



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN



Institut National de Statistique
et d'Economie Appliquée

RAPPORT DE PROJET DE FIN D'ETUDES

Organisme d'accueil : AXA ASSURANCE MAROC

Etude de la rentabilité du portefeuille Décès-
Incapacité-Maladie

Préparé par : ALOUAN Imad Eddine (Actuariat-Finance)
BENZINEB Mariam (Actuariat-Finance)

Sous la direction de : M. CHAOUBI Abdelaziz (INSEA)
M. MOUATASSIM Younes (AXA)

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

Diplôme d'Ingénieur d'Etat

Option : Actuariat-Finance

Devant le jury composé de :

M. LACHGAR Jamal (INSEA)

M. CHAOUBI Abdelaziz (INSEA)

M. MOUATASSIM Younes (AXA)

*Année universitaire
2013-2014*

Remerciements

Nos plus profonds remerciements vont tout d'abord à notre encadrant interne, Monsieur le professeur Abdelaziz CHAOUBI, pour la disponibilité dont il a fait preuve à notre égard, pour les conseils qu'il nous prodigués et pour ses encouragements.

Nous remercions également Mr. Jamal LACHGAR pour l'intérêt qu'il a porté à ce travail en acceptant de siéger parmi les membres du jury. Veuillez accepter l'expression de notre haute considération.

Nous tenons aussi à exprimer notre profonde gratitude envers Mr. Abderrahim DBICH, directeur marché des particuliers d'AXA Assurance Maroc, pour nous avoir accueillis au sein de son service.

Nous exprimons toute notre reconnaissance à Mr. Younes MOUATASSIM pour son encadrement et le soutien chaleureux dont il a toujours fait preuve.

Enfin, un grand merci à tous les membres de l'équipe actuariat et l'équipe de souscription pour leur convivialité, leur bonne humeur et leur coopération dans l'aboutissement de notre mission.

Dédicace

Je dédie ce travail,

A mes chers parents

Aucun mot, aucune dédicace ne saurait exprimer mon respect, ma considération et l'amour éternel pour les sacrifices que vous avez consentis pour mon instruction et mon bien être.

Votre générosité et votre bonneté ont toujours été un exemple pour nous tous.

Trouvez en ce travail le fruit de votre dévouement et l'expression de ma gratitude et mon profond amour.

A mes sœurs et frère,

Vous m'avez toujours soutenu durant toutes mes études, je vous souhaite beaucoup de bonheur et de réussite.

A tous mes amis et amies

ALOUAN Imad Eddine

Dédicace

Je dédie ce travail,

A mes très chers parents

Qui n'ont économisé aucun effort pour que je puisse être ce que je suis.

Aucun mot n'exprimera ma profonde gratitude pour tout l'amour, Le soutien, les sacrifices et la confiance qu'ils me font.

*A ma sœur Nawal et mes deux frères Moncef et Tarik,
A mon oncle Hassan pour son soutien qu'il n'a cessé d'apporter au cours de ma formation.*

A toute ma famille

A mon cher binôme IMAD, pour son travail remarquable, son soutien, et surtout pour avoir été le meilleur collègue et ami qu'on pouvait avoir.

Et à tous mes amis avec qui j'ai partagé de meilleurs souvenirs.

BEN ZINEB Mariam

RESUME

Le but de notre projet de fin d'étude est d'élaborer une étude de la rentabilité de la branche maladie d'AXA ASSURANCE MAROC constitué actuellement de plusieurs produits et ceci à travers les indicateurs S/P (Loss Ratio), coût moyen de sinistre, et la fréquence de sinistre. Ces indicateurs nous permettrons d'identifier l'ampleur des variables étudiées aux différents niveaux. Pour cela il a fallu établir en premier temps une modélisation des différentes composantes des risques liés à cette branche afin de modéliser la prime pure puis calculer et estimer les réserves nécessaires pour assurer l'équilibre de ce portefeuille.

Nous commencerons donc dans un premier lieu par une élaboration des différentes classes tarifaires via des variables tarifaires ensuite nous allons utiliser les modèles linéaires généralisés (**GLM**) pour la modélisation de la fréquence et du coût moyen des sinistres ordinaires et les distributions Pareto généralisées (**GPD**) pour la modélisation des sinistres graves des différentes garanties.

Après, nous essayerons de calculer les provisions pour sinistres à payer (PSAP) liés à ce portefeuille en utilisant plusieurs méthodes et approches déterministes et stochastiques.

Mots clés :

- S/P (Loss ratio)
- Segmentation
- Fréquence
- Coût moyen
- Prime pure
- GLM (Generalized linear model)
- POT (Peaks over threshold)

Liste d'abréviations

AD	Anderson Darling
AIC	Akaike Information Criterion
ALC	<i>Affection lourde et couteuse</i>
ALD	<i>Affection de Longue Durée</i>
AMC	Assurance Maladie Complémentaire
AMO	Assurance Maladie Obligatoire
ANAM	Agence National de l'Assurance Maladie
ANOVA	Analyse de la variance
BIC	Bayesian Information Criterion
CNOPS	Caisse nationale des organismes de prévoyance sociale
CNSS	Caisse Nationale de Sécurité Sociale
CVM	Cramer Von Mises
CHAID	Chi-squared Automatic Interaction Detector
DF	Degree of freedom (Degré De Liberté)
DH	Dirham
DIM	Décès Invalidité Maladie
FME	Fonction Moyenne des Excès
GLM	Modèles Linéaires Généralisés
GPD	Generalized Pareto distribution
INSEA	Institut national de statistique et d'économie appliquée
POT	Peaks over threshold
QQ-PLOT	Quantile-Quantile plot
RAMED	Régime d'Assistance Médicale
SAS	Statistical Analysis System
TNR	Tarif National de Référence
TVE	Théorie des valeurs extrêmes
ZINB	<i>Zero-inflated negative binomial</i>
ZIP	<i>Zero-Inflated Poisson</i>

LISTE DES TABLEAUX

<i>Tableau 1 : Types de garantie en cas de décès</i>	<i>14</i>
<i>Tableau 2 : Types de garantie en cas d'incapacité/Invalidité.....</i>	<i>15</i>
<i>Tableau 3 : Types de garantie en cas d'assurance maladie.....</i>	<i>16</i>
<i>Tableau 4 : Panier de soins de la CNOPS</i>	<i>17</i>
<i>Tableau 5 : Panier de soins de la CNSS.....</i>	<i>19</i>
<i>Tableau 6 : Chiffre d'affaire de la maladie des compagnies d'assurances (en 1000DHS).....</i>	<i>20</i>
<i>Tableau 7 : La part du marché de la maladie des différentes compagnies d'assurance</i>	<i>21</i>
<i>Tableau 8 : L'évolution du chiffre d'affaire de la maladie des différentes compagnies d'assurance</i>	<i>22</i>
<i>Tableau 9 : matrice de corrélation des variables.....</i>	<i>34</i>
<i>Tableau 10: Les statistiques AD et CvM des différents seuils.....</i>	<i>41</i>
<i>Tableau 11: Estimation des coefficients du modèle GPD des charges extrêmes</i>	<i>42</i>
<i>Tableau 12: Estimation du modèle Poisson pour les sinistres graves</i>	<i>43</i>
<i>Tableau 13: Estimation du modèle Binomiale Négative pour les sinistres graves</i>	<i>44</i>
<i>Tableau 14: Comparaison des modèle Poisson et Binomiale négative.....</i>	<i>44</i>
<i>Tableau 15 : Fréquence des médicaments pour l'âge.....</i>	<i>45</i>
<i>Tableau 16 : Fréquence des médicaments pour le lien.....</i>	<i>46</i>
<i>Tableau 17 : Fréquence des médicaments pour le sexe</i>	<i>46</i>
<i>Tableau 18 : Fréquence des médicaments pour la profession</i>	<i>46</i>
<i>Tableau 19 : Fréquence des médicaments pour le taux de remboursement.....</i>	<i>47</i>
<i>Tableau 20 : Fréquence des médicaments pour la situation matrimoniale</i>	<i>47</i>
<i>Tableau 22 : Tableau d'analyse de type 3 pour la fréquence des médicaments</i>	<i>49</i>
<i>Tableau 23 : AIC et la déviance des lois ZINB et Négative Binomiale</i>	<i>50</i>
<i>Tableau 24 : Le coût moyen des médicaments par âge.....</i>	<i>51</i>
<i>Tableau 25 : Le coût moyen des médicaments par sexe.....</i>	<i>51</i>
<i>Tableau 26 : Le coût moyen des médicaments par lien.....</i>	<i>51</i>
<i>Tableau 27: Le coût moyen des médicaments par taux de remboursement</i>	<i>51</i>
<i>Tableau 28 : Le coût moyen des médicaments par catégorie socio professionnelle...</i>	<i>52</i>
<i>Tableau 29 : Le coût moyen des médicaments par situation matrimoniale</i>	<i>52</i>
<i>Tableau 30 : Le coût moyen des médicaments par profession</i>	<i>52</i>
<i>Tableau 31 : Tableau d'analyse de type 3 pour le coût moyen des médicaments pour les variables restantes</i>	<i>54</i>

<i>Tableau 32 : Prime pure des médicaments selon l'âge et le taux de remboursement.</i>	55
<i>Tableau 33 : Prime totale de toutes les garanties par âge et taux de remboursement</i>	56
<i>Tableau 34 : S/P estimé pour chaque garantie</i>	57
<i>Tableau 35 : S/P estimé par âge et taux de remboursement</i>	57
<i>Tableau 36 : Triangle des incréments de paiements.....</i>	60
<i>Tableau 37 : Facteurs de développement de Chain Ladder.....</i>	65
<i>Tableau 38 : Total des réserves par Chain Ladder.....</i>	66
<i>Tableau 39 : Les facteurs de London Chain.....</i>	66
<i>Tableau 40 : Comparaison des réserves estimés de London Chain et Chain Ladder.</i>	67
<i>Tableau 41 : Comparaison des réserves estimées de London Chain avec Taylor et De Vylder</i>	67
<i>Tableau 42 : les réserves estimées selon Mack.....</i>	72
<i>Tableau 43 : Les réserves estimées selon Bootstrap</i>	73
<i>Tableau 44 : Comparaison entre Mack et Bootstrap pour les réserves estimées.....</i>	73
<i>Tableau 45 : Comparaison des écart-type de Mack et Bootstrap</i>	74
<i>Tableau 46 : Comparaison des intervalles de confiance pour Mack et Bootstrap.....</i>	74
<i>Tableau 47 : Meilleure estimation des réserves.....</i>	75

LISTE DES FIGURES

<i>Figure 1 : évolution du chiffre d'affaire maladie.....</i>	<i>21</i>
<i>Figure 2 : l'évolution de part de marché AXA Assurance Maroc.....</i>	<i>22</i>
<i>Figure 3 : l'évolution du chiffre d'affaire de la maladie d'AXA / marché.....</i>	<i>23</i>
<i>Figure 4 : Q-Q plot Log normal.....</i>	<i>38</i>
<i>Figure 5 : Q-Q plot gamma.....</i>	<i>38</i>
<i>Figure 6 : Courbe de concentration de Lorenz.....</i>	<i>38</i>
<i>Figure 7 : La fonction moyenne des excès</i>	<i>39</i>
<i>Figure 8: Le Hill-plot.....</i>	<i>40</i>
<i>Figure 9: P-P plot des charges extrêmes</i>	<i>42</i>
<i>Figure 10: Fonctions de répartition des charges extrêmes (GPD et observé).....</i>	<i>43</i>
<i>Figure 11 : histogramme de nombre des sinistres pour les médicaments.....</i>	<i>48</i>
<i>Figure 12 : distribution des différentes lois et la distribution réelle.....</i>	<i>48</i>
<i>Figure 13 : Q-Q plot Gamma pour le coût moyen des médicaments</i>	<i>53</i>
<i>Figure 14 : Q-Q plot Log normal pour le coût moyen des médicaments</i>	<i>53</i>
<i>Figure 15 : diagramme de Lexis pour la dynamique de vie des sinistres</i>	<i>59</i>

TABLE DES MATIERES

Remerciements	3
Dédicace	4
Dédicace	5
RESUME	6
Liste d'abréviations	7
INTRODUCTION	13
CHAPITRE 1 : La description du marché d'assurance maladie au Maroc	14
I. La description du produit Décès-Incapacité-Maladie	14
II. L'assurance maladie obligatoire au Maroc	16
1. Assurance Maladie Obligatoire géré par la CNOPS	17
2. Assurance Maladie Obligatoire géré par la CNSS :	18
3. Assurance Maladie Obligatoire gérée par les compagnies d'assurances	19
III. Axa assurance Maroc dans le marché de la maladie	19
1. Offre d'Axa assurance Maroc	19
2. Compétitivité du marché	20
CHAPITRE 2 : Modélisation de la prime pure	24
I. Cadre théorique	24
1. les modèles linéaires généralisés	24
2. Théorie des valeurs extrêmes (Peaks over threshold)	28
3. Évaluation et adéquation du modèle	30
II. Analyse des données	32
1. Travaux d'épuration et la fiabilisation de la base de données	32
2. Statistiques descriptives	33
3. Segmentation et codification des variables	33
III. Modélisation des sinistres graves :	37
1. Modélisation des sinistres graves de la garantie « Médicaments »	37
2. Modélisation des sinistres graves de la garantie « Consultation » et « Autres »	44
IV. Modélisation de la fréquence et du coût moyen	45
1. Approche :	45
2. Modélisation de la garantie « Médicaments »	45
3. Modélisation de la garantie « Consultation » et la garantie « Autres »	55
4. prime totale	55

5. <i>L'indicateur de rentabilité S/P :</i>	56
CHAPITRE 3 : <i>Provisionnement</i>	58
I. <i>Généralité</i>	58
1. <i>La dynamique de vie des sinistres</i>	58
2. <i>Formalisation du problème de provisionnement</i>	59
II. <i>Les méthodes déterministes</i>	61
1. <i>La méthode de Chain Ladder</i>	61
2. <i>Méthode de London Chain</i>	62
3. <i>Méthode des moindres carrés de De Vylder</i>	62
4. <i>Méthode de séparation de Taylor</i>	63
5. <i>Application des méthodes déterministes</i>	64
5.1 <i>Méthode de Chain Ladder</i>	65
5.2. <i>Méthode de London Chain</i>	66
III. <i>Les méthodes stochastiques</i>	68
1. <i>Modèle de Mack</i>	68
2. <i>La méthode de Bootstrap :</i>	70
3. <i>Application des méthodes stochastiques :</i>	72
3.1. <i>Modèle de Mack</i>	73
3.2. <i>Bootstrap</i>	73
Conclusion	76
BIBLIOGRAPHIE.....	77
<i>Annexe 1 : Schéma de la couverture médicale de base au Maroc</i>	78
<i>Annexe 2 : Modélisation des sinistres graves pour la garantie «Consultation»</i>	79
<i>Annexe 3 : Modélisation des sinistres graves pour la garantie «Autres»</i>	80
<i>Annexe 4 : Estimation des paramètres du modèle ZINB de la fréquence de la garantie « Médicaments »</i>	81
<i>Annexe 5 : Estimation des paramètres du modèle Gamma du coût moyen de la garantie « médicaments »</i>	83
<i>Annexe 6 : Modélisation des sinistres ordinaires pour la garantie «Consultation»</i>	85
<i>Annexe 7 : Modélisation des sinistres ordinaires pour la garantie «Autres»</i>	87
<i>Annexe 8 : Triangles des règlements cumulés</i>	89
<i>Annexe 9 : Vérification de l'hypothèse H2 pour l'application de la méthode Chain Ladder</i>	90

INTRODUCTION

Assurer l'équilibre des portefeuilles des différents produits d'assurance est cruciale pour garantir une continuité de ces produits sur le marché, c'est pour cela le suivi fréquent de la rentabilité de chaque produit est nécessaire.

Dans le cas de l'assurance maladie, qui a devenu une assurance obligatoire depuis 2005, il est très important d'assurer la rentabilité du portefeuille et de déterminer les différents indicateurs (S/P, fréquence et coût moyen) permettrons d'évaluer l'équilibre et la rentabilité de cette branche dans un marché très concurrents est parfaitement libéral en terme de primes appliquées.

En effet, la première partie de cette étude est consacré à l'élaboration d'un modèle de la prime et la segmentation du portefeuille qui se fait à l'aide de variables jugées fortement corrélées avec la sinistralité. Une fois la segmentation réalisée, on peut construire des modèles statistiques reliant les variables tarifaires aux variables de sinistralité.

La sinistralité en assurance maladie est caractérisée par plusieurs risques et garanties c'est pour cette raison que nous établirons des modèles de la fréquence et du coût moyen pour les différents types de sinistres (graves et ordinaires) et de risques pour calculer la prime pure correspondante à chaque risque et finalement déduire la prime pure totale comme étant la somme de ces dernières.

Les modèles utilisés sont: les modèles linéaires généralisés (GLM) pour les sinistres ordinaires et les distributions Pareto généralisés (GPD) pour les sinistres graves.

Dans la deuxième partie nous allons estimer les différentes réserves et provisions, spécialement les provisions des sinistres à payer (PSAP) en utilisant des méthodes déterministes (Chain Ladder, London Chain , De Vylder et Taylor) et des méthodes stochastiques (modèle de Mack et méthode Bootstrap) afin de savoir les engagements de l'assurance envers ces assurés.

CHAPITRE 1 : La description du marché d'assurance maladie au Maroc

Dans ce chapitre, nous allons donner un aperçu sur le marché d'assurance maladie au Maroc. Tout d'abord nous allons présenter le produit qu'offrent les assurances privées pour la maladie, puis nous allons détailler la couverture médicale de base et finalement comparer l'évolution du chiffre d'affaire de la maladie d'AXA Assurance Maroc par rapport à celle du marché des assurances.

1. La description du produit Décès-Incapacité-Maladie

Le produit « **D.I.M** » est un contrat de groupe, a pour objet la prise en charge partielle ou totale des frais engagés par les affiliés et les personnes qu'ils ont à charge (conjoint et enfants) suite à un Décès, Incapacité - Invalidité et Maladie – Maternité. Nous allons d'abord définir chaque composante du produit « **D.I.M** » et leurs types de garantie.

1. Assurance décès

L'assurance décès est une assurance qui a pour objet le paiement d'un capital au profit des bénéficiaires ou ayants droit en cas de décès de l'assuré. Sur le marché, les types de garantie qui existent en cas de décès sont la garantie décès toutes causes et la garantie décès accidentel qui sont bien définie dans le tableau suivant :

Les types de garantie en cas de décès	
Garantie Décès toutes Causes	Garantie Décès accidentel
Garantie qui a pour objet de garantir en cas de décès de l'Assuré, le paiement au(x) bénéficiaire(s) désigné(s) d'un capital dont le montant est fixé aux conditions particulières du contrat.	Garantie qui a pour objet le paiement d'un capital supplémentaire en cas de décès accidentel de chaque assuré déjà garanti en cas de décès. Le décès est dit accidentel lorsqu'il est provoqué par une cause extérieure, soudaine et involontaire, à la condition qu'il survienne dans les six mois suivant l'accident

Tableau 1 : Types de garantie en cas de décès

2. Assurance Incapacité/Invalidité

L'assurance Incapacité/Invalidité est une assurance contre la perte de revenus qui couvre l'assuré en cas de survenance d'une maladie ou d'un accident entraînant une incapacité de travail et donc une réduction de ses revenus professionnels. On peut distinguer deux états d'invalidité qu'un assuré pourrait avoir et qui sont :

-Invalidité permanente : Etat d'invalidité dans lequel aucun traitement ultérieur ne permettra d'apporter une amélioration notable à la capacité de l'assuré. Le degré d'invalidité est alors considéré comme définitif.

-Invalidité temporaire : Etat d'invalidité transitoire susceptible soit de s'achever par une fin d'incapacité, soit de se transformer en invalidité permanente. Il y a deux types de garanties pour l'invalidité qui sont définies dans le tableau suivant :

Les types de garantie pour Incapacité/Invalidité	
Invalidité totale	Invalidité partielle
Garantie dont les prestations ne sont versées qu'en cas où le degré d'invalidité est de 100%. Souvent une invalidité à partir de 67% est considérée comme totale à 100%.	Garantie dont les prestations sont versées proportionnellement au degré d'invalidité.

Tableau 2 : Types de garantie en cas d'incapacité/Invalidité

3. Assurance maladie

L'assurance maladie est une assurance contre les dépenses de soin de santé qui a pour objet de couvrir l'assuré en cas de survenance d'une maladie, d'un accident ou d'un accouchement entraînant des frais d'hospitalisation comme les nuitées, les honoraires médecins, produits pharmaceutiques, etc. Le tableau suivant contient les types de garantie pour l'assurance maladie.

Types de garantie pour l'assurance maladie						
Consultation et visite	médicament	laboratoire et radiologie	Hospitalisations	Optique	Soins et prothèses dentaires	Maternité
Toutes consultations, visite de nuit ou de dimanche	tous frais pharmaceutiques pour les médicaments.	Toutes analyses médicales et de Radiologie.	Toutes hospitalisations chirurgicales dans les cliniques privées, dans les hôpitaux publics ou les polycliniques	Tous frais de verres et de montures.	Tous soins conservateurs et les soins chirurgicaux.	Tous frais d'accouchement normal ou suite à une césarienne.

			C.N.S.S. Et toutes hospitalisation médicales comme les honoraires des médecins et frais de séjour.		
--	--	--	--	--	--

Tableau 3 : Types de garantie en cas d'assurance maladie

II. L'assurance maladie obligatoire au Maroc

Le régime marocain d'assurance maladie couvre les salariés du secteur public et ceux du secteur privé. Il assure aux intéressés une protection contre les risques de différentes maladies.

On distingue deux types d'assurance maladie qui sont :

- **L'assurance maladie obligatoire (AMO) :** Ce régime est entré en vigueur le 18 août 2005 a été institué en 2002 par la loi 65.00 portant code de la couverture médicale, cette assurance garantit les prestations qui comprennent en général les types de frais qu'on a cité dans la maladie.
- **L'assurance maladie complémentaire (AMC) :** Elle garantit une couverture des soins exclus de l'AMO comme elle présente un complément du taux de remboursement des soins de l'AMO.

Le régime d'assurance maladie obligatoire est géré par les organismes suivants :

- **La Caisse Nationale de Sécurité Sociale (CNSS) :** elle gère l'assurance maladie obligatoire du secteur privé.
- **La caisse nationale des organismes de prévoyance sociale (CNOPS) :** elle gère l'assurance maladie obligatoire du secteur public.
- **Les compagnies d'assurance :** elles gèrent aussi l'assurance maladie obligatoire (article 114 de la loi 65.00) ainsi que l'assurance maladie complémentaire du secteur public et privé.

Le régime d'assistance médicale (RAMED) est géré par les organismes suivants :

- **l'état et les collectivités locales :** ils couvrent les personnes démunies non couvertes par le régime d'assurance maladie obligatoire de base.

Pour veiller au bon fonctionnement de la couverture médicale de base de ces organismes, l'agence national de l'assurance maladie (ANAM) est faite pour cela et a pour mission le contrôle et l'encadrement technique de l'AMO, l'architecture de la couverture médicale de base est bien expliqué dans le schéma de l'annexe 1.

1. Assurance Maladie Obligatoire géré par la CNOPS

1.1. La population éligible

Les agents du secteur public sont soumis obligatoirement à l'assurance maladie obligatoire, ils sont:

Les actifs : les fonctionnaires et les agents de l'Etat, des collectivités locales, des établissements publics et des personnes morales de droit public.

Les retraités: les titulaires de pensions du secteur public (y compris les pensions de reversions).

Les Ayants droits : les Handicapés à vie, les enfants à l'âge de 21 ans, les étudiants jusqu'à l'âge de 26 ans et les conjointes.

1.2. Panier de soins

Dans le panier de soins, nous avons cité les types de soins que CNOPS couvre ainsi que les prestations qu'elle donne, ceci est présenté dans le tableau suivant :

Le panier de soins	
les types de soins	les prestations
<ul style="list-style-type: none"> - Actes de médecine générale et de spécialités médicales et chirurgicales, actes paramédicaux, de rééducation fonctionnelle et de kinésithérapie délivrés à titre ambulatoire hors médicaments. - Soins liés à l'hospitalisation et aux interventions chirurgicales y compris les actes de chirurgie réparatrice et le sang et ses dérivés labiles. - Médicaments admis au remboursement. - Lunetterie médicale, dispositifs médicaux et implants nécessaires aux actes médicaux et chirurgicaux. - Appareils de prothèse et d'orthèse médicales admis au remboursement. - Soins bucco-dentaires. - Orthodontie médicalement requise pour les enfants. 	<ul style="list-style-type: none"> ALD et ALC : 100 % sur la base de la tarification nationale de référence (TNR). - Hospitalisation dans les hôpitaux publics : 100 % de la TNR. - Hospitalisation dans les cliniques privées : 90 % de la TNR. - Médicaments : 70 % (Le remboursement des dépenses de médicaments s'effectue sur la base du PPM du médicament générique, lorsqu'il existe.) - Soins ambulatoires : 80 % de la TNR. - Appareillage, lunetterie, prothèses, orthèses, implants et dispositifs médicaux : sous forme de forfaits prévus dans la TNR.

Tableau 4 : Panier de soins de la CNOPS

1.3. Le taux de cotisation :

Le taux de cotisation pour l'assurance maladie n'est pas le même pour les actifs que pour les pensionnés du secteur public, on a donc pour :

Les actifs : Le taux de cotisation est de 5% de l'ensemble des rémunérations, 2,5 % à la charge de l'employeur et 2,5 % qui reste est à la charge du salarié.

Les retraités : Le taux de cotisation est de 2,5% du montant global des pensions de base.

Pour la CNOPS, toutes les cotisations doivent atteindre un seuil minimal mensuel de 70 DH et plafonné à 400 DH.

2. Assurance Maladie Obligatoire géré par la CNSS :

2.1. La population éligible :

Les agents du secteur privé sont soumis obligatoirement à l'assurance maladie obligatoire de la CNSS, ils sont:

Les actifs : les salariés du secteur privé.

Les retraités : les titulaires de pensions du secteur privé.

Les ayants droits : les handicapés à vie, l'enfant à l'âge de 21 ans, les étudiants jusqu'à l'âge de 26 ans et les conjointes.

2.2. Le panier de soins :

Dans le panier de soins, nous avons cité les types de soins que la CNSS couvre ainsi que les prestations qu'elle donne à ses assurés et ceci est présenté dans le tableau suivant :

Le panier de soins	
les types de soins	les prestations
<ul style="list-style-type: none"> - ALD et ALC - Suivi de la grossesse, l'accouchement et ses suites, les actes médicaux et chirurgicaux. - Les médicaments admis au remboursement, le sang et ses dérivés labiles, les actes paramédicaux et les actes de rééducation fonctionnelle et de kinésithérapie. - Hospitalisation: l'ensemble des prestations et 	<ul style="list-style-type: none"> 90% pour les prestations assurées par les professionnels et établissements de soins du secteur public (TNR). - 70% pour les prestations assurées par les professionnels et établissements de soins du

soin rendus dans ce cadre y compris les actes de chirurgie réparatrice. - Soins Ambulatoires.	secteur privé (TNR). - 90% pour les ALD et ALC (TNR).
--	--

Tableau 5 : Panier de soins de la CNSS

2.3. Le taux de cotisation

Le taux de cotisation pour l'AMO change selon les actifs et les pensionnés du secteur privé, on a donc pour :

Les actifs : le taux de cotisation est de 4% de l'ensemble des rémunérations réparties en deux, 2 % à la charge de l'employeur et 2 % à la charge du salarié qui est majoré de 1,50% de l'ensemble de la rémunération brute mensuelle du salarié.

Les retraités : Le taux de cotisation due par les titulaires de pensions est fixé à 4% sur le montant global des pensions de base.

Contrairement à la CNOPS, les cotisations pour la maladie à la CNSS ne sont pas plafonnées.

3. Assurance Maladie Obligatoire gérée par les compagnies d'assurances

L'assurance maladie obligatoire qui est gérée par les assurances concerne les salariés du secteur privé, les fonctionnaires de l'état et les titulaires de pensions du secteur privé et public qui sont connus par l'article 114 ainsi que leurs ayants droits qui sont les handicapés à vie, les enfants à l'âge de 21 ans, les étudiants jusqu'à l'âge de 25 ans et leurs conjointes. Les compagnies d'assurance couvrent tous les types de soins qui sont couverts par la CNSS et la CNOPS avec des prestations qui varient généralement de 70% à 90% des charges selon le panier choisi par l'assuré en contrepartie d'une cotisation qui est soit forfaitaire soit avec un taux de cotisation sur l'ensemble de la masse salariale.

III. Axa Assurance Maroc dans le marché de la maladie

Axa Assurance Maroc est un groupe d'assurances s'adressant aux particuliers et aux entreprises en répondant à leurs besoins de services en matière d'assurance, de prévoyance, d'épargne et de transmission de patrimoine.

1. Offre d'Axa assurance Maroc

Axa Assurance Maroc met à la disposition de ses clients de différentes couvertures principalement les soins médicaux, Hospitalisation, maternité, dentaire et optique. Pour la couverture DIM, AXA dispose des deux produits suivants :

- **Prévoyance Santé Entreprise** : c'est une assurance maladie obligatoire, elle concerne les entreprises, identique à celle de la CNSS, de la CNOPS ou d'un autre régime de base, elle couvre en cas de maladie, décès et invalidité.
- **Vitalis** : c'est une assurance complémentaire au Régime Collectif de Prévoyance souscrit auprès d'AXA. elle permet de prendre en charge les frais engagés en cas de maladie ou d'hospitalisation suite à une maladie ou à un accident, elle n'intervient qu'après épuisement d'un plafond de 1.000.000 Dhs par bénéficiaire et par période de 12 mois pour l'ensemble des prestations.

2. Compétitivité du marché

Le marché d'assurance au Maroc devient de plus en plus concurrent spécialement celui de la maladie.

2.1 Evolution du chiffre d'affaire maladie

Le tableau suivant comporte les données annuelles de 2008 à 2013 du chiffre d'affaire de la branche maladie des différentes compagnies d'assurance. Axa Assurance Maroc se situe la quatrième en termes de chiffre d'affaire Maladie en 2013 (394,89 Millions Dhs).

	2008	2009	2010	2011	2012	2013
Cinia Saada Assurance	125041	450065	459858	480242	503682	527567
RMA Watanya	436248	500543	491890	509099	489118	483267
Wafa Assurance	129267	145301	301899	349318	383459	414383
Axa Assurance Maroc	471943	496691	395090	384689	389992	394890
Atlanta	219387	222672	243718	207719	326481	346459
Sanad	111255	118808	122145	136622	116235	137528
Zurich	107733	121147	121230	113016	117572	108621
Marocaine Vie	24831	19980	18821	16727	18370	22128
MCMA	2337	3639	5325	5984	11390	11454
MAMDA	5930	6354	5580	5901	6660	7586
Marché	1633972	2085200	2165556	2209317	2362959	2453883

Tableau 6 : Chiffre d'affaire de la maladie des compagnies d'assurances (en 1000DHS)

Le graphe suivant illustre l'évolution du chiffre d'affaire Axa Assurance Maroc par rapport à l'évolution du marché d'assurance.

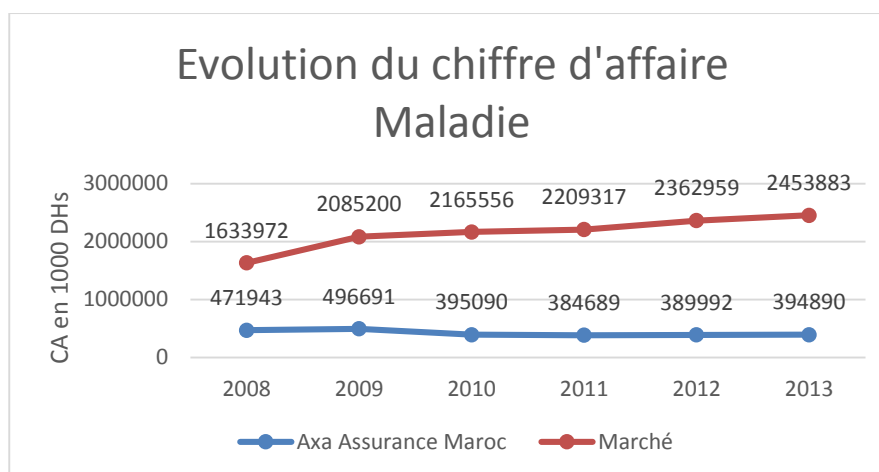


Figure 1 : Evolution du chiffre d'affaire maladie

Nous remarquons donc que le marché a connu une croissance en volume de 50.2% entre 2008 et 2013 pour atteindre un chiffre d'affaire globale de 2 453MDhs contre AXA qui a connu une baisse de 16.3% sur cette même période.

2.2.Evolution du part marché maladie

Le tableau suivant contient la part du marché de la branche maladie des différentes compagnies d'assurance.

	2008	2009	2010	2011	2012	2013
Cnia Saada Assurance	7,7%	21,6%	21,2%	18,9%	21,3%	21,5%
RMA Watanya	26,7%	24,0%	22,7%	4,5%	20,7%	19,7%
Wafa Assurance	7,9%	7,0%	13,9%	0,7%	16,2%	16,9%
Axa Assurance Maroc	28,9%	23,8%	18,2%	18,6%	16,5%	16,1%
Atlanta	13,4%	10,7%	11,3%	9,4%	13,8%	14,1%
Sanad	6,8%	5,7%	5,6%	0,9%	4,9%	5,6%
Zurich	6,6%	5,8%	5,6%	0,2%	5,0%	4,4%
Marocaine Vie	1,5%	1,0%	0,9%	7,8%	0,8%	0,9%
MCMA	0,1%	0,2%	0,2%	5,1%	0,5%	0,5%
MAMDA	0,4%	0,3%	0,3%	18,8%	0,3%	0,3%
Marché	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tableau 7 : La part du marché de la maladie des différentes compagnies d'assurance

D'après les données du tableau, le graphique suivant montre l'évolution de la part du marché d'AXA Assurance Maroc sur la même période.

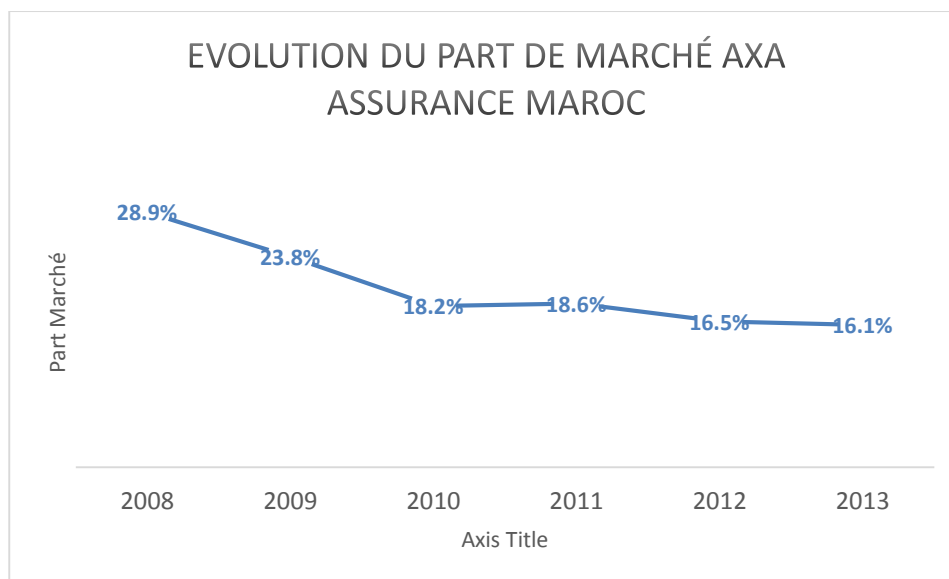


Figure 2 : L'évolution de part de marché AXA Assurance Maroc

Nous remarquons que la part du marché d'AXA Assurance Maroc ne fait que diminuer durant cette période en passant de 28,9% en 2008 à 16,1% en 2013, ceci est à cause du marché qui devient de plus en plus concurrent.

2.3. Evolution AXA/marché

Le tableau suivant contient l'évolution du chiffre d'affaire de la branche maladie des différentes compagnies d'assurance.

	2009/2008	2010/2009	2011/2010	2012/2011	2013/2012
Cnia Saada Assurance	72,2%	2,1%	4,2%	4,7%	4,7%
RMA Watanya	12,8%	-1,8%	3,4%	-4,1%	-1,2%
Axa Assurance Maroc	5,0%	-25,7%	-2,7%	1,4%	1,3%
Wafa Assurance	11,0%	51,9%	13,6%	8,9%	8,1%
Atlanta	1,5%	8,6%	-17,3%	36,4%	6,1%
Zurich	11,1%	0,1%	-7,3%	3,9%	-7,6%
Sanad	6,4%	2,7%	10,6%	-17,5%	18,3%
Marocaine Vie	-24,3%	-6,2%	-12,5%	8,9%	20,5%
MCMA	35,8%	31,7%	11,0%	47,5%	0,6%
MAMDA	6,7%	-13,9%	5,4%	11,4%	13,9%
Marché	21,6%	3,7%	2,0%	6,5%	3,8%

Tableau 8 : L'évolution du chiffre d'affaire de la maladie des différentes compagnies d'assurance

Nous avons illustré l'évolution d'Axa Assurance Maroc par rapport au marché de la période donnée dans le graphique suivant.

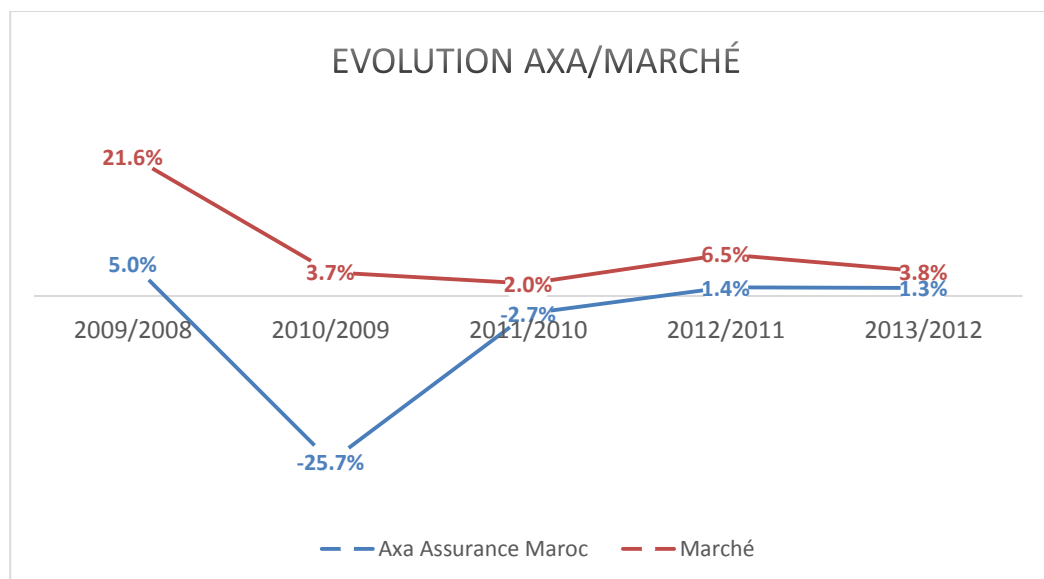


Figure 3 : L'évolution du chiffre d'affaire de la maladie d'AXA par rapport au marché

Le marché de l'assurance maladie a connu une évolution annuelle moyenne entre 2008 et 2013 de 7.52% contre une baisse annuelle moyenne de 4.1 % couru par AXA Assurance Maroc à cause d'une baisse importante en 2009 de l'ordre de -25.7%.

CHAPITRE 2 : Modélisation de la prime pure

La modélisation de la prime pure nécessite une méthodologie bien choisie, pour ce faire nous allons présenter le cadre théorique que nous allons adopter et mettre en pratique l'approche théorique adopté pour l'élaboration du modèle le plus adéquat. Nous allons commencer par modéliser la prime pure des sinistres graves par l'approche de la théorie des valeurs extrêmes et puis nous allons appliquer les principes du modèle linéaire généralisé pour la modélisation des sinistres ordinaires afin d'avoir la prime pure estimée totale de la maladie. Finalement nous allons analyser la rentabilité de la maladie par le ratio S/P.

I. Cadre théorique

1. les modèles linéaires généralisés

Actuellement, les problèmes statistiques deviennent de plus en plus complexes, ce qui nécessite des modèles tenant mieux compte de la réalité de l'assurance, chose qui n'est pas pris en considération par le modèle linéaire classique. En effet, le modèle gaussien impose une série de limitations peu conciliables avec la réalité des nombres ou des coûts de sinistres tels que la densité de probabilité normale, la linéarité des variables explicatives vis-à-vis de la variable réponse ainsi que l'homoscédasticité. Toutefois, il est possible de transformer la variable réponse à l'aide de fonctions bien choisies mais ceci s'accompagne de nombreux désavantages à savoir le travail sur une échelle artificielle, la difficulté de revenir aux quantités initiales, etc.

Les modèles linéaires généralisés nous permettent de résoudre ce problème et de s'affranchir de l'hypothèse de normalité en traitant de manière unifiée des réponses dont la loi fait partie de la famille exponentielle. La régression de poisson (et les modèles apparentés ; tel que la régression binomiale négative) font partie de cette famille et offrent un outil très intéressant pour l'analyse et la modélisation du nombre de sinistres. Ajoutant à cela, les techniques « GLM » qui permettent l'analyse d'un très grand nombre de phénomène dans une optique de quasi-vraisemblance en ne spécifiant que la structure moyenne-variance.

Intéressons-nous pour le moment à cette famille dite exponentielle. Nombreux sont les lois de probabilités qu'elle décrit qu'elles soient discrètes ou continues. Définies à partir de deux paramètres \emptyset et θ , leur densité peut s'écrire sous la forme :

$$f(y; \theta, \emptyset) = \exp \left\{ \frac{\theta y - \mathbf{b}(\theta)}{\emptyset} + c(y, \emptyset) \right\}, y \in S$$

Où θ est un paramètre réel, appelé paramètre naturel, $\phi > 0$ (ϕ éventuellement donné) est un paramètre de dispersion. Les fonctions \mathbf{b} et \mathbf{c} sont spécifiques à la distribution, \mathbf{b} étant « régulière ». Le support \mathbf{S} est un sous-ensemble de \mathbf{N} ou de \mathbf{R} .

On peut facilement montrer que :

$$\begin{aligned}\mu &= \mathbb{E}(Y) = b'(\theta) \\ \sigma^2 &= \text{Var}(Y) = \phi b''(\theta) \\ \text{Var}(\mu) &= b''(\theta) = b''(b'^{-1}(\mu))\end{aligned}$$

La fonction V est appelée fonction variance de la distribution et joue un rôle essentiel dans ces modèles.

Souvent une pondération est nécessaire, on remplace alors ϕ par ϕ/w où w est un poids connu à priori.

Exposons quelques modèles dont on aura besoin par la suite en spécifiant les caractéristiques et les propriétés spécifiques de chaque loi, on commence par la loi de poisson, ensuite on présente la loi binomiale Négative, puis la loi Normale et Gamma et enfin les modèles zero-inflated ZIP et ZINB:

1.1. Distributions discrètes

– Loi de Poisson $P(\lambda)$:

La distribution de probabilité d'une loi poisson de paramètre λ est donnée par :

$$P(Y = y) = \exp(-\lambda) \frac{\lambda^y}{y!}$$

Dans ce cas les paramètres de la famille exponentielle sont :

$$\begin{aligned}\theta &= \ln \lambda ; \phi = 1 \\ \mu &= E(Y) = \text{Var}(Y) = \lambda \\ c(y, \phi) &= -\ln(y!) ; b(\theta) = \exp(\theta)\end{aligned}$$

L'égalité $\text{Var}(Y) = E(Y)$ introduit une limitation forte dans l'application de ce modèle en assurance non-vie. On s'en libère en utilisant le modèle de Poisson « sur-dispersé » $P_{surd}(\lambda, \phi)$, ϕ étant un paramètre additionnel strictement positif. Celui-ci s'introduit formellement ainsi :

$$Y \sim P_{surd}(\lambda, \phi) \text{ ssi } \frac{Y}{\phi} \sim P(\lambda)$$

- *Loi binomiale Négative* $BN\left(r, \frac{r}{r+\mu}\right)$

Une introduction alternative de la sur-dispersion reposerait sur la loi Binomiale Négative à deux paramètres r et μ . Cependant celle-ci n'appartient pas au type exponentiel des **GLM**. On a en effet :

$$P(Y = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!} \frac{r^y \mu^y}{(r + \mu)^{r+y}} \text{ avec } y \in \mathbb{N}$$

Que l'on peut écrire sous la forme :

$$P(Y = y) = \exp\left[(y + r)\ln\left(\frac{\mu}{r + \mu}\right) + c(y, r)\right]$$

Cependant si r est donné, on récupère une distribution exponentielle avec :

$$\theta = \ln\left(\frac{\mu}{r + \mu}\right); \phi = 1; E(Y) = \mu; V(\mu) = \mu + \frac{\mu^2}{r}$$

1.2. Distributions continues

- *Loi Normale* $N(\mu, \sigma^2)$

Cette distribution rentre dans le type exponentiel car :

$$f(y) = \exp\left\{\frac{1}{\sigma^2}\left(\mu y + \frac{\mu^2}{2}\right) + c(y, \sigma^2)\right\} \text{ avec } y \in \mathbb{R}$$

Ainsi on a $\theta = \mu = E(y); \text{Var}(Y) = \sigma^2; V(\mu) = 1$

- *Loi Gamma* $\gamma\left(v, \frac{v}{\mu}\right)$

$$f(y) = \exp\left[\left(-\frac{y}{\mu} - \ln(\mu)\right)v + c(y, v)\right] \text{ avec } y > 0$$

$$\theta = -\frac{1}{\mu}, \phi = \frac{1}{v}, E(Y) = \mu, V(\mu) = \mu^2$$

Remarquons que le paramètre de dispersion ϕ contrôle uniquement la variance et pas l'espérance de la distribution. Par conséquent, la prime pure se détermine seulement à l'aide du paramètre θ alors que ϕ est considéré comme un paramètre de

nuisance. Il reste toutefois important dans la mesure où il contrôle la variance et par suite le risque.

1.3. Les modèles zero-inflated

La motivation principale pour le modèle zero-inflated est que dans les données on constate principalement une sur-dispersion et un excès des zéros, ce modèle qui est un mélange entre une masse en 0 et un modèle classique de comptage typiquement un modèle de Poisson ou binomial négatif, donne la possibilité de modéliser l'excès des zéros en plus de permettre la sur-dispersion.

Pour modéliser la probabilité de ne pas avoir de sinistre (et donc d'avoir un surpoids en 0), considérons un modèle logistique par exemple,

$$\pi_i = \frac{\exp[X_i' \beta]}{1 + \exp[X_i' \beta]}$$

Pour le modèle de comptable, on note $p_i(k)$ la probabilité que l'individu i ait k sinistres.

Donc on a

$$P(N_i = k) = \begin{cases} \pi_i + [1 - \pi_i]p_i(0) & \text{si } k = 0 \\ [1 - \pi_i]p_i(k) & \text{si } k = 1, 2, \dots \end{cases}$$

Si p_i correspond à un modèle de Poisson (modèle ZIP), on peut alors montrer facilement que :

$$\mathbb{E}(N_i) = [1 - \pi_i]\mu_i \quad \text{et} \quad \text{Var}(N_i) = \mu_i[1 - \pi_i][1 + \mu_i\pi_i]$$

Si p_i correspond à un modèle Binomial Négatif (modèle ZINB), on a :

$$\mathbb{E}(N_i) = [1 - \pi_i]\mu_i \quad \text{et} \quad \text{Var}(N_i) = \mu_i[1 - \pi_i][1 + \mu_i[\pi_i + \alpha]]$$

Dans le cadre des modèles linéaires généralisés, pour chaque individu appartenant à une population donnée dont on dispose d'informations sous forme d'un ensemble de variables x_i^t , on cherche à déterminer des paramètres inconnues β tel que :

$$\mu_i = g^{-1}(X_i' \beta)$$

g est appelée la fonction de lien. Pour chaque type de distribution, il y a une fonction de lien canonique qui est définie par :

$$g = b'$$

Les fonctions les plus utilisés sont la fonction logarithmique, la fonction puissance et la fonction Identité.

En bref, un modèle GLM se résume donc à :

Une distribution de probabilité, pour la variable réponse, une fonction variance V et un paramètre de dispersion \emptyset avec une fonction lien g qui lie la moyenne au prédicteur.

2. Théorie des valeurs extrêmes (Peaks over threshold)

La théorie des valeurs extrêmes a pour but d'étudier la loi des extrêmes ou des maximum d'une suite des variables aléatoires réelles même si, et spécialement si, la loi du phénomène n'est pas connue. En fait seules les données extrêmes sont utilisées pour estimer les paramètres des modèles de la TVE ce qui assure un meilleur ajustement du modèle à la queue de la loi.

En effet, les outils probabilistes traditionnels (développés dans un univers gaussien) sont inadaptés à l'appréhension de ces comportements extrêmes : la moyenne n'existe plus dans le nouvel univers des risques.

Il existe plusieurs méthodes de modélisation des valeurs extrêmes dans la TVE : GEV, POT, BLOCK MAXIMA ...etc. Nous nous intéresserons à la méthode des excès au-delà d'un seuil (Peak Over Threshold).

2.1. La méthode des excès au-delà d'un seuil

La méthode des excès au-delà d'un seuil (ou Peak Over Threshold, POT) repose sur le comportement des valeurs observées au-delà d'un seuil donné. En d'autres termes, elle consiste à observer non pas le maximum ou les plus grandes valeurs mais toutes les valeurs des réalisations qui excèdent un certain seuil élevé. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil.

On définit un seuil $u \in \mathcal{R}$, $N_u = \text{Card}\{i, i = 1, \dots, n, X_i > u\}$ et $Y_i = X_i - u > 0$ pour $1 \leq i \leq N_u$ où N_u est le nombre des dépassements du seuil u par les $(X_i)_{1 \leq i \leq N_u}$ et Y_1, \dots, Y_{N_u} les excès correspondants.

On cherche à partir de la loi F de X à définir une loi conditionnelle F_u par rapport au seuil u pour les variables aléatoires dépassant ce seuil.

On définit alors la loi conditionnelle des excès F_u par :

$$F_u = \Pr(X - u \leq y \mid X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad y \geq 0$$

La forme de la loi limite pour les valeurs extrêmes, sous certaines conditions de convergence, est une loi de Pareto généralisée que l'on notera GPD.

En effet, on a

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - G_{\xi\beta(u)}(y)| = 0$$

Où $F_u(y)$ est la fonction de répartition conditionnelle des excès pour u élevé, x_F est le point terminal de F , $x_F = \{x \in \mathcal{R}: F(x) < 1\}$ et $G_{\xi\beta(u)}(y)$ est la GPD donnée par :

$$G_{\xi\beta(u)}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\beta(u)}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta(u)}\right), & \xi = 0 \end{cases}$$

Où : $y \geq 0$ Pour $\xi \geq 0$ et $0 \leq y \leq -\frac{\beta(u)}{\xi}$ Pour $\xi \leq 0$

Donc le cas où $\xi > 0$ correspond aux lois à queues épaisses, pour lequel 1-G se comporte comme une puissance, $x^{-\frac{1}{\xi}}$ pour x assez élevé. Si $\xi = 0$, on trouve :

$1 - \exp\left(-\frac{y}{\beta(u)}\right)$, c'est la loi exponentielle de paramètre β et enfin pour $\xi < 0$ c'est la loi Pareto de type II à support borné.

La GPD a les propriétés suivantes :

$$\mathbb{E}(y) = \frac{\beta}{1 - \xi}, (\xi < 1) \text{ et } Var(y) = \frac{\beta^2}{(1 - \xi)(1 - 2\xi)}, \left(\xi < \frac{1}{2}\right)$$

En pratique le choix du seuil constitue une difficulté. En fait, u doit être assez grand pour que l'approximation GPD soit valide, mais pas trop élevé pour garder un nombre suffisant de dépassements pour estimer les paramètres du modèle. Le seuil doit être choisi de façon à faire un arbitrage, traditionnel en statistiques, entre le biais et la variance.

Généralement, u est déterminé graphiquement en exploitant la linéarité de la fonction d'excès moyenne $e(u)$ pour la GPD.

En effet, pour une variable aléatoire qui suit une $G_{\xi\beta}$ avec ($\xi < 1$), $e(u)$ est linéaire avec u :

$$e(u) = E(X - u | X > u) = \frac{\beta + \xi u}{1 - \xi}, (\xi < 1)$$

Cette technique fournit une aide précieuse, cependant, il ne faut pas attendre d'elle la bonne valeur de u .

2.2. Estimation des paramètres de la GPD par Maximum de vraisemblance

En se référant à la littérature, on note diverses méthodes qui ont été proposées pour estimer les paramètres des lois GPD. On trouve la méthode basée sur le maximum de vraisemblance [Smith, 1987], la méthode des moments pondérées [Greenwood *et al.* 1979], la méthode des moments [Christophe, 1994] ou encore par des méthodes bayésiennes [Lye *et al.* 1993]. Il existe également des approches non paramétriques pour l'estimation de l'indice de queue. Mais la méthode la plus populaire et qui sous certaines conditions est la plus efficace est la méthode de maximum de vraisemblance. Dans ce qui suit, nous allons présenter donc la méthode de maximum de vraisemblance d'estimation paramétrique pour la GPD.

Considérons à nouveau la loi GPD, la densité bi-paramétrique de la GDP est donnée par :

$$g(y) = \frac{1}{\beta} \left(1 + \xi \left(\frac{y}{\beta} \right) \right)^{-\frac{1}{\xi}-1}$$

Où : $y \geq 0$ si $\xi \geq 0$ et $0 \leq y \leq -\frac{\beta}{\xi}$ si $\xi \leq 0$

L'estimation de la GPD, par la méthode de maximum de vraisemblance, porte sur l'indice de queue ξ ainsi que sur le paramètre d'échelle β . L'expression de la log-vraisemblance est donc :

$$l((\beta, \xi); Y) = -N_u \ln \beta - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\beta} y_i \right)$$

Pour un échantillon Y_1, \dots, Y_{N_u} des excès. À partir d'ici on peut dériver par rapport à chaque paramètre, obtenir les équations à résoudre, et dégager l'estimateur de maximum de vraisemblance (EMV) $\hat{\theta}_{ML}$ de $\theta = (\beta, \xi)$.

Pour $\xi > -0.5$, Smith [1985, 1987], Hosking et Wallis [1987] prouvent que les conditions de régularités de la fonction de vraisemblance sont remplies et que l'estimateur du maximum de vraisemblance conduit à un estimateur non biaisé, asymptotiquement normal.

3. Évaluation et adéquation du modèle

3.1. Test de Wald

Soit le test suivant : $H_0: \beta_k = 0$ contre $H_1: \beta_k \neq 0$

Ainsi la statistique de Wald s'écrit :

$$W = \frac{\widehat{\beta}_k}{s(\widehat{\beta}_k)}$$

Avec $s(\widehat{\beta}_k)$: l'écart type estimé de $\widehat{\beta}_k$.

Sous H_0 : W suit approximativement la loi normale $N(0,1)$.

– **Règle de décision :**

Au niveau de signification α : on rejette H_0 si $W > \phi(1-\alpha)$ avec ϕ : la fonction de réciproque de la loi normale $N(0,1)$.

3.2. Test du rapport de vraisemblance

Pour tester l'adéquation d'un modèle, on utilise une approche qui consiste à comparer un modèle sans paramètre autre que la constante, et le modèle avec tous les paramètres que l'on veut tester. On calcule pour chacun des deux modèles, la -2LL (c'est-à-dire -2 Log-vraisemblance). Le test consiste simplement à calculer la différence entre les deux valeurs, celle-ci se distribuant comme le chi2 à J degré de liberté correspondant au nombre de paramètres que comprend le modèle. L'hypothèse correspondante au test étant que la différence est nulle. Autrement dit que l'incorporation des variables n'apporte rien à la vraisemblance.

$$H_0: G = -2 \log \left(\frac{\text{Vraisemblance du modèle sans les variables}}{\text{Vraisemblance du modèle avec les variables}} \right) = 0$$

– **Règle de décision :**

Au niveau de signification α : on rejette H_0 si $G > \chi_{J,1-\alpha}$.

3.3. Test de Hosmer-Lemshow

Parmi les tests pour s'assurer de la qualité d'ajustement du modèle il y a le test de Hosmer-Lemshow:

La population est divisée en g classes d'effectifs marginaux égaux en se basant sur les percentiles de la probabilité estimée. Les observations sont triées selon la probabilité de réalisation de l'événement dans un ordre croissant.

On définit la statistique de Pearson :

$$X_{HL} = \sum_{i=1}^g \frac{(O_i - N_i \bar{q}_i)^2}{N_i \bar{q}_i (1 - N_i \bar{q}_i)}$$

Avec :

- O_i : Nombre d'événement dans le groupe i ;
- N_i : Nombre d'individus dans le groupe i ;
- \bar{q}_i : La probabilité moyenne estimée d'un événement dans le groupe i.

Cette statistique suit une loi khi-deux χ^2 de $g - 2$ degrés de liberté. Elle mesure l'ajustement global du modèle.

Pour une p-value supérieure à 0.05, on ne rejette pas l'hypothèse H_0 : le modèle est adéquat.

II. Analyse des données

1. Travaux d'épuration et la fiabilisation de la base de données

L'étude a concerné les fichiers relatifs aux exercices 2007, 2008, 2009, 2010, 2011 et 2012.

Après la collecte des données vient la phase de l'épuration à l'aide des tests de fiabilité. Cette phase consiste à effectuer une opération d'audit pour détecter l'ensemble des anomalies pouvant exister dans les fichiers production et sinistre et ceci avant de procéder à les fusionner.

La première chose à faire était, pour chaque exercice donné, d'agrèger les lignes par numéro de police et par numéro d'affilié. En effet, les tables qui nous ont été données contiennent, pour le même individu, plusieurs lignes indiquant chacune un évènement qui a survenu durant cet exercice. Le but était de ne garder qu'une seule ligne par individu sommant ainsi les nombres de sinistres et les charges correspondantes.

Une fois l'agrégation faite, nous sommes passés aux tests de fiabilité. Les tests effectués sont au nombre de huit, et ils concernent les problèmes se rapportant aux :

- doublons;
- date de fin de garantie antérieure à date début de garantie;
- date de naissance aberrante ;
- charge < 0 ;
- nombre de sinistres > 0 et charge nulle;

Afin de ne pas biaiser notre étude, lorsqu'on détecte une des anomalies ci-dessus, il ne faut en aucun cas supprimer la ligne correspondante, sauf pour les doublons, mais simplement remplacer la valeur par un ".", ce qui la qualifie de valeur manquante.

Après la fiabilisation des bases de données nous avons fusionné les fichiers production et sinistres et parallèlement, nous avons écarté un certain nombre de variables tout en créant d'autres qui sont nouvelles et qui seront utiles pour la suite de l'étude. Ainsi les variables retenues sont:

Les variables explicatives :

- L'âge de l'affilié.
- Le sexe.

- Le lien de parenté avec l'assuré.
- La situation matrimoniale.
- Taux de remboursement.
- Plafond de la garantie.
- La catégorie socio-professionnelle.
- La profession de l'affilié.

Les variables d'intérêt (à expliquer) :

- le nombre de sinistres.
- la charge des sinistres.

Le travail le plus pénible est la construction et l'épuration de la base de données, car il influence toute l'étude de la tarification, surtout la segmentation et la codification. Finalement, nous avons gardé que les variables qui vont expliquer les variables d'intérêt, ce qui va être explicité à partir du chapitre suivant.

2. Statistiques descriptives

Les statistiques descriptives préliminaires liées à la description des différentes composantes du portefeuille (nombre des affiliés, prime totale, charge totale, S/P etc...) ne seront pas élaborer dans cette étude et ceci pour des raisons de confidentialité de ces informations et leurs appartenances à AXA Assurance Maroc seule.

Par la suite, toutes informations utiles présentées et utilisées dans le reste de ce chapitre ne représentent pas les vraies informations et valeurs, et sont utilisées seulement afin de faciliter l'explication et la démonstration des méthodes et des approches adoptées et de mieux garder la logique des étapes utilisées.

3. Segmentation et codification des variables

3.1. Matrice de corrélation des variables

La matrice de corrélation regroupe les corrélations de plusieurs variables entre elles, les coefficients de cette matrice indiquent l'influence que les variables ont les unes sur les autres. L'âge et le sexe deux variables qui expliquent bien le nombre de sinistre, ajoutant d'autres variables qu'on supposera explicatives et qu'on testera leurs significativité selon chaque modèle par la suite.

Ces variables sont :

- **lien**: cette variable représente le type de relation entre l’affilié et les personnes à charges : Assuré, conjoint ou enfant.
- **Profession** : Cette variable représente le type de profession de l’assuré. Il existe plus de 15 types de profession dans notre base de données.
- **Taux de remboursement** : il s’agit du taux de remboursement des charges réelles des sinistres dans le cas de leurs survenances, ce taux se diffèrent d’une police à une autre et se détermine dans les conditions particulières des polices.
- **Catégorie socio-professionnelle** : Cette variable représente la catégorie socio-professionnelle de l’assuré au sein de l’entreprise dont il travail, nous trouvons qu’il existe des sociétés qui préfère opter pour des différentes caractéristiques de l’assurance maladie pour leurs collaborateurs et ceci selon leurs positions, ou salaires etc., au sein de ces sociétés. Ces caractéristiques se portent surtout sur le taux de remboursement et le plafond de la garantie.
- **Situation matrimoniale** : il s’agit de la situation conjugale de l’affilié : célibataire, marié, veuf ou divorcé.

D’après la base de données de la maladie nous avons établi la matrice de corrélation suivante entre les variables pour voir est ce qu’il y a dépendance ou non entre ces variables.

	âge	sexe	lien	profession	catégorie socio professionnelle	taux de remboursement	situation matrimoniale
âge	1.0000	-0.232	-0.3852	0.2540	-0.0538	0.0321	-0.3418
sexe	-0.232	1.0000	0.1272	-0.1291	0.0286	0.0708	-0.0370
lien	-0.3852	0.1272	1.0000	-0.2730	0.0821	-0.0053	0.3709
profession	0.2540	-0.1291	-0.2730	1.0000	-0.0230	0.0089	-0.0997
catégorie socio professionnelle	-0.0538	0.0286	0.0821	-0.0230	1.0000	0.0015	-0.0031
taux de remboursement	0.0321	0.0708	-0.0053	0.0089	0.0015	1.0000	-0.0978
situation matrimoniale	-0.3418	-0.0370	0.3709	-0.0997	-0.0031	-0.0978	1.0000

Tableau 9 : matrice de corrélation des variables

Comme nous voyons, les coefficients de la matrice sont compris entre -1 et 1, plus ce coefficient pour deux variables tend vers 0 plus nous avons indépendance entre ces variables sinon si ce coefficient est proche de 1 ou -1, donc il existe une dépendance linéaire entre ces variables. Nous constatons en général une indépendance linéaire

presque totale entre les variables par exemple, entre âge et sexe, âge et taux de remboursement, catégorie socio professionnelle et taux de remboursement tandis que l'âge montre des similitudes avec la situation matrimoniale et le lien.

3.2. *Algorithme CHAID*

La segmentation des variables est une étape importante du processus de modélisation. En effet elle nous permet de passer d'une infinité de classes de risques à un nombre fini de classes de risque facilement interprétables. Pour faire la segmentation nous allons utiliser les arbres de décision. Il existe plusieurs algorithmes de construction de ces arbres : l'algorithme CRT, l'algorithme CHAID etc....

Chaid est l'acronyme de -Chi-square Automatic Interaction Detected, il désigne comme on dit tantôt un algorithme de segmentation de variable. Il se base pour cela sur une variable dépendante qui lui servira de guide. L'algorithme subdivise la population initiale en plusieurs ensembles distincts en se basant sur les variables indépendantes les plus significatives, chacun des groupes ainsi obtenu est aussitôt subdivisé en deux sous-groupes distincts ou plus en considérant les variables restantes. On répète ce processus jusqu'à la fin ; soit toutes les variables sont utilisées, soit une condition spécifiée par l'utilisateur est réalisée. Le résumé de toutes ces démarches est donné sous forme d'un arbre de décision où chaque nœud représente un regroupement d'observations.

Il y a trois étapes importantes lors de l'exécution de l'algorithme :

- *Etape de fusion*

C'est l'étape la plus importante de l'algorithme. Elle consiste à mélanger les modalités des différentes variables indépendantes. Notons que ces variables doivent être nominales ou ordinales. Si des prédicateurs sont continus, ils sont transformés avant que l'algorithme ne commence. Les phases les plus importantes de la fusion sont donc :

- Mélanger deux modalités « vraisemblables » du prédicateur. Vraisemblables en ce sens que si la variable est ordinale, ce sont des modalités adjacentes et si elle est nominale, n'importe quelles modalités.
- Choisir le mélange ayant la p-value la plus élevée. Si cette p-value est supérieure au niveau spécifiée par l'utilisateur, le mélange est conservé et on crée ainsi une nouvelle modalité. Le calcul du p-value est décrite plus bas. Toutefois toute modalité ayant une faible quantité d'observations est mélangée avec une autre modalité.

- *Etape de la séparation*

Le meilleur découpage pour chaque prédicateur est obtenu lors du processus de fusion. En effet, à cette étape de l'algorithme, on sélectionne la variable qui a la plus petite p-

value ajustée -calculée lors du processus de fusion. C'est celle qui discrimine le mieux le nœud courant. Si cette p-value est inférieure au niveau de confiance spécifié par l'utilisateur, l'algorithme s'arrête et le nœud devient un nœud terminal.

- **Etape de l'arrêt de l'algorithme**

Le processus d'arrêt de l'algorithme se fait en accord avec un certain nombre de règles prédéfinies. Une seule de ces règles suffit.

- Si tous les nœuds sont purs, autrement dit si ils contiennent des modalités identiques de la variable dépendante.
- Si l'arbre a atteint la limite de croissance spécifiée par l'utilisateur.
- Si la taille d'un nœud est inférieure à la taille minimale spécifiée par l'utilisateur.

La p-value ajustée est un élément important de l'algorithme, il sert tant à fusionner des modalités qu'à les séparer. Cette p-value est calculée de deux manières différentes selon que la variable dépendante est une variable catégorielle ou une variable continue :

Dans le cas où c'est une variable continue, la p-value est obtenue à partir d'un test d'ANOVA. En effet il s'agira de déterminer si les moyennes dans les différentes catégories testées sont égales ou non. La statistique calculée est le rapport entre la moyenne des carrés intra-groupes et la moyenne des carrés intergroupes. Elle suit une Fisher $F(g - 1, N - g)$ où g est le nombre de groupes et N le nombre d'observations

$MC_{inter} = \frac{1}{g - 1} \sum n_i \sum (y_{i.} - y_{..})^2 \quad \text{avec } y_{ij} \text{ } j - \text{eme observation du groupe } i$ $MC_{intra} = \frac{1}{N - g} \sum \sum (y_{ij} - y_{i.})^2 \quad n_i \text{ nombre d'observations du groupe } i$ $F = MC_{inter} / MC_{intra} \quad y_{i.} = \frac{1}{n_i} \sum_I y_{ij} \quad \text{et } y_{..} = \frac{1}{N} \sum_{ij} y_{ij}$

Dans le cas où c'est une variable nominale, la p-value est calculée à partir de la statistique du KHI-DEUX obtenue lors du test d'indépendance entre la variable dépendante et le prédicateur. Si la variable dépendante a **I modalités** et le prédicateur **J modalités** le KHI-DEUX de Pearson se calcule comme suit :

$\sum_I \sum_J \frac{(n_{..}n_{ij} - n_{i.}n_{.j})^2}{n_{..}n_{i.}n_{.j}}$	<p>n_{ij} est le nombre d'observations de niveau i pour la variable dépendante et de niveau j pour le prédicteur</p>
$n_{..} = \sum n_{ij} ; n_{i.} = \sum_J n_{ij} ; n_{.j} = \sum_I n_{ij}$	

III. Modélisation des sinistres graves :

Quand nous considérons une série de sinistres d'assurance, il est habituel de la considérer comme étant une seule distribution. C'est parce que la philosophie classique de l'assurance est centrée autour d'un certain nombre de principes; en particulier le principe de l'indépendance des sinistres et le principe de la mutualisation, ou plus précisément, une homogénéité en terme de sinistralité entre les individus ou particulièrement entre les classes de risques.

La présence de sinistres graves (rares) vient perturber cette hypothèse d'homogénéité des classes et de stabilité des indicateurs de risque comme la prime pure. En général, face à tels événements, les assureurs répartissent ces charges sur l'ensemble du portefeuille (principe d'écrêtement), mais la question qui se pose c'est : quand est-ce qu'un sinistre est considéré comme étant un sinistre grave ? La détection et la modélisation des ses sinistres graves aidera à assurer une certaine stabilité des indicateurs de sinistralité et donc une adéquation entre la prime de référence et la sinistralité.

Parmi les théories qui ont traité cette problématique est celle des valeurs extrêmes et en particulier la méthode basée sur la distribution de Pareto généralisée (GPD), plusieurs méthodes sont proposées (moyenne des excès, estimateur de Hill) pour la détermination d'un seuil à partir duquel un événement est considéré comme atypique. Elles permettent de prévoir des sinistres graves pour une probabilité d'occurrence donnée (très faible) et un intervalle de confiance fixé.

Dans cette partie nous allons déterminer ces seuils pour les différentes garanties (Médicaments, Consultation et Autres) et modéliser la charge moyenne et la fréquence des sinistres dépassants ces seuils (sinistres graves).

1. Modélisation des sinistres graves de la garantie « Médicaments »

D'après les deux graphiques du QQ-plot ci-dessous nous pouvons facilement constater l'existence des sinistres graves qui se présente comme étant des points très éloignés des

droites de la loi et aussi d'après le graphe de Lorenz nous remarquons que presque 2% des sinistres représentent plus de 15% de la charge totale (illustré sur le graphe par les flèches rouges), donc ceci nous confirme l'existence des sinistres graves dans le portefeuille.

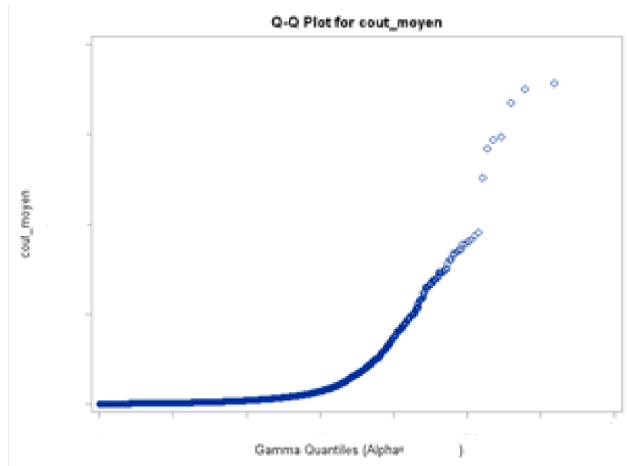


Figure 4 : Q-Q plot Log normal

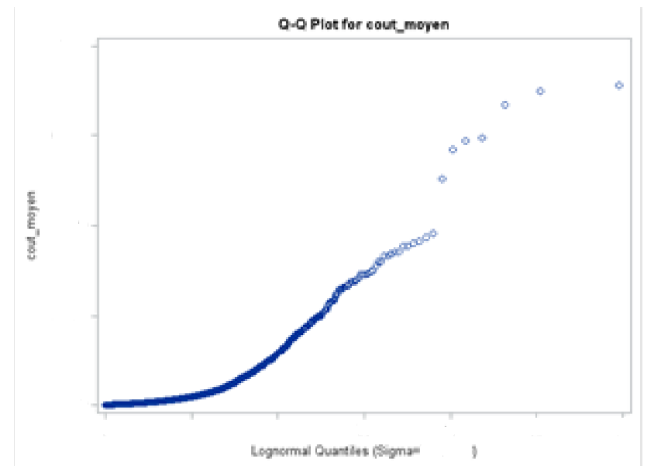


Figure 5 : Q-Q plot gamma

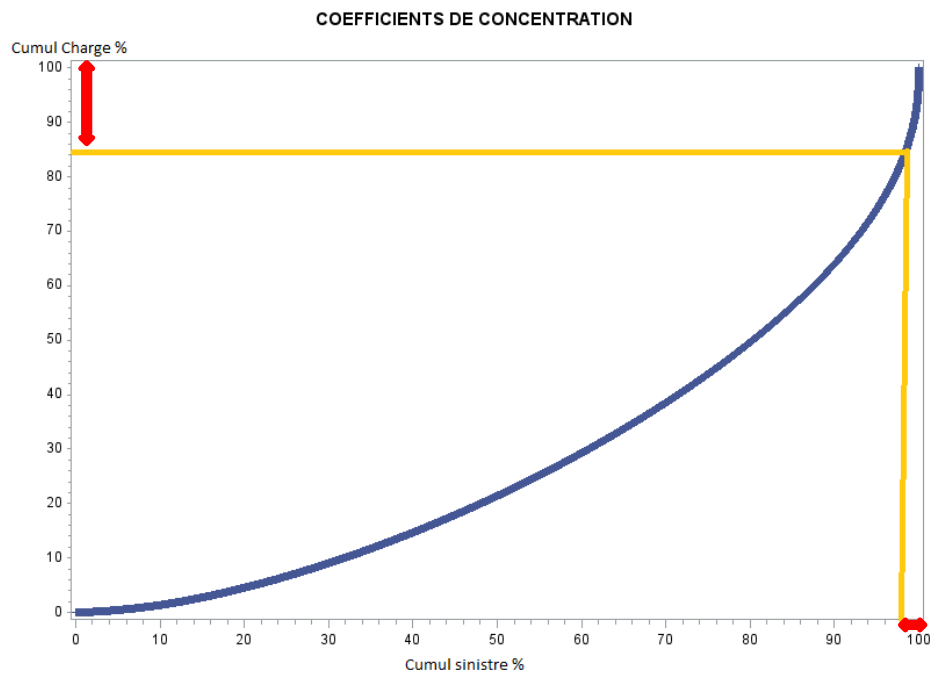


Figure 6 : Courbe de concentration de Lorenz

1.1.L'estimation du seuil

Il existe plusieurs méthodes pour déterminer le seuil au-delà duquel les charges sont considérées de valeurs extrêmes et graves. Nous allons se baser sur deux méthodes graphique pour déterminer le seuil la première méthode est celle de la moyenne des excès et la deuxième est celle de l'estimateur de Hill.

1.1.1. La fonction moyenne des excès

La fonction moyenne des excès est par définition la somme des excès dépassant un certain seuil élevé, noté u , divisé par le nombre de points des données qui dépassent ce seuil.

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n I_{\{X_i > u\}}}$$

$$\text{avec } I_{\{X_i > u\}} = \begin{cases} 1 & \text{si } X_i > u \\ 0 & \text{sinon} \end{cases}$$

Autrement dit, il s'agit d'une estimation de la fonction moyenne des excès (notée FME) qui permet de décrire la prédiction du dépassement du seuil lorsqu'un excès se produit.

Trois cas peuvent se présenter :

Si à un certain seuil, la FME empirique est marquée par une pente positive, les données suivent la distribution GPD avec un paramètre ξ positif.

Si la fonction moyenne des excès est horizontale, les données suivent une distribution exponentielle.

Si la FME empirique est marquée par une pente négative, les données suivent une distribution à queue légère.

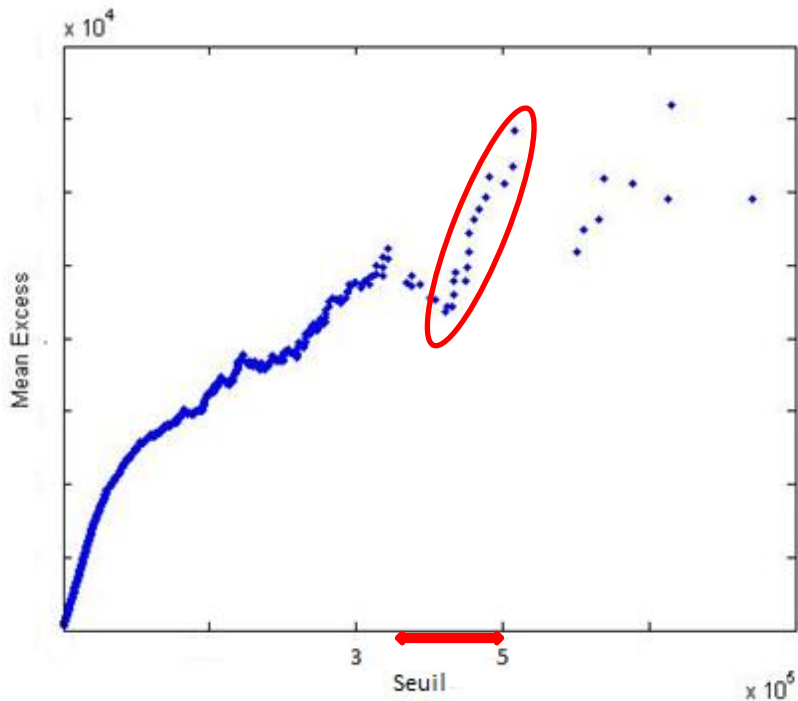


Figure 7 : La fonction moyenne des excès

La figure ci-dessus présente la fonction moyenne des excès d'une distribution GPD avec un paramètre ξ positif. On constate que la FME devient stable pour

un seuil de l'ordre de 350000 à 500000 Dhs, cette stabilité est représentée sous formes d'une droite de pente positive (la partie du courbe entourée par le cercle rouge).

1.1.2. Estimateur de Hill

Hill (1975) a proposé l'estimateur suivant de la distribution GPD :

$$\hat{\xi} = \frac{1}{k-1} \sum_{i=1}^{k-1} (\ln X_{i,n} - \ln X_{k,n}) \text{ pour } k \geq 2$$

Avec k , l'ordre statistique le plus élevé (le nombre des excès), n est la taille de l'échantillon et $\alpha = \frac{1}{\xi}$ est l'indice de la queue de distribution.

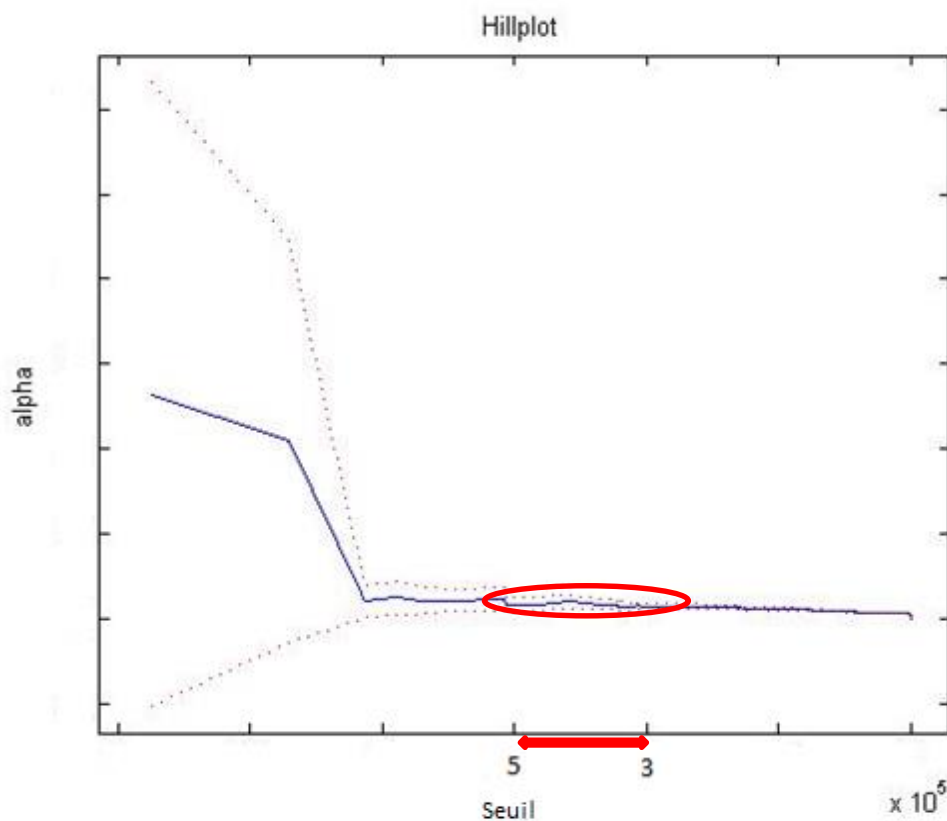


Figure 8: Le Hill-plot

Le graphique Hill-plot nous permet d'avoir des estimations du paramètre α en fonction des différents seuils possible, nous choisissons ainsi l'indice le plus stable.

Dans ce graphique, la stabilité est atteinte pour les seuils entre 300000 et 500000Dhs (la partie de la courbe entourée par le cercle rouge).

On peut aussi représenter, le Hill plot en utilisant les nombres des valeurs extrêmes au lieu des valeurs des seuils.

La représentation ci-dessus nous permet d'estimer le quantile extrême (seuil) en se servant de l'indice de queue de la distribution retrouvé par la même méthode.

Le Hill-plot est donc un outil à double utilité :

- L'estimation de l'indice de la queue de la distribution (α)
- L'estimation du seuil.

1.2.Choix du seuil

D'après les deux méthodes de la moyenne des excès et d'estimateur Hill, nous nous concentrons sur les charges entre 300000 et 500000 DH comme un intervalle contenant les niveaux des seuils possibles. Pour ces différents niveaux, nous utiliserons les tests d'Anderson Darling et Cramer-von-Mises comme un outil pour choisir le seuil le plus adéquat.

Pour cela nous allons suivre les étapes suivantes pour le calcul de ces deux statistiques :

- Pour les charges données $x_{(1)} \leq \dots x_{(i)} \leq \dots x_{(n)}$, nous appliquons la transformation suivante $z_{(i)} = F(x_{(i)})$ en utilisant les paramètres estimés, afin d'avoir un échantillon 'z' uniformément distribué.
- Calculer les statistiques d'Anderson Darling (AD) et Cramer-von-Mises (CvM) de la manière suivante :

$$AD = -n - \left(\frac{1}{n}\right) \sum_{i=1}^n (2i - 1) [\log(z_{(i)}) + \log(1 - z_{(n+1-i)})]$$

$$CvM = \sum_{i=1}^n \left[\frac{(z_{(i)} - (2i - 1))}{2n} \right]^2 + \frac{1}{12n}$$

<i>Tests d'ajustement</i>		
<i>Seuil</i>	<i>Cramer-von-Mises</i>	<i>Anderson-Darling</i>
300000	0.61591	3.57831
350000	0.41828	2.41158
400000	0.35265	1.98574
450000	0.21124	1.18652
500000	0.76282	3.5323

Tableau 10: Les statistiques AD et CvM des différents seuils

D'après le tableau ci-dessus nous allons choisir le seuil 450000 Dhs car celui-ci permet d'avoir la valeur la plus faible pour les deux statistiques d'Anderson Darling (AD) et Cramer-von-Mises (CVM).

1.3. Modélisation et résultat

1.3.1. Modélisation de la charge des sinistres graves

Estimation des paramètres du modèle GPD				
Parameter	Estimate	Standard error	t Value	Approx Pr > t
Theta	****	5465	3.61	0.0036
Xi	****	.	.	.

Tableau 11: Estimation des coefficients du modèle GPD des charges extrêmes

Nous remarquons d'après ce tableau que les p-values sont <5% donc le modèle est significative.

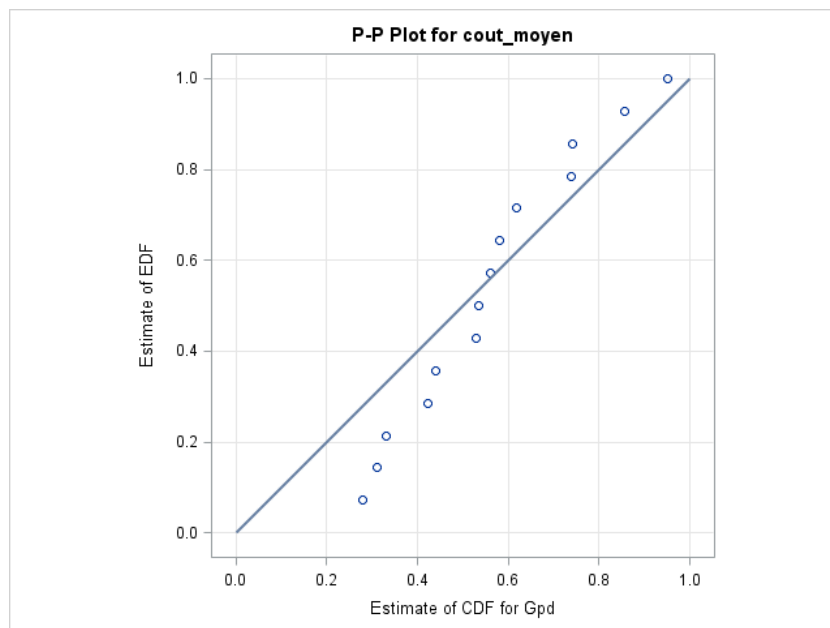


Figure 9: P-P plot des charges extrêmes

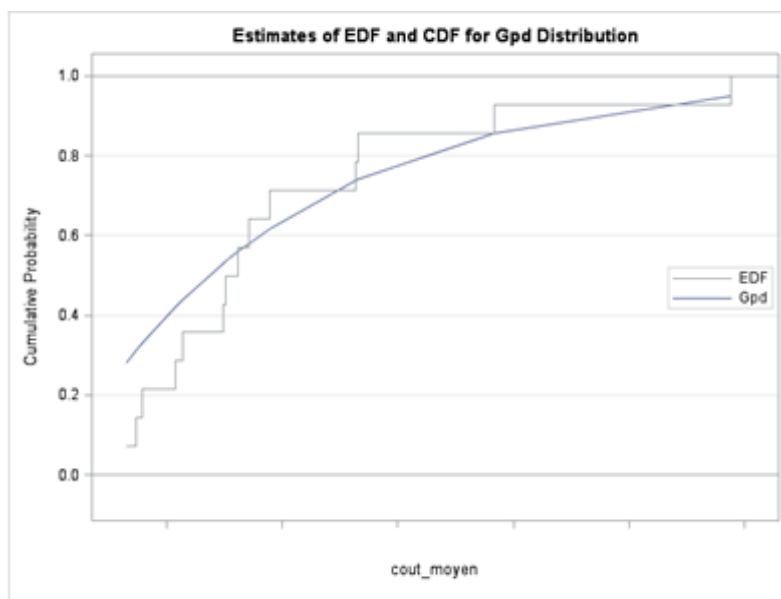


Figure 10: Fonctions de répartition des charges extrêmes (GPD et observé)

Les deux graphiques ci-dessus de PP-plot et de fonctions de distributions nous confirment que la charge des sinistres graves s’ajuste avec un modèle de Pareto généralisé (GPD).

1.3.2. Modélisation de la fréquence des sinistres graves

Après la modélisation du coût moyen des sinistres graves nous abordons maintenant la partie de la modélisation de la fréquence des sinistres graves.

Nous avons choisi les modèles poisson et binomiale négative comme des candidats possible pour modéliser la fréquence des sinistres graves

D’après les deux tableaux ci-dessous nous remarquons que les deux modèles sont significatifs.

Estimation des paramètres du modèle Poisson					
Parameter	DF	Estimate	Standard error	t Value	Approx Pr > t
Intercept	1	****	****	-756.58	<.0001

Tableau 12: Estimation du modèle Poisson pour les sinistres graves

Estimation des paramètres du modèle Binomiale négative					
Parameter	DF	Estimate	Standard error	t Value	Approx Pr > t
Intercept	1	****	****	-624.37	<.0001
Alpha	1	****	****	47.87	<.0001

Tableau 13: Estimation du modèle Binomiale Négative pour les sinistres graves

Donc pour choisir le modèle qui ajuste plus notre modèle, nous allons comparer les Akaike information criterion(AIC) et les Bayesian information criterion(BIC) de ces deux modèles.

Modèle	AIC	BIC
Poisson	233538	233551
Binomiale négative	207396	207423

Tableau 14: Comparaison des modèle Poisson et Binomiale négative

D'après le tableau ci-dessus nous remarquons que le modèle binomiale négative est celui qui s'ajuste plus à notre modèle car celui-ci a les valeurs d'AIC et BIC la plus faible.

2. Modélisation des sinistres graves de la garantie « Consultation » et « Autres »

La même méthodologie et étapes effectués dans cette partie sur la garantie consultation pour la détermination du seuil des sinistres graves et la modélisation de leurs fréquences et charges seront appliqués sur la garantie « Consultation » et «Autres », les résultats et les sorties correspondantes se trouvent dans la partie annexes.

Les seuils des sinistres graves estimés pour les garanties «Consultation » et « Autres » par les deux méthodes 'Mean excess' et 'Hill-plot' sont respectivement 70000 Dhs et 360000 Dhs.

Pour les garanties «Consultation » et « Autres », les résultats des modèles de la fréquence et du coût moyen sont portés respectivement dans les Annexes 2 et 3.

IV. Modélisation de la fréquence et du coût moyen

1. Approche :

Comme nous l'avons mentionné déjà le risque maladie comporte plusieurs garanties (prestations) comme médicament, consultation, hospitalisation dentaire et optique etc.... pour modéliser le risque maladie nous allons modéliser d'abord la garantie médicaments et puis celle de la consultation et pour les prestations restantes, un modèle commun sera élaboré. Cette décision est basée sur le nombre des sinistres survenus pour chaque prestation et aussi sur les limites de la modélisation GLM qui consiste à avoir un nombre minimale des observations pour avoir une régression correcte..

Donc nous allons élaborer dans cette partie trois modèles de la fréquence et trois modèles du coût moyen pour les garanties médicaments, consultation et autres.

2. Modélisation de la garantie « Médicaments »

2.1. Modélisation de la fréquence

2.1.1. Analyse descriptive pour la fréquence

Dans cette partie nous allons donner la fréquence des sinistres pour chaque modalité de chaque variable qui pourrait être significative pour notre modèle et grâce à l'algorithme de «CHAID » nous avons pu avoir des classes de modalités indépendantes entre elle pour chaque variable.

- Age

La variable « âge » est celle qui explique mieux la sinistralité. L'algorithme « CHAID » permet un regroupement des tranches d'âge selon leurs degrés de sinistralité.

Le tableau ci-dessous représente la fréquence des sinistres de chaque modalité et leur poids, donc les modalités 3 et 5 sont les plus sinistrés.

<i>Modalité</i>	<i>Fréquence(%)</i>	<i>Poids(%)</i>
1	20.9%	20.88%
2	15.3%	38.62%
3	80.1%	16.37%
4	53.3%	5.72%
5	70.8%	18.41%

Tableau 15 : Fréquence des médicaments pour l'âge

- **Lien**

Le tableau ci-dessous représente la fréquence des sinistres de chaque modalité et leur poids, on remarque donc que la modalité 2 est plus risquée en termes de sinistres que les autres modalités avec un poids de 50 %.

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	16.40%	20%
2	60.23%	50%
3	35.43%	30%

Tableau 16 : Fréquence des médicaments pour le lien

- **Sexe**

Cette variable est pertinente pour décrire la sinistralité des assurés, d'après le tableau ci-dessous, nous remarquons que la modalité 1 est la plus sinistrée avec une fréquence qui représente presque le double de la modalité 2 avec un poids de 40 %.

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	35.93%	39.02%
2	68.96%	60.98%

Tableau 17 : Fréquence des médicaments pour le sexe

- **Profession**

Toujours en utilisant l'algorithme «CHAID » pour le regroupement de ces types, nous avons obtenu cinq modalités pour la profession. Le tableau ci-dessous représente la fréquence et le poids de chaque modalité. Nous remarquons que la modalité 2 est celle la moins risquée en termes de sinistralité contrairement à la modalité 5.

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	31.11%	18.16%
2	10.43%	30.61%
3	36.00%	19.25%
4	30.04%	5.14%
5	62.14%	26.85%

Tableau 18 : Fréquence des médicaments pour la profession

- **Taux de remboursement**

Après avoir regroupé le taux de remboursement en 5 modalités avec des fréquences différentes, Nous remarquons dans le tableau suivant que la modalité 2 est la plus sinistrée avec un poids élevé de 40 % :

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	37.4%	38.46%
2	86.7%	40.87%
3	30.4%	4.18%
4	40.4%	14.28%
5	50.4%	2.21%

Tableau 19 : Fréquence des médicaments pour le taux de remboursement

- *Situation matrimoniale*

La modalité 2 a une fréquence de 90%, elle est plus sinistrée que les autres d'après le tableau suivant :

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	50.3%	40.37%
2	90.6%	58.80%
3	30.7%	0.83%

Tableau 20 : Fréquence des médicaments pour la situation matrimoniale

- *Catégorie socio-professionnelle*

D'après le tableau suivant la modalité 1 est plus sinistrée que l'autre modalité presque 100% mais avec un poids ne dépassant pas 10%.

<i>Modalité</i>	<i>Fréquence</i>	<i>Poids(%)</i>
1	99.3%	9.22%
2	40.8%	90.78%

Tableau 21 : Fréquence des médicaments pour la catégorie socio-professionnelle

2.1.2. Ajustement et choix du modèle

Avant de commencer la modélisation « GLM » il faut tout d'abord choisir la loi à laquelle s'ajuste mieux avec le modèle, le graphique ci-dessous qui représente l'histogramme de la variable nombre de sinistre nous montre qu'il y a un excès de zéros dans notre base de données, c'est-à-dire la probabilité d'avoir zéro sinistre dépasse 68% donc il faut penser à ajuster notre modèle à un modèle qui prend en considération les excès de zéro comme le ZIP ou ZINB.

Nous remarquons aussi que les probabilités d'avoir plus de quatre sinistres sont presque nuls, ceci est normale puisque l'évènement d'avoir plusieurs risques est considéré rare.

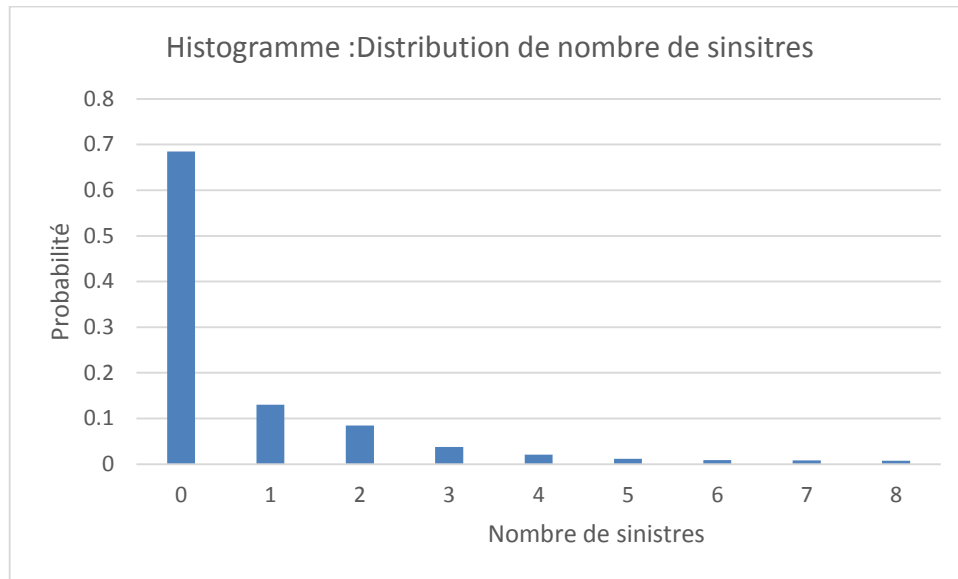


Figure 11 : Histogramme de nombre des sinistres pour les médicaments

Le graphique suivant contient les distributions Poisson, Binomiale négative, ZIP, ZINB et la distribution réelle observée.

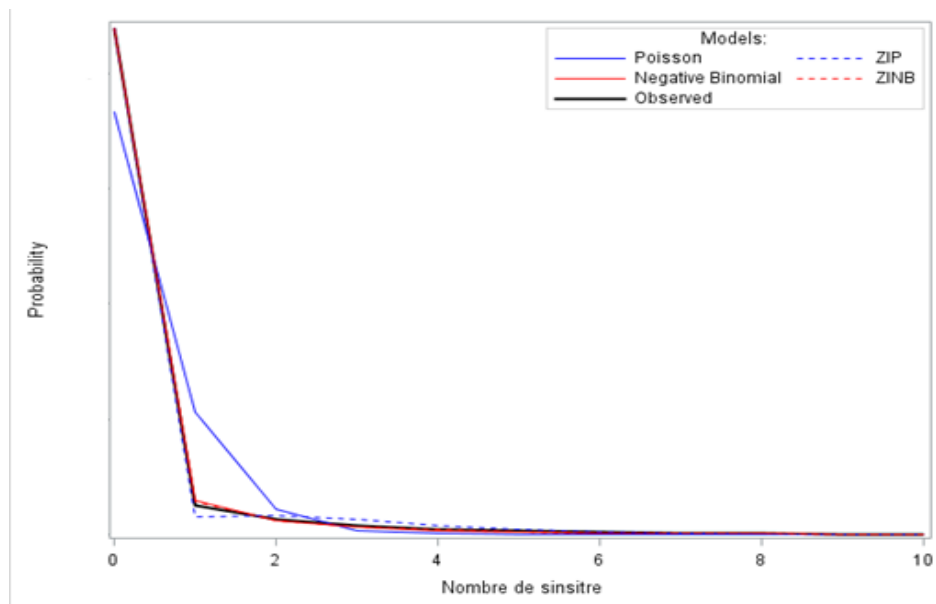


Figure 12 : Distribution des différentes lois et la distribution réelle

Nous constatons que les modèles Poisson et ZIP (les deux courbes bleues), ne s'ajustent pas avec le modèle observé (la courbe noire) surtout la loi de poisson qui nous sous-estime la probabilité d'avoir aucun sinistre. La loi qui s'ajuste le plus avec notre modèle est celle de ZINB car elle suit très bien la courbe du modèle observé, aussi nous

pouvons considérer la loi Binomiale négative comme un candidat possible qui s'ajuste avec notre modèle.

2.1.3. Résultat

Dans ce paragraphe, nous allons donner les résultats de l'estimation des paramètres avec les sorties SAS. L'analyse de type III permet d'examiner la contribution de chacune des variables explicatives, par rapport à un modèle ne la contenant pas. Le tableau ci-dessous nous fournit le résultat de cette analyse. Dans la colonne « khi-deux de Wald », est calculée pour chaque variable deux fois la différence entre la log-vraisemblance obtenue pour le modèle contenant toutes les variables, et le modèle sans la variable en question. La statistique est asymptotiquement distribuée selon une Khi-deux avec DF degrés de liberté, où DF (degree of freedom ou degré de liberté) est le nombre de paramètres associés à la variable explicative examinée. La dernière colonne nous fournit la p-value associée au test du rapport de vraisemblance.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
âge	4	17662.1	<.0001
sexe	1	4603.37	<.0001
lien	2	1004.51	<.0001
sit_mat	2	2712.27	<.0001
taux_remb	4	15860.5	<.0001
cat_socio_prof	1	12231	<.0001
profession	3	13383.9	<.0001
LR Statistics For Type 3 Analysis of Zero Inflation Model			
Source	DF	Chi-Square	Pr > ChiSq
âge	4	14548.3	<.0001
sexe	1	24156.3	<.0001
lien	2	1545.16	<.0001
sit_mat	2	32219.1	<.0001
taux_remb	4	6057.85	<.0001
cat_socio_prof	1	16897.7	<.0001
profession	3	29200.2	<.0001

Tableau 22 : Tableau d'analyse de type 3 pour la fréquence des médicaments

Nous remarquons ainsi, que toutes les variables explicatives sont statistiquement significatives (p-valeur <5%) et l'omission d'une d'entre elle détériorerait

significativement le modèle. Il reste à vérifier la significativité de chaque modalité prise dans le modèle tout entier.

Aussi nous constatons que la variable qui explique plus le modèle est celle d'âge suivie par la variable taux de remboursement.

Par contre nous remarquons que les variables qui expliquent plus la probabilité d'avoir aucun sinistre sont l'état matrimonial et la profession.

D'après les deux tableaux des estimations des paramètres du modèle ZINB portés en Annexe 4 nous remarquons que toutes les modalités des différentes variables sont significatives dans le modèle binomiale négative ainsi que dans le modèle des zéros.

2.1.4 Comparaison et choix du modèle de la fréquence

	Distribution	
	ZINB	Négative binomiale
AIC	7685635	7699835
déviance	59682	77566

Tableau 23 : AIC et la déviance des lois ZINB et Négative Binomiale

D'après le résultat ci-dessus, nous remarquons que l'AIC et la déviance de la loi négative binomiale sont supérieurs à celui de la loi ZINB, nous allons donc choisir le modèle ZINB car il est le meilleur d'après ces deux indicateurs.

2.2. Modélisation du coût moyen

2.2.1. Analyse descriptive pour le coût moyen

Dans ce paragraphe, nous allons donner le coût moyen des sinistres pour chaque modalité de chaque variables qui pourrait être significative pour notre modèle, le choix des modalités est basé généralement sur l'algorithme de «CHAID» pour éviter la dépendance entre les modalités de chaque variable.

Age :

Après avoir regroupée l'âge selon des modalités indépendantes, le tableau suivant nous fournit les résultats du coût moyen et le poids qui correspondent à chaque modalité, le coût moyen le plus élevé correspond à la modalité 4 et la modalité 1 contient le coût moyen le moins faible.

<i>Modalité</i>	<i>coût moyen</i>	<i>Poids(%)</i>
1	190,72	2,31%
2	267,31	3,95%
3	326,8	13,49%
4	556,8	59,63%
5	445,8	20,62%

Tableau 24 : Le coût moyen des médicaments par âge

Sexe :

Le tableau suivant contient le coût moyen des sinistres selon le sexe, la modalité 1 comporte ceux qui ont le coût moyen le plus élevé avec un poids de 60%.

<i>Modalité</i>	<i>coût moyen</i>	<i>Poids(%)</i>
1	313,08	60%
2	283,37	40%

Tableau 25 : Le coût moyen des médicaments par sexe

Lien :

Selon le tableau suivant le coût moyen le plus pesant correspond à la modalité 1 avec un poids de 26 % mais 50 % du poids correspond au coût moyen le moins coûteux de la modalité 3.

<i>Modalité</i>	<i>coût moyen</i>	<i>Poids(%)</i>
1	394,15	25,12%
2	300,03	25,15%
3	201,27	49,73%

Tableau 26 : Le coût moyen des médicaments par lien

Taux de remboursement :

Les taux de remboursement ont été regroupés en quatre classes indépendantes, d'après le tableau le coût moyen le plus faible correspond à la modalité 1 mais avec un poids ne dépassant pas 5%, le coût moyen le plus élevé est pour la modalité 4 avec un poids de 24 %.

<i>Modalité</i>	<i>coût moyen</i>	<i>Poids(%)</i>
1	79,57	4,56%
2	229,71	11,31%
3	303,9	59,59%
4	378,9	24,54%

Tableau 27: Le coût moyen des médicaments par taux de remboursement

Catégorie socio professionnelle :

Selon le tableau suivant, le coût moyen le plus élevé est celui de la modalité 2 avec un poids très important de l'ordre de 78%.

<i>Modalité</i>	<i>Coût moyen</i>	<i>Poids(%)</i>
1	200,2	22.22%
2	321,1	77.78%

Tableau 28 : Le coût moyen des médicaments par catégorie socio professionnelle

Situation matrimoniale :

Pour la situation matrimoniale, le tableau suivant donne les coûts moyens selon les modalités choisies, on a la modalité 3 est celle qui donne le coût moyen le plus élevé mais avec un poids très petit de 2%.

<i>Modalité</i>	<i>Coût moyen</i>	<i>Poids(%)</i>
1	196,12	41,20%
2	256,18	56.80%
3	351,59	2%

Tableau 29 : Le coût moyen des médicaments par situation matrimoniale

Profession :

Comme les autres variables, le coût moyen est donné par le tableau suivant selon les cinq modalités choisies, le coût moyen le plus élevé est celui de la modalité 4 et le plus faible est celui de la dernière modalité avec un poids identique ne dépassant pas 4 %.

<i>Modalité</i>	<i>coût moyen</i>	<i>Poids(%)</i>
1	326,56	12,3%
2	267,31	19,6%
3	201,3	60,6%
4	406,89	3,9%
5	197,26	3,6%

Tableau 30 : Le coût moyen des médicaments par profession

2.2.2. Ajustement et choix du modèle

Les deux modèles les plus classiques permettant de modéliser le coût moyen des sinistre sont le modèle Gamma et le modèle log-normal. Les deux graphiques suivants montrent l'ajustement des deux lois avec la distribution réelle des charges.

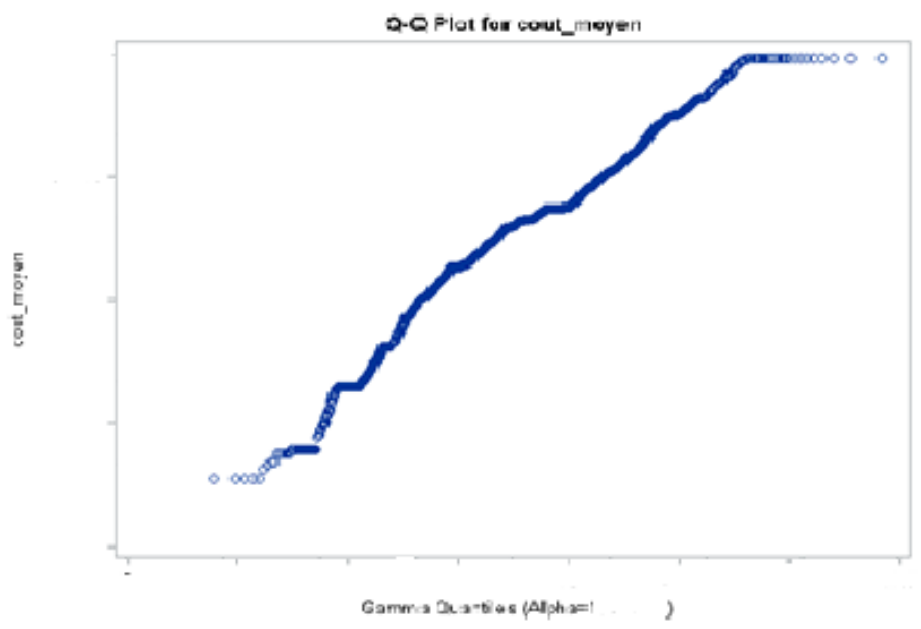


Figure 13 : Q-Q plot Gamma pour le coût moyen des médicaments



Figure 14 : Q-Q plot Log normal pour le coût moyen des médicaments

Nous remarquons donc que les deux lois s'ajustent avec notre modèle mais dans cette partie nous allons représenter que les résultats du modèle Gamma.

2.2.3. Résultat

D'après le tableau des estimations des paramètres du modèle Gamma « Tableau 1 » porté en Annexe 5, nous remarquons que la modalité 2 de la variable lien et la modalité 3 de la variable profession ne sont pas significatives ($p\text{-value} > 0.5$) ce qui nous a emmené à enlever de notre modèle les deux variables qui contiennent ces modalités pour avoir le résultat final suivant selon les sorties SAS ci-dessous :

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
âge	4	75017.2	<.0001
sexe	1	318.78	<.0001
sit_mat	2	94.86	<.0001
taux_remb	3	107534	<.0001
cat_socio_prof	1	59.42	<.0001

Tableau 31 : Tableau d'analyse de type 3 pour le coût moyen des médicaments pour les variables restantes

Nous constatons d'après l'analyse de type 3 que la variable qui explique plus la charge des sinistres est le taux de remboursement suivi par l'âge.

Nous remarquons que ces deux variables sont les plus explicatives dans les deux modèles de la fréquence et la charge moyenne des sinistres, donc elles sont des variables tarifaires importantes dans la détermination des classes tarifaires.

D'après le tableau des estimations des paramètres du modèle Gamma pour les variables significatives restantes, « Tableau 2 » porté en Annexe 5, nous constatons que toutes les variables sont significatives ainsi que chacune de leurs modalités.

Donc nous allons retenu le modèle Gamma comme un modèle du coût moyen de la garantie « Médicaments ».

2.3.calcul de la prime pure

La prime pure de la garantie « Médicament » est défini comme suit :

$$PP = f_{sinistre\ grave} \times CM_{sinistre\ grave} + f \times CM$$

PP : la prime pure

$f_{sinistre\ grave}$: Fréquence des sinistres graves

$CM_{sinistre\ grave}$: Coût moyen des sinistres graves

f : fréquence des sinistres ordinaires

CM : coût moyen des sinistres ordinaires

Comme nous avons vu, le taux de remboursement et l'âge sont deux critères qui expliquent le plus la sinistralité, les résultats de la prime pure pour les médicaments sont donnés selon ces deux critères par le tableau suivant :

	<i>Modalité</i>	<i>Prime pure</i>
Age	1	257,472
	2	360,8685
	3	441,18
	4	751,68
	5	601,83
Taux de remboursement	1	107,4195
	2	310,1085
	3	410,265
	4	511,515

Tableau 32 : Prime pure des médicaments selon l'âge et le taux de remboursement

On a donc la prime pure la plus élevée correspond à la modalité 4 de l'âge et du taux de remboursement, d'autre part la prime pure de la modalité 1 est la moins élevée pour les deux critères.

3. Modélisation de la garantie « Consultation » et la garantie « Autres »

La garantie « consultation » comprend toutes les consultations généralistes et spécialistes, nous allons modéliser sa fréquence et son coût moyen de la même manière que nous avons déjà modéliser la garantie « médicaments ». Aussi pour la garantie « Autres » qui comprennent toutes les garanties restantes comme optique, dentaire, maternité, et hospitalisation nous élaborons des modèles pour la fréquence et la charge moyenne de ses sinistres.

Comme dans le cas de la garantie « médicaments », le modèle adéquat pour la fréquence de la « consultation » et la garantie « autre » est le modèle ZINB, il est le plus ajusté à notre modèle réelle observée. Pour le coût moyen nous avons choisi le modèle Gamma pour les deux garanties car il est le mieux ajusté à nos données.

Les sorties SAS des résultats finaux pour les modèles retenus pour le coût moyen et la fréquence des deux garanties sont données par les Annexes 6 et 7.

4. prime totale

La prime totale de la maladie est la somme des primes pures de toutes les garanties, ici nous avons trois garanties : médicaments, consultation et autres. La prime pure totale obtenu de la maladie selon les deux variables les plus explicatives, l'âge et le taux de remboursement est donnée par le tableau suivant :

	<i>Modalité</i>	<i>Prime pure</i>
Age	1	1621,12
	2	2272,135
	3	2777,8
	4	4732,8
	5	3789,3
Taux de remboursement	1	716,13
	2	2067,39
	3	3410,1
	4	2735,1

Tableau 33 : Prime totale de toutes les garanties par âge et taux de remboursement

Finalement nous avons calculé les primes estimées pour la maladie selon plusieurs critères et variables tarifaires.

5. L'indicateur de rentabilité S/P :

Le Ratios S/P est un indicateur calculé par l'assureur afin de vérifier la rentabilité généralement d'une police, Il consiste à calculer le rapport entre le montant des sinistres et celui des primes. Il peut être pris de manière individuelle pour chaque assuré, ou de manière globale pour déterminer la rentabilité de l'activité. Si le ratio est supérieur à un, cela signifie que la compagnie d'assurance est déficitaire. En effet, nous avons calculé les S/P estimés pour chaque garantie afin de mesurer la rentabilité de la maladie dans notre portefeuille DIM, ici nous donnons les S/P par garantie, les S/P par âge et taux de remboursement pour l'ensemble des garanties de la maladie.

Par garantie :

Après avoir calculé les primes estimées et les montants des sinistres pour chaque garantie, nous avons obtenu le ratio S/P selon le tableau suivant :

Garantie	montant des sinistres	primé pure estimée	S/P
médicaments	8354320	9531455	87,65%
consultation	2329519	3478971	66,96%
autre	14496246	20582487	70,43%

Tableau 34 : S/P estimé pour chaque garantie

Nous remarquons que le ratio S/P ne dépasse pas en général 88% pour ces garanties, nous avons donc une rentabilité pour la consultation proche de la rentabilité de la garantie « autre » et beaucoup mieux que celle des médicaments.

Par âge et taux de remboursement :

Le tableau suivant contient la rentabilité pour les modalités des deux variables tarifaires « âge » et «taux de remboursement» :

	Modalité	montant des sinistres	Prime pure	S/P
Age	1	917240702,8	1621139453	56,58%
	2	2027677606	2272162266	89,24%
	3	1922538450	2777833334	69,21%
	4	4696413796	4732856794	99,23%
	5	3169408552	3789345472	83,64%
Taux de remboursement	1	329495366,9	716138593,6	46,01%
	2	1977275523	2067414809	95,64%
	3	2727430709	3410140921	79,98%
	4	2233509462	2735132821	81,66%

Tableau 35 : S/P estimé par âge et taux de remboursement

Nous constatons d'après ces résultats que le S/P de la modalité 3 de l'âge et le S/P de la modalité 2 du taux de remboursement est compris entre 95% et 100 %, d'où la rentabilité est presque négligeable pour cette catégorie d'où il faut penser à des solutions pour que nous ayons plus de rentabilité comme cibler commercialement à intégrer plus des individus et des entreprises dont les affiliés sont plus de type autres classes que celle-là.

CHAPITRE 3 : Provisionnement

Dans cette partie nous allons s'intéresser à estimer les provisions de la branche maladie et plus particulièrement aux provisions pour sinistres à payer (PSAP) définies par l'arrêté réglementaire du 10 juin 1996 comme la valeur estimative des dépenses en frais et en principal, tant internes qu'externes nécessaire au règlement de tous les sinistres survenus à laquelle on ajoute les frais de gestion des sinistres ainsi que la charge des sinistres tardifs.

Nous allons traiter dans un premier lieu les méthodes déterministes à savoir la méthode de Chain Ladder et sa correction par London Chain (éléments de la théorie et exemple d'application sous tableur Excel) et aussi les deux méthodes de DeVyllder et Taylor. Dans un second lieu nous allons s'intéresser aux méthodes stochastiques à savoir le modèle de Mack et la méthode Bootstrap.

I. Généralité

Le provisionnement est une technique d'estimation des provisions correspondant aux charges à prévoir pour faire face à la sinistralité déclarée mais non encore réglée et à la sinistralité non encore déclarée mais prévisible des contrats en cours. Le provisionnement n'est pas un problème de modélisation complexe et sophistiqué, mais plutôt un exercice de choix de modèle ; quel modèle dois-je utiliser compte tenu de mes données ? En ce sens, le provisionnement n'est plus une science, mais davantage un art, où le savoir-faire et l'expérience de l'actuaire l'emportent. Il existe deux types de méthodes pour le provisionnement que nous allons utiliser : les méthodes déterministes et les méthodes stochastiques, leur but est de prévoir combien il reste à payer pour les sinistres survenus et de quantifier l'incertitude.

1. La dynamique de vie des sinistres

A la date t , la compagnie d'assurance est tenue de constituer une provision pour les sinistres survenus avant la date t qu'elle sera tenu d'indemniser. Elle doit donc estimer le coût des sinistres survenus, et retrancher les montants déjà versés. Il s'agit donc fondamentalement d'un problème de prévision. En effet et contrairement à l'hypothèse faite dans la plupart des modèles actuariels, les coûts de sinistres ne sont pas connus le jour de la survenance du sinistre. Il y a tout d'abord un délai avant que le sinistre ne soit déclaré à la compagnie d'assurance par l'assuré, puis un temps (plus ou moins long) de gestion du sinistre, d'expertises, de paiements, avant de le clôturer plusieurs mois, ou plusieurs années plus tard.

La vie d'un sinistre peut être plus ou moins simple, mais on peut schématiquement la décrire sous la forme présentée sur la figure suivante, avec un diagramme de Lexis.

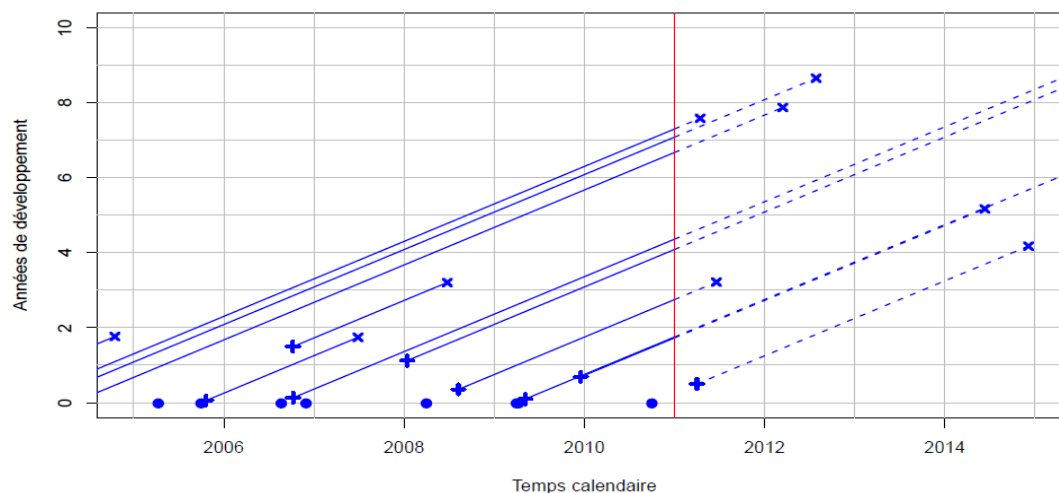


Figure 15 : Diagramme de Lexis pour la dynamique de vie des sinistres

Les sinistres surviennent à la date *, sont déclarés à l'assureur à la date + et clôturés à la date *. L'exercice de provisionnement consiste à estimer à une date donnée (ici 2010, correspondant au trait plein vertical), le montant des paiements restant à faire pour l'ensemble des sinistres survenus (déclarés ou pas).

En pratique, le jour de la déclaration du sinistre à l'assureur (+), la gestion des sinistres est tenu d'estimer un montant pour le sinistre (à l'aide de facture à sa disposition, ou de coûts moyens de sinistres similaires). Le montant réel du sinistre ne sera connu que le jour de la clôture (*). Entre ces deux dates, le gestionnaire de sinistre peut réviser ses estimations de coûts, mais aussi effectuer des paiements. Toutefois, au lieu de travailler sur des données individuelles, les données sont agrégées par années: on s'intéresse à l'année de survenance du sinistre (notée i) et l'année du paiement (par rapport à l'année de la survenance, notée j).

2. Formalisation du problème de provisionnement

Les méthodes de provisionnement sont toutes basées sur les triangles des incréments de paiements reflétant la dynamique des sinistres et permettant d'avoir une vision agrégée de ceux-ci, les notations sont les suivantes :

- i (en ligne) l'année de survenance, $i = 1, \dots, n$.
- j (en colonne) l'année de développement, $j = 0, \dots, n - 1$.

- $y_{i,j}$ les incréments de paiements, pour l'année de développement j et pour l'année de survenance i .

Année de survenance	Développement				
	1	...	j	...	n
1	$y_{1,1}$...	$y_{1,j}$...	$y_{1,n}$
.	.				.
.	.			.	.
i	.			$y_{i,1}y_{i,j}$.
.	.		.		.
.	.		.		.
.	.		.		.
n	$y_{n,1}$.		.

Tableau 36 : Triangle des incréments de paiements

Le provisionnement est un problème de prédiction, conditionnelle à l'information dont on dispose à la date n . On notera H_n l'information disponible à la date n , soit formellement :

$$H_n = \{(y_{i,j}), i+j \leq n\} = \{(C_{i,j}), i+j \leq n\}$$

On cherche à étudier, par année de survenance, la loi conditionnelle de $C_{i,1}$,

Sachant H_n , ou plutôt, si l'on suppose les sinistres clos au bout de n années

la loi de $C_{i,n}$ sachant H_n . Si l'on se focalise sur une année de survenance particulière

- $c_{i,j}$ les paiements cumulés, au sens ou $c_{i,j} = y_{i,0} + y_{i,1} + \dots + y_{i,j}$,

, on pourra noter :

$$F_{i,n-i} = \{(y_{i,j}), j=0, \dots, n-i\} = \{(C_{i,j}), j=0, \dots, n-i\}$$

Cette notation permet de prendre en compte que l'information disponible change d'une ligne à l'autre, on cherchera par la suite à prédire le montant des sinistres à payer pour l'année i , i.e.:

$$\hat{C}_{i,n}^{(n-i)} = E[C_{i,n}/F_{i,n-i}]$$

Et la différence entre ce montant et le montant déjà payé constituera la provision pour sinistres à payer :

$$\hat{R}_i = \hat{C}_{i,n}^{(n-i)} - C_{i,n-i}$$

II. Les méthodes déterministes

Les méthodes déterministes reposent sur l'hypothèse de stabilité du délai s'écoulant entre la survenance d'un sinistre et le(s) règlement(s), quel que soit l'exercice de survenance, en absence d'inflation, de changement de structure de portefeuille, des garanties des contrats, des franchises, et plus généralement de la gestion des sinistres. Si toutes des hypothèses sont vérifiées sur une période suffisamment longue (au moins 5 ans), les méthodes déterministes peuvent être un premier outil intéressant pour prévoir la charge finale, en utilisant les cadences de règlement observées sur le passé. Nous allons définir les méthodes déterministes les plus utilisés que nous allons appliquer par la suite.

1. La méthode de Chain Ladder

Cette méthode compte parmi les plus populaires, car elle est facile à mettre en œuvre, et facile à comprendre. L'idée est que le déroulement des paiements est gouverné par des facteurs de déroulement (notés λ_j) qui ne dépendent que de l'année de déroulement : le modèle sous-jacent est alors de la forme :

$$C_{i,j} = \lambda_j C_{i,j-1}.$$

Les paramètres intervenant dans la méthode Chain Ladder présentent l'avantage d'avoir une interprétation claire, et sont facilement estimable. L'inconvénient étant que cette estimation est relativement peu robuste. Cette méthode ne fait aucune hypothèse quant à la loi que peuvent suivre les coûts des sinistres, ou leur fréquence.

La méthode Chain Ladder est fondée sur l'utilisation de Link-ratios encore appelés coefficients de passage, coefficients de déroulement ou coefficients de développement, entre les différents années de développement.

Les hypothèses sous-jacentes sont :

Hypothèse1 : les années de survenance sont indépendantes entre elles.

Hypothèse2 : les années de développement sont les variables explicatives du comportement des sinistres futurs.

Cette méthode consiste à supposer que les $C_{i,j}$ sont liés par un modèle de la forme :

$$C_{i,k+1} = \lambda_k C_{i,k} \quad \text{Pour tout } i, k = 1, \dots, n$$

Les coefficients λ_k sont appelés link-ratios. Ils peuvent être estimés, à l'aide des observations, par le rapport des totaux relatifs aux éléments communs de deux colonnes successives, c'est-à-dire par :

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}$$

Et à partir de ces coefficients de passage, il est alors possible d'obtenir une estimation des montants de provisions, en considérant :

$$\hat{C}_{i,j} = (\hat{\lambda}_{n+1-i} \dots \hat{\lambda}_{j-1}) C_{i,n+1-i}$$

2. Méthode de London Chain

Dans la méthode dite London Chain, on suppose que la dynamique des $C_{i,j}$ est donnée par un modèle de la forme :

$$C_{i,k+1} = \lambda_k C_{i,k} + \alpha_k \text{ Pour tout } i, k = 1, \dots, n$$

Dans ce modèle, on a alors $2n$ paramètres à identifier :

$$\lambda_k \text{ et } \alpha_k \text{ pour } k = 1, \dots, n.$$

La méthode la plus naturelle consiste à estimer ces paramètres à l'aide des moindres carrés, c'est-à-dire que l'on cherche, pour tout k :

$$(\hat{\lambda}_k, \hat{\alpha}_k) = \arg \min \left\{ \sum_{i=1}^{n-k} (C_{i,k+1} - \alpha_k - \lambda_k C_{i,k})^2 \right\}$$

Ce qui donne :

$$\hat{\lambda}_k = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} C_{i,k} C_{i,k+1} - \bar{C}_k^{(k)} \bar{C}_{k+1}^{(k)}}{\frac{1}{n-k} \sum_{i=1}^{n-k} C_{i,k}^2 - \bar{C}_k^{(k)2}}$$

Où :

$$\bar{C}_k^{(k)} = \frac{1}{n-k} \sum_{i=1}^{n-k} C_{i,k} \text{ Et } \bar{C}_{k+1}^{(k)} = \frac{1}{n-k} \sum_{i=1}^{n-k} C_{i,k+1}$$

Et où la constante est donnée par :

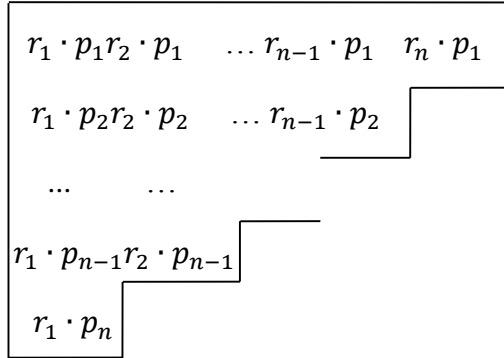
$$\hat{\alpha}_k = \bar{C}_{k+1}^{(k)} - \hat{\lambda}_k \bar{C}_k^{(k)}$$

3. Méthode des moindres carrés de De Vylder

Cette méthode repose sur une modélisation des incréments de la forme :

$$y_{i,j} = r_j \cdot p_i$$

Où p_i correspond à la charge ultime des sinistres survenus au cours de l'année i , et r_j la proportion du montant p_i payé l'année j . le triangle des paiements s'écrit alors comme suit :



Les coefficients r_j et p_i sont obtenus en minimisant la somme des carrés des écarts entre les valeurs observées $Y_{i,j}$ et leur forme théorique $r_j p_i$ soit :

$$\sum_{i+j \leq n} (y_{i,j} - r_j \cdot p_i)^2$$

Avec contrainte d'identifiabilité $r_1 + \dots + r_n = 1$. Ceci donne que :

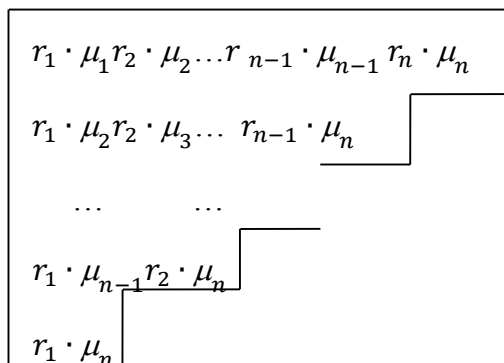
$$\hat{p}_i = \frac{\sum_j \hat{r}_j y_{i,j}}{\sum_j \hat{r}_j^2} \text{ Et } \hat{r}_i = \frac{\sum_i \hat{p}_j y_{i,j}}{\sum_i \hat{p}_j^2}$$

4. Méthode de séparation de Taylor

L'idée dans cette méthode est de considérer l'inflation comme un facteur endogène aux triangles. On travaille pour cela sur les triangles non-agrégés, et l'on suppose que le montant des paiements $y_{i,j}$, relatifs aux sinistres survenus l'année i , et payés l'année $i+j-1$ est de la forme suivante :

$$y_{i,j} = r_j \cdot \mu_{i+j-1} \quad \text{pour tout } i,j.$$

Les triangles des paiements s'écrit alors :



Le but est alors d'estimer les coefficients de développement r_1, r_2, \dots, r_n et les facteurs d'inflation $\mu_1, \mu_2, \dots, \mu_n$. Pour estimer ces $2n$ coefficients, on fera comme précédemment l'hypothèse de normalisation des r_j , à savoir $r_1 + r_2 + \dots + r_n = 1$. Aussi la somme sur la diagonale est :

$$d_n = y_{1,n} + y_{2,n-1} + \dots + y_{n,1} = \mu_n (r_1 + r_2 + \dots + r_n) = \mu_n$$

Pour la première sur-diagonale, on peut noter que

$$d_{n-1} = y_{1,n-1} + y_{2,n-2} + \dots + y_{n-1,1} = \mu_{n-1} (1 - r_n)$$

Et de plus, en considérant la n ème colonne, $\gamma_n = y_{1,n} = r_n \cdot \mu_n$ d'où

$$r_n = \frac{\gamma_n}{\mu_n} \text{ Et } \mu_{n-1} = \frac{d_{n-1}}{1 - r_n}$$

De façon générale, pour la i ème sur-diagonale, on peut noter que

$$\begin{aligned} d_{n-i} &= y_{1,n-i} + y_{2,n-i+1} + \dots + y_{n-i,1} \\ &= \mu_{n-i} (r_1 + r_2 + \dots + r_{n-i}) \\ &= \mu_{n-i} (1 - (r_n + r_{n-1} + \dots + r_{n-i+1})) \end{aligned}$$

Et de plus, en considérant la $n-i+1$ ème colonne,

$$\begin{aligned} \gamma_{n-i+1} &= y_{1,n-i+1} + y_{2,n-i+1} + \dots + y_{i-1,n-i+1} \\ &= r_{n-i+1} \cdot \mu_{n-i+1} + \dots + r_{n-i+1} \cdot \mu_{n-1} + r_{n-i+1} \cdot \mu_n \end{aligned}$$

D'où

$$r_{n-i+1} = \frac{\gamma_{n-i+1}}{\mu_n + \mu_{n-1} + \dots + \mu_{n-i+1}}$$

Et

$$\mu_{k-i} = \frac{d_{n-i}}{1 - (r_n + r_{n-1} + \dots + r_{n-i+1})}$$

5. Application des méthodes déterministes

Dans cette partie nous allons appliquer les différentes méthodes déterministes présentées avant. Nous allons appliquer ces méthodes déterministes à partir de l'exercice de survénance 2004 pour la branche maladie. Les règlements cumulés par année de survénance et année comptable pour la période (2004-2013) sont résumés dans le Tableau 1 porté en Annexe 8.

Pour des raisons de confidentialité des données réelles appartenant à la compagnie d'accueil seule, les données représentées dans ce tableau ne représentent pas les vraies valeurs et ne sont que des valeurs utilisées pour appliquer les différentes méthodes traités dans ce chapitre.

5.1 Méthode de Chain Ladder

La méthode de Chain Ladder cherche à remplir le triangle de Run-Off en un rectangle. Pour cela, nous suivons ces étapes : estimation des facteurs de développement, puis l'extrapolation du triangle de données et enfin la déduction des réserves.

Etape 1 : Estimation des facteurs de développement par la méthode de Chain Ladder

	1	2	3	4	5	6	7	8	9
LINK RATIO	2.05351	1.39078	1.00147	1.00047	1.00024	1.00010	1.00005	1.00003	1
Moyenne	2.06650	1.39437	1.00193	1.00058	1.00026	1.00011	1.00005	1.00003	1.000
Mediane	2.04133	1.39974	1.00033	1.00037	1.00024	1.00010	1.00004	1.00003	
Variance	0.00471362	0.00066972	0.00000689	0.00000033	0.00000002	0.00000000	0.00000000	0.00000000	
Ecart-type	0.0686558	0.02587886	0.002625303	0.000573863	0.000123414	4.71941E-05	1.78381E-05	1.95134E-06	
Coefficient de variation	0.03322	0.01856	0.00262	0.00057	0.00012	0.00005	0.00002	0.00000	

Tableau 37 : Facteurs de développement de Chain Ladder

Nous constatons ici que les coefficients de variation sont très petits, ce qui montre que les facteurs de développement (Link ratio) sont stables. Nous remarquons aussi que seuls les facteurs de développement des années de développement 1 et 2 qui sont supérieures à 1 les autres sont presque égalent à un, ceci est logique car le temps de règlement des sinistres liées à l'assurance maladie ne dépassent pas 3 années et plus de 80% des sinistres sont payé durant les deux premières années.

Etape 2 : Extrapolation du triangle des règlements cumulés

Maintenant nous allons remplir le tableau des règlements cumulés, ces estimations sont représentées par la partie en jaune dans le Tableau 2 porté en Annexe 8, elle représente les montants de sinistres qu'on a retrouvés avec la méthode d'extrapolation de Chain Ladder.

Etape 3 : calcul des réserves

Le tableau ci-dessous représente les réserves par exercice de survenance :

Année de survenance	Réserve
2003	
2004	0
2005	3
2006	7
2007	25
2008	58
2009	351
2010	353
2011	45,946
2012	110,399
TOTAL DES RESERVES	157,142

Tableau 38 : Total des réserves par Chain Ladder

Vérification de l'hypothèse H2 de la méthode de Chain Ladder

L'observation des graphiques portés en Annexe 9 ne montre pas une bonne adéquation des données avec le modèle de Chain Ladder, car les couples $(C_{i,j+1}, C_{i,j})$ et surtout pour $j=1$ et $j=2$ ne sont pas alignés selon une droite passant par l'origine des axes. Donc il faut penser à utiliser la méthode de London Chain pour corriger ce problème.

5.2. Méthode de London Chain

Le tableau ci-dessous représenté les estimations des facteurs en appliquant la méthode de London Chin sur le triangle de règlement présenté précédemment.

	1	2	3	4	5	6	7	8	9
<i>Facteurs</i>	1.9368	1.3551	0.9964	0.9990	0.9999	0.9999	0.9999	1.0000	1.0000
\hat{a}_j	4898.81	2942.15	553.12	145.37	33.42	19.14	11.34	-1.11	0.00

Tableau 39 : Les facteurs de London Chain

L'équation qu'on va utiliser pour l'extrapolation :

$$C_{i,j+1} = f_j * C_{ij} + a_j \text{ Pour tout } i = 1, \dots, n \text{ et } j = 1, \dots, n-1.$$

Le triangle des règlements cumulés estimés est représenté par la partie en jaune dans le Tableau 3 porté en Annexe 8.

Nous trouvons donc les réserves suivantes :

	<i>London chain</i>	<i>Chain Ladder</i>	
<i>Année de survenance</i>	<i>Réserves</i>	<i>Réserves</i>	<i>Différence (%)</i>
2004			
2005	0	0	0.0%
2006	3	3	1.2%
2007	6	7	-17.9%
2008	6	25	-77.7%
2009	22	58	-62.9%
2010	110	351	-96.9%
2011	39	353	-89.1%
2012	44,327	45,946	-3.5%
2013	105,826	110,399	-4.1%
Reserve totale	155,237	157,142	-4.4%

Tableau 40 : Comparaison des réserves estimés de London Chain et Chain Ladder

Nous remarquons que les réserves estimées par la méthode de Chain Ladder sont plus grandes que celles estimées par la méthode de London Chain de presque 4.4%. Ceci est due au fait que l'hypothèse H2 n'était pas vérifiée par la méthode de Chain Ladder.

5.3 Méthode des moindres carrés de De Vylder et de séparation de Taylor

Pour ces deux méthodes nous allons seulement présenter les réserves par années calculées par chaque méthodes et les comparés avec celles obtenus par Chain Ladder.

	<i>DeVylder</i>	<i>Taylor</i>	<i>Chain Ladder</i>	<i>Difference DeVylder/Chain Ladder</i>	<i>Difference Taylor/Chain Ladder</i>
<i>Année de survenance</i>	<i>Réserve</i>	<i>Réserve</i>	<i>Réserves</i>		
2004	-	-			
2005	0	0	0	-	-
2006	3	3	3	1.39%	23.56%
2007	7	8	7	-1.01%	19.82%
2008	25	21	25	-1.07%	-15.73%
2009	55	54	58	-5.19%	-7.99%
2010	120	128	351	-65.69%	-63.50%
2011	275	386	353	-22.05%	9.49%
2012	45,620	51,629	45,946	-0.71%	12.37%
2013	109,351	131,260	110,399	-0.95%	18.90%
TOTAL DES RESERVES	155,456	183,490	157,142	-1.07%	16.77%

Tableau 41 : Comparaison des réserves estimées de London Chain avec Taylor et De Vylder

Nous remarquons ici que la méthode de De Vylder a donné presque les mêmes résultats que Chain Ladder, contrairement à la méthode de Taylor qui a surestimé la valeur des réserves par rapport aux autres valeurs estimés par les autres méthodes avec une différence qui dépasse 16 % par rapport à la méthode de Chain Ladder.

III. Les méthodes stochastiques

L'idée dans cette partie est de pouvoir quantifier la variabilité des réserves estimées, notamment par la construction d'intervalles de confiance, et ainsi d'obtenir une marge d'erreur sur le montant des provisions. Pour cela, les méthodes stochastiques sont adaptées, puisqu'elles considèrent les réserves sous un angle probabiliste en considérant leur distribution.

Introduisons la notion d'information disponible par les données observées avant l'année de développement $I - i$, pour l'année de survenance i :

$$\mathcal{F}_{i,I-i} := \{C_{i,k}, k = 0, \dots, I - i\}$$

Posons également l'ensemble d'information suivant :

$$\mathcal{B}_k = \{C_{i,j}, j \leq k, i + j \leq I\}$$

Qui correspond à l'information disponible à la date actuelle et qui concerne une année de développement antérieure à k .

1. Modèle de Mack

1.1. Présentation du modèle

Le modèle proposé par Mack (1993) est un modèle non paramétrique, conditionnel à la réalisation du triangle, et qui permet d'estimer les erreurs commises lors du calcul des réserves. Il s'agit en fait d'un modèle stochastique pour la méthode de Chain-Ladder.

Les hypothèses sous-jacentes à ce modèle sont les suivantes :

Hypothèse 1 : Les paiements cumulés $C_{i,k}$ sont indépendants par année d'accident i.e $\forall i, j \{C_{i,1}, \dots, C_{i,J}\}, \{C_{j,1}, \dots, C_{j,J}\}$ sont indépendants.

Hypothèse 2 : Il existe des paramètres $f_1, \dots, f_J > 1$ et $\sigma_1^2, \dots, \sigma_J^2$ tels que :

$$\forall i \in \{0, \dots, I\}, k \in \{0, \dots, J\}$$

$$\mathbb{E}(C_{i,k+1} | \mathcal{F}_{i,k}) = C_{i,k} f_k$$

$$\text{Var}(C_{i,k+1} | \mathcal{F}_{i,k}) = C_{i,k} \sigma_k^2$$

Sous ces deux hypothèses, Mack a montré que ce modèle stochastique permet d'obtenir exactement les mêmes montants de provision que la méthode standard de Chain-Ladder, qui est la référence en matière de provisionnement.

On introduit les termes $\gamma_0, \dots, \gamma_J$ qui vérifient : $\forall k \in \{1, \dots, J - 1\}$

$$\begin{cases} \gamma_k = \frac{\gamma_{k+1}}{f_{k+1}} \\ \gamma_J = 1 \end{cases}$$

Ainsi, $\forall i \in \{0, \dots, I\}, k \in \{0, \dots, J\}$:

$$\mathbb{E}\left(\frac{C_{i,k}}{\gamma_k} \mid \mathcal{F}_{i,k-1}\right) = \frac{C_{i,k-1}}{\gamma_{k-1}}$$

1.2. Estimateurs du modèle

On remarque d'après la formule précédente que la famille $\left(\frac{C_{i,k}}{\gamma_k}\right)_{k=1, \dots, J}$ forme une \mathcal{F} -martingale. Nous pouvons donc en déduire l'expression suivante :

$$\mathbb{E}(C_{i,k+1} \mid \mathcal{F}_{i,k}) = C_{i,I-i} \prod_{i=I-i+1}^k f_i$$

On va alors estimer les f_0, \dots, f_J par les estimateurs de Chain-Ladder données par l'expression suivante :

$$\forall k \geq I - i + 1, \hat{F}_k = \frac{\sum_{j=0}^{I-k} C_{j,k+1}}{\sum_{j=0}^{I-k} C_{j,k}}$$

On peut donc définir l'estimateur suivant :

$$\hat{C}_{i,k}^M = C_{i,I-i} \prod_{i=I-i+1}^k \hat{F}_i$$

Et on trouve les propriétés suivantes :

- Les estimateurs \hat{F}_k sont non biaisés : $\mathbb{E}(\hat{F}_k) = f_k$
- Les estimateurs \hat{F}_k sont non corrélés : $\mathbb{E}(\hat{F}_0 \dots \hat{F}_k) = \mathbb{E}(\hat{F}_0) \dots \mathbb{E}(\hat{F}_k)$
- L'estimateur $\hat{C}_{i,k}^M$ est non biaisé pour $\mathbb{E}(C_{i,I} \mid \mathcal{B}_I)$: $\mathbb{E}(\hat{C}_{i,k}^M) = \mathbb{E}(C_{i,I} \mid \mathcal{B}_I)$

2. La méthode de Bootstrap :

Les méthodes de ré échantillonnage, telle que le Jacknife et le Bootstrap (Efron, 1979), permettent de remplacer les déductions théoriques en analyse statistique par des simulations de type Monte-Carlo. Reposant principalement sur des simulations, le Bootstrapping jouit d'une mise en œuvre relativement simple. Depuis quelques années, la technique du Bootstrap est communément utilisée pour analyser la variabilité des montants des sinistres et obtenir des erreurs de prédiction pour différentes méthodes de provisionnement (principalement pour les méthodes basées sur Chain Ladder et sur les modèles linéaires généralisées)

2.1.Méthodologie

La technique de Bootstrap consiste à partir d'un échantillon originel unique, à créer par tirage aléatoire avec remise de nouveaux échantillons .Ces nouveaux échantillons sont constitués de « pseudo données » ayant la même loi sous-jacente. La répétition du ré échantillonnage à partir de données « sources » permet d'estimer la variabilité d'un paramètre déterminer.

La méthodologie doit être adaptée à chaque situation, pour un modèle linéaire (classique ou généralisé), il est commun d'adopter l'une des deux approches suivantes :

- ✓ Le ré échantillonnage effectue directement sur les observations.
- ✓ Le ré échantillonnage appliqué sur les résidus du modèle.

Bien que la première approche soit plus robuste que le Bootstrap sur les résidus, seule cette dernière peut être implémentée en matière de provisionnement. En effet, la technique du Bootstrap nécessite le caractère i.i.d des données. Or, dans un modèle de régression les variables réponses sont indépendantes mais non équidistribuées. Pour respecter le domaine d'application du Bootstrap, il est impératif d'appliquer cette méthode aux résidus du modèle choisi. Ceux-ci ont un caractère « plus » i.i.d que les variables originelles.

L'étape suivante consiste à définir un résidu adéquat. Dans le cadre des modèles linéaires généralisées, il existe principalement deux types de résidus : les résidus bruts et les résidus de Pearson.

- Les résidus bruts correspondent à la différence entre la valeur observée et la valeur espérée : $r_{ij} = Y_{ij} - m_{ij}$

Ils sont d'espérance nulle, de variance non constante et ne sont généralement pas retenus pour l'utilisation de la technique Bootstrap (qui nécessite des résidus i.i.d).

- Les résidus de Pearson ont pour expression : $r_{ij}^{(p)} = \frac{Y_{ij} - m_{ij}}{\sqrt{V(m_{ij})}}$, de variance constante et sont considérés, à cet égard, comme des résidus « adéquats » pour la méthodologie Bootstrap.
- L'expression des résidus empiriques est alors : $\hat{r}_{ij}^{(p)} = \frac{Y_{ij} - \hat{m}_{ij}}{\sqrt{V(\hat{m}_{ij})}}$

Y_{ij} : C'est le montant des paiements connus

m_{ij} : C'est le montant des paiements estimés en utilisant la méthode de Chain Ladder rétrospective.

2.2. Hypothèses de base pour l'utilisation du Bootstrap :

L'hypothèse de base sur laquelle repose le Bootstrap est l'indépendance (d'où le tirage avec remise), ainsi que la distribution identique des lois de chacun des éléments composants l'échantillon.

La démarche du ré échantillonnage :

A partir du triangle des incréments, on constitue un triangle des résidus. Ce triangle sert ensuite de base à la création de k (plus le nombre est grand plus on aura une idée claire de la distribution des réserves) nouveaux triangles des résidus, en effet les valeurs du triangle initial sont « casses » aléatoirement dans chaque nouveau triangle. Plus intuitivement chaque nouveau triangle correspond à une permutation du triangle originel.

La formule : $\hat{r}_{ij}^{(p)} = \frac{Y_{ij} - \hat{m}_{ij}}{\sqrt{V(\hat{m}_{ij})}}$ permet de constituer k triangles des incréments à partir des

k triangles des résidus. L'étape suivante consiste à estimer la partie supérieure du triangle M en utilisant la méthode de Chain Ladder standard rétrospective.

Les étapes de l'application du Bootstrap au calcul des provisions se déroulent comme suit:

- Le point de départ est le triangle de développement des montants cumulés C_{ij} , à partir duquel on va calculer les facteurs de développement à l'aide de la méthode de Chain Ladder.
- Le triangle de règlement et les coefficients de Chain Ladder permettent de créer un nouveau triangle appelé triangle prédit en procédant par BACKWARD-engineering, c'est-à-dire par récursion arrière à partir des valeurs de la diagonale. Les valeurs de la diagonale sont divisées par les coefficients de Chain Ladder.

- Le triangle des résidus de Pearson $r^{(p)}_{ij}$ se calcule à partir du triangle des montants non cumulés Y_{ij} et du triangle des valeurs non cumulées prédites \hat{m}_{ij} par le modèle.
$$\hat{r}_{ij}^{(p)} = \frac{Y_{ij} - \hat{m}_{ij}}{\sqrt{V(\hat{m}_{ij})}}$$
- Après calcul des résidus de Pearson on génère K séries des résidus en utilisant Excel ou R partir de la série des résidus originale.
- Pour chaque série générée des résidus on calcul les Y_{ij} en utilisant la formule suivante : $Y_{ij} = \hat{r}^{(p)} \sqrt{V(\hat{m}_{ij})} + \hat{m}_{ij}$ Pour $i + j \leq n + 1$.
- Après avoir obtenu les K triangles, on refait la méthode de Chain Ladder standard pour estimer la partie inférieure du triangle et déduire les réserves pour chaque année de survenance, ainsi que la charge totale.

3. Application des méthodes stochastiques :

Dans cette partie on va donner pour les deux méthodes (Mack et Bootstrap) les montants de réserves calculées, les écarts-types ainsi que les intervalles de confiances correspondants.

3.1. Modèle de Mack

Année de survenance	Réserve (Ri)	Intervalle de confiance à 95%		
		Ecart-type	borne inf	borne sup
2004				
2005	0.0	0.046	0.0	0.0
2006	2.6	0.277	2.3	2.8
2007	7.0	2.132	4.8	9.1
2008	24.8	7.248	17.6	32.1
2009	58.4	18.560	39.9	77.0
2010	351.1	81.998	269.1	433.1
2011	352.7	366.293	-13.6	719.0
2012	45,946.0	3024.278	42,921.7	48,970.3
2013	110,399.2	6026.808	104,372.4	116,426.0
TOTAL DES RESERVES	157,141.8	6863.139	150,278.6	164,004.9

Tableau 42 : les réserves estimées selon Mack

Donc la réserve totale selon le modèle de Mack s'élève à 157 141 800 Dhs.

3.2. Bootstrap

Année de survenance	Réserve (Ri)	Ecart-type	Erreur (%)	Intervalle de confiance à 95%	
				borne inf	borne sup
2004					
2005	0	0.36	-31333.65%	-0.4	0.4
2006	3	0.43	16.77%	2.1	3.0
2007	7	2.16	31.13%	4.8	9.1
2008	25	7.23	29.13%	17.6	32.1
2009	58	18.53	31.68%	40.0	77.0
2010	139	82.05	58.88%	57.3	221.4
2011	351	365.09	104.06%	-14.2	715.9
2012	45969	3026.18	6.58%	42,943.2	48,995.6
2013	110431	5988.89	5.42%	104,442.0	116,419.8
TOTAL DES RESERVES	156983.4	6916.68	4.41%	150,066.7	163,900.1

Tableau 43 : Les réserves estimées selon Bootstrap

Donc la réserve totale selon la méthode de Bootstrap est de 156 983400 Dhs.

3.3. Comparaison des deux méthodes

Nous allons faire une comparaison des deux méthodes stochastiques afin de trouver les meilleures estimations des réserves par années de survenance et de la réserve totale, pour cela nous présenterons tout d'abord les réserves calculées par les deux méthodes, ensuite nous calculerons les écarts types pour enchaîner avec les intervalles de confiance que nous allons croiser pour trouver à la fin la meilleure estimation des réserves.

Les réserves par année de survenance par le modèle de Mack et Bootstrap sont données par le tableau suivant :

Année de survenance	Mack	Bootstrap
	Réserve (Ri)	Réserve (Ri)
2004		
2005	0.0	0.0
2006	2.6	2.6
2007	7.0	7.0
2008	24.8	24.8
2009	58.4	58.5
2010	351.1	139.4
2011	352.7	350.9
2012	45,946.0	45969.4
2013	110,399.2	110430.9
TOTAL DES RESERVES	157,141.8	156983.4

Tableau 44 : Comparaison entre Mack et Bootstrap pour les réserves estimées

	Mack	Bootstrap
Année de survenance	Ecart-type	Ecart-type
2004		
2005	0.046	0.36
2006	0.277	0.43
2007	2.132	2.16
2008	7.248	7.23
2009	18.560	18.53
2010	81.998	82.05
2011	366.293	365.09
2012	3024.278	3026.18
2013	6026.808	5988.89
TOTAL DES RESERVES	6863.139	6916.68

Tableau 45 : Comparaison des écart-type de Mack et Bootstrap

Dans l'ensemble nous remarquons que les deux modèles Mack et Bootstrap ont presque mêmes réserves par année de survenance et mêmes tendance en termes d'écart-types.

	Mack		Bootstrap	
Année de survenance	borne inf	borne sup	borne inf	borne sup
2004				
2005	0.0	0.0	-0.4	0.4
2006	2.3	2.8	2.1	3.0
2007	4.8	9.1	4.8	9.1
2008	17.6	32.1	17.6	32.1
2009	39.9	77.0	40.0	77.0
2010	269.1	433.1	57.3	221.4
2011	-13.6	719.0	-14.2	715.9
2012	42,921.7	48,970.3	42,943.2	48,995.6
2013	104,372.4	116,426.0	104,442.0	116,419.8
TOTAL DES RESERVES	150,278.6	164,004.9	150,066.7	163,900.1

Tableau 46 : Comparaison des intervalles de confiance pour Mack et Bootstrap

Nous constatons des intervalles de confiance presque équivalente pour la méthode de Mack par rapport à la méthode du Bootstrap. En comparant ces intervalles de confiance nous allons retenir les réserves qui ont l'intervalle le moins large pour chaque année de survenance.

La meilleure estimation des réserves est donnée par le tableau suivant :

Année de survenance	borne inf	borne sup	Best estimate
2004			
2005	0.0	0.0	0.0
2006	2.3	2.9	2.6
2007	4.8	9.1	7.0
2008	17.6	32.1	24.8
2009	40.0	77.0	58.5
2010	269.1	433.1	245.2
2011	-14.2	715.9	350.9
2012	42,921.7	48,970.3	45969.4
2013	104,442.0	116,419.8	110431
TOTAL DES RESERVES	147,683.3	166,660.2	157,089.3

Tableau 47 : Meilleure estimation des réserves

L'objectif de cette partie était de déterminer les meilleures estimations des réserves par années de survenance et de la réserve totale. Pour cela nous avons effectué pour chaque année de survenance une intersection des intervalles de confiances données par les différentes méthodes et nous avons considéré les réserves appartenant à ces intervalles. Puisque plusieurs réserves peuvent appartenir à un intervalle de confiance, nous considérons comme meilleure estimation la moyenne des réserves incluses dans l'intervalle de confiance.

De 2003 à 2010 l'intervalle de confiance le plus court intègre le modèle de Mack et celui du Bootstrap et donc nous avons pris en considération la réserve moyenne calculée à partir de ces deux méthodes.

A partir de 2010 on ne retiendra que la méthode du Bootstrap puisque le modèle de Mack aura des intervalles de confiances larges.

CONCLUSION

Durant ce stage qui a duré quatre mois et qui a été pour nous aussi bénéfique que plaisant, nous avons pu élaborer une étude de la rentabilité de la branche Maladie de la compagnie AXA ASSURANCE MAROC.

Dans le secteur des assurances, chaque portefeuille doit être en équilibre et doit garder un certain seuil de rentabilité afin d'être capable d'honorer ses engagements envers ses assurés.

La modélisation et la segmentation de la prime pure est l'un des outils très importants pour contrôler et suivre le comportement de la sinistralité de la population couverte.

En effet, pour élaborer ces modèles, nous avons cherché les fortes corrélations qui pourraient exister entre les variables donnant des informations sur les assurés et les variables utilisés pour calculer la prime pure.

Grâce à ce travail, nous avons pu calculer de la prime d'un assuré, mais aussi comparer sa sinistralité avec la sinistralité moyenne de la classe dont il appartient. Ce qui nous permet en fin de compte de déterminer les « bons » et les « mauvais » assurés.

Aussi, et par l'application des différentes méthodes déterministes et stochastiques nous avons conduit à la détermination d'intervalles de confiance pour les réserves par année de survenance et les réserves totales.

En général, les résultats sont cohérents. Les indicateurs qui nous ont servies pour détecter et évaluer la rentabilité du portefeuille sont la fréquence, le coût moyen ainsi que le Loss Ratio S/P.

Ainsi, il serait préférable et plus bénéfique aussi bien pour les assureurs que pour les assurés, de lancer plus des produits d'assurance maladie individuelle avec des tarifs qui prend en compte l'âge comme variables tarifaire, car elle est la variable qui explique plus la sinistralité du risque maladie.

BIBLIOGRAPHIE

LIVRES

- **CHRISTIAN GOURIEROUX**, Statistiques de l'assurance, Economica, 1999.
- **CHRISTIAN PARTRAT ET JEAN-LUC BESSON**, Assurance Non vie : Modélisation, Simulation, Edition Economica, 2004.
- **Fernandes L.B.** , Extreme value theory and value at risk, Revista de Analisis Economico Vol 18 pp 57-85, 2003.
- **MICHEL DENUIT ET ARTHUR CHARPENTIER**, Mathématiques de l'assurance non-vie Tome 1 et 2, Economica, 2004.
- **McCullagh P. & Nelder J.A.** Generalized linear models. Chapman & Hall/CRC, 2 ed, 1989.

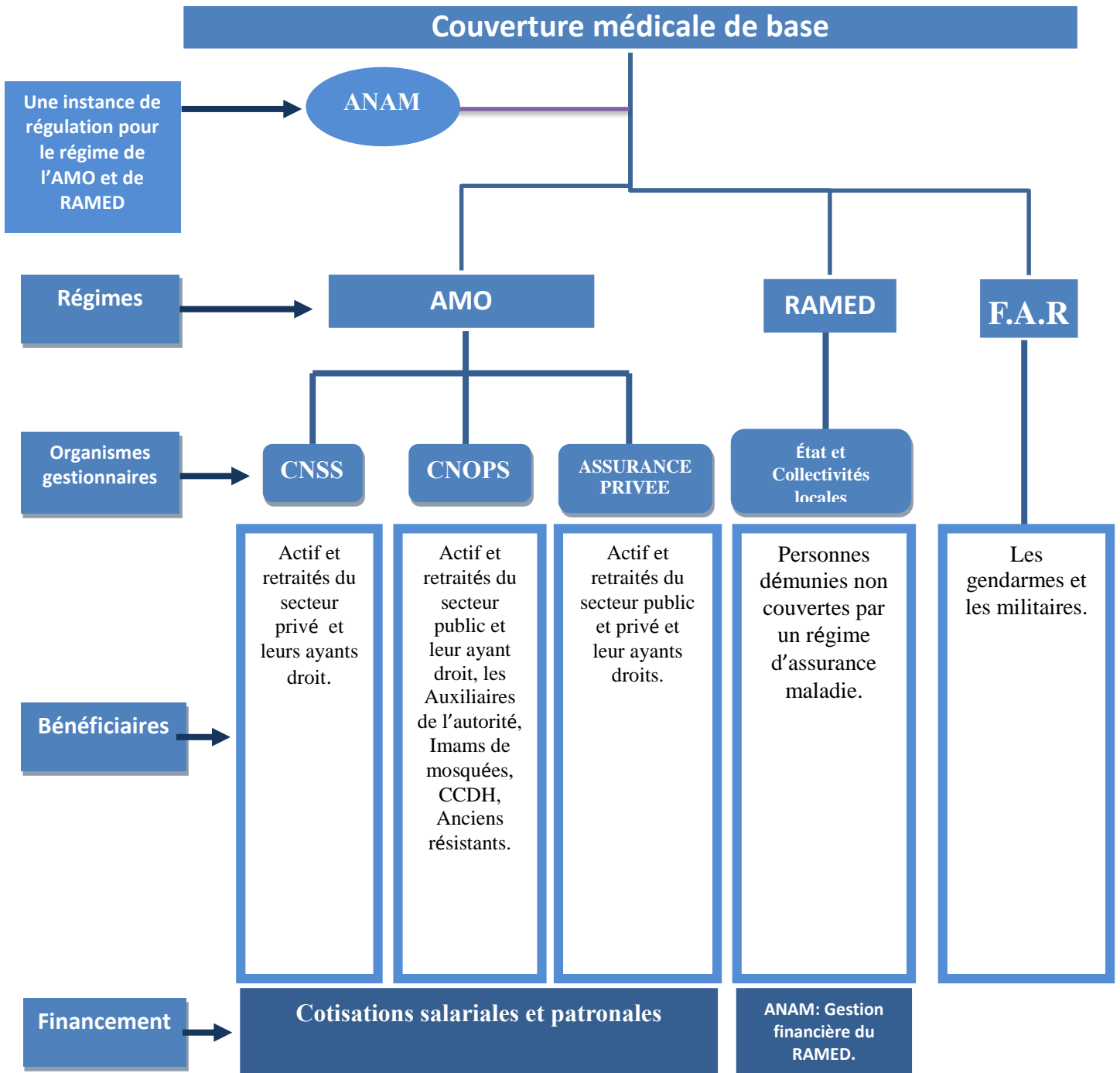
PFEs et Mémoires

- **OLIVIER Jean-Baptiste**, Tarification santé en Allemagne pour un produit soins complets, PFE, **UNIVERSITÉ PARIS IX DAUPHINE**, Novembre 2003.
- **P.D. England & R.J. Verall**. Stochastic claims reserving in general insurance. *Presented to the Institute of Actuaries, 28 Janvier 2002.*
- **Thomas CHAPUIS**, Provisionnement en assurance construction MLG et mesure de l'aléa, PFE, **CENTRE D'ÉTUDES ACTUARIELLES**, 2003.

ARTICLES

- **Alexander J., McNeil, Saladin T.** *The peaks over thresholds for estimating high quantiles of loss distribution*, International ASTIN Colloquium, pp 70-94. 1997.
- **Dipankar Bandyopadhyay, Ph.D**, Lecture 13: GLM for Poisson Data, University of South Carolina, 2011.
- **Duncan Anderson & Sholom Feldblum & Claudine Modlin & Doris Schrimacher & Ernesto Schirmacher & Neeza Thandi**, Practitioner's Guide to Generalized Linear Models, Mai 2004.
- **Efron B.**, **Bootstrap** methods: another look at the Jackknife, The Annals of Statistics, 7, 1-26, 1997.
- **NICHOLAS J. HORTON AND STUART R. LIPSITZ**, Review of Software to Fit Generalized Estimating Equation Regression Models, The American Statistician, Vol 23, May 1999.

Annexe 1 : Schéma de la couverture médicale de base au Maroc



Annexe 2 : Modélisation des sinistres graves pour la garantie «Consultation»

D'après les graphes ci-dessous, on conclue que les modèles de la fréquence (Binomiale négative) et du coût moyen (GPD) sont significatives et s'ajuste avec nos modèles des sinistres graves de la garantie « Consultation».

Fit Statistics for Gpd Distribution	
-2 Log Likelihood	473.78301
Akaike's Information Criterion	477.78301
Corrected Akaike's Information Criterion	478.41459
Schwarz's Bayesian Information Criterion	479.98510
Kolmogorov-Smirnov Statistic	1.50482
Anderson-Darling Statistic	1.98574
Cramer-von Mises Statistic	0.35265

Parameter Estimates for Gpd Distribution				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Theta			4.61	0.0002
Xi		.	.	.

Tableau : Estimation des paramètres du modèle GPD de la charge des sinistres graves « Consultation »

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1			-824.37	<.0001
_Alpha	1			47.87	<.0001

Tableau : Estimation du modèle Binomiale négative de la fréquence des sinistres graves « Consultation »

Annexe 3 : Modélisation des sinistres graves pour la garantie «Autres»

D'après les graphes ci-dessous, on conclue que les modèles de la fréquence (Binomiale négative) et du coût moyen (GPD) sont significatives et s'ajuste avec nos modèles des sinistres graves de la garantie «Autres».

Fit Statistics for Gpd Distribution	
-2 Log Likelihood	473.78301
Akaike's Information Criterion	477.78301
Corrected Akaike's Information Criterion	478.41459
Schwarz's Bayesian Information Criterion	479.96510
Kolmogorov-Smirnov Statistic	1.50482
Anderson-Darling Statistic	1.98574
Cramer-von Mises Statistic	0.35265

Parameter Estimates for Gpd Distribution				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Theta			4.61	0.0002
Xi		.	.	.

Tableau : Estimation des paramètres du modèle GPD de la charge des sinistres graves « Autres »

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1			-624.37	<.0001
_Alpha	1			47.87	<.0001

Tableau : Estimation du modèle Binomiale négative de la fréquence des sinistres graves « Autres»

Annexe 4 : Estimation des paramètres du modèle ZINB de la fréquence de la garantie « Médicaments »

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.0229		1052.14	<.0001
class_age	1	1		0.0099		6434.52	<.0001
class_age	2	1		0.0087		11573.3	<.0001
class_age	3	1		0.0088		5044.95	<.0001
class_age	4	1		0.0091		1216.99	<.0001
class_age	5	0		0.0000		-	-
sexe	1	1		0.0040		4665.60	<.0001
sexe	2	0		0.0000		-	-
lien	1	1		0.0073		548.27	<.0001
lien	2	1		0.0126		127.02	<.0001
lien	3	0		0.0000		-	-
sit_mat	1	1		0.0161		73.68	<.0001
sit_mat	2	1		0.0167		758.95	<.0001
sit_mat	3	0		0.0000		-	-
taux_remb	1	1		0.0094		2741.85	<.0001
taux_remb	2	1		0.0116		9326.18	<.0001
taux_remb	3	1		0.0096		1540.67	<.0001
taux_remb	4	1		0.0098		1404.52	<.0001
taux_remb	5	0		0.0000		-	-
cat_socio_prof	1	1		0.0099		65.08	<.0001
cat_socio_prof	2	0		0.0000		-	-
class_profess	1	1		0.0082		117.85	<.0001
class_profess	2	0		0.0000		-	-
class_profess	3	1		0.0096		349.03	<.0001
class_profess	4	1		0.0108		457.05	<.0001
class_profess	5	0		0.0000		-	-
Dispersion		1	7.6413	0.0220	7.5983 7.6845		

Tableau 1 : Estimation des paramètres du modèle (ZINB)

Annexe 4 : (suite)

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.1774		3829.82	<.0001
class_age	1	1		0.0200		2850.71	<.0001
class_age	2	1		0.0185		1137.39	<.0001
class_age	3	1		0.0173		483.72	<.0001
class_age	4	1		0.0168		82.22	<.0001
class_age	5	0		0.0000		.	.
sexe	1	1		0.0152		19274.1	<.0001
sexe	2	0		0.0000		.	.
lien	1	1		0.0373		7658.12	<.0001
lien	2	1		0.0453		7917.08	<.0001
lien	3	0		0.0000		.	.
sit_mat	1	1		0.0439		445.37	<.0001
sit_mat	2	1		0.0441		988.79	<.0001
sit_mat	3	0		0.0000		.	.
taux_remb	1	1		0.0440		384.19	<.0001
taux_remb	2	1		0.0575		23.16	<.0001
taux_remb	3	1		0.0438		1176.42	<.0001
taux_remb	4	1		0.0453		234.11	<.0001
taux_remb	5	0		0.0000		.	.
cat_socio_prof	1	1		0.1548		505.71	<.0001
cat_socio_prof	2	0		0.0000		.	.
class_profess	1	1		0.0195		12959.1	<.0001
class_profess	2	0		0.0000		.	.
class_profess	3	1		0.0172		10592.0	<.0001
class_profess	4	1		0.0203		12225.6	<.0001
class_profess	5	0		0.0000		.	.

Tableau 2 : Estimation des paramètres du modèle des zéros (ZINB)

Annexe 5 : Estimation des paramètres du modèle Gamma du coût moyen de la garantie « médicaments »

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1		0.0109		337142	<.0001
class_age_sin	1		0.0067		36598.3	<.0001
class_age_sin	2		0.0059		23871.6	<.0001
class_age_sin	3		0.0061		11572.1	<.0001
class_age_sin	4		0.0070		2632.74	<.0001
class_age_sin	5	0	0.0000		.	.
sexe	1		0.0019		52.00	<.0001
sexe	2	0	0.0000		.	.
lien	1		0.0036		175.67	<.0001
lien	2		0.0065		0.27	0.6010
lien	3	0	0.0000		.	.
sit_mat	1		0.0076		34.48	<.0001
sit_mat	2		0.0078		40.65	<.0001
sit_mat	3	0	0.0000		.	.
taux_sin	1		0.0048		130757	<.0001
taux_sin	2		0.0031		10911.9	<.0001
taux_sin	3		0.0020		5258.66	<.0001
taux_sin	4	0	0.0000		.	.
cat_socio_prof	1		0.0050		38.68	<.0001
cat_socio_prof	2	0	0.0000		.	.
class_profess	1		0.0036		70.87	<.0001
class_profess	2	0	0.0000		.	.
class_profess	3		0.0046		1.03	0.3104
class_profess	4		0.0047		377.78	<.0001
class_profess	5	0	0.0000		.	.
Scale	1		0.0026			

Tableau 1 : Estimation des paramètres du modèle Gamma

Annexe 5 : (suite)

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1		0.0102		393034	<.0001
class_age_sin	1	1	0.0063		45215.1	<.0001
class_age_sin	2	1	0.0058		25434.0	<.0001
class_age_sin	3	1	0.0061		12068.4	<.0001
class_age_sin	4	1	0.0070		2700.58	<.0001
class_age_sin	5	0	0.0000		.	.
sexe	1	1	0.0017		318.56	<.0001
sexe	2	0	0.0000		.	.
sit_mat	1	1	0.0075		84.67	<.0001
sit_mat	2	1	0.0077		92.94	<.0001
sit_mat	3	0	0.0000		.	.
taux_sin	1	1	0.0044		159550	<.0001
taux_sin	2	1	0.0030		11428.4	<.0001
taux_sin	3	1	0.0020		5130.70	<.0001
taux_sin	4	0	0.0000		.	.
cat_socio_prof	1	1	0.0050		58.73	<.0001
cat_socio_prof	2	0	0.0000		.	.
Scale	1		0.0028			

Tableau 2 : Estimation des paramètres du modèle Gamma pour les variables restantes

Annexe 6 : Modélisation des sinistres ordinaires pour la garantie

« Consultation »

D'après les graphes ci-dessous, on conclue que les modèles de la fréquence (ZINB) et du coût moyen (Gamma) sont significatives et s'ajuste avec nos modèles de la garantie « Consultation ».

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.0464		8.47	0.0036
class_age	1	1		0.0190		709.58	<.0001
class_age	2	1		0.0156		2789.82	<.0001
class_age	3	1		0.0151		1635.45	<.0001
class_age	4	1		0.0153		192.27	<.0001
class_age	5	0		0.0000		.	.
sexe	1	1		0.0087		847.00	<.0001
sexe	2	0		0.0000		.	.
lien	1	1		0.0174		123.76	<.0001
lien	2	1		0.0246		276.94	<.0001
lien	3	0		0.0000		.	.
cat_socio_prof	1	1		0.0265		19.36	<.0001
cat_socio_prof	2	0		0.0000		.	.
class_profess	1	1		0.0159		69.27	<.0001
class_profess	2	0		0.0000		.	.
class_profess	3	1		0.0143		1654.36	<.0001
class_profess	4	1		0.0317		12.69	0.0004
class_profess	5	0		0.0000		.	.
Dispersion		1	5.6048	0.0322	5.5421 5.6682		

Tableau : Estimation des paramètres du modèle de la fréquence (ZINB) de la garantie « Consultation »

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.5990		486.88	<.0001
class_age	1	1		0.0587		2606.02	<.0001
class_age	2	1		0.0525		695.40	<.0001
class_age	3	1		0.0545		353.97	<.0001
class_age	4	1		0.0551		71.21	<.0001
class_age	5	0		0.0000		.	.
sexe	1	1		0.0356		3909.99	<.0001
sexe	2	0		0.0000		.	.
lien	1	1		0.0635		3115.27	<.0001
lien	2	1		0.0635		3929.73	<.0001
lien	3	0		0.0000		.	.
sit_mat	1	1		0.5573		5.66	0.0173
sit_mat	2	1		0.5579		72.25	<.0001
sit_mat	3	0		0.0000		.	.
cat_socio_prof	1	1		0.1473		84.72	<.0001
cat_socio_prof	2	0		0.0000		.	.
class_profess	1	1		0.0759		31.36	<.0001
class_profess	2	0		0.0000		.	.
class_profess	3	1		0.0445		2488.82	<.0001
class_profess	4	1		0.0613		3859.69	<.0001
class_profess	5	0		0.0000		.	.

Tableau : Estimation des paramètres du modèle des zéros de la fréquence (ZINB) de la garantie « Consultation »

Annexe 6 : (suite)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1		0.0272			54843.2	<.0001
class_age_sin	1	1		0.0130			8887.49	<.0001
class_age_sin	2	1		0.0108			6130.57	<.0001
class_age_sin	3	1		0.0116			2548.19	<.0001
class_age_sin	4	1		0.0145			342.52	<.0001
class_age_sin	5	0		0.0000			.	.
sexe	1	1		0.0050			101.81	<.0001
sexe	2	0		0.0000			.	.
sit_mat	1	1		0.0158			120.08	<.0001
sit_mat	2	1		0.0166			173.13	<.0001
sit_mat	3	0		0.0000			.	.
taux_sin	1	1		0.0086			32781.6	<.0001
taux_sin	2	1		0.0163			275.03	<.0001
taux_sin	3	1		0.0082			48.00	<.0001
taux_sin	4	0		0.0000			.	.
cat_socio_prof	1	1		0.0205			48.83	<.0001
cat_socio_prof	2	0		0.0000			.	.
Scale		1	1.6561	0.0068	1.6428	1.6695		

**Tableau : Estimation des paramètres du modèle Gamma du coût moyen de la garantie
« Consultation »**

Annexe 7 : Modélisation des sinistres ordinaires pour la garantie «Autres»

D'après les graphes ci-dessous, on conclue que les modèles de la fréquence (ZINB) et du coût moyen (Gamma) sont significatives et s'ajuste avec nos modèles de la garantie « Autre ».

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.1251		63.58	<.0001
class_age	1	1		0.0480		92.34	<.0001
class_age	2	1		0.0450		82.78	<.0001
class_age	3	1		0.0460		37.76	<.0001
class_age	4	1		0.0488		12.58	0.0004
class_age	5	0		0.0000		.	.
sexe	1	1		0.0142		806.27	<.0001
sexe	2	0		0.0000		.	.
sit_mat	1	1		0.0607		14.55	0.0001
sit_mat	2	1		0.0612		44.88	<.0001
sit_mat	3	0		0.0000		.	.
cat_socio_prof	1	1		0.0306		46.59	<.0001
cat_socio_prof	2	0		0.0000		.	.
Dispersion		1	1.9629	0.0596	1.8494	2.0832	

Tableau : Estimation des paramètres du modèle de la fréquence (ZINB) de la garantie « Autres »

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.3051		77.79	<.0001
sexe	1	1		0.0153		1137.99	<.0001
sexe	2	0		0.0000		.	.
sit_mat	1	1		0.0669		123.52	<.0001
sit_mat	2	1		0.0664		5.93	0.0149
sit_mat	3	0		0.0000		.	.
taux_remb	1	1		0.2782		47.42	<.0001
taux_remb	2	1		0.2827		44.46	<.0001
taux_remb	3	1		0.2788		67.05	<.0001
taux_remb	4	1		0.2787		40.04	<.0001
taux_remb	5	0		0.0000		.	.
cat_socio_prof	1	1		0.0544		363.15	<.0001
cat_socio_prof	2	0		0.0000		.	.
class_profess	1	1		0.0291		188.71	<.0001
class_profess	2	0		0.0000		.	.
class_profess	3	1		0.1310		46.10	<.0001
class_profess	4	1		0.1569		98.60	<.0001
class_profess	5	0		0.0000		.	.

Tableau : Estimation des paramètres du modèle des zéros de la fréquence (ZINB) de la garantie « Autres »

Annexe 7 : (suite)

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1		0.0272		54843.2	<.0001
class_age_sin	1	1		0.0130		8887.49	<.0001
class_age_sin	2	1		0.0108		6130.57	<.0001
class_age_sin	3	1		0.0116		2548.19	<.0001
class_age_sin	4	1		0.0145		342.52	<.0001
class_age_sin	5	0		0.0000		.	.
sexe	1	1		0.0050		101.81	<.0001
sexe	2	0		0.0000		.	.
sit_mat	1	1		0.0158		120.08	<.0001
sit_mat	2	1		0.0166		173.13	<.0001
sit_mat	3	0		0.0000		.	.
taux_sin	1	1		0.0086		32781.6	<.0001
taux_sin	2	1		0.0163		275.03	<.0001
taux_sin	3	1		0.0082		48.00	<.0001
taux_sin	4	0		0.0000		.	.
cat_socio_prof	1	1		0.0205		48.83	<.0001
cat_socio_prof	2	0		0.0000		.	.
Scale		1	1.6561	0.0068	1.6428	1.6695	

*Tableau : Estimation des paramètres du modèle Gamma du coût moyen de la garantie
« Autres »*

Annexe 8 : Triangles des règlements cumulés

Année de survenance	Développement									
	1	2	3	4	5	6	7	8	9	10
2004	23,758	49,114	68,582	69,000	69,120	69,152	69,164	69,169	69,171	69,171
2005	31,245	63,741	90,775	91,325	91,378	91,406	91,412	91,415	91,418	
2006	26,312	57,779	82,451	82,500	82,562	82,573	82,578	82,581		
2007	30,470	65,482	90,973	90,975	90,983	91,005	91,017			
2008	49,756	101,587	136,854	136,897	136,919	136,937				
2009	50,420	102,735	140,056	140,102	140,123					
2010	56,762	112,223	157,459	157,479						
2011	53,726	106,526	149,632							
2012	55,326	116,598								
2013	59,268									

Tableau 1 : Triangle des incréments de paiements cumulé pour la maladie

Année de survenance	Développement									
	1	2	3	4	5	6	7	8	9	10
2004	23,758	49,114	68,582	69,000	69,120	69,152	69,164	69,169	69,171	69,171
2005	31,245	63,741	90,775	91,325	91,378	91,406	91,412	91,415	91,418	91,418
2006	26,312	57,779	82,451	82,500	82,562	82,573	82,578	82,581	82,584	82,584
2007	30,470	65,482	90,973	90,975	90,983	91,005	91,017	91,021	91,024	91,024
2008	49,765	101,587	136,854	136,897	136,919	136,937	136,951	136,958	136,962	136,962
2009	50,420	102,735	140,056	140,102	140,123	140,156	140,171	140,177	140,181	140,181
2010	56,762	112,223	157,459	157,479	157,553	157,590	157,606	157,614	157,618	157,830
2011	53,726	106,526	149,632	149,852	149,922	149,958	149,973	149,980	149,985	149,985
2012	55,326	116,598	162,162	162,400	162,476	162,515	162,532	162,539	162,544	162,544
2013	59,268	121,708	169,268	169,517	169,597	169,636	169,654	169,662	169,660	169,667

Tableau 2 : les règlements cumulés estimés par Chain Ladder

Année de survenance	Développement									
	0	1	2	3	4	5	6	7	8	9
2004	23758	49114	68582	69000	69120	69152	69164	69169	69171	69171
2005	31245	63741	90775	91325	91378	91406	91412	91415	91418	91418
2006	26312	57779	82451	82500	82562	82573	82578	82581	82584	82584
2007	30470	65482	90973	90975	90983	91005	91017	91020	91023	91023
2008	49765	101587	136854	136897	136919	136937	136939	136937	136943	136943
2009	50420	102735	140056	140102	140123	140140	140141	140140	140145	140145
2010	56762	112223	157459	157479	157473	157488	157487	157484	157490	157490
2011	53726	106526	149632	149650	149652	149667	149668	149665	149671	149671
2012	55326	116598	160941	160918	160909	160923	160922	160919	160925	160925
2013	59268	119690	165130	165093	165079	165093	165092	165088	165094	165094

Tableau 3 : les règlements cumulés estimés par London Chain

**Annexe 9 : Vérification de l'hypothèse H2 pour l'application de la méthode
Chaine Ladder**

D'après les graphes ci-dessous, on conclue que l'hypothèse 2 n'est pas vérifiée surtout pour la première et la deuxième année, donc on aura besoin d'effectuer quelques corrections.

