ROYAUME DU MAROC
--*-*-*
HAUT COMMISSARIAT AU PLAN
--*-*-*-*-*-*
INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE
INSEA

Projet de Fin d'Etudes

*****

Modélisation de la consommation médicale et évaluation de l'engagement médical de l'OCP sous la norme IFRS/IAS19

Préparé par : *M^elle ELKARIM Imane (AF)*
M^elle SGHIOURI Soukaina (SE)

Sous la direction de : ***M. MARRI Fouad*** **(INSEA)**
M. ABERKANE Massin **(OCP)**

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Filières : Statistique-économie appliquée
Actuariat et Finance

Dédicace

Je dédie ce travail à tous les chauffeurs de taxis qui m'ont dit « Non » au trajet de l'OCP. Grâce à eux, j'ai internalisé la vertu de la persévérance.

Je le dédie aussi à mes très chers parents, ma petite sœur et mes amis, qui sans eux, je ne serais pas ce que je suis aujourd'hui.

IMANE

Au meilleur des pères, et à ma très chère maman. Qu'ils trouvent en moi la source de leur fierté à qui je dois tout.

A mes frères : Khalid, Zakaria et Yahya, à qui je souhaite un avenir radieux plein de réussite.

A mes Amis source intarissable de support et d'inspiration.

A tous ceux qui me sont chers.

SOUKAINA

Remerciements

Avant tout, merci Allah !

Au terme de notre stage, nous avons bénéficié de l'appui de plusieurs personnes auxquelles nous voulons exprimer ici nos sincères remerciements.

Nous tenons à remercier notre encadrant, M. Fouad Marri, pour son soutien et sa disponibilité ainsi que pour ses précieuses directives.

Nous remercions très sincèrement M. Younes LHALOUANI et M. Massin ABERKANE qui nous ont encadrés et assurés les conditions nécessaires afin que nous puissions effectuer ce stage dans les meilleures conditions possibles.

Nous tenons à exprimer notre gratitude envers les belles dames Mme Nawal EZZIRARI, Mme Hasna MOURID et Mme ERRIANI Meriem, ainsi que tous les autres membres de l'équipe Prévoyance, pour leur amabilité et leur convivialité, dignes des plus grands et plus sincères remerciements.

Nous remercions tous ceux qui ont contribué à l'aboutissement de ce travail de près et de loin.

Avertissement

Pour des raisons de confidentialité, les valeurs numériques présentes dans ce mémoire ont été modifiées. Les ordres de grandeur ont néanmoins été conservés.

Résumé

Un régime collectif de frais de santé mis en place par l'OCP afin d'accroître la couverture médicale de ses salariés, a un rôle majeur dans le système de la protection médicale. Dans le cadre de la norme IAS 19, lorsqu'une entreprise participe au financement de ce type de régime, celle-ci doit évaluer et provisionner des engagements au titre du financement à long terme qu'elle apporte au régime de ses salariés retraités.

C'est dans ce contexte que rentre notre projet, consistant à étudier les engagements médicaux du Groupe OCP au titre du financement des avantages postérieurs à l'emploi. Plusieurs paramètres étaient à définir ou modéliser : le taux d'actualisation, le taux d'inflation médicale et les flux financiers utilisés dans la projection des prestations médicaux.

Dans un premier temps, nous avons pris connaissance de la norme IAS19 et des méthodes d'évaluation des engagements adaptées. Dans un second temps il était question de retrouver tous les paramètres cités. Pour cette fin, nous avions eu besoin d'une courbe de taux zéro coupon pour définir les taux d'actualisation. Ce dernier constitue un élément central de l'évaluation des engagements d'un régime de frais de santé (et plus généralement d'un régime à prestations définies). Ce paramètre permet de connaître la valeur économique actuelle des flux financiers futurs du régime et est, à ce titre, spécifiquement encadré par la norme IAS 19. Une prochaine étape consiste à utiliser les données HCP de la section santé pour définir un taux d'inflation médical global aussi utilisé durant la projection.

L'étude de la consommation médicale de l'OCP nous servira par la suite à projeter l'évolution future du montant des provisions Cette projection sera faite tête par tête comme le stipule la norme IAS19. Afin d'estimer ces montants-là, nous devions donc retrouver un « tarif technique » en utilisant le modèle le plus répandu en tarification qui est le modèle fréquence coût. Cette modélisation usera du GLM, puis avec la technique de Machine Learning « forêts aléatoires » qui se dit plus performante.

Une fois les dépenses estimées, et les hypothèses actuarielles fixées pour notre date d'évaluation des engagements, nous avons créé un outil de calcul sur Excel et VBA Excel afin de calculer l'engagement médical de l'OCP. Etant donné que nous ne disposons que de la liste de la population OCP 2017, la date d'évaluation de l'engagement sera naturellement de 2017.

Mots clés : IAS19, engagement social, modélisation, hypothèses actuarielles, consommation médicale, tarification, Machine Learning.

Abstract

A collective health insurance plan put in place by OCP to increase the medical coverage of its employees has a major role in the system of medical protection. Under IAS 19, when an enterprise participates in the financing of this type of plan, it must evaluate and provision long-term commitments meant for its retired employees.

Our work is about studying OCP Group's medical commitments that post-employment benefits encompass. Several parameters were to be defined or modeled: the discount rate, the medical inflation rate and the financial flows used in the projection of medical benefits.

At first, we took note of the IAS19 standard and methods for evaluating the appropriate commitments. Then it was about finding all the parameters mentioned. For this purpose, we needed a zero coupon yield curve to set the discount rates. The latter constitutes a central element in the evaluation of the commitments of a health expense plan (and more generally of a defined benefit plan). This parameter makes it possible to know the current economic value of the future financial flows of the plan and is, in this respect, specifically governed by IAS 19. A next step is to use the HCP data of the health section to define a medical inflation rate, also used during the projection.

The medical consumption will then be used to project the future evolution of the provisions. This projection will be done head-to-head as stipulated in IAS19. In order to estimate these amounts, we had to find a "technical tariff" using the most popular pricing model which is the loss cost model. This modeling will use the GLM, then with the machine learning technique "random forests" which is said to be more efficient.

Once the expenses are estimated, and the actuarial assumptions are fixed for our commitment valuation date, we have created an Excel and VBA Excel calculation tool to calculate OCP's medical commitment. Since we only have the 2017 OCP population list, the commitment evaluation date will naturally be 2017.

Key words : IAS19, social commitment, modeling, Actuarial assumptions, medical consumption, princing, Machine Learning.

Liste des abréviations :

AIC : Akaike Information Criterion

AICC : Akaike Information Criterion Corrected (for small sample size)

ALD : Affection de Longue Durée

ALC : Affection Lourde et Couteuse

AMB : Assurance Maladie de Base

AMC : Assurance Maladie Complémentaire

AMMC : Autorité Marocaine du Marché des Capitaux

AMO : Assurance Maladie Obligatoitre

ANOVA : ANalysis Of Variance

BAM : Bank-al-Maghrib

BIC : Bayesian Information Criterion

BIS : Bank for International Settlments

CART : Classification And Regression Trees

CIR : Caisse Inter-générationnelle de Retraite

CNOPS : Caisse Nationale des Organismes de Prévoyance Sociale

CNSS : Caisse Nationale de Sécurité Sociale

DBO : Defined Benefits Obligation

FASB : Financial Accounting Standards Board

GLM : General Linear Model

HCP : Haut Commissariat au Plan

IAS : International Accounting Standards

IASB : International Accounting Standards Borad

IASC : International Accounting Standards Committee

IFRS : International Financial Reporting Standards

IFRSIC : International Financial Reporting Standards Interpretations Committe

IPC : Indice de Prix à la Consommation

FMI : Fonds Monétaire International

MCO : Moindre Carrées Ordinaire

MSE : Mean Squarred Error

OCP : Office Chérifien des Phosphates

OOB : Out Of Bag

PUCM : Projected Unit Credit Method

VAN : Valeur Actuelle Nette

VAPF : Valeur Actuelle Probable Future

VP : Valeur Probable

ZC : Zero Coupon

Liste des figures

Figure 1 : Période de jouissement des avantages court terme et postérieurs à l'emploi 24

Figure 2 : Historique des taux (2010-2017) de référence marocain TMP. Source : BAM 33

Figure 3 : Courbe des taux de rendement de l'Etat marocain pour l'année 2016 . Source :BAM , Réalisée par nos soins. 34

Figure 4 : Différentes catégories de courbe de taux 35

Figure 5 : Sortie graphique du lissage de la courbe de taux en utilisant l'interpolation linéaire et cubique avec pour la date du 1/03/2018 50

Figure 6 : Sortie graphique du lissage de la courbe de taux avec les méthodes indirectes pour la date du 01/03/2018 50

Figure 7 : Impact des facteurs $\boldsymbol{\beta 1}, \boldsymbol{\beta 2}, \boldsymbol{\beta 3}$ sur la formation de la courbe des taux pour la date du 01/03/2018 51

Figure 8 : Evolution de la courbe des taux marocaine du Janvier 2004 au Juillet 2014. Source : FMI 52

Figure 10 : Echantillon de données du marché obligataire collecté par nos soins. 55

Figure 11 : Courbe de taux augmentée de la prime de risque calculée à base du modèle de régression. 56

Figure 12 : Comparaison de la courbe du taux sans risque et la courbe de taux augmentée de la prime de risque calculée. 58

Figure 13 : Evolution de l'IPC annuel par sections pour la division santé (Base 100 : 2006). 60

Source : HCP 60

Figure14 : Evolution des prestations des organismes d'assurance maladie (remboursements et tiers payant), 1997/98-2013. Source : Comptes nationaux de santé 2013 60

figure 15: Réseau neuronal artificiel 67

Figure16 : Schéma du Bagging avec pour règle de base un arbre CART 68

Figure 17: La part de la consommation totale par type de bénéficiaire 72

Figure 18: La part du nombre de dossiers par type de bénéficiaire 73

Figure19 : Répartition de la population selon le sexe du pensionné 74

Figure20: Répartition du nombre de dossiers par sexe du pensionné 74

Figure21 : La part du montant remboursé total par sexe du pensionné 75

Figure22 : Répartition de la population des pensionnés selon la situation familiale 76

Figure 23: Répartition des dossiers par Situation familiale 76

Figure24 : Découpage de la charge des sinistres selon situation familiale 77

Figure 25: Répartition des dossiers médicaux par intervalle de pension 79

Figure 26: Répartition du remboursement total par intervalle de pension 79

Figure 27: Le nombre de dossiers médicaux par âge du pensionné 80

Figure 28: La courbe de consommation médicale par âge du pensionné 81

Figure 29: Remboursement moyen par dossier (à gauche) et par personne (à droite) des ALD et des Non ALD 83

Figure30: fréquence moyenne par an, selon la présence ou pas d'ALD 83

Figure 31: Comparaison des consommations médicales moyennes selon la présence ou non de l'ALD_ALC 84

Figure 32: Le remboursement moyen par âge du pensionné selon la présence d'ALD_ALC, toutes les années confondues. 85

Figure 33: Le remboursement moyen par âge du pensionné selon la présence d'ALD_ALC, après nettoyage de la base. 86

Figure 34: Comparaison des courbes de consommation par âge des dossiers ALD et Non ALD 88

Figure 35: Test d'ajustement graphique de la loi de poisson 91

Figure 36: Test d'ajustement graphique de la loi Négative binomiale 93

Figure 37:Ajustement courbe de la loi Normale sur l'histogramme du cout moyen. 101

Figure 38: L'ajustement graphique par la loi log-normale, Gamma et Weibull 102

Figure 39: Quantiles de la distribution Gamma des coûts moyens 103

Figure 40: Table de score issue des résultats GLM 107

Liste des tableaux

Tableau1 : liste des pays et leur méthode de modélisation de leur courbe des taux. Source : BIS (Bank For International Settlments) 47

Tableau 2 : Tableau de calcul des taux de référence des bons de trésor du 01/03/2018.Source : BAM 48

Tableau 3 : Calcul du taux de référence TMP pour les maturités pleines 3mois-30ans pour la date du 01/03/2018 49

Tableau 4 : Tests de validation du modèle de régression du Spread par le TSR et la Maturité56

Tableau 5 : La volatilité de la part de la prime de risque dans le taux de rendement, calculée par maturité 57

Tableau 6: Tableau de calcul de l'inflation médicale 61

Tableau6: Répartition de l'effectif et des dossiers par sexe du pensionné 74

Tableau7: Répartition des montants remboursés par sexe du pensionné 75

Tableau8: La répartition de l'effectif et des dossiers par situation familiale du pensionné 75

Tableau9 : Répartition des montants remboursés par situation familiale du pensionné 77

Tableau 10: Répartition de l'effectif et des dossiers par intervalle de pension 78

Tableau 11: L'évolution de la moyenne des remboursements par présence d'ALD_ALC au titre des années 2015, 2016 et 2017. 82

Tableau12: Localisation des valeurs extrêmes 85

Tableau13: Les résultats du test ANOVA à deux facteurs 87

Tableau 14: Quantiles de la distribution historique des coûts moyens 89

Tableau 15: Quantiles de la distribution historique de la fréquence des sinistres 90

Tableau 16: La segmentation de la variable "Age" 95

Tableau 17: La segmentation de la variable Pension 95

Tableau 18: Récapitulatif des informations sur le modèle utilisé 96

Tableau 19: Classes de variables GLM (fréquence) 96

Tableau 20: Résultat de validation du modèle GLM pour la fréquenc 97

Tableau 21: Sortie SAS du résultat GLM pour la fréquence 97

Tableau 22: Classes de variables retenus (pour le coût moyen) 103

Tableau 23: Résultat de validation du modèle GLM pour le coût moyen 104

Tableau 24: Sortie SAS du résultat GLM pour le coût moyen 105

Tableau 25: Récapitulatif des hypothèses démographiques..... 115

Table des matières

Liste des abréviations : 8

Liste des figures 10

Liste des tableaux 12

Introduction générale 17

Partie 1 : Contexte général 18

Chapitre 1 : Contexte économique et social : La protection médicale 18

1.1 La couverture médicale au Maroc 18

1.1.1 L'assurance Maladie de la sécurité sociale : l'AMO 18

1.1.2 Les organismes d'assurance santé complémentaire 19

1.2 Couverture médicale : Cas de l'OCP 19

1.2.1 Basculement vers la CNOPS 19

1.2.2 Le taux de cotisation 20

1.2.3 Nécessité d'une Assurance complémentaire 20

Chapitre 2 : Contexte financier : Evaluation des engagements sociaux en norme IAS 19. 22

2.1 Les Normes IAS/IFRS 22

2.1.1 Le Timeline 22

2.1.2 Structure organisationnelle de l'IASB 22

2.1.3 Objectifs de l'IASB 23

2.2 Le champ d'application de la norme IAS 19 23

2.3 Les objectifs de la norme IAS 19 24

2.4 Les avantages postérieurs à l'emploi 24

2.4.1 Les régimes à cotisations définies 25

2.4.2 Les régimes à prestations définies 25

2.4.3 Les hypothèses actuarielles 25

2.5 L'évaluation des engagements 27

2.5.1 La formulation générale 27

2.5.2 LA méthode d'évaluation et comptabilisation des engagements 28

Partie 2 : Modélisation des hypothèses actuarielles : taux d'actualisation et inflation médicale 30

CHAPITRE 1 Taux d'actualisation : avant-propos 30
1.1Concepts fondamentaux 31
1.1.1Valeur Actuelle Nette 31
1.2 Les Taux de base 32
1.3Taux spot de référence au Maroc 33
1.4 Structure par terme des taux 34
1.5 Taux à terme (*forward*) 37
CHAPITRE 2 Lissage de la courbe des taux de rendements 38
2.1 Quelques notions importantes 38
2.1.1 Les obligations 38
2.1.2 Les coupons 39
2.1.3 Les catégories d'obligations 39
2.1.4 Evaluation d'un produit obligataire : Le « Zéro Coupon » 39
2.1.5 Evaluation d'une obligation à coupon 40
2.1.6 La prime de risque sur un placement à revenu fixe 40
2.2 Classification des courbes des taux 42
2.2.1 Les courbes de marché ou directes 42
2.2.2 Les courbes implicites ou indirectes 42
2.3 Méthodes de construction de la courbe des taux 42
2.3.1 La Méthode directe : Le Bootstrapping 42
2.3.2Les Méthodes indirectes 44
2.4Elaboration d'un outil métrique sur EXCEL & VBA Excel 47
2.4.1 La courbe de taux de référence pour une date donnée : Exemple du 01/03/2018 47
2.4.2 La courbe des taux historique 51
2.4.3 La courbe de taux forward 52
2.4.4 La courbe des taux augmentée de la prime de risque sur obligations. 54
Chapitre 3 : Le taux d'inflation médicale 59
3.1 La consommation de soins et de biens médicaux 59
3.2 Détermination du taux d'inflation médical global 60
Partie 3 : Modélisation de la consommation médicale & calcul de l'engagement médical de l'OCP 62
Chapitre 1 : Modélisation de la consommation médicale 62
1.1La méthode de tarification santé 62
1.1.1 Le tarif technique : dépense médicale estimée 62

1.1.2 Modélisation GLM (General Linear Model) 63
1.1.3Une alternative au GLM : l'apprentissage statistique 66
1.1.4 Cadre théorique : Les forêts aléatoires (Randomforests) 67
1.2Mise en œuvre de la tarification 70
1.2.1 La base de données médicale de l'OCP 70
1.2.2 Analyse univariée des variables explicatives 71
1.2.3 Ecrêtement des grands coûts et fréquences 88
1.2.4 Modélisation de la fréquence de sinistralité 90
1.2.5 Modélisation du coût moyen 99
1.3 Exploitation du GLM : Table de score 106
1.4 l'apprentissage statistique pour la tarification 107
1.4.1 Les forêts aléatoires (Random Forest-RI) : implémentation 108
1.4.2 Modélisation de la fréquence 109
1.4.3 Modélisation du coût moyen 110
1.5 GLM et Machine Learning : Comparaison des performances 112
1.5.1 Analyse Graphique 112
1.5.2 Calcul des MSE de chaque modèle de tarification 113
Chapitre 2 : Calcul de l'engagement médical de l'OCP 115
2.1. Les hypothèses de calcul 115
2.1.1 Les hypothèses démographiques 115
2.1.2 Les hypothèses financières 115
2.2 Calcul de l'engagement 117
2.2.1 La formulation 117
2.2.2 Résultats 118
Conclusion générale 119
Bibliographie 120
ANNEXES 121

Introduction générale

Le groupe OCP a opté pour la norme IFRS depuis 2009 avec la date du 1er Janvier 2008 comme point de transition. Le groupe représente un cas d'école en matière de reporting financier. Le phosphatier marocain compte 37 filiales spécialisés dans l'extraction, l'exploitation, le service, le trading, etc. Le périmètre de consolidation concerne une trentaine de sociétés dont une vingtaine à travers le monde et une dizaine en Afrique, rattachées à OCP Africa.

C'était donc dans l'intérêt de L'OCP d'évoluer vers le référentiel international du fait de la diversification de ses partenaires et de son développement à l'international. Ainsi, le groupe produit ses comptes selon la norme IFRS sur une base trimestrielle et selon le référentiel national avec une périodicité semestrielle. A partir de 2016, l'entreprise décide de publier uniquement des comptes IFRS chaque trimestre et de publier dans la presse nationale le package complet avec une $fréquence$ semestrielle. Ceci garantira à l'OCP une information financière de qualité à l'instar de ses concurrents internationaux.

La norme IAS 19 étant la norme du référentiel IFRS qui s'applique à la comptabilisation de tous les avantages du personnel, nous étions amenées, dans le cadre de notre projet de fin d'études, à calculer l'engagement social lié au frais de santé de l'OCP. Cet engagement constitue une dette sociale que le groupe OCP S.A. doit évaluer selon cette norme internationale.

Afin de pouvoir mener cette évaluation, nous étions appelées à nous familiariser avec la norme IAS19 et d'appréhender les méthodes de calcul actuariels des engagements qui nous intéressent et imposés par celle-ci. Ceci fera donc l'objet de la première partie de ce mémoire.

A travers nos lecture des normes, plusieurs hypothèses actuarielles dont nous avons besoin pour l'évaluation, nécessitent d'être déterminées soit au préalable (cas de l'inflation médicale), soit au fur et à mesure des calculs (le cas du taux d'actualisation). La $2^{ème}$ partie de ce mémoire traitera de la modélisation de deux hypothèses importantes : La première est le taux d'actualisation : Stipulé par la norme IAS19, Le taux d'actualisation sera défini à partir de la courbe des taux ZC (courbe de taux sans risque), dont nous exhiberons la théorie dans cette partie. La deuxième hypothèse que nous chercherons à définir est le taux d'inflation médicale.

Le montant de la prestation médicale à provisionner et qui représentent le coût du risque auquel l'OCP devrait faire face dans les cadre des avantages postérieurs à l'emploi est aussi un paramètre que nous devions définir et modéliser. C'est d'ailleurs la partie la plus cruciale du calcul de l'engagement. Dans la dernière partie de ce mémoire, nous exhiberons les techniques de modélisation statistiques de la consommation médicale, sous forme de tarif technique d'assurance frais de santé, avant de pouvoir exploiter les résultats obtenus dans le calcul de l'engagement médical de l'OCP sous la norme IAS19.

Partie 1 : Contexte général

Dans cette 1ère partie, nous introduisons le contexte général de notre projet de fin d'études. Il s'inscrit d'une part dans un registre social puisqu'il concerne la protection médicale offerte par l'entreprise à ses salariés, et un registre technique puisque nous serons amenées à évaluer l'engagement ou le coût de cette protection médicale sous une norme financière qui est l'IAS19.

Chapitre 1 : Contexte économique et social : La protection médicale

1.1 La couverture médicale au Maroc

Le système de la protection médicale marocain est basé sur deux types d'acteurs :

- **L'Assurance Maladie de la Sécurité Sociale** : elle offre aux assurés sociaux une couverture médicale obligatoire de base et donc une couverture médicale minimale. C'est bien le cas de l'AMO (Assurance Maladie Obligatoire) promulguée en Novembre 2002.

- **Les organismes d'assurance santé complémentaire** : ils permettent aux assurés sociaux de compléter leur couverture médicale de base, en intervenant sur la part des coûts laissée par l'Assurance Maladie à la charge de l'assuré, garantissant ainsi aux assurés une protection renforcée.

1.1.1 L'assurance Maladie de la sécurité sociale : l'AMO

L'Assurance Maladie Obligatoire (AMO) est une prestation sociale marocaine garantissant l'accès universel au soin et le remboursement des prestations, des soins ainsi que des médicaments. Instituée par la loi 65.00

Le principal objectif de l'AMO est, d'une part, d'assurer une certaine égalité quant à l'accès aux soins et ce, à tout individu résidant au Maroc et, d'autre part, de mettre en place une prise en charge collective et solidaire des dépenses relatives à la santé. Ceci, en impliquant aussi bien les acteurs économiques et sociaux que les professionnels de santé.

L'AMO vise tous les salariés et tous les pensionnés des deux secteurs d'activité (public et privé). Les bénéficiaires, dont il est question, sont les fonctionnaires, les agents et les employés des établissements régis par l'AMO :

« *Les salariés du régime public sont gérés par la caisse nationale des organismes de prévoyance sociale (CNOPS) et ceux du régime privé par la caisse nationale de sécurité sociale (CNSS*) »

(Réglementation AMO loi 65.00)

Depuis 2015, des efforts sont fournis afin de généraliser de manière progressive l'assurance maladie obligatoire.

1.1.2 Les organismes d'assurance santé complémentaire

Il existe deux façons d'être assuré par une complémentaire santé :

- Assurance complémentaire à titre individuel.

Il est possible de souscrire à une complémentaire santé individuelle, dit « contrat individuel ». Dans ce cas, le tarif est calculé en fonction des garanties choisies par le souscripteur qui paye intégralement sa cotisation.

- Assurance complémentaire par contrat collectif

Il est également possible d'être couvert par une complémentaire santé dans son entreprise soit par un contrat issu de la convention collective (branche professionnelle) soit par un contrat négocié par l'entreprise elle-même (mis en place par accord collectif, référendum ou décision unilatérale de l'employeur). Dans ce cas le tarif est calculé en fonction des garanties choisies, et peut l'être aussi selon la classification des salariés (Cadres, Non Cadres). Mais l'employeur prend en charge une partie des cotisations. On parle alors de « contrat collectif ». Celui-ci est assorti d'exonérations sociales pour l'employeur (ce qui constitue une incitation à sa mise en place) et d'exonérations fiscales pour les salariés.

1.2 Couverture médicale : Cas de l'OCP

1.2.1 Basculement vers la CNOPS

Depuis l'application de l'AMO en 2005, le Maroc a fait un pas significatif sur le chantier de l'accès universel à la santé par la loi N°65-00, soumettant toute entreprise disposant d'une loi interne à passer vers l'AMO. C'est bien le cas de l'OCP.

Avant de l'externaliser au profit de l'RCAR en 2008, L'OCP se dotait d'une caisse interne (CIR), elle devait donc passer vers l'AMO. Puisque c'est est un organisme public, ses salariés sont donc gérés par le CNOPS.

Il est à noter que l'OCP avait plusieurs caisses pour couvrir ces différentes prestations notamment, celles relatives à la prévoyance sociale. Nous énumérons :

- La caisse de retraite
- La caisse médicale pour les retraités
- La caisse médicale pour les actifs
- La caisse accident de travail.

Ces prestations constituent une dette sociale que le Groupe doit évaluer selon la norme IAS19 que nous verrons ultérieurement dans ce mémoire.

L'année 2010 a vu le lancement d'un projet d'envergure en matière de couverture médicale. Ce projet porte sur la refonte des régimes de couverture médicale existants afin d'offrir au personnel OCP une meilleure couverture médicale. L'une des ambitions de ce projet est la préparation de l'externalisation de la caisse interne d'assurance maladie et l'amélioration des prestations et de la qualité de la prise en charge du personnel dans les structures externes. Le processus d'externalisation est toujours en cours, et le calcul des engagements médicaux se fait à la lumière de celui-ci.

1.2.2 Le taux de cotisation

En ce qui concerne l'OCP, les cotisations étaient fixées à 3% de la pension, représentant la part salariale et d'une part patronale de 9%, soit à un total de 12%.Quant à la CNOPS, ses retraités doivent cotiser à 2.5% de leur pension (avec un plafond de400 Dhs et un minimum de 70 Dhs)

1.2.3 Nécessité d'une Assurance complémentaire

Nous pouvons remarquer que le panier de soins proposé par l'OCP est, en termes de pourcentage de remboursement, meilleur que celui relatif à la CNOPS.

Le panier de soin relatif à l'OCP se présente comme suit :

- 80% pour le médicament ambulatoire sur la base PPM (Prix Public Maroc) du médicament générique s'il existe et sur la base de la liste des médicaments remboursables.
- 100 % des frais engagés pour les autres soins ambulatoires (biologie, radiologie, dentaire, consultations) sur la base du tarif national de référence ou le tarif adopté par l'OCP
- 100% des frais d'hospitalisation dans une structure privée sur la base du tarif national de référence
- 100% en cas d'hospitalisation dans le secteur public
- 100% pour les affections de longue durée et coûteuses (ALD et ALC).

Tandis que celui de la CNOPS se présente comme suit :

- 70% pour le médicament ambulatoire sur la base du médicament générique s'il existe et sur la base de la liste des médicaments remboursables
- 80% des frais engagés pour les autres soins ambulatoires (biologie, radiologie, dentaire, consultations) sur la base du tarif national de référence ou le tarif adopté par la CNOPS
- 90% des frais d'hospitalisation dans une structure privée sur la base du tarif national de référence
- 100% en cas d'hospitalisation dans le secteur public
- 100% pour les affections de longue durée et coûteuses (ALD et ALC).

Par conséquent, afin d'offrir une meilleure couverture médicale à son personnel comme incitait le projet de refonte de son régime de couverture médical, et surtout de maintenir le même niveau de prestation à ses employés après l'externalisation, l'OCP doit donc offrir une assurance complémentaire.

Les adhérents à cette Assurance complémentaire bénéficient ainsi que les membres de leurs familles, d'une Assurance Médicale Complémentaire qui leur permet d'obtenir des remboursements et / ou des prises en charge en complément, à ceux octroyés par le régime de couverture médicale de base

Il est à noter que L'AMC ne couvre pas les maladies courantes ou soins ambulatoires. Elle ne couvre pas non plus les maladies non prises en charge par la Caisse Nationale des Organismes de Prévoyance Sociale (**CNOPS**).

Chapitre 2 : Contexte financier : Evaluation des engagements sociaux en norme IAS 19

2.1 Les Normes IAS/IFRS

2.1.1 Le Timeline

Le pays qui fut le premier à développer une profession comptable indépendante est le Royaume Uni (à partir de 1850). En 1970, l'«Accounting Standards Committee» (ASC),organisme chargé de développer de nouvelles normes, est créé par l'« Institute of Chartered Accountants of England and Wales ».

En 1973, l'«International Accounting Standards Committee» (IASC) fut fondé, à Londres, sur la base du modèle anglo-saxon susmentionné et sous l'influence américaine du « Financial Accounting Standards Board » (FASB) et ce par un accord entre les institutions comptables de dix pays à savoir, le Canada, l'Australie, la France, le Japon, l'Allemagne, les Pays-Bas le Mexique, le Royaume-Uni, l'Irlande et les Etats-Unis.

En 2001, l'IASC devient « l'International Accounting Standards Board » (IASB) suite à une réorganisation ambitionnant de donner un nouvel élan au projet des normes internationales .Les normes publiées ultérieurement ont pris le nom d'IFRS (International Financial Reporting Standard), celles édictées avant la date du 1er avril 2001restent intitulées IAS.

2.1.2 Structure organisationnelle de l'IASB

- Un organe de surveillance sous forme de fondation appelé «International Financial Reporting Standard Fondation» (IFRS Foundation) : Il s'agit d'une entité privée à but non lucratif qui rassemble des fonds servant à financer les opérations de l'IASB dans le cadre de ses fonctions de normalisateur comptable indépendant.
- Un organe de conseil (IFRS Advisory Council), dont l'objectif est de fournir un forum à différentes organisations et personnes qui s'intéressent à l'information financière internationale et qui souhaitent s'exprimer dans le cadre du processus de normalisation
- Une instance interprétative « International Financial Reporting Standards Interpretations Committee»(IFRSIC) dont le rôle est d'interpréter l'application des Normes internationales d'information financière.

2.1.3 Objectifs de l'IASB

Les objectifs de l'IASB sont les suivants :

- Elaborer, dans l'intérêt général, un jeu unique de normes comptables de haute qualité, compréhensibles et que l'on puisse faire appliquer dans le monde entier, imposant la fourniture dans les états financiers et autres informations financières, d'informations de haute qualité, transparentes et comparables, de manière à aider les différents intervenants sur les marchés de capitaux du monde, ainsi que les autres utilisateurs de ces informations dans leur prise de décisions économiques
- Promouvoir l'utilisation et l'application rigoureuse de ces normes
- Travailler activement avec des normalisateurs nationaux pour porter la convergence des normes comptables nationales et des Normes internationales d'information financière vers des solutions de haute qualité.

Il est à préciser que L'IASB est un organisme privé, indépendant dans la mesure où il n'a de compte à rendre qu'aux fondations qui le financent.

2.2 Le champ d'application de la norme IAS 19

La norme IAS 19 est la norme du référentiel IFRS qui s'applique à la comptabilisation de tous les avantages du personnel, c'est-à-dire à « toute forme de contrepartie donnée par une entité au titre des services rendus par son personnel ».

Les avantages du personnel sont répartis en quatre 4 grandes catégories :

- Les indemnités de fin de contrat de travail ou "***Terminationbenefits***" : il s'agit d'avantages perçus à la rupture du contrat de travail avant la date normale de départ à la retraite ;
- Les avantages postérieurs à l'emploi ou "***Post-employmentbenefits*** " : ils désignent les avantages obtenus après la cessation d'activité des salariés (pensions, retraites, assurances vie, régimes de frais de santé des retraités, etc.) ;
- Les avantages à court terme ou ***"short-termemployeebenefits*** " : ils désignent les avantages obtenus dans les douze mois qui suivent la période d'acquisition des droits nécessaires à l'obtention de ces avantages (congés payés, congés maladie, salaires, cotisations sociales, régimes de frais de santé des actifs, etc.)
- Les avantages à long terme ou "***other long-termemployeebenefits*** " : ils désignent les avantages du personnel qui n'entrent pas dans les trois premières catégories (congés liés à l'ancienneté, congés sabbatiques, jubilés, etc.).

2.3 Les objectifs de la norme IAS 19

La norme IAS 19 vient compléter les textes en vigueur en matière de comptabilisation, d'information et d'évaluation des engagements relatifs aux avantages du personnel quelle que soit leur origine (conventions collectives, accords nationaux, etc.). Elle poursuit notamment les objectifs suivants :

- Améliorer la qualité de l'information transmise aux investisseurs et favoriser la comparabilité des états financiers de l'entreprise ;
- Déterminer l'impact économique des engagements sociaux à moyen et à long terme
- Obliger l'entreprise à surveiller ses engagements sociaux à long terme ;
- Comptabiliser le coût réel des engagements futurs à moyen et long terme par les méthodes décrites par la norme (notamment les méthodes ***d'actualisation*** et de ***probabilisation*** ainsi que la méthode ***des unités de crédits projetées*** que nous aborderons plus loin dans ce chapitre).

Ces objectifs visent à conserver la confiance des investisseurs et à assurer la pérennité des régimes sociaux de l'entreprise.

2.4 Les avantages postérieurs à l'emploi

Comme nous l'avant cité en introduction, les régimes de frais de santé peuvent appartenir à la catégorie des avantages à court terme lorsqu'ils s'appliquent à la période d'activité des salariés. Cependant, ils entrent dans la catégorie des avantages postérieurs à l'emploi dès lors que les salariés profitent de ces régimes durant leur période de retraite :

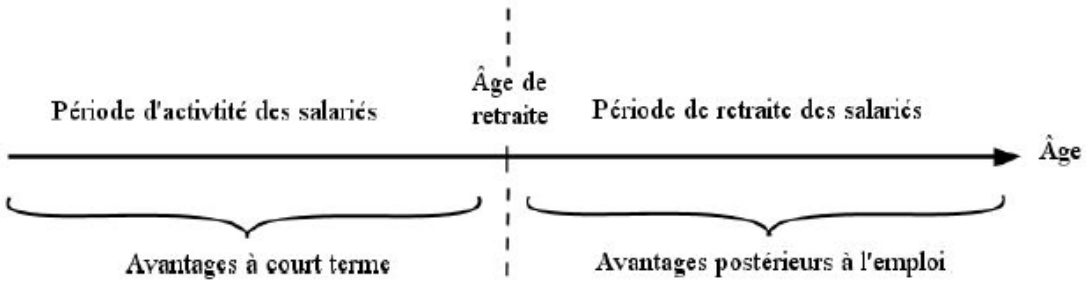


Figure 1 : Période de jouissement des avantages court terme et postérieurs à l'emploi

L'évaluation des engagements relatifs aux avantages postérieurs à l'emploi nécessite l'utilisation d'hypothèses actuarielles (contrairement aux avantages à court terme). Ce type d'avantages servant de cadre de base pour l'évaluation de l'engagement social de l'OCP que nous allons étudier, nous nous limiterons donc à ce type d'avantages pour toute la suite de cette étude.

Les régimes d'avantages postérieurs à l'emploi sont classés en deux types de régimes : les régimes à cotisations définies et les régimes à prestations définies. Ce classement est censé refléter la réalité économique du régime, et les obligations des différents acteurs du régime (l'entreprise, l'organisme gestionnaire du régime, les adhérents, etc.).

2.4.1 Les régimes à cotisations définies

Dans un régime à cotisations définies, l'entreprise s'engage juridiquement sur un montant à verser à l'organisme gestionnaire du régime pour les adhérents dudit régime au titre des avantages postérieurs à l'emploi prévus par le régime. L'ensemble des cotisations versées (par l'entreprise ou les adhérents) à l'organisme gestionnaire servira à verser les différentes prestations prévues par le régime.

De plus, les obligations sont évaluées sans aucune hypothèse actuarielle. En effet, « le risque actuariel (risque que les prestations soient moins importantes que prévu) et le risque de placement (risque que les actifs investis ne soient pas suffisants pour faire face aux prestations prévues) incombent » uniquement à l'organisme gestionnaire du régime.

2.4.2 Les régimes à prestations définies

Dans un régime à prestations définies, l'entreprise s'engage juridiquement sur le niveau des prestations versées aux adhérents du régime (salariés actifs et retraités). Ce type de régime peut être partiellement ou totalement financé par les cotisations de l'entreprise et/ou des adhérents du régime.

Comme dans les régimes à cotisations définies, l'organisme gestionnaire du régime est chargé d'évaluer et de recevoir les cotisations nécessaires au bon fonctionnement du régime.

Au vu de l'importance des enjeux financiers (actifs de couverture du fond et engagements sociaux de l'entreprise), un tel régime doit impérativement être suivi de manière quasi-continue, si l'entreprise veut éviter de se retrouver en difficulté.

Ainsi, « le risque actuariel (risque que les prestations coûtent plus cher que prévu) et le risque de placement incombent » entièrement à l'entreprise. L'entreprise est tenue par une obligation de résultat.

Les engagements de tels régimes portant sur du moyen et long terme, leur évaluation est complexe et soumise à de nombreux aléas actuariels (mortalité, taux de turn-over, etc.). Les montants évalués sont caractérisés par une forte volatilité liée à la sélection des hypothèses actuarielles.

Ce type de régime est le seul nécessitant l'utilisation d'hypothèses actuarielles et représente la majeure partie des régimes encadrés par la norme IAS 19. Nous limiterons le reste de la présentation d'IAS 19 à ce régime.

2.4.3 Les hypothèses actuarielles

L'évaluation d'engagements, portant sur des avantages postérieurs à l'emploi de régimes à prestations définies, nécessite la détermination et l'utilisation d'hypothèses actuarielles que la norme IAS 19 pose comme devant être « objectives et mutuellement compatibles », c'est-à-dire

qu'elles doivent être ***ni trop risquées ni trop prudentes*** et qu'elles « traduisent les rapports économiques existant » entre les différents facteurs entrant en jeu (inflation, taux d'actualisation, etc.). La détermination des hypothèses actuarielles ne doit donc pas présenter de contradictions.

De manière générale, ces hypothèses sont décrites par la norme comme devant être « les meilleures estimations faites par l'entité des variables qui détermineront le coût final des avantages postérieurs à l'emploi ». Elles sont réparties en deux grandes catégories : les hypothèses démographiques et les hypothèses financières.

1) Les hypothèses démographiques

Les hypothèses démographiques traduisent les caractéristiques démographiques actuelles et futures du personnel actif ou retraité (ou de leurs ayants-droit) « réunissant les conditions requises pour bénéficier des avantages » du régime.

Les régimes considérés portant sur des avantages à long terme, on exclura de l'évaluation toute personne étant dans l'entreprise pour du court terme (les stagiaires, etc.).

Dans le cadre d'IAS 19, la population étudiée doit être considérée comme étant fermée. Ces hypothèses permettent ainsi de caractériser le « volume » des personnes ayant droit aux prestations et la durée sur laquelle ces personnes auront accès aux prestations. Elles sont généralement propres à l'entreprise ou à un secteur d'activité. Parmi ces hypothèses, on peut notamment citer :

- La mortalité du personnel actif et retraité ;
- Le taux de rotation du personnel ;
- L'âge de départ à la retraite ;
- La proportion de membres affiliés au régime et de personnes à leur charge respectant les conditions nécessaires pour disposer des prestations ;
- Le taux de demandes d'indemnisation en vertu de régimes médicaux ;
- etc.

2) Les hypothèses financières

Les hypothèses financières permettent d'intégrer le contexte économique global dans lequel est réalisée l'évaluation des engagements et permettent notamment de comparer des valeurs économiques ayant cours à des dates différentes. Les hypothèses les plus courantes sont les suivantes :

- Le taux d'actualisation ;
- Le taux d'augmentation des salaires ;
- L'évolution du plafond annuel de la Sécurité Sociale ;
- L'évolution des coûts médicaux ;
- Le taux attendu du rendement des actifs du régime ;
- etc.

2.5 L'évaluation des engagements

Dans le cadre d'une évaluation des engagements à long terme, les flux à verser n'étant pas certains et n'ayant pas la même valeur économique au cours du temps, il est nécessaire de recourir aux mécanismes de l'actualisation et de la probabilisation. De plus, le cadre d'IAS 19 impose une évaluation tête par tête des engagements.

2.5.1 La formulation générale

1) La valeur probabilisé

Dans un régime à prestations définies, l'entreprise s'engage envers un adhérent n ∈ [1 : N] (N étant le nombre total d'adhérents du régime) du régime à lui verser une prestation C_t^n, à une date t donnée, dès que l'adhérent a acquis définitivement les droits pour bénéficier de cette prestation.

On place l'évaluation à un instant $t_0 \leq$ t. Les droits n'étant pas nécessairement acquis à l'adhérent n en t_0, il est nécessaire de probabiliser le montant C_t^n futur entre la date t_0 et la date t. On calcule par conséquent la valeur probabilisée (VP) de C_t^n en t_0 :

$$VP_{t0}^n = P(t_0, t)^n \times C_t^n$$

Avec :

$P(t_0, t)^n$: la probabilité, en se plaçant en t_0, que la prestation C_t^n soit effectivement versée à l'adhérent n en t, c'est-à-dire la probabilité que la personne puisse bénéficier du régime en t. Ce facteur a deux composantes :

- La probabilité que la personne survive entre $t_0$et t ;
- La probabilité que la personne reste dans l'entreprise entre t_0 et t,

2) La valeur actuelle probable future

La valeur économique en t du montant C_t^n n'étant pas la même que celle qu'il a en t_0, le calcul des engagements doit prendre en compte un facteur d'actualisation, c'est-à-dire que le montant calculé précédemment (VP) doit être actualisé en t_0 via un taux d'actualisation i (pour simplifier les expressions on le suppose constant au cours du temps). On obtient ainsi la valeur actuelle probable future (VAPF) du montant C_t^n :

$$VAPF_{t_0}^{n} = P(t_0, t)^n \times C_t^n \times (1 + i)^{t_0 - t}$$

La VAPF totale à t_0 est alors égale à la somme des $VAPF_{t_0}^{n}$, pour n ∈ [1 : N] :

$$VAPF_{t_0} = \sum_{n=1}^{N} VAPF_{t_0}^{n}$$

2.5.2 LA méthode d'évaluation et comptabilisation des engagements

La valeur actuelle de l'obligation ou Defined Benefit Obligation (DBO) désigne « la valeur actuelle, avant déduction des actifs du régime, des paiements futurs attendus qui sont nécessaires pour éteindre l'obligation résultant des services rendus au cours de la période en cours et des périodes antérieures ». Ce montant est obtenu à partir de la **VAPF** calculée précédemment selon une méthode appelée « méthode des unités de crédits projetés » ou **Projected Unit Credit Method** (**PUCM**).

Cette méthode considère que « chaque période de service donne lieu à une unité supplémentaire de droits à prestations ». La VAPF est de ce fait répartie au prorata des unités de services rendus par l'adhérent du régime, définissant ainsi la valeur de l'engagement au moment de l'évaluation sur la base du rythme d'acquisition des droits prévus par le régime.

De plus, il faut évaluer « séparément chacune de ces unités pour obtenir l'obligation finale », c'est-à-dire que le calcul doit être fait tête par tête, comme nous l'avons initialisé précédemment dans le calcul de la VAPF. L'engagement de l'entreprise sera alors la somme des engagements individuels.

On obtient une expression de la DBO du type :

$$DBO_{t0}^{n} = VAPF_{t0}^{n} \times \alpha_{t0}^{k}$$

Où $\alpha_{t0}^{k} \in$]0 : 1] caractérise la méthode k utilisée (avec k ∈ [1;3])

Les méthodes pouvant être utilisées sont :

$$\alpha_{t0}^{1} = \frac{Ancienneté actuelle}{Ancienneté à terme}$$

$$\alpha_{t0}^{2} = \frac{droits acquis actuels}{droits acquis à terme}$$

$$\alpha_{t0}^{3} = \frac{Ancienneté actuelle}{Ancienneté plafond}$$

Dans notre cas nous retiendrons la première qui est utilisée à l'OCP.Afin d'illustrer la suite de ce mémoire, on calculera ainsi la DBO de la manière suivante :

$$DBO_{t0}^{n} = VAPF_{t0}^{n} \times \frac{Ancienneté actuelle_{t0}^{n}}{Ancienneté à terme^{n}}$$

Avec :

- $Ancienneté_{t0}^{n}$: l'ancienneté actuelle du salarié n ;
- $Ancienneté à terme^{n}$: l'ancienneté au terme (ancienneté à l'age de départ à la retraite) du salarié n.

De même que pour la VAPF, La DBO_{t0} totale en $t_0$est alors égale à la somme des DBO_{t0}^{n}, pour n∈ [1 : N], c'est-à-dire :

$DBO_{t0} = \sum_{n=1}^{n=N} DBO_{t0}^{n}$

- Ce montant sera par conséquent la valeur comptabilisée au titre des engagements de l'entreprisse, en occurrence l'OCP.
- Les avantages post emploi à prestations définies sont constatés sous forme de *« dette long terme »* au bilan.

Conclusion

Nous avons montré à travers ce chapitre le contexte dans lequel s'inscrit notre projet de fin d'études. Après avoir pris connaissance de la norme IAS19, les technicités qu'elle enferme devraient être mises en application. Nous commencerons donc par expliciter les hypothèses financières dans une deuxième partie.

Partie 2 : Modélisation des hypothèses actuarielles : taux d'actualisation et inflation médicale

Après avoir pris connaissance des hypothèses financières dans le 2ème chapitre de la 1ère partie (voir 2.4.3), nous serons amenées à définir celles-ci, à côté des hypothèses démographiques. Cette dernière ne demandera pas un grand effort de notre part, puisque nous userons, sous les directives, de la table de mortalité TV 73 -77 pour modéliser la mortalité d'une part, et le taux de rotation étant faible ici à l'OCP, ce paramètre ne fera pas objet d'une modélisation dans ce mémoire.

Ainsi dans ce qui suit, un essai de notre part à comprendre la structure du taux d'actualisation et le taux d'inflation.

CHAPITRE 1 Taux d'actualisation : avant-propos

Le taux d'actualisation est une hypothèse qui a un très fort impact sur la valeur des engagements. Le taux d'actualisation à appliquer pour l'évaluation des engagements sociaux est encadré de manière bien précise par les articles 78 à 81 de la norme IAS19.

Concernant la détermination de ce taux, l'article 78 indique que le taux à appliquer « doit être déterminé par référence à un taux de marché, à la date de clôture, fondé sur les obligations d'entités de première catégorie. Dans les pays où ce type de marché n'est pas actif, il faut prendre le taux des obligations d'État ».

Pour notre cas, on retiendra les taux d'obligations d'État qui sont assimilés à des taux sans risques augmentés d'une prime de risque. De plus,. L'article 79 définit le taux d'actualisation comme étant le taux qui « traduit la valeur temps de l'argent ».

L'article 80 indique que la norme tolère que l'on puisse utiliser « un taux d'actualisation moyen, unique et pondéré », c'est-à-dire une structure plate des taux d'actualisation. Cependant, l'article81précise qu'il est préférable que l'entreprise estime le taux d'actualisation à long terme

« par extrapolation des taux actuels du marché à l'aide de la courbe des taux de rendement », c'est-à-dire en travaillant sur une structure non plate des taux.

Dans le 3ème chapitre de cette partie, nous essayerons de construire la courbe des taux de rendement afin de pouvoir estimer le taux d'actualisation qui va être utilisé par ailleurs dans le calcul de l'engagement médical de l'OCP. Mais avant d'entamer cette modélisation, Il nous a paru nécessaire de présenter en premier lieu quelques outils du calcul financier basiques du taux d'intérêt, et donner une brève présentation du taux de référence marocain et de la structure par terme des taux.

1.1Concepts fondamentaux

L'intérêt est fréquemment appelé la valeur temps de l'argent, c'est à dire la valeur accordée aujourd'hui au fait de ne pas utiliser l'argent dont on dispose pour une utilisation immédiate (i.e. la consommation). Le fait de renoncer à consommer aujourd'hui pour prêter de l'argent a un prix, et ce prix peut être assimilé au taux d'intérêt. On revient dans ce qui suit sur les principes de base que nous devrions assimiler autour de la notion du taux d'intérêt et ce avant de pouvoir l'exploiter plus étroitement dans la modélisation de la courbe des taux Zéro-Coupon.

De façon général, en investissant un montant M à un taux r (en %), vous vous retrouvez avec M(1 + r) à la fin de l'année. Ce principe relativement simple, se complexifie légèrement lorsque l'on dépasse l'horizon d'investissement d'un an. On distingue alors les intérêts simples des intérêts composés

1.1.1Valeur Actuelle Nette

La Valeur Actuelle Nette est une méthode simple permettant de juger de l'intérêt d'un placement. C'est une méthode largement présentée, sinon utilisée, lors de la sélection de différents investissements. L'idée est simple : on actualise les flux de revenu futurs occasionnés par le placement et on les compare au montant initial de l'investissement.

La formule générale de la VAN est la suivante :

$$VAN = -I + \sum_{p=1}^{n} \frac{F_p}{(1+r)^p}$$

Avec :

F_p : flux net de trésorerie de la période p

I : Investissement Initial

n : durée de vie du projet

r : taux d'intérêt

Ce principe n'est pas utile en tant que tel pour les produits de taux, mais sa démarche est essentielle pour comprendre le fonctionnement des principes de base de l'évaluation des produits de taux et notamment dans la reconstitution de la courbe de taux avec la méthode du Bootstrapping

1.2 Les Taux de base

L'outil de base d'un modèle de taux est l'obligation zéro-coupon, produit financier versant 1 MAD à la maturité T. On notera P (t, T) son prix à la date t ∈ [0, T].

A partir du prix de l'obligation zéro-coupon on peut définir plusieurs types de taux :

1. Taux d'intérêt instantané continûment composé

C'est la quantité R(t, T) définie par :

$$R(t,T) = -\frac{\ln(P(t,T)}{T-t}$$

R(t,T) est parfois dit "taux zéro-coupon" : c'est le taux continu constant sur la période [t,T] tel que

$$P(t,T) = e^{-R(t,T)(T-t)}$$

2. Taux d'intérêt simplement composé

C'est la quantité L(t, T) définie par :

$$L(t,T) = \frac{1-P(t,T)}{P(t,T)(T-t)}$$

Tel que : $P(t,T) = \frac{1}{1+L(t,T)(T-t)}$

Il est à noter que les taux interbancaires sont définis de la sorte.

3. Taux d'intérêt composé annuellement

C'est la quantité Y (t, T) définie par :

$$Y(t,T) = \frac{1}{P(t,T)^{\frac{1}{T-t}} - 1}$$

Y (t, T) est aussi appelé taux actuariel

4. Taux spot

On appelle taux court (ou taux instantané) la limite de tous ces taux lorsque T tend vers t. On la note r(t)

5. Conventions

Les définitions précédentes montrent que le temps restant à maturité T - t est une grandeur essentielle pour la modélisation de taux d'intérêts. Différentes conventions sont utilisées sur les marchés de taux afin que tous les intervenants mesurent cette quantité de la même manière. Les conventions les plus courantes sont les suivantes :

- Actuel/365 : une année compte 365 jours
- Actuel/360 : une année compte 360 jours
- 30/360 : une année compte 360 jours et un mois 30 jours.

1.3Taux spot de référence au Maroc

Des taux de référence sont publiés quotidiennement dans le site de Bank-Al-Maghrib. Ce sont des taux interbancaires calculés à partir de taux effectivement pratiqués par un panel des banques exerçant au Maroc.

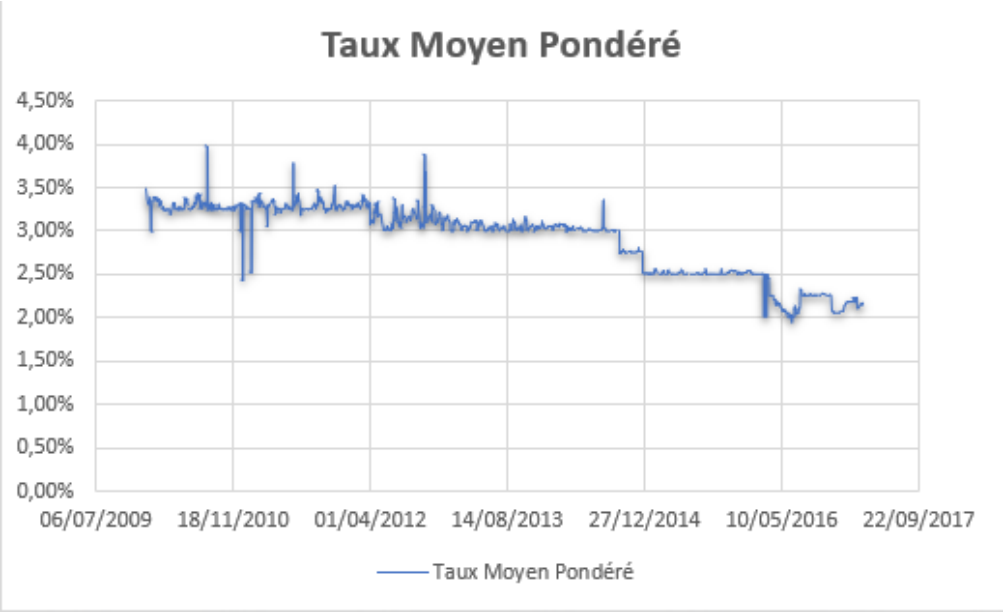


Figure 2 : Historique des taux (2010-2017) de référence marocain TMP. Source : BAM

1.4 Structure par terme des taux

On appelle courbe des taux ou structure par terme des taux la fonction associant à une maturité T un taux R, L ou Y. De telles courbes sont construites à partir des prix d'obligations observés sur les marchés étudiés. La courbe de référence est la courbe dite "zéro-coupon".

Etant des instruments théoriques (ou très peu liquides s'ils existent), on reconstruit cette courbe à partir des obligations réelles (couponnées) échangées sur les marchés. Les problématiques essentiellement rencontrées lors de cette reconstruction sont le "stripping" des obligations (i.e. le calcul théorique du zéro-coupon à partir d'une obligation couponnée) et le lissage de la courbe obtenue

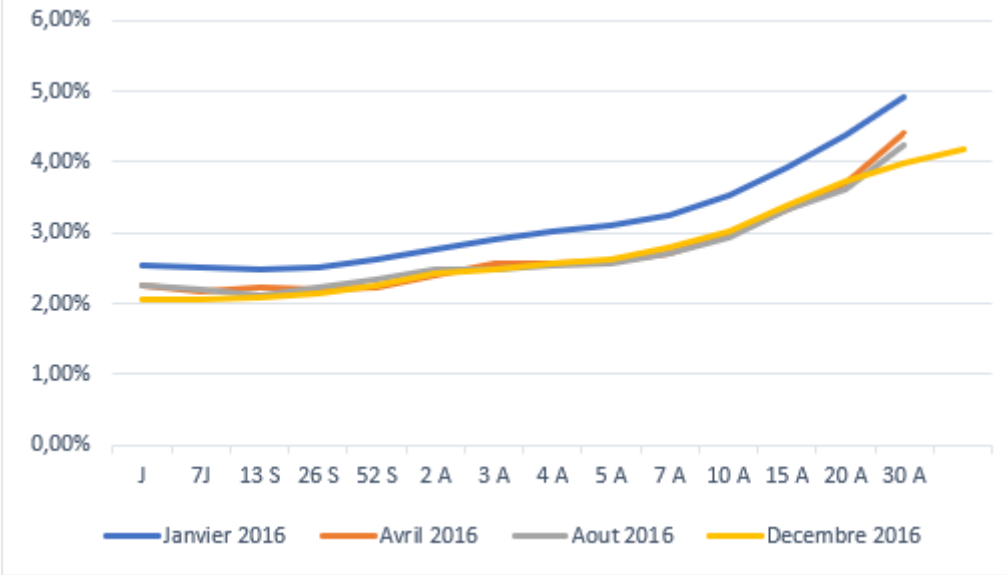


Figure 3 : Courbe des taux de rendement de l'Etat marocain pour l'année 2016 . Source :BAM , Réalisée par nos soins.

1) La courbe des taux

Comme mentionné précédemment, la représentation des taux d'intérêt d'obligations de maturités différentes, toutes choses égales par ailleurs, s'appelle **courbe des taux** (*yieldcurve*).

La courbe des taux peut être:

- **croissante**(taux à long terme supérieurs aux taux à court terme),
- **plate**(taux à long terme égaux aux taux à court terme),
- **décroissante ou inversée** (taux à long terme inférieurs aux taux à

court terme),

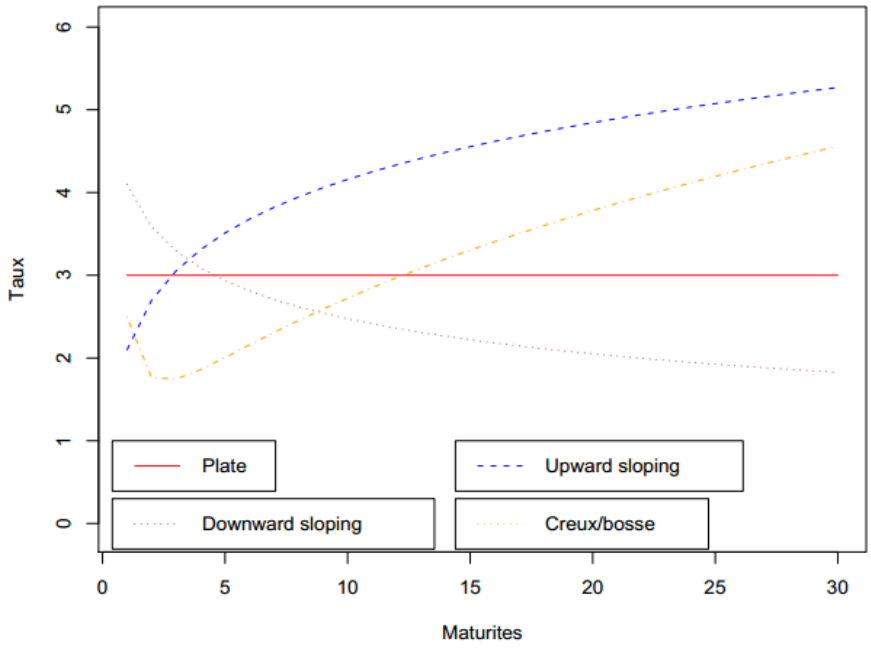


Figure 4 : Différentes catégories de courbe de taux

ou prendre une forme plus compliquée.

Les études empiriques ont mis en évidence les trois faits stylisés suivants :

1. les taux d'intérêt pour des obligations de maturités différentes varient habituellement ensemble dans le temps ;
2. la courbe des taux est plus fréquemment croissante lorsque le taux à court terme est bas, plus fréquemment inversée lorsque le taux à court terme est élevé (ou, de manière équivalente, les taux à court terme fluctuent en général davantage que les taux à long terme) ;
3. la courbe des taux est habituellement croissante.

2) Théorie pour expliquer la structure par terme des taux d'intérêt

Quatre théories ont été mises en avant pour expliquer la structure par terme des taux d'intérêt :

- la théorie des anticipations, qui peut expliquer les faits 1 et 2 ;
- la théorie des marchés segmentés, qui peut expliquer le fait 3 ;
- la théorie de la prime de liquidité et la théorie de l'habitat préféré, qui peuvent expliquer les faits 1, 2, et 3.

Ces quatre théories répondent à la question suivante : comment les résultats obtenus en univers certain se modifient-ils, lorsqu'on prend en compte l'incertitude sur les taux futurs ? De ce point de vue, il se dégage deux positions extrêmes et une position médiane.

La théorie des anticipations occupe une position extrême, en faisant l'hypothèse que les agents souhaitant prêter pendant n périodes n'ont pas de préférence entre acheter des obligations d'une maturité plutôt que d'une autre de même rendement; ainsi les obligations de maturité différentes sont **parfaitement substituables**. En d'autres termes, l'introduction de l'incertitude ne modifie pas substantiellement les résultats en monde certain.

La théorie de l'habitat préféré occupe une position médiane, en précisant que l'attitude des agents vis-à-vis du risque justifie l'existence de primes de risque. Plusieurs facteurs déterminent alors la structure des taux : les anticipations, mais aussi la distribution des habitats sur l'échelle des maturités, ainsi que l'aversion des individus pour le risque.

La théorie de la prime de liquidité fait l'hypothèse que les obligations de maturités différentes sont imparfaitement substituables, les investisseurs préférant les obligations à court terme car elles présentent moins de risque de taux d'intérêt. Pour que les obligations à long terme soient détenues, il faut donc que leur rendement comporte une prime de liquidité positive qui compense leurs inconvénients par rapport aux obligations de court terme.

Enfin, la théorie de la segmentation occupe l'autre extrême, dans le sens où elle soutient l'idée suivant laquelle les obligations de maturités différentes ne sont **pas du tout substituables**. Les marchés d'obligations de maturités différentes sont donc entièrement séparés (ou "segmentés") et indépendants entre eux. L'idée est que les investisseurs ont des raisons importantes de préférer une maturité à une autre. Par exemple, ils peuvent avoir une durée précise de placement en tête (suivant l'horizon de leurs dépenses prévues) et souhaiter que la maturité de leur placement soit identique pour supprimer le risque.

Les acteurs sont ainsi si adverses à la spéculation que les anticipations ne constituent plus un facteur déterminant. Chaque taux d'intérêt résulte de la confrontation d'une offre et d'une demande spécifique.

1.5 Taux à terme (*forward*)

On appelle le taux forward instantané à la date t pour une maturité T la quantité :

$$f(t,T) = -\frac{\partial \ln \mathrm{P(t,T)}}{\partial T}$$

i.e. tel que :

$$P(t,T) = \exp\left(-\int_t^T f(t,u)du\right)$$

On peut définir alors de façon cohérente les prix de zéro-coupon forward à la date t pour l'expiration T et la maturité S par

$$P(t,T,S) = \frac{P(t,S)}{P(t,T)}$$

P(t, T, S) est la valeur vue à la date t du montant à payer à la date T pour acheter l'obligation zéro-coupon qui versera un dirham à la date S.

Dans ce chapitre, nous avons abordé les différentes notions-clés qui nous aiderons à comprendre la structure par terme des taux d'intérêt ,avant de comprendre comment se construit une courbe de taux. Le chapitre suivant ce fait avec plus de détails avec un cas pratique que nous expliquerons l'intérêt.

CHAPITRE 2 Lissage de la courbe des taux de rendements

L'heuristique de reconstruire la courbe des taux coupon zéro à partir des obligations cotées et liquides est que dans le cas de l'utilisation d'une structure de taux réelle, on actualise un flux à partir du taux d'actualisation ayant la même maturité que ce flux. Afin de déterminer ces taux d'actualisation, nous avons besoin de la courbe des taux zéro-coupon calculée à partir des données de marché marocain.

Cependant, reconstruire la courbe des taux coupon zéro à partir des obligations cotées et liquides ne nous était pas possible, puisque le département où nous étions assignées ne disposait pas de la liste (ou d'une base) des obligations cotées et liquides sur le marché marocain. Nous allons pourtant, à la demande de notre encadrant, introduire le principe de la constitution de la courbe des taux zéro coupon mais nous ne la construirons pas par nous-mêmes.

D'une autre part, l'article 80 de la norme IAS19 indique que la norme tolère que l'on puisse utiliser « un taux d'actualisation moyen, unique et pondéré », c'est-à-dire une structure plate des taux d'actualisation. Le principe de la détermination de ce taux moyen sera exhibé dans la prochaine partie quand nous introduirons le principe de la duration.

Il est à savoir que le taux d'actualisation utilisé pour l'évaluation des titres émis par l'Etat est le taux fourni par la courbe des taux de référence publiée quotidiennement par Bank Al Maghrib. L'estimation de chaque taux de référence équivalent à une maturité sur laquelle la courbe des taux publiée ne donne pas information est effectué par une méthode d'interpolation. Nous essaierons donc d'utiliser les méthodes indirectes de constitution de la courbe des taux pour retrouver ces taux-là.

2.1 Quelques notions importantes

2.1.1 Les obligations

Les obligations sont des titres de créance détenus par un ou plusieurs porteurs à l'encontre d'un emprunteur. Derrière cette définition se cache une idée très simple : quand un particulier souhaite emprunter de l'argent pour financer l'achat de sa maison de son appartement, la somme empruntée est dérisoire, au regard des réserves dont dispose la banque. Lorsqu'une entreprise souhaite financer un projet d'investissement, le coût est très largement supérieur. La banque n'est alors plus en mesure de fournir seule la somme requise sans se mettre dramatiquement en danger. Une solution simple à ce problème consiste à découper la somme à prêter en autant de

petites sommes qu'il le faut, et à proposer ces micro-prêts sur un marché organisé : le marché obligataire. Chacun de ces micro-prêts est appelé obligation et porte des intérêts appelés coupons.

2.1.2 Les coupons

Les coupons correspondent, dans le cas d'une obligation, au versement des intérêts à intervalles fixes. Les coupons correspondent donc aux tombées d'intérêts régulières. Les intervalles entre deux coupons peuvent être variables. Ils sont en général semi-annuels ou annuels. Le taux de coupon, i.e. le taux d'intérêt de référence pour calculer le coupon, est négocier à l'émission de l'obligation.

2.1.3 Les catégories d'obligations

Il existe de multiples catégories d'obligations segmentées selon différents critères :

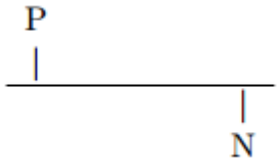
- Les obligations à taux fixe / les obligations à taux variable selon la nature fixe ou variable du taux de coupon
- Les obligations d'Etat /les obligations Corporate selon la nature publique ou privée de l'émetteur
- Les obligations à coupon /les obligations zéro-coupon selon l'existence ou non de coupons intermédiaires dans l'échéancier de remboursement
- Les obligations sans clause optionnelle / les obligations à clause optionnelle (obligations convertibles...) selon l'existence ou non d'options associées au produit purement obligataire

2.1.4 Evaluation d'un produit obligataire : Le « Zéro Coupon »

Commençons simplement par s'interroger sur la valeur d'une obligation. A la question "quelle valeur a l'obligation X ?", il est possible de fournir deux types de réponses : d'une part, il est possible de s'interroger sur le prix de marché de cette obligation. Les obligations sont cotées en continu et il est possible à tout moment d'une journée de trading d'obtenir le prix de marché, sorte de consensus autour de la valeur d'un titre.

Le prix d'une obligation ZC emprunte à ce qui a été présenté dans le cadre de la VAN :

Le prix d'un ZC est l'actualisation de son unique flux en fin de vie au taux correspondant. Entrons néanmoins dans les détails : l'étalement des flux dans le temps se présente comme suit:



Où P est le prix payé pour acheter l'obligation ZC, N est le nominal et C l'unique coupon qui lui est attaché.

$$VAN = -P + \frac{N}{(1+r)^n}$$

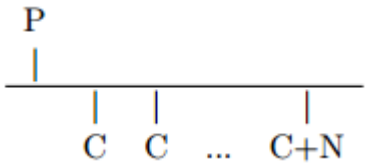
Le prix d'un ZC s'établit de façon à ce que sa VAN =0. Le taux d'actualisation permettant d'obtenir cette VAN nulle n'est donc rien d'autre que le TIR* (Taux Interne de Rentabilité). Dans le cas du calcul obligataire, on parle non pas de TIR mais de *Yield to Maturity* (YTM). Il s'agit donc du taux solution de :

$$P = \frac{N}{(1+YTM)^n}$$

*TIR : C'est le taux d'actualisation qui permet d'´égaliser la somme des cash flows futurs actualisés et le montant initialement investi. Il s'agit donc du taux d'actualisation pour lequel la VAN = 0

2.1.5 Evaluation d'une obligation à coupon

On sait à présent valoriser une obligation ZC : il est assez aisé de passer des ZC aux obligations à coupons. On rappelle que la différence fondamentale entre les deux types d'obligations est quele ZC ne délivre pas de coupons avant la maturité. Pour donner un prix à une obligation à coupon, il suffit d'observer le timing des versements attachés à une obligation à coupon. C'est ce qu'on représente sur la figure suivante :



On sait que le prix de l'obligation à coupon est nécessairement une actualisation de ses versements futurs. Le problème est alors de connaitre le taux à utiliser pour actualiser ces flux : Ce sont les taux ZC pour des maturités correspondantes aux versements qui doivent être utilisées pour l'actualisation. D'où le théorème suivant :

$$P = \frac{C}{1+r(1)} + \frac{C}{(1+r(2))^2} + \frac{C}{(1+r(3))^3} + \ldots + \frac{C+N}{(1+r(n))^n}$$

Où r(i) est le taux ZC correspondant à la maturité i

Cette fois-ci, le YTM ne correspond pas le moins du monde au taux ZC : il s'agit simplement du taux de rendement interne de l'investissement.

2.1.6 La prime de risque sur un placement à revenu fixe

1) Définition de la prime de risque

En Finance, les notions de rendement et de risque sont intimement liées. Les investisseurs ont la possibilité d'investir sur des actifs sans risque à faible rendement. Le taux sans risque est généralement assimilé au taux des obligations d'Etats correctement notés. Pour détenir un actif risqué, telles les actions et les obligations, les investisseurs exigent un rendement plus ou moins important. C'est ainsi que les prix se forment sur les marchés par rapport aux espérances de gain.

2) La prime de risque sur une obligation

La prime de risque sur une obligation est l'écart de rendement entre celle-ci et une obligation sans risque (Obligation émise par l'Etat marocain). En effet, On utilise par convention le taux des obligations d'état comme référence de taux sans risque. Au Maroc, on se base donc sur le taux de référence du marché secondaire des bons de trésor.

EXEMPLE :

Si une obligation d'Etat de maturité 10 ans offre par exemple 3,26% de rendement cela signifie qu'un investisseur achetant celle-ci pourra toucher 3,26% de rémunération par an avec un risque nul (la probabilité de défaut de l'Etat étant quasi inexistante).

Un placement promettant 4% de rendement proposera dès lors une prime de risque de 0,74%soit 74 points de base en rémunération du risque pris.

N.B : Le terme « prime de risque de marché » est très souvent lié au marché des actions. Pour ce qui nous intéresse, ce terme peut être utilisé pour nommer l'écart de rendement entre deux actifs de même classe.

2) Décomposition de la prime de risque

- **Risque de perte Sur défaut** : Les pertes attendues sur défaut constituent l'une des composantes évidentes des marges. Les estimations relatives à ces pertes sont calculées à l'aide des matrices de transition (non conditionnelle) des notations sur un an des agences de notation (comme Moody's et Fitch)
- **Risque de volatilité du marché** : La volatilité de la part non expliquée de la marge accentue encore le risque inhérent aux obligations d'entreprises
- **La Fiscalité** : les obligations d'entreprises sont assujetties à l'impôt au niveau des États, mais pas les valeurs du Trésor. Puisque les investisseurs comparent les rendements après impôts respectifs des divers instruments, le rendement de la dette d'entreprise devra donc être supérieur, pour des considérations d'arbitrage, afin de compenser le prélèvement fiscal
- **La prime de liquidité** : la plupart des obligations d'entreprises s'échangent sur des marchés relativement étroits et sont ainsi habituellement plus coûteuses à négocier que les actions et les valeurs du Trésor. Il faut donc rémunérer les investisseurs en conséquence. Plusieurs études récentes avancent même que les primes de liquidité peuvent constituer la deuxième composante des marges, après la fiscalité.

2.2 Classification des courbes des taux

A une date donnée et dans un pays ou une zone économique unifiée, il existe une multitude de courbes de taux. On distingue : les courbes de marché et les courbes implicites.

2.2.1 Les courbes de marché ou directes

Les courbes de marché sont construites directement à partir des cotations de marché d'instruments comme les obligations et les swaps. Parmi les courbes de marché:

- la courbe des taux de rendement à maturité: elle est construite à partir des taux de rendement des obligations
- la courbe des taux de swaps: elle est construite à partir des taux de swaps.

2.2.2 Les courbes implicites ou indirectes

Les courbes implicites sont dérivées indirectement à partir des cotations de marché d'instruments comme les obligations et les swaps. Parmi les courbes implicites:

- La courbe des taux zéro-coupon
- La (les) courbe(s) de taux forward
- La courbe des taux forward instantanés,
- La courbe des taux de rendement au pair

On s'intéressera dans ce qui suit aux différentes techniques qui permettent de construire une courbe des taux à partir des données de marché.

2.3 Méthodes de construction de la courbe des taux

2.3.1 La Méthode directe : Le Bootstrapping

Il s'agit d'une procédure en plusieurs étapes qui permet de reconstituer une courbe zéro-coupon au comptant "pas à pas", i.e. segment par segment de maturité

a) Pour le segment de la courbe inférieur à 1 an :

Extraction des taux zéro-coupon grâce aux prix des titres zéro-coupon cotés sur le marché puis obtention d'une courbe continue par interpolation.

b) Pour le segment de la courbe allant de 1 an à 2 ans :

Parmi les obligations de maturité comprise entre 1 an et 2 ans, on choisit l'obligation à l'échéance la plus rapprochée. Ce titre verse deux flux. Le facteur d'actualisation du premier flux est connu grâce à l 'étape 1. Le facteur d'actualisation du second flux est solution de l'équation non linéaire :

$$P = CB(0,t_1) + (N+C)B(0,t_2) \quad avec\ t_1 \leq 1\ \&\ 1 < t_2 \leq 2$$

On obtient alors un premier point de courbe sur ce segment. On réitère alors le même procédé avec l'obligation de maturité immédiatement supérieure mais toujours inférieure à 2 ans.

c) Pour le segment de la courbe allant de 2 ans à 3 ans :

On réitère l'opération précédente à partir des titres ayant une maturité comprise entre 2ans et 3 ans.

etc...

On comprend bien que lors des calculs conduisant à tracer la courbe des taux, nous sommes contraints par la gamme des échéances disponibles. Pour obtenir un taux zéro-coupon entre deux dates disponibles, il est possible de procéder à une interpolation. Deux sont particulièrement utilisées : les interpolations linéaires et cubiques.

Interpolation linéaire :

On connait les taux zéro-coupon de maturités t_1 et t_2. On souhaite interpoler le taux de maturité t avec $t_1 \leq t \leq t_2$. La formule générale est la suivante :

$$r(0,t) = \frac{(t_2 - t)r(0,t_1) + (t - t_1)r(0,t_{2)}}{t_2 - t_1}$$

Interpolation cubique :

On procède à une interpolation cubique par segment de courbes. On définit un premier segment entre t_1 et t_4 où l'on dispose de 4 taux $\mathrm{r}(0,t_1)$, $\mathrm{r}(0,t_2)$, $\mathrm{r}(0,t_3)$, $\mathrm{r}(0,t_4)$. Le taux r(0,t) est défini par :

$$r(0,t) = at^3 + bt^2 + ct + d$$

Sous contrainte que la courbe passe par les quatre points de marché $\mathrm{r}(0,t_1)$, $\mathrm{r}(0,t_2)$, $\mathrm{r}(0,t_3)$, $\mathrm{r}(0,t_4)$. D'où le système à résoudre :

$$r(0,t_1) = at_1{}^3 + bt_1{}^2 + ct_1 + d$$

$$r(0,t_2) = at_2{}^3 + bt_2{}^2 + ct_2 + d$$

$$r(0,t_3) = at_3{}^3 + bt_3{}^2 + ct_3 + d$$

$$r(0,t_4) = at_4{}^3 + bt_4{}^2 + ct_4 + d$$

2.3.2Les Méthodes indirectes

1) Les Splines Polynômiaux

Les splines polynomiaux consistent à estimer la courbe de taux par des polynômes définie par morceaux. On considère l'écriture standard comme dans l'exemple qui suit :

$$B(0,s) = \begin{cases} B_0(s) = d_0 + c_0 s + b_0 s^2 + a_0 s^3 & s\epsilon[0,5] \\ B_5(s) = d_1 + c_1 s + b_1 s^2 + a_1 s^3 & s\epsilon[5,10] \\ B_{10}(s) = d_2 + c_2 s + b_2 s^2 + a_2 s^3 & s\epsilon[10,20] \end{cases}$$

La fonction d'actualisation compte ici 12 paramètres. On rajoute des contraintes de régularité sur cette fonction qui garantissent la continuité de la fonction, la continuité de la dérivée première et de la dérivée seconde de cette fonction aux points de raccord 5 et 10.

Pour i=0, 1 et 2

$$\begin{cases} B_0^{(i)}(5) = B_5^{(i)}(5) \\ B_{10}^{(i)}(5) = B_5^{(i)}(10) \end{cases}$$

Et la contrainte qui porte sur le facteur d'actualisation $B_0(0) = 1$: En utilisant l'ensemble de ces 7 contraintes, le nombre de paramètres à estimer tombe à 5 :

$$B(0,s) = \begin{cases} B_0(s) = d_0 + c_0 s + b_0 s^2 + a_0 s^3 \\ B_5(s) = 1 + c_0 s + b_0 s^2 + a_0[s^3 - (s-5)^3] + a_1(s-5)^3 \quad s\epsilon[0,5] \\ B_5(s) = 1 + c_0 s + b_0 s^2 + a_0[s^3 - (s-5)^5] + a_1[(s-5)^3 - (s-10)^5] \\ + a_2(s-10)\ s\epsilon[0,5] \end{cases}$$

2)Méthode de Nelson Siegel et Nelson Siegel Svensson

a) Méthode Nelson Siegel

Cette méthode est très utilisée par les banques centrales Nelson Siegel présenta en 1987, un modèle original en formulant une expression mathématique décrivant la dynamique des taux à

terme instantanés $f_t(m)$. Cette expression est solution d'une équation différentielle du second ordre dans le cadre d'une racine double :

$$f_t(m) = \beta_1 + \beta_2 \exp\left(\frac{\tau}{\lambda}\right) + \beta_3 \left(\frac{\tau}{\lambda}\right) exp\left(-\frac{\tau}{\lambda}\right)$$

Nous déduisons une expression fermée du taux zéro à la maturité m :

$$y(\tau) = \beta_1 + \beta_2 \left[\frac{1-EXP\left(-\frac{\tau}{\lambda}\right)}{\frac{\tau}{\lambda}}\right] + \beta_3 \left[\frac{1-exp\left(-\frac{\tau}{\lambda}\right)}{\frac{\tau}{\lambda}} - exp\left(-\frac{\tau}{\lambda}\right)\right]$$

Ainsi, pour une section donnée des rendements, nous devons estimer quatre paramètres: $\boldsymbol{\beta_1, \beta_2}$, $\boldsymbol{\beta_3}$ **et** $\boldsymbol{\lambda}$. pour les taux observés, avec m différentes maturités $\boldsymbol{\tau_1, \tau_2, \dots, \tau_m}$.

Où :

- $\boldsymbol{\beta_1}$ **: Facteur de niveau** ; ilest indépendant du temps de maturité, et donc il est souvent interprété comme le niveau de rendement à long terme
- $\boldsymbol{\beta_2}$: **Facteur de pente** ou de rotation , il est pondéré par une fonction du temps jusqu'à maturité . Cette fonction est l'unité pour t = 0 et décroît exponentiellement vers zéro quand t croit, d'où l'influence de β_2 est seulement ressenti à l'extrémité courte de la courbe.
- $\boldsymbol{\beta_3}$**Facteur de courbure :** il est également pondéré par une fonction de t, mais cette fonction est nulle pour t = 0, augmente, puis redescend à zéro à mesure que t augmente en ajoutant ainsi une bosse à la courbe.
- $\boldsymbol{\lambda}$ affecte les fonctions de poids pour β_2 et β_3; plus particulièrement, il détermine la position de la bosse.

Il existe une stratégie simple pour obtenir les paramètres pour ce modèle: fixer λ, puis estimer les valeurs de β par MCO. Nous ne supposons pas que les paramètres du modèle sont constants, mais ils peuvent changer avec le temps. Pour simplifier la notation, nous n'ajouterons pas d'indices pour la période.

Les paramètres du modèle ont donc, dans une certaine mesure, une interprétation directe, ce qui entraîne ***les contraintes***:

$$\boldsymbol{\beta_1 > 0}$$

$$\boldsymbol{\beta_1 + \beta_2 > 0}$$

$$\boldsymbol{\lambda > 0}$$

b) Nelson Siegel Svensson

Le modèle Nelson-Siegel-Svensson (NSS ultérieurement) ajoute un second terme de bosse au modèle NS(on notera le facteur de **, $\boldsymbol{\beta_4}$**. Soit encore y (t) le taux zéro pour la maturité t, alors:

$$y(\tau) = \beta_1 + \beta_2 \left[\frac{1-EXP(-\frac{\tau}{\lambda_1})}{\frac{\tau}{\lambda_1}}\right] + \beta_3 \left[\frac{1-EXP(-\frac{\tau}{1})}{\frac{\tau}{\lambda_1}} - \exp(-\frac{\tau}{1})\right] + \beta_4 \left[\frac{1-EXP(-\frac{\tau}{\lambda_2})}{\frac{\tau}{\lambda_2}} - exp(-\frac{\tau}{\lambda_2})\right]$$

Ainsi on aura à estimer six paramètres $\boldsymbol{\beta_1, \beta_2, \beta_3, \beta_4, \lambda_1}$ **et** $\boldsymbol{\lambda_2}$**.**

Pour le deuxième modèle (Nelson Siegel Svensson) nous avons en plus$\lambda_{1,2}$> 0. Comme pour NS, nous pourrions fixer les valeurs de λ c'est-à-dire, utiliser une grille de différentes valeurs - , puis exécutez un algorithme des moindres carrés pour obtenir des estimations de paramètres, même si nous ne pouvons pas gérer les contraintes d'inégalité avec un solveur standard des moindres carrés. Plus généralement, les paramètres des modèles peuvent être estimés en minimisant la différence entre les taux du modèle y, et les taux observés y_M(taux du marché). Un problème d'optimisation peut être énoncé comme suit:

$$Min_{\beta,\lambda}\left[\sum (y - y_M)^2\right]$$

- Il existe de nombreuses variantes de cette fonction objective : nous pourrions utiliser des valeurs absolues à la place de carrés, ou une fonction d'échelle plus robuste. De même, nous pourrions utiliser les prix des obligations à la place des taux, et ainsi de suite.
- Il est à noter que l'outil que nous utiliserons dans ce mémoire pour l'estimation des paramètres est le solveur d'EXCEL (automatisé aussi sous VBA EXCEL pour nos calculs)
- Nous avons remarqué que la plupart des pays développés utilisent le modèle NS ou NSS pour la modélisation de leur courbe de taux, ci-dessous (voir Figure) une liste de ces pays et les méthodes de reconstitution utilisées :

Country	Yield curve fitting model
Belgium	N-SModel, Svensson Model
Canada	Svensson Model
Finland	N-S Model
France	N-S Model, Svensson Model
German	Svensson Model
Italy	N-S Model
Japan	Maximum smoothing spline，i.e. FNZ Model
Norway	Svensson Model
Spain	N-S Model (by 1995), Svensson Model
Sweden	Svensson Model
England	Svensson Model
U.S.A.	Maximum smoothing spline，i.e. FNZ Model

Tableau1 : liste des pays et leur méthode de modélisation de leur courbe des taux.

Source : BIS (Bank For International Settlments)

2.4Elaboration d'un outil métrique sur EXCEL & VBA Excel

2.4.1 La courbe de taux de référence pour une date donnée : Exemple du 01/03/2018

La disposition d'une base contenant L'ensemble des obligations en circulation sur le marché obligataire marocain est nécessaire pour constituer les premiers points de notre courbe de taux pour ce jour-là.

Un tel jeu de données n'était pas disponible au département de prévoyance où nous sommes affectées. D'une part, parce qu'ils ne reçoivent que les inputs (taux d'actualisation i.e. taux ZC pour des maturités bien définis)

La constitution des premiers points usant de la méthode du Bootstrapping ne serait donc pas possible vu l'indisponibilité d'outils afin de l'initier par nous-même. Cependant, nous avons automatisé le calcul des taux de référence pour toutes les maturités pleines entre 3 mois et 30 ans sur la base des données de BAM, en utilisant les méthodes d'interpolation linéaire et cubique ; puis avec les méthodes indirectes à titre d'illustration.

a) Importation des données journaliers de BAM

INSERER DONNEES BAM ICI			
En millions de dirhams			
Date d'Ã©chÃ©ance	Maturité en années	Taux moyen pondÃ©rÃ©	Date de la valeur
16/04/2018	0,133333333	2,25%	27/02/2018
14/05/2018	0,211111111	2,26%	27/02/2018
22/08/2018	0,483333333	2,28%	01/03/2018
17/09/2018	0,558333333	2,29%	28/02/2018
05/12/2018	0,777777778	2,32%	28/02/2018
20/05/2019	1,236111111	2,39%	01/03/2018
16/09/2019	1,566666667	2,42%	01/03/2018
21/10/2019	1,663888889	2,41%	01/03/2018
20/04/2020	2,169444444	2,47%	01/03/2018
18/05/2020	2,247222222	2,48%	01/03/2018
17/10/2022	4,7	2,74%	28/02/2018
19/06/2023	5,380555556	2,81%	28/02/2018
02/06/2025	7,363888889	3,00%	28/02/2018
19/04/2027	9,269444444	3,18%	28/02/2018
14/06/2027	9,425	3,21%	28/02/2018
16/04/2029	11,28888889	3,39%	01/03/2018
06/08/2029	11,6	3,49%	01/03/2018
18/07/2031	13,575	3,59%	01/03/2018
04/02/2036	18,19166667	3,91%	01/03/2018
19/02/2046	28,38055556	4,36%	01/03/2018

Tableau 2 : Tableau de calcul des taux de référence des bons de trésor du 01/03/2018.Source : BAM

La maturité en année est calculée par nos soins tel que :

Maturité en année = Date échéance – Date valeur

b) Interface de calcul du taux de référence avec les méthodes directes & indirectes

Calculer taux par interpolation linéaire

Calculer taux par interpolation cubique

Calculer taux par spline polynomial

Calculer taux par NS-Svensson

Maturité pleine	taux_interp_li	taux_cubique	taux_splin	taux_NS_Svens
0,25	0,022618571	0,022628105	0,0226228	0,022991114
0,5	0,022814444	0,022874147	0,022812	0,022863916
1	0,023525455	0,023371535	0,023522	0,023016694
2	0,024465653	0,02438481	0,0243453	0,023630722
3	0,025597961	0,025417556	0,0259179	0,024410136
4	0,026657984	0,026463541	0,0268116	0,025284651
5	0,027726204	0,027516535	0,0277107	0,026216983
6	0,028720924	0,028570308	0,0288344	0,027180713
7	0,029658739	0,029618629	0,0297665	0,028156064
8	0,030584184	0,030655267	0,0302903	0,029128116
9	0,031502551	0,031673992	0,0312275	0,030085699
10	0,032652206	0,032668572	0,0327321	0,031020587
11	0,033612563	0,033632778	0,0332703	0,031926883
12	0,035072532	0,034560379	0,0357596	0,032800524
13	0,035578861	0,035445143	0,0361219	0,033638887
14	0,036168267	0,036280841	0,0357868	0,034440476
15	0,036870072	0,037061242	0,0360514	0,035204663
16	0,037571877	0,037780115	0,0367965	0,035931485
17	0,038273682	0,038431229	0,0378174	0,036621482
18	0,038975487	0,039008354	0,0389096	0,03727556
19	0,039468593	0,039505253	0,0398878	0,037894896
20	0,039912214	0,039915714	0,0407873	0,038480844
21	0,040355834	0,040233487	0,0413847	0,039034881
22	0,040799455	0,040452349	0,0419369	0,039558547
23	0,041243075	0,040566068	0,0423811	0,040053413
24	0,041686696	0,040568414	0,042734	0,040521047
25	0,042130316	0,040453157	0,0430127	0,040962991
26	0,042573937	0,040214065	0,043234	0,041380751
27	0,043017557	0,039844308	0,043415	0,041775775
28	0,043461178	0,039339455	0,0435726	0,042149454
29	0,043904798	0,038691476	0,0437238	0,042503111
30	0,044348419	0,03789474	0,0438854	0,042838004

Tableau 3 : Calcul du taux de référence TMP pour les maturités pleines 3mois-30ans pour la date du 01/03/2018

Les fonctions utilisées sur VBA Excel ainsi que sur EXCEL pour automatiser les calculs et obtenir les résultats sont en ANNEXE . Il suffirait d'importer le tableau du site de BAM et lancer le calcul.

c) Analyse graphique/comparative de la courbe des taux

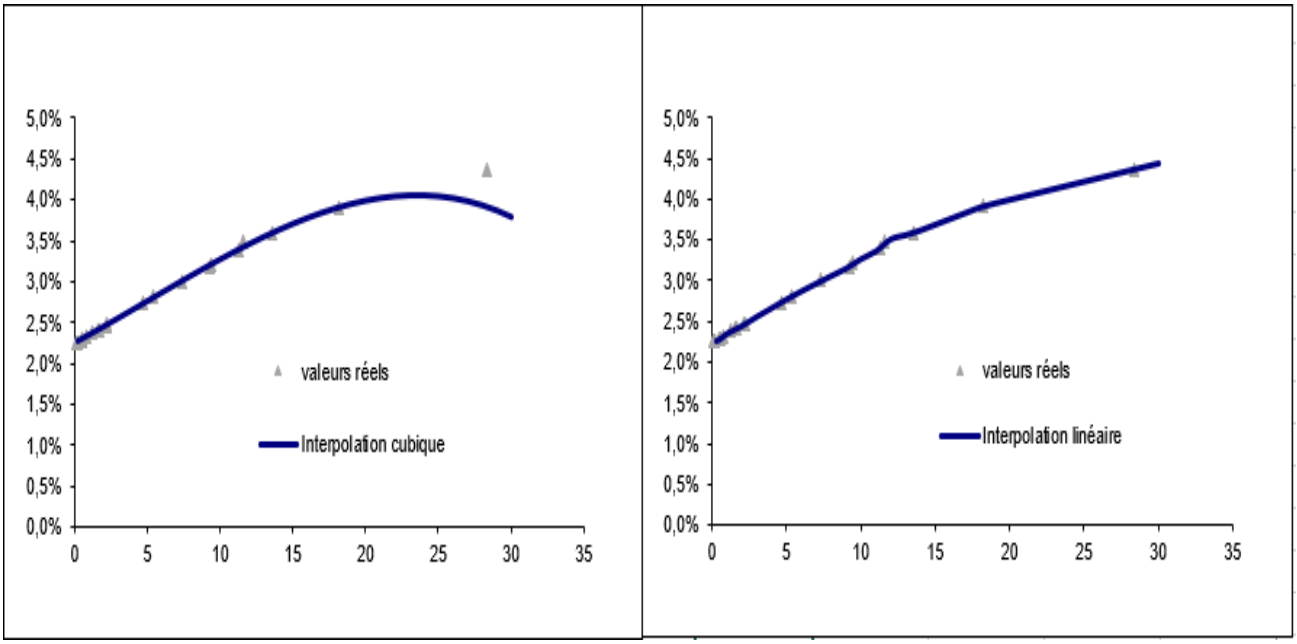


Figure 5 : Sortie graphique du lissage de la courbe de taux en utilisant l'interpolation linéaire et cubique avec pour la date du 1/03/2018

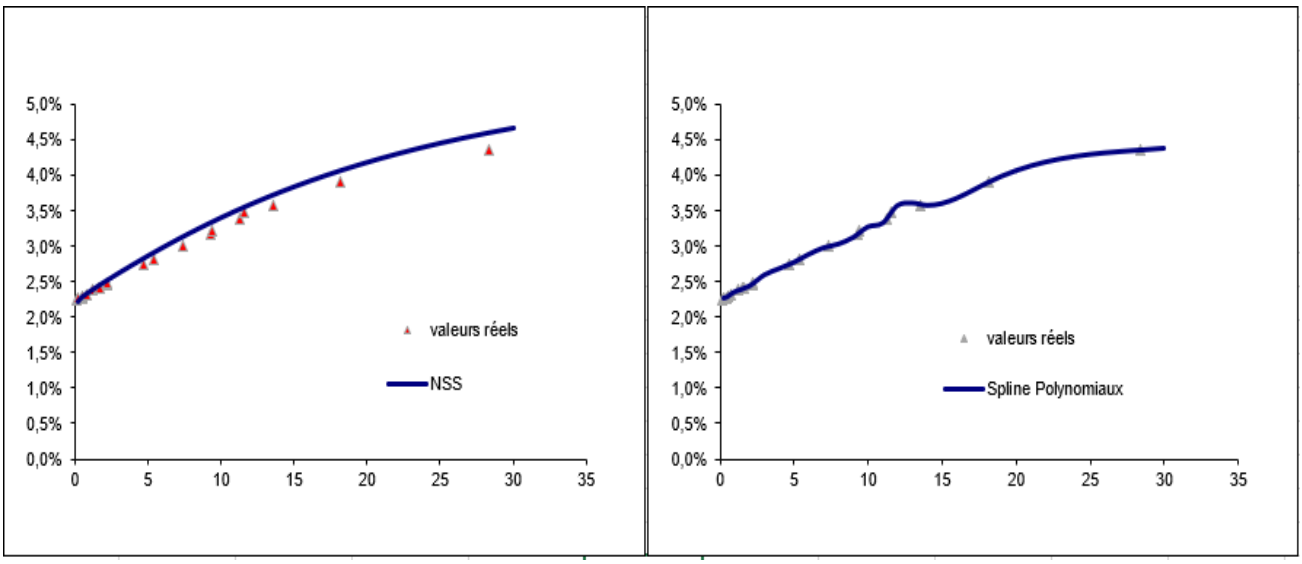


Figure 6 : Sortie graphique du lissage de la courbe de taux avec les méthodes indirectes pour la date du 01/03/2018

- Les résultats de notre implémentation sont représentés dans les figures ci-dessus. Avec les diverses méthodes, nous avons réussi à modéliser la courbe de taux marocaine.
- L'interpolation linéaire reste la plus utilisée pour les calculs des taux de rendement des obligations émises au marché obligataire marocain, vu la simplicité de son utilisation.
- La méthode de reconstitution indirecte Nelson Siegel reste cependant largement utilisé dans le monde, et surtout par les grandes puissances économiques. Un rapport du FMI semble travailler sur la matière et a démontré l'efficacité de cette méthode à approcher au mieux les données de marché marocaines.

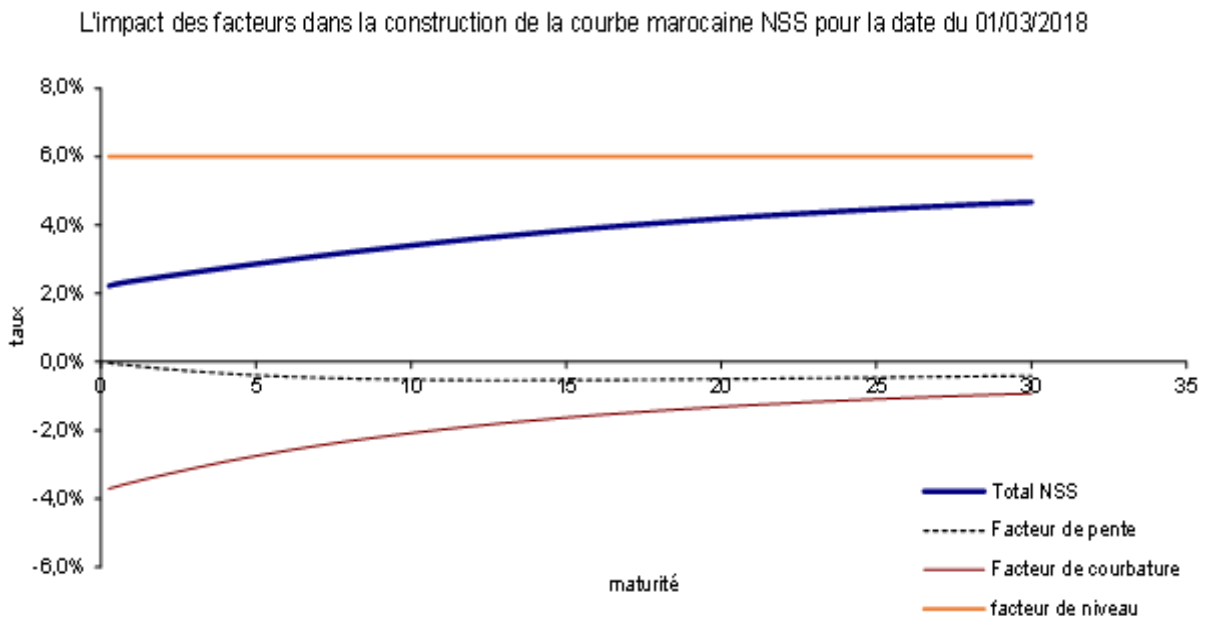


Figure 7 : Impact des facteurs $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}$ sur la formation de la courbe des taux pour la date du 01/03/2018

2.4.2 La courbe des taux historique

Nous sommes sortis dans la section précédente avec un modèle de courbe de taux relatif à une journée de calcul daté du 01/03/2018. Il serait ainsi intéressant d'introduire la courbe des taux pour toutes les dates précédentes afin de construire une courbe de taux historique. Cette courbe sera donc tridimensionnelle, avec la coordonnée « date » comme troisième dimension.

Nous avons déniché une courbe de taux historique propre au Maroc dans un rapport du FMI et dont l'allure de l'évolution des taux pour une date donnée semble bien pareille aux résultats trouvés dans la section précédente.

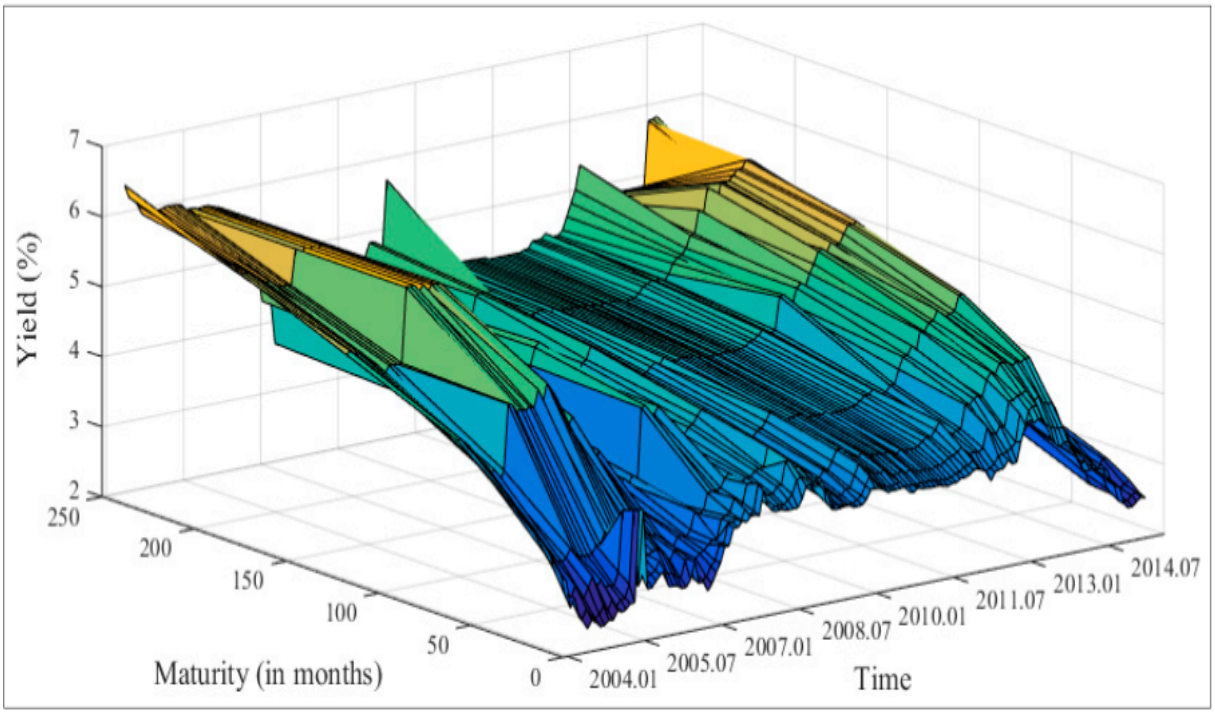


Figure 8 : Evolution de la courbe des taux marocaine du Janvier 2004 au Juillet 2014. Source : FMI

2.4.3 La courbe de taux forward

Nous rappelons encore une fois que le taux forward (ou taux forward zéro-coupon) $F(t,x,y-x)$, déterminé en t, démarrant en x et d'échéance y, est défini par :

$$F(t,x,y-x) = [\frac{\big(1+R(t,y)\big)^{y-t}}{\big(1+R(t,x)\big)^{x-t}}]^{\frac{1}{y-x}} - 1$$

Pour un emprunt avec remboursement des intérêts et du capital à l'échéance, $F(t,x,y-x)$ est le taux d'intérêt auquel on peut signer un contrat aujourd'hui, avec un démarrage en x et l 'échéance en y

Nous avons créé en outre de la courbe de taux, la courbe de taux forward. Son calcul est donc simple et se fait à partir des taux retrouvés dans la section précédente. La figure suivante représente l'interface que nous avons implémenté pour calculer le taux forward pour n'importe quelle date de démarrage T et date d'échéance de T+ m avec m ∈[0,25, 30] ans, à partir d'une date donné (Settlement date)

N.B : Nous pouvons appliquer la formule ci-dessus sous l'hypothèse de l'absence d'opportunité d'arbitrage (A0A**)

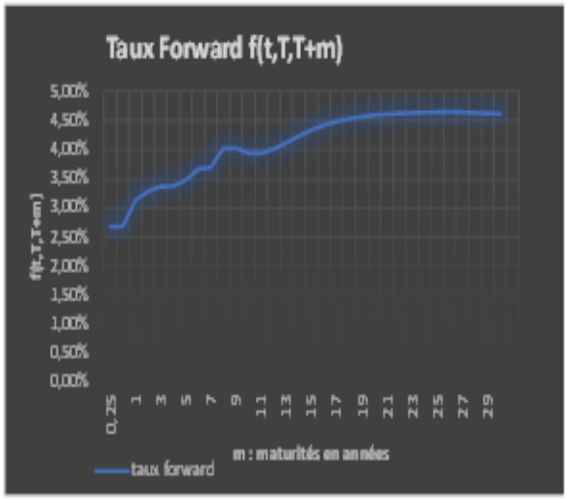
AOA** : Soit un portefeuille quelconque de titres, de coût nul ou négatif. On ditqu'il y a absence d'opportunité d'arbitrage si et seulement s'il n'est pas possible de réaliser à terme un

profit positif presque surement. Autrement dit, si θ représente les quantités d'actifs en portefeuille et Pi leur prix en date i, alors :

$$\theta\, P_0 \leq 0 \Rightarrow P\,(\,\theta P_t > 0\,) = 0$$

m : maturité pleine (années)	(r(t,T) (NSiegel)	P(t,T) : prix ZC	forward matuirté m dans T ans
0,25	0,022991114	0,994268708	2,68%
0,5	0,022863916	0,988633139	2,68%
1	0,023016694	0,97724617	3,13%
2	0,023630722	0,95383799	3,29%
3	0,024410136	0,929386666	3,37%
4	0,025284651	0,903807752	3,38%
5	0,026216983	0,877143288	3,48%
6	0,027180713	0,84951959	3,67%
7	0,028156064	0,821114719	3,70%
8	0,029128116	0,792133826	4,03%
9	0,030085699	0,762790937	4,03%
10	0,031020587	0,733295979	3,94%
11	0,031926883	0,703845989	3,94%
12	0,032800524	0,674619611	4,01%
13	0,033638887	0,645774094	4,12%
14	0,034440476	0,617444144	4,24%
15	0,035204663	0,589742105	4,34%
16	0,035931485	0,562759023	4,42%
17	0,036621482	0,536566277	4,48%
18	0,03727556	0,511217528	4,53%
19	0,037894896	0,486750786	4,57%
20	0,038480844	0,463190489	4,59%
21	0,039034881	0,440549487	4,61%
22	0,039558547	0,418830894	4,62%
23	0,040053413	0,398029762	4,63%
24	0,040521047	0,37813458	4,64%
25	0,040962991	0,359128583	4,65%
26	0,041380751	0,34099089	4,65%
27	0,041775775	0,323697474	4,64%
28	0,042149454	0,307221989	4,63%
29	0,042503111	0,291536452	4,63%
30	0,042838004	0,276611823	4,62%

Date expiration T (ans)	4
Date échéance T+m (ans)	8
calcul taux forward	*0,033780742*
Settlement date :(t=0)	27/02/2018



$$F(t_1,t_2) = \left[\frac{(1+r_2)^{d_2}}{(1+r_1)^{d_1}}\right]^{\frac{1}{d_2-d_1}} - 1$$

Avec :

- $F(t_1,t_2)$: taux forward entre t_1 et t_2
- r_1 : taux zéro-coupon d'échéance t_1
- r_2 : taux zéro-coupon d'échéance t_2
- d_1 : nombre d'années entre la date initiale et l'échéance t_1
- d_2 : nombre d'années entre la date initiale et l'échéance t_2

Figure 9 : Interface de calcul du taux forward pour la date du 01/03/2018

2.4.4 La courbe des taux augmentée de la prime de risque sur obligations.

1) Problématique

Il est à rappeler que typiquement, le taux sans risque du marché est donné par la courbe des taux des bons de trésor notées BBB- au Maroc, construite par interpolation linéaire sur toutes les différentes maturités.

La prime de risque pour une maturité et une obligation émise bien précise sera déduite en calculant la différence entre le taux de rendement de cette obligation et le taux sans risque correspondant à sa maturité. Mais pourrions-nous nous prononcer sur une prime de risque d'une obligation quelconque du marché des obligations marocain étant donnée sa maturité ? Cette généralisation est-t-elle possible dans un premier lieu ? Etant donné les circonstances du marché marocain ?

Il paraît donc intéressant d'implémenter un modèle qui prévoit la prime de risque suivant les caractéristiques liée à chaque obligation ; Et à la base de cette modélisation, construire un outil d'aide à la décision se prononçant sur quelles obligations l'OCP pourra y investir pour couvrir les engagements de couverture médicale que nous entamerons dans la deuxième partie de ce mémoire.

A première vue, le travail paraît implémentable. Mais le jeu de données que nous avons réussi à collecter concernant le marché des obligations marocain n'était pas assez large pour considérer un travail de modélisation statistique (Au total, 39 obligations des divers secteurs de maturités majoritairement de 5 ,7 et 10 ans ou perpétuelle carrément). Même si nous effectué cette modélisation, les résultats restent discutables.

Par contre, en analysant la structure du marché des obligations marocains à base des informations collectées, nous nous rendons compte à quel point le marché est dominé par des émissions des banques. La performance des banques marocaines sont bonnes mais surtout proches (coté rating et primes de risque offertes). Une analyse de la volatilité de la part de la prime de risque dans le taux de rendement de l'obligation pour chaque maturité nous a paru plus efficace quoique simple.

Nous nous sommes exclusivement intéressées au rendement fixe offert par ces obligations (Un travail similaire pourra se faire en calculant les primes de risque en cas de taux variables, mais nous avons estimé que le travail sur le taux fixe est suffisant). Dans ce qui suit les résultats de notre travail sur la prime de risque sur obligations à rendement fixe.

2) Jeu de données utilisé

Notons :

Spread = Taux du rendement fixe – taux sans risque correspondant à la date d'émission et la maturité de l'obligation

Les données sont collectées à partir du site de MaroClear, L'AMMC et de la bourse des valeurs de Casablanca.

libelle	date emission	date échéance	Emetteru	secteur	notation	montant émi
28DEC2017 3.69% 7A 100K ATW	28/12/2017	26/12/2024	ATW	Banques	BB+	150000
29JUN17 3,63% 7A 100K ATW	29-juin-17	27/06/2024	ATW	Banques	BB+	300000
23DEC16 3.44% 7A 100K ATW	23/12/2016	22/12/2023	ATW	Banques	BB+	50000

Maturité	taux fixe	Taux sans	Flux	Duration	Spread	Spread/TF*
7	3,69%	0,0301081	5535	6,3965998	0,68%	18,41%
7	3,63%	0,0302041	10890	6,2464088	0,61%	16,79%
7	3,44%	0,0287438	1720	6,2541992	0,57%	16,44%

Figure 10 : Echantillon de données du marché obligataire collecté par nos soins.

*TF : Taux fixe

3) 1ère approche : Modèle de régression

Retenons le modèle de régression linéaire suivant :

$$Spread = \alpha + \beta_1 * TSR + \beta_2 * Maturité + \varepsilon$$

La régression implémenté avec le logiciel stata a donné les résultats suivants (les sorties stata sont en ANNEXE) :

Test	Résultat	Conclusion

R^2 ajusté	R^2 ajusté = 37,74%	Le pouvoir explicatif du modèle est de 37,74%
Fisher	p-value = 0.0001	Le modèle est globalement significatif
Student(constante)	p-value = 0.000	La constante est significative
Student(variable)	p-value= 0.000	Le coefficient de la variable est significatif
Moyenne des résidus	Root MSE = 0.00742	Les résidus sont de moyenne presque nulle

Tableau 4 : Tests de validation du modèle de régression du Spread par le TSR et la Maturité

Ainsi, le modèle obtenu est le suivant :

Spread = 0,0305775 -0,5980101*TSR + 0,0002205* Maturité

Autrement dit, une augmentation du taux sans risque de 1% fait baisser le spread de 0,5980101, et un prolongement de la maturité d'un an fait augmenter le spread de 0,0002205.

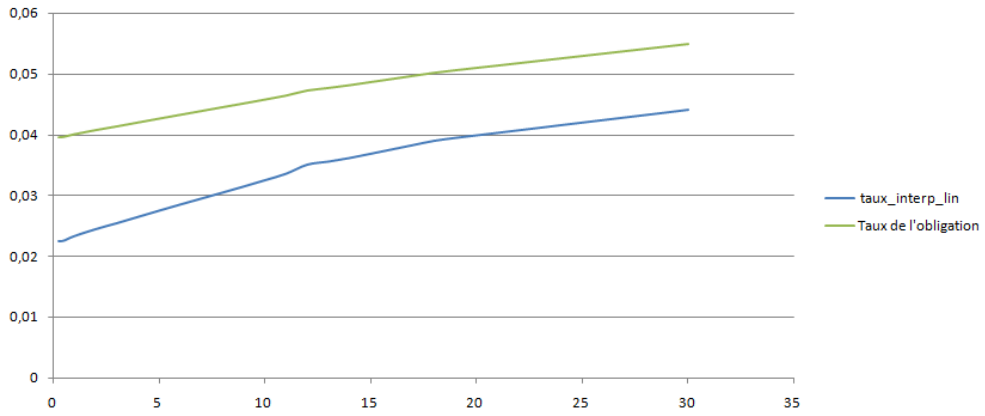


Figure 11 : Courbe de taux augmentée de la prime de risque calculée à base du modèle de régression.

4) 2ème approche : Analyse de la volatilité de la part de la prime du risque dans le taux de rendement.

Comme nous avons cité dans la problématique, une deuxième approche pour retrouver la prime de risque est de faire une analyse de la volatilité de la part de la prime de risque dans le taux de rendement de l'obligation pour chaque maturité. Une méthode qui nous a paru simple ,efficace et qui répond à nos besoins de calcul, et ce à partir de la base que nous avons pu recueillir.

Ainsi , nous calculerons la quantité $\frac{Spread}{Taux\ Fixe}$ pour définir la part de la prime de risque dans le taux d'une maturité donnée, puis analyser l'écart type de celle-ci . Le résultat obtenu est dans le tableau suivant :

Maturité	Spread moyen	Spread/TF moyen	rendement espéré du marché	Ecart-type des (Spread/TF)
5	0,009675333	0,199880733	0,0461	0,082292035
7	0,008357113	0,199014837	0,041967	0,064516914
10	0,010888611	0,222707271	0,049617	0,052358448
15	0,009328892	0,167304178	0,05584	0,035510969
perpetuel	0,018470439	0,303968466	0,05655	0,153151412

Tableau 5 : La volatilité de la part de la prime de risque dans le taux de rendement, calculée par maturité

L'écart type de la quantité $\frac{Spread}{Taux\ Fixe}$ est très petit pour les divers maturités(de l'ordre de 0,08 pour les taux de rendement des obligations de maturité 5 ans , 0,06 pour ceux de 7 ans et 0,03 pour ceux de 10 ans).

A la base de cette analyse, nous déciderons d'ajouter le spread moyen (2ème colonne du tableau) au taux de rendement obtenu à partir de la courbe de taux. Ainsi par exemple si :

$$\text{Maturité} = 5\ \text{ans} \rightarrow Prime\ de\ risque = 0{,}9\%\ (99\ \text{points de base})$$

$$\text{Maturité} = 7\ \text{ans} \rightarrow Prime\ de\ risque = 0{,}8\%\ (83\ points\ de\ base\ )$$

$$\text{Maturité} \ = \ 10 \text{ ans} \rightarrow \ \textit{Prime de risque} = 1\% \ (100 \ \textit{points de base} \ )$$

Etc.

Nous utiliserons l'interpolation linéaire pour constituer le reste des spreads correspondant à leurs maturités. L'outil Excel utilisé précédemment nous aidera pour faire cela.

Le résultat final est exhibé ci-dessous :

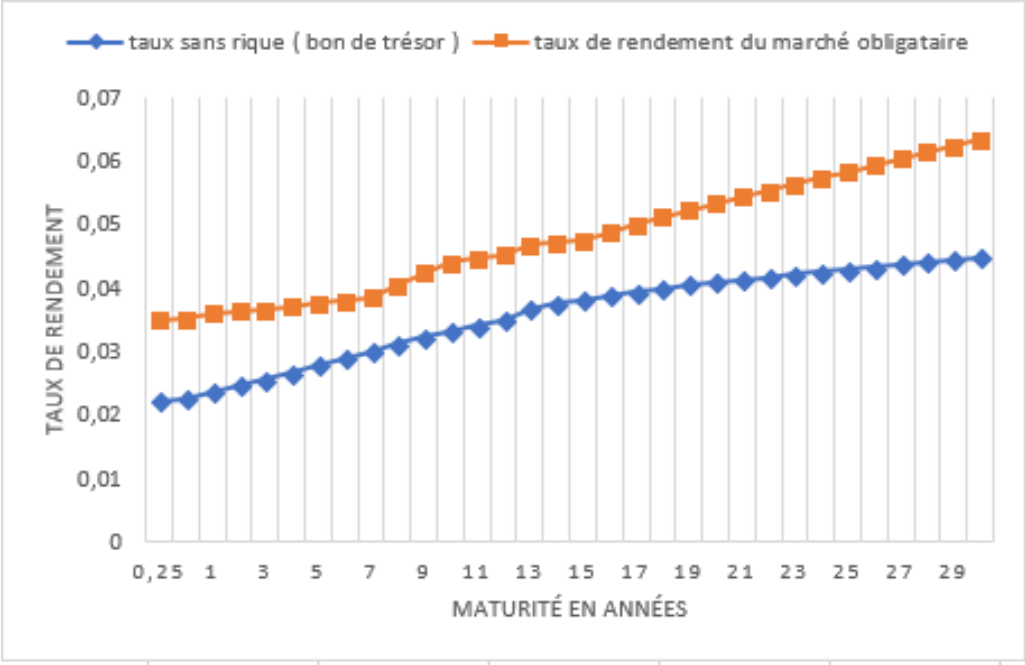


Figure 12 : Comparaison de la courbe du taux sans risque et la courbe de taux augmentée de la prime de risque calculée.

Chapitre 3 : Le taux d'inflation médicale

Notre intérêt se concentre ici sur la modélisation d'une deuxième hypothèse actuarielle : l'inflation médicale.

L'IAS 19 s'attarde ainsi sur les spécificités de ce paramètre dans les articles 87 à 91.Son estimation doit tenir compte de :

- L'évolution du niveau des prestations d'assurance santé offertes par l'Assurance Maladie etl'État, le niveau des prestations des régimes de frais de santé en dépendant

- De paramètres globaux tels que l'inflation et l'évolution spécifique aux coûts médicaux

Nous allons essayer de déterminer ce deuxième paramètre aussi important et qui a un très fort impact sur la valeur de l'engagement

Comme nous n'avions pas eu accès aux données relatives aux garanties (types de soins) , nous n'aurons pas l'opportunité de déterminer l'inflation médicale relative aux couvertures de soins propre à l'OCP. Nous nous sommes tournées donc vers les statistiques nationales de l'inflation médicale.

3.1 La consommation de soins et de biens médicaux

Il est à savoir que la consommation de soins varie selon :

- Le revenu des individus
- Le niveau d'études
- Les catégories socioprofessionnelles.

La consommation de soins et de biens médicaux se décompose en :

1- Soins hospitaliers (public et privé)
2- Soins ambulatoires : Médecine, chirurgie (dont dentaire) Auxiliaires médicaux (infirmiers, masseurs kinésithérapeutes, orthophonistes, orthoptistes, …) Analyses biologiques
3- Transports sanitaires
4- Biens médicaux (médicaments, optique, prothèses, petit matériel et pansements)

Nous avons cherché le taux d'inflation selon le type de soins. Ci-dessous l'évolution de l'IPC santé par section entre 2008 et 2017 retrouvé sur le site de l'HCP.

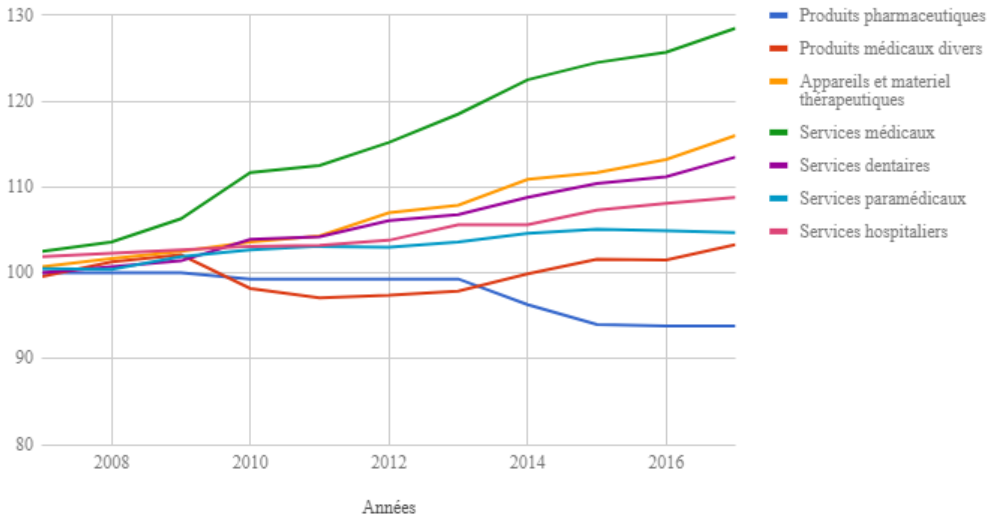


Figure 13 : Evolution de l'IPC annuel par sections pour la division santé (Base 100 : 2006).

Source : HCP

3.2 Détermination du taux d'inflation médical global

Après avoir analysé l'évolution de l'IPC santé par section (types de soins) et dégager l'inflation liée, la deuxième étape est de définir le poids de chacun de ces types de soins dans l'ensemble des prestations fournies par les organismes d'assurance maladie (notamment la CNOPS vers laquelle l'OCP va céder complètement la gestion de son personnel d'ici 2021). Une telle information a pu être trouvé dans les rapport de santé marocaines

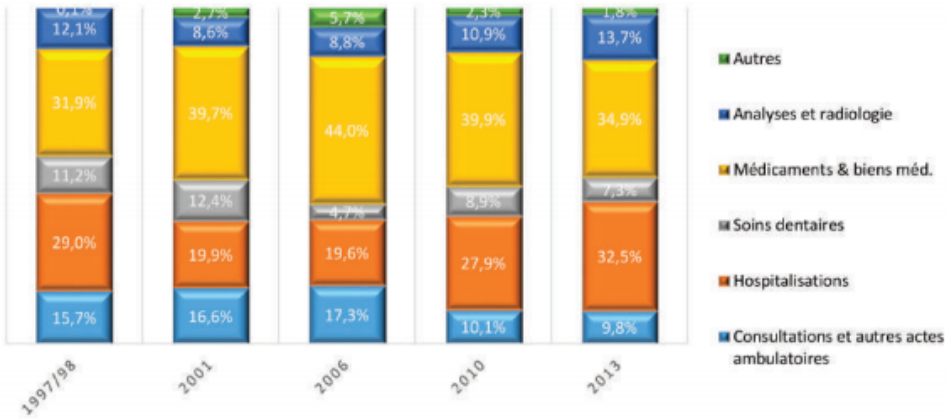


Figure14 : Evolution des prestations des organismes d'assurance maladie (remboursements et tiers payant), 1997/98-2013. Source : Comptes nationaux de santé 2013

Malheureusement, les données que nous avons pu collecter n'étaient pas récentes (les derniers chiffre relèvent de 2013), nous allons pourtant les utiliser, puisque c'est les derniers chiffres officielles publiés par le ministère de santé !

L'Estimation de l'inflation se fera de manière simple : le calcul d'une moyenne de l'inflation par section (type de soins), pondérée par le poids de prestations accordé à chaque type de soin.

Le résultat de notre travail est comme suit :

Inflation médicale Année	Produits pharmaceutiques +Appareil et matériel	Produits médicaux divers	Service médicaux et paramédicaux	Services dentaires	Services hospitaliers	Inflation médicale globale
poids (Rapport santé 2013)	34,90%	15,50%	9,80%	7,30%	32,50%	100,00%
2008	0,49%	1,69%	0,48%	0,70%	0,39%	0,66%
2009	0,39%	0,79%	2,03%	0,69%	0,39%	0,63%
2010	0,18%	-3,89%	2,87%	2,44%	0,39%	0,05%
2011	0,34%	-1,13%	0,55%	0,29%	0,10%	0,05%
2012	1,28%	0,31%	1,14%	1,81%	0,58%	0,93%
2013	0,42%	0,51%	1,70%	0,66%	1,72%	1,00%
2014	-0,16%	2,02%	2,14%	1,86%	0,00%	0,60%
2015	-0,85%	1,69%	1,05%	1,46%	1,60%	0,69%
2016	0,56%	-0,10%	0,38%	0,72%	0,74%	0,51%
2017	1,22%	1,76%	1,01%	2,05%	0,65%	1,16%

Tableau 6: Tableau de calcul de l'inflation médicale

Pour l'année 2017, nous constatons une inflation médicale globale de 1,16%. Vu la simplicité de ce modèle et l'insuffisance des données pour pouvoir mener une étude statistique prévisionnelle plus pointue, nous supposerons ce taux fixe par la suite. Il sera utilisé dans le calcul des flux futurs pour les engagements médicaux de la population OCP.

Conclusion

Dans cette 2ème partie nous avons essayer de comprendre et de déterminer quelques hypothèses financières dont nous aurons besoin vers la 3ème partie de ce mémoire pour évaluer l'engagement médical. Un autre paramètre important se doit d'être modélisé qui est la consommation médicale. Nous verrons dans la prochaine partie les approches et les méthodes de tarification que nous userons pour définir ce paramètre.

Partie 3 : Modélisation de la consommation médicale & calcul de l'engagement médical de l'OCP

Chapitre 1 : Modélisation de la consommation médicale

Afin de pouvoir calculer l'engagement médical, une modélisation de la consommation en soins de santé est requise. En effet, l'évaluation des engagements doit tenir compte de l'évolution permanente de la situation financière du régime qui dépend elle-même de l'évolution des consommations médicales du régime.

Cette partie aura comme finalité d'effectuer une analyse de la consommation médicale propre à l'OCP et de définir le coût du risque auquel l'OCP devrait faire face en termes de couverture médicale. L'OCP devra donc provisionner pour couvrir les frais des avantages post-emploi liés à la couverture des soins médicaux de la population OCP éligible d'accéder à cette prestation(les retraités, leurs conjoints et leurs enfants)

1.1La méthode de tarification santé

1.1.1 Le tarif technique : dépense médicale estimée

Le risque santé est la somme de plusieurs composantes : consultations médicales, hospitalisation, pharmacie, soins dentaires, optique… Les méthodologies usuelles de tarification sont distinguées entre individuel (statistiques tête par tête) et collectif (pilotage du S/P par entreprise).

Le tarif technique tient compte des spécificités du contrat, des garanties, mais également des caractéristiques de l'entreprise et de chaque assuré : localisation des bénéficiaires, âge moyen des assurés, proportion de conjoints, proportion d'hommes, nombre d'enfants. Il s'agit des facteurs influençant la tarification.

Le calcul de cette prime s'exécute généralement selon deux modèles, le modèle fréquence coût et un modèle dit de barème. Le modèle Fréquence cout étant le plus répondu, nous l'utiliserons donc dans un premier temps dans notre étude de tarification.

Par conséquent, pour la tarification des contrats individuels, la prime pure est calculée selon l'approche :

Dépense estimée = Prime pure = Fréquence ×Coût moyen

Avec :

Fréquence : Fréquence annuelle de survenance d'un sinistre par bénéficiaire, c'est le nombre d'actes total d'une population considérée sur le nombre de bénéficiaires de cette population

Coût moyen : En assurance santé, le coût moyen dépend de la garantie. Il correspond, en fonction de la dispersion des actes et d'une population, au montant total des remboursements complémentaires divisé par le nombre d'actes

- Sous réserve de l'indépendance entre la fréquence et le coût moyen du sinistre, la prestation probable par bénéficiaire est le produit de la fréquence par bénéficiaire par le coût moyen du sinistre.

N.B : pour notre étude, nous n'avons pas pu recevoir les différentes garanties (les prestations par type de soins : hospitalisation , optique, chirurgie..) Par conséquent, il n'y aura pas de distinction des coût selon ces garanties (type de soins) La dépense estimée sera donc unique.

Les facteurs discriminants au remboursement sont pris en compte : régime, nature de l'acte… Les paramètres démographiques discriminent également le tarif : âge, composition familiale , région , catégorie socio-professionnelle, etc.

Il est donc possible de recourir au modèle linéaire généralisé pour concevoir les coefficients d'ajustement du tarif en fonction des variables discriminantes.

1.1.2 Modélisation GLM (General Linear Model)

1) Notions importantes

Les modèles linéaires généralisés (GeneralizedLinearModels, GLM) constituent le cadre de référence pour modéliser l'effet des variables de segmentation sur le tarif.

Dans un modèle, on cherche à expliquer une variable d'intérêt dite variable dépendante ou variable expliquée par les variations d'un ensemble d'autres variables dites explicatives ou

facteurs ou variables indépendantes. C'est cette explicitation qui permet de prédire les valeurs que peut prendre la variable expliquée connaissant celles prises par les facteurs.

Dans la modélisation, on essaie d'ajuster une forme mathématique à la variation, « en moyenne », du phénomène, en ayant recours à des méthodes. C'est dans ce cadre que s'inscrit le Modèle Linéaire Généralisé (GLM) qui est sans doute l'outil le plus général, le plus utile et, par conséquent, le plus utilisé de la panoplie des instruments dévolus à la modélisation.

Avant de définir le GLM, il convient de préciser quelques notations et définitions :

- Soit $Y = (Y_1, Y_2, \dots, Y_i, \dots, Y_n)$ est le vecteur des variables aléatoires, une grandeur à expliquerY_i, i de 1 à n, les valeurs prises par la grandeur y lors de n mesures.
- Soient X= (X_1,X_2 …, X_p) p grandeurs explicatives. La variable générique $X_j$prend n valeursX_{ij} , i de 1 à n.

Le modèle GLM s'écrit sous la forme :

$Y \sim L_{exp}$ et $\mu = E(Y)$

$\eta = g(\mu)$

$\eta = X\beta$

Avec :

- $\mu = (\mu_1, \mu_2, ..., \mu_i, ..., \mu_n)$' est le vecteur des espérances,
- $\eta = (\eta_1, \eta_2, ..., \eta_i, ..., \eta_n)$' est le vecteur des prédicteurs,
- $\beta = (\beta_1, \beta_2, ..., \beta_i, ..., \beta_n)$' est le vecteur des paramètres à déterminer

Dans la forme générale du GLM, la loi de probabilités de la réponse Y_i doit appartenir à ***la famille exponentielle***. En effet, la famille exponentielle est un ensemble de lois dont l'écriture est résumée par une formule globale. Les lois de probabilités, telles que la loi normale, la loi binomiale, la loi de Poisson, la loi Gamma et la loi de Gauss inverse, appartiennent à la famille exponentielle décrite précédemment.

On reconnaît ainsi deux parties dans l'expression du GLM :

➢**Une partie déterministe** : précise quels sont les prédicteurs et exprimée sous forme d'une combinaison linéaire $\beta 0+ \beta 1\ X1+...+ \beta pXp$ (appelée aussi prédicteur linéaire)

➢**Une partie aléatoire** : c'est la distribution de la variable expliquée

Et entre les deux :

➢**La fonction de lien** : spécifie comment l'espérance mathématique de Y notée μ est liée au prédicteur linéaire construit à partir des variables explicatives.

2) Estimation des paramètres

Les estimateurs des paramètres βdu modèle du modèle linéaire généralisé seront obtenus en maximisant la vraisemblance du modèle. Cela revient à maximiser la fonction log-vraisemblance vu que la fonction logarithmique est une fonction croissante.

La fonction log-vraisemblance du modèle pour les réponses $(Y_1, Y_2, \dots, Y_i, \dots, Y_n)$:

$$L\ (\theta\varphi/y) = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i)$$

Où :$E\ (y_i) = \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}$ et $g\ (\mu_i) = x_i^T \beta = \eta_i$

La famille des distributions exponentielles satisfait les conditions de régularité qui permettent d'assurer que le max global du logarithme de la fonction de vraisemblance est unique.

3) Validation du modèle

Deux statistiques sont utiles pour juger de l'adéquation du modèle aux données : la déviance normalisée et la statistique du Khi-deux de Pearson.

Pour mesurer l'adéquation du modèle étudié aux données, on construit tout d'abord un modèle saturé, modèle basé sur la même loi de probabilité et la même fonction de lien, mais contenant autant de variables explicatives indépendantes que de données : ce modèle permet de reconstruire parfaitement les données. On note b_{max}l'estimation du vecteur des paramètres β pour ce modèle.

- **La déviance normalisée :**

$$D* = 2\log\left(\frac{L(bmax\,;y)}{L(b;y)}\right) = 2[l(bmax;y) - l(b;y)]$$

Lorsque le modèle étudié est exact, la déviance normalisée D* suit approximativement une loi de khi-deux à n-k degrés de liberté.

- **La statistique de Khi-deux de Pearson:**

$$x^2 = \sum \frac{(Yi - \hat{\mu}i)^2}{V(\hat{\mu}i)}$$

Comme la déviance normalisée, cette statistique est distribuée approximativement selon e loi de khi-deux à n-k degrés de liberté si le modèle étudié est exact.

1.1.3 Une alternative au GLM : l'apprentissage statistique

Une question naturelle se pose : est-il possible d'améliorer les prévisions de la dépense médicale par une alternative basée sur l'apprentissage statistique ?

Théoriquement, OUI, car cela cela est déjà possible dans le domaine médical, qui dispose de beaucoup plus de données sur les individus. Il faudrait cependant que l'assureur dispose de données suffisamment porteuses d'information sur les assurés en matière de santé.

Mais ceci est conditionnée par la possibilité ou pas de l'Actuaire d'accéder à des données nombreuses et intéressantes sur les bénéficiaires.

Il existe plusieurs méthodes d'apprentissage nouvelles qui ont prouvé leur efficacité. Nous citons la méthode neuronale et les arbres de régression, notamment les forêts aléatoires.

Pour le réseau de neurones par exemple, Un réseau de neurones structuré en couches permet de définir un modèle non linéaire entre des variable explicatives (quantitatives ou non) et une variable à expliquer qui peut être :

Quantitative → Régression non linéaire

Qualitative → Classification

Ce modèle s'inspire directement du mécanisme de pensée du cerveau. L'information se propage de cellule en cellule au travers des différentes couches du réseau. Des exemples sont

présentés au modèle qui ajuste ses coefficients de manière à réduire l'erreur entre les prédictions et les valeurs attendues.

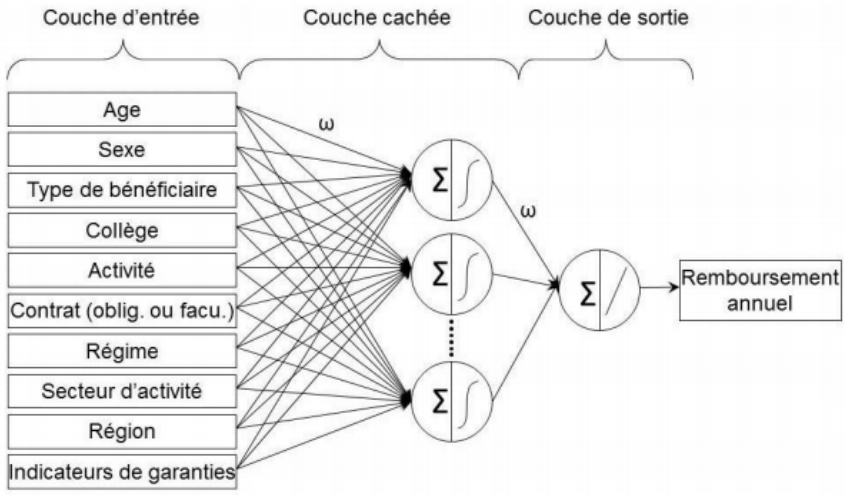


figure 15: Réseau neuronal artificiel

1.1.4 Cadre théorique : Les forêts aléatoires (Randomforests)

1) Les méthodes dites d'ensemble ou d'arbres

Les forêts aléatoires combinent des prédicteurs ou estimateurs de base que sont les arbres, donnant lieu à ce que l'on appelle aujourd'hui les méthodes d'arbres. Plus largement ce sont des méthodes d'ensemble dont le principe général est de construire une collection de prédicteurs, pour ensuite agréger l'ensemble de leurs prédictions.

En régression, agréger les prédictions de q prédicteurs revient par exemple à en faire la moyenne: chaque prédicteur fournit un $\hat{y}l$, et la prédiction finale est alors $\frac{1}{q}\sum_1^q \hat{y}_l$. En classification, l'agrégation consiste par exemple à faire un vote majoritaire parmi les labels des classes fournis par les prédicteurs

Ainsi, au lieu d'essayer d'optimiser une méthode "en un coup", les méthodes d'ensemble génèrent plusieurs règles de prédiction et mettent ensuite en commun leurs différentes réponses. L'heuristique de ces méthodes est qu'en générant beaucoup de prédicteurs, on explore largement l'espace des solutions, et qu'en agrégeant toutes les prédictions, on dispose d'un prédicteur qui rend compte de cette exploration. Ainsi, on s'attend à ce que le prédicteur final soit meilleur que chacun des prédicteurs individuels.

En pratique, il apparaît que “perturber puis agréger” améliore les performances d’une méthode de prédiction donnée. Cette remarque ne vaut que si dans la collection construite, les prédicteurs sont différents les uns des autres. C’est pourquoi ces stratégies sont appliquées à des règles de basequi sont des méthodes dites “instables”. Une méthode est instable si de petites perturbations de l’échantillon d’apprentissage peuvent engendrer de grandes modifications du prédicteur obtenu. Par exemple, les arbres de décision sont instables au contraire des méthodes linéaires(régression ou analyse discriminante) qui sont stables et donc jamais utilisées comme règle de base dans les méthodes d’ensemble.

2) Cas particulier des forêts aléatoires : Le bagging

La méthode du Bagging a été introduite par Breiman (1996). Le mot Bagging est la contraction des mots **B**ootstrap et **Agg**regat**ing**. Étant donné un échantillon d’apprentissage *Ln* et une méthode de prédiction (appelée règle de base, par exemple CART), qui construit sur *Ln* un prédicteur $\hat{h}(. ; Ln)$.

Le principe du Bagging est de tirer indépendamment plusieurs échantillons bootstrap ($L_n^{\Theta_1}$; …… ;$L_n^{\Theta_q}$) d’appliquer la règle de base sur chacun d’eux pour obtenirune collection de prédicteurs ($\hat{h}(., L_n^{\Theta_1})$; ... … ; $\hat{h}(., L_n^{\Theta_q})$) et enfin d’agréger ces prédicteurs de base. La collection de prédicteurs est alors agrégée en faisant simplement une moyenne ou un vote majoritaire.

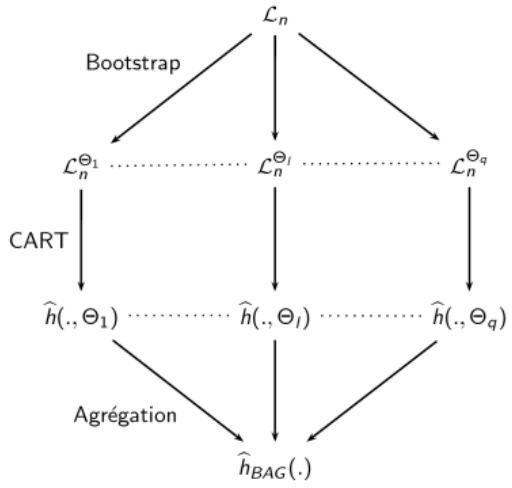


Figure16 : Schéma du Bagging avec pour règle de base un arbre CART

3) Les forêts aléatoires « Random Input »

Random Forests-RI signifie “forêts aléatoires à variables d’entrée aléatoires” (Random Forests with Random Inputs) et le principe de leur construction est tout d’abord de générer plusieurs échantillons bootstrap ($L_n^{\Theta_1}$; …… ;$L_n^{\Theta_q}$) (comme dans le Bagging). Ensuite, surchaque échantillon $L_n^{\Theta_l}$ une variante de CART est appliquée .Plus précisément, un arbre est, ici, construit de la façon suivante : Pour découper un nœud, on tire aléatoirement un nombre *m* de variables, et on cherche la meilleure coupure uniquement suivant les *m* variables sélectionnées.

Le tirage des *m* variables à chaque nœud représente un aléa supplémentaire, par rapport au Bagging. Pour les Random Forests-RI, il y a donc deux sources d’aléas pour générer la collection des prédicteurs individuels : l’aléa dû au bootstrap et l’aléa du choix des variables pour découper chaque nœud d’un arbre. Ainsi, on perturbe à la fois l’échantillon sur lequel on lance la règle de base et le coeur de la construction de la règle de base.

En pratique, les Random Forests-RI (avec le paramètre *m* bien choisi) améliorent les performances du Bagging. L’explication heuristique de ces améliorations est que le fait de rajouter un aléa supplémentaire pour construire les arbres, rend ces derniers encore plus différents les uns des autres, sans pour autant dégrader de façon significative leurs performances individuelles. Le prédicteur agrégé est alors meilleur. C’est d’ailleurs la méthode utilisée dans le Package de R *Random Forest*. Et dans la procédure *HPRFOREST* de SAS aussi.

4) L’erreur OOB

En plus de construire un prédicteur, l’algorithme des Random Forests-RI calcule une estimation de son erreur de généralisation : l’erreur Out-Of-Bag (OOB) où “Out-Of-Bag” signifie “en dehors du bootstrap”. Cette erreur était déjà calculée par l’algorithme du Bagging, d’où la présence du mot “Bag”. Le procédé de calcul de cette erreur est le suivant et il est très astucieusement intriqué dans l’algorithme de construction d’une RF.

Fixons une observation (*Xi; Yi*) de l’échantillon d’apprentissage *Ln* et considérons l’ensemble des arbres construits sur les échantillons bootstrap ne contenant pas cette observation, c’est-à-dire pour lesquels cette observation est “Out-Of-Bag”. Nous agrégeons alors uniquement les prédictions de ces arbres pour fabriquer notre prédiction $\hat{Y}_i$de *Yi*. Après avoir fait cette opération pour toutes les données de *Ln*, nous calculons alors l’erreur commise :l’erreur quadratique moyenne en régression ($\frac{1}{n}\sum_1^n(\hat{Y}_i - Y_i)^2$), et la proportion d’observations mal classées en classification ($\frac{1}{n}\sum_1^n 1_{\hat{Y}_i \neq Y_i}$). Cette quantité est appelée erreurOOB

1.2Mise en œuvre de la tarification

1.2.1 La base de données médicale de l'OCP

1) Les bénéficiaires

On entend par bénéficiaire :

- Un ex-agent OCP mis en retraite titulaire d'une pension dite directe
- Le/les conjoints d'un ex-agent
- Les enfants d'un ex-agent. (Âgés de moins de 26 ans)

Il est à noter que nous ne traiterons que des retraités et leurs conjoints par la suite. L'OCP réserve un traitement ou un calcul spécial pour les enfants. Une fois le calcul de l'engagement fait, le provisionnement médical est multiplié par le pourcentage des enfants bénéficiaires vu que leur poids est assez faible. Nous exclurons donc les enfants de notre étude et nous ajouterons leur part à la fin du calcul de l'engagement (la part des enfants est estimé à 8% de l'engagement total par l'OCP, nous avons retrouvé comme nous verrons par la suite, un poids de 11%pour la part de la consommation des enfants)

2) Les variables disponibles

Les données sont présentées dans un fichier de consommation médicale contenant plusieurs variables: le code ou l'identifiant de l'assuré, le sexe, la date de naissance des bénéficiaires, le type d'assuré (selon qu'il est actif ou retraité), le lien ou le type de bénéficiaire (Assuré, conjoint ou enfant), le sort de dossier (réglé ou non), le type des dossiers (dossiers normaux ou non), le remboursement AMB (Assurance maladie de base) AMC (Assurance maladie complémentaire) et le remboursement total, un code ALD-ALC (renvoyant si l'assuré est atteint ou non d'une maladie de longue durée ou une affection lourde et couteuse) .

La Base contient 1 million 803 milles observations de remboursement de dossiers médicaux des trois dernières années (2015-2017). Notre étude de tarification portera cependant sur les deux années de survenance 2016-2017.

3) Nettoyage et préparation de la base

Les données fournies ont exigé un premier traitement :

- Traitement des données manquantes
- Traitement des données non significatifs.
- Traitement des données mal saisies (le regroupement des villes et des sites par exemple a été avorté par après, vu que les données étaient insuffisantes pour qu'on puisse l'utiliser ultérieurement).

Le nettoyage de la base de données et sa préparation pour l'analyse ont été effectué sur le logiciel SAS.

Cette étape était cruciale avant d'entamer notre analyse. Cependant, elle a consommé une bonne partie de la durée du stage ; entre attente de la réception de la base de données de chez l'entreprise gestionnaire des dossiers médicaux, et des erreurs opérationnelles commise de temps à autre.

Ainsi nous aurons construit une base de données de la forme suivante :

1.2.2 Analyse univariée des variables explicatives

La base de données sur laquelle nous mènerons notre étude, représente la consommation médicale individuelle des pensionnés relative aux années 2016 et 2017. On rappelle les variables qui la constituent :

- Le code du pensionné
- Le numéro de dossier médical de chaque bénéficiaire
- Le sexe du pensionné
- L'âge du pensionné

- La situation familiale du pensionné
- La pension annuelle que le pensionné reçoit
- Le nombre de dossiers par pensionné
- Le montant total remboursé par pensionné (tenant compte de l'inflation médicale).

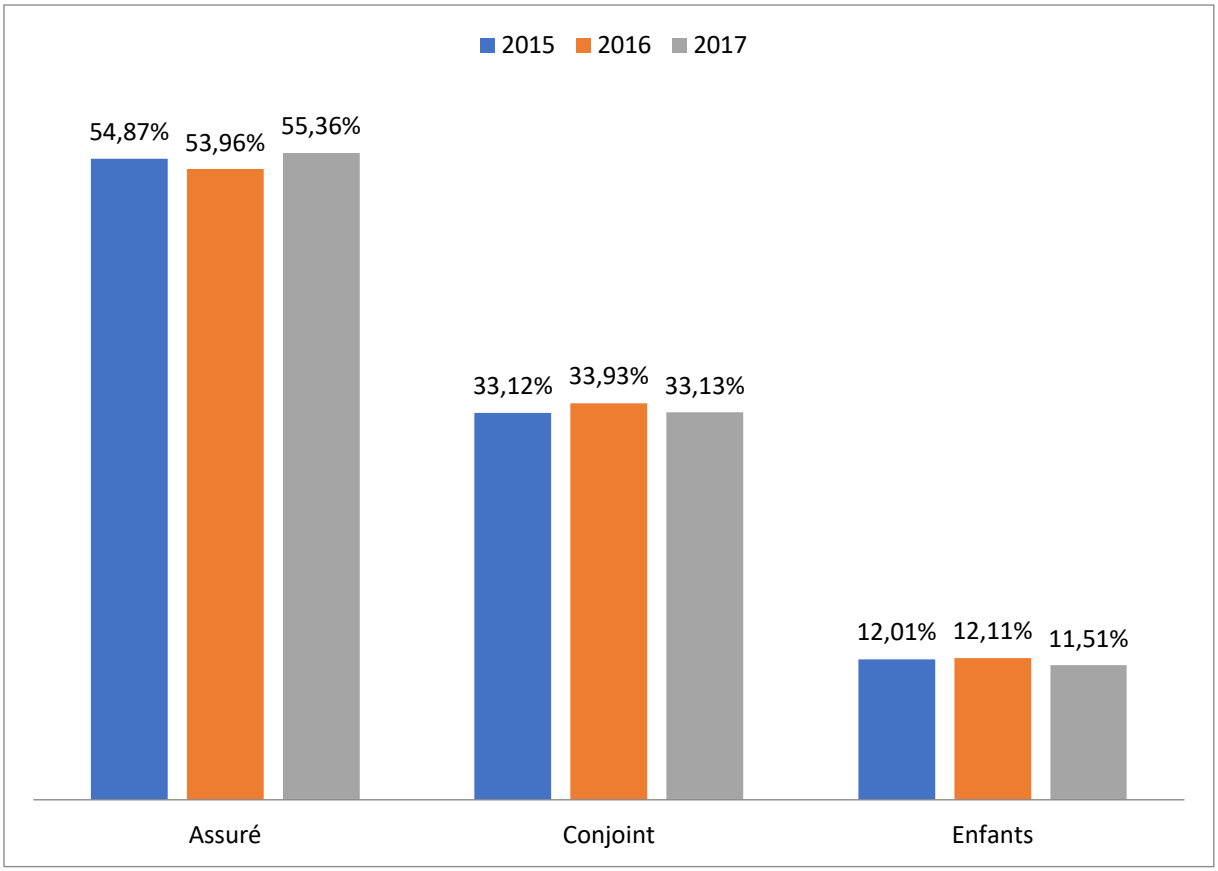


Figure 17: La part de la consommation totale par type de bénéficiaire

Dans notre analyse, nous allons nous contenter d'étudier le comportement de la consommation médicale seulement chez les assurés directs et leurs conjoints. Quant aux enfants, leur dépense médicale est conclue par rapport à la dépense médicale globale(car elle représente une part constante du remboursement total).

En effet, l'analyse par type de bénéficiaire, sur les trois années, montre qu'environ la moitié de montants remboursés bénéficie aux assurés directs, suivis par leurs conjoints avec un pourcentage d'à peu près 33%. Quant aux enfants, ils détiennent une part de l'ordre de 12%.

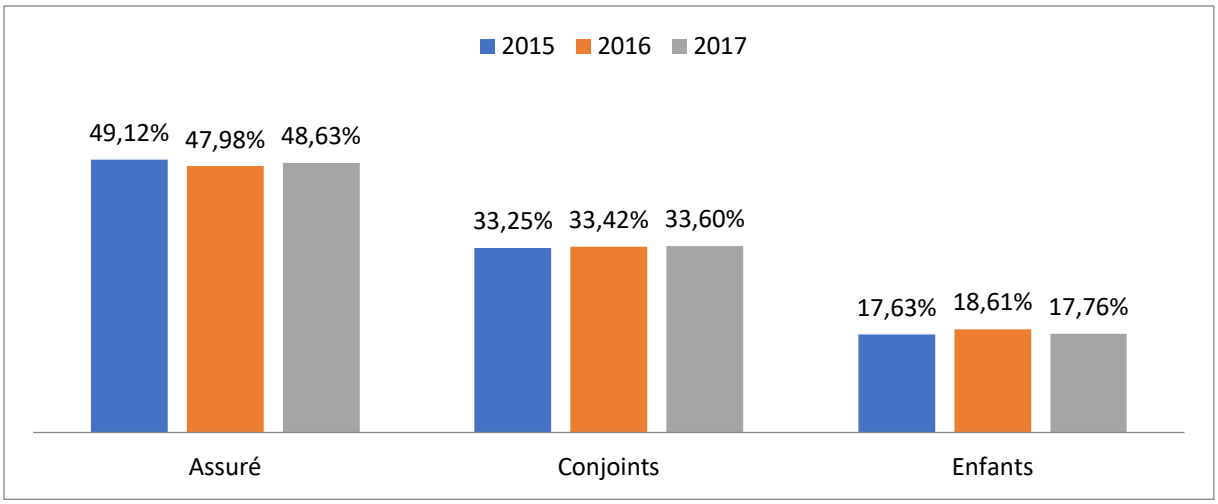


Figure 18: La part du nombre de dossiers par type de bénéficiaire

On retrouve la même hiérarchie pour le nombre de dossiers remboursés par type de bénéficiaire, ainsi les assurés détiennent une part importante de dossiers (de l'ordre de 48%) suivis par leurs conjoints (à peu près 33%) et finalement les enfants ont un nombre moins importants de dossiers avec un pourcentage environnant 17 à 18% .

1) Analyse par sexe du pensionné :

Nous allons, dans ce qui suit, considérer les données relatives aux deux années 2016 et 2017.

SEXE	Effectif	Total de dossiers	Moyenne de dossiers
F	41,44%	35,78%	4,99900128
M	58,56%	64,22%	6,34904265

Tableau6: Répartition de l'effectif et des dossiers par sexe du pensionné

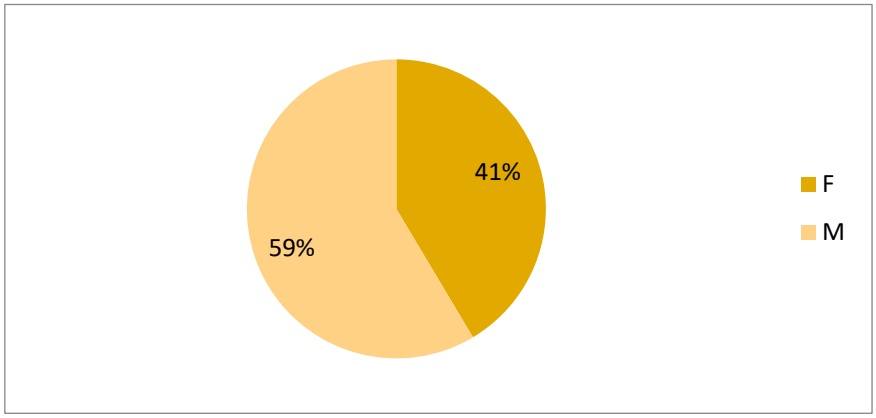


Figure19 : Répartition de la population selon le sexe du pensionné

Nous remarquons que la population masculine est importante par rapport au pensionnées féminins. En effet, ces derniers représentent 59% de la population globale, ce qui explique que leur consommation en terme de nombre de dossiers ainsi que les montants moyens remboursés sont importants également.

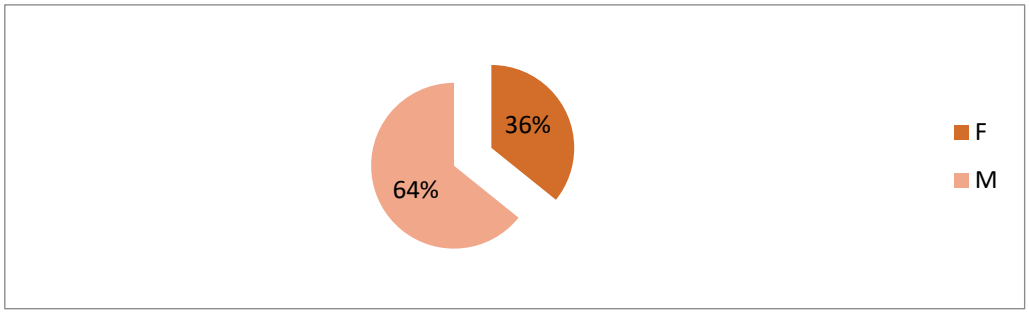


Figure20: Répartition du nombre de dossiers par sexe du pensionné

Effectivement, les hommes s'accaparent de 64% de dossiers avec une moyenne de plus de 6 dossiers par pensionné masculin et à peu près 5dossiers par pensionnée féminin .

SEXE	Montant total remboursé	Montant moyen remboursé
F	38,69%	327,9366519
M	61,31%	381,3956073

Tableau7: Répartition des montants remboursés par sexe du pensionné

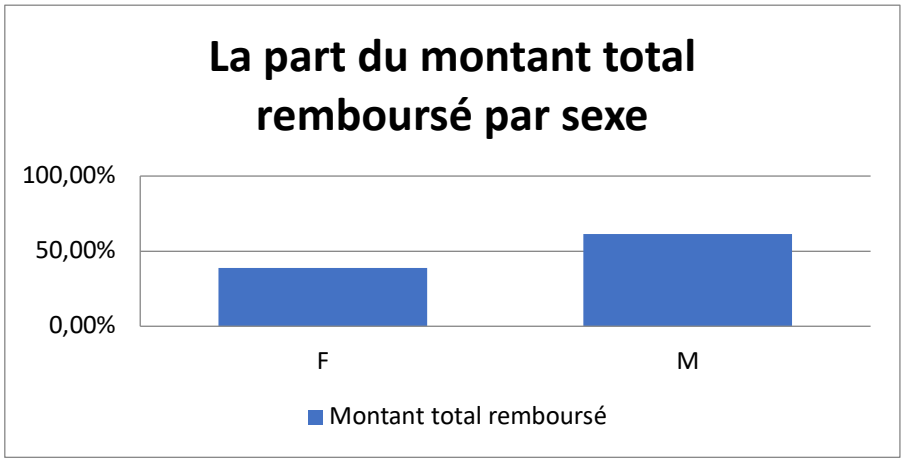


Figure21 : La part du montant remboursé total par sexe du pensionné

En terme de remboursement, les hommes consomment 61,31% du montant total remboursé, avec une part moyenne plus élevée (327,9366519 Dhs VS 381,3956073Dhs par dossier) par rapport aux femmes.

2) Analyse par situation familiale du pensionné:

Situation familiale	Effectif	Total de dossiers	Moyenne de dossiers
C	0,6%	0,4%	3,594059406
D	0,7%	0,6%	5,498245614
M	79,4%	71,5%	5,212797361
V	19,3%	27,5%	8,247167336

Tableau8: La répartition de l'effectif et des dossiers par situation familiale du pensionné

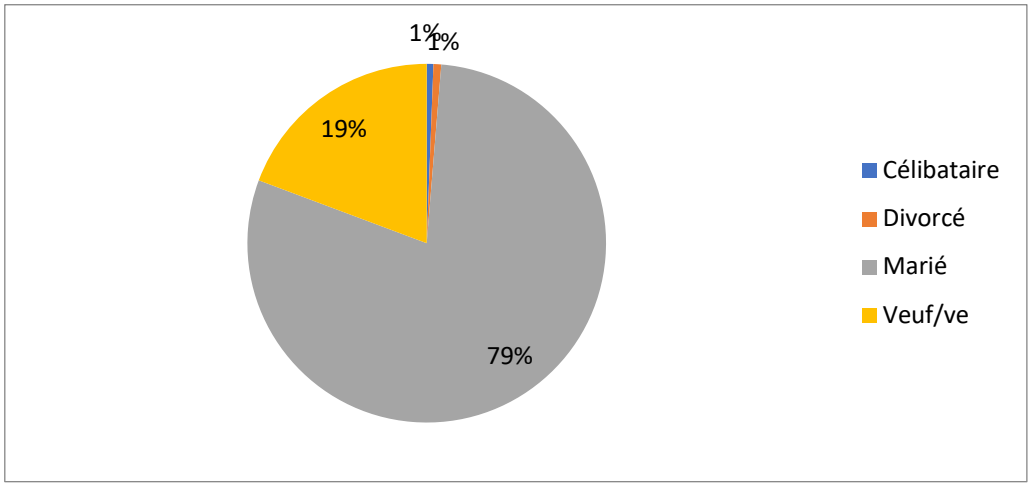


Figure22 : Répartition de la population des pensionnés selon la situation familiale

La population des pensionnés est majoritairement dominée par les mariés(79%), suivis par les veufs(19%), quant aux célibataires et les divorcés ils se partagent les 2% de la population totale des pensionnés.

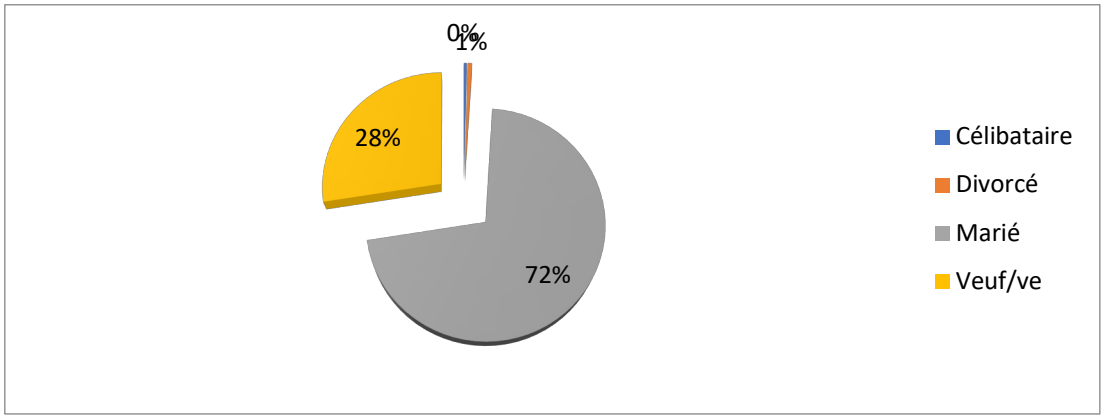


Figure 23: Répartition des dossiers par Situation familiale

les mariés représentent la partie majoritaire de la population, avec une consommation totale de 82%, et une consommation par dossier de 72% . Arrivent, en 2ème position, les veufs/ves recourant à 17% de la totalité des prestations médicales, et à 28% du nombre de dossiers total. tandis que les divorcés et les célibataires consomment moins de 1% de la consommation globale.

Situation familiale	Montant total remboursé	Montant moyen remboursé
C	1%	323,8544956
D	1%	341,1690116
M	82%	371,7181686
V	17%	309,5785666

Tableau9 : Répartition des montants remboursés par situation familiale du pensionné

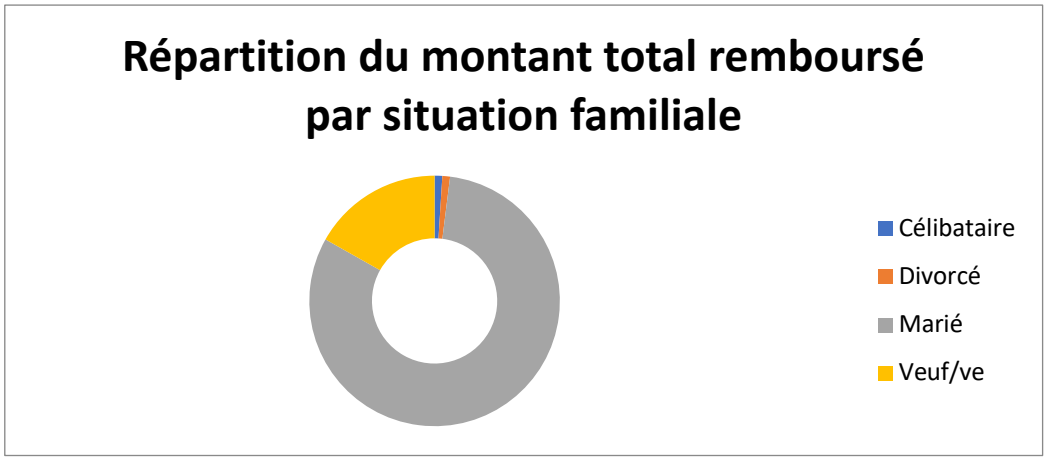


Figure24 : Découpage de la charge des sinistres selon situation familiale

La répartition des montants remboursés montre une prépondérance des mariés, suivis par les veufs/ves puisqu'ils représentent la partie majoritaire de la population des pensionnés. Les pensionnés célibataires et divorcés arrivent en dernier classement avec une part de consommation de 1% chacun.

3) Analyse par pension annuelle:

Intervalle de pensions	effectif	nombre moyen de dossiers	Montant moyen remboursé
<10000	8,2%	7	283,6408741
10000-20000	7,3%	6	327,7608587
20000-30000	14,5%	6	360,007367
30000-40000	7,4%	5	418,2778356
40000-50000	5,1%	5	367,091733
50000-60000	2,6%	5	386,84799
60000-70000	2,1%	5	407,1328393
70000-80000	0,7%	5	404,4095717
80000-90000	0,5%	5	407,9453951
90000-100000	0,4%	5	397,2887585
>100000	51,0%	5	465,0776684

Tableau 10: Répartition de l'effectif et des dossiers par intervalle de pension

on remarque que plus la pension augmente la fréquence de consommation diminue, notons que la consommation par dossier se fige dans une moyenne de 5 dossiers par bénéficiaire à partir d'une pension annuelle de l'ordre de 30 000 de dirhams.

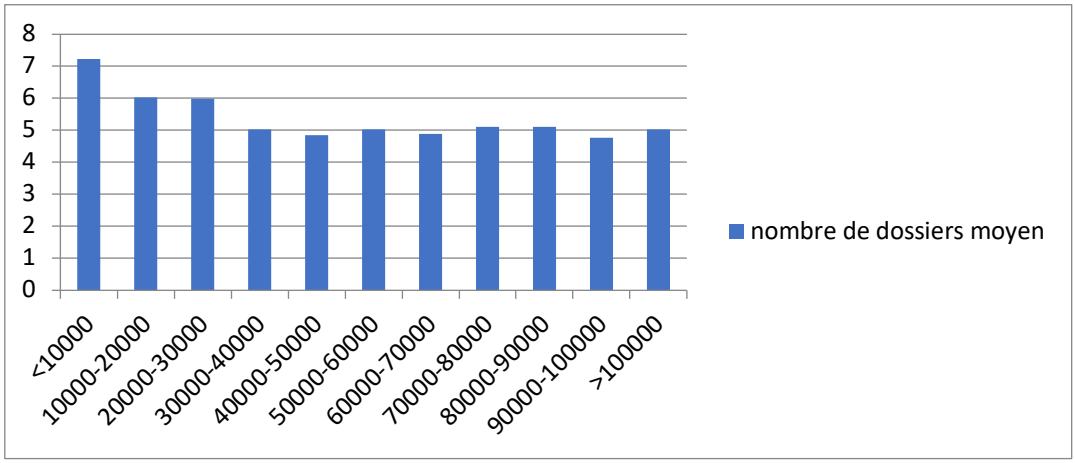


Figure 25: Répartition des dossiers médicaux par intervalle de pension

Pour ce qui est du montant moyen remboursé, le remboursement total augmente et atteint sa limite dans l'intervalle de pension de 2000 à 3000 de dirhams, puis diminue au fur et à mesure que la pension augmente, comme le montre la figure ci dessous.

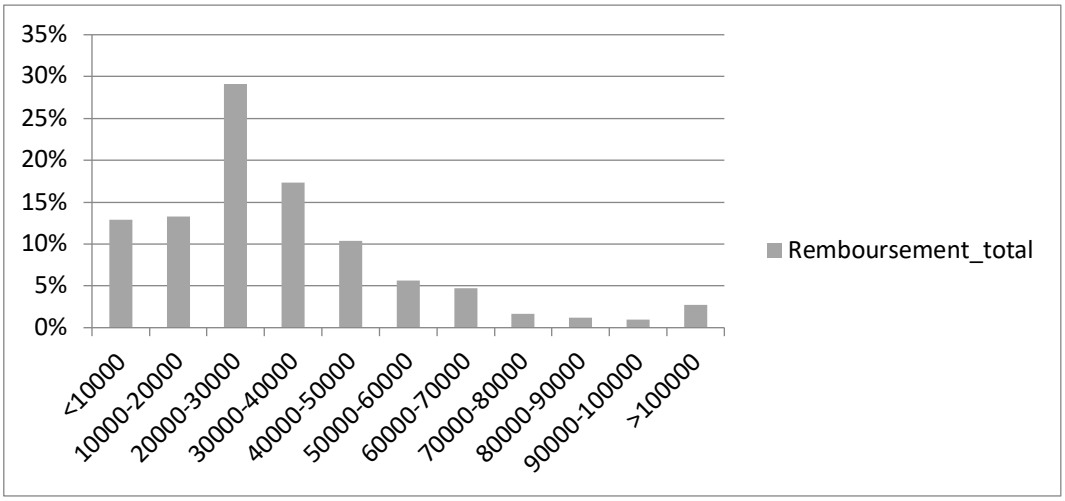


Figure 26: Répartition du remboursement total par intervalle de pension

4) Analyse par tranche d'âge:

L'analyse par âge montre que la consommation des dossiers médicaux augmente au fil de l'âge pour atteindre son apogée à l'âge de 62 ans puis s'éteint progressivement pour atteindre une consommation de l'ordre de 1%.

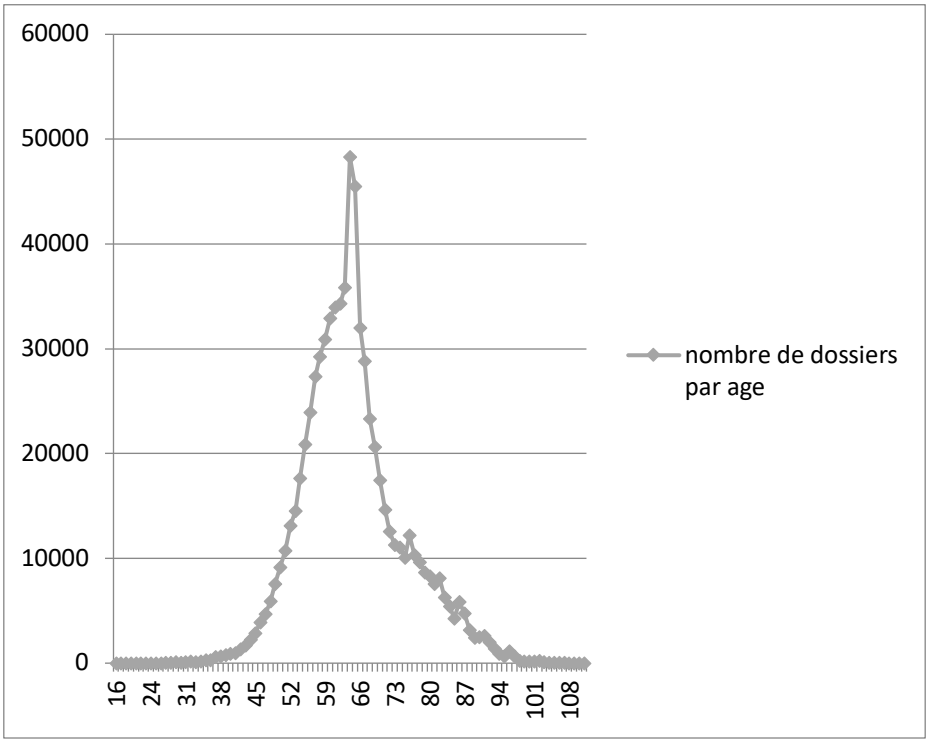


Figure 27: Le nombre de dossiers médicaux par âge du pensionné

En ce qui concerne la courbe de la consommation par âge, la consommation prend des valeurs élevées dans les extrémités de la distribution(moins 23 ans et plus de 108 ans) , puis augmente au fur et à mesure que l'âge augmente, et atteint son pique à l'âge de 65 ans, et enfin diminue au delà de cet âge comme le montre la figure ci dessous.

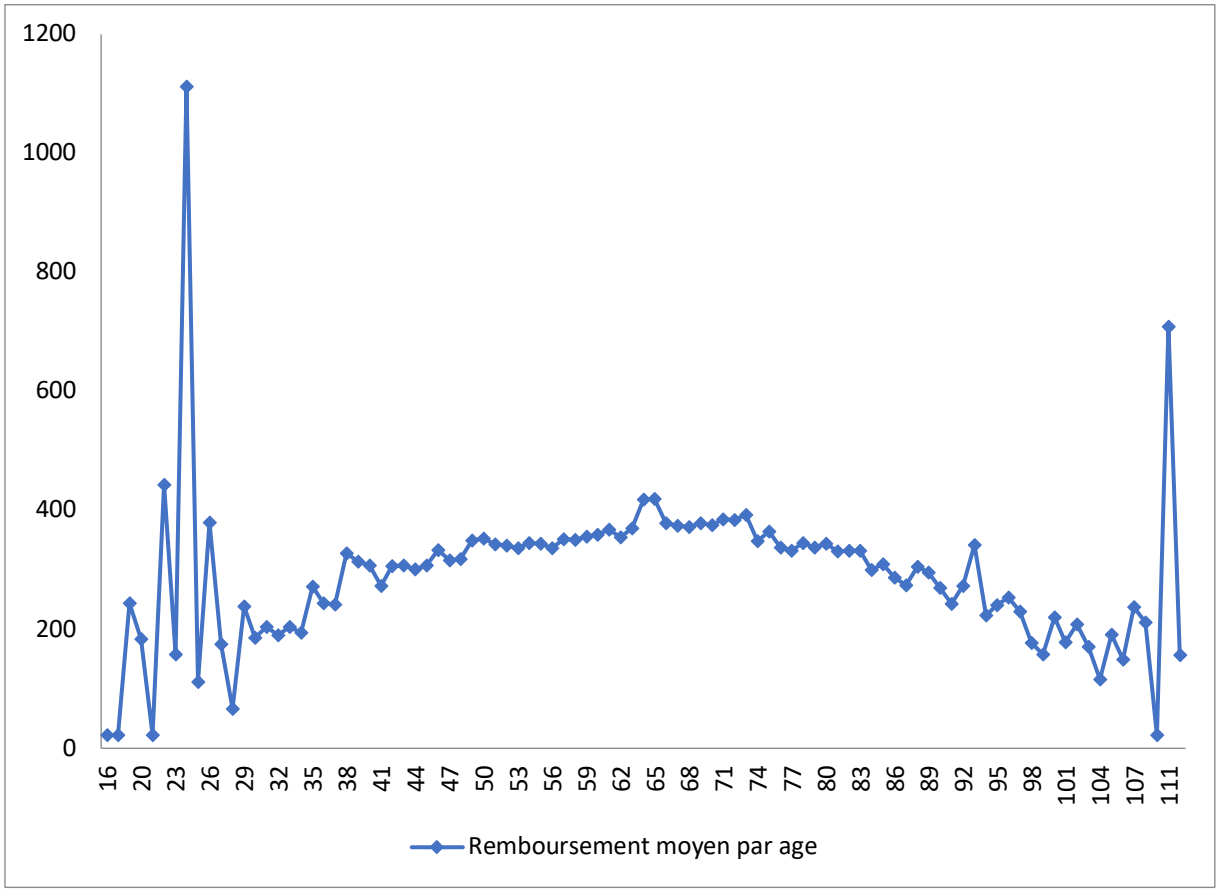


Figure 28: La courbe de consommation médicale par âge du pensionné

5) Analyse du cas de l'ALD ALC

On entend par « dossier normal », un dossier déclaré sous l'étiquette « Non » pour la variable ALD_ALC

Si « Non » veut dire la non présence de maladie grave, couteuse ou de longue durée , nous avons néanmoins retrouvé dans notre jeu de données des bénéficiaires ayant déclaré au cours de l'année des dossiers des deux types : Oui et Non. Comment doit-t-on réagir face à une telle situation ?

L'intérêt de cette partie est donc de sortir avec une décision simple : amalgamer ou pas les dossiers normaux des personnes atteintes d'une ALD avec les dossiers normaux des personnes non atteints d'une ALD.

Ce travail est une étape préliminaire avant d'entamer la tarification que nous effectuerons par la suite. Elle nous garantira la solidité et la fiabilité des résultats futurs de notre étude.

La question à laquelle il faudra répondre est : y-a-t-il une différence significative entre la moyenne de remboursement de ces dossiers et la moyenne de remboursement des dossiers non ALD.

Une analyse de la variance nous paraît donc la démarche la plus appropriée à suivre.

Notons dans ce qui suit les dossiers des personnes atteintes d'une ALD tout simplement par ALD.

Nous procédons par la suite à une analyse comparative entre le remboursement des ALD et non ALD :

Notations du tableau et des graphes:

- Non (ou 1) : les dossiers normaux des non ALD
- Oui (ou 2) : les dossiers normaux des ALD

Variable d'analyse : TOTAL_REMBOURSEMENT TOTAL_REMBOURSEMENT							
YEAR_REMBOUR	ALD_ALC	N Obs	N	Moyenne	Ecart-type	Minimum	Maximum
2015	Non	402424	402424	372.2971709	1093.83	0.3813300	147229.35
	Oui	50954	50954	404.5700192	1694.62	0.3813300	213929.92
2016	Non	504377	504377	367.4391534	1144.93	0.3389600	158101.54
	Oui	34262	34262	441.6815923	1573.54	3.4234960	110594.53
2017	Non	463563	463563	414.0715129	1417.98	0.3389600	119714.66
	Oui	17104	17104	458.3246996	1589.64	4.3386880	87578.79

Tableau 11: L'évolution de la moyenne des remboursements par présence d'ALD_ALC au titre des années 2015, 2016 et 2017.

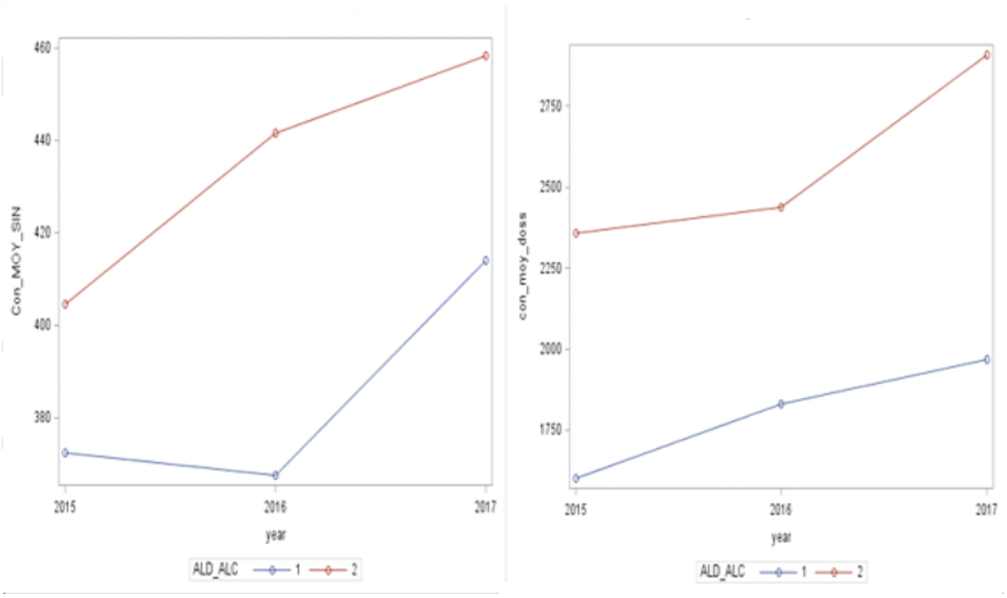


Figure 29: Remboursement moyen par dossier (à gauche) et par personne (à droite) des ALD et des Non ALD

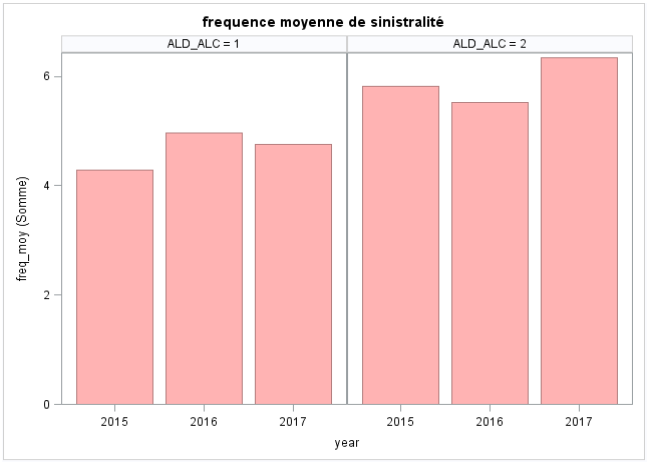


Figure30: fréquence moyenne par an, selon la présence ou pas d'ALD

La fréquence de sinistralité d'un seul ALD est supérieurs à celle d'un non ALD et la moyenne de remboursement de son dossier est nettement supérieur à la moyenne des non ALD.

Même si le nombre de dossiers décroit d'une année à l'autre, le remboursement lui, accroît. La question qui se pose donc est : est-ce que cet accroissement est dû à une augmentation de la fréquence de sinistralité (nombre de dossiers par personnes) ou à un effet d'inflation des médicaments ?

On remarque que la fréquence a légèrement baissé, (de l'ordre de 6 par an), ce qui nous laisse penser que la cause de l'augmentation du coût du remboursement est dû à l'augmentation des couts médicaux (étant donné que même le nombre de dossiers médicaux a diminué d'une année à l'autre). Cette augmentation est-elle peut être relative aux coûts de soins exclusivement liés aux ALD ?

Les coûts de remboursement des non ALD a par contre a augmenté parallèlement avec l'ascension de la fréquence de sinistralité d'un dossier Non ALD. Ce qui nous laisse penser que pour les Non ALD, c'est l'effet de volume qui a le plus impacté le niveau de remboursement. (Ceci ne veut pas pour autant dire qu'il n'y a pas un effet d'inflation)

Quelle conclusion ? Les écarts types sont très grands, et ne nous permet pas de trancher. Nous chercherons donc une variable qui nous permettra d'obtenir des groupes plus homogènes. La variable qui nous paraît la plus adéquate est naturellement l'âge des collaborateurs. L'âge a-t-il donc un impact ? Existe-t-il une différence significative entre les moyennes de remboursement par groupes d'âge ?

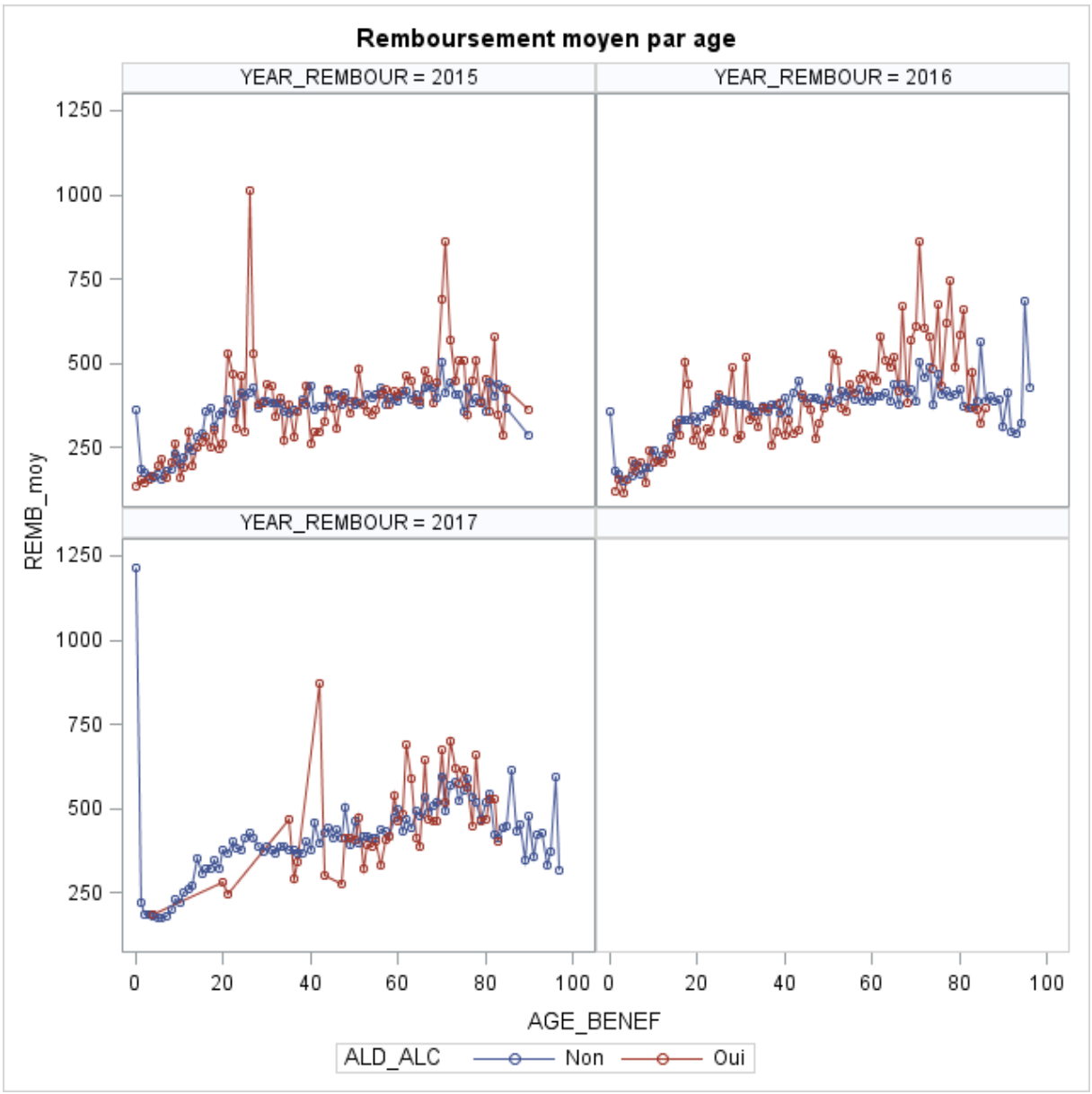


Figure 31: Comparaison des consommations médicales moyennes selon la présence ou non de l'ALD_ALC

Cette répartition, ainsi qu'une analyse d'ANOVA permet de montrer l'existence d'une différence significative entre les moyennes de remboursement suivant l'âge. Pour appuyer ce résultat, nous combinerons les 3 années d'observation après avoir pris en considération l'inflation médicale pour les années 2016 et 2017. Les résultats sont donc les suivantes :

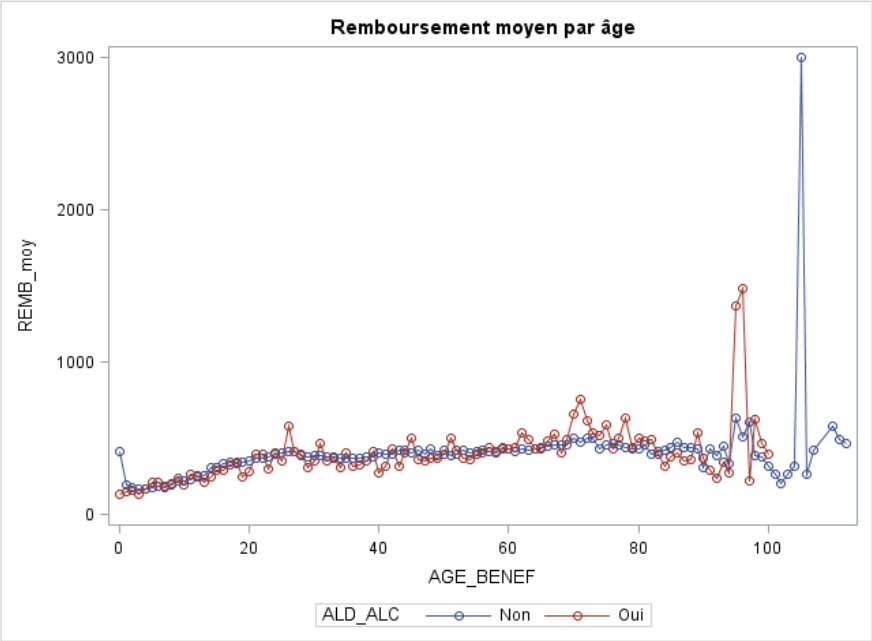


Figure 32: Le remboursement moyen par âge du pensionné selon la présence d'ALD_ALC, toutes les années confondues.

Les valeurs extrêmes ont été localisées puis traitées, vu que la fréquence n'est pas assez importante pour les prendre en considération comme moyenne représentative de la classe d'âge associée .

	TOTAL_REMBOURSEMENT							
	ALD_ALC							
	Non				Oui			
	N	Mean	Std	Max	N	Mean	Std	Max
95	332	636.77	1541.06	13663.71	72	1366.80	2119.60	8974.34
96	479	510.67	1570.91	23852.42	46	1481.01	1800.52	6824.52
97	454	605.96	4753.36	98581.43	19	225.20	176.26	607.09
98	92	384.17	867.56	6563.72	10	623.93	1316.51	4335.40
99	91	383.26	1067.85	6892.81	26	462.89	1219.84	6334.32
100	127	321.23	598.64	3784.83	7	393.95	308.59	874.14
101	112	267.86	495.57	5084.40	.	.	.	.
102	120	205.76	196.27	1285.84	.	.	.	.
103	64	260.49	699.09	5694.53	.	.	.	.
104	10	318.11	165.42	482.14	.	.	.	.
105	5	3003.30	4756.48	11118.97	.	.	.	.

Tableau12: Localisation des valeurs extrêmes

Après avoir nettoyer la base, nous obtenons le résultat suivant :

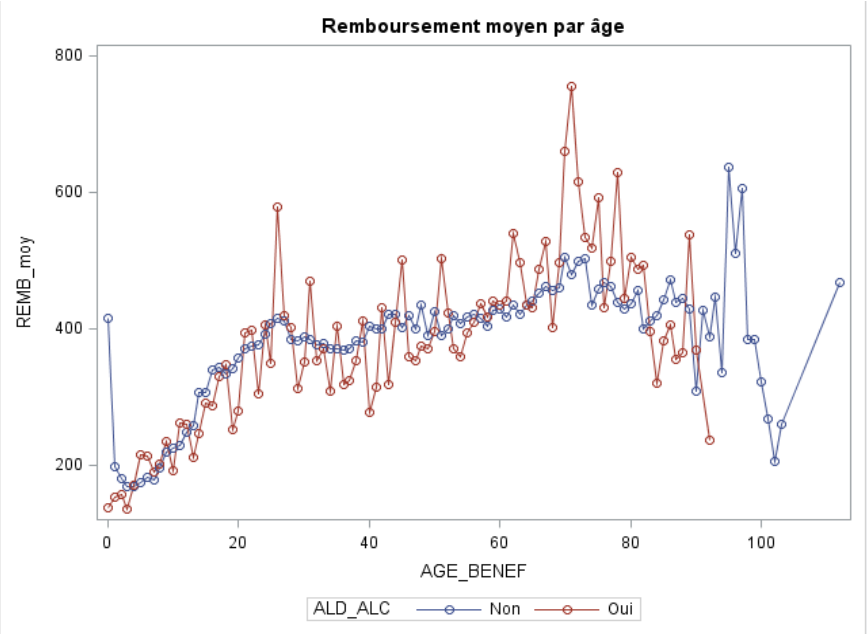


Figure 33: Le remboursement moyen par âge du pensionné selon la présence d'ALD_ALC, après nettoyage de la base.

Le test ANOVA final est un test à deux facteurs : présence ALD et âge du bénéficiaire, On vérifiera l'existence d'une différence significative entre les moyennes de remboursement suivant la présence d'ALD chez la personne et suivant son âge . Le résultat ci-dessous montre que la différence est significative.

Remboursement moyen par age

The ANOVA Procedure

_REMBOURSEMENT TOTAL_REMBOURSEMENT

Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Model	211	9796823031.3	46430440.907	29.19	<.0001
Error	1.44E6	2.2863272E12	1590377.8267		
Corrected Total	1.44E6	2.296124E12			

r-carré	Coef de Var	Racine MSE	TOTAL_REMBOURSEMENT Moyenne
0.004267	324.4847	1261.102	388.6475

Source	DDL	Anova SS	Moyenne quadratique	Valeur F	Pr > F
ALD_ALC	1	196986038	196986038	123.86	<.0001
AGE_BENEF	110	9278982547	84354387	53.04	<.0001
AGE_BENEF*ALD_ALC	100	320854447	3208544	2.02	<.0001

Tableau13: Les résultats du test ANOVA à deux facteurs

Conclusion finale :

Nous ne devrons pas amalgamer les dossiers normaux des ALD avec les dossiers normaux des Non ALD. Ils sont à considérer comme une classe propre qu'on notera par la suite O_N.

D'une autre part, ils ne seront sûrement pas amalgamés avec les dossiers déclarés ALD. La preuve est tout simplement devant nous comme suit :

Notons ici :

Oui : dossier ALD

Non : dossier normal d'une personne ayant au moins déclaré une ALD

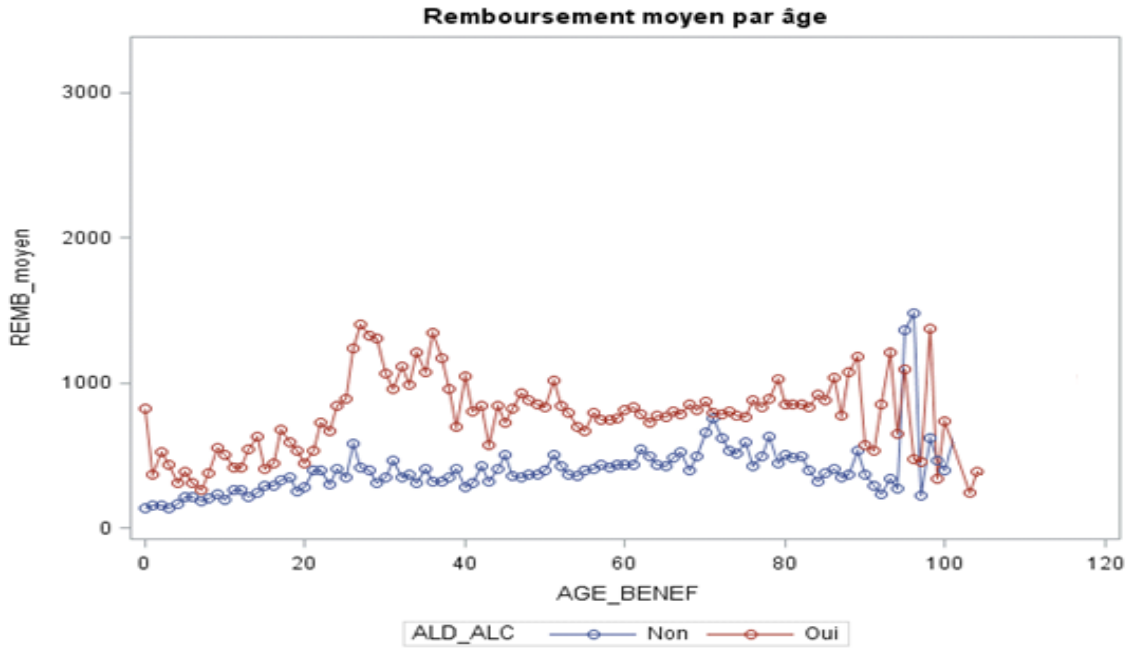


Figure 34: Comparaison des courbes de consommation par âge des dossiers ALD et Non ALD

Ce résultat reste tout de même discutable. On s'assurera de la justesse de cette classification à partir des résultats de la modélisation GLM. (Le groupe de classe reste-t-il significatif ou faut il après tout le regrouper avec les dossiers Non ALD).

1.2.3 Ecrêtement des grands coûts et fréquences

Dans le cadre de la tarification (et du provisionnement) en assurance non vie, notamment l'assurance maladie, une hypothèse classique est celle selon laquelle le portefeuille est constitué de risques similaires. Un problème pour que cette hypothèse soit vérifiée est le poids important des sinistres « graves ». Pour le résoudre, les sinistres observés sont souvent écrêtés : ils sont plafonnés à un niveau maximum M Dhs. La charge résiduelle étant répartie sur une assiette plus large que le segment dont elle est issue (ex : répartition à l'ensemble de la mutualité de risques). Le choix de ce plafond peut être délicat et peut conduire à une sous-estimation ou à une surestimation des sinistres ordinaires. Quelques techniques classiques d'écrêtement :

- Choisir un seuil tel que la sur-crête représente un certain pourcentage de la charge totale de sinistres.
- choisir un certain quantile de la distribution historique des coûts unitaires de sinistres (typiquement 0,5 %, 1 % ou 2 %).

1)Ecrêtement des grands coûts

Nous opterons pour la deuxième technique pour retrouver le seuil d'écrêtement des grands coûts de sinistres.

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	43975.8908
99%	2770.8433
95%	1137.5647
90%	784.2524
75% Q3	432.7503
50% Médiane	217.7828
25% Q1	75.4509
10%	0.0000
5%	0.0000
1%	0.0000
0% Min	0.0000

Tableau 14: Quantiles de la distribution historique des coûts moyens

Ainsi, le seuil choisi sera de 2770 Dhs, et la charge résiduelle sera répartie sur toute la population afin de mutualiser le risque. Nous aurons donc une augmentation de 22 dhs sur le coût moyen de chaque bénéficiaire sur la base de données.

2) Ecrêtement des grandes fréquences

Quantiles (Définition 5)	
Niveau	Quantile
100Max 100%	144
99%	36
95%	20
90%	14
75% Q3	8
50% Médiane	4
25% Q1	1
10%	0
5%	0
1%	0
0% Min	0

Tableau 15: Quantiles de la distribution historique de la fréquence des sinistres

La plupart ont un nombre moyen de sinistres inférieurs à 4 par ans quand certains peuvent dépasser les 100, nous avons jugé donc qu'il fallait faire un écrêtement des fréquences aussi , et de distribuer les fréquences excédentaires après avoir trié la population par fréquence descendante. Le seuil d'écrêtement correspond à 36.

1.2.4 Modélisation de la fréquence de sinistralité

Nous devrons estimer la loi du nombre de sinistre par an. Plusieurs lois discrètes vont être testés.

1) Loi de Poisson

Dans un modèle de Poisson, la probabilité pour qu'une variable aléatoire Y (nombre de sinistres responsables déclarés) prennent la valeur y_i(y_i =0,1, 2...) pour un assuré iest donnée par :

$$\mathrm{P}(\mathrm{Y} = y_i \,/X_i\,) = \frac{e^{-\lambda_i} * \lambda_i{}^{y_i}}{y_i\,!}$$

Où le paramètre $\lambda_i$dépend du vecteur $X_i$des caractéristiques (régresseurs) del'assuré i

On vérifie aisément que dans la loi de Poisson (1), l'espérance est égale à la variance.

$$E(y_i / X_i) = Var(y_i / X_i) = \lambda_i$$

En choisissant la fonction de lien log linéaire, on aura :

$$Ln(\lambda_i) = X'_i\beta$$

oùβ est le vecteur descoefficients à estimer.

Cependant, nous devrons s'assurer de l'ajustement de la loi de poisson à la distribution historique de la fréquence avant d'aller plus loin dans la modélisation GLM. Le résultat sous SAS donne :

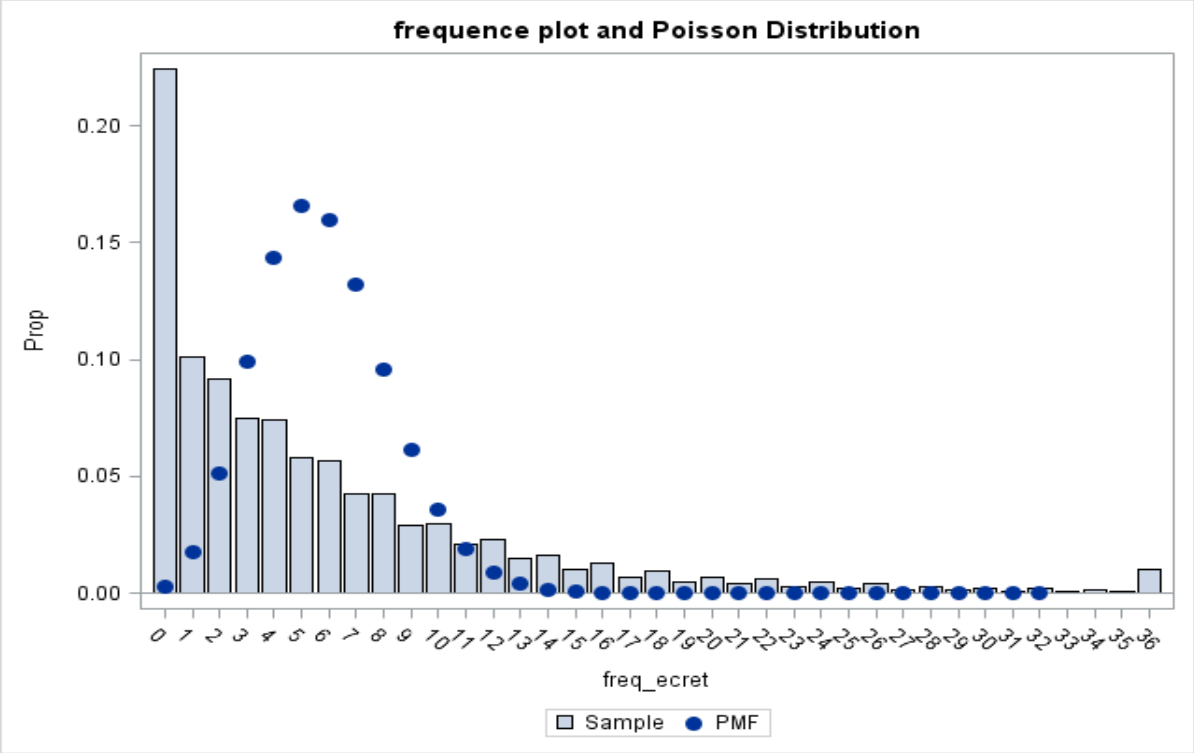


Figure 35: Test d'ajustement graphique de la loi de poisson

Le test chi-2 de Pearson montre aussi que la loi de poisson n'ajuste pas nos données

Cette hypothèse d'équidispersion (homogénéité de la population par rapport au risque maladie) est très restrictive. Mais dans la pratique, du fait d'une abondance de valeurs nulles et de la présence de quelques valeurs extrêmes, la variance est supérieure à la moyenne. Dans ce cas, on parle d'une sur-dispersion de la variable Y. Cette situation implique une sous-estimation des écarts types et on rejette trop souvent l'hypothèse

nulle de non significativité des coefficients β du modèle. D'où l'idée d'utiliser un modèle de comptage alternatif, basé sur la loi binomiale négative, qui prend en compte cette sur-dispersion par l'introduction d'un paramètre supplémentaire (α) qui permet, en outre, de capter l'hétérogénéitéinobservée de la variable endogène (qui peut impliquer la surdispersion observée).

2) Loi Binomiale négative

Dans un modèle binomial négatif, on définit la probabilité pour que Y prenne la valeur $Y_I$par:

$$P(Y = y_i / X_i) = \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1).\Gamma(\nu)} \cdot \left(\frac{\nu}{\lambda_i + \nu}\right)^{\nu} \cdot \left(\frac{\lambda_i}{\lambda_i + \nu}\right)^{y_i}$$

En posant$\nu = \frac{1}{\alpha}$, l'espérance et la variance s'expriment ainsi :

$$E(y_i / X_i) = e^{X'_i\beta} = \lambda_i$$

$$Var(y_i / X_i) = \lambda_i(1+\alpha\lambda_i)$$

La variance est donc différente de l'espérance et le paramètre α traduit une sur-dispersion (ou une sous dispersion) des données. Si α=0 , le modèle binomial se réduit au modèle de Poisson. Si α>0 , le modèle de Poisson est rejeté au profit du modèle négatif binomial. La sur-dispersion peut être testée soit par le ratio D / (n -p) , où D désigne la déviance, n le nombre d'observations et p le nombre de paramètres dans le modèle, soit parle ratio X^2/ (n -p) , où X^2correspond à la statistique du chi-deux de Pearson. Ladéviance est définie comme 2 fois la différence entre le maximum possible de la logvraisemblance et le maximum atteint sur le modèle estimé.

Si ces ratios sont supérieurs à 1, les données présentent une sur dispersion (et une sous dispersion si ces ratios sont inférieurs à 1).

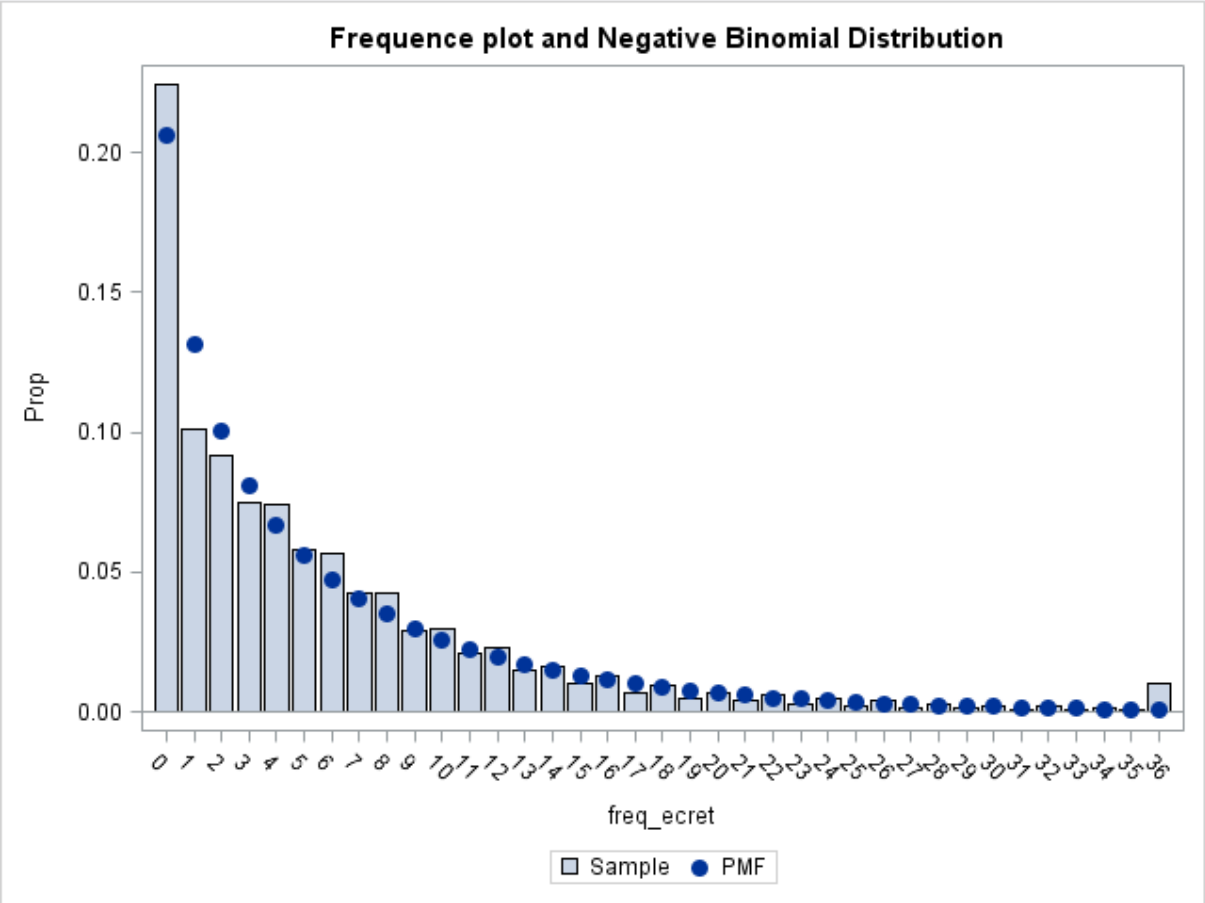


Figure 36: Test d'ajustement graphique de la loi Négative binomiale

Malgré l'ajustement meilleur que nous donne la loi négative binomiale , le test chi-2 de Pearson annonce qu'elle n'ajuste pas correctement nos données. Cependant elle reste la meilleure distribution à utiliser par la suite dans le modèle GLM.

3) Utilisation d'une variable offset

Si on veut tenir compte de l'exposition au risque d, on sait que l'espérance l de la loi binomiale devient $\boldsymbol{\lambda d}$. La régression se réécrit alors :

$$\boldsymbol{E[N/\,x, d\,] = d \times \exp\left(\sum_{k=1}^{p} \beta_k x_k\right) = \exp\left(\beta' x + \ln(d)\right)}$$

Tout se passe donc comme si l'on ajoutait une variable explicative pour laquelle le coefficient β est connu (ici égal à 1) et ne doit donc pas être estimé. La variable s'appelle une variable offset.

Mais dans notre cas, tous les bénéficiaires ont un code de pensionné, ce qui veut dire que l'exposition au risque est étalée sur toute l'année et donc est égale à 1 . La variable offset devient par conséquent nulle. Les coefficients ne seront donc pas contraints (ln(d) =0) .

4) Segmentation des variables tarifaire

La codification des variables qualitatives étant fait (SEXE, ALD_ALC, Situation familiale)
La segmentation des variables quantitatives s'impose aussi . Deux variables quantitatives sont concernés :

- Le niveau de la pension
- L'âge du bénéficiaire.

L'objectif est de produire dans l'espace de la variable explicative des régions où la valeur de la variable à expliquer est homogène (variance faible) avec le minimum de perte d'information.

Il faudra donc former des classes d'âge et de pension regroupant les observations de fréquence selon le critère de la variance.

Nous utiliserons pour cette fin les arbres de décision :Le logiciel utilisé est SAS miner. Il va nous générer, en plus d'un arbre de décision graphique (Voir ANNEXE), un code de segmentation qui sera utilisé ensuite pour produire les classes des variables tarifaires sur notre base de données sur SAS.

Le résultat de la segmentation des variables tarifaires est comme suit :

CLASSE d'Age	Intervalle
1	< 47 ans
2	[47, 50[ans
3	[50,55 [ans
4	[55,61 [ans
5	[61,65 [ans
6	[65,72 [ans

7	[72,83 [ans
8	[83,93 [ans
9	>= 93

Tableau 16: La segmentation de la variable "Age"

En ce qui concerne la segmentation de la pension, nous retrouvons :

CLASSE de Pension	Intervalle (Dhs)
1	< 2875,7
2	[2875,7 ; 2876 [
3	[2876 ; 3300 [
4	[3300 ; 5753,4 [
5	[5753,4 ; 9481 [
6	[9481 ; 11281 [
7	[11281 ; 18862 [
8	[18862 ; 21988 [
9	[21988 ; 27222 [
10	[27222 ; 31258 [
11	[31258 ; 31273 [
12	>= 31273

Tableau 17: La segmentation de la variable Pension

5) Résultat de la modélisation GLM avec la loi négative binomiale

a) 1er modèle (ALD-ALC inclus comme variable tarifaire)

Informations sur le modèle	
Data Set	BD.RETFR
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Frequence

Tableau 18: Récapitulatif des informations sur le modèle utilisé

Après avoir regroupé les classes de variables qui n'étaient pas significatifs, nous restons avec 5 classes d'âges et 10 classes de pension :

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
F_AGE_CLASS	5	7 8 9 16 17
F_PENSION_CLASS	10	7 12 14 16 17 18 19 20 22 23
SIT_FAM	4	C D M V
ALD_ALC	3	Non O_N Oui

Tableau 19: Classes de variables GLM (fréquence)

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	13E4	149505.4462	1.1460
ScaledDeviance	13E4	149505.4462	1.1460
Pearson Chi-Square	13E4	125568.0203	0.9625
Scaled Pearson X2	13E4	125568.0203	0.9625
Log Likelihood		904938.3566	
Full Log Likelihood		-362668.5113	
AIC (smallerisbetter)		725379.0225	
AICC (smallerisbetter)		725379.0296	
BIC (smallerisbetter)		725584.3807	

Tableau 20: Résultat de validation du modèle GLM pour la fréquenc

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	2.9953	0.0260	2.9442	3.0463	13224.6	<.0001
SEXE	F	1	-0.0907	0.0074	-0.1052	-0.0763	151.79	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
F_AGE_CLASS	7	1	-0.1756	0.0433	-0.2605	-0.0906	16.42	<.0001
F_AGE_CLASS	8	1	-0.1185	0.0252	-0.1678	-0.0691	22.14	<.0001
F_AGE_CLASS	9	1	0.0432	0.0182	0.0074	0.0789	5.60	0.0180
F_AGE_CLASS	16	1	0.0751	0.0193	0.0373	0.1130	15.12	0.0001
F_AGE_CLASS	17	0	0.0000	0.0000	0.0000	0.0000	.	.
F_PENSION_CLASS	7	1	-0.1332	0.0098	-0.1524	-0.1139	183.37	<.0001
F_PENSION_CLASS	12	1	-0.1050	0.0163	-0.1370	-0.0730	41.39	<.0001
F_PENSION_CLASS	14	1	-0.1683	0.0255	-0.2182	-0.1183	43.61	<.0001
F_PENSION_CLASS	16	1	-0.5169	0.0335	-0.5826	-0.4512	237.78	<.0001
F_PENSION_CLASS	17	1	-0.2540	0.0212	-0.2955	-0.2125	143.83	<.0001
F_PENSION_CLASS	18	1	-0.3110	0.0179	-0.3461	-0.2759	302.03	<.0001
F_PENSION_CLASS	19	1	-0.1667	0.0230	-0.2118	-0.1217	52.60	<.0001
F_PENSION_CLASS	20	1	-0.1728	0.0145	-0.2013	-0.1443	141.04	<.0001
F_PENSION_CLASS	22	1	0.0633	0.0123	0.0391	0.0875	26.28	<.0001
F_PENSION_CLASS	23	0	0.0000	0.0000	0.0000	0.0000	.	.
SIT_FAM	C	1	-0.9401	0.0451	-1.0286	-0.8516	433.68	<.0001
SIT_FAM	D	1	-0.5030	0.0428	-0.5869	-0.4191	138.14	<.0001
SIT_FAM	M	1	-0.6655	0.0136	-0.6922	-0.6389	2402.62	<.0001
SIT_FAM	V	0	0.0000	0.0000	0.0000	0.0000	.	.
ALD_ALC	Non	1	-0.7862	0.0121	-0.8099	-0.7625	4223.23	<.0001
ALD_ALC	O_N	1	0.0746	0.0157	0.0438	0.1055	22.52	<.0001
ALD_ALC	Oui	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	1.2361	0.0061	1.2242	1.2481		

Note: The negative binomial dispersion parameterwasestimated by maximum likelihood.

Tableau 21: Sortie SAS du résultat GLM pour la fréquence

Nous observons, à travers le tableau ci-dessus que tous les p-values du test de Wald sont inférieures à 0.05. l'effet explicatif des variables sexe, âge, pension et situation familiale est confirmé, d'où l'adéquation de notre modèle.

b) 2ème approche (ALD_ALC disjoints)

Pour ne pas encombrer le mémoire, nous n'exhiberons que la modélisation de la fréquence pour cas l'ALD est présente (ALD_ALC = Oui) , celui-ci ayant donné un résultat (comme nous le verrons dans la modélisation du coût) particulier. La modélisation pour les autres catégories d'ALD ALC sont en annexe.

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	13E4	24301.1967	0.1863
Scaled Deviance	13E4	24301.1967	0.1863
Pearson Chi-Square	13E4	69636.9640	0.5337
Scaled Pearson X2	13E4	69636.9640	0.5337
Log Likelihood		119351.4823	
Full Log Likelihood		-71144.8977	
AIC (smaller is better)		142307.7954	
AICC (smaller is better)		142307.7968	
BIC (smaller is better)		142395.8060	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	-0.4607	0.0753	-0.6082	-0.3132	37.48	<.0001
AGE_CLASS	5	1	0.3073	0.0774	0.1556	0.4589	15.77	<.0001
AGE_CLASS	6	1	0.6854	0.0738	0.5407	0.8301	86.20	<.0001
AGE_CLASS	7	1	0.3533	0.1101	0.1375	0.5690	10.30	0.0013
AGE_CLASS	8	1	-0.5451	0.1163	-0.7730	-0.3172	21.97	<.0001
AGE_CLASS	9	0	0.0000	0.0000	0.0000	0.0000	.	.
PENS_CLASS	2	1	-0.5691	0.0548	-0.6765	-0.4618	107.93	<.0001

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
PENS_CLASS	7	1	-0.2861	0.0453	-0.3748	-0.1974	39.95	<.0001
PENS_CLASS	8	1	-0.4107	0.0751	-0.5579	-0.2635	29.91	<.0001
PENS_CLASS	9	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	42.5343	0.5010	41.5635	43.5277		

1.2.5 Modélisation du coût moyen

1) Loi Log-normale

La variable X suit une loi log normale de paramètres μ et σ, (X ~ ln N(μ, σ)) si le logarithme de X suit une loi normale (ln(X) ~ N(μ, σ)).

Sa densité est donnée par l'expression suivante :

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

où x > 0, μ ∈ R et σ > 0.

Les moments de la loi log normale sont donnés par :

$$E(X) = \exp(\mu + \sigma^2/2)$$

$$Var(X) = [\exp(2\mu + \sigma^2)][\exp(\sigma^2) - 1]$$

2) Loi Gamma

La variable X suit une loi Gamma de paramètres α et β, (X↷Ga(α, β))

si sa densité est de la forme :

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)}{\Gamma(\alpha)} 1_{IR+ (x)}$$

où $\alpha > 0$, $\beta > 0$ et $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$.

Les moments de la loi Gamma sont donnés par :

$$E(X) = \alpha\beta$$

$$Var(X) = \alpha\beta^2$$

3) Loi Weibull

La variable X suit une loi de Weibull de paramètres α et β, ($X \sim W(\alpha, \beta)$) si sa densité est de la forme :

$$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} \exp(-(\frac{x}{\beta})^{\alpha}) 1_{IR+ (x)}$$

et sa fonction de répartition :

$$F(x) = 1 - \exp(-(x\beta)^{\alpha})$$

où $\alpha > 0$ et $\beta > 0$.

Les moments de la loi de Weibull sont donnés par :

$$E(X) = \beta \alpha \Gamma(\frac{1}{\alpha})$$

$$Var(X) = \frac{\beta^2}{\alpha}[2\Gamma(\frac{2}{\alpha}) - \frac{1}{\alpha}\Gamma(\frac{1}{\alpha})^2$$

4) Segmentation des variables tarifaires

De la même manière que précédemment, nous segmenterons les 2 variables quantitatives AGE et PENSION , avant de faire la modélisation en GLM.

Les résultats de segmentation sont en ANNEXE avec son Code flux obtenu à partir de SAS miner.

5) Résultat de la modélisation GLM du coût moyen avec la loi Gamma

a) Test de normalité

Nous allons tester l'hypothèse suivante :

H_0 : les montants remboursés moyens suivent une loi Normal N(μ,σ^2)

Vs

H_1 : les montants remboursés moyens ne suivent pas une loi Normal N(μ,σ^2)

Le résultat du test Shapiro wilk sous le logiciel stata est le suivant:

```
. swilk REMBOURSEMENT_DOSS_moy

                    Shapiro-Wilk W test for normal data

    Variable |   Obs        W           V          z       Prob>z
-------------+--------------------------------------------------
REMBOURSEM~y |130478     0.35991    2.3e+04    28.287     0.00000
```

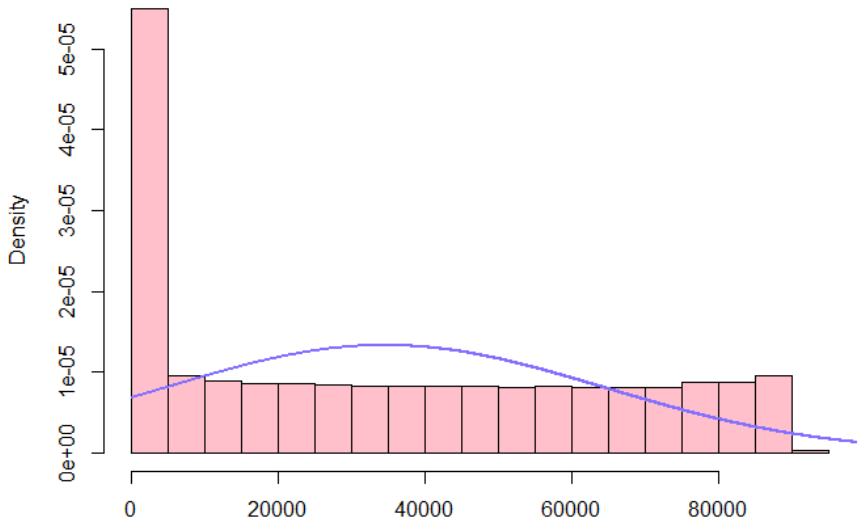


Figure 37:Ajustement courbe de la loi Normale sur l'histogramme du cout moyen.

La p-value du test de Shapiro Wilk est inférieur à 5% donc on rejette l'hypothèse nulle de normalité.

b) Test d'ajustement graphique : choix de la loi la plus adaptée

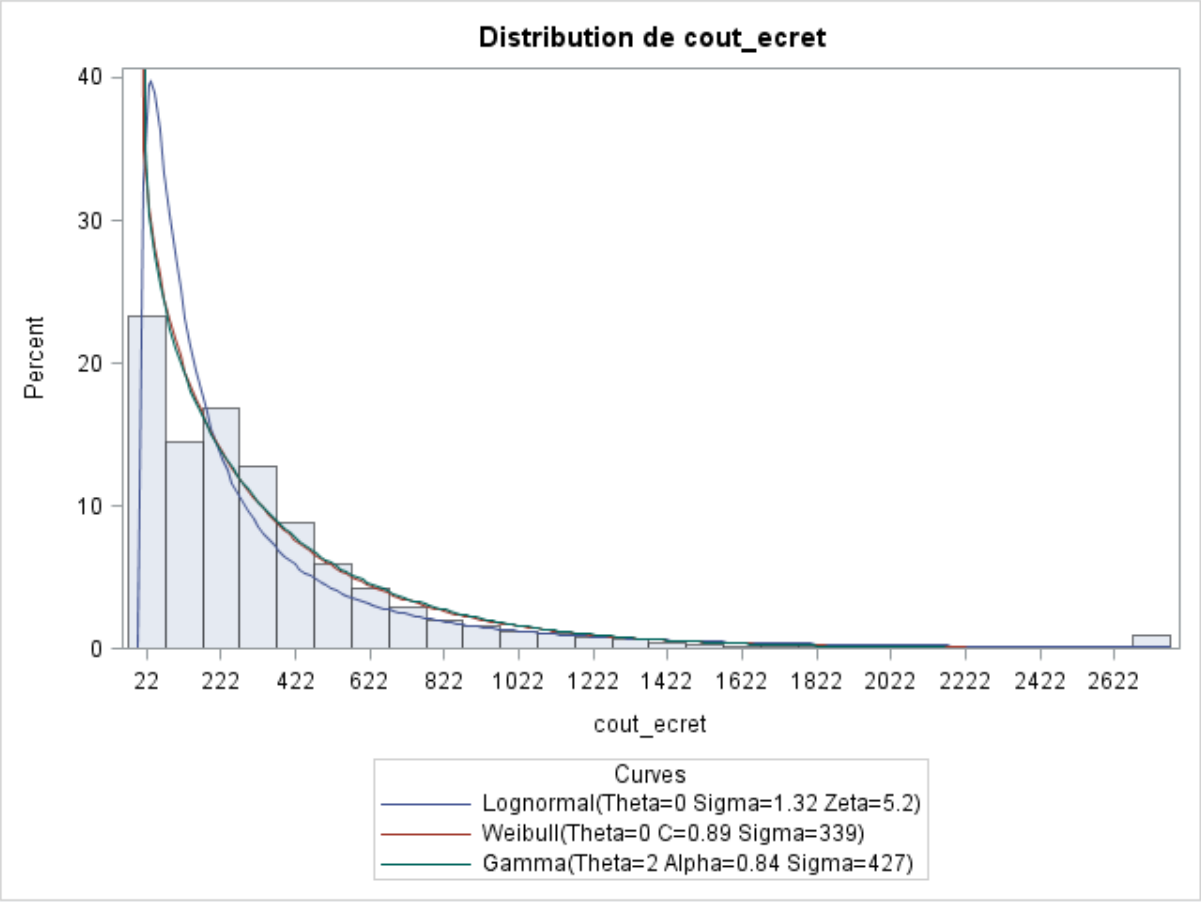


Figure 38: L'ajustement graphique par la loi log-normale, Gamma et Weibull

A partir du test graphique et les QQ-plots des 3 lois , La loi Gamma semble la meilleure possible à ajuster la distribution du coût moyen. (Les Résultats d'ajustement et les QQ-plots des autres lois sont en ANNEXE)

Quantiles for Gamma Distribution		
	Quantile	
Pourcentage	Observé	Estimé
1.0	22.1614	3.62305
5.0	22.1614	13.24085
10.0	22.1614	28.22467
25.0	93.5515	86.24075
50.0	241.6323	230.97839
75.0	453.6914	496.47545
90.0	799.2345	861.22758
95.0	1142.6135	1142.25025
99.0	2686.5056	1803.56348

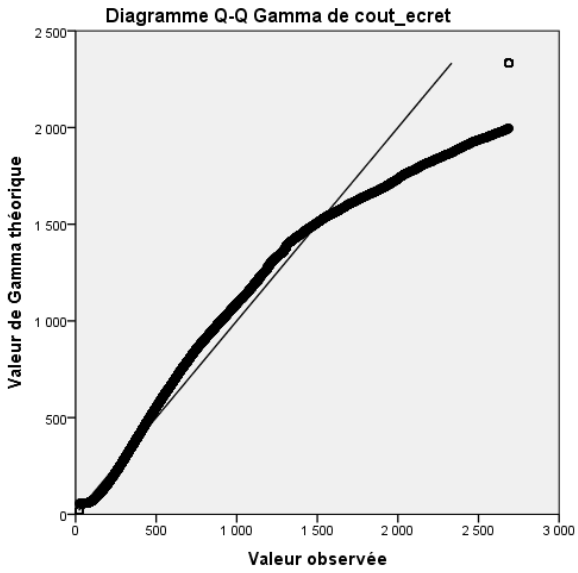


Figure 39: Quantiles de la distribution Gamma des coûts moyens

c) 1er Modèle GLM pout le coût moyen

En fusionnant les classes pour obtenir de nouvelles classes significatives, nous ressortons avec :

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
AGE_CLASS	6	8 9 10 12 14 15
PENS_CLASS	7	4 6 8 9 14 16 17
SIT_FAM	4	C D M V
ALD_ALC	2	Non Oui

Tableau 22: Classes de variables retenus (pour le coût moyen)

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	13E4	170826.7432	1.3149
ScaledDeviance	13E4	152119.1651	1.1709
Pearson Chi-Square	13E4	184875.6378	1.4231
Scaled Pearson X2	13E4	164629.5371	1.2672
Log Likelihood		-891436.8212	

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Full Log Likelihood		-891436.8212	
AIC (smallerisbetter)		1782909.6425	
AICC (smallerisbetter)		1782909.6477	
BIC (smallerisbetter)		1783085.5881	

Tableau 23: Résultat de validation du modèle GLM pour le coût moyen

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	6.6223	0.0179	6.5872	6.6574	136619	<.0001
SEXE	F	1	-0.1003	0.0071	-0.1142	-0.0864	200.17	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
AGE_CLASS	8	1	-0.1696	0.0153	-0.1996	-0.1397	122.98	<.0001
AGE_CLASS	9	1	-0.0859	0.0091	-0.1036	-0.0681	89.90	<.0001
AGE_CLASS	10	1	-0.0588	0.0086	-0.0756	-0.0420	46.87	<.0001
AGE_CLASS	12	1	-0.1472	0.0151	-0.1767	-0.1177	95.61	<.0001
AGE_CLASS	14	1	-0.0612	0.0113	-0.0833	-0.0392	29.58	<.0001
AGE_CLASS	15	0	0.0000	0.0000	0.0000	0.0000	.	.
PENS_CLASS	4	1	-0.5096	0.0161	-0.5412	-0.4781	1004.11	<.0001
PENS_CLASS	6	1	-0.1568	0.0099	-0.1762	-0.1375	252.50	<.0001
PENS_CLASS	8	1	-0.3579	0.0152	-0.3877	-0.3282	556.37	<.0001
PENS_CLASS	9	1	-0.2430	0.0126	-0.2677	-0.2183	372.25	<.0001
PENS_CLASS	14	1	0.2711	0.0201	0.2316	0.3106	181.19	<.0001
PENS_CLASS	16	1	-0.1044	0.0103	-0.1247	-0.0841	101.88	<.0001
PENS_CLASS	17	0	0.0000	0.0000	0.0000	0.0000	.	.
SIT_FAM	C	1	-0.1576	0.0386	-0.2333	-0.0820	16.66	<.0001
SIT_FAM	D	1	-0.1274	0.0378	-0.2016	-0.0532	11.33	0.0008
SIT_FAM	M	1	-0.0873	0.0117	-0.1103	-0.0643	55.27	<.0001
SIT_FAM	V	0	0.0000	0.0000	0.0000	0.0000	.	.
ALD_ALC	Non	1	-0.4436	0.0110	-0.4651	-0.4220	1627.46	<.0001
ALD_ALC	Oui	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	0.8905	0.0030	0.8845	0.8965		

Tableau 24: Sortie SAS du résultat GLM pour le coût moyen

d) Résultat GLM de la 2ème approche

La modélisation GLM pour les atteints d'une ALD montre qu'il n'y a aucune différence significative entres les modalités des variables tarifaires, à part pour la variable SEXE. les variables tarifaires n'ont ainsi aucune influence sur le coût. Cela semble bien logique. Une Affection de longue durée (prenons le diabète par exemple) nécessite un même traitement tel que soit les caractéristiques de la personne(l'âge ou la situation familière etc.) ; le coût du traitement étant le même.

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		DDL	Valeur estimée	Erreur type	Wald 95% intervalle de confiance		Khi-2 de Wald	Pr > Khi-2
Intercept		1	6.3383	0.0533	6.2338	6.4427	14144.1	<.0001
SEXE	F	1	-0.1428	0.0150	-0.1722	-0.1134	90.75	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
C_AGE_CLASS	12	1	0.0875	0.0455	-0.0018	0.1768	3.69	0.0547
C_AGE_CLASS	13	0	0.0000	0.0000	0.0000	0.0000	.	.
C_PENS_CLASS	6	1	-0.0202	0.0249	-0.0690	0.0287	0.65	0.4188
C_PENS_CLASS	7	0	0.0000	0.0000	0.0000	0.0000	.	.
SIT_FAM	C	1	-0.0558	0.0322	-0.1189	0.0072	3.01	0.0826
SIT_FAM	V	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.8557	0.0241	1.8091	1.9034		

Le modèle n'inclura donc que le sexe pour la modélisation :

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	1E4	5962.1935	0.5865
Scaled Deviance	1E4	11058.0595	1.0877
Pearson Chi-Square	1E4	8252.3003	0.8118
Scaled Pearson X2	1E4	15305.5125	1.5056
Log Likelihood		-73328.5127	
Full Log Likelihood		-73328.5127	
AIC (smaller is better)		146663.0253	
AICC (smaller is better)		146663.0277	
BIC (smaller is better)		146684.7063	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	6.3699	0.0097	6.3509	6.3890	428583	<.0001
SEXE	F	1	-0.1442	0.0147	-0.1729	-0.1154	96.56	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.8547	0.0240	1.8082	1.9024		

La même procédure sera appliquée sur les deux autres catégories d'ALD_ALC. Les résultats sont en Annexe.
Le tarif final (coût moyen × fréquence) est la somme probabilisée des 3 tarifs liés aux catégories de l'ALD que nous avons distinctement modélisés.

1.3 Exploitation du GLM : Table de score

Les résultats de la modélisation de la fréquence et le coût moyen étant exhibés, nous avons utilisé ceux-ci pour créer une table de score qui définit la dépense estimée (Fréquence × Coût moyen) pour chaque cas possible de bénéficiaire.

Nous rappelons que les variables tarifaires sont :

- Classe d'âge
- Le sexe
- La situation familiale
- Présence d'ALD
- Classe de pension

Chaque combinaison possible entre les modalités de ces variables correspond à une dépense remboursée médicale estimée. Puisque nous avons besoin de l'utiliser pour calculer l'engagement médical sous la norme IAS19.L'estimation devra se faire pour chaque âge. Ce qui a été pris en considération.

Le logiciel SAS nous permet facilement de prédire ces dépenses-là. Il ne fallait que préparer notre carte de score.

Nous montrerons dans la figure ci-dessous la table de score pour le 1er modèle de tarification utilisé.

AGE_BENEFI	SEXE	SIT_FAM	PENS_CLASS	ALD_ALC	Remboursement estimé
16	F	C	4	Non	397,0007725
17	F	C	4	Non	397,0007725
19	M	V	4	Non	1653,432282
19	M	V	8	Non	1918,912927
20	F	V	6	Non	2270,064244
20	M	V	8	Non	1924,710521
20	M	V	16	Non	2568,274509
21	F	V	6	Non	2270,064244
21	M	V	16	Non	2568,274509
22	F	V	6	Non	2270,064244
22	M	V	8	Non	1666,11834
23	F	V	8	Non	1568,758638
23	F	V	9	Non	1753,270637
23	M	V	4	Non	1517,641114
23	M	V	8	Non	1924,710521
24	F	V	6	Non	2270,064244
24	M	V	8	Non	1924,710521

. . . .

60	F	C	6	O_N	2145,297889
60	F	C	16	Non	928,9918844
60	F	C	16	Oui	3184,454775
60	F	C	17	Non	1030,97494
60	F	D	4	Non	829,697767
60	F	D	6	Non	1638,33623
60	F	D	8	Non	1114,002359
60	F	D	9	Non	1576,497195
60	F	D	16	Non	1486,490477
60	F	D	17	Non	1649,674724
60	F	M	4	Non	733,5694553
60	F	M	4	O_N	1735,588935
60	F	M	4	Oui	2515,338998

Figure 40: Exemple de la table de score issue des résultats GLM de la 1ère approche de tarification

1.4 l'apprentissage statistique pour la tarification

La modélisation de la fréquence et le coût moyen est généralement faite par des modèles GLM pour tous ces avantages. Pour notre cas, les modèles n'ajustaient pas idéalement la distribution historique des coûts et fréquence dont nous disposions. L'intérêt de cette partie est d'explorer la possibilité d'améliorer notre modèle statistique de tarification en utilisant les techniques d'apprentissage statistique.

Comme nous l'avons cité plus haut, il existe des modèles de régression non linéaire issus de l'apprentissage statistique. Les forêts aléatoires entre autres, ont prouvé leur efficacité par rapport au modèles linéaires généralisé surtout en présence de grands volumes de données (méta Data).

Il est à rappeler que la base de données sur laquelle nous travaillons est celle des bénéficiaires de la population des pensionnées OCP. Les données que nous possédons sont de l'ordre de 62000 observations (Les observations sur les deux années 2016 et 2017).

1.4.1 Les forêts aléatoires (Random Forest-RI) : implémentation

Nous implémenterons la méthode d'ensemble forêts aléatoires sur SAS dans un premier lieu. L'avantage c'est qu'elle nous produira un fichier binaire qui calcule les prédictions. Il suffira d'exploiter celui-ci pour faire notre carte de score. Puis nous utiliserons SAS-Miner pour comparer les performances avec ceux d'un arbre de décision simple. Dans ce qui suit, nous exhiberons les résultats pour le modèle où l'ALD-ALC est incluse comme variable tarifaire. La même procédure sera poursuivie pour le cas où on ne la considère pas comme variable tarifaire. Faute de redondance, nous l'exhiberons pas dans ce sous chapitre.

```
proc hpforest data = BD.RET_

 maxtrees = 500 vars_to_try=4

 seed=600 trainfraction=0.6

 maxdepth = 50 leafsize = 6

 alpha =0.1 ;

 target cout_ecret / level= interval ;
 input AGE_BENEFI PENSION  / level= interval ;
 input SEXE ALD_ALC SIT_FAM / level=nominal ;

 ods output fitstatistics = fitstats ;

 save file ='C:\Users\imane\Desktop\S5\PFE\Partie Actuariat\RET\coutPourScorSA.bin';
 run;
```

L'intérêt de montrer ici un exemple de la procédure SAS utilisée est d'expliciter les différentes options utilisées pour implémenter notre Random Forest RI :

- **TRAINFRACTION** spécifie la fraction des observations réelles utilisée pour le Bootstrapping de chaque arbre. Nous choisissons 60%
- **SEED** option de la randomisation pour le Bootstrapping
- **MAXTREES** spécifie le nombre maximum d'arbres à utiliser. Nous choisissons 500 arbres
- **VARS_TO_TRY** : Puisque nous travaillons avec le Random Forest RI, cette option spécifie le nombre parmi lequel les variables seront aléatoirement choisies à chaque nœud. Nous choisissons 4 (< 5 nombre total des variables explicatives)
- **LEAFSIZE** indique le nombre minimum d'observation qu'on pourra tolérer pour chaque branche
- **ALPHA** spécifie la p-value seuil qu'une variable candidate devrait avoir pour qu'un nœud soit divisé. La p-value que nous avons choisi est 1%

- **MAXDEPTH** spécifie le nombre maximal de règles de division pour un nœud.

1.4.2 Modélisation de la fréquence

1) Résultats

Notons que les forêts aléatoires perdent cette propriété d'interprétabilité puisqu'il n'existe pas d'arbre associé à une forêt, cependant les sorties montrent l'amélioration de l'erreur OOB au fur et à mesure que nous augmentons le nombre d'arbres.

Fit Statistics			
Nombre d'arbres	Nombre de feuilles	Moyenne Carré Erreur (Données complètes)	Moyenne Carré Erreur (OOB)
496	27239	44.0891	44.2479
497	27300	44.0891	44.2479
498	27350	44.0891	44.2479
499	27396	44.0893	44.2480
500	27450	44.0894	44.2482

Tableau : L'erreur OOB par nombre d'arbres de décision utilisés

N.B : Les tableaux entier sont en ANNEXE

Nous prenons compte aussi de l'importance des variables explicatives durant la construction de la forêt. Nous remarquons que la situation ALD_ALC et la situation familiale sont les plus importante lors du découpage des premiers nœuds des arbres, suivi du sexe, du niveau de la pension et finalement l'âge du bénéficiaire.

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
ALD_ALC	2284	4.830739	3.206613	0.277227	0.184376
SIT_FAM	3379	2.246581	1.489616	0.098478	0.065206
SEXE	3535	0.227744	0.146249	0.011757	0.007533
PENSION	7844	0.323584	0.136866	0.021250	0.010101
AGE_BENEFI	9908	0.283042	0.119398	0.020476	0.009971

Tableau : Classement des variables par le critère VI (Variable Importance)

Le résultat de la forêt aléatoire sera enregistré dans un fichier binaire qui servira à la prédiction ou notre cas créer une carte de score.

2) Une petite analyse comparative

Sur SAS-Miner, il est possible de comparer les performances des modèles. Nous l'utiliserons pour comparer la forêt aléatoire avec un simple arbre de décision.

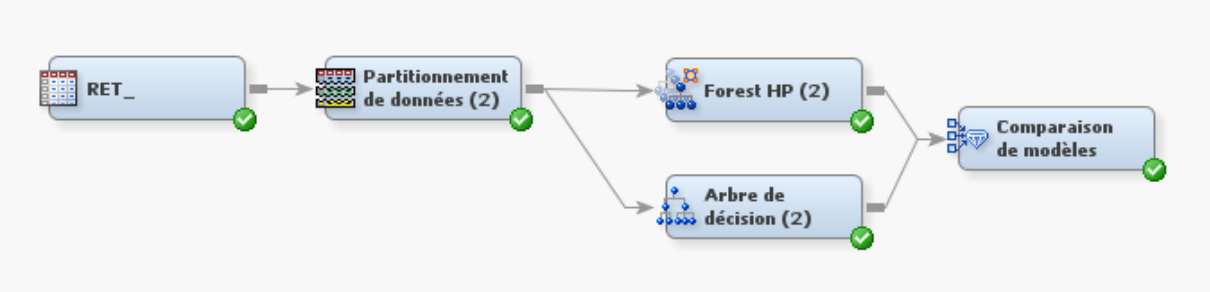


Figure : Diagramme affichant les étapes précédant la comparaison des modèles sur SAS-Miner

Modèle sélectionné	Noeud prédécesseur	Noeud du modèle	Description du modèle	Apprentissage: Target Variable	Libellé cible	Critère de sélection : Validation: Erreur quadratique approchée (ASE : Average Squared Error)
Y	Tree2	Tree2	Arbre de dé...	freq		44.5363
	HPDMFore...	HPDMFore...	Forest HP (2)	freq		44.95036

Tableau : Comparaison de l'erreur OOB des RF et un arbre de décision

Il parait que l'utilisation d'un seul arbre de décision minimise aussi bien l'erreur OOB

1.4.3 Modélisation du coût moyen

Avec le même mode opératoire, nous obtenons un fichier binaire qui servira pour la calcul des prédictions. Nous obtenons les outputs suivants :

Fit Statistics			
Nombre d'arbres	Nombre de feuilles	Moyenne Carré Erreur (Données complètes)	Moyenne Carré Erreur (OOB)
496	24868	173541	174219
497	24915	173542	174219
498	24958	173542	174219
499	25011	173542	174220
500	25059	173542	174219

Tableau : Minimisation de l'erreur OOB par ordre croissant du nombre d'arbres de décision utilisés (les derniers500). Modélisation coût moyen

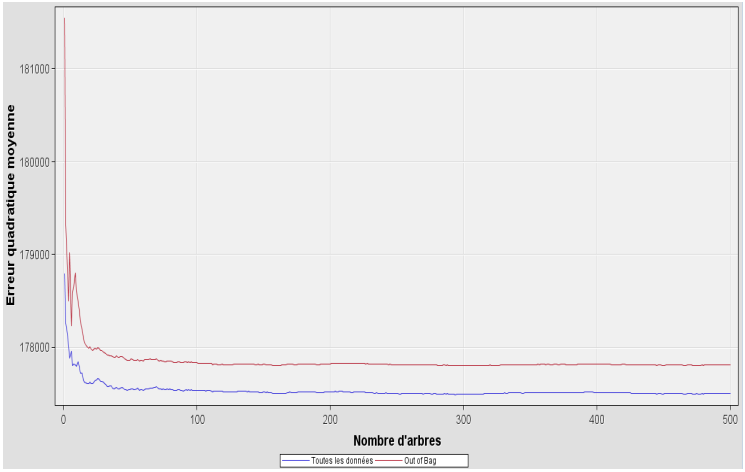


Figure: Amélioration de l'erreur OOB

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
ALD_ALC	1183	7574.536	5040.687	7.104930	4.718842
PENSION	9783	2304.158	1175.011	2.765263	1.571656
SEXE	2442	797.223	511.470	0.575419	0.374529
AGE_BENEFI	7182	634.209	180.297	0.356177	0.063805
SIT_FAM	3969	312.134	109.681	0.411349	0.202924

Tableau : Classement des variables par le critère VI (*Variable Importance*)

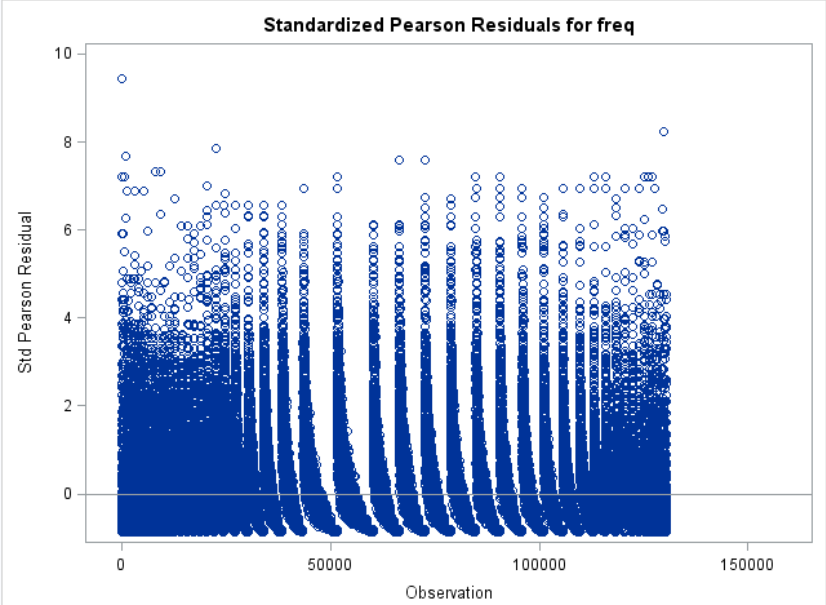
Dans le cas de la modélisation du coût, nous voyons que cette fois-ci, c'est naturellement la situation ALD_ALC et le niveau de la pension qui pèsent le plus dans la détermination de ce coût moyen. Suivi du sexe, de l'âge et finalement la situation familiale.

1.5 GLM et Machine Learning : Comparaison des performances

La question qui se pose est si l'utilisation des méthodes de Machine Learning comme les forêts aléatoires peuvent améliorer les estimations que nous utiliserons. Il faudra donc user des moyens de comparaison connues : les résidus. Nous n'exhiberons ici que les résultats concernant la 1ère approche de tarification (ALD incluse comme variable tarifaire)

1.5.1 Analyse Graphique

1) Analyse des résidus et des résidus standardisés de Pearson pour le modèle GLM de la fréquence et le coût moyen.



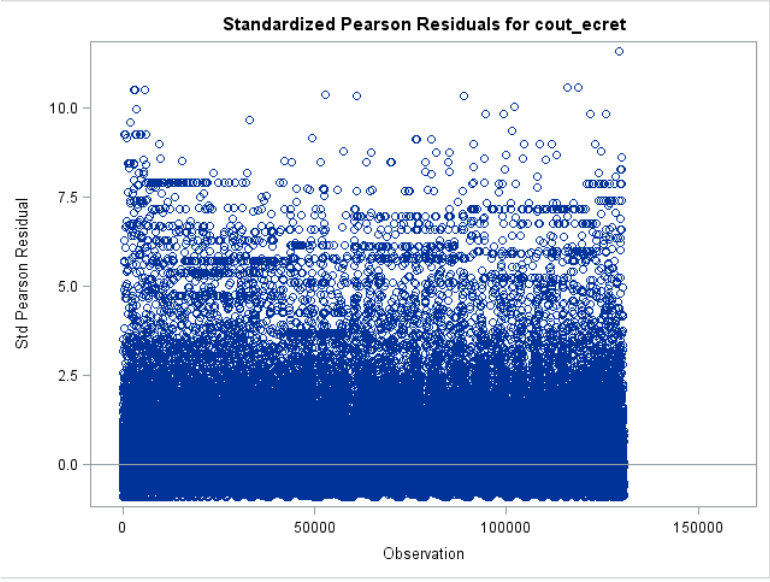


Figure : Analyse graphique des résidus simple les résidus standardisés du modèle de la fréquence et le coût moyen. Sortie SAS

2) Performance de la forêt aléatoire

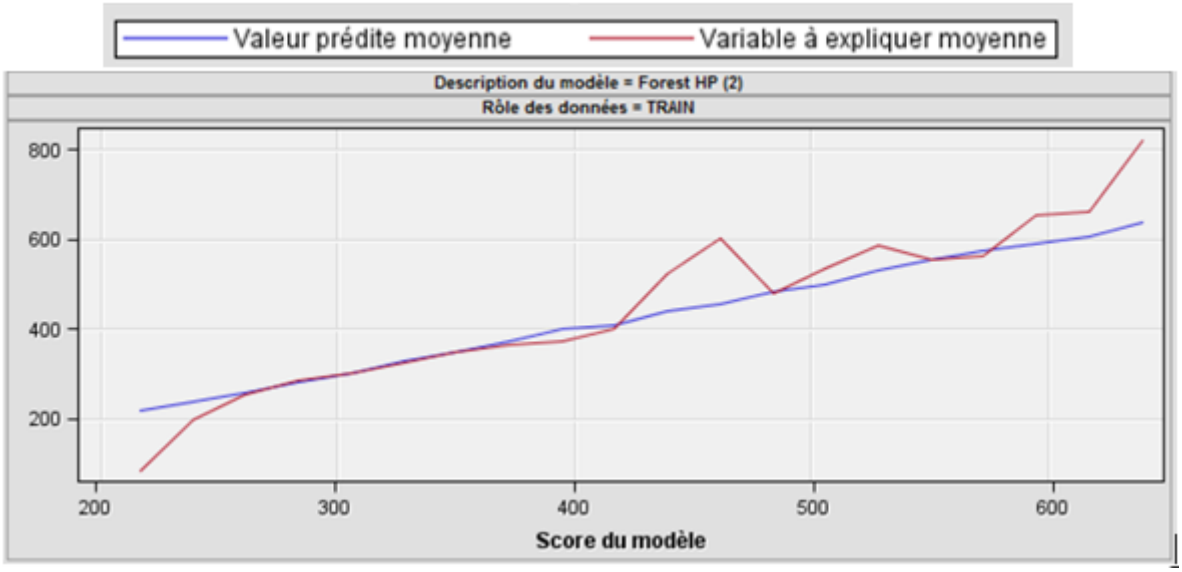


Figure : Score de la forêt aléatoire pour les données d'apprentissage

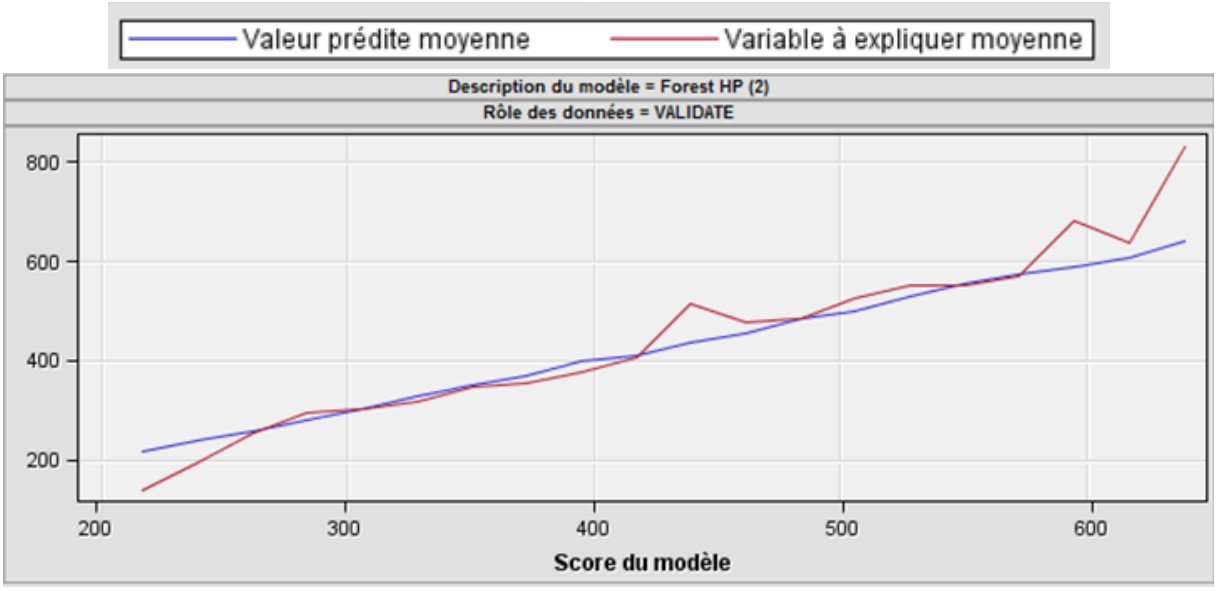


Figure : Performance de la forêt aléatoire sur les données de validation du modèle

1.5.2 Calcul des MSE de chaque modèle de tarification

Les graphiques sont très utiles pour voir d'un coup d'œil la distance entre valeurs observées et valeurs calculées par le modèle. Pourtant, on a besoin aussi de valeurs numériques comme mesures simples de cette distance.

MSE est probablement le critère quantitatif le plus utilisé pour comparer valeurs calculées et valeurs observées, et donc, entre les modèles.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

Avec

N : nombre d'observations

$\hat{Y}_i$: Notre tarif technique ou dépense médicale estimée

Y_i : Dépense médicale réelle

Nous calculerons donc la MSE pour les deux approches, celui où la variable ALD_ALC est incluse comme variable tarifaire, et l'autre ù ce n'est pas le cas. Nous obtenons :

	MSE du modèle de tarification avec GLM	MSE du modèle avec Random Forest
1ère approche de tarification (ALD_ALC incluse comme variable tarifaire)	17,40	17,11
2ème approche de tarification (probabilisation du tarif suivant la catégorie de l'ALD)	18,24	18,41

- La 1ère approche de tarification est meilleure en termes de précision par rapport à la 2ème approche. Ce qui est normale car nous avons introduis l'ALD comme variable tarifaire ce qui augmente la précision du modèle.
- Nous obtenons une MSE avec la forêt aléatoire qui est meilleure que celle du GLM dans la 1ère approche de la tarification. Le contraire est perçu pour la 2ème approche.

Conclusion

Chapitre 2 : Calcul de l'engagement médical de l'OCP

Il est à rappeler que dans le cadre d'IAS 19, la population étudiée doit être considérée comme étant fermée. Ceci permet de caractériser le volume des personnes ayant droit aux prestations et la durée sur laquelle ces personnes auront accès aux prestations. Pour les avantages post-emploi les prestations sont accessible dès la retraite et jusqu'à extinction du bénéficiaire.

La population recensée à la date d'évaluation atteint les 98603 bénéficiaire (dont les 62000 retraités qui bénéficient pleinement de ces avantages, prorata = 1), qu'on devra leur provisionner pour leurs avantages post-emploi en matière de couverture médicale. Les montants de cotisations (ou tarif estimé) seront issus des scores que nous avons retrouvés dans le chapitre précédent selon la 1ère approche de tarification qui nous a donné une meilleure précision.

Nous expliciterons dans ce qui suit les hypothèses démographiques et financières utilisés, avant de pouvoir entamer nos calculs automatisés sur Excel et Excel VBA afin de calculer l'engagement et la duration du régime pour pouvoir définir le taux d'actualisation adéquat.

2.1. Les hypothèses de calcul

2.1.1 Les hypothèses démographiques

Les hypothèses démographiques traduisent les caractéristiques démographiques actuelles et futures du personnel actif ou retraité (ou de leurs ayants-droit) réunissant les conditions requises pour bénéficier des avantages du régime.

La mortalité du personnel actif et retraité	Table TV-73-77
Le taux de rotation du personnel	1%
L'âge de départ à la retraite	60 ans

Tableau 25 : Récapitulatif des hypothèses démographiques

2.1.2 Les hypothèses financières

1) Le taux d'actualisation

Pour déterminer les taux d'actualisation que nous utiliserons pour actualiser les différents flux des engagements de l'entreprise, nous retiendrons la courbe des taux zéro-coupon pour la date d'évaluation adéquate. Elle sera générée à partir de l'outil Excel que nous avons élaboré dans la deuxième partie de ce mémoire.

Si dans le cas de l'utilisation d'une structure de taux réelle, on actualise un flux à partir du taux d'actualisation ayant la même maturité que ce flux, dans le cas de l'utilisation d'un taux moyen comme nous ferons, il faut déterminer une maturité moyenne pour l'engagement global de l'entreprise. La notion qui semble alors la plus naturelle, pour caractériser cette durée, est la notion de « duration » des engagements (ou duration de l'obligation). L'application d'un taux moyen conduit à des engagements très proches de ceux obtenus avec une structure de taux réelle. Ce résultat provient des fortes maturités des engagements (supérieurs à 15 ans) , qui correspondent à des taux se trouvant sur la partie plate de la courbe des taux utilisée.

Donnons une acception de la duration :

a) Duration

La duration se définit comme étant le barycentre des maturités des flux financiers du régime.

$\boldsymbol{F(x,i)}$: Flux financier actuel et futur probable du régime d'un individu n (de maturité *i*)
$\boldsymbol{a_i}$: taux d'actualisation retenu pour actualiser les flux à leur date de maturité i

$$\boldsymbol{Duration = \frac{\sum_{n=1}^{N}\sum_{i=0}^{w_x_n} iF_{(x,i)}/(1+a_i)^i}{\sum_{n=1}^{N}\sum_{i=0}^{w_x_n} F_{(x,i)}/(1+a_i)^i}}$$

Dans notre cas d'étude d'un régime de couverture médicale, la formulation appropriée pour calculer la duration, pour une structure des taux plate, est alors la suivante :

$$\boldsymbol{Duration = \frac{\sum_{n=1}^{N}\sum_{i=0}^{w_x_n} iF_{(x,i)}/(1+a)^i}{\sum_{n=1}^{N}\sum_{i=0}^{w_x_n} F_{(x,i)}/(1+a)^i}}$$

b) Méthode de détermination du taux d'actualisation

En calculant la duration des engagements à partir de la formule et des hypothèses initiales (avec un taux d'actualisation choisi arbitrairement). Supposant que ce taux est de 4% ; on obtient une duration de 19 ans. En se reportant à la courbe des taux d'actualisation, on obtient un nouveau taux d'actualisation correspondant à cette maturité qui est de 4,5%. On recalcule la duration

avec ce taux qui est alors égale à 18 ans. On obtient à nouveau un taux d'actualisation de 4,5% qui est donc le taux d'actualisation final que nous prendrons.

2) Le taux d'inflation médicale

Le taux d'inflation médicale retenu sera de 1,16% que nous considérerons fixe durant la projection des dépenses médicales.

2.2 Calcul de l'engagement

2.2.1 La formulation

Il convient de rappeler que le calcul se fait tête par tête, que la population est fermée et que nous utiliserons un seul taux d'actualisation moyen déterminée à partir de la duration du régime de frais de santé pour les avantages postérieurs à l'emploi.

A la lumière des formules de calculs cités dans les normes IAS19, nous explicitons les formules de calculs utilisé pour calculer le provisionnement pour les avantages postérieurs à l'emploi :

Il est à rappeler que la VAPF est calculée selon une méthode appelée « méthode des unités de crédits projetés » ou Projected Unit Credit Method (PUCM).

$$VAPF_{t_0}^{n} = \sum_{n=1}^{N} \sum_{i=0}^{w_x_n} F_{(x,i)} / (1+a)^i$$

Où

$F_{(x,i)} = p_x^{x+i} \times CM_{x+i} \times (1+inf)^i$

x_n : l'âge du bénéficiaire n à la date d'évaluation

inf : taux d'inflation médical

p_x^{x+i} : probabilité de survie entre $x_n$et $x_n + i$

CM_{x+i} : La consommation médicale estimé du bénéficiaire n à l'âge x_n+ i

w : Age maximal de survie (106 ans dans le table TV73-77)

Nous calculerons la DBO pour chaque tête, avant de sommer le tout et retrouver le montant à provisionner, après avoir trouvé le taux d'actualisation adéquat :

$$DBO_{t0}^{n} = VAPF_{t0}^{n} \times \frac{Ancienneté actuelle_{t0}^{n}}{Ancienneté à terme^{n}} \times (1 - tr)$$

2.2.2 Résultats

Concernant Le calcul sur Excel, il concernait le calcul de la VAPF puis la DBO pour un bénéficiaire quelconque, étant donné ses caractéristiques.

Le calcul était automatisé par la suite afin de calculer les VAPF pour chacun des bénéficiaires et futurs bénéficiaires au titre des avantages postérieurs à l'emploi. (soit pour les 98603 individus recensés à la date du 01/07/2017)

Pour le cas du GLM, le taux d'actualisation moyen avec lequel nous avons initié le calcul était de 4,5%, nous obtenons une duration de 14,81 ans. Ce qui correspond à un taux d'intérêt dans courbe ZC augmenté d'une prime de risque de 4,12%, nous relançons le calcul avec ce taux la, et nous retrouvons une duration de 15 ans. Le taux correspondant est de 4,12%. C'est le taux moyen adéquat. Le bon choix du taux moyen au départ nous a permis de rapidement retrouver le taux adéquat. La même procédure sera appliquée aussi pour le cas de la 2ème modélisation.

Nous obtenons donc :

	Tarification avec GLM	Tarification avec Random Forest
DBO	1 993 631 797,90 Dhs	2 161 803 689,14 Dhs

Tableau : Engagement DBO calculé avec les deux modèles (GLM et Random Forest)

En utilisant la carte de score issue du modèle GLM de la première approche, nous retrouvons en définitive un engagement (n'incluant pas les frais de gestion et autres et en ajoutant l'engagement des enfants (11%)) une DBO égalant 1 993 631 797,90 de Dhs. Soit environ 2 milliards de dirhams que l'OCP devra provisionner pour financer le régime de frais de santé relevant des avantages postérieurs à l'emploi. Le montant de la DBO que nous retrouvons par contre si on utilise une carte de score issue du modèle Random Forest est de 2 161 803 689,14 Dhs.

Conclusion générale

Nous avons pu constater à travers ce mémoire que la norme IAS19, relative aux avantages du personnel, dicte les méthodes et les paramètres nécessaires pour le calcul des engagements sociaux. Le montant des cotisations (ou de la consommation médicale estimée) étant le paramètre le plus important à définir.

La norme IAS19 possède en outre un grand nombre de spécificités techniques que nous étions dans l'obligation de déterminer ou de modéliser comme nous l'avons fait avec quelques hypothèses financières et démographiques trouvées à notre portée. Le choix de ces paramètres (choix du taux d'actualisation, taux de rotation, mortalité, prorata d'ancienneté etc.) était décisif quant à la détermination de l'engagement social que nous avons retrouvé vers la fin de ce mémoire, tout en utilisant les diverses méthodes de modélisation de la consommation médicale puis comparer les résultats entre eux.

Le taux d'actualisation moyen se définie à partir de la courbe des taux ZC augmentée d'une prime de risque, ou comme le stipule la norme, « se référer aux taux des obligations d'État si le marché des obligations de 1ère catégorie n'est pas actif ». Cette courbe ZC n'était pas possible à modéliser faute de manque de données pratiques ici au département de prévoyance. Nous avons exploré tout de même la théorie et nous avons implémenté un outil de lissage des divers courbes de taux à titre d'illustration, tandis que la courbe des taux ZC avec laquelle nous avons évaluer l'engagement médical était celle utilisée par l'OCP.

Le taux d'inflation était aussi un paramètre que nous devions intégrer dans l'évaluation, nous avons usé cette fois-ci des données sur l'inflation médicale nationale pour chacune des garanties santé comme les soins ambulatoires, l'hospitalisation, les frais de matériels et transports etc. Le résultat aurait été plus affine si nous avions à utiliser les données internes de l'OCP quant à la consommation médicale pour chaque garantie santé, mais nous n'avons pas reçu cette information. Le taux que nous avons retrouvé pour l'année 2017 était cependant proche de celui utilisé par l'OCP pour la même année (1,16% contre 1%).

L'étape suivante était d'estimer la consommation médicale. La méthode classique de la détermination des cotisations (qu'elle soient versées par les bénéficiaires ou par l'OCP elle-même au profit de ses bénéficiaires) est de tarifier le contrat pour le régime frais de santé. Nous avons dans un premier lieu utilisé le GLM très prôné pour ses résultats de modélisation, puis nous avons exploré une technique avancée de tarification qui use des prouesses de la Big Data : l'apprentissage statistique ou Machine Learning. Nous avons utilisé spécifiquement la technique des forêts aléatoires à cette fin. Cette technique, comme tout autre algorithme de classification avancé, s'impose d'autant plus que les données prennent masse.

La dernière étape était d'évaluer l'engagement médical de l'OCP qui s'inscrit dans le cadre de la dette long terme au titre des avantages postérieurs à l'emploi (et donc concernant les retraités). Tous les paramètres étaient définis, nous retrouvons un engagement premier (sans prendre compte des frais supplémentaires comme les frais de gestion) était de l'ordre d 2 milliards de dirhams.

Bibliographie

IFRS Foundation, *Norme comptable internationale 19 Avantages du personnel.*

Bernard BRICE, *L'évaluation des engagements IAS 19 des régimes de frais de santé*, Euro Institut D'actuariat, 2011

Chaoubi, *Modèles linéaires Généralisés (GLM)*, cours INSEA, Rabat, 2017

Marri, Fouad, *Assurance non vie 2*, cours INSEA, Rabat, 2017

Robin Genuer, Jean-Michel Poggi, *Arbres CART et Forêts aléatoires, Importance et sélection de variables,* 2017. <hal-01387654v2>

Frédéric PLANCHET, Antoine MISERAY, *Tarification IARD ,Introduction aux techniques avancées,* cours ISFA, France, 2017

Franck Vermet (EURIA) et Jean-Marc Aouizerate (Aon France), « *Mieux prévoir la charge sinistre en santé ?* » ,Colloque, 2015.

Arthur Charpentier. *Statistique de l'assurance. 3rd cycle*. Université de Rennes 1 et Université de Montréal, 2010, pp.133. <cel-00550583>

TAGUEMOUNT, Asma, *La structure par terme des taux d'intérêt, gestion et modélisation des titres financiers à revenu fixes,* cours ENSSEA, France.

BANK-AL-MAGHRIB,taux de référence des bons de trésor, marché obligataire/marché des bons de trésor/ Marché secondaire.

Ministère de la santé, *Rapport de la santé,* Maroc,2013

Site HCP, IPC santé par section 2008-2017.

SAS community Help.

ANNEXES

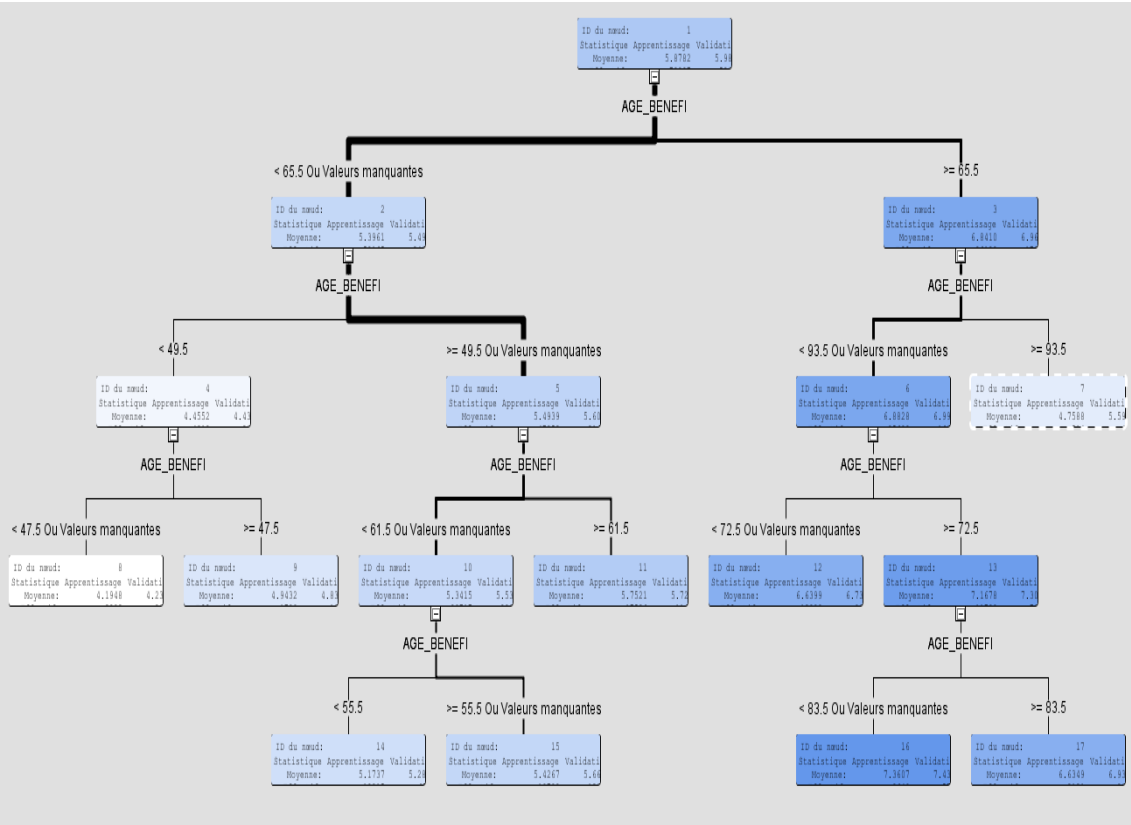
ANNEXE A : La courbe de taux ZC + spread OCP

ANNEXE B : Fonctions VBA du lissage de la courbe

Interpolation linéaire :

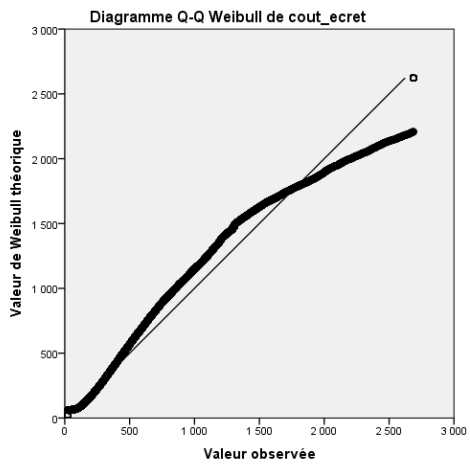
```
Function LinInterpolate(xVal As Double, _
xRange As Variant, _
yRange As Variant, _
Optional isSorted As Long = 1) As Double
Dim yVal As Double
Dim xBelow As Double, xAbove As Double
Dim yBelow As Double, yAbove As Double
Dim testVal As Double
Dim High As Long, Med As Long, Low As Long

Low = 1
High = xRange.Cells.Count

If isSorted <> 0 Then
Rem binary search sorted range
Do
Med = Int((Low + High) \ 2)
If (xRange.Cells(Med).Value) < (xVal) Then
Low = Med
Else
High = Med
End If
Loop Until Abs(High - Low) <= 1
Else
Rem search every entry
xBelow = -1E+205
xAbove = 1E+205

For Med = 1 To xRange.Cells.Count
testVal = xRange.Cells(Med)
If testVal < xVal Then
If Abs(xVal - testVal) < Abs(xVal - xBelow) Then
Low = Med
xBelow = testVal
End If
Else
If Abs(xVal - testVal) < Abs(xVal - xAbove) Then


High = Med
xAbove = testVal
End If
End If
Next Med
End If

xBelow = xRange.Cells(Low): xAbove = xRange.Cells(High)
yBelow = yRange.Cells(Low): yAbove = yRange.Cells(High)

LinInterpolate = yBelow + (xVal - xBelow) * (yAbove - yBelow) / (xAbove - xBelow)
End Function
```

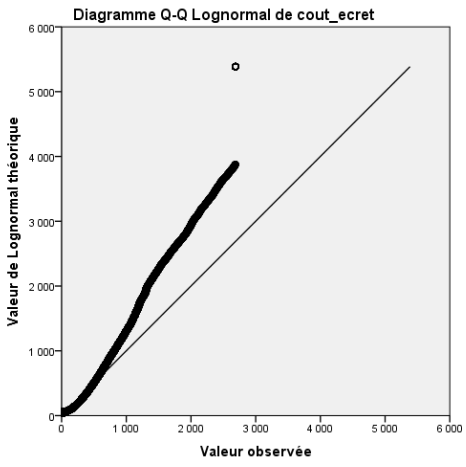
ANNEXE C : Exemple d'arbre de décision utilisé pour segmenter une variable tarifaire (ici la variable AGE_BENEFI suivant la fréquence) :



Le code est généré automatiquement par SAS Miner. Portable et utilisable dans la segmentation de la variable tarifaire dans la base de données.

ANNEXE D : Ajustement de la distribution historique du coût moyen par les différentes loi(Weibull et Lognormale)





ANNEXE E : 2ème approche de tarification , les résultats du GLM pour la catégorie ALD=Non et ALD=O_N

ALD =Non

Modélisation GLM du cout :

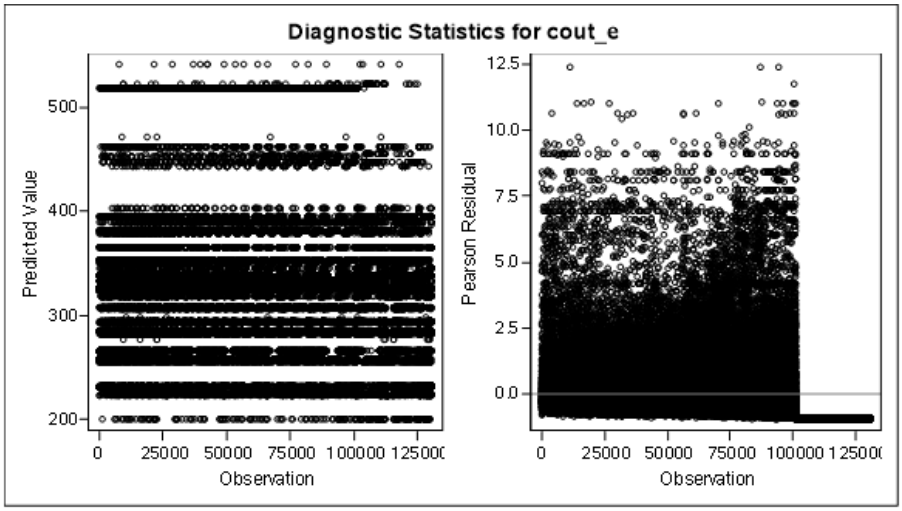
Informations sur le modèle	
Data Set	BD.RETN
Distribution	Gamma
Link Function	Log
Dependent Variable	cout_ecret

Number of Observations Read	130478
Number of Observations Used	108821
Number of Invalid Responses	21657

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
C_AGE_CLASS	4	4 5 6 7
C_PENS_CLASS	6	4 5 6 10 12 13

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	11E4	156167.0544	1.4352
Scaled Deviance	11E4	128671.7060	1.1825
Pearson Chi-Square	11E4	167988.5205	1.5439
Scaled Pearson X2	11E4	138411.8412	1.2720
Log Likelihood		-733692.3294	
Full Log Likelihood		-733692.3294	
AIC (smaller is better)		1467406.6588	
AICC (smaller is better)		1467406.6612	
BIC (smaller is better)		1467512.2309	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	5.9920	0.0310	5.9312	6.0528	37298.1	<.0001
SEXE	F	1	-0.1377	0.0075	-0.1523	-0.1231	340.61	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
C_AGE_CLASS	4	1	0.1101	0.0169	0.0771	0.1432	42.58	<.0001
C_AGE_CLASS	5	1	0.1022	0.0193	0.0644	0.1401	28.09	<.0001
C_AGE_CLASS	6	1	0.1444	0.0164	0.1122	0.1765	77.37	<.0001
C_AGE_CLASS	7	0	0.0000	0.0000	0.0000	0.0000	.	.
C_PENS_CLASS	4	1	-0.5524	0.0281	-0.6076	-0.4973	385.64	<.0001
C_PENS_CLASS	5	1	-0.3126	0.0270	-0.3656	-0.2596	133.65	<.0001
C_PENS_CLASS	6	1	0.1576	0.0338	0.0914	0.2237	21.80	<.0001
C_PENS_CLASS	10	1	-0.2341	0.0278	-0.2886	-0.1795	70.71	<.0001
C_PENS_CLASS	12	1	-0.1567	0.0279	-0.2114	-0.1020	31.52	<.0001
C_PENS_CLASS	13	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	0.8239	0.0031	0.8180	0.8299		



Modélisation de la fréquence

Informations sur le modèle	
Data Set	BD.RETN
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	freq

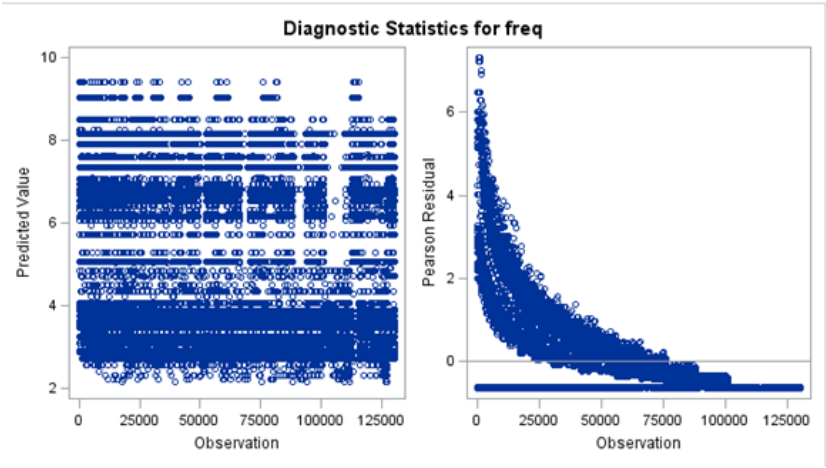
Number of Observations Read	130478
Number of Observations Used	130478

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
AGE_CLASS	2	7 9
PENS_CLASS	9	6 10 12 14 15 16 17 20 21
SIT_FAM	4	C D M V

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	13E4	136185.9122	1.0439
Scaled Deviance	13E4	136185.9122	1.0439
Pearson Chi-Square	13E4	105574.0942	0.8092
Scaled Pearson X2	13E4	105574.0942	0.8092
Log Likelihood		522293.1726	
Full Log Likelihood		-312445.3334	
AIC (smaller is better)		624920.6669	
AICC (smaller is better)		624920.6705	
BIC (smaller is better)		625067.3513	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	2.0332	0.0261	1.9821	2.0843	6070.33	<.0001
SEXE	F	1	-0.0729	0.0096	-0.0916	-0.0541	57.85	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
AGE_CLASS	7	1	-0.0398	0.0201	-0.0792	-0.0003	3.90	0.0482
AGE_CLASS	9	0	0.0000	0.0000	0.0000	0.0000	.	.
PENS_CLASS	6	1	-0.1166	0.0294	-0.1743	-0.0590	15.74	<.0001
PENS_CLASS	10	1	-0.1767	0.0214	-0.2187	-0.1347	68.08	<.0001
PENS_CLASS	12	1	0.1053	0.0109	0.0839	0.1267	93.06	<.0001
PENS_CLASS	14	1	-0.0777	0.0347	-0.1458	-0.0096	5.00	0.0253

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
PENS_CLASS	15	1	0.2077	0.0570	0.0960	0.3195	13.27	0.0003
PENS_CLASS	16	1	-0.3718	0.0427	-0.4554	-0.2881	75.86	<.0001
PENS_CLASS	17	1	-0.0942	0.0257	-0.1446	-0.0437	13.39	0.0003
PENS_CLASS	20	1	0.0753	0.0154	0.0451	0.1056	23.83	<.0001
PENS_CLASS	21	0	0.0000	0.0000	0.0000	0.0000	.	.
SIT_FAM	C	1	-0.9781	0.0592	-1.0941	-0.8621	273.11	<.0001
SIT_FAM	D	1	-0.5224	0.0566	-0.6333	-0.4115	85.26	<.0001
SIT_FAM	M	1	-0.7452	0.0172	-0.7789	-0.7114	1871.45	<.0001
SIT_FAM	V	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	2.2718	0.0116	2.2491	2.2948		



ALD =O_N

Modélisation de la fréquence

Informations sur le modèle	
Data Set	BD.RETO_N
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	freq

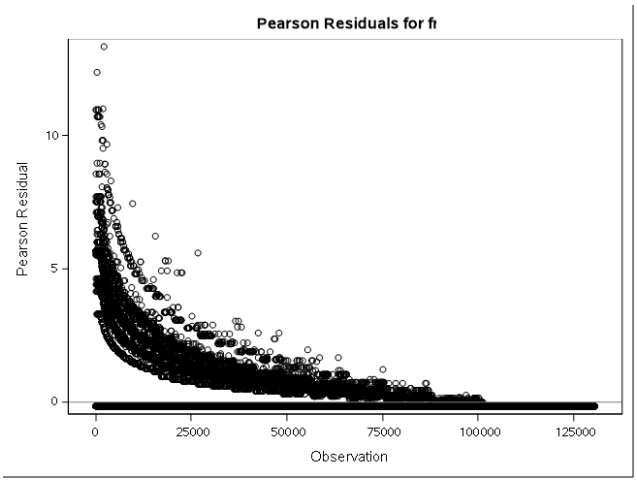
Number of Observations Read	130478
Number of Observations Used	130478

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
AGE_CLASS	4	4 5 6 7
PENS_CLASS	6	4 6 11 12 14 19
sit	2	M V

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	13E4	27114.0093	0.2078
Scaled Deviance	13E4	27114.0093	0.2078
Pearson Chi-Square	13E4	61149.9442	0.4687
Scaled Pearson X2	13E4	61149.9442	0.4687
Log Likelihood		161733.6369	
Full Log Likelihood		-80638.3449	
AIC (smaller is better)		161296.6898	
AICC (smaller is better)		161296.6915	
BIC (smaller is better)		161394.4794	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	0.5370	0.0642	0.4112	0.6628	69.98	<.0001
AGE_CLASS	4	1	-0.6566	0.0635	-0.7812	-0.5321	106.79	<.0001
AGE_CLASS	5	1	-0.3039	0.0491	-0.4001	-0.2076	38.26	<.0001
AGE_CLASS	6	1	-0.2263	0.0482	-0.3208	-0.1317	22.00	<.0001
AGE_CLASS	7	0	0.0000	0.0000	0.0000	0.0000	.	.
PENS_CLASS	4	1	-0.5308	0.0709	-0.6698	-0.3918	56.04	<.0001
PENS_CLASS	6	1	-0.5065	0.0682	-0.6402	-0.3728	55.16	<.0001
PENS_CLASS	11	1	-0.2831	0.0745	-0.4292	-0.1370	14.42	0.0001
PENS_CLASS	12	1	-1.3107	0.1664	-1.6369	-0.9845	62.02	<.0001
PENS_CLASS	14	1	-0.3293	0.0962	-0.5179	-0.1407	11.71	0.0006

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
PENS_CLASS	19	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	38.7699	0.4275	37.9411	39.6169		



Modélisation du coût moyen

Informations sur le modèle	
Data Set	BD.RETO_N
Distribution	Gamma
Link Function	Log
Dependent Variable	cout_ecret

Number of Observations Read	130478
Number of Observations Used	11489
Number of Invalid Responses	118989

Informations sur les niveaux de classe		
Classe	**Niveaux**	**Valeurs**
SEXE	2	F M
C_AGE_CLASS	5	4 5 7 8 9
C_PENS_CLASS	2	4 5
sit	2	M V

Critères d'évaluation de l'adéquation			
Critère	**DDL**	**Valeur**	**Valeur/DDL**
Deviance	11E3	6882.2186	0.5994
Scaled Deviance	11E3	12513.8427	1.0899
Pearson Chi-Square	11E3	10091.5814	0.8789
Scaled Pearson X2	11E3	18349.3827	1.5981
Log Likelihood		-83121.5806	
Full Log Likelihood		-83121.5806	
AIC (smaller is better)		166259.1613	
AICC (smaller is better)		166259.1738	
BIC (smaller is better)		166317.9545	

Paramètres estimés par l'analyse du maximum de vraisemblance								
Paramètre		**DDL**	**Valeur estimée**	**Erreur type**	**Wald 95% intervalle de confiance**		**Khi-2 de Wald**	**Pr > Khi-2**
Intercept		1	6.4772	0.0242	6.4298	6.5246	71760.8	<.0001
SEXE	F	1	-0.1202	0.0163	-0.1521	-0.0884	54.71	<.0001
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
C_AGE_CLASS	4	1	-0.0944	0.0316	-0.1563	-0.0326	8.95	0.0028
C_AGE_CLASS	5	1	-0.0813	0.0206	-0.1217	-0.0409	15.57	<.0001
C_AGE_CLASS	7	1	-0.1601	0.0422	-0.2428	-0.0774	14.40	0.0001
C_AGE_CLASS	8	1	-0.0701	0.0177	-0.1048	-0.0353	15.61	<.0001
C_AGE_CLASS	9	0	0.0000	0.0000	0.0000	0.0000	.	.
C_PENS_CLASS	4	1	-0.0648	0.0246	-0.1131	-0.0165	6.92	0.0085
C_PENS_CLASS	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.8183	0.0221	1.7754	1.8622		

ANNEXE F : 2ème approche de tarification : Outputs du Random Forest : VI

ALD =Non

modélisation du cout :

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
PENSION	8231	1509.316	780.6513	1.729193	0.968504
SEXE	1002	348.470	228.1816	0.193267	0.125445
AGE_BENEFI	6333	310.082	46.2721	0.304130	0.071103
SIT_FAM	3882	165.949	37.4760	0.364388	0.182261

modélisation de la fréquence :

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
SIT_FAM	2176	1.779561	1.175607	0.082442	0.054409
SEXE	1462	0.114758	0.073953	0.004547	0.002858
AGE_BENEFI	6465	0.118468	0.045496	0.012043	0.006071
PENSION	4592	0.089388	0.025075	0.007261	0.002836

ALD =Oui

modelisationfrequence

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
AGE_BENEFI	5311	0.053529	0.027393	0.007029	0.004203
PENSION	3855	0.034408	0.009309	0.005065	0.002692
SIT_FAM	1672	0.014016	0.007207	0.004840	0.003086
SEXE	2359	0.008253	0.003286	0.001522	0.000885

Modelisation cout

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
AGE_BENEFI	4080	166.1677	83.70917	0.446819	0.271213
SIT_FAM	558	114.0550	75.27227	0.440405	0.291396
PENSION	3827	157.7966	42.36912	0.397400	0.211258
SEXE	1871	13.6243	0.23814	0.050209	0.025625

BD.RET ALD=O_N

Modelisation cout

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
AGE_BENEFI	4079	174.4215	83.14012	0.455956	0.275652
SIT_FAM	896	77.1858	40.41538	0.278346	0.177258
SEXE	1140	27.8735	16.14553	0.075035	0.047866
PENSION	1420	35.6848	-0.59762	0.069677	0.028109

Modelisation frequence

Loss Reduction Variable Importance					
Variable	Nombre de règles	MSE	OOB MSE	Absolute Error	OOB Absolute Error
AGE_BENEFI	4918	0.053799	0.02593	0.006405	0.003705
SEXE	1318	0.006756	0.00332	0.000784	0.000436
SIT_FAM	2096	0.009124	0.00189	0.002334	0.001326
PENSION	3897	0.028540	-0.00142	0.002766	0.000831