, il suffit d'ajouter autant de variables explicatives que de modalités et qui valent 1 ou 0 selon la modalité prise

Maintenant que le modèle est exposé, il est nécessaire d'étudier la spécification de ce modèle.

La spécification du modèle logit passe par l'estimation des paramètres. Ici les paramètres se résument au vecteur β. A ce stade, plusieurs méthodes statistiques sont à disposition pour estimer les paramètres : méthode des moments, maximisation de la vraisemblance, minimisation de l'erreur quadratique, etc. La méthode retenue est celle de la maximisation de la vraisemblance car elle permet la mise en place d'un algorithme de type Newton facile à exprimer.

Définition 4.7 (Fonction de vraisemblance)**.**

La fonction de vraisemblance associée à un modèle $\{F_\beta; \beta \in \mathbb{R}^{n+1}\}$ est la fonction définie par :

$$L_N(\beta) = \prod_{i=1}^{N} f_\beta(x_i), \beta \in \mathbb{R}^{n+1}$$

où f_β est la densité du modèle

Dans le cadre du modèle logit, la densité de $Y|X = x$ est celle d'une loi de Bernoulli :

$$f_\beta(x) = p_\beta(x)^y.(1-p_\beta(x))^{1-y}, \beta \in \mathbb{R}^{n+1}$$

Cela donne finalement la fonction de vraisemblance suivante :

$$L_N(\beta) = \prod_{i=1}^{N} p_\beta(x_i)^{y_i}.(1-p_\beta(x_i))^{1-y_i}, \beta \in \mathbb{R}^{n+1}$$

Définition 4.8 (Fonction de log-vraisemblance)**.**

La fonction de log-vraisemblance associée à un modèle $\{F_\beta; \beta \in \mathbb{R}^{n+1}\}$ est une transformée de la vraisemblance :

$$\mathcal{L}_N(\beta) = -\frac{1}{N}\ln(L_N(\beta)), \beta \in \mathbb{R}^{n+1}$$

où f_β est la densité du modèle

Remarque 4.6.
En pratique, la log-vraisemblance est plus simple à manipuler car une somme est souvent plus facile à dériver qu'un produit.

La log-vraisemblance associée au modèle logit est donc par application du logarithme :

$$\begin{aligned}
-N.\mathcal{L}_N(\beta) &= \ln\left(\prod_{i=1}^{N} p_\beta(x_i)^{y_i}.(1-p_\beta(x_i))^{1-y_i}\right) \\
&= \sum_{i=1}^{N} y_i.\ln(p_\beta(x_i)) + (1-y_i).\ln(1-p_\beta(x_i)) \\
&= \sum_{i=1}^{N} y_i.\ln\left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right) + (1-y_i).\ln\left(1-\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right) \\
&= \sum_{i=1}^{N} y_i.(x_i'\beta - \ln(1+e^{x_i'\beta})) - (1-y_i).\ln(1+e^{x_i'\beta}) \\
&= \sum_{i=1}^{N} y_i x_i'\beta - \ln(1+e^{x_i'\beta})
\end{aligned}$$

Et donc finalement :

$$\mathcal{L}_N(\beta) = -\frac{1}{N}\sum_{i=1}^{N} y_i x_i'\beta - \ln(1+e^{x_i'\beta})$$

4.4 Calibrage

La maximisation de la vraisemblance permet d'obtenir un estimateur des paramètres. Cette maximisation est complètement équivalente à la minimisation de la log-vraisemblance. La première étape est donc d'exprimer les conditions aux premier et second ordre afin de vérifier l'existence d'un tel extremum.

La première étape est de calculer le gradient de la log-vraisemblance :

$$\begin{aligned}
\forall j \in [\![0;n]\!] \quad \frac{\partial \mathcal{L}_N(\beta)}{\partial \beta_j} &= -\frac{1}{N}\sum_{i=1}^{N} y_i x_{i;j} - \frac{x_{i;j}e^{x_i'\beta}}{1+e^{x_i'\beta}} \\
&= -\frac{1}{N}\sum_{i=1}^{N} x_{i;j}(y_i - p_\beta(x_i))
\end{aligned}$$

Il vient donc en écriture matricielle :

$$\begin{aligned}
\nabla \mathcal{L}_N(\beta) &= -\frac{1}{N}X'(Y-P_\beta) \\
\text{où } P_\beta &= (p_\beta(x_1);...;p_\beta(x_N))'
\end{aligned}$$

La deuxième étape est de calculer la Hessienne de la log-vraisemblance afin d'étudier sa convexité. Il est démontré en annexe A que la log-vraisemblance du modèle logit est convexe.

Maintenant que l'existence du maximum de vraisemblance est vérifiée, il reste à en trouver l'expression. Pour rappel, ce maximum est obtenu en annulant le gradient de la log-vraisemblance.

$$\nabla \mathcal{L}_N(\beta) = -\frac{1}{N} X'(Y - P_\beta) = 0$$

La difficulté est d'expliciter la solution de cette équation. A l'opposé du modèle linéaire, il n'est pas possible de donner de forme analytique de la solution.

Il est nécessaire de se tourner vers des méthodes numériques de résolution d'équation. Ici la fonction de log-vraisemblance est deux fois dérivable et convexe, un algorithme de type Newton-Raphson est particulièrement adapté.

Un autre algorithme de la même famille a été choisi : l'algorithme des moindres carrés itérativement repondérés (Iteratively Reweighted Least Squares (IRLS)). Il est possible de démontrer qu'un même algorithme est applicable à l'ensemble des MLG[Nelder and Wedderburn, 1972].

L'algorithme se décompose en trois étapes principales : l'initialisation, l'itération et la convergence. L'atteinte du minimum est un point à discuter. Il est ici considéré que l'algorithme a convergé si le vecteur β n'est pas significativement modifié lors d'une itération. Le niveau de significativité est laissé au choix de l'utilisateur mais un $\epsilon = 10^{-5}$ a été utilisé. Un nombre maximum d'itération est aussi fourni.

Algorithm 1 IRLS

Ensure: $\beta = \hat{\beta}^{\mathrm{MV}}$
  iterator $\leftarrow 0$
  $z \leftarrow g(y)$
  $\Omega \leftarrow \mathrm{diag}\left(\frac{1}{b"(\theta_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right)$
  $\beta \leftarrow (X'\Omega X)^{-1} X'\Omega z$
  $\beta_{\mathrm{old}} \leftarrow \beta + 2.\epsilon$
  while (iterator $<$ nbIterMax) && $(||\beta - \beta_{\mathrm{old}}||^2 > \epsilon)$ **do**
    $\beta_{\mathrm{old}} \leftarrow \beta$
    iterator $\leftarrow$ iterator $+ 1$
    $\mu \leftarrow g^{-1}(X\beta)$
    $\Omega \leftarrow \mathrm{diag}\left(\frac{1}{b"(\theta_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right)$
    $z \leftarrow X\beta + g'(\mu).(y - \mu)$
    $\beta \leftarrow (X'\Omega X)^{-1} X'\Omega z$
  end while

4.5 Pertinence dans le cadre du rachat

Le modèle logit permet de modéliser une variable binaire de réponse sachant un ensemble de variables explicatives. La décision de rachat est exactement une variable binaire. Il est possible d'associer à l'évènement de rachat la valeur 1 et 0 au non rachat.

Le point le plus pertinent a priori est la modélisation à partir de variables explicatives. Il a été dit précédemment que le rachat est une décision comportementale de l'assuré qui réagit face à l'environnement législatif, économique et à des besoins de liquidité personnels. Cette prise de décision de l'assuré passe donc naturellement par une observation des indicateurs de l'environnement. Il apparaît donc nécessaire de prendre en compte ces indicateurs dans la modélisation du rachat pour tenter de reproduire fidèlement le comportement de l'assuré.

De plus, ce type de modélisation a sa place au sein d'un modèle de projection comme l'outil DEvent présenté dans la partie 7 à la page 62. Ceci peut permettre d'améliorer la fiabilité de l'outil de projection et de gestion des risques.

Un dernier atout est la forme sigmoïde de l'espérance conditionnelle dans le modèle logit. En effet, la présentation des ONC du QIS 5 dans la partie 3.2 à la page 20 montre que la forme proposée pour la loi de rachats dynamiques ressemble à une sigmoïde affine par morceaux.

4.6 Limites a priori

Si la modélisation par régression logistique possède des atouts de taille, elle possède également des points faibles a priori.

D'un point de vue statistique, la méthode du maximum de vraisemblance requiert l'hypothèse d'indépendance des observations. Or l'indépendance n'est a priori pas vérifiée : l'exemple de rachats massif en cas de moins-value latente globale va baisser le rendement de l'assureur et donc potentiellement induire d'autres rachats. Dans cet exemple l'hypothèse d'indépendance n'est pas valide.

D'autre part, seule la décision de rachat total est modélisée et non le rachat partiel. Cependant la modélisation du rachat total peut être réutilisée dans le cadre du rachat partiel en intégrant la modélisation du taux de rachat.

De plus la régression logistique est un MLG et suppose donc un comportement pseudo-linéaire du rachat devant les variables explicatives. La simple variable d'ancienneté montre directement le biais de ce type d'hypothèse : deux pics de rachats sont observés à 4 ans et 8 ans d'ancienneté sur les produits d'épargne en France et une dépendance d'ordre 3 paraît plus appropriée dans cette situation. D'autres variables connues pour leur incidence sur le rachat induisent la même critique.

Une dernière faiblesse a priori du modèle logit est le choix des variables explicatives qui peut rapidement devenir une limite du modèle. Il est important d'arbitrer entre un nombre de variables suffisant pour expliquer au mieux le rachat et assez faible pour garder un pouvoir d'analyse et des temps de calcul raisonnables. La présentation du GSE dans la partie 8 à la page 68 montrera aussi une limite dans le choix des variables économiques mais cette fois la limite est liée au GSE et non au modèle logit.

Chapitre 5

Régression non paramétrique par projection

5.1 Apports du non paramétrique

La régression logistique repose sur une dépendance pseudo-linéaire vis-à-vis des variables explicatives. Plusieurs exemples permettent a priori de réfuter cette hypothèse dans le cadre du rachat. Par exemple, toutes choses égales par ailleurs, l'ancienneté n'a pas une incidence linéaire sur le taux de rachat (l'âge de l'assuré est un autre exemple pertinent). L'exemple est développé pour comprendre l'intérêt d'une amélioration de la modélisation.

Soit un ensemble d'assurés dans un environnement économique stable et dont les caractéristiques sont identiques (hors ancienneté). Il est possible d'obtenir une courbe de rachat en fonction de l'ancienneté du contrat : la tendance de cette courbe est assez connue des assureurs en raison des différents paliers fiscaux que propose l'assurance vie en France. Elle se décompose en quatre parties :

i) Une phase de rachat faible et stable en raison de la forte imposition
ii) Un pic de rachat au bout de 4 ans en raison du premier pallier fiscal
iii) Un pic de rachat au bout de 8 ans en raison du second et dernier pallier fiscal
iv) Une baisse du rachat jusqu'à extinction du portefeuille

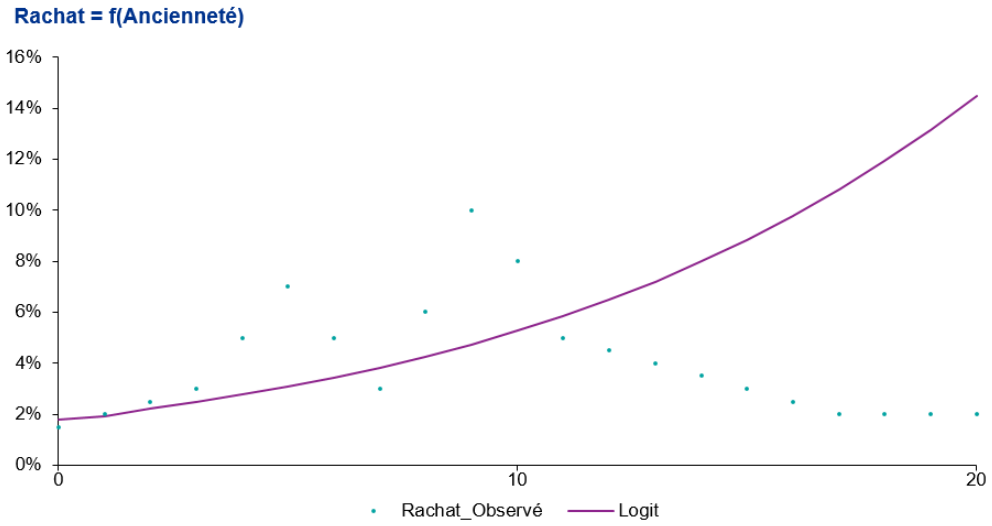

FIGURE 5.1 – Données et comportement du modèle logit en fonction de l'ancienneté

Ce graphique montre une faiblesse de cette proposition de modélisation. Il apparaît donc nécessaire de se soustraire à l'hypothèse de linéarité dans la fonction logit.

Deux approches classiques sont envisageables dans ce cadre : une approche paramétrique non linéaire oet une approche non paramétrique. Dans les deux cas, le modèle logit est réécrit comme :

$$g(\mathbb{E}_\beta[Y|X=x]) = f(x)$$

où f est une fonction à estimer

L'approche paramétrique non linéaire revient à supposer que la fonction f s'écrit comme une fonction d'un paramètre β et dont la forme est supposée connue. Par exemple, pour la dépendance vis-à-vis de l'ancienneté, une forme polynomiale de degré 3 peut être envisagée.

La faiblesse de ce type d'approche réside dans le choix de la forme paramétrique. En effet ce type de décision repose sur le jugement d'expert et n'est pas générique face aux données. Autrement dit, pour deux jeux de données différents, deux formes paramétriques différentes peuvent être retenues en raison des différences entre les observations. De plus l'ajout d'une variable explicative nécessiterait une modification du code et présenterait un risque opérationnel fort.

Une autre approche possible est l'approche non paramétrique. Cette approche allie deux avantages forts : elle est générique et elle ne suppose aucune forme a priori.

Il existe différentes méthodes d'estimation non paramétriques et le choix effectué dans cette étude est la méthode de régression par projection. Ce choix se justifie par la forme même de la méthode étant donné l'utilisation finale au sein du modèle logit. In fine, le calibrage du modèle hybride ne diffère en rien de celle présentée pour le modèle logit classique dans la partie 4.4 à la page 31.

5.2 Présentation de la régression par projection

La théorie ici développée s'appuie sur les ouvrages de A.B. Tsybakov[Tsybakov, 2003] et de T. Hastie, R. Tibshirani et J. Friedman[Hastie T., Tibshirani R., Friedman J., 2009].

Soient X et Y deux variables aléatoires à support respectivement dans $\mathbb{R}^n$ et $\mathbb{R}$. Le problème ici étudié est celui de la régression exprimé dans la partie 4.1 à la page 26.

Une première hypothèse restrictive est faite dans l'estimation par projection : X est supposée de loi uniforme sur $[0;1]^n$. Il est possible d'adapter les données en les normalisant.

Soient $(X^k;Y^k)_{k\in[\![1;N]\!]} \in ([0;1]^n)^N$ N observations indépendantes des variables aléatoires X et Y.

Une seconde hypothèse restrictive est réalisée : la fonction de régression f est supposée dans $L^2([0;1]^n)$.

L'espace $L^2([0,1]^n)$ est un espace vectoriel. Soit $(\varphi_j)_{1\leqslant j}$ une base orthonormale de cet espace. La fonction f admet une unique décomposition dans cette base. Elle s'exprime donc comme :

$$f = \sum_{j=1}^{\infty} \theta_j \varphi_j$$

$$\text{où} \quad \begin{cases} \forall j \in \mathbb{N}\ \theta_j = \langle f|\varphi_j\rangle = \int\limits_{[0,1]^n} g\varphi_j \\ \sum\limits_{j=1}^{\infty} \theta_j^2 < \infty \end{cases}$$

L'idée de l'estimation non paramétrique par projection est d'exploiter cette décomposition dans une base orthonormale afin de remplacer l'estimation de f par l'estimation des coefficients dans la cette base.

Cependant, l'espace $L^2([0,1]^n)$ étant un espace hilbertien et non euclidien, il est de dimension infinie et la base orthonormale est une base hilbertienne. Il est en fait impossible d'estimer numériquement une famille de coefficients dont le cardinal est infini.

L'estimation de f par projection repose finalement sur une approximation de f par une fonction f_M définie par :

$$f_M = \sum_{j=1}^{M} \theta_j \varphi_j$$

M est appelé niveau de troncature puisqu'il correspond à une troncature de la série. Tout le travail repose in fine sur le choix de ce niveau de troncature afin de garder une approximation de f de qualité suffisante.

Le modèle s'apparente alors à un modèle de régression linéaire.

5.3 Calibrage

5.3.1 Modèle approché

Le modèle approché est d'abord considéré :

$$Y = f_M(X) + \xi$$

Soit ϕ_M la matrice de $\mathcal{M}_{N\text{x}M}(\mathbb{R})$ telle que :

$$\begin{bmatrix} \varphi_1(X_1) & \varphi_2(X_1) & ... & \varphi_M(X_1) \\ \varphi_1(X_2) & & & \\ & & (...) & \\ & & & \\ \varphi_1(X_N) & \varphi_2(X_N) & ... & \varphi_M(X_N) \end{bmatrix}$$

Soit θ_M le vecteur $(\theta_1; ...; \theta_M)'$.

Afin de calculer l'estimateur des moindres carrés ordinaire de θ_M, la matrice $\phi'_M\phi_M$ est supposée non dégénérée strictement définie positive.

Remarque 5.1.
La matrice $\phi'_M\phi_M$ est toujours symétrique et semi-définie positive.

Comme vu dans la partie 4.1 à la page 26, l'estimateur des moindres carrés ordinaire s'écrit :

$$\hat{\theta}_M^{\text{MC}} = (\phi'_M\phi_M)^{-1}\phi'_M Y$$

Deux hypothèses supplémentaires sont réalisées :

$$\mathbb{E}[\xi] = 0$$

$$\mathbb{V}[\xi] = \sigma^2 I_M$$

Les variables X et ξ sont supposées indépendantes.

Sous ces hypothèses, deux propriétés statistiques de l'estimateur des moindres carrés viennent immédiatement :

$$\mathbb{E}[\hat{\theta}_M^{\text{MC}}] = \theta$$

$$\mathbb{V}[\hat{\theta}_M^{\text{MC}}] = \sigma^2(\phi'_M\phi_M)^{-1}$$

En notant $\phi_M(x) = (\varphi_1(x); ...; \varphi_M(x))'$, il vient immédiatement par linéarité de l'espérance que :

$$\forall x \in [0;1]^N \ \mathbb{E}[\hat{\theta}_M^{\text{MC}'}\phi_M(x)] = \mathbb{E}[\hat{\theta}_M^{\text{MC}}]'\phi_M(x) = \theta'_M\phi_M(x) = g_M(x)$$

Il est possible d'en déduire un estimateur sans biais de $f_M(x)$:

$$\hat{f}_M(x) = \hat{\theta}_M^{\mathrm{MC}'} \phi_M(x)$$

Soit $A_M = \phi_M(\phi'_M\phi_M)^{-1}\phi'_M$. Il vient avec cette notation :

$$(\hat{f}_M(X_1); ...; \hat{f}_M(X_N))' = \phi_M \hat{\theta}_M^{\mathrm{MC}'} = A_M Y$$

Proposition 5.1.
La matrice A_M est un projecteur orthogonal.

Démonstration.
La démonstration est faite dans l'annexe B.1 à la page 108. □

Proposition 5.2.
La matrice $B_N = \frac{1}{N}\phi'_M\phi_M$ converge vers la matrice identité lorsque $N \to \infty$.

Démonstration.
Le produit matriciel donne pour chaque élément de B_N :

$$\forall (i;j) \in [\![1;M]\!]^2 \; B_N(i;j) = \frac{1}{N}\sum_{k=1}^{N} \varphi_i(X^k)\varphi_j(X^k)$$

Or les variables X^k sont supposées i.i.d. de loi uniforme sur $[0;1]^n$ (une variable générique X de même loi peut être introduite). Comme φ est une base orthonormale de L^2, il vient que quelque soit $(i;j) \in [\![1;M]\!]^2$ et quelque soit $k \in [\![1;N]\!]$, les variables $\varphi_i(X^k)\varphi_j(X^k)$ sont de carrés intégrables et donc intégrables.

La loi des grands nombres assure donc que :

$$\forall (i;j) \in [\![1;M]\!]^2 \; B_N(i;j) = \frac{1}{N}\sum_{k=1}^{N} \varphi_i(X^k)\varphi_j(X^k) \xrightarrow[N\to\infty]{} \mathbb{E}[\varphi_i(X)\varphi_j(X)] \text{ p.s.}$$

Or, par expression de l'espérance et propriété de la base orthonormale :

$$\forall (i;j) \in [\![1;M]\!]^2 \; \mathbb{E}[\varphi_i(X)\varphi_j(X)] = \int_{[0,1]^n} \varphi_i(x)\varphi_j(x).1\mathrm{d}x = \delta_{i;j}$$

Il vient donc en écriture matricielle :

$$B_N \xrightarrow[N\to\infty]{} I_M$$

□

Il est donc possible d'introduire cette approximation de $\phi'_M\phi_M$ dans l'expression de $\hat{f}_M(x)$ afin d'obtenir un second estimateur de $g_M(x)$:

$$\widetilde{f}_M(x) = (N.I_M)^{-1}\phi'_M Y\phi_M(x) = \frac{1}{N}\phi'_M Y\phi_M(x)$$

Proposition 5.3.
L'estimateur par la méthode des moments de θ est :

$$\forall j \in [\![1; M]\!]\ \widetilde{\theta}_j = \frac{1}{N}\sum_{k=1}^{N} Y_k\varphi_j(X_k)$$

Démonstration.
Soit $j \in [\![1; M]\!]$. Le calcul de l'espérance

$$\mathbb{E}[Y.\varphi_j(X)] = \mathbb{E}[f_M(X).\varphi_j(X) + \xi.\varphi_j(X)] = \mathbb{E}[f_M(X).\varphi_j(X)] + \mathbb{E}[\xi.\varphi_j(X)]$$

Or, par indépendance des variables X et ξ, il vient :

$$\mathbb{E}[Y.\varphi_j(X)] = \mathbb{E}[f_M(X).\varphi_j(X)] + \mathbb{E}[\xi].\mathbb{E}[\varphi_j(X)] = \mathbb{E}[f_M(X).\varphi_j(X)]$$

Finalement la substitution de f_M par son expression donne :

$$\mathbb{E}[Y.\varphi_j(X)] = \sum_{k=1}^{N} \theta_k.\mathbb{E}[\varphi_i(X).\varphi_j(X)]$$

Et comme vu dans la démonstration de la propriété 5.2 à la page 38 :

$$\mathbb{E}[Y.\varphi_j(X)] = \sum_{k=1}^{N} \theta_k.\delta_{j;k} = \theta_j$$

La méthode des moments donne donc finalement l'estimateur suivant :

$$\widetilde{\theta}_j = \frac{1}{N}\sum_{k=1}^{N} Y_k\varphi_j(X_k)$$

Il suffit finalement de généraliser sur j. □

Remarque 5.2.

i) Le calcul de cet estimateur nécessite que la matrice $\phi'_M\phi_M$ soit inversible
ii) Mais il a l'avantage de ne demander aucune inversion in fine et donc permet d'éviter les problèmes numériques de mauvais conditionnement de matrice
iii) En plus d'éviter les problèmes numériques, l'algorithme de calcul sera plus rapide sans inversion (procédé souvent chronophage)
iv) L'hypothèse d'uniformité de X est nécessaire à l'obtention de cet estimateur, mais elle n'est en rien nécessaire au calcul de celui des moindres carrés

Cet estimateur des moments permet d'écrire :

$$\widetilde{f}_M(x) = \sum_{j=1}^{M} \widetilde{\theta}_j.\varphi_j(x)$$

Maintenant que le modèle approché est connu et que des estimateurs de la fonction de régression sont exprimés, un retour au modèle général est nécessaire pour mesurer la qualité de l'approximation.

5.3.2 Modèle général

Soit le modèle général :

$$Y = f(X) + \xi$$

Le premier indicateur de qualité d'un estimateur est son biais. Soit $b_M(x)$ le biais de l'estimateur $\widetilde{f}_M(x)$. Il vient alors que :

$$b_M(x) = \mathbb{E}[f(X) - \widetilde{f}_M(x)] = \mathbb{E}\left[\sum_{j=1}^{M}(\theta_j - \widetilde{\theta}_j).\varphi_j(x) + \sum_{j=M+1}^{\infty} \theta_j.\varphi_j(x)\right]$$

Proposition 5.4.

$$b_M(x) \xrightarrow[M\to\infty]{} 0$$

Démonstration.

Comme, par définition de la base hilbertienne, la convergence de la séries $\sum_{j=1}^{\infty} \theta_j\varphi_j$ a lieu dans L^2, il est possible d'intervertir intégrale et somme et donc espérance et somme :

$$b_M(x) = \sum_{j=1}^{M} \mathbb{E}[(\theta_j - \widetilde{\theta}_j).\varphi_j(x)] + \sum_{j=M+1}^{\infty} \mathbb{E}[\theta_j.\varphi_j(x)]$$

Or, comme vu précédemment, l'estimateur par la méthode des moments est sans bais, donc :

$$b_M(x) = \sum_{j=M+1}^{\infty} \mathbb{E}[\theta_j.\varphi_j(x)] = \sum_{j=M+1}^{\infty} \theta_j.\varphi_j(x)$$

Comme la série $\sum_{j=1}^{\infty} \theta_j\varphi_j$ converge dans L^2, la suite des restes de cette série converge vers 0 et donc :

$$b_M(x) \xrightarrow[M\to\infty]{} 0$$

□

Remarque 5.3.

L'étude du modèle pourrait presque s'arrêter ici et le niveau de troncature pourrait être choisi en calculant les erreurs quadratiques pour les premières valeurs et le minimum serait gardé. Cependant une méthode de choix du niveau de troncature plus économe en calcul aurait sa place dans la mise en pratique du modèle.

Première majoration de l'erreur quadratique intégrée

Définition 5.1 (Erreur quadratique intégrée)**.**

L'erreur quadratique intégrée associée à l'estimateur $\widetilde{f}_M$ est définie par :

$$R(\widetilde{f}_M, f) = \mathbb{E}\left[\left\|f - \widetilde{f}_M\right\|^2\right] = \mathbb{E}\left[\int_{[0;1]^n} (f - \widetilde{f}_M)^2\right]$$

La première étape vers l'expression de $R(\widetilde{f}_M, f)$ est le calcul de la partie intégrale :

$$\begin{aligned}\int_{[0;1]^n} (f - \widetilde{f}_M)^2 &= \int_{[0;1]^n} (f - \widetilde{f}_M)^2 \\ &= \int_{[0;1]^n} \left(\sum_{j=1}^{M} (\theta_j - \widetilde{\theta}_j).\varphi_j + \sum_{j=M+1}^{\infty} \theta_j.\varphi_j\right)^2\end{aligned}$$

La convergence dans L^2 permet de développer le carré et d'intervertir les signes somme et intégrale. De plus, la définition d'une base orthonormale permet d'annuler les termes croisés (doubles sommes) et il vient alors :

$$\int_{[0;1]^n} (f - \widetilde{f}_M)^2 = \sum_{j=1}^{M} (\theta_j - \widetilde{\theta}_j)^2 + \sum_{j=M+1}^{\infty} \theta_j^2$$

En passant à l'espérance et en inversant les signes somme et espérance grâce à la convergence L^2, il est possible de réécrire cette égalité comme :

$$\begin{aligned}R(\widetilde{f}_M, f) &= \sum_{j=1}^{M} \mathbb{E}[(\theta_j - \widetilde{\theta}_j)^2] + \sum_{j=M+1}^{\infty} \mathbb{E}[\theta_j^2] \\ &= \sum_{j=1}^{M} \mathbb{V}[\widetilde{\theta}_j] + \sum_{j=M+1}^{\infty} \theta_j^2\end{aligned}$$

Soit $j \in [\![1; M]\!]$. Il est possible de majorer la variance de $\widetilde{\theta}_j$ par :

$$V[\widetilde{\theta}_j] \leqslant \frac{1}{N}.(\|f\|_\infty^2 + \sigma^2)$$

Ce résultat est démontrer dans l'annexe B.2 à la page 109.

En réintégrant ce résultat, il vient finalement la majoration du risque quadratique intégré suivante :

$$R(\widetilde{f}_M, f) \leqslant \frac{M}{N}.(\|f\|_\infty^2 + \sigma^2) + \sum_{j=M+1}^{\infty} \theta_j^2$$

Cette majoration, sous condition de connaître à l'avance certaine propriétés de la fonction f, permet de choisir par approximation un niveau de troncature $\widetilde{M}^{\text{Opt1}}$ optimal. Ce niveau optimum est donné par :

$$\widetilde{M}^{\text{Opt1}} = \underset{M\in\mathbb{N}^*}{\operatorname{argmin}} \left\{ \frac{M}{N}.(\|f\|_\infty^2 + \sigma^2) + \sum_{j=M+1}^{\infty} \theta_j^2 \right\}$$

Remarque 5.4.
Parmi les propriétés à connaître, la norme infinie est la moins contraignante. En pratique, un majorant est souvent connu et l'inégalité peut être adaptée. Cependant la connaissance des restes de la série θ est bien plus difficile à connaître. Finalement s'ils sont connus d'avance, il est légitime de supposer que toute la fonction f est connue.

Cette première majoration paraît donc difficilement utilisable en pratique, du moins dans la majeure partie des cas.

Deuxième majoration de l'erreur quadratique intégrée

Deux hypothèses sont réalisées pour obtenir un autre majorant du risque quadratique intégré :

i) f est bornée par un réel F

ii) $\exists l \in \mathbb{N}^* \ \exists L \in \mathbb{R}_+^* \ \sum_{j=1}^{\infty} j^{2l}.\theta_j^2 \leqslant L$

Soient l et L ainsi choisis.

Quelque soit le niveau de troncature choisi M, ce dernier et toujours supérieur à 1 et donc strictement positif, donc sous les hypothèses précédentes :

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{j}{M}^{2l}.\theta_j^2 &\leqslant L.M^{-2l} \\ \sum_{j=M+1}^{\infty} \frac{j}{M}^{2l}.\theta_j^2 &\leqslant L.M^{-2l} \end{aligned}$$

Or : $\forall j > M \frac{j}{M} \leqslant 1$.

Soit : $\forall j > M \frac{j}{M}^{2l} \leqslant 1$.

Ainsi, par inégalité et comme la série θ converge bien dans L^2 :

$$\sum_{j=M+1}^{\infty} \theta_j^2 \leqslant L.M^{-2l}$$

Il vient alors par substitution dans le premier majorant du risque quadratique :

$$R(\widetilde{f}_M, f) \leqslant M.\frac{1}{N}.(F^2+\sigma^2) + L.M^{-2l}$$

Soit r la fonction définie par :

$$\begin{array}{rccl} r: & \mathbb{N}^* & \to & \mathbb{R}_+ \\ & x & \mapsto & M.\frac{1}{N}.(F^2+\sigma^2) + L.M^{-2l} \end{array}$$

De la même manière que pour $\widetilde{M}^{\mathrm{Opt}1}$, il est possible de définir $\widetilde{M}^{\mathrm{Opt}2}$ tel que :

$$\widetilde{M}^{\mathrm{Opt}2} = \underset{M\in\mathbb{N}^*}{\operatorname{argmin}}\{r(M)\}$$

La condition du premier ordre s'écrit :

$$r'(M) = \frac{1}{N}.(F^2+\sigma^2) - 2l.L.M^{-2l-1} = 0$$

Et la condition du second ordre s'écrit :

$$r''(M) = 2l.(2l+1).L.M^{-2l-2} \geqslant 0$$

La condition du second ordre est toujours vraie car tous les termes sont positifs, la fonction r est donc strictement convexe et admet un unique minimum. Ceci assure l'existence de $\widetilde{M}^{\mathrm{Opt}2}$. Ce dernier se calcul grâce à la condition du premier ordre :

$$\begin{aligned} r'(\widetilde{M}^{\mathrm{Opt}2}) &= 0 \\ \frac{1}{N}.(F^2+\sigma^2) - 2l.L.(\widetilde{M}^{\mathrm{Opt}2})^{-2l-1} &= 0 \\ 2l.L.(\widetilde{M}^{\mathrm{Opt}2})^{-2l-1} &= \frac{1}{N}.(F^2+\sigma^2) \\ \widetilde{M}^{\mathrm{Opt}2} &= \left[\frac{1}{N.2l.L}.(F^2+\sigma^2)\right]^{2l+1} \end{aligned}$$

Remarque 5.5.

Là encore plusieurs caractéristiques de la fonction f doivent être connues d'avance pour pouvoir appliquer la méthode. Cette fois la norme infinie est directement remplacée par un majorant mais il est remarquable que plus ce majorant est grand plus le niveau de troncature retenu est grand. De plus la série des résidus n'a plus besoin d'être connue et est remplacée par la connaissance de l et L. En revanche, si ces deux éléments représentent beaucoup moins d'information que la série complète, ils restent difficiles à obtenir en pratique.

Minimisation pénalisée de l'erreur quadratique intégrée

Proposition 5.5.

S'il existe un entier $M_0 < N$ tel que le vecteur $(f(X_k); k \in [\![1; N]\!])'$ appartienne à l'espace vectoriel engendré par les colonnes de ϕ_{M_0}, alors un estimateur sans biais de σ^2 est défini par :

$$\hat{\sigma}^2_{M_0} = \frac{1}{N-M_0}.\|(I_N - A_{M_0})Y\|^2$$

Démonstration.
La démonstration est faite dans l'annexe B.3 à la page 110.

□

La connaissance d'un estimateur sans biais de la variance permet de proposer une troisième méthode pour le choix du niveau de troncature. Ainsi une méthode proposée[A. Dalalyan, 2008] est de minimiser le risque quadratique en utilisant un estimateur de ce dernier.

Cependant l'estimateur du risque quadratique est biaisé. Une méthode consiste donc à remplacer le programme d'optimisation du risque quadratique par un programme pénalisé.

Un M_0 maximum est d'abord choisi. Il représente la limite supérieure du niveau de troncature acceptable du point de vue de l'effort de calcul. Il est tout de même nécessaire d'avoir $N \ll M_0$. La variance est ensuite supposée égale à $\hat{\sigma}^2_{M_0}$. Le nouveau programme d'optimisation donne donc :

$$\widetilde{M}^{\mathrm{Opt3}} = \underset{M\in[\![1;M_0]\!]}{\operatorname{argmin}} \left\{ \|(I_N - A_M)Y\|^2 - (N - 2.M).\hat{\sigma}^2_{M_0} \right\}$$

Remarque 5.6.
Cette optimisation présente l'avantage de ne demander aucune connaissance sur la fonction f mais elle demande M_0 calculs et inversions de matrices. Sans une maîtrise de ces deux points, la méthode peut rapidement devenir chronophage. Elle est également peu robuste devant les problèmes mal conditionnés.

Utilisation de la vitesse de convergence

Une quatrième et dernière méthode proposée repose sur l'utilisation de la vitesse de convergence de l'estimation des coefficients[D. Bosq, 2002].

Un seuil de significativité γ_N des coefficients est laissé au choix de l'utilisateur. Ce seuil s'exprime comme :

$$\gamma_N = \min_{c\in\mathbb{R}^*_+} \left\{ c.\sqrt{\frac{\ln(N)}{N}}; 1 \right\}$$

Le terme $\sqrt{\frac{\ln(N)}{N}}$ représente la vitesse de convergence acceptée pour les estimateurs des coefficients $\widetilde{\theta}$.

Un niveau de troncature maximum M_0 est également choisi a priori par l'utilisateur. Ce choix doit être fait en tenant compte des contraintes temporelles et physiques de la machine.

Le niveau de troncature optimal est ensuite choisi selon la règle suivante :

$$\widetilde{M}^{\mathrm{Opt4}} = \underset{M\in[\![1;M_0]\!]}{\operatorname{argmax}} \left\{ |\widetilde{\theta}_j| \geqslant \gamma_N \right\}$$

Le graphique suivant illustre la situation avec $M_0 = 20$ et $\widetilde{M}^{\mathrm{Opt4}} = 12$:

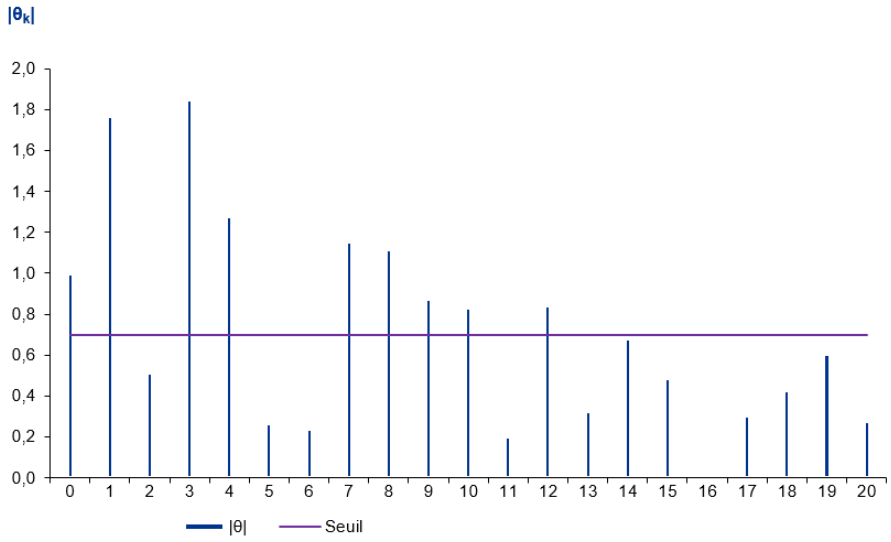

FIGURE 5.2 – Exemple d'utilisation de la vitesse de convergence

5.4 Intégration au modèle logit

Pour pallier les contraintes linéaires du modèle logit classique, un modèle hybride s'inspirant de la structure du logit et de la souplesse de l'estimation par projection est proposé.

En effet, un taux de rachat est par définition un élément de $[0;1]$, or la théorie développée sur l'estimation par projection peut être fragilisée par ce point : il n'y a aucune raison pour que la variable réponse soit dans cet intervalle.

De plus, c'est la forme sigmoïdale du modèle logit qui a amené à le choisir dans un premier temps puisqu'elle est en accord avec les ONC et facilite la compréhension du phénomène de rachat.

L'avantage de l'estimation par projection est sa forme finale qui ressemble fortement à celle d'un modèle linéaire. La connexion entre les deux modèles est donc facilement concevable.

L'idée revient donc à écrire le modèle hybride suivant :

$$g(\mathbb{E}_\beta[Y|X=x]) = f(x)$$

$$\text{où} \begin{cases} f \text{ est une fonction à estimer} \\ g = \text{logit}^{-1} \end{cases}$$

Là deux hypothèses sont possibles : soit la fonction f est supposée suivre le modèle approché, soit le modèle général.

Si elle suit le modèle approché, alors le niveau de troncature et la base doivent être choisis par l'utilisateur. L'estimation se fait ensuite comme pour le modèle logit classique : l'estimateur des moindres carrés ordinaire étant un estimateur sans biais des coefficients, l'algorithme IRLS permet de conclure.

Finalement le modèle approché s'apparente à une simple transformation de la matrice X par le passage dans une base hilbertienne tronquée.

Si le modèle général est considéré, une méthode de choix du niveau de troncature doit être choisie. Dans le cadre du modèle hybride, c'est la quatrième méthode qui est retenu car elle permet de se rapprocher du modèle approché tout en choisissant le niveau de troncature par seuil de significativité et donc laisse moins de place au jugement d'expert.

Cependant l'approche par projection simple mérite d'être testée sur le cas du rachat et reste un outil actuariel puissant, facile à mettre en œuvre et généralisant les modèles linéaires.

5.5 Limites a priori

L'ajout de la régression non paramétrique par projection pallie à plusieurs faiblesses du modèle logit classique.

Le modèle classique est a priori fragile avec une hypothèse forte de linéarité et donc un manque de liberté dans la modélisation. Le modèle hybride donne cette fois un grand nombre de libertés supplémentaires et parfois même un trop grand nombre.

Le premier choix capital est celui de la base. La partie 5.3 à la page 37 montre qu'en théorie, quelque soit la base hilbertienne orthonormale choisie, le modèle converge. En pratique, la convergence dépend de l'adéquation entre la base et la forme réelle de la fonction à estimer. Un point de vue métier et une analyse a priori des données peut donner une intuition sur le choix juste.

La base trigonométrique est un exemple usuel de base de $L^2([0;1])$: c'est une combinaison de sinus et de cosinus. Si la fonction à estimer est affine, le modèle est clairement inadapté et même s'il converge en théorie, il sera bien moins efficace que la version classique du modèle logit.

De plus la théorie est mise à mal par les contraintes physiques de la machine : la nécessité de choisir un niveau de troncature apparaît. La convergence des estimateurs reste vérifiée, mais un biais est induit par cette approximation.

Le choix du niveau de troncature amène dès lors un deuxième degré de liberté. Quatre méthodes ont été exposées pour le choix de ce niveau de troncature mais chacune amène son lot d'hypothèses sur la fonction à estimer. Des compromis sont à réaliser et la difficulté de la méthode d'estimation par projection repose sur la capacité à choisir la méthode la plus adaptée au problème étudié.

Chapitre 6

Modèle à états et SVM

6.1 Besoin d'un modèle à état

Les modèles proposés pour la modélisation des rachats possèdent une limite commune : ils supposent tous que la sensibilité de l'assuré est constante et que son comportement est et sera toujours rationnel.

Un constat macroéconomique met en péril cette hypothèse : les situations de crises économiques montrent que les corrélations peuvent s'inverser et que la sensibilité des assurés aux variables économiques peut être modifiée.

Les ONC[ACPR, 2013] montrent également une prise en compte de ce phénomène. Les zones de différence extrêmes entre le taux servi et le taux attendu sont caractérisées par une indifférence aux variables économiques et des taux de rachats dynamiques constants. En effet, lors de crise économiques majeures, les assurés ne cherchent plus à faire fructifier leur épargne mais plutôt à la sauvegarder. Dans ce cas le comportement est indifférent à un mouvement restreint des conditions économiques.

En cas de situation favorable, un comportement semblable est observable : plus ou moins 1% de rendement n'est pas identique entre un niveau moyen à $0,5\%$ et un niveau moyen à 5%.

De plus, en cas de situation de crise avérée, l'environnement médiatique se fait plus présent, les assurés connaissent parfois une sur-repréentation médiatique de la crise et les comportements tendent à devenir plus risquophobes.

Ces constats a priori laissent penser qu'un modèle à états permettrait de capter les différences de comportement en situation de crise. Ainsi différents calibrages de la loi selon les états économiques permettrait d'affiner la modélisation du rachat.

L'utilisation des Support Vector Machines (SVM) est ici envisagé pour la reconnaissance des états du monde.

6.2 Présentation des SVM

La théorie des SVM a été introduite pour la première fois par Vapnik en 1982[V. Vapnik, 1982].

Le problème initial qui a conduit au développement des SVM est celui de la classification binaire. Ce dernier consiste en la prévision d'une variable aléatoire binaire y dont les valeurs possibles sont dans $\{-1;1\}$. Cette prévision se fait par la construction d'une règle de décision D à partir de la réalisation d'un vecteur aléatoire x à support dans $\mathbb{R}^p$ qui est supposer influencer la valeur de y. Un échantillon d'apprentissage est disponible (ensemble de réalisations des couples $(x;y)$) et l'estimation de la règle de décision est effectuée sur cet échantillon. L'idée est de minimiser la probabilité d'erreur $\mathbb{P}(D(x) \neq y|x)$.

L'innovation des SVM, par rapport aux approches type logit, est de construire une règle de décision géométrique, et non probabiliste. La règle consiste en la caractérisation d'une frontière qui sépare les points au mieux selon la valeur prise par la variable y. Les SVM sont parfois appelés Séparateurs à Vaste Marge en français et cette appellation fait référence à la caractérisation de la frontière par maximisation de marge : le but est de faire passer la frontière le plus loin possible de chaque point. Ainsi il est supposé qu'en maximisant la marge, l'erreur est minimisé et donc la qualité de prévision est optimale.

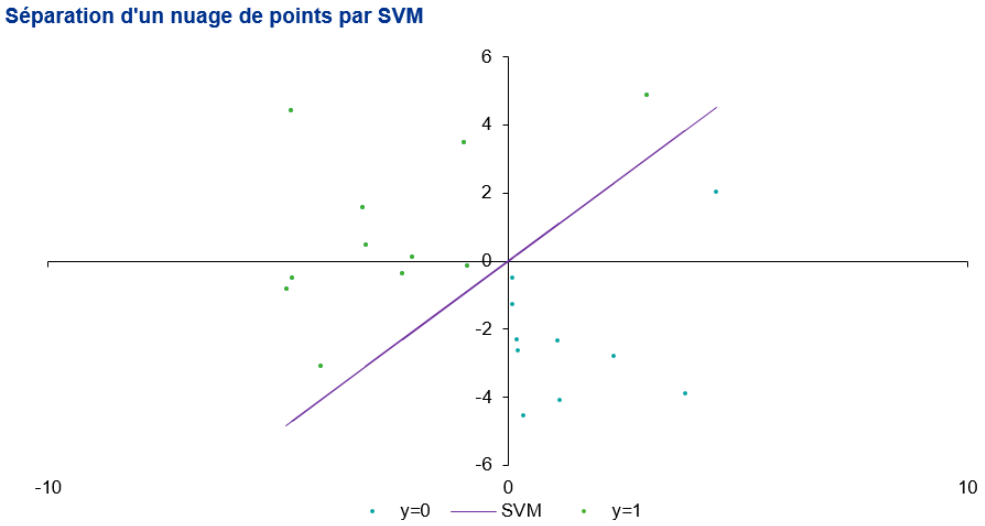

FIGURE 6.1 – Exemple de classification linéaire par SVM

L'approche par SVM parait convenir à la recherche des états du monde économiques : la distinction des états en crise et non crise est une classification binaire de l'espace fonction des variables économiques observables.

6.3 Etude théorique

6.3.1 Cas linéaire séparable

Formulation primale

Soit le problème de classification binaire évoqué dans la partie 6.2.

Définition 6.1 (Problème de classification linéairement séparable)**.**
Un problème de classification binaire est dit linéairement séparable s'il existe une fonction de décision D de la forme :

$$\begin{aligned} D: \quad \mathbb{R}^p &\to \mathbb{R} \\ x &\mapsto \operatorname{sgn}(f(x)) \\ &\quad \text{où } f(x) = v'x + a \end{aligned}$$

qui classe correctement toutes les observations de l'échantillon d'apprentissage.

Remarque 6.1.
La fonction f est appelée fonction caractéristique.

L'hypothèse ici réalisée peut paraître très restrictive, mais elle permet d'introduire les notions importantes et facilitent l'introduction des SVM. De plus ce cas est probable dans le cadre restreint de l'utilisation faite des SVM dans cette étude.

Il est possible d'associer une frontière de décision à toute fonction de décision.

Définition 6.2 (Frontière de décision)**.**
La frontière de décision associée à la fonction de décision D est définie par :

$$\Delta(v; a) = \left\{x \in \mathbb{R}^p | v'x + a = 0\right\}$$

Remarque 6.2.
La fonction de décision (et a fortiori la frontière) sont définies à une constante multiplicative près. L'unicité peut être obtenue en imposant $\|v\| = 1$.

Une notion essentielle dans l'étude des SVM est la marge. En effet tout le calibrage repose sur la maximisation de cette marge.

Définition 6.3 (Marge)**.**
La marge m d'un classifieur est égale à la plus petite distance d'un élément de l'échantillon à la frontière de décision :

$$m = \min_{i \in [\![1;n]\!]} \{\operatorname{dist}(x_i; \Delta(v; a)\}$$

Le principe de choix du classifieur repose donc sur la maximisation de la marge et le problème d'optimisation suivant est à résoudre :

$$\max_{v\in\mathbb{R}^p;a\in\mathbb{R}}\{m\}$$

De plus la distance du point x_i à la frontière linéaire Δ s'exprime comme :

$$\text{dist}(x_i;\Delta(v;a)=\frac{|v'x_i+a|}{\|v\|}$$

Ce problème peut se réécrire comme un problème d'optimisation sous contraintes équivalent :

$$\begin{cases}\max\limits_{v\in\mathbb{R}^p;a\in\mathbb{R}}\{m\}\\ \min\limits_{i\in[\![1;n]\!]}\left\{\dfrac{|v'x_i+a|}{\|v\|}\right\}\geqslant m\end{cases}$$

Comme remarqué précédemment, le problème doit être caractérisé pour dégager une solution unique. Une solution classique est d'effectuer le changement de variable suivant en posant :

$$\begin{cases}w=\dfrac{v}{m.\|v\|}\\ b=\dfrac{a}{m.\|v\|}\end{cases}$$

Le programme d'optimisation est donc équivalent à :

$$\begin{cases}\max\limits_{w\in\mathbb{R}^p;b\in\mathbb{R}}\left\{\dfrac{1}{\|w\|}\right\}\\ \forall i\in[\![1;n]\!]\ y_i.(w'x+b)\geqslant 1\end{cases}$$

Une formulation classique dans la littérature propose de minimiser $\dfrac{\|w\|^2}{2}$ au lieu de maximiser $\dfrac{1}{\|w\|}$. Le programme équivalent s'écrit donc finalement :

$$\begin{cases}\min\limits_{w\in\mathbb{R}^p;b\in\mathbb{R}}\left\{\dfrac{\|w\|^2}{2}\right\}\\ \forall i\in[\![1;n]\!]\ y_i.(w'x+b)\geqslant 1\end{cases}$$

Ce programme d'optimisation peut s'écrire comme un programme quadratique :

$$\begin{cases}\min\limits_{z}\left\{\dfrac{z'Az}{2}-d'z\right\}\\ Bz\leqslant e\end{cases}$$

Avec :

$$\begin{cases} z &= (\beta;\beta_0)' \in \mathbb{R}^{p+1} \\ d &= (0;...;0)' \in \mathbb{R}^{p+1} \\ A &= \begin{bmatrix} I_p & 0_{px1} \\ 0_{1xp} & 0 \end{bmatrix} \in \mathcal{M}_{(p+1)}(\mathbb{R}) \\ y &= (y_1;...;y_n)' \in \mathbb{R}^n \\ X &= (x_1';...;x_n')' \in \mathcal{M}_{nxp}(\mathbb{R}) \\ B &= -[\text{diag}(y)X, y] \\ e &= -(1;...;1)' \in \mathbb{R}^n \end{cases}$$

Le problème est bien convexe car la matrice A est semi-définie positive [1]. Ce problème est dit primal.

La première étape vers la résolution du programme d'optimisation est l'expression du Lagrangien associé :

$$\mathcal{L}(w;b;\alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i.(y_i.(w'x_i + b) - 1)$$

où $\forall i \in [\![1;n]\!] \alpha_i \geqslant 0$

Les conditions de Karush, Kuhn et Tucker permettent de caractériser la solution $(w^*;b^*)$ et les multiplicateurs de Lagrange α^* du problème primal :

$$\begin{aligned} \text{Stationnarité :} \qquad w^* - \sum_{i=1}^{n} \alpha_i^* y_i x_i &= 0 \\ \sum_{i=1}^{n} \alpha_i^* y_i &= 0 \\ \text{Complémentarité :} \qquad \forall i \in [\![1;n]\!] \alpha_i^*.(y_i.(w^{*'}x_i + b^*) - 1) &= 0 \\ \text{Admissibilité primale :} \qquad \forall i \in [\![1;n]\!] y_i.(w^{*'}x_i + b^*) &\geqslant 1 \\ \text{Admissibilité duale :} \qquad \forall i \in [\![1;n]\!] \alpha_i^* &\geqslant 0 \end{aligned}$$

Les conditions de complémentarité donnent l'ensemble des contraintes actives :

$$\mathcal{A} = \left\{ i \in [\![1;n]\!];\ y_i.(w^{*'}x_i + b^*) = 1 \right\}$$

Les multiplicateurs de Lagrange dont l'indice se trouve dans $\mathcal{A}$ sont donc strictement positives. Les multiplicateurs dont l'indice est dans $\overline{\mathcal{A}}$ sont nécessairement nuls. De plus :

$$\forall i \in \overline{\mathcal{A}}\ y_i.(w^{*'}x_i + b^*) > 1$$

1. La démonstration a déjà été faite dans un cas similaire dans la calibrage de la régression logistique

La solution vérifie finalement le système linéaire suivant :

$$\begin{cases} w - X'_{\mathcal{A}} D_y \alpha_{\mathcal{A}} &= 0 \\ -D_y X_{\mathcal{A}} w - b.y_{\mathcal{A}} &= -e_{\mathcal{A}} \\ y'_{\mathcal{A}} \alpha_{\mathcal{A}} &= 0 \end{cases}$$

Avec :

$$\begin{cases} \alpha_{\mathcal{A}} &= \alpha(\mathcal{A}) \\ y_{\mathcal{A}} &= y(\mathcal{A}) \\ X_{\mathcal{A}} &= X(\mathcal{A};) \\ D_y &= \mathrm{diag}(y_{\mathcal{A}}) \end{cases}$$

Ainsi, si l'ensemble $\mathcal{A}$ est connu, la solution s'exprime comme :

$$\beta = X'_{\mathcal{A}} D_y \alpha_{\mathcal{A}} = \sum_{i \in \mathcal{A}} \alpha_i.y_i.x_i$$

Définition 6.4 (Vecteurs supports)**.**
La solution w est finalement une combinaison linéaire des observations x_i dont les contraintes sont actives. Elles sont appelées vecteurs supports puisque ce sont elles qui supportent la frontière de décision.

Remarque 6.3.
Seuls les vecteurs supports sont finalement utilisés pour la résolution du programme d'optimisation.

Ainsi la fonction caractéristique f peut s'écrire comme :

$$f(x) = w'x + b = \sum_{\alpha \in \mathcal{A}} \alpha_i.y_i(x'x_i) + b$$

Il vient ensuite :

$$\begin{cases} D_y X_{\mathcal{A}} X'_{\mathcal{A}} D_y \alpha_{\mathcal{A}} + b.y_{\mathcal{A}} &= e_{\mathcal{A}} \\ y'_{\mathcal{A}} \alpha_{\mathcal{A}} &= 0 \end{cases}$$

Et donc finalement :

$$\begin{cases} b &= \dfrac{y'_{\mathcal{A}} (D_y X_{\mathcal{A}} X'_{\mathcal{A}} D_y)^{-1} e_{\mathcal{A}}}{y'_{\mathcal{A}} (D_y X_{\mathcal{A}} X'_{\mathcal{A}} D_y)^{-1} y_{\mathcal{A}}} \\ \alpha_{\mathcal{A}} &= (D_y X_{\mathcal{A}} X'_{\mathcal{A}} D_y)^{-1} (e_{\mathcal{A}} - b.y_{\mathcal{A}}) \end{cases}$$

Ainsi, la difficulté de la résolution du programme d'optimisation réside dans la recherche de l'ensemble $\mathcal{A}$.

Formulation duale

Cependant, ce programme d'optimisation admet une formulation duale qui est souvent plus pratique à manipuler. Cette formulation sera celle retenue pour la mise en pratique des SVM[2].

Le dual du programme d'optimisation s'écrit :

$$\begin{cases} \min\limits_{\alpha\in\mathbb{R}^n} \left\{ \dfrac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i' x_j}{2} - \sum\limits_{i=1}^{n}\alpha_i \right\} \\ \sum\limits_{i=1}^{n} \alpha_i y_i = 0 \\ \forall i \in [\![1;n]\!] \alpha_i \geqslant 0 \end{cases}$$

Ceci se traduit en écriture matricielle comme :

$$\begin{cases} \min\limits_{\alpha\in\mathbb{R}^n} \left\{ \dfrac{\alpha' G\alpha}{2} - e'\alpha \right\} \\ y'\alpha = 0 \\ \forall i \in [\![1;n]\!] \alpha_i \geqslant 0 \end{cases}$$

Avec :

$$G = D_y X X' D_y$$

6.3.2 Cas linéaire non séparable

En pratique, le cas de données linéairement séparables est très restrictif. La frontière des problèmes de classification n'est jamais nette. Cette situation s'illustre par le graphique suivant :

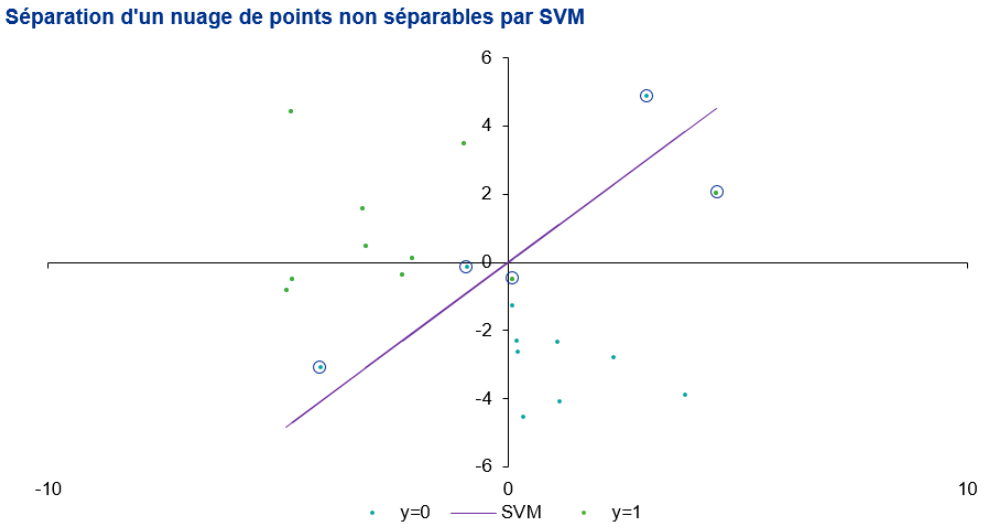

FIGURE 6.2 – Exemple de données non linéairement séparables

2. c.f. partie 6.4 à la page 57

Il est donc nécessaire d'adapter l'étude à ce cas afin d'assurer la pérennité de la modélisation.

Pour se faire, il est nécessaire d'introduire des variables d'écart $(\xi_i)_{i\in[\![1;n]\!]}$ dans $\mathbb{R}_+^n$. Dans le cas où le point i se trouve du bon côté de la frontière de décision (i.e. $y_i(w'x_i+b)\geqslant 1$), alors $\xi_i = 0$.

En cas d'erreur (i.e. $y_i(w'x_i+b)<1$), la variable d'écart est donnée par : $\xi_i = 1-y_i(w'x_i+b) > 0$.

Le programme d'optimisation se réécrit alors comme :

$$\begin{cases} \min\limits_{w\in\mathbb{R}^p;b\in\mathbb{R}} \left\{\dfrac{\|w\|^2}{2}\right\} \\ \min\limits_{\xi\in\mathbb{R}_+^n} \left\{\sum\limits_{i=1}^n \xi_i\right\} \\ \forall i\in[\![1;n]\!]\ y_i.(w'x+b)\geqslant 1-\xi_i \end{cases}$$

Remarque 6.4.
Si toutes les variables d'écart sont nulles, la formulation du problème redonne bien le primal du cas linéaire séparable.

La résolution de se problème s'effectue en introduisant un paramètre de pénalisation $C\in\mathbb{R}_+^*$. Il permet de pénaliser un dépassement de la marge afin d'arbitrer entre maximisation de la marge et minimisation des erreurs. Le programme se réécrit maintenant :

$$\begin{cases} \min\limits_{w\in\mathbb{R}^p;b\in\mathbb{R};\xi\in\mathbb{R}_+^n} \left\{\dfrac{\|w\|^2}{2} + C.\sum\limits_{i=1}^n \xi_i\right\} \\ \forall i\in[\![1;n]\!]\ y_i.(w'x_i+b)\geqslant 1-\xi_i \\ \forall i\in[\![1;n]\!]\ \xi_i\geqslant 0 \end{cases}$$

Le Lagrangien s'exprime cette fois comme :

$$\mathcal{L}(w;b;\alpha;\beta) = \frac{\|w\|^2}{2} + C.\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i.(y_i.(w'x_i+b)-1+\xi_i) - \sum_{i=1}^n \beta_i.\xi_i$$

où $\forall i\in[\![1;n]\!]\alpha_i\geqslant 0$

où $\forall i\in[\![1;n]\!]\beta_i\geqslant 0$

Et le dual s'écrit :

$$\begin{cases} \min\limits_{\alpha\in\mathbb{R}^n} \left\{\dfrac{\sum\limits_{i=1}^n\sum\limits_{j=1}^n \alpha_i\alpha_j y_i y_j x_i' x_j}{2} - \sum\limits_{i=1}^n \alpha_i\right\} \\ \sum\limits_{i=1}^n \alpha_i y_i = 0 \\ \forall i\in[\![1;n]\!] 0\leqslant\alpha_i\leqslant C \end{cases}$$

6.3.3 Cas non linéaire

Cependant, les SVM linéaires supposent que la classification se fait selon une séparation parfaitement droite de l'espace. Il n'est pas difficile d'imaginer un contre-exemple dont la frontière de décision n'est pas linéaire. Dans le cadre de la détection des états de crise, le suivi de l'inflation le montre : toute chose égale par ailleurs, une trop forte inflation est synonyme de crise mais une trop forte déflation [3] l'est également. Voici un exemple graphique de la situation :

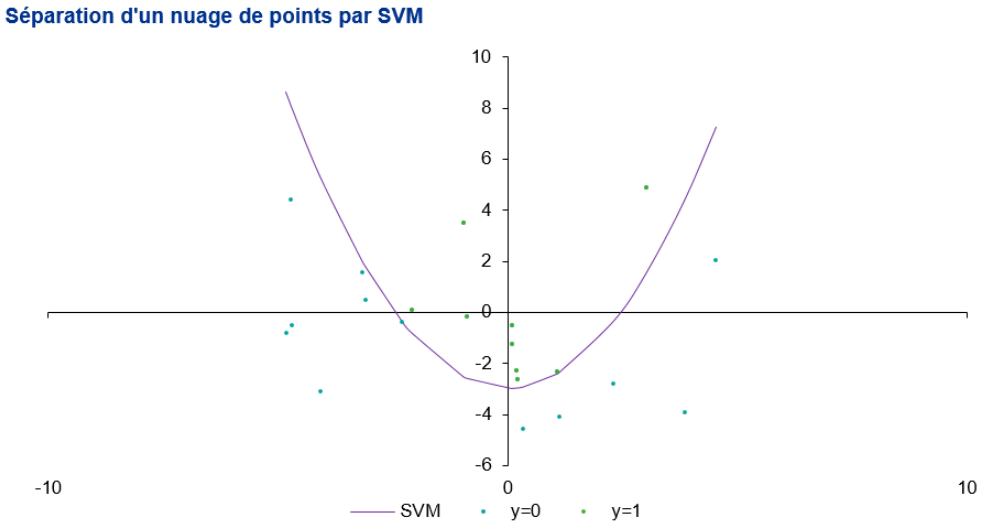

FIGURE 6.3 – Exemple de classification non linéaire par SVM

Ceci montre le besoin de généraliser la théorie des SVM au cas non linéaire. La généralisation peut se faire par l'introduction des noyaux [4].

Le principe est de réécrire la fonction caractéristique f comme :

$$f(x) = \sum_{\alpha \in \mathcal{A}} \alpha_i . y_i k(x_i; x) + b$$

Définition 6.5 (Noyau)**.**
Soit χ l'ensemble auquel appartiennent les éléments de l'échantillon. Un noyau k est défini par :

$$\begin{array}{rccc} k: & \chi^2 & \to & \mathbb{R} \\ & (x_1; x_2) & \mapsto & k(x_1; x_2) \end{array}$$

Remarque 6.5.
L'introduction des noyaux permet une dépendance non linéaire et elle généralise l'ensemble d'appartenance de l'échantillon. En effet il n'est plus limité au cas réel. Cette généralisation peut permettre par exemple de considérer une courbe de taux complète pour la distinction des états de crise.

3. Inflation négative durablement
4. Cette méthode est appelée "kernel trick" en anglais

Définition 6.6 (Matrice de Gram).
Soit k un noyau. Soit $(x_i)_{i\in[\![1;n]\!]} \in \chi^n$ un échantillon associé. Sa matrice de Gram est la matrice K de terme général $K_{i;j} = k(x_i; x_j)$.

Définition 6.7 (Noyau positif).
Soit k un noyau. Il est dit positif si :

$$\forall n \in \mathbb{N}^* \ \forall (x_i)_{i\in[\![1;n]\!]} \in \chi^n \begin{cases} K \text{ est symétrique} \\ K > 0 \end{cases}$$

Remarque 6.6.
Un noyau positif et défini sur un compact est dit de Mercer.

La construction de tels noyaux est une étape importante dans le développement des SVM. Cependant seuls des noyaux usuels seront utilisés dans le cadre de l'étude. Il est tout de même bon de donner des pistes sur la construction des noyaux. Il existe deux grandes méthodes.

La première méthode est l'utilisation d'un espace intermédiaire muni d'un produit scalaire. Une première transformation ϕ permet de passer de χ à l'espace intermédiaire. Le noyau est ensuite défini par : $k(x_1; x_2) = \langle x_1 | x_2 \rangle$.

La deuxième méthode se repose sur les propriétés algébriques des noyaux positifs. Ainsi la somme, le produit, la somme directe, le passage à la limite, etc. conservent la positivité et permettent de construire de nouveaux noyaux positifs.

Définition 6.8 (Espace de Hilbert à Noyau Reproduisant).
Soit H un espace de Hilbert muni de son produit scalaire $\langle \bullet | \bullet \rangle$. H est dit Espace de Hilbert à Noyau Reproduisant (EHNR) si :

$$\forall x \in \chi \ \exists k(x; \bullet) \in H \text{ symétrique } \forall f \in H \ f(x) = \langle f | k(x; \bullet) \rangle$$

Théorème 6.1 (Théorème de Mercer [5]).
Il est possible d'associer un EHNR à tout noyau positif et réciproquement.

Définition 6.9 (SVM dans le cas non linéaire).
Soit H un EHNR de noyau k. Un SVM est une fonction de décision de la forme :

$$\begin{aligned} D : \ \chi &\to \mathbb{R} \\ x &\mapsto \operatorname{sgn}(f(x) + b) \\ &\quad \text{où } f \in H \text{ et } b \in \mathbb{R} \end{aligned}$$

5. La démonstration n'est pas donnée ici mais elle est réalisée dans l'ouvrage de Mercer[J. Mercer, 1909]

Où f et b sont donnés par la résolution du programme d'optimisation suivant pour un $C \geqslant 0$ donné :

$$\begin{cases} \min\limits_{w\in\mathbb{R}^p;b\in\mathbb{R};\xi\in\mathbb{R}_+^n} \left\{ \dfrac{\|f\|^2}{2} + C.\sum\limits_{i=1}^{n} \xi_i \right\} \\ \forall i \in [\![1;n]\!]\ y_i.(f(x_i)+b) \geqslant 1-\xi_i \\ \forall i \in [\![1;n]\!]\ \xi_i \geqslant 0 \end{cases}$$

Théorème 6.2 (Solution des SVM[6])**.**
La solution des SVM dans le cas non linéaire s'exprime comme :

$$f(x) = \sum_{i\in\mathcal{A}} \alpha_i.y_i.k(x;x_i)$$

Où α est solution de programme d'optimisation suivant :

$$\begin{cases} \min\limits_{\alpha\in\mathbb{R}^n} \left\{ \dfrac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} \alpha_i\alpha_j y_i y_j.k(x_i;x_j)}{2} - \sum\limits_{i=1}^{n}\alpha_i \right\} \\ \sum\limits_{i=1}^{n} \alpha_i y_i = 0 \\ \forall i \in [\![1;n]\!] 0 \leqslant \alpha_i \leqslant C \end{cases}$$

Et $\mathcal{A}$ est l'ensemble des contraintes actives.

Les conditions de Karush, Kuhn et Tucker donnent cette fois :

$$\begin{aligned} \alpha_i = 0 &\Leftrightarrow y_i.(\sum_{j=1}^{n} y_j\alpha_j k(x_i;x_j-b) \geqslant 1 \\ 0 < \alpha_i < C &\Leftrightarrow y_i.(\sum_{j=1}^{n} y_j\alpha_j k(x_i;x_j-b) = 1 \\ \alpha_i = C &\Leftrightarrow y_i.(\sum_{j=1}^{n} y_j\alpha_j k(x_i;x_j-b) \leqslant 1 \end{aligned}$$

6.4 Calibrage par algorithme SMO

Maintenant que le problème des SVM est exposé, la difficulté réside dans la résolution des programmes d'optimisation. Les résolutions ne peuvent cependant pas se faire de manière analytique et des algorithmes d'optimisation sont nécessaires pour la mise en pratique des SVM. Avant de présenter ces algorithmes, il est important d'être sensibilisé aux contraintes physiques des machines.

6. La démonstration n'est pas donnée ici mais elle repose sur la même méthode que pour le cas linéaire

Les SVM nécessitent une base d'apprentissage suffisamment large pour atteindre un niveau de précision satisfaisant et donc un pouvoir de généralisation correct. Mais qui dit base d'apprentissage suffisante, dit gros volume de données. Les contraintes de mémoire des ordinateurs peuvent donc limiter (voir même empêcher) l'utilisation des SVM. De plus, l'utilisation du kernel trick peut mener au passage dans un espace intermédiaire de dimension bien supérieure à celle de l'espace contenant l'échantillon. Donc une base déjà volumineuse nécessite parfois d'être multipliée pour appliquer la théorie des SVM non linéaires. Ainsi l'équilibre entre performance et précision est constamment recherché.

Ces pour ces raisons d'optimisation de la mémoire que tous les algorithmes se basent sur une réduction de la dimension. De plus ils possèdent souvent une composante heuristique [7] qui les rend plus difficile à appréhender.

Il en existe trois principaux : l'algorithme de Chunking[V. Vapnik, 1982] proposé par Vapnik lui même, l'algorithme d'Osuna[E. Osuna et al., 1998] et l'algorithme Sequential Minimal Optimization (SMO)[J.C. Platt, 1997] proposé par John C. Platt. C'est ce dernier qui est ici développé puisqu'il est l'un des plus récent et des plus reconnus.

L'algorithme SMO repose sur la résolution de beaucoup de sous-problèmes plus simples que le problème initial. Sa convergence est assurée par le théorème d'Osuna[E. Osuna et al., 1998].

Il se décompose en trois étapes :

i) Le choix de deux multiplicateurs de Lagrange α_i et α_j

ii) L'optimisation pour ces deux multiplicateurs

iii) La mise à jour du SVM

L'algorithme boucle sur ces trois étapes jusqu'à convergence.

6.4.1 Heuristique de choix de deux multiplicateurs

L'algorithme SMO se différencie par l'optimisation de seulement deux multiplicateurs à la fois.Le choix des deux multiplicateurs s'effectue en deux étapes distinctes.

Choix du premier multiplicateur

L'algorithme SMO commence par parcourir l'ensemble de l'échantillon d'apprentissage afin de vérifier quels éléments ne respectent pas les conditions KKT. Chaque élément de l'échantillon d'apprentissage qui viole ces conditions est éligible pour le premier choix et la deuxième heuristique de choix est alors lancée.

Une fois que la première heuristique à parcouru l'ensemble de l'échantillon d'apprentissage, elle parcourt tous les éléments qui ne sont pas aux bornes de conditions KKT (i.e. $\alpha_i \in]0; C[$). Là encore, tous ceux qui ne respectent pas les conditions KKT sont éligibles. Cet ensemble est parcouru

7. Algorithme approché d'optimisation utilisé pour améliorer les complexités spatiales et temporelles des algorithmes classiques

tant qu'il reste des éléments qui ne vérifient pas les conditions KKT.

Une fois qu'ils vérifient tous les conditions, l'algorithme repart sur l'ensemble de l'échantillon et continue et réitère le processus. L'algorithme s'arrête une fois que tous les éléments de l'échantillon d'apprentissage vérifient les conditions.

La notion de respect des conditions KKT est ici relative : un seuil de précision ϵ est choisi afin d'arbitrer entre performance de calcul et précision. $\epsilon = 0,001$ est souvent suffisant[J.C. Platt, 1997].

Choix du second multiplicateur

Une fois le premier multiplicateur choisi, la deuxième heuristique choisi le second multiplicateurs de façon à effectuer la plus grande avancée possible. Ainsi elle parcourt l'ensemble des éléments qui ne sont pas aux limites des conditions KKT et choisit le multiplicateur qui maximisera l'avancée dans l'optimisation présentée dans la partie 6.4.2.

Cependant, l'évaluation de la fonction objectif peut être coûteuse en temps de calcul et une approximation peut s'avérer intéressante en temps de calcul sans pour autant être trop grossière.Ainsi si α_1 a été choisi par la première heuristique, alors α_2 est choisi de manière à maximiser $|E_1 - E_2|$ [8].

Pour des soucis de performance, l'algorithme peut conserver l'erreur E pour les éléments qui ne sont pas aux bornes afin d'éviter des calculs supplémentaires.

Dans certains cas, la deuxième heuristique ne permet pas d'effectuer d'avancée dans l'optimisation. L'algorithme prévoit alors une hiérarchie de second choix pour éviter de stagner : l'ensemble des éléments qui ne sont pas aux bornes sont d'abord testés en partant d'un rang arbitraire puis l'ensemble complet et testé de la même manière. Si aucun choix n'a permit une avancée dans l'optimisation, alors l'heuristique de premier choix est relancée en passant le premier choix précédent.

6.4.2 Optimisation pour deux multiplicateurs

En se limitant à seulement deux multiplicateurs, il est possible de résoudre analytiquement le problème. Soit α_1 et α_2 les deux multiplicateurs sélectionnés. Les contraintes de boîtes affirment que : $0 \leqslant \alpha_1 \leqslant C$ et $0 \leqslant \alpha_2 \leqslant C$. L'ensemble de recherche est donc réduit à un carré. La contrainte d'égalité donne :

$$y_1.\alpha_1 + y_2.\alpha_2 = \sum_{i=3}^{n} y_i.\alpha_i = k$$

k représente un terme constant et il vient au signe près :

$$y_1 = y_2 \quad \Rightarrow \quad \alpha_1 + \alpha_2 = k$$

$$y_1 \neq y_2 \quad \Rightarrow \quad \alpha_1 - \alpha_2 = k$$

8. La notation est introduite dans la partie 6.4.2

Soient L et H définis par :

$$\begin{aligned} y_1 \neq y_2 \quad &\Rightarrow \quad L = \max(0; \alpha_2 - \alpha_1) \qquad H = \min(C; C + \alpha_2 - \alpha_1) \\ y_1 = y_2 \quad &\Rightarrow \quad L = \max(0; \alpha_2 + \alpha_1 - C) \quad H = \min(C; \alpha_2 + \alpha_1) \end{aligned}$$

La dérivée second de la fonction objectif s'écrit alors :

$$\eta = k(x_1; x_1) + k(x_2; x_2) - 2.k(x_1; x_2)$$

Soient :

$$\forall i \in [\![1; n]\!] E_i = \sum_{j=1}^{n} y_j \alpha_j k(x_i; x_j - b - y_i$$

La nouvelle valeur de α_2 est donnée par :

$$\alpha_2^{\text{new}} = \alpha_2 + \frac{y_2(E_1 - E_2)}{\eta}$$

Ainsi, pour s'assurer de respecter les contraintes de KKT :

$$\alpha_2^{\text{new ; borné}} = \min(\max(\alpha_2^{\text{new}}; L); H)$$

Soit maintenant $s = y_1 y_2$, la mise à jour de α_1 est donnée par :

$$\alpha_1^{\text{new}} = \alpha_1 + s(\alpha_2 - \alpha_2^{\text{new ; borné}})$$

Le terme de dérivée seconde η doit être strictement positif cependant il peut ne pas l'être dans certains cas exceptionnels : si le noyau ne respecte pas les conditions de Mercer ou si deux éléments de l'échantillon sont strictement identiques par exemple. L'algorithme SMO peut tout de même converger dans ce genre de situation. Il suffit d'évaluer la fonction objectif aux deux extrémités L et H et d'affecter aux deux multiplicateurs la valeur qui minimise la fonction objectif. Ainsi, en notant Ψ la fonction objectif, les étapes de calcul sont :

$$\begin{aligned} f_1 &= y_1(E_1 + b) - \alpha 1.k(x_1; x_1) - s.\alpha_2.k(x_1; x_2) \\ f_2 &= y_2(E_2 + b) - s.\alpha 1.k(x_1; x_2) - \alpha_2.k(x_2; x_2) \\ L_1 &= \alpha_1 + s.(\alpha_2 - L) \\ H_1 &= \alpha_1 + s.(\alpha_2 - H) \\ \Psi_L &= L_1 f_1 + L f_2 + \frac{L_1^2 k(x_1; x_1)}{2} + \frac{L^2 k(x_2; x_2)}{2} + sLL_1 k(x_1; x_2) \\ \Psi_H &= H_1 f_1 + H f_2 + \frac{H_1^2 k(x_1; x_1)}{2} + \frac{H^2 k(x_2; x_2)}{2} + sHH_1 k(x_1; x_2) \end{aligned}$$

6.4.3 Mise à jour du SVM

Si, après l'étape d'optimisation, $0 < \alpha_1 < C$, alors la valeur de b est mise à jour comme :

$$b_1 = E_1 + y_1(\alpha_1^{\text{new}} - \alpha_1)k(x_1; x_1) + y_2(\alpha_2^{\text{new ; borné}} - \alpha_2)k(x_1; x_2) + b$$

Si c'est α_2 qui vérifie cette condition, alors :

$$b_2 = E_2 + y_1(\alpha_1^{\text{new}} - \alpha_1)k(x_1; x_2) + y_2(\alpha_2^{\text{new ; borné}} - \alpha_2)k(x_2; x_2) + b$$

Si les deux conditions sont vérifiées en même temps, la nouvelle valeur de b est $\frac{b_1 + b_2}{2}$.

6.5 Application dans le cadre des rachats

Dans le cadre de la modélisation des rachats, l'introduction des SVM se fait en deux étapes. La première consiste en l'application des SVM au sens de la classification alors que la deuxième consiste en l'adaptation du problème de modélisation du rachat.

Tout d'abord l'échantillon est découpé en deux parties : une partie est marquée crise ($y = -1$) et l'autre on crise ($y = 1$). Cette séparation se fait par jugement d'expert selon les périodes connues.

Une fois que tout l'échantillon d'apprentissage est marqué, la théorie des SVM est appliquée sur les variables économiques uniquement. Ainsi une règle de décision est construite afin de classer, selon l'environnement économique, les états de crise et de non crise.

Sur chacun des deux sous-échantillons marqués, la modélisation du rachat est faite de manière indépendante comme vu dans les parties 4, 5 et 5.4.

Ainsi, lors de l'utilisation au sein du modèle de projection, sur chaque scénario et à chaque pas de temps, le SVM calibré permettra de choisir la loi de rachat à appliquer. Le modèle de rachat possède donc deux états.

6.6 Limites a priori

L'utilisation des SVM permet ici de séparer l'espace des variables économiques. La théorie présentée permet une certaine perméabilité lors du calibrage : des points peuvent traverser la frontière et se trouver du "mauvais côté". Cette possibilité permet aux SVM d'être plus réalistes car s'il est possible que les données soient tout à fait séparables, cette situation est peu probable.

Cependant, lors de l'étape de généralisation (i.e. de prévision), l'espace est considéré comme franchement découpé et cette notion de perméabilité est terminée. Finalement 1 bp d'écart sur l'OAT 10 ans peut faire basculer d'un état non crise à un état crise. Cela fait partie même du fondement des SVM avec la recherche de maximisation de la marge, mais l'analyse doit être faite avec précaution sur les points alentours à la frontière.

De plus la mise en place de telles méthodes est lourde en termes de compréhension et d'implémentation. Si les SVM sont de puissants outils, il faut s'assurer de leur nécessité avant de se lancer dans leur utilisation.

Si la généralisation est souvent meilleure, l'analyse peut être plus difficile lors de prévision par SVM avec un effet boîte noire conséquence de l'utilisation des noyaux et des heuristiques de calibrage.

Cependant les SVM trouvent beaucoup d'applications au sein même du domaine assurantiel. Un mémoire s'est déjà intéressé au sujet dans le cadre de l'accélération du calcul du besoin en capital dans le cadre d'un modèle interne[M. Niedzwiedz, 2015].

Chapitre 7

Outil ALM DEvent

7.1 Projection de flux de trésorerie

7.1.1 Outil de 3ème génération

Le calcul des éléments du bilan économique nécessite l'utilisation d'un outil Asset and Liabilities Management (ALM). Le rôle de ce type d'outil est de projeter les flux de trésorerie du business de l'assureur afin d'évaluer les engagements, les profits et les taxes futurs. L'objectif principal de ce type d'outil sous Solvabilité II est d'évaluer les éléments du bilan économique. Pour rappel, l'évaluation de certaines garanties telles que le TMG nécessitent l'utilisation de scénarios économiques simulés afin de capter les asymétries de ces options. Un exemple simple a été donné dans la figure 2.5 à la page 9.

L'outil DEvent est un outil dit de 3ème génération : c'est un outil de projection de flux qui repose sur l'utilisation d'un GSE. Il est peut être utilisé de manière déterministe ou stochastique. Voici l'organisation globale des deux outils internes :

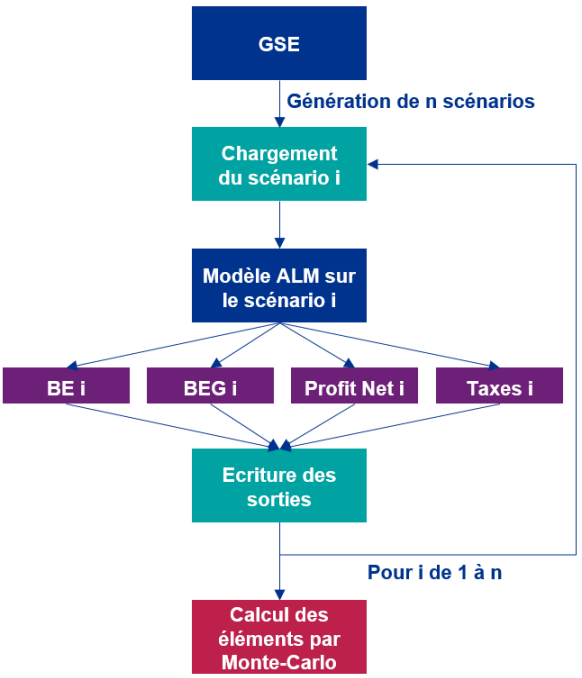

FIGURE 7.1 – Organisation des outils internes

L'outil DEvent est donc construit de façon à projeter les flux de trésorerie d'un assureur de manière déterministe (i.e. sur un seul scénario à la fois). Cette partie se concentre sur la présentation déterministe de l'outil. L'utilisation du GSE sera développée dans la partie 8 à la page 68.

L'architecture globale de DEvent repose sur une vision programmation objet de l'outil et permet d'avoir rapidement une vue d'ensemble :

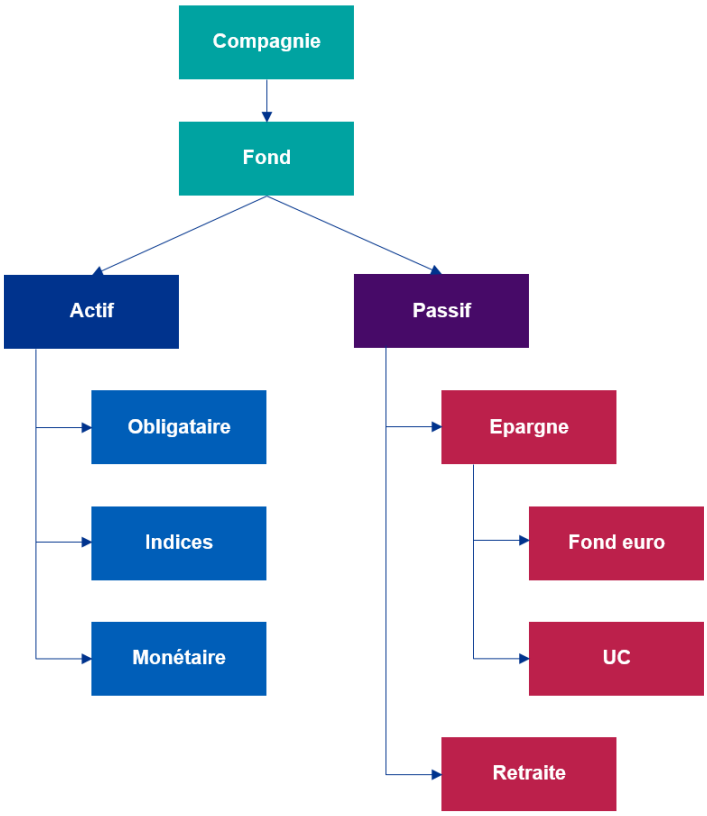

FIGURE 7.2 – Architecture de l'outil DEvent

La majorité du business est géré au niveau fond et cette architecture laisse la possibilité d'une gestion multi-fonds (cantons, compagnie à plusieurs portefeuilles, etc.).

7.1.2 Actif

L'actif modélisé se décompose en trois grandes classes : obligataire, indices et monétaire. Chacune de ces classes peut se subdiviser en plusieurs sous-classes. Le schéma de projection est cependant le même pour chaque classe :

i) Initialisation de l'actif (valeur de marché, valeur comptable, plus ou moins value latente)

ii) Projection des flux (coupons, nominaux, loyers, dividendes)

iii) Intégration des flux d'actif et de passif

iv) Réallocation (achats/ventes) selon les allocations cible et tactique données en entrée du modèle

v) Gestion des frais d'investissements au niveau fond

Si le schéma est globalement identique, chaque étape se confronte aux spécificités des classes. De plus, dans un environnement épargne français, les normes françaises sont toujours applicables et reflètent la gestion réelle de la compagnie. Ainsi, même sous Solvabilité II, des éléments Solvabilité I persistent et sont projetés dans l'outil DEvent. Ces éléments sont maintenant détaillés pour chaque classe mais tous les concepts ne sont pas rappelés ; le but de cette présentation est de comprendre l'utilisation d'un modèle ALM et les enjeux qu'il représente dans la modélisation du rachat.

Obligataire

Les actifs obligataires représentent en général environ 80% du portefeuille d'un assureur vie. Ils constituent une solide base de revenues récurrents qui sont utilisés pour financer les taux garantis.

Les caractéristiques (date d'achat, valeur d'achat, taux de coupon, nominal, fréquence de coupon) et la valeur de marché de chaque obligation sont directement lues en entrée du modèle. Cependant, une modification importante est opérée sur chaque actif obligataire : la risque-neutralisation. La projection stochastique qui est faite est dite risque-neutre ; c'est-à-dire que chaque actif ramène en moyenne le taux sans risque. Il est donc nécessaire de s'assurer que les actifs modélisés sont en accord avec les scénarios générés. Ainsi, à l'instant initial, le nominal de chaque obligation est retouché de façon à ce que sa valorisation à la courbe de taux de référence soit égale à la valeur de marché d'entrée.

Le taux actuariel à l'achat et la valeur nette comptable sont recalculés à partir des éléments précédents après risque-neutralisation. Les plus ou moins values latentes peuvent maintenant être calculées.

Lors de la projection, les coupons et nominaux tombent au fil de l'avancée du temps. Leur tombées est supposée en fin de période. Si les titres présentent un risque de défaut, les défauts sont appliqués à chaque fin de période par le biais d'une intensité de défaut fournie par le GSE. La valeur de marché est mise à jour selon le scénario de taux

Si des achats sont réalisés, ils sont supposés au pair (i.e. valeur de marché = nominal et TRA = taux de coupon).

L'outil modélise également la réserve de capitalisation qui est une réserve en norme française mais qui impact les revenus comptables lors de vente d'actifs relevant de l'article R343-9 du code des Assurances. La Provision pour Dépréciation Durable (PDD) est également modélisée et elle peut être liée aux actifs présentants un risque de crédit.

Indices

Les actifs relatifs à des indices représentent environ 15% du portefeuille. Ils permettent de diversifier les investissements et de s'exposer un peu plus aux risques de marché et donc d'augmenter les rendements économiques et comptables. Les classes indices correspondent aux actions et à l'immobilier.

Les valeurs de marché et comptables sont des entrées du modèles. Cette fois aucune risque-neutralisation n'est nécessaire et les plus ou moins values latentes peuvent directement être calculées.

Lors de la projection, les dividendes et loyers tombent puis la valeur de marché est mise à jour selon la performance de l'indice correspondant du scénario.

Deux provisions relatives à ces actifs sont modélisées : la PDD et la Provision pour Risque d'Exigibilité (PRE). Ces deux provisions correspondent aux normes françaises et sont relatives aux actifs relevant de l'article R343-10 du code des Assurances.

Monétaire

La classe monétaire représente les liquidités de l'entreprise. La valeur de marché et comptable sont données en entrée du modèle et ont la particularité d'être constamment égales. En effet, cette classe capitalise le taux sans risque réalisé sur le scénario et est donc supposée sans risque.

Cependant cette classe joue un rôle particulier dans la projection ALM puisqu'elle comptabilise tous les flux d'actif (coupons, dividendes, loyers, surcotes/décotes, nominaux, frais) et de passif (présentés dans la partie 7.1.3).

7.1.3 Passif

Trois types de produits sont aujourd'hui gérés par le modèle : l'épargne, les unités de compte et la retraite. Là encore la trame générale de la modélisation est commune aux trois produits. Les principales étapes sont :

i) Initialisation des paramètres et de provisions
ii) Intégration des primes
iii) Intégration des arrivées à maturité, des décès et des rachats
iv) Mise à jour des provisions et intégration des frais
v) Intégration des flux de passif aux actifs monétaires

Les produits en unités de compte et de retraite ne sont pas développés ici puisqu'ils sortent du cadre de l'étude.

Les caractéristiques des assurés sont fournies en entrée du modèle : la date de naissance, le sexe, l'ancienneté, la date de maturité du contrat, le TMG, la table du tarif, la clause de PB contractuelle, le taux de chargement sur encours, les primes, le taux de chargement sur prime, le type de réseau de distribution, la réserve mathématique et le taux servi l'année précédente.

Une table de projection différente de celle du tarif peut être utilisée. Les primes futures peuvent être nulles notamment dans le cadre du calcul du BE qui est réalisé en vision run-off.

Les décès sont calculés par rapport à la table de projection. Les rachats sont calculés selon les ONC du QIS 5.

Les provisions mathématiques sont ensuite abondées des produits de trois natures différentes : le TMG, la PB minimale et la PB discrétionnaire. L'abondement du TMG se fait en deux parties : une première partie en milieu de période et une seconde en fin de période. Les résultats de flux sont financés par les produits financiers et les ventes d'actifs si besoin.

Lorsque plusieurs produits sont présents dans un même fond, des interactions entre produits peuvent avoir lieu : de bons résultats techniques sur un produit peuvent donner lieu à de la PB discrétionnaire sur un autre ou à des compensations de résultats. Dans ce cadre, la modélisation fine des management actions est indispensable.

7.1.4 Interactions Actif-Passif

Pour un produit d'épargne simple, il existe trois principales interactions actif-passif : l'évaluation du taux de rachat, le calcul du minimum de PB réglementaire et la distribution des produits.

L'évaluation du taux de rachat dans le cadre du QIS 5 se décompose en deux parties : les rachats structurels et les rachats dynamiques. La première partie est souvent déterministe et ne dépend que de l'ancienneté. La seconde dépend de l'écart entre le taux servi et un taux compétiteur. Il dépend en fait de ces taux avec un an de latence en raison du temps de réaction des assurés. Ainsi le taux servi l'année N aura un impact sur le taux de rachats dynamiques observés l'année N+1. Donc finalement, la performance de l'actif mélangée au résultat technique et aux management actions de participation aux bénéfices décideront de ce taux de rachat.

La deuxième interaction est le résultat d'un calcul réglementaire qui fait intervenir les résultats technique et financier. La formule est :

$$\mathrm{PB}_{\min} = 90\%.\mathrm{Rés}_{\mathrm{tech}}.1_{\{\mathrm{Rés}_{\mathrm{tech}}>0\}} + 100\%.\mathrm{Rés}_{\mathrm{tech}}.1_{\{\mathrm{Rés}_{\mathrm{tech}}<0\}} + 85\%.\mathrm{Rés}_{\mathrm{fi}}.1_{\{\mathrm{Rés}_{\mathrm{fi}}>0\}}$$

Cette relation met en évidence l'enjeu actif-passif généré par la participation minimum aux bénéfices : le pilotage des résultats techniques et financiers ne sont pas sans conséquences sur les produits à redistribuer et ces mêmes produits ont un impact sur le taux de rachats comme vu précédemment.

La distribution des produits est la distribution effective du maximum entre le TMG et le taux de PB réglementaire majoré de la PB discrétionnaire. La PB discrétionnaire est un levier de pilotage des résultats futurs selon le mode de distribution : direct ou passage en Provision pour Participation aux Bénéfices (PPB). Là encore, le modèle doit tenir compte des management actions pour la modélisation de cette provision dont la gestion aura des impacts sur les rachats (passif) et donc sur l'actif a fortiori (réalisation de plus ou moins values).

Ces enjeux sont exposés succinctement afin de sensibiliser aux règles principales imposées aux assureurs et aux leviers dont ils disposent pour piloter leur business. Des analyses plus détaillées seront faites dans la partie IV à la page 92.

7.2 Calcul du BE

Maintenant que les méthodes et règles de projection des flux de trésorerie sont exposées, le calcul du BE repose sur la bonne utilisation de ces flux. Ainsi les flux sont décomposés en plusieurs

groupes et sont répartis dans les différents éléments constituants le bilan économiques.

Les flux relatifs au BE sont donc stockés et le modèle donne en sortie une chronique de flux de BE pour chaque scénario. Le BE déterministe sur un scénario est donc la valeur actuelle au taux sans risque du scénario de l'ensemble des flux de la chronique.

Le BE stochastique est finalement la moyenne empirique de chacun des BE déterministes. L'ANR peut ensuite être déduit.

7.3 Calcul du SCR

Le calcul du SCR repose à la fois sur le calcul du BE vu précédemment et sur la formule standard. C'est cette dernière qui est mise en place dans le cadre de l'étude et non un modèle interne.

Ainsi chaque sous-module se calculant comme une variation d'ANR donne lieu à deux calculs de BE : le BE dit de base et le BE choqué. Le BE de base est commun à tous les calculs de SCR. Le BE choqué est calculé de la même manière que le BE de base mais des chocs calibrés par l'EIOPA sont appliqués lors de la projection comme présenté dans la partie 2.3 à la page 10.

Chapitre 8

Utilisation du GSE

L'outil ALM repose entièrement sur le GSE pour une utilisation stochastique. Dans le cadre du calcul du BE, la projection est risque-neutre : tous les actifs doivent en moyenne rapporter le taux sans-risque.

Pour se faire, le GSE se repose sur différents modèles d'actifs et sur des corrélations entre ces modèles. Le but, dans le cadre d'une étude sur la modélisation des rachats, n'est pas de présenter les différents modèles utilisés. La validation de ce type d'outil est une étude à part entière. C'est le lien entre la modélisation des variables économiques et la modélisation du rachat qui fait sens.

En effet toute l'étude se repose sur l'explication du rachat à partir des variables économiques entre autres. Avant de voir quel modèle de rachat est choisi, il est nécessaire d'étudier au préalable l'utilisation possible du GSE dans ce cadre.

Le GSE est le point de pivot entre les chroniques historiques des variables économiques (partie de calibrage de la loi de rachat) et les chroniques projetées des variables économiques (partie de projection ALM pour l'évaluation du bilan économique). Ainsi, pour que ce point de pivot ne se transforme pas en point de rupture, il est nécessaire d'adapter les variables économiques choisies lors du calibrage de la loi pour qu'elle correspondent à celles qui seront projetées. Pour s'en assurer, il suffit de connaître les variables qui servent au calibrage du GSE et savoir comment se comporte les modèles entre eux.

Il est claire qu'en supposant que les assurés se réfèrent au niveau du CAC 40 pour prendre leur décision de rachat et projeter en supposant qu'ils surveillent désormais l'évolution d'un panier d'actions non cotées ne ferait aucun sens. De même le taux de chômage parait être un bon indicateur du rachat puisqu'en cas de hausse du chômage, le besoin de liquidité augmente et donc le rachat aussi. Mais le GSE ne modélisant pas le chômage et l'outil ALM non plus, cette variable est difficilement utilisable. L'utilisation du GSE peut donc induire une contrainte dans la modélisation et nécessite une rationalisation dans le choix du modèle pour pouvoir être exploiter dans une de projection ALM.

Le modèle d'indices et un modèle Black & Scholes et il est calibré sur l'EUROSTOXX 50 pour le type 1. Le type 2 et l'immobilier suivent le même modèle avec des volatilités égales à un coefficient près à celle du type 1.

La diffusion du taux est faite initialement sur la courbe EIOPA avec ajustement pour volatilité.

C'est donc un taux sans risque qui est diffusé. Ce modèle est calibré sur les prix d'options sur produits de taux (type swaptions).

Le modèle d'inflation est quant à lui définit par une élasticité du modèle de taux. Il est calibré sur les swaps d'inflation.

Des ajustements seront présentés dans la partie IV à la page 92 pour pouvoir utiliser correctement la loi de rachat.

Troisième partie

Mise en œuvre de l'analyse

Chapitre 9

Etude de données

9.1 Caractéristiques de l'échantillon

L'ensemble des données se décompose en trois parties : l'échantillon historique de rachats et des variables relatives aux assurés, l'historique des variables économiques et le portefeuille utilisé pour la projection ALM. La dernière catégorie de données ne sera utilisée que dans la partie IV à la page 92.

9.1.1 Données de rachats

L'échantillon des données de rachats est constitué de près de 100 000 individus. Voici la liste des variables observées :

Variables obsrvées pour l'explication du rachat	
Variables des rachats	Rachat total ou partiel Part rachetée (si rachat partiel) Date de rachat
Variables de l'assuré	Age Ancienneté Sexe Revenu annuel Catégorie socio-professionnelle Provision mathématique
Variables de l'assureur	TMG Taux de rendement comptable Taux servi aux assurés

FIGURE 9.1 – Liste des variables de rachats observées

Les individus de l'échantillon correspondent à des contrats commercialisés en France uniquement. Le choix de nombreuses variables explicatives repose sur ce fait.

La base de données contient des variables qualitatives : le sexe, la situation maritale, le revenu annuel et la situation locative. Pour les besoins de modélisation, ces variables qualitatives seront remplacées par autant de variables qualitatives que de modalités prises. La modalité prise par un individu correspond à la valeur 1 et les autres à la valeur 0.

Le taux de données manquantes s'exprime comme le rapport entre le nombre d'individus dont au moins une variable est manquante et le nombre d'individus total. Il est ici de 45%.

Les individus dont certaines données sont manquantes nécessitent un traitement spécifique. Un exemple de traitement sera précisé dans la partie 9.3 à la page 78.

Dans un soucis de cohérence des données, les PM sont retraitées de l'inflation. L'indice d'inflation utilisé est décrit dans la partie 9.1.2 à la page 74.

9.1.2 Données économiques

Les données économiques sont utilisées de deux façons : pour le calibrage des SVM et pour le calibrage de la loi de rachat.

L'historique économique s'étend du 27/07/1998 au 31/12/2015.

Le choix des variables économiques doit être en accord avec les caractéristiques des contrats considérés. Ici, les contrats sont uniquement commercialisés en France. Des variables économiques représentatives du marché européen sont donc à privilégier. L'indice CAC 40, le taux actuariel de l'OAT 10 ans française, un indice immobilier européen et l'indice des prix à la consommation semblent être de bons candidats.

Action

Cependant le choix des variables ne peut se faire indépendamment du GSE. Ainsi comme le modèle d'action type 1 est calibré sur les volatilités de l'EUROSTOXX 50 fournies par Bloomberg, il serait incohérent de considérer le CAC 40 dans la partie historique. Comme le montre le graphique suivant, les deux indices sont fortement corrélés et il est donc possible de sélectionner l'EUROSTOXX 50 uniquement :

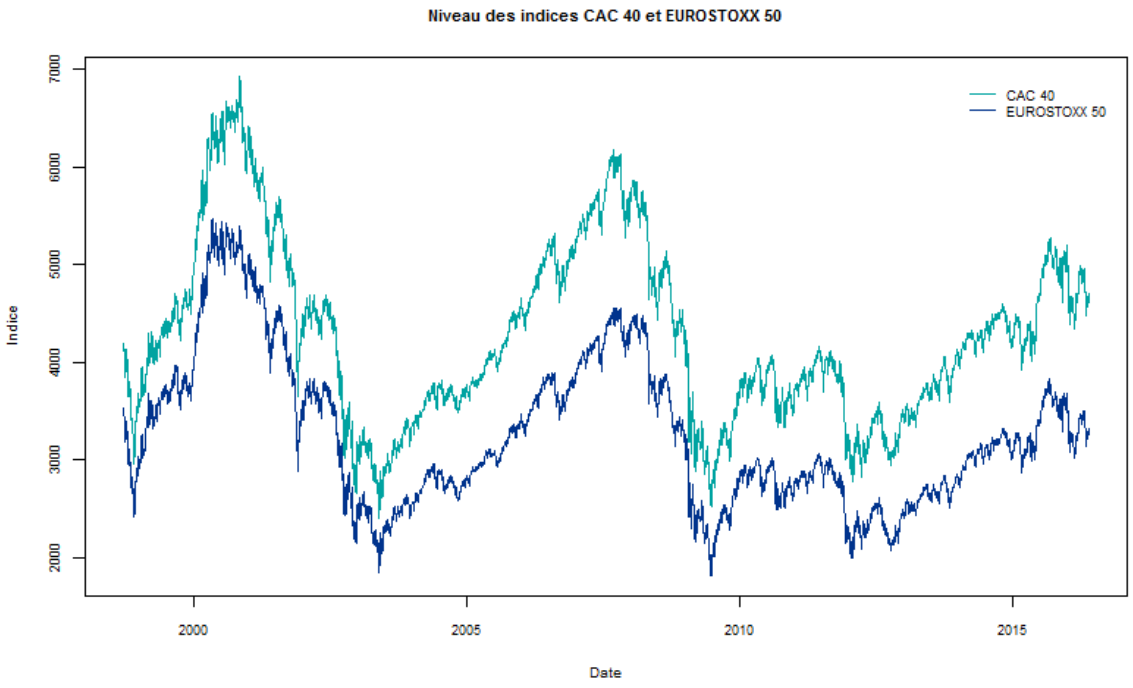

FIGURE 9.2 – Evolution des indices CAC 40 et EUROSTOXX 50

Source : Bloomberg - CAC Index et SX5E Index

Immobilier

Le choix de l'indice immobilier ne peut bénéficier des mêmes remarques que l'indice action. En effet, le GSE se base sur une volatilité proportionnelle à celle du modèle action et donc aucun indice de référence n'est utilisé. Le choix de l'indice se fait ici par jugement d'expert de la façon la plus réaliste possible pour l'explication du rachat. Ainsi l'indice MSCI Real Estate a été retenu. Voici l'historique de l'indice immobilier :

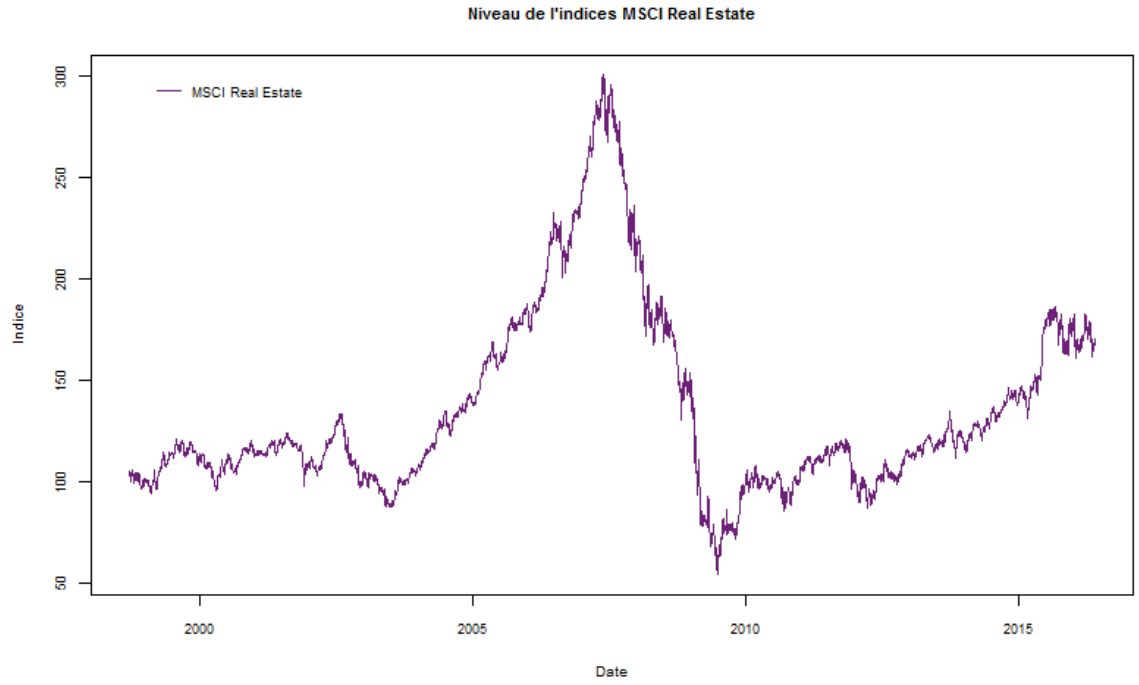

FIGURE 9.3 – Evolution de l'indice MSCI Real Estate

Source : Bloomberg - MXEU0RE Index

Obligataire

Le GSE diffuse la courbe de taux sans risque EIOPA avec ajustement de volatilité au 31/12/2015. Il paraît peu pertinent de supposer qu'un assuré s'appuie sur le taux sans risque EIOPA pour prendre la décision de racheter ou non.

Le taux actuariel de l'OAT 10 ans est un indicateur à la fois accessible et communiqué dans de nombreux. Cependant il ne peut être utilisé en l'état puisqu'il n'est pas comparable aux informations de taux fournies par le GSE lors de la phase de projection.

Pour ce faire, un spread entre l'OAT 10 ans au 31/12/2015 et la courbe EIOPA avec ajustement de volatilité au 31/12/2015 est calculé. Ce spread est ensuite supposé constant dans la projection ALM. Cette hypothèse est justifiée par le fait que le risque de crédit n'est pas modélisé dans la

projection.

Le taux actuariel de l'OAT 10 ans est au 31/12/2015 de $0,995\%$. Le taux au pair d'une obligation sans risque de crédit vis-à-vis de la courbe EIOPA à la même date est de $1,116\%$. Ce taux au pair est défini comme le taux de coupon fixe d'une obligation de nominal 1 et de valeur de marché 1. Le spread ainsi obtenu est la différence de ces deux taux soit $-0,121\%$. Un spread négatif n'est pas un problème étant donné que le spread est une mesure relative du risque de crédit. La courbe EIOPA avec ajustement de volatilité est souvent assez élevée et donne des spreads négatifs sur les titres peu risqués comme l'OAT 10 ans.

Voici l'historique du taux actuariel :

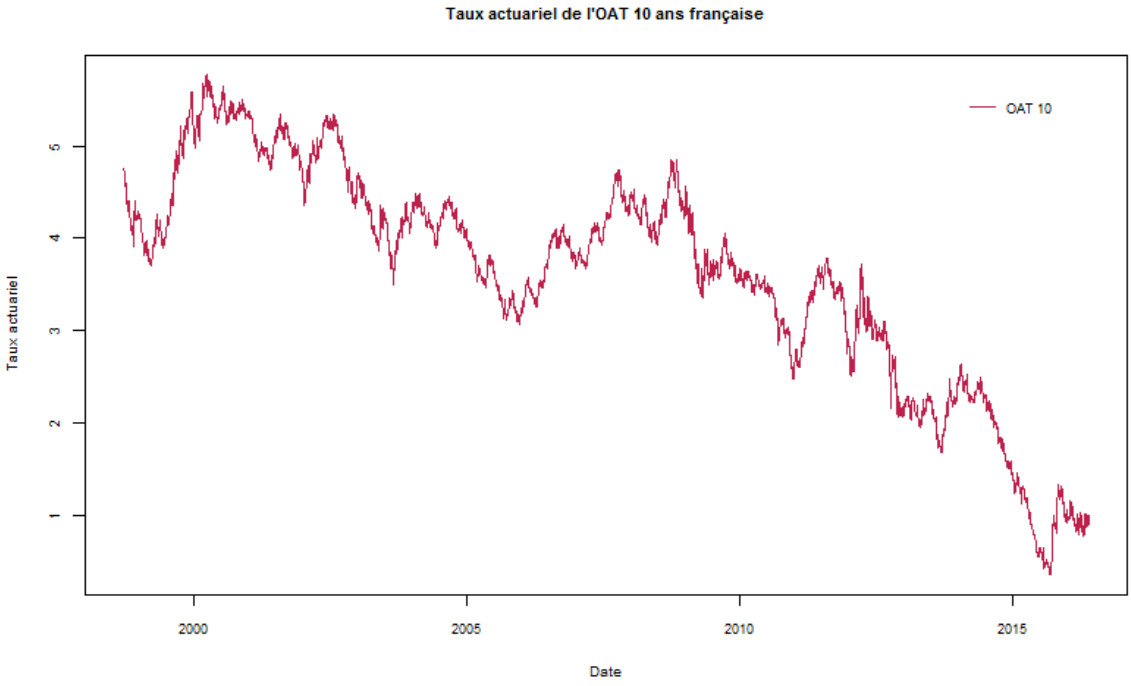

FIGURE 9.4 – Evolution du taux actuariel de l'OAT 10 ans française

Source : Banque de France - Taux quotidiens

Inflation

L'inflation est projetée par le GSE cà l'aide d'un taux d'élasticité et du taux sans risque sur la période de projection. Le modèle est calibré sur les swaps d'inflation. Un indice des prix à la consommation français est ici retenu car il traduit une inflation réalisée et non une inflation anticipée par le marché (à l'inverse des swaps d'inflation). Voici son historique :

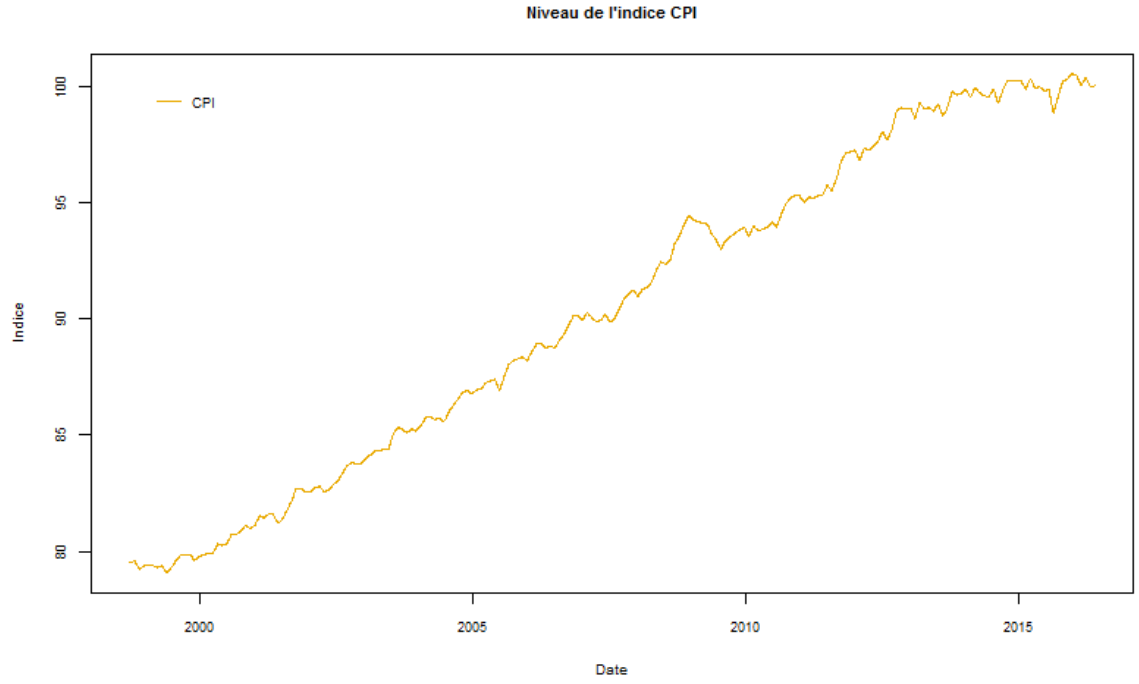

FIGURE 9.5 – Evolution de l'indice CPI français

Source : Agence France Trésor

Périodes de crise de l'historique

Maintenant que les variables économiques ont été sélectionnées, il est nécessaire de repérer les périodes de crise perceptibles sur l'historique retenu.

Trois crises majeurs sont visibles sur l'historique :

i) La crise de la bulle internet

ii) La crise des subprimes

iii) La crise de la dette de la zone euro

Le première crise est la conséquence d'une bulle spéculative. A la fin des années 1990, les valeurs boursières des actions relatives aux technologies de l'information, aux nouveaux médias et aux télécommunications augmentent fortement consécutivement à l'excès d'intérêt et de confiance des investisseurs. De nombreuses entrées en bourse et des fusions de grande ampleur comme celle entre AOL et Time Warner stimulent ce phénomène d'hyper-spéculation. La bulle éclate en mars 2000 avec la remontée des taux et la prise de conscience des dérives de ces nouvelles compagnies. Elle se termine sur l'historique en février 2003.

La seconde crise est due à une prise de conscience de la mauvaise qualité des produits financiers dans lesquels de nombreuses institutions américaines et mondiales ont investi (principalement des CDO). Elle s'étend de juillet 2007 à décembre 2008.

La troisième crise est la conséquence d'une perte de confiance en la gestion et la capacité de remboursement des dettes des pays de la zone euro. Elle a débuté par la dette grecque. Elle s'étend de décembre 2009 à décembre 2011. Cependant les impacts ne sont visibles sur l'historique de l'EUROSTOXX 50 et du MSCI que de mai 2011 à décembre 2011. Elle n'est cependant que peu visible sur l'historique de l'OAT 10 ans. La baisse des taux sur la fin de l'historique est la conséquence du Quantitative Easing mené par la Banque Centrale Européenne.

Le graphique suivant montre les périodes de crise et leur représentation sur l'historique :

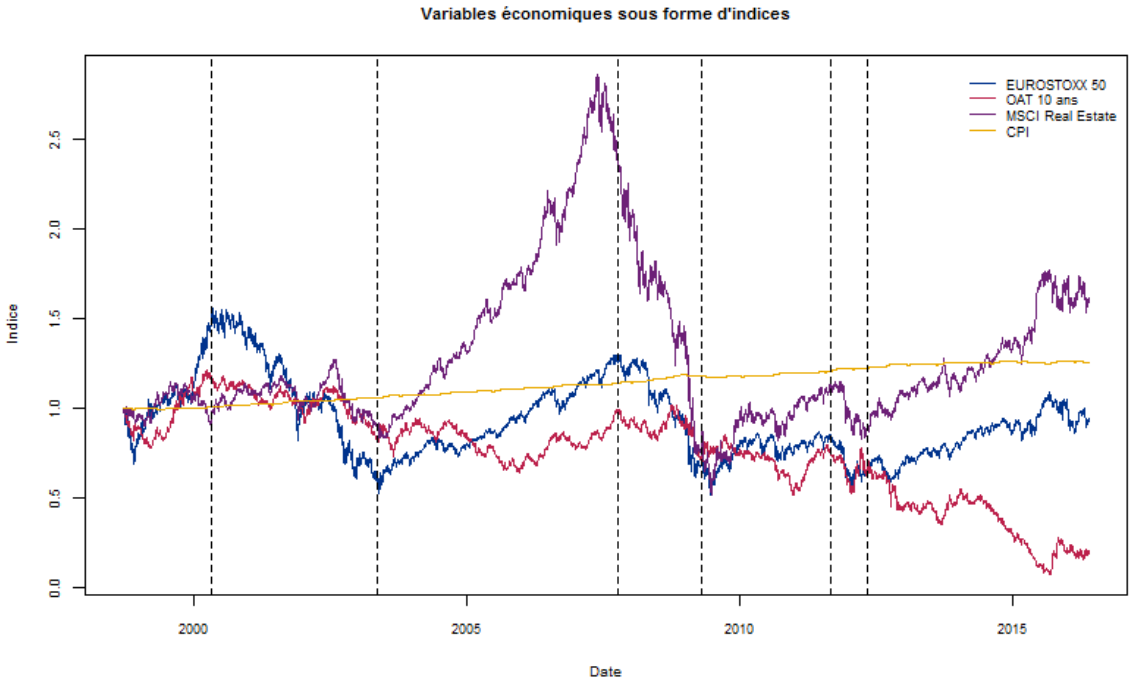

FIGURE 9.6 – Evolution des variables sous forme d'indice et représentation des crises

Il est remarquable que le niveau de l'indice d'inflation et du taux de l'OAT 10 ans ne reflètent pas les périodes de crise. Les différentes crises touchent souvent les marchés de manière isolée et il est rare qu'une crise soit généralisée. La zone observée ne présente pas de crise de taux ou d'inflation et sur la période observée. La crise de la dette de la zone euro n'a pas impactée suffisamment l'OAT française pour déceler une période de crise sur cette variable. Seul le Quantitative Easing est observable mais peut difficilement être interprété comme une crise financière.

Ainsi seuls l'EUROSTOXX 50 et le MSCI sont représentatifs des crises financières :

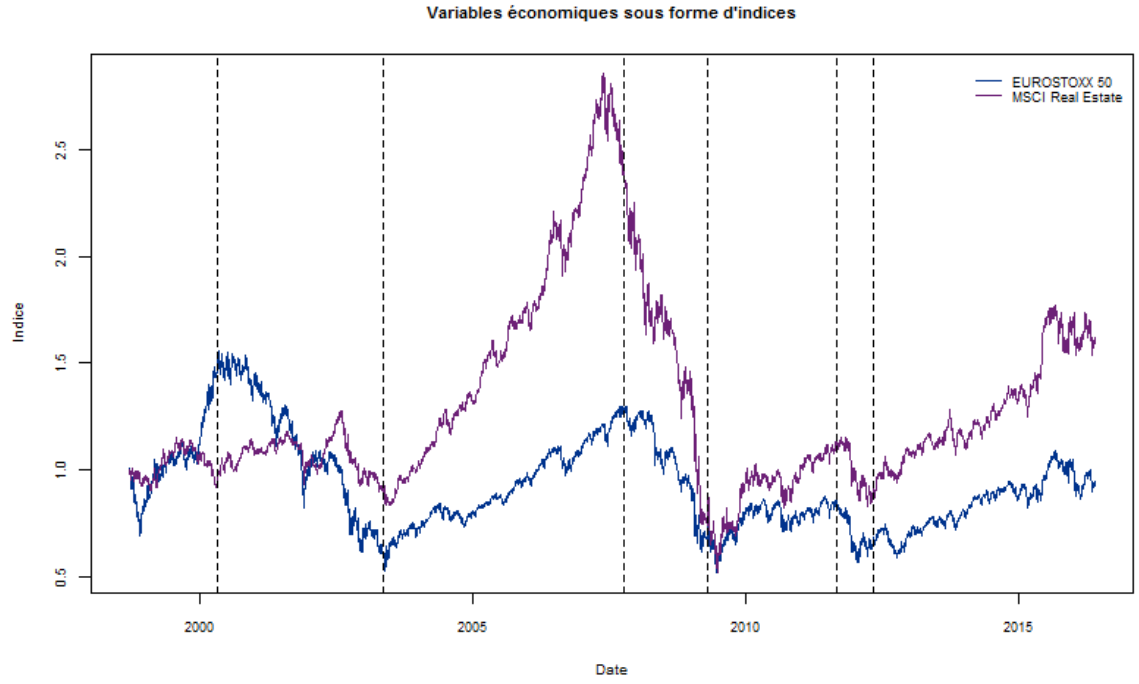

FIGURE 9.7 – Evolution des variables représentatives des crises

9.2 Classification des données manquantes

Lors d'extractions de données de systèmes informatiques, il est fréquent de rencontrer un problème de données manquantes.

Il est nécessaire d'être prudent avec ce type de données lors de leur utilisation dans un cadre de régression et de prévision. En effet, la raison de l'absence de certaines données peut être fondée et dans ce cas le calibrage d'un modèle de régression serait biaisé. Un exemple compréhensible est celui du questionnaire sur les revenus : les individus à revenus élevés ne souhaitent en général pas le divulguer et biaisent donc le questionnaire.

Ceci montre qu'ignorer simplement les données manquantes pourrait conduire à une généralisation aberrante.

Les données manquantes se classent en trois catégories. Pour chacune des catégories, des méthodes d'imputation sont à privilégier.

La première catégorie est celle des données Missing Completely At Random (MCAR). Ce type de donnée manque de façon totalement aléatoire et l'absence est complètement indépendante de la réalisation de la variable. Lors de problèmes de migration de bases de données ou de fiabilité de bases, des données peuvent se perdre et elles seront alors classées MCAR. Si la fréquence de ce type de manquement est maîtrisée, ignorer les individus correspondants ne biaisera pas la régression.

La deuxième catégorie est celle des données Missing At Random (MAR). Le cas de données MCAR étant idéaliste, celui ci est plus réaliste : les données manquantes dépendent de la réalisation

des autres variables de l'individu. Des méthodes statistiques doivent être alors utilisées pour éviter de biaiser l'étude.

La dernière catégorie est celle des données Missing Not At Random (MNAR). Ce cas correspond aux manquements déterministes dépendants de la variable en question. C'est le cas le plus difficile à gérer en pratique.

Une méthode d'imputation de données manquantes peut être appliquée dans le cadre de l'étude : les k plus proches voisins.

9.3 Algorithme des k plus proches voisins

Lors de données manquantes de type MCAR ou MAR, une méthode naïve consiste en l'imputation de la moyenne de la variable manquante sur l'échantillon. Cependant cette méthode amène un fort biais dans le cadre d'un modèle de régression avec une sur-représentation possible de certaine valeur et une modification de la structure même des données. Elle peut cependant être affinée pour faire plus de sens lors de l'utilisation en régression.

L'idée supplémentaire est de sélectionner les individus sur lesquels la moyenne sera faite au lieu de considérer tout l'échantillon à chaque fois. L'algorithme d'imputation se décompose en trois étapes.

Tout d'abord l'échantillon doit être séparé en deux parties : l'ensemble des individus complets (i.e. sans données manquantes) et l'ensemble des individus dont au moins une donnée est manquante.

Ensuite un entier k est une distance doivent être choisis. L'entier k doit être inférieur au cardinal de l'ensemble des individus complets. Pour chaque individu incomplet, l'ensemble des individus complets est parcouru et la distance entre l'individu complet et l'individu incomplet est calculée (sur les variables connues uniquement). Les k plus proches voisins sont finalement conservés.

Maintenant que les k plus proches voisins sont identifiés, les données manquantes sont imputées par la moyenne de la variable sur ces voisins.

Finalement c'est une méthode d'imputation non paramétrique qui repose sur l'hypothèse de similarité des individus. Cependant le choix de k et de la distance sont des paramètres laissés au choix de l'utilisateur. La méthode n'est pas facilement utilisable pour des variables qualitatives.

La base de données étudiée dans le cadre de l'étude présente un fort taux de données manquantes : 46%. La grande majorité de ces données manquantes sont des données qualitatives (catégorie socio-professionnelle et revenu annuel). La méthode est donc difficilement applicable. Ainsi les données manquantes sont simplement ignorées pour l'application de l'étude.

Chapitre 10

Application de la régression logistique

10.1 Résultats

Le calibrage du modèle logit permet de définir l'espérance conditionnelle du rachat en fonction des variables explicatives retenues. La forme du modèle permet d'interpréter le poids et le sens qu'exerce chaque variable explicative sur le rachat.

Pour se faire, il est important de remarquer que la fonction $\frac{e^{\cdot}}{1+e^{\cdot}}$ est croissante sur $\mathbb{R}$. Ainsi, en raison de la forme du modèle logit, plus $\beta' x$ est grand et plus l'assuré rachète, vice-versa.

La normalisation des variables permet de simplifier cette analyse [1]. Ainsi toutes les variables ont les mêmes poids et signe. Un β_i positif signifie donc que la croissance d'une variable, toutes choses égales par ailleurs, augmentera le rachat et inversement. Et plus un β_i est grand en valeur absolue, plus le modèle sera sensible à la variable associée.

Les résultats du calibrage permettent ici de valider quelques idées a priori sur le rachat. Ainsi les facteurs dominants (en termes de valeurs absolues) sont le taux servi brut, l'ancienneté, le TMG et le taux actuariel de l'OAT 10 ans. Il est aussi remarquable que le sexe ne joue presque aucun rôle sur la décision de rachat. Les signes permettent de montrer que plus le taux servi brut, le TMG et le taux actuariel de l'OAT 10 ans sont élevés et moins l'assuré aura tendance à racheter (β positifs). Et inversement pour l'ancienneté ; plus elle est élevée et plus l'assuré rachète (β négatif).

Si ces variables valident des idées a priori, d'autres vont dans le sens contraire. C'est le cas notamment de l'indice EUROSTOXX 50. Il serait légitime de penser qu'une hausse du marché action conduirait à des rachats afin d'améliorer le rendement de l'épargne des assurés. Finalement c'est le cas contraire qui est traduit dans le portefeuille étudié. Ceci peut s'expliquer par la périodicité des performances action retenue. Une performance trimestrielle peut paraître trop court-terme pour observer des rachats sensibles au niveau action.

Si les résultats paraissent cohérent a priori, une étude plus quantitative de la qualité du modèle s'impose.

1. Elle permet également de préparer l'arrivée du modèle hybride

10.2 Qualité du modèle

Lors de la mise en œuvre de modèle de classification ou de régression il est important de ne pas négliger la validation du modèle. L'étude de qualité possède en général trois objectifs : vérifier que le modèle traduit bien une relation de causalité, tester le pouvoir de généralisation et choisir un modèle parmi plusieurs propositions. Quelques indicateurs tels que le taux de mauvais classés ou l'erreur quadratique viennent rapidement à l'esprit. Si ce sont des indicateurs qu'il est nécessaire de maîtriser et de surveiller, ils ne sont souvent pas suffisants et des méthodes de validation plus élaborées sont préférables.

En effet, si un faible taux de mauvais classés en classification est important, il ne garanti en rien la qualité de généralisation du modèle. L'exemple d'une situation où les classes ne sont pas représentées en proportions équivalentes le montre facilement. Soit un échantillon est composé de 1000 individus et que parmi ces individus 980 prennent la modalité 1 alors que les 20 derniers prennent la modalité -1. Si le modèle de classification renvoi toujours 1, le taux de mauvais classés sera de 2% : un modèle qui visiblement n'est pas adapté à la situation montre un taux tout à fait acceptable.

Il en est de même pour l'évaluation de la qualité d'un modèle de régression par la simple erreur quadratique. L'utilisation d'une seule erreur est bien trop dépendante du choix de l'échantillon d'apprentissage et la qualité de généralisation (i.e. de prévision) n'est pas vérifiée lors de l'évaluation de l'erreur quadratique. Une méthode de back-testing[2] simple améliore ce point et vérifie la qualité de généralisation mais reste cependant fragile.

Des méthodes de validation plus sophistiquées sont ici présentées et seront appliquées à chaque étape de modélisation.

10.2.1 Taux d'erreur et Cross-Validation

Présentation du concept

Dans le cadre d'une classification, le taux de mauvais classés est donc insuffisant. La première amélioration est d'estimer quatre grandeurs pour éviter les problèmes d'échelle en cas de représentation inégales des classes. Ces quatre grandeurs sont :

i) Le nombre de vrais positifs : nombre d'individus dont la prévision est 1 et la vraie valeur est 1

ii) Le nombre de vrais négatifs : même chose que le taux de vrais positifs mais dont la valeur est -1

iii) Le nombre de faux positifs : nombre d'individus dont la prévision est 1 et la vraie valeur est -1

iv) Le nombre de faux négatifs : même chose que le taux de vrais positifs en inversant les valeurs 1 et -1

Dans l'exemple simple d'introduction, le nombre de vrais négatifs est de 0 soit 100% d'erreur sur ce segment d'individus. Ce deuxième indicateur améliore bien la compréhension de la qualité

2. Le back-testing consiste en la séparation de l'échantillon en deux sous-échantillon (souvent 80% - 20%) afin d'effectuer le calibrage sur une part de l'échantillon (80%) et de tester le pouvoir de généralisation sur l'autre part de l'échantillon (20%)

de généralisation.

Deux taux sont définis pour évaluer la qualité : la sensibilité (vrais positifs rapportés au nombre de positifs initiaux) et la spécificité (vrais négatifs rapportés au nombre de négatifs initiaux). La sensibilité de l'exemple est donc de 100% mais la spécificité de 0% : le modèle n'est donc pas adapté.

Cependant, ces indicateurs sont fortement dépendants des échantillons initiaux d'apprentissage et de test. Une seconde amélioration est d'estimer l'espérance de l'erreur ou même estimer la distribution de l'erreur. La méthode de Cross-Validation permet d'estimer l'espérance de l'erreur. Le principe de la méthode est de séparer aléatoirement l'échantillon initial en 2 parties : un échantillon d'apprentissage et un échantillon de test. La loi est calibrée sur l'échantillon d'apprentissage et les taux d'erreur sont calculés sur l'échantillon de test. Finalement k erreurs sont calculées et la moyenne empirique est un estimateur de l'espérance de l'erreur. Le choix de l'entier k dépend principalement de la taille de l'échantillon initial. L'erreur finalement obtenue est plus robuste car elle n'est pas calculée sur un unique cas particulier.

Application à la régression logistique

La méthode de Cross-Validation a été appliquée pour l'évaluation de la qualité du modèle logit dans le cadre du rachat. L'application a été réalisée à l'aide du logiciel R. Le choix du logiciel ou langage doit tenir compte du nombre de données.

A chaque itération, l'échantillon a été séparé aléatoirement en deux parites : 80% de l'échantillon initial rempli le rôle d'échantillon d'apprentissage et 20% le rôle d'échantillon de test. 20 itérations ont été réalisées pour estimer les différents taux d'erreur. Voici les résultats obtenus :

```
[1] "Le taux d'erreur de prediction du modele est de : 7.45 %."
[1] "Le taux de vrais negatifs est de : 93.97 %."
[1] "Le taux de vrais positifs est de : 90.53 %."
```

FIGURE 10.1 – Taux d'erreur estimés pour le modèle logit

Le taux d'erreur global du modèle est de $7,45\%$. Communément, un taux d'erreur est acceptable s'il est inférieur à 5%. Le modèle logit paraît donc trop imprécis sur ce seul critère. L'analyse des taux de vrais négatifs et vrais positifs montre que le modèle sous-estime plus souvent le rachat qu'il ne le sur-estime.

Ces résultats doivent être interpréter avec précaution. Si les taux d'erreurs paraissent trop élevé, ils ne veulent pas pour autant dire que le modèle logit est inadapté à la modélisation du rachat. En effet, le cadre de l'étude introduit des contraintes de choix de variables. Ces contraintes ont forcément un impact sur la qualité du modèle. Le taux de chômage est un exemple de contrainte. Différentes études sur la modélisation du rachat ont montré que le taux de chômage explique en partie le phénomène. L'erreur est donc mécaniquement majorée dans le cas présent.

10.2.2 Courbe ROC

Présentation du concept

Si les deux indicateurs que sont la sensibilité et la spécificité améliore le simple taux de bien classés, il est nécessaire de savoir arbitrer entre deux indicateurs, autrement dit réaliser une optimi-

sation multi-objectif. Un exemple permet de comprendre l'enjeu : soient un modèle A de sensibilité 95% et de spécificité 85% et un modèle B de sensibilité 90% et de spécificité 90%. A partir de ces deux informations, comment choisir le meilleur modèle ? Au sens de Pareto, les deux modèles sont équivalents et aucun n'est meilleur que l'autre.

Cependant, la courbe Receiver Operating Characteristics (ROC) propose une méthode de choix un peu plus sophistiquée entre deux modèles qui allie sensibilité et spécificité. Dans le cadre de la régression logistique par exemple, la classification repose sur un score : une espérance conditionnelle est modélisée et une règle de classification est fixée pour connaître la valeur prise par la variable réponse. Ainsi si $\mathbb{E}[Y|X = x] > 0,5$, alors $Y = 1$, sinon $Y = -1$. La courbe ROC repose sur le choix du seuil $0,5$. Elle est définie de $[0;1]$ dans $[0;1]$ et elle associe la sensibilité en ordonnée à $(1 -$ spécificité$)$ en abscisse. Pour déterminer cette courbe, il suffit finalement de calculer chacune des deux grandeurs pour plusieurs valeurs de seuil entre 0 et 1. De plus elle passe nécessairement par les points $(0;0)$ et $(1;1)$.

Une fois la courbe construite, l'interprétation de la qualité se fait grâce à la position de la courbe : plus elle est proche de la première bissectrice et moins le modèle est de bonne qualité, plus elle se rapproche de la droite $y = 1$ et plus le modèle est de bonne qualité.

Ainsi un score est meilleur qu'un autre si ça courbe ROC est toujours au dessus. Il est bon de noté que la position relative de deux courbes ROC peut s'inverser et donc que tous les modèles ne peuvent pas être ordonnés.

Un dernier indicateur lié à la courbe ROC est l'Area Under the Curve (AUC). L'AUC est l'air sous la courbe ROC : elle mesure la performance du score. En cas d'intersection entre deux courbes ROC, l'AUC peut permettre de choisir le modèle et cette fois l'optimisation est mono-objectif.

Application à la régression logistique

Sur le même principe de Cross-Validation, la courbe ROC associée au modèle logit a été estimée à l'aide du logiciel R. L'intervalle $[0;1]$ a été subdivisé uniformément avec un pas de $0,05$. La sensibilité et la spécificité ont été estimées pour chacun des seuils de la subdivision. Voici la courbe ROC obtenue :

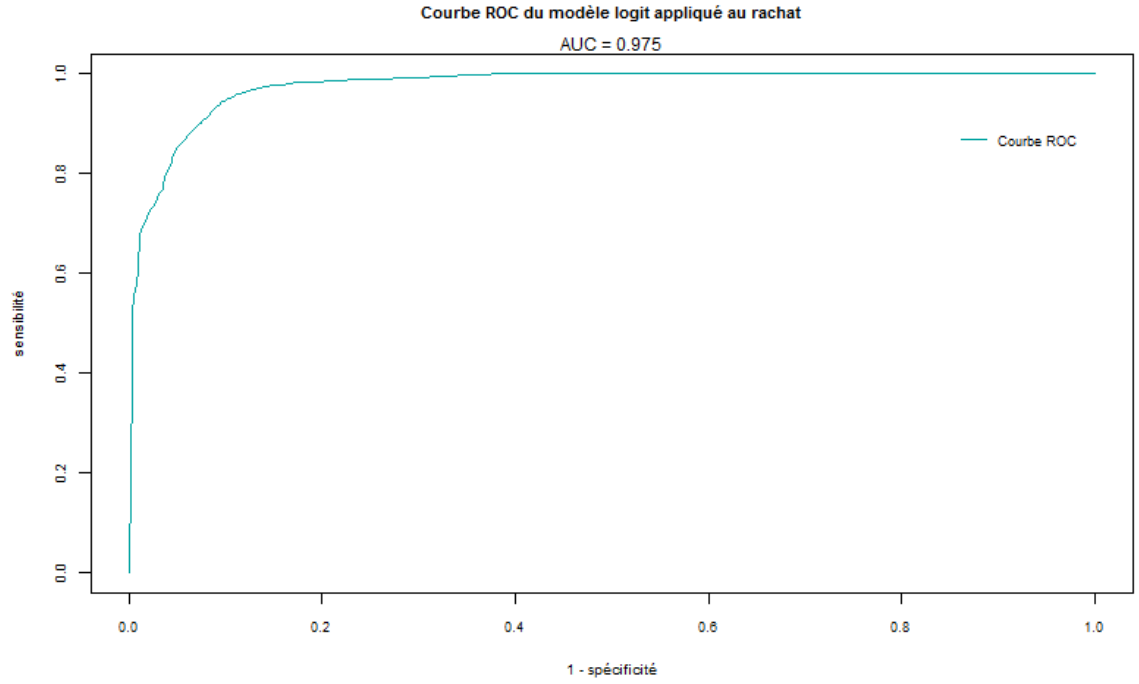

FIGURE 10.2 – Courbe ROC estimée pour le modèle logit

Plus la courbe ROC s'éloigne de la première bissectrice, plus le modèle est efficace. Ici la courbe ROC est clairement éloignée de la première bissectrice : le modèle paraît adapté à la situation. De plus le coefficient AUC est de $0,975$. En pratique un score AUC supérieur à $0,9$ est acceptable ; le modèle logit valide donc ce critère.

Finalement le modèle logit paraît adapté à la situation mais il reste trop imprécis pour être accepté. Ces imprécisions peuvent venir de deux sources principales : le choix des variables explicatives ou la structure même du modèle avec des hypothèses de symétrie de la fonction de lien et de linéarité. Ce dernier point a mené à la proposition d'un modèle hybride.

Chapitre 11

Mise en place de la proposition de modèle hybride

11.1 Difficultés pratiques

Afin d'améliorer la qualité du modèle logit, un modèle hybride est ici proposé. L'idée est de combiner estimation non paramétrique par projection et modèle logit pour s'affranchir de la linéarité du modèle.

La première difficulté pratique rencontrée est le choix de la base hilbertienne. De nombreuses bases hilbertiennes sont connues en dimension 1 : la base trigonométrique, la base de Haar, la base des polynômes de Tchebychev, etc. Cependant des bases similaires en dimension supérieure sont bien moins connues. Une solution est de supposer le modèle additif. Dans ce cas la fonction à estimer de n variables s'exprime comme la somme de n fonctions à 1 variable. La projection s'exprime alors comme une somme de projection et une base hilbertienne en dimension 1 suffit à l'étude. Dans ce cadre la base trigonométrique a été retenue.

Une limite associée est l'ensemble de valeur des variables explicatives. En effet la projection nécessite que l'ensemble des valeurs prises par chaque variable explicative soit un convexe fermé. L'utilisation de la base trigonométrique requiert que les variables soient à valeurs dans $[0;1]$. Une normalisation de la base initiale est possible pour le calibrage.

Voici la définition de la base trigonométrique :

$$\forall j \in \mathbb{N}^* \ \varphi_j(x) = \begin{cases} \sqrt{2}.\cos(2\pi.j.x) \text{ si j est pair} \\ \sqrt{2}.\sin(2\pi.j.x) \text{ si j est impair} \end{cases}$$

La deuxième difficulté pratique est la taille des données. Le passage au modèle hybride multiplie par N [1] la taille de la matrice de données. L'algorithme de calibrage peut être glouton en mémoire vive et un travail d'amélioration de l'algorithme serait une piste d'amélioration a priori.

De plus les variables qualitatives ont été transformées en variables binaires, la transformation par la base trigonométrique n'est donc pas applicable. Par exemple, si le revenu mensuel est de

1. Niveau de troncature retenu

"Moins de 50 Keuros", la variable quantitative associée vaut 1 et les coefficients de la base valent $\sqrt{2}$ si j est pair et 0 si j est impair. Ces variables ne sont donc pas transformées et leur valeur initiale est conservée.

Pour des raisons de simplifications, le modèle est supposé additif et le niveau de troncature est supposé égal pour chaque variable transformée.

11.2 Résultats

La transformation des variables explicatives par la base hilbertienne rend ici impossible toute analyse a priori analogue à celle pratiquée pour le modèle logit.

Cette transformation permet d'augmenter la dimension de l'espace des variables explicatives et capter des tendances non linéaires. En revanche, une variable n'est plus associée à un seul β_i mais à autant de β_i que le niveau de troncature retenu. Ainsi l'ancienneté est maintenant transformée en M variables et est associée à M coefficients.

L'analyse des coefficients a priori était un point confortable pour un assureur puisqu'elle permettait de vérifier que la loi calibrée était en accord avec les idées a priori. Le modèle hybride introduit un effet "boîte noire" qui certes améliore le modèle mais le rend moins facile à comprendre et à analyser.

Finalement seule une étude quantitative de la qualité du modèle permettra de le juger.

11.3 Taux d'erreur et Cross-Validation

La mise en pratique a été réalisée dans les mêmes conditions que la mise en pratique du modèle logit.

Le niveau de troncature retenu est $N = 5$. Il a été choisi de façon à optimiser la significativité des coefficients.

Voici les taux obtenus :

```
[1] "Le taux d'erreur de prediction du modele est de : 3.18 %."
[1] "Le taux de vrais negatifs est de : 97.49 %."
[1] "Le taux de vrais positifs est de : 95.82 %."
```

FIGURE 11.1 – Taux d'erreur estimés pour le modèle hybride

Le premier constat est l'amélioration significative du taux d'erreur par rapport au modèle logit simple : passage de $7,45\%$ à $3,18\%$. De même les taux de vrais négatifs et de vrais positifs sont améliorés de respectivement $4,5\%$ et $5,3\%$.

Cette amélioration s'explique principalement par le relâchement de l'hypothèse de linéarité.

Seule une base hilbertienne est ici testée, mais les paramètres sont clairement sensibles au choix de la base. Si théoriquement le choix de la base n'a pas d'impact sur la convergence du modèle, en

pratique le choix du niveau de troncature, la précision et les paramètres dépendent fortement de la base hilbertienne retenue. La méthode de choix de la base est donc une piste d'optimisation de ce modèle.

Il est également important de retenir qu'un taux d'erreur de $3,18\%$ reste élevé en pratique. Des études et mémoires qui se sont intéressés à la modélisation du taux de rachat ont déjà affiché des taux d'erreur inférieurs. Mais le perfectionnement de l'erreur de prévision se fait souvent au prix de la mise en production du modèle de rachat au sein d'un modèle de projection ALM. Il est en fait nécessaire d'arbitrer entre pratique et précision, mais le taux d'erreur reste ici tout de même maîtrisé.

11.4 Courbe ROC et coefficient AUC

La courbe ROC a été estimée dans les mêmes conditions que celle du modèle logit. Voici la courbe ROC obtenue :

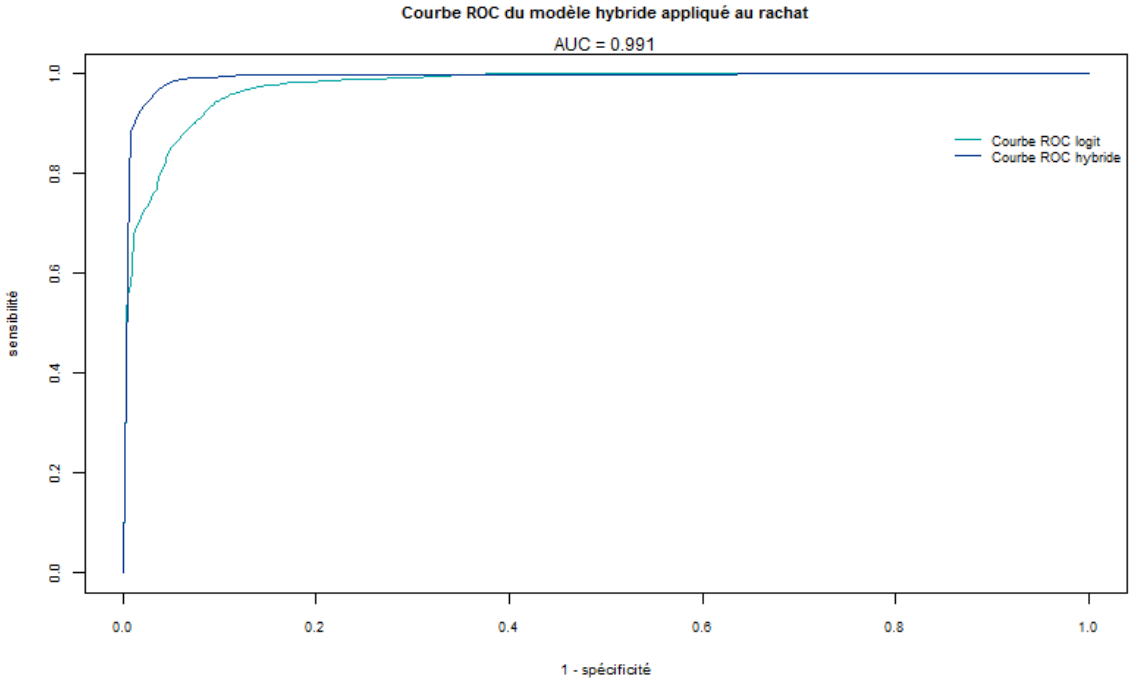

FIGURE 11.2 – Courbe ROC estimée pour le modèle hybride

L'amélioration de la précision se retrouve également sur la courbe ROC : elle est plus éloignée de la première bissectrice que la courbe ROC du modèle logit. Elle est même toujours au dessus de la courbe ROC du modèle logit. Le modèle hybride est donc meilleur.

Le coefficient AUC passe également de $0,975$ à $0,991$. Sur ces critères le modèle hybride est clairement acceptable d'un point de vue opérationnel.

Chapitre 12

Mise en place du modèle à états

12.1 Calibrage du SVM

La théorie des SVM est ici utilisée dans le but de séparer les états crise et non-crise. Pour se faire seules les performances trimestrielles de l'EUROSTOXX 50 et du MSCI Real Estate ont été considérées. En effet, comme vu dans la partie 9.1.2 à la page 75, seuls ces deux indices permettent de détecter certaines périodes de crise dans l'historique considéré.

Une fois les variables sélectionnées, il est nécessaire de choisir les paramètres de calibrage du SVM à savoir le noyau, les paramètres associés et le coût. Deux noyaux ont été testés dans le cadre de l'étude : le noyau linéaire et le noyau gaussien.

La qualité du SVM a été évaluée à travers les taux d'erreur, de vrais positifs et de vrais négatifs par Cross-Validation. La courbe ROC ne peut être définie dans le cadre des SVM puisque ce ne sont pas des modèles probabilistes.

Voici les taux d'erreur obtenus pour le noyau linéaire :

```
[1] "Le taux d'erreur de prediction du modele est de : 10.2 %."
[1] "Le taux de vrais negatifs est de : 94.11 %."
[1] "Le taux de vrais positifs est de : 80 %."
```

FIGURE 12.1 – Taux d'erreur estimés pour le SVM linéaire

Le taux d'erreur du modèle est de $10,2\%$. Ce taux d'erreur est trop élevé mais il doit être analysé avec attention. Tout d'abord le taux de vrais négatifs est de $94,11\%$ ce qui reste acceptable et le taux de vrais positifs est de 80% ; c'est donc la prévision des états de crise qui est mauvaise.

Ici seules deux dimensions sont retenus pour le SVM et ceci permet d'étudier graphiquement les résultats :

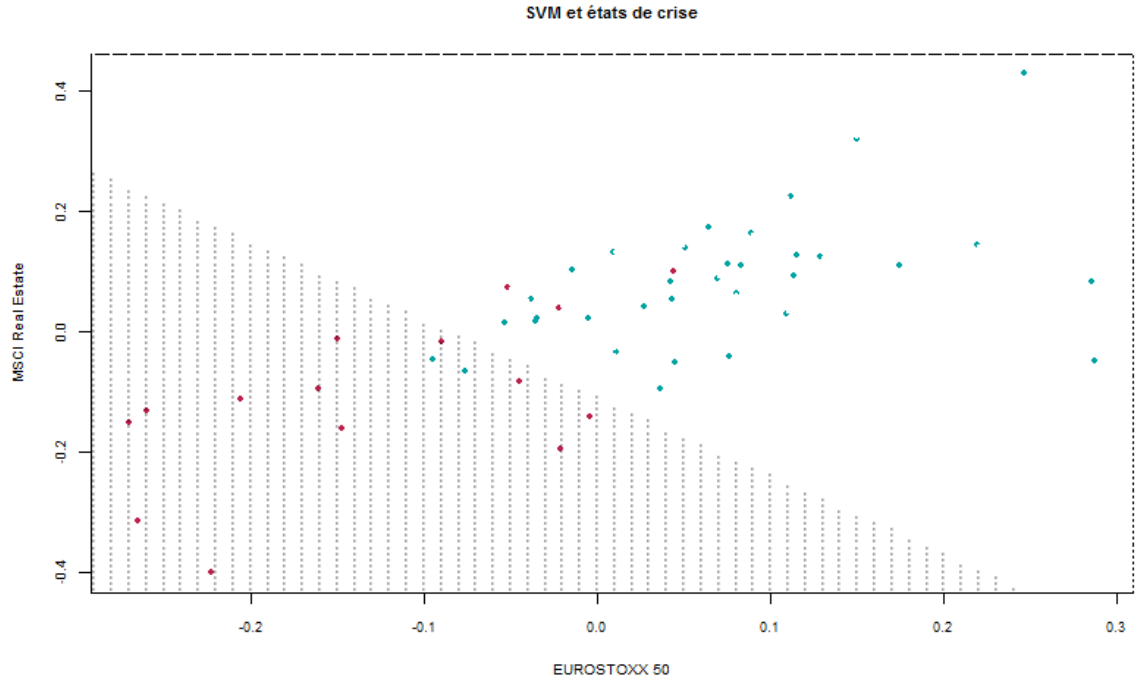

FIGURE 12.2 – Tracé du nuage de points et de la frontière linéaire

Ce graphique montre en rouge les points de crise, en vert les points de non-crise et en grisé la zone de crise de la frontière linéaire.

Finalement seuls 5 points sont mal classés par le SVM et la frontière paraît adaptée à la situation. En fait l'historique ne contient qu'une cinquantaine de points et le moindre point mal classé augmente significativement l'erreur du modèle.

Ceci incite à porter un regarde critique sur les résultats de calibrage du SVM et sur la prédiction future des états de crise. Cependant, même si l'historique disponible dans le cadre de l'étude est restreint, une autre étude s'est déjà intéressée à l'utilisation des SVM pour la détection des états de crise[A. Chaudhuri et al., 2004]. Cette étude s'intéresse plus particulièrement aux crises relatives au change de monnaie et les SVM donnent des résultats prometteurs.

la limite la plus importante est en fait la diversité des types de crises économiques et leur faible représentation dans l'échantillon. Ainsi les crises de change ou de crédit ne sont pas du tout prises en compte par le SVM. Et elles peuvent pourtant avoir un impact significatif sur le rachat.

Il est cependant intéressant de regarder les résultats obtenus avec le noyau gaussien :

```
[1] "Le taux d'erreur de prediction du modele est de : 14.28 %."
[1] "Le taux de vrais negatifs est de : 94.11 %."
[1] "Le taux de vrais positifs est de : 66.66 %."
```

FIGURE 12.3 – Taux d'erreur estimés pour le SVM gaussien

Le taux d'erreur est détériorer de 4% et le taux de vrais négatifs est identique tandis que le taux de vrais positifs baisse d'un peu plus de 13%.

Les mêmes remarques que pour le noyau linéaire peuvent être faites dans le cas gaussien. Finalement la différence de qualité entre les deux modèles est difficile à évaluer au regard du faible nombre de points du nuage.

L'étude graphique amène cependant des remarques plus intéressantes :

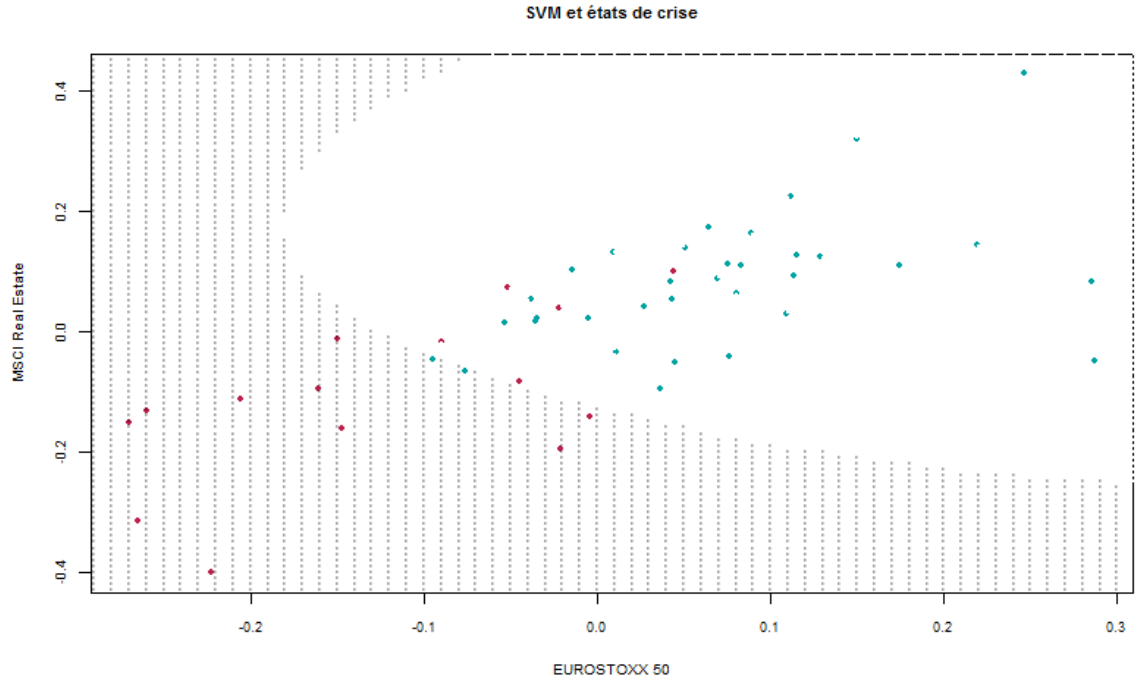

FIGURE 12.4 – Tracé du nuage de points et de la frontière gaussienne

La différence de taux d'erreur s'explique par la mauvaise classification de deux points de crise supplémentaires. Cependant la forme de la frontière caractéristique des fonctions gaussiennes montre un pouvoir de généralisation a priori trop faible. En effet la forme en cloche se referme sur elle même et un point de forte performance action et immobilier sera considéré comme un état de crise. Cette généralisation va contre toute intuition a priori et s'explique par le manque de points dans la partie haute du graphique. un regarde critique doit donc être porté sur les résultats et ce graphique montre que le noyau gaussien n'est pas adapté à la situation.

Ces donc finalement le noyau linéaire qui est retenu pour la détection des états de crise futurs.

12.2 Intégration au modèle hybride

12.2.1 Taux d'erreur et Cross-Validation

En parallèle du calibrage du SVM, le modèle hybride a été calibré deux fois sur deux sous-échantillons séparés afin de donner lieu au modèle à états.

Ainsi l'échantillon initial a été séparé en deux morceaux distincts : l'échantillon de crise et l'échantillon de non-crise.

Le modèle hybride a ensuite été calibré sur chacun de ces sous-échantillons dans les mêmes conditions que le modèle à hybride simple. Les erreurs de prévisions ont été estimées par Cross-Validation en utilisant le bon modèle hybride pour chaque point de test.

Voici les résultats obtenus :

```
[1] "Le taux d'erreur de prediction du modele est de : 2.32 %."
[1] "Le taux de vrais negatifs est de : 97.99 %."
[1] "Le taux de vrais positifs est de : 97.22 %."
```

FIGURE 12.5 – Taux d'erreur estimés pour le modèle à états

Finalement le passage au modèle à états permet d'améliorer sensiblement le taux d'erreur avec un passage de $2,32\%$ à $2,93\%$. Le taux de vrais négatifs est quasi inchangé alors que le taux de vrais positifs est amélioré de $1,4\%$. Ces améliorations montrent que la sensibilité des assurés aux crises économiques est mieux captée par le modèle à états.

L'amélioration est finalement sensible mais l'impact sur le bilan économique sera à surveiller.

12.2.2 Courbe ROC et coefficient AUC

La courbe ROC a été estimée dans les mêmes conditions que celle du modèle logit. Voici la courbe ROC obtenue :

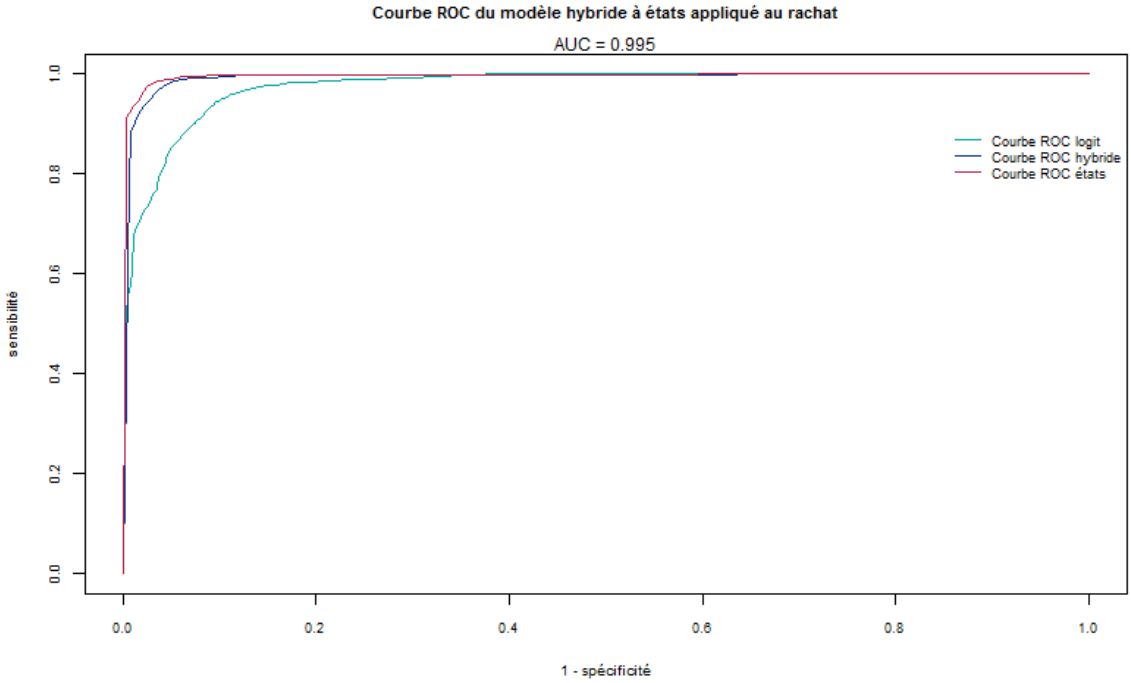

FIGURE 12.6 – Courbe ROC estimée pour le modèle à états

L'amélioration de la précision sur la courbe ROC est très sensible. Elle est toujours au dessus de la courbe ROC du modèle hybride mais elles sont difficiles à distinguer.

Le coefficient AUC passe également de $0,991$ à $0,995$. Il est légèrement au dessus de celui du modèle hybride. La faible amélioration est assez logique étant donné que les modèles sont de même typologie.

Finalement le calibrage du modèle à états peut bénéficier des mêmes remarque que le modèle hybride. Mais sont utilisation dans la projection ALM nécessite un regarde critique lié au SVM.

Quatrième partie

Etude d'impacts

Chapitre 13

Résultats de la modélisation type QIS 5

13.1 Best Estimate

La loi de rachat initialement utilisée dans la projection ALM est l'ajout d'une composante structurelle et d'une composante conjoncturelle. La composante structurelle est calibrée par ancienneté sur l'historique : elle correspond au rachat moyen. La composante conjoncturelle est quant à elle basée sur le QIS5 : une loi moyenne est retenue.

La projection est réalisée sur un horizon de 40 ans à pas semestriel. Le calcul est réalisé sur la base de 500 scénarios économiques. Voici le BE obtenu :

Modélisation	BE
QIS 5	2 579 240 648

FIGURE 13.1 – BE calculé avec la modélisation type QIS 5

Un moniteur de flux de BE projetés est également donné, il représente la moyenne sur l'ensemble des scénarios des flux de BE déflatés :

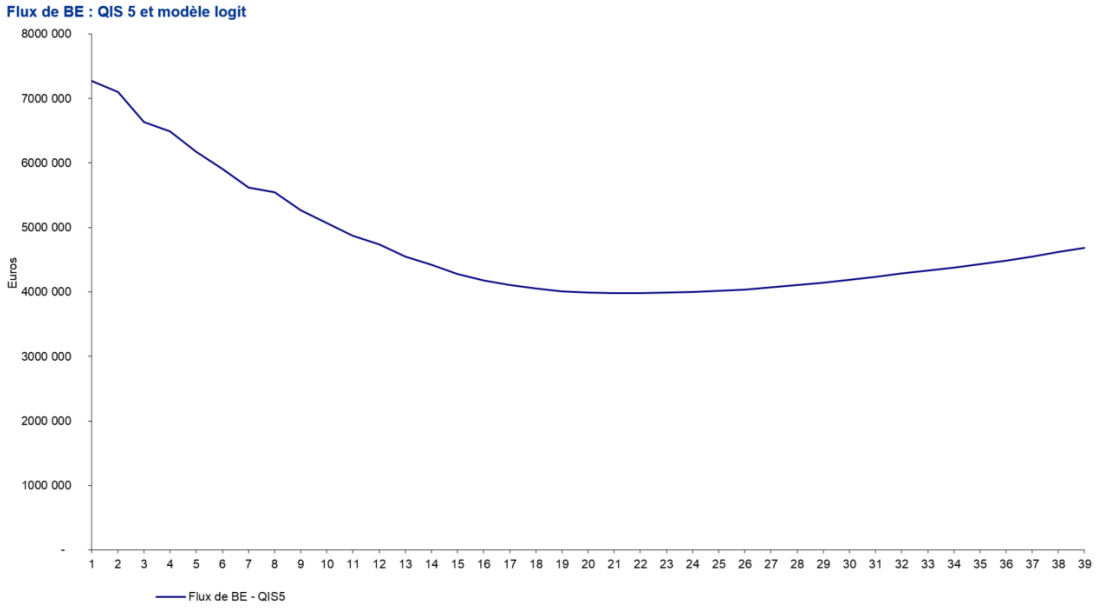

FIGURE 13.2 – Graphique des flux de BE - Modélisation QIS 5

Chapitre 14

Impacts sur le Best Estimate

Si les modèles proposés arrivent à prévoir le rachat avec des taux d'erreur satisfaisants, il est nécessaire d'étudier leur comportement au sein d'un modèle de projection de flux. Le but de l'étude étant de proposer une méthode de meilleure estimation du rachat en environnement Solvabilité II, le BE est au centre des études d'impacts.

Ainsi chaque loi est implémentée dans l'outil ALM DEvent et une sensibilité du BE au choix de la loi est réalisée. La sensibilité est évaluée sur la base du BE et de l'évolution des flux de BE.

14.1 Modèle logit

Voici le BE obtenu pour le modèle logit :

Modélisation	BE
QIS 5	2 579 240 648
Modèle logit	2 590 283 213

FIGURE 14.1 – BE calculé avec le modèle logit

L'impact sur le BE est de +11 M €. L'impact relatif sur le BE de référence [1] est de $+0,5\%$. Cet impact traduit une sous-estimation du risque de rachat dans la modélisation type QIS 5 mais reste toutefois contrôlé en termes de montant.

Une étude du moniteur de flux permet de visualiser la modification du BE de manière plus fine. Voici le moniteur obtenu :

1. Le BE de référence est le BE QIS 5

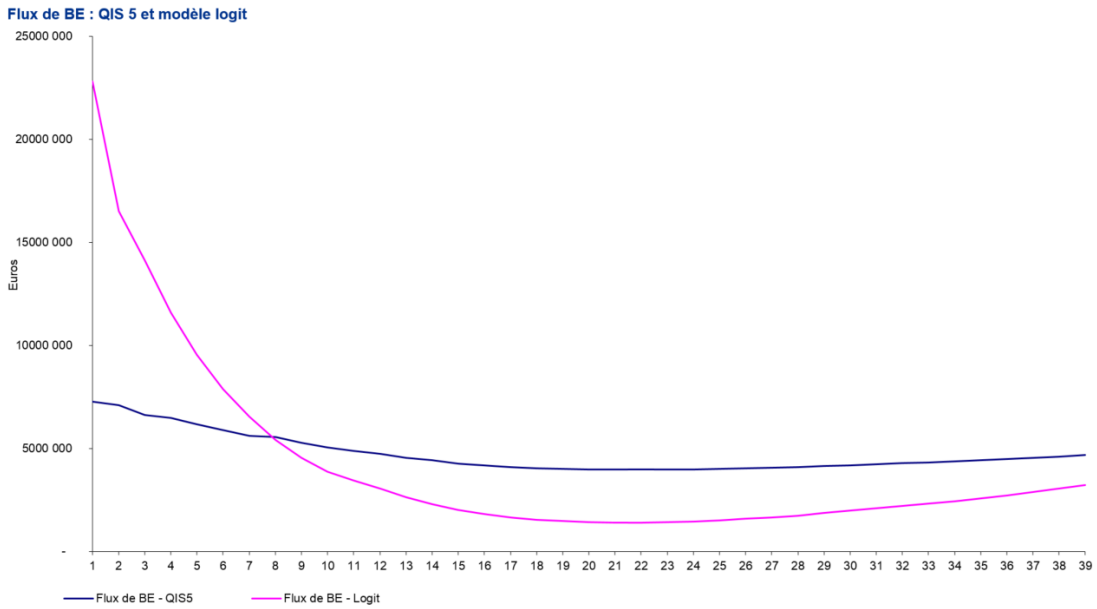

FIGURE 14.2 – Graphique des flux de BE - Modèle logit

Si l'impact sur le montant du BE paraît contrôlé, la modification de la typologie des flux qui le compose est quant à elle plus importante. Une hausse des flux en début de projection est remarquable.

Cette hausse s'explique par la différence des taux de rachats en QIS 5 (en moyenne 4%) contre les taux de rachats fournis par le modèle logit. Le modèle logit est à la base un modèle de classification ; c'est-à-dire un modèle de prédiction de l'évènement de rachat. Cela se traduit par des espérances conditionnelles souvent proches de 0 ou de 1. Ainsi un model point donné aura tendance à racheter intégralement. Le modèle logit amène finalement une projection bien tranchée par model point alors que la modélisation type QIS 5 correspond à une hypothèse moyenne. Finalement, la PM étant un driver du rachat, plus elle est élevée et plus l'espérance conditionnelle est élevée. Ce point amène à de forts taux de rachats sur les model point à forte PM.

De plus certains points en portefeuille sont vus comme "aberrants" par le modèle logit : des model points à forte ancienneté seront très vite rachetés en quelques années à peine alors que ce rachat sera progressif dans le BE de référence.

Au bout de 8 ans les flux repassent en dessous de ceux du BE QIS 5. Ils y restent ensuite jusqu'à la fin de la projection. Ceci s'explique par l'extinction assez forte du portefeuille en début de projection pour les raisons citées précédemment. Ainsi la projection en run-off amène à un portefeuille beaucoup plus faible. Deux effets expliquent donc l'inversion d'ordre entre les flux des deux BE : une assiette de PM moindre sur le logit et des taux de rachats plus maîtrisés une fois que les points "aberrants" du point de vue logit sont retirés du portefeuille.

Finalement la différence de répartition des flux amène un effet compensation dans le résultat global, mais cette différence de typologie amène une toute autre interprétation avec une vision rachat total par model point et non rachat moyen.

14.2 Modèle hybride

Voici le BE obtenu pour le modèle hybride :

Modélisation	BE
QIS 5	2 579 240 648
Modèle hybride	2 583 161 315

FIGURE 14.3 – BE calculé avec le modèle hybride

L'impact sur le BE est de +4 M €. L'impact relatif sur le BE de référence est de +0, 15%. Cet impact traduit une très faible sous-estimation du risque de rachat.

De même l'étude du moniteur de flux permet de visualiser la modification du BE de manière plus fine. Voici le moniteur obtenu :

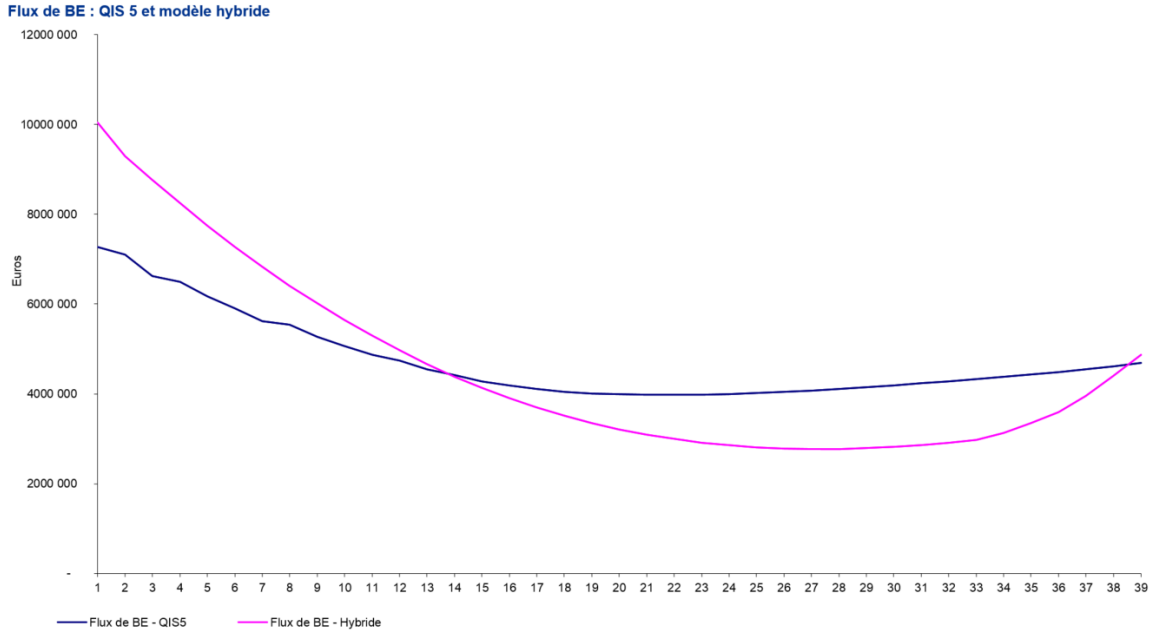

FIGURE 14.4 – Graphique des flux de BE - Modèle hybride

Les tendances du modèle logit se retrouvent dans le modèle hybride. Elles sont cependant atténuées. Les mêmes commentaires sont de nouveaux valables ici, mais la modélisation est plus fine avec la généralisation de la partie linéaire du modèle logit. La prise en compte des facteurs explicatif du rachat est ici plus nuancée et les forts taux de rachats induits par de fortes PM sont réduits dans le modèle hybride.

Cette réduction des écarts amène une inversion des flux à 14 ans. Cette inversion s'explique toujours par la réduction de l'assiette de PM. Cependant les flux du modèle hybride deviennent bien inférieurs à ceux du modèle QIS de 20 à 32 ans. Ceci traduit un effet régime de croisière du portefeuille restant. Les rachats sont donc quasi au point mort sur cette période.

Les flux remontent ensuite brusquement en fin de projection en raison d'une ancienneté et d'un âge moyens de plus en plus importants.

Finalement les flux du modèle hybride suivent la même tendance que ceux du modèle type QIS 5, mais la nuance introduite par l'hypothèse moyenne du modèle de référence a tendance à lisser le tout alors que le modèle hybride amène des taux de rachats très peu nuancés.

14.3 Modèle à états

Voici le BE obtenu pour le modèle à états :

Modélisation	BE
QIS 5	2 579 240 648
Modèle à états	2 590 651 344

FIGURE 14.5 – BE calculé avec le modèle à états

L'impact sur le BE est de $+11,4$ M €. L'impact relatif sur le BE de référence est de $+0,5\%$. Cet impact traduit une sous-estimation du risque de rachat dans la modélisation type QIS 5. Cet écart reste encore contrôlé pour le modèle à états.

De même l'étude du moniteur de flux permet de visualiser la modification du BE de manière plus fine. Voici le moniteur obtenu :

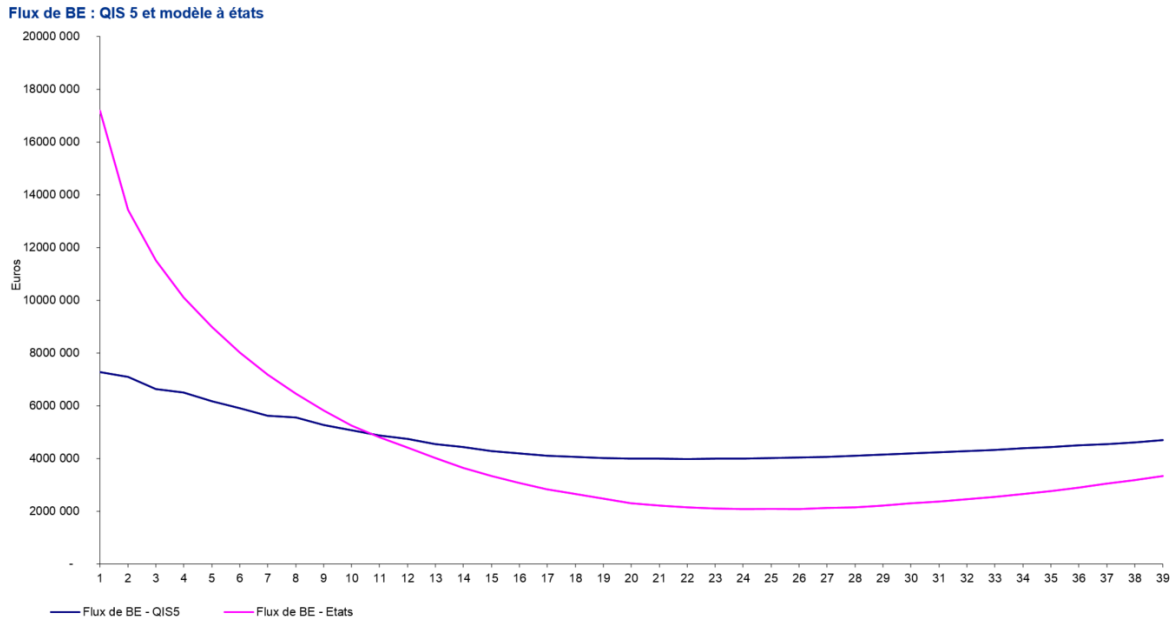

FIGURE 14.6 – Graphique des flux de BE - Modèle à états

Le modèle à états se place plus près du modèle logit que du modèle hybride. Il reste tout de même plus nuancé que le modèle logit. Le modèle à états ayant une philosophie relativement proche de celle du modèle hybride, il est important de comprendre cette différence significative.

Pour rappel, le modèle à états est une combinaison de deux modèles hybrides calibrés sur deux parties distinctes de l'échantillon. Cependant l'utilisation de l'une ou de l'autre loi dépend de la prédiction du SVM sur l'état de crise. La loi des états de crise traduit une agressivité accrue des

assurés face au rachat. L'explication des forts taux de rachats en début de période se trouve dans les scénarios économiques. La courbe de référence pour la génération des scénarios est le courbe EIOPA au 31/12/2015. Les taux sur courtes maturités étant très faibles, la diffusion risque neutre du GSE implique qu'en moyenne les performances action et immobilier sont égales au taux sans risque. Une proportion non négligeable de performances négatives sont donc produites par le GSE ce qui entraîne un passage en état de crise assez régulier. La loi de rachat la plus agressive est donc souvent utilisée et les flux sont mécaniquement plus élevés que ceux du modèle hybride.

14.4 Synthèse

Voici un graphique résumant les flux de BE de chaque modèle :

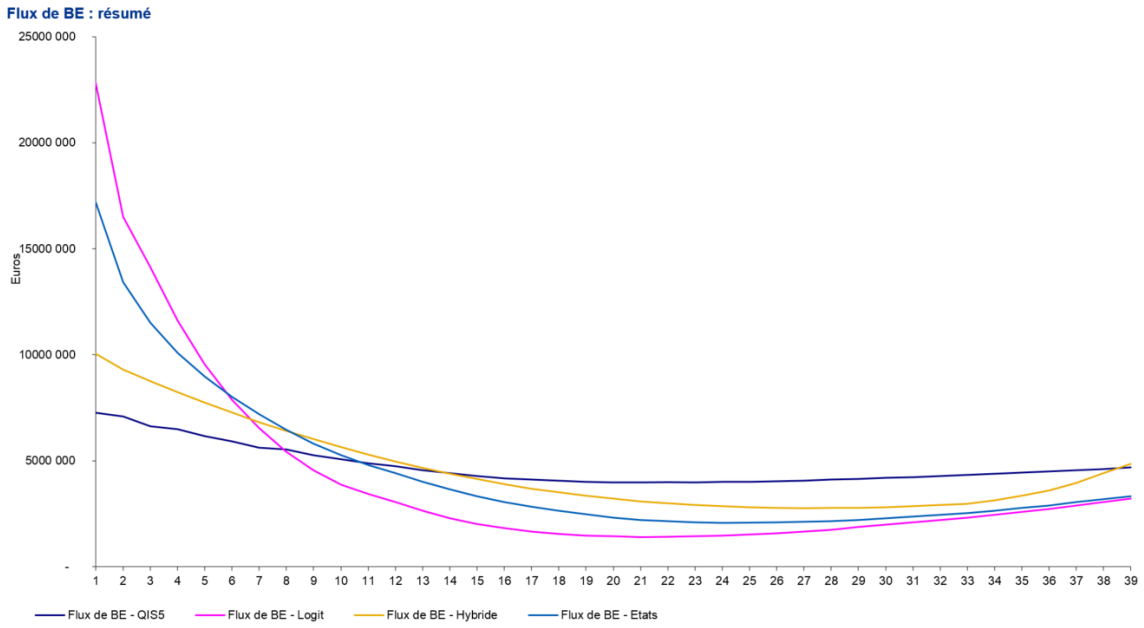

FIGURE 14.7 – Graphique des flux de BE des différents modèles

Finalement, l'application de taux de rachats assez tranchés par model point et non moyennés sur l'ensemble des model points [2] se traduit par une déformation de la typologie des flux de BE. La politique de rachat est finalement plus agressive.

Cependant l'impact sur le BE reste toujours maîtrisé (moins de $0,5\%$ à chaque fois). Si l'impact absolu sur le BE n'est pas insoutenable pour un assureur, la sensibilité du BE au choix de la loi peut poser des problèmes de stabilité du modèle ALM.

Les taux d'erreurs obtenus étant très encourageants, l'étude mérite une application supplémentaire sur diverses bases de données plus fournies.

Un approfondissement par une modélisation séparée de la décision de rachat et du taux de rachat peut être une alternative profitable. Ceci permettrait à la fois une amélioration significative du rachat et une maîtrise de la sensibilité du modèle avec une application moyennée du rachat.

2. Approche de type structurel et QIS 5

Conclusion

Le contexte réglementaire et économique que les assureurs connaissent créé de nouveaux besoins tant en compréhension qu'en modélisation du risque.

Les montants engagés par l'épargne sur fonds euro en France (1375 Mds €) explique à lui seul les enjeux que représente la gestion de ce type de produit. Les prestations afférentes aux produits d'épargne sur fonds euro sont de deux natures : décès et rachats (une troisième nature est distinguée sur les contrats multi-supports : les arbitrages). Si les premiers sont désormais maîtrisés par les assureurs, les seconds demandent encore approfondissements pour respecter les objectifs réglementaires : l'ACPR a demandé aux assureurs de ne plus se contenter de la modélisation ONC lors de la commission Solvabilité II du 10 mars 2016[Institut des Actuaires - ACPR, 2016].

Une étude macro de la décision de rachat des assurés met en exergue le besoin de prendre en compte nombre de facteurs explicatifs dans la modélisation du comportement humain. La décision de l'assuré traduit sa réaction à l'environnement personnel, fiscal et économique dans lequel il évolue.

Les modélisations explorées ces dernières années ne reposent pas toujours sur l'utilisation de variables explicatives (modèle de Vasicek, modélisation comme une option financière, Chain-Ladder, modèle de durée) ou ne sont pas adaptés à une utilisation au sein d'un modèle de projection (variables explicatives non projetées). Ceci mène à la recherche d'un modèle qui améliore la compréhension du rachat et qui soit conforme aux attentes réglementaires.

Ainsi une modélisation de type régression logistique paraît adaptée puisqu'elle concilie une variable réponse binaire (rachat ou non) et la prise en compte de facteurs explicatifs. L'analogie avec le taux de rachat est ainsi facilité puisque la variable réponse qui est l'espérance conditionnelle du rachat peut directement être interprétée comme un taux de rachat.

Ce type de modèle répond à nombre d'objectifs de départ, mais il présente plusieurs limites a priori : il repose sur une dépendance pseudo-linéaire du rachat vis-à-vis des variables explicatives et le calibrage de ce modèle repose sur une hypothèse d'indépendance des données observées, ce qui est rarement le cas. Cependant les coefficients calibrés peuvent être interprétés et améliorent la compréhension du rachat.

Si la deuxième limite est difficile à corriger, la première peut l'être en adaptant le modèle. Un modèle mêlant régression logistique et régression non paramétrique par projection est proposé dans l'étude (qualifié de modèle hybride).

Le modèle hybride permet cette fois de s'affranchir de l'hypothèse de pseudo-linéarité du mo-

dèle logit tout en conservant le socle du modèle. Il présente l'avantage d'être quasi aussi simple à calibrer que le modèle logit.

Là encore le modèle présente des limites a priori. Deux limites techniques apparaissent : le support des variables explicatives et le choix du niveau de troncature. Il présente également des limites pratiques : les variables sont supposées indépendantes et le comportement des assurés est supposé constant dans le temps. Le rachat dépend fortement de l'environnement économique de l'assuré : ce dernier cherche à optimiser son couple rendement/risque et n'hésite pas à arbitrer entre les produits financiers disponibles pour ce faire. La modélisation de ce changement de comportement peut être réalisée par l'intermédiaire d'un modèle à états.

La construction du modèle à états est rendue difficile par le manque d'informations et un modèle probabiliste de type chaîne de Markov n'est pas envisageable. Le but étant de classifier les états de crise et de non crise, un modèle de classification tel que les SVM paraît adapté. Le choix s'est finalement porté sur ces derniers puisque les informations nécessaires à leur mis en place sont disponibles. Le modèle à états est ensuite construit en calibrant deux lois de rachats sur chacun des sous-échantillons obtenus. Le SVM permet de détecter les états de crise dans la projection et ainsi d'utiliser la loi adéquate.

Le modèle à états proposé se rapproche du modèle hybride et permet finalement de prendre en compte le changement de comportement de l'assuré. La composante SVM présente cependant une limite a priori car de nombreux choix de paramètres doivent être opérés.

Une fois les différents modèles construits, la mise en place de ces modèles sur un cas pratique permet de valider leur adéquation à la modélisation du rachat. Les différents modèles ont donc été calibrés sur les données d'un assureur français.

Un choix de variables explicatives a dû être opéré. Cependant, quelque soit le modèle considéré, le cadre de Solvabilité II nécessite de modéliser le rachat au sein d'un modèle de projection de flux. Ainsi les variables considérées pour la mise en application doivent être disponibles lors de la projection. Par exemple, le taux de chômage qui constitue a priori une variable explicative du rachat ne peut être considéré puisqu'il ne peut être projeté correctement dans l'outil DEvent. Les variables retenues sont finalement celles disponibles dans l'outil de projection et qui semblent faire sens quant à l'explication du phénomène de rachat.

La mise en application des modèles s'est faite par Cross-Validation afin d'obtenir des indicateurs plus robustes. Différents taux d'erreurs ont été considérés ainsi que les courbes ROC et les coefficients AUC.

Les modèles ont montré des résultats satisfaisants avec un taux d'erreur trop important mais contenu pour la régression logistique ($7,45\%$) et des taux d'erreurs tout à fait acceptables pour le modèle hybride et le modèle à états (respectivement $3,18\%$ et $2,32\%$). En revanche, le calibrage du SVM était moins satisfaisant avec un taux d'erreur de prédiction de $10,2\%$. Ce taux s'explique principalement par le manque de données économiques avec un historique trop court pour un calibrage robuste.

Les taux d'erreurs n'étant pas des indicateurs assez robustes, les taux de vrais positifs et de vrais négatifs ont été estimés. De manière générale, les taux de vrais positifs observés sont inférieurs

aux taux de vrais négatifs ce qui traduit une difficulté plus importante des modèles à prédire les évènements de rachat que de non rachat. Une amélioration des deux taux est à chaque fois constatée entre le modèle logit, le modèle hybride et le modèle à états.

Un dernier indicateur de qualité de modèle à était utilisé : la courbe ROC. La courbe traduit le pouvoir de généralisation d'un modèle probabiliste de classification. Ce pouvoir peut se résumer en un coefficient : le coefficient AUC qui est l'aire sous la courbe. Les résultats obtenus traduisent un fort pouvoir de classification des modèles avec une espérance conditionnelle assez tranchée et quasi-binaire. Une amélioration du coefficient AUC est à chaque fois constatée comme pour les autres taux d'erreur.

Une fois les modèles calibrés et les taux d'erreurs calculés, il était important de mesurer l'impact du choix de la loi sur l'un des éléments principaux du bilan économique : le BE. Le BE étant le plus sensible au choix de la loi, la mesure d'impact sur ce dernier donne une vision importante de la sensibilité du bilan économique au choix de la loi.

Les lois ont donc été implémentées dans l'outil de projection DEvent. Les différentes projections réalisées ont toutes montré le même impact sur le BE : la modélisation de type QIS 5 tend à sous-estimer le risque de rachat. Cependant l'écart relatif de BE ne dépasse jamais $0,5\%$. Pour vérifier ces éléments, un suivi des flux de BE déflatés moyens a été réalisé.

Ce suivi des flux a montré une modification de la typologie des flux du BE avec des flux plus importants en début de projection et plus faibles en fin de projection relativement à la modélisation initiale de l'assureur. Cette augmentation de la dispersion temporelle des flux cashflows peut avoir des conséquences sur les gestions de taux et de liquidité. Elle constitue un approfondissement supplémentaire possible à la suite de ce mémoire. Ceci s'explique par l'utilisation de modèles de classification et donc l'utilisation de taux de rachat bien plus tranchés que dans la vision rachats structurels et QIS 5. Ainsi le taux de rachat appliqué n'est pas un taux moyenné mais plutôt un taux proche de 0 ou 1. Le calibrage en ligne à ligne et la projection model points par model points est également une limite de l'utilisation des modèles proposés. Ceci limite l'utilisation des lois en l'état pour un assureur.

Finalement, l'étude montre des résultats positifs et encourageants et elle mérite d'être relancée sur diverses bases de données plus fournies. Les résultats obtenus peuvent également conduire à une modélisation séparée de la décision de rachat et du taux de rachat afin de revenir à une hypothèse moyennée comme proposé par les ONC du QIS 5.

La législation évolue également et le projet de loi "Sapin 2" adopté le 8 novembre 2016 en est un exemple d'actualité. Cette loi donne plus d'importance à la modélisation du comportement des assurés en état de crise et en sortie de crise. Les SVM et la détection des états de crise forment une autre piste d'approfondissement de l'étude.

Glossaire

ACPR Autorité de Contrôle Prudentiel et de Résolution. 4
ALM Asset and Liabilities Management. 62
ANR Actif Net Réévalué. 9
AUC Area Under the Curve. 82

BE Best Estimate. 7

CRDS Contribution au Remboursement de la Dette Sociale. 6
CSG Contribution Sociale Généralisée. 6

EHNR Espace de Hilbert à Noyau Reproduisant. 56
EIOPA European Insurance and Occupational Pensions Authority. 7

FP Fonds Propres. 10

GSE Générateur de Scénarios Economiques. 9

IRLS Iteratively Reweighted Least Squares. 32

MAR Missing At Random. 77
MCAR Missing Completely At Random. 77
MLG Modèles Linéaires Généralisés. 28
MNAR Missing Not At Random. 78
MP Model Point. 20

OCDE Organisation de Coopération et de Développement Économiques. 13
ONC Orientations Nationales Complémentaires. 20
ORSA Own Risk and Solvency Assessment. 7

PDD Provision pour Dépréciation Durable. 64
PFL Prélèvement Forfaitaire Libératoire. 6
PM Provision Mathématique. 18
PPB Provision pour Participation aux Bénéfices. 66
PRE Provision pour Risque d'Exigibilité. 65

QIS Quantitative Impact Studies. 20

RM Marge de Risque. 16
ROC Receiver Operating Characteristics. 82

SCR Capital de Solvabilité Requis. 7
SMO Sequential Minimal Optimization. 58
SVM Support Vector Machines. 47

TMG Taux Minimum Garanti. 8

UC Unités de Compte. 5

VaR Value-at-Risk. 10

Bibliographie

[A. Chaudhuri et al., 2004] A. Chaudhuri, C.M. Rocco, and J.A. Moreno (2004). A Support Vector Machine Model for Currency Crisis Discrimination. Springer Berlin Heidelberg.

[A. Dalalyan, 2008] A. Dalalyan (2008). Estimation d'une fonction de régression. Projet Polytechnique.

[A. Suru, 2011] A. Suru (2011). Modélisation du rachat et parallèle avec la Physique. Ecole Centrale de Paris, Université Paris Dauphine, Allianz.

[ACPR, 2013] ACPR (2013). Orientations Nationales Complèmentaires aux Spécifications Techniques. Publié sur le site de l'ACPR.

[Banque De France, 2016] Banque De France (2016). Epargne des Ménages. Publié sur le site de la Banque de France.

[Code Des Assurances, 2016a] Code Des Assurances (2016a). Article L132-21-1. Publié sur le site de Legifrance.

[Code Des Assurances, 2016b] Code Des Assurances (2016b). Article R132-5-3. Publié sur le site de Legifrance.

[Code Des Impôts, 2016] Code Des Impôts (2016). Article 125-0 A. Publié sur le site de Legifrance.

[D. Bosq, 2002] D. Bosq (2002). Estimation localement suroptimale et adaptative de la densité. C. R. Acad. Sci. Paris, Ser. I 334 591-595.

[E. Osuna et al., 1998] E. Osuna, R. Freund, and F. Girosi (1998). Improved Training Algorithm for Support Vector Machines. Technical Report MSR-TR-98-14 Microsoft Research.

[EIOPA, 2014] EIOPA (2014). Technical Specifications. https://eiopa.europa.eu{..}technical-specifications.

[FSA, 2016] FSA (2016). Assurance Vie : Collecte Nette. Publié sur le site de la FSA.

[Hastie T., Tibshirani R., Friedman J., 2009] Hastie T., Tibshirani R., Friedman J. (2009). The Elements of Statistical Learning : Data Mining, Inference, and Prediction.

[Institut des Actuaires - ACPR, 2016] Institut des Actuaires - ACPR (2016). Commission Solvabilité II du 10 mars 2016. Publié sur le site de l'Institut des Actuaires.

[J. Mercer, 1909] J. Mercer (1909). Functions of positive and negative type, and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society of London.

[J.C. Platt, 1997] J.C. Platt (1997). A Fast Algorithm For Training Support Vector Machines. Proc. IEEE NNSP 97.

[M. Charrel, 2016] M. Charrel (2016). La BCE dévoile un nouvel arsenal de mesures pour soigner l'économie européenne. http://www.lemonde.fr/crise-de-l-euro{..}.

[M. Niedzwiedz, 2015] M. Niedzwiedz (2015). Utilisation des supports vecteurs machines pour l'accélèration du calcul du capital économique. ENSAE, Actuaris.

[Nelder and Wedderburn, 1972] Nelder and Wedderburn (1972). Generalized linear models. Journal of the Royal Statistical Society.

[Parlement et Conseil Européens, 2009] Parlement et Conseil Européens (2009). Directive 2009/138/CE. Journal officiel de l'Union Européenne.

[S. Detroulleau and S. Mouret, 2013] S. Detroulleau and S. Mouret (2013). Modèle ALM : Apport de la Logique Floue dans la modélisation des comportements. ENSAE ParisTech, Mazars.

[Tsybakov, 2003] Tsybakov, A. (2003). Introduction à l'estimation non paramétrique.

[V. Vapnik, 1982] V. Vapnik (1982). Estimation of Dependences Based on Empirical Data. Springer-Verlag.

Annexe A

Etude de la vraisemblance du modèle logit

A.1 Calcul de la Hessienne

La log-vraisemblance du modèle logit étant deux fois dérivable, il convient de calculer la Hessienne afin d'étudier se convexité :

$$\begin{aligned}
\forall j \in [\![0;n]\!]\ \forall k \in [\![0;n]\!] \quad \frac{\partial^2 \mathcal{L}_N(\beta)}{\partial\beta_j \partial\beta_k} &= -\frac{1}{N}\sum_{i=1}^{N} \frac{\partial x_{i;j}(y_i - p_\beta(x_i))}{\partial\beta_k} \\
&= \frac{1}{N}\sum_{i=1}^{N} x_{i;j}\frac{\partial p_\beta(x_i)}{\partial\beta_k} \\
&= \frac{1}{N}\sum_{i=1}^{N} x_{i;j}\frac{\partial (1+e^{x'_i\beta})^{-1}}{\partial\beta_k} \\
&= \frac{1}{N}\sum_{i=1}^{N} x_{i;j}x_{i;k}e^{x'_i\beta}(1+e^{x'_i\beta})^{-2}
\end{aligned}$$

Le passage en écriture matricielle donne donc la Hessienne :

$$\begin{aligned}
\nabla^2\mathcal{L}_N(\beta) &= X'\Omega_\beta X \\
\text{où } \Omega_\beta &= \text{diag}(\{p_\beta(x_i)(1-p_\beta(x_i)); i \in [\![1;N]\!]\})
\end{aligned}$$

A.2 Démonstration du caractère convexe

Maintenant que la Hessienne est connue, il reste à démontrer qu'elle est semi-définie positive pour conclure sur son caractère convexe.

Tout d'abord, étant donné que la matrice Ω est diagonale dont les éléments diagonaux sont tous strictement positifs. Il est donc possible de réécrire la Hessienne comme :

$$\begin{aligned}
\nabla^2\mathcal{L}_N(\beta) &= X'\Omega_\beta X \\
&= X'(\Omega_\beta^{\frac{1}{2}})'\Omega_\beta^{\frac{1}{2}}X \\
&= (\Omega_\beta^{\frac{1}{2}}X)'\Omega_\beta^{\frac{1}{2}}X
\end{aligned}$$

Il vient donc finalement :

$$\begin{aligned} \forall u \in \mathbb{R}^n \quad \left\langle \nabla^2 \mathcal{L}_N(\beta) u \middle| u \right\rangle &= u'(\Omega_\beta^{\frac{1}{2}} X)' \Omega_\beta^{\frac{1}{2}} X u \\ &= ||\Omega_\beta^{\frac{1}{2}} X u||^2 \geq 0 \end{aligned}$$

La Hessienne est donc semi-définie positive et la log-vraisemblance du modèle logit est donc convexe. Ceci conclut sur l'existence du maximum de vraisemblance.

Annexe B

Démonstrations relatives au modèle non-paramétrique

B.1 Démonstration du caractère projecteur orthogonal

Il est bon de rappeler qu'une matrice est un projecteur orthogonal si et seulement si elle est symétrique et idempotente.

$$\begin{aligned} A_M^2 &= \phi_M(\phi'_M\phi_M)^{-1}\phi'_M.\phi_M(\phi'_M\phi_M)^{-1}\phi'_M \\ &= \phi_M(\phi'_M\phi_M)^{-1}(\phi'_M\phi_M)(\phi'_M\phi_M)^{-1}\phi'_M \\ &= \phi_M(\phi'_M\phi_M)^{-1}\phi'_M \\ &= A_M \end{aligned}$$

La matrice A_M est idempotente : c'est donc un projecteur.

$$\begin{aligned} A'_M &= (\phi_M(\phi'_M\phi_M)^{-1}\phi'_M)' \\ &= (\phi'_M)'(\phi'_M\phi_M)^{-1'}\phi'_M \\ &= \phi_M((\phi'_M\phi_M)')^{-1}\phi'_M \\ &= \phi_M(\phi'_M\phi_M)^{-1}\phi'_M \\ &= A_M \end{aligned}$$

Elle est de plus symétrique, c'est donc un projecteur orthogonal.

B.2 Majoration de la variance de l'estimateur des moments

Soit $j \in [\![1; M]\!]$. Il est alors possible de remplacer $\widetilde{\theta}_j$ par son expression :

$$\begin{aligned} V[\widetilde{\theta}_j] &= \mathbb{V}\left[\frac{1}{N}\sum_{k=1}^{N} Y_k \varphi_j(X_k)\right] \\ &= \mathbb{V}\left[\frac{1}{N}\sum_{k=1}^{N} (g(X_k)+\xi_k).\varphi_j(X_k)\right] \\ &= \frac{1}{N^2}.\mathbb{V}\left[\sum_{k=1}^{N} (g(X_k)+\xi_k).\varphi_j(X_k)\right] \end{aligned}$$

De plus les X_k et ξ_k sont supposés indépendants entre eux. Il vient donc :

$$\begin{aligned} V[\widetilde{\theta}_j] &= \frac{1}{N^2}.\sum_{k=1}^{N} \mathbb{V}[(g(X_k)+\xi_k).\varphi_j(X_k)] \\ &= \frac{1}{N}.\mathbb{V}[(g(X)+\xi).\varphi_j(X)] \\ &= \frac{1}{N}.\mathbb{V}[(g(X)+\xi).\varphi_j(X)] \\ &= \frac{1}{N}.(\mathbb{V}[g(X).\varphi_j(X)] + \mathbb{V}[\xi.\varphi_j(X)] + 2.\mathrm{Cov}[g(X).\varphi_j(X); \xi.\varphi_j(X)]) \end{aligned}$$

L'indépendance donne finalement :

$$\begin{aligned} V[\widetilde{\theta}_j] &= \frac{1}{N}.(\mathbb{V}[f(X).\varphi_j(X)] + \mathbb{V}[\xi.\varphi_j(X)] + 2(\mathbb{E}[f(X).\varphi_j(X).\xi.\varphi_j(X)] \\ &\quad -\mathbb{E}[f(X).\varphi_j(X)].\mathbb{E}[\xi.\varphi_j(X)])) \\ &= \frac{1}{N}.(\mathbb{V}[f(X).\varphi_j(X)] + \mathbb{V}[\xi.\varphi_j(X)] + 2(\mathbb{E}[\xi].\mathbb{E}[f(X).\varphi_j(X).\varphi_j(X)] \\ &\quad -\mathbb{E}[f(X).\varphi_j(X)].\mathbb{E}[\xi].\mathbb{E}[\varphi_j(X)])) \\ &= \frac{1}{N}.(\mathbb{V}[f(X).\varphi_j(X)] + \mathbb{V}[\xi.\varphi_j(X)]) \\ &= \frac{1}{N}.(\mathbb{E}[(f(X).\varphi_j(X))^2] - \mathbb{E}[f(X).\varphi_j(X)]^2 + \mathbb{E}[(\xi.\varphi_j(X))^2] - \mathbb{E}[\xi.\varphi_j(X)]^2) \\ &= \frac{1}{N}.(\mathbb{E}[(f(X).\varphi_j(X))^2] - \mathbb{E}[f(X).\varphi_j(X)]^2 + \mathbb{E}[(\xi.\varphi_j(X))^2] - \mathbb{E}[\xi]^2.\mathbb{E}[\varphi_j(X)]^2) \\ &= \frac{1}{N}.(\mathbb{E}[(f(X).\varphi_j(X))^2] - \mathbb{E}[f(X).\varphi_j(X)]^2 + \mathbb{V}[\xi].\mathbb{E}[\varphi_j(X)^2]) \\ &= \frac{1}{N}.(\mathbb{E}[(f(X).\varphi_j(X))^2] - \mathbb{E}[f(X).\varphi_j(X)]^2 + \mathbb{V}[\xi]) \end{aligned}$$

Comme le terme $\mathbb{E}[f(X).\varphi_j(X)]^2$ est toujours positif, il vient l'inégalité suivante :

$$\begin{aligned}
V[\widetilde{\theta}_j] &\leqslant \frac{1}{N}.(\mathbb{E}[(f(X).\varphi_j(X))^2] + \mathbb{V}[\xi]) \\
&\leqslant \frac{1}{N}.\left(\int_{[0;1]^n} (f(x).\varphi_j(x))^2 \mathrm{d}x + \mathbb{V}[\xi]\right) \\
&\leqslant \frac{1}{N}.\left(\int_{[0;1]^n} (\|f\|_\infty.\varphi_j(x))^2 \mathrm{d}x + \mathbb{V}[\xi]\right) \\
&\leqslant \frac{1}{N}.\left(\|f\|_\infty^2.\int_{[0;1]^n} \varphi_j(x)^2 \mathrm{d}x + \sigma^2\right) \\
&\leqslant \frac{1}{N}.(\|f\|_\infty^2 + \sigma^2)
\end{aligned}$$

B.3 Calcul du biais de l'estimateur de la variance

$(f(X_k); k \in [\![1; N]\!])'$ appartient à l'espace vectoriel engendré par les colonnes de ϕ_{M_0} par hypothèse. Cela revient à dire que le modèle approché est considéré. Et cette hypothèse amène que :

$$(I_N - A_{M_0})Y = \xi = (I_N - A_{M_0})\xi$$

De plus A_{M_0} est un projecteur orthogonal, donc $I_N - A_{M_0}$ l'est également sur l'orthogonal de l'espace vectoriel engendré par les colonnes de ϕ_{M_0}. Il est donc idempotent et symétrique et il vient alors :

$$\begin{aligned}
\|(I_N - A_{M_0})\xi\|^2 &= ((I_N - A_{M_0})\xi)'.(I_N - A_{M_0})\xi \\
&= \xi'.(I_N - A_{M_0})'.(I_N - A_{M_0})\xi \\
&= \xi'.(I_N - A_{M_0}).(I_N - A_{M_0})\xi \\
&= \xi'.(I_N - A_{M_0})^2.\xi \\
&= \xi'.(I_N - A_{M_0}).\xi
\end{aligned}$$

Il suffit ensuite d'appliquer l'espérance et d'utiliser les propriétés de la trace :

$$\begin{aligned}
\mathbb{E}[\xi'(I_N - A_{M_0}).\xi] &= \mathrm{Tr}(\mathbb{E}[\xi'(I_N - A_{M_0}).\xi]) \\
&= \mathbb{E}[\mathrm{Tr}(\xi'.(I_N - A_{M_0}).\xi)] \\
&= \mathbb{E}[\mathrm{Tr}((I_N - A_{M_0}).\xi'.\xi)] \\
&= \mathrm{Tr}(\mathbb{E}[(I_N - A_{M_0}).\xi'.\xi]) \\
&= \mathrm{Tr}((I_N - A_{M_0}).\mathbb{E}[\xi'.\xi]) \\
&= \mathrm{Tr}((I_N - A_{M_0}).\sigma^2) \\
&= \mathrm{Tr}(I_N - A_{M_0}).\sigma^2
\end{aligned}$$

Finalement il reste à calculer la trace de la matrice $I_N - A_{M_0}$:

$$\begin{aligned}
\mathrm{Tr}(I_N - A_{M_0}) &= \mathrm{Tr}(I_N) - \mathrm{Tr}(A_{M_0}) \\
&= N - \mathrm{Tr}(\phi_{M_0}(\phi'_{M_0}\phi_{M_0})^{-1}\phi'_{M_0}) \\
&= N - \mathrm{Tr}((\phi'_{M_0}\phi_{M_0})(\phi'_{M_0}\phi_{M_0})^{-1}) \\
&= N - \mathrm{Tr}(I_{M_0}) \\
&= N - M_0
\end{aligned}$$

Ceci permet de conclure sur le biais de l'estimateur proposé $\hat{\sigma}^2_{M_0}$.