

Mémoire présenté le :

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : GOUBER Aïcha

Titre MISE EN PLACE D'UNE VISION PROSPECTIVE DES PRESTATIONS SANTE

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

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

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

signature

Entreprise :

Nom : KLESIA

Signature :

Directeur de mémoire en entreprise :

Nom : Sobidaa Varatharajan

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

<p>GIE Klesia ADP 4 rue Georges Picquart 75017 Paris - France <i>Groupement d'intérêt économique</i> Siret 752 610 147 00239 - RCS Paris N° TVA intracommunautaire - FR 59 752610147</p>

Signature du candidat

Résumé

Dans le cadre du plan prévisionnel d'activité, réalisé chaque année par Klesia, des ratios S/P correspondant à la charge de sinistralité rapportée aux chiffres d'affaires sont estimés. La méthode utilisée actuellement consiste à appliquer une dérive au ratio observé sur la survenance antérieure. Cette méthode a l'avantage d'être facile d'application, cependant elle ne repose que sur des hypothèses potentielles fournies par le comité des risques et solvabilité (CRS).

Ce mémoire a pour objectif la modélisation d'un outil de projection permettant d'estimer la charge de sinistralité santé d'un portefeuille sur une survenance complète. Étant donné que la santé représente un risque à court terme, l'estimation sera basée sur deux années.

Deux modèles seront utilisés à cette fin : les modèles de type SARIMA et le lissage exponentiel de Holt-Winters. Ces deux modèles font partie des méthodes de prévision des séries chronologiques. Le lissage exponentiel de Holt-Winters accorde un poids plus important aux données les plus récentes et se révèle très efficace lorsque la tendance et la saisonnalité sont stables. Le modèle SARIMA, quant à lui, offre une plus grande flexibilité.

Pour ces méthodes, la qualité des données est cruciale. C'est pourquoi, des retraitements seront effectués afin d'atténuer les effets de la Covid, entraînant des perturbations dans les remboursements.

L'application de ces méthodes a permis d'obtenir une estimation satisfaisante de la charge de sinistralité pour l'entité Klesia Prévoyance, membre du groupe Klesia, ainsi que pour plusieurs portefeuilles au sein de l'entité.

Pour approfondir notre analyse, nous avons souhaité examiner les impacts des changements de ratio S/P sur les résultats en Solvabilité 2. Nous avons ainsi constaté que certains éléments S2 subissent naturellement des évolutions, tels que les provisions techniques, mais également les BE de sinistres et les fonds propres.

Mots Clés : SARIMA, Série temporelle, Solvabilité 2, Lissage Exponentiel, Prédiction, Covid, Santé, Ratio



Abstract

As part of the annual business plan carried out by Klesia, S/P ratios corresponding to loss ratios reported to revenue figures are estimated. The current method used involves applying a trend to the ratio observed in the previous occurrence. This method has the advantage of being easy to apply. However, it is based solely on potential assumptions provided by the Risk and Solvency Committee (RSC).

The objective of this thesis is to model a projection tool to estimate the health loss ratio of a portfolio over a complete occurrence. As healthcare risk is short-term, the estimation will be based on two years.

Two models will be discussed further: SARIMA-type models and Holt-Winters exponential smoothing. These two models are part of time series forecasting methods. The latter allows for greater weight to be given to the most recent data, but it is very effective only when trend and seasonality are stable. The SARIMA model, on the other hand, is more flexible.

For these methods, data quality is very important. That's why reprocessing will be carried out to eliminate Covid-related effects that cause disruptions in reimbursements.

These models allowed for a good estimation of the loss ratio on the “Klesia Prevoyance” entity, a member of the Klesia group, as well as on several portfolios of the entity.

To go further, we wanted to analyze the impacts of changes in the S/P ratio on Solvency 2 results. We were thus able to observe that certain Solvency 2 elements naturally evolves, such as technical provisions, as well as claims BE and equity.

Keywords: SARIMA, Time Series, Solvency 2, Exponential Smoothing, Prediction, Covid, Health, Ratio



Remerciements

Je tiens à remercier en premier lieu ma tutrice de stage, Sobidaa VARATHARAJAN, sans qui je n'aurais pu aboutir à ce mémoire. Elle a su guider mes recherches, répondre avec patience à toutes mes questions et m'apporter tout son soutien.

Toute ma gratitude va à Luzia Hubinois, Directrice de l'actuariat, pour ses conseils avisés.

Je souhaite témoigner ma sympathie et ma gratitude à toute l'équipe Actuariat de Klesia pour leur soutien durant la rédaction de ce mémoire.

J'adresse également mes remerciements à mon tuteur académique Pierre RIBERAU.

Enfin, j'exprime toute ma reconnaissance à ma famille et mes amis. Je les remercie de m'avoir soutenu sans faillir dans toutes les étapes qui m'ont permis de réussir jusqu'à présent.

Sommaire

Résumé	1
Abstract	2
Remerciements	3
Sommaire	4
Introduction	6
Partie 1 : Contexte	8
1. La santé en France	8
1.1. Présentation de la sécurité sociale	8
1.2. Remboursement des frais de santé par la sécurité sociale	9
1.3. Complémentaire santé	10
1.3.1. Réforme 100% santé.....	10
1.3.2. Contrat responsable	11
2. La santé au sein de Klesia Prévoyance	12
2.1. Présentation de Klesia.....	12
2.2. Présentation de la base	12
2.3. Statistiques descriptives et analyses	14
2.4. Présentation des indicateurs de prestations	17
2.5. Limite de l'outil	19
3. Modèles utilisés dans le domaine de l'actuariat	19
3.1. Les méthodes exogènes	20
3.2. Adéquation avec nos enjeux.....	21
Partie 2 : Théorie sur la projection des prestations santé	23
1. Les Séries Temporelles	23
1.1. Généralités	23
1.2. Processus stationnaire : modèles ARMA.....	24
1.3. Processus non stationnaires	25
1.3.1. Modèles ARIMA et SARIMA	25
1.3.2. Etapes de sélection du modèle	26
1.3.3. Validation du modèle	28
2. Le Lissage Exponentiel	31
2.1. Lissage exponentiel simple.....	31
2.2. Lissage exponentiel Double.....	31
2.3. Lissage exponentiel Holt-Winters.....	32
Partie 3 : Mise en œuvre et application des modèles	34
1. Modèles SARIMA	34
1.1. Type de modèle : additive ou multiplicative.....	34
1.2. Sélection du modèle	37
1.3. Mise en application.....	41
1.4. Retraitement Covid.....	44



1.4.1. Impact du covid sur les institutions de prévoyance	44
1.4.2. Retraitement avec utilisation de l'historique	45
1.4.3. Retraitement des résidus	47
1.4.4. Mise en application	48
2. Lissage Holt-Winters.....	50
2.1. Mise en application	50
3. Comparaison avec le Plan Prévisionnel d'Activité.....	52
4. Application du modèle aux branches de Klesia Prévoyance	54
4.1. Etude des portefeuilles.....	54
4.2. Résultat de la modélisation	57
Partie 4 : Impact des prévisions sur Solvabilité 2	61
1. Présentation Solvabilité 2.....	61
4.3. Pilier 1 : Exigence Quantitative	61
4.4. Pilier 2 : Exigence Qualitative & Pilier 3 : Communication	63
2. Présentation du modèle.....	64
2.1. Modélisation de la charge de sinistralité.....	64
3. Impact sur le résultat Solvabilité 2	64
3.1. Impact sur les BE et le bilan S2	65
3.2. Impact sur le SCR.....	67
Conclusion	71
Bibliographie	72
Table des Figures.....	73
Annexes.....	75

Introduction

Chez un organisme assureur, il est nécessaire d'effectuer un Plan Prévisionnel d'Activité chaque année afin d'estimer son chiffre d'affaires ainsi que sa charge de sinistralité.

Le chiffre d'affaires peut être projeté sur un horizon de cinq années en appliquant des taux de résiliation, d'indexation et d'affaires nouvelles à partir du chiffre d'affaires de l'année courante.

Quant à la charge de sinistralité, actuellement, un taux de dérive est appliqué aux ratio S/P de l'année courante. Il s'agit du principe de base. Ensuite, une phase pilotage permettant d'ajuster ces indicateurs peut avoir lieu.

La question se pose désormais : peut-on développer un nouvel outil permettant d'allier modèle mathématique, taux de dérive et pilotage ?

Ce mémoire s'articule donc autour du développement d'un modèle SARIMA et d'un modèle de lissage Holt-Winters afin de projeter les prestations risque santé sur deux années et d'estimer la charge de sinistralité pour la survenance de 2021. A partir du chiffre d'affaires projeté lors du Plan Prévisionnel d'Activité (PPA), il sera donc possible d'établir un ratio S/P.

Dans un premier temps, les modèles de prévision de type ARIMA seront abordés. Il s'agit des modèles les plus utilisés dans le cadre des projections. Ces types de prévision ont l'avantage d'être compréhensibles et applicables grâce à la méthode Box-Jenkins. Cette technique permet de calibrer un modèle ARIMA ou SARIMA en cinq étapes : suppression de la tendance et de la saisonnalité, estimation du modèle, puis validation de ce dernier et enfin sélection du meilleur modèle à travers les critères AIC et BIC.

Il existe également des méthodes déterministes telles que les méthodes de lissage exponentiel. Cette technique se divise en trois méthodes, allant des plus classiques au plus complexes : lissage exponentiel simple, double et de Holt-Winters. Les deux premières peuvent être automatiquement rejetées en raison de leurs hypothèses restrictives.

Ces deux méthodes seront bien évidemment implémentées afin de comparer leur pouvoir prédictif ainsi que leur adéquation à nos données.

Les méthodes de prévision dépendent nécessairement des données en entrée. Dans les organismes assureurs, ces données peuvent souvent être de moindre qualité. Dans une entreprise telle que Klesia, qui a connu une fusion en 2012 et qui connaît depuis de nombreux changements tels que le regroupement et la création de nouvelles entités, il n'est pas rare que les cadences de règlement fluctuent d'une année à l'autre. C'est ce qui est observable dans nos données avant 2017.

A cela s'ajoutent des événements extérieurs tels que la Covid, qui a nécessité un retraitement de cette période. Deux méthodes ont ainsi été appliquées : l'utilisation de l'historique et le retraitement des résidus.

Les résidus, définis par des événements aléatoires ne pouvant être anticipés, rendent cohérente l'hypothèse d'annuler l'effet Covid en retraçant les données des résidus sur une période donnée. La deuxième méthode, plus classique, permet de transposer, à partir de l'historique, les cadences de règlement des prestations santé.

Dans une première partie, nous présentons la santé en France et chez Klesia, accompagnée d'une analyse sur la base de données Klesia Prévoyance.



Dans une seconde partie, la théorie sur les deux modèles utilisés sera exposée. Il s'agira des modèles SARIMA ainsi que des modèles de lissage exponentiel. Ces modèles seront ensuite appliqués à la base de données précédemment présentée afin d'estimer la sinistralité pour l'année 2021. Cette projection sera réalisée à la fois sur l'entité Klesia Prévoyance et sur certains de ses portefeuilles. De plus, des méthodes de retraitement de la période Covid seront abordées et mises en œuvre.

Une étude d'impact S2 viendra conclure ce mémoire, ayant pour objectif de déterminer de quelle manière le changement d'un ratio S/P affecte les éléments S2.

Partie 1 : Contexte

1. La santé en France

1.1. Présentation de la sécurité sociale

En 1945, la sécurité sociale est mise en place dans le but d'offrir une protection contre les conséquences d'événements imprévus, souvent qualifiés de risques sociaux.

Il s'agit encore aujourd'hui d'un fondement de notre système français, classé en quatre branches :

- Maladie (maladie, maternité, invalidité, décès)
- Accident du travail et maladie professionnelle
- Retraite
- Famille

En 2018, la sécurité sociale versait 742 milliards de prestations, avec 81% des versements répartis entre les branches Retraite et Santé.

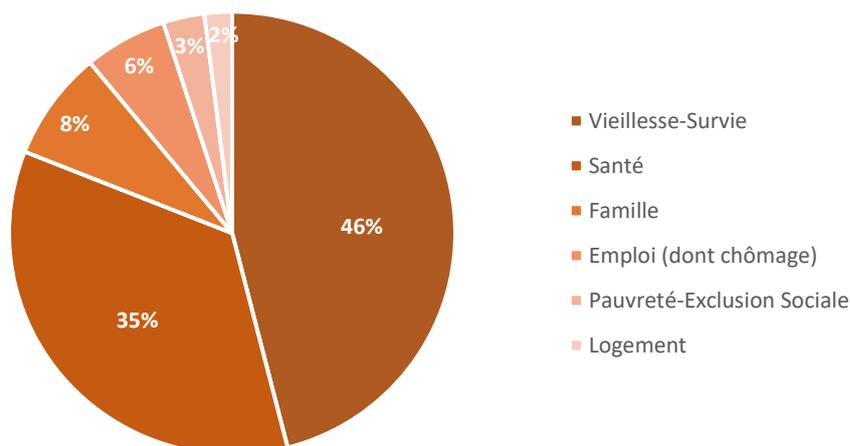


Figure 1 : Répartition des remboursements de la sécurité sociale en 2018

Les remboursements des frais de santé, mais également de prévoyance, sont gérés par la branche maladie. Celle-ci se charge d'indemniser l'ensemble des citoyens contre les aléas de la vie et favorise de cette manière l'accès aux soins pour tous grâce à son action sociale.

L'offre du groupe Klesia s'articule autour des risques de santé et prévoyance. Les remboursements viennent donc compléter ceux versés dans un premier temps par la sécurité sociale via la branche Maladie.

Une explication des remboursements de la sécurité sociale est alors nécessaire avant d'aborder les remboursements complémentaires.

1.2. Remboursement des frais de santé par la sécurité sociale

Dans le domaine de la santé, la Sécurité Sociale met en place un tarif servant de base pour le remboursement des soins de santé, il s'agit de la base de remboursement (BR). Un taux de remboursement (TR) est par la suite appliqué à ce dernier afin d'obtenir le montant remboursé par la sécurité sociale. Il varie en fonction des garanties. Voici un tableau illustrant les taux de remboursements :

Garanties	Général	Régime Alsace – Morelle
Frais d'hospitalisation	80%	100%
Médecins	70%	90%
Optique, prothèses auditives	60%	90%
Frais de transport	65%	100%
Cure thermale avec hospitalisation	80%	100%

Figure 2 : Illustration des taux de remboursement

A partir de ces deux éléments, on peut déduire les frais remboursés par la SS, le ticket modérateur ainsi que le dépassement. Le schéma suivant permet d'illustrer le remboursement de la sécurité sociale :

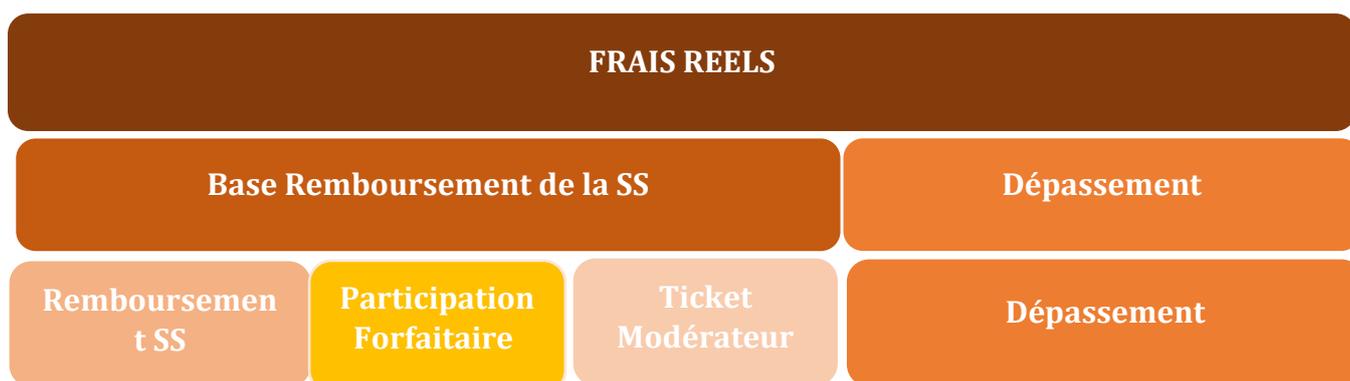


Figure 3 : Schéma de remboursement de la sécurité sociale

Sur ce schéma, la participation forfaitaire correspond au reste à charge de 1 euro obligatoire depuis 2005 pour le régime général. Il s'applique sur les consultations médicales, les examens de radiologie ainsi que les actes de biologie médicale. Un plafond est tout de même appliqué avec un maximum de 50 euros par an et un de 4 euros par jour pour les actes réalisés par le même praticien.

Pour certains postes, la base de remboursement n'est pas assez suffisante pour couvrir l'ensemble des frais c'est notamment le cas pour les équipements optiques, dentaires et auditifs.

Dans ce contexte, il est primordial pour un individu d'avoir une complémentaire santé qui permet de réduire les coûts réels de santé. Notre étude porte sur les contrats de groupe (assurance collectives) et notamment la santé.

En France, il existe deux types d'assurances : individuelles ou collectives. La principale différence réside dans le fait que la première est souscrite de la propre initiative de l'assuré tandis que la seconde est souscrite par une personne morale généralement le chef d'entreprise.

L'article L141-1 du code de l'assurance définit l'assurance collective comme :

« un contrat d'assurance de groupe souscrit par une personne morale en vue de l'adhésion d'un ensemble de personnes répondant à des conditions définies au contrat, pour la couverture des risques dépendant de la durée de la vie humaine, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité, des risques d'incapacité de travail ou d'invalidité ou du risque de chômage. [...] »

L'assurance collective octroie des avantages aux salariés puisque l'employeur est contraint de cotiser une partie des cotisations de ces salariés à hauteur de 50% minimum.

La loi du 14 juin 2013 oblige depuis janvier 2016 les employeurs à proposer une complémentaire en santé et en prévoyance. D'autres lois et décrets viennent règlementer l'application de celle-ci.

Certaines auront un impact sur notre modèle et sont détaillées ci-dessous.

1.3. Complémentaire santé

1.3.1. Réforme 100% santé

L'offre 100% santé entre en application à partir de 2021 afin de permettre l'accès aux soins à tous et plus facilement. Cette réforme concerne trois postes clés : optique, dentaire et audiologie. Avant 2021, certains de ces soins nécessitaient une prise en charge plus ou moins importante de la part de l'assuré. Aujourd'hui, sur ces trois postes, chaque assuré peut bénéficier s'il le souhaite d'un reste à charge nul s'il bénéficie d'une complémentaire responsable.



Figure 4 : Schéma de remboursement 100% Santé



Cette réforme a entraîné une implication de la part de tous les acteurs de la santé : l'assurance maladie (hausse de la base de remboursement), les complémentaires (hausse des remboursements) et les professionnels (baisse des tarifs).

Le 100% santé concerne des paniers de soins spécifiques :

- Optique :
 - Montures de classe A : minimum 17 modèles adultes et 10 pour enfant
 - Verres de classe A : qualité minimum
 - Tarifs limités (monture : 30 euros)
- Dentaire :
 - BR revalorisées
 - Tarifs limités
 - Panier 100% Santé concerne environ 46% des actes
- Audioprothèse :
 - Aide auditive « Classe 1 » : tous types d'appareils
 - Tarifs limités
 - Renouvellement tous les 4 ans

Tous les professionnels sont dans l'obligation de fournir aux assurés un devis avec un équipement du panier 100% santé.

1.3.2. Contrat responsable

Le Contrat responsable fait référence aux contrats liés à la garantie santé, il bénéficie d'avantages fiscaux sous condition de respecter un cahier des charges explicité par le décret du 18 novembre 2014.

Les obligations du contrat responsable sont les suivantes :

- Prise en charge du ticket modérateur sauf :
 - Médicaments remboursés à 15% et 30%
 - Cures thermales
- Interdiction pour le contrat responsable de la prise en charge de :
 - La participation forfaitaire
 - La majoration de participation (sanction liée à l'absence médecin traitant ou non recours à ce dernier)
 - Le dépassement d'honoraire si non-respect du parcours soins

D'autres mesures viennent compléter cela, notamment pour le poste optique avec un remboursement plafonné à 100 euros pour la monture et une limitation de l'équipement à un tous les 2 ans (sauf cas particulier). Un plafond est également appliqué sur le remboursement des verres optiques.

Ces lois ont un impact direct sur les règlements en santé puisque leur application engendre une augmentation des remboursements c'est notamment le cas de la réforme 100% santé. La seconde loi quant à elle permet de diminuer la part de l'assureur, notamment en optique.



Il est donc nécessaire de suivre l'évolution des prestations engendrées par le 100% santé ainsi que le contrat responsable.

2. La santé au sein de Klesia Prévoyance

2.1. Présentation de Klesia

Klesia est un groupe paritaire de protection sociale, il s'agit d'un ensemble structuré de personnes morales, ayant entre elles des liens étroits et durables, piloté et contrôlé par les partenaires sociaux.

L'activité liée à l'assurance de personnes, que nous étudierons dans ce mémoire, est composée de deux types d'organismes assureurs :

- Relevant du code de la sécurité sociale avec trois instituts de prévoyance : Klesia Prévoyance, Carcept Prévoyance, Ipriac.
- Relevant du code de la mutualité : Klesia Mutuelle, JAJI, MACP.

Les organismes assureurs sont agréés pour des branches d'activités, il en existe 26 couvrant des risques liés aux personnes et aux biens.

Dans ce mémoire, l'étude sera menée sur l'institut de prévoyance Klesia Prévoyance. Elle a pour objet d'assurer la couverture du risque décès, des risques liés à la maladie ou à la maternité, des risques d'incapacité de travail, d'invalidité ou d'inaptitude.

Klesia Prévoyance est une institution mixte (vie et non vie) agréée pour les branches d'activités suivantes :

- Les accidents (branche 1)
- La maladie (branche 2)
- La vie - décès (branche 20)

Elle assure, en termes de couverture complémentaire santé et/ou prévoyance, des offres dédiées à différentes branches professionnelles : Hôtel Café Restaurant, Immobilier, Fleuristes, Pharmacie d'Officine, Laboratoire de biologie médicale, Horlogerie de gros, etc.

2.2. Présentation de la base

La base utilisée dans la suite sera uniquement composée des montants remboursés par l'institution de prévoyance Klesia Prévoyance (KP), entité du groupe Klesia.

Sa charge de sinistralité, fluctuant mensuellement entre 5 millions et 20 millions d'euros, en fait l'entité la plus importante du groupe, générant le plus de remboursements.

En 2021, Klesia Prévoyance a enregistré une charge de sinistralité de plus de 261 millions d'euros liée à la santé, faisant de cette entité la plus importante du groupe.

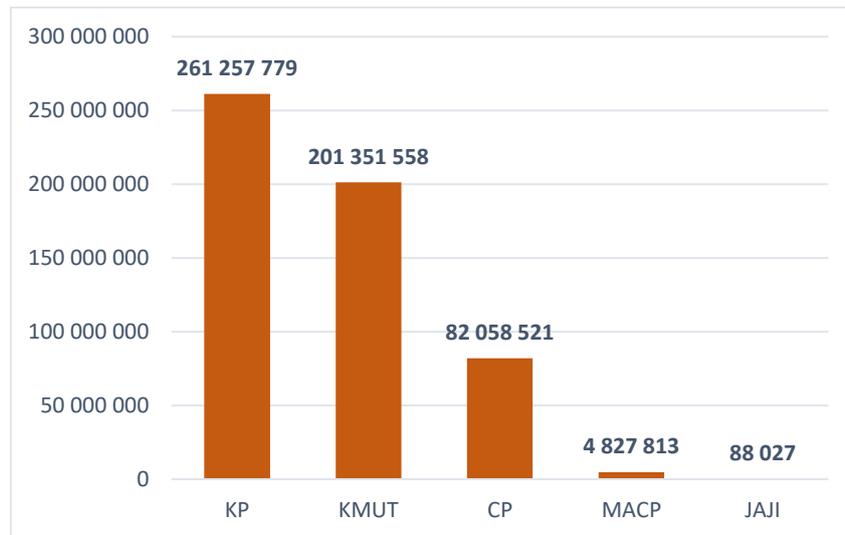


Figure 5 : Remboursements des soins santé en 2021

Notre étude porte donc sur l'entité la plus volumineuse et pourra ensuite être étendue à l'ensemble du groupe.

Elle consistera dans un premier temps à « prédire » les résultats de l'année comptable et de la survenance 2021 afin de contrôler l'ajustement de notre modèle.

Pour approfondir notre analyse de l'entité, Klesia Prévoyance est composée de différents portefeuilles ayant chacun un poids plus ou moins important au sein de l'entité. Après avoir appliqué notre modèle à Klesia Prévoyance, une analyse plus détaillée sera réalisée sur les portefeuilles relevant des branches : IMMO, PHIE, PHAC, LABO, FLEU et CNVS. Ces derniers représentent plus de 50% de l'entité en termes de montants de prestations à savoir 110 millions d'euros sur la survenance 2021.

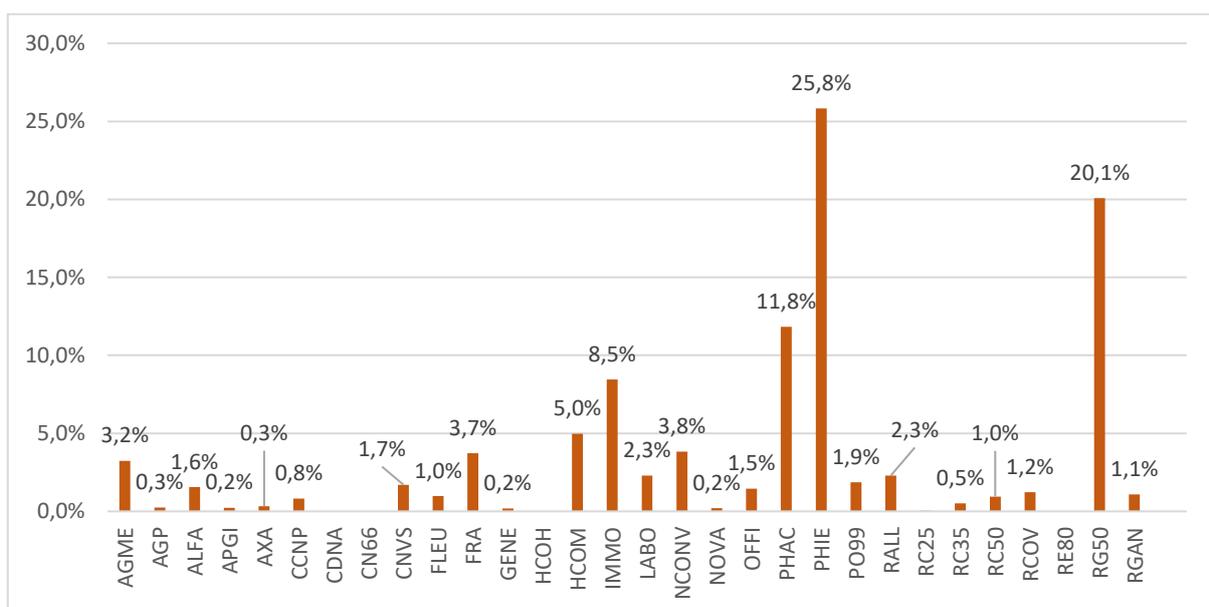


Figure 6 : Répartition des remboursements santé de Klesia Prévoyance par portefeuille

Notre étude sera réalisée selon l'axe de survenance et du mois relatif au remboursement. Ainsi un remboursement effectué en janvier 2022 concernant une prestation survenue en janvier 2021 correspondra au 13e mois. L'axe temporel de notre étude sera noté dans la suite : « 2021_13 », pour un tel remboursement. En effet, l'objectif est d'obtenir un S/P par survenance ce qui nous oblige à inclure cette variable, dans notre étude.

Le périmètre s'étend de la survenance 2011 à 2022, en notant que 2022 est incomplet au moment de la réalisation de ce mémoire puisque les données du 4e trimestre sont manquantes pour cette dernière année. Cependant cette profondeur d'historique nous semble adéquate et pertinente pour notre étude.

La représentation graphique des prestations de l'entité Klesia Prévoyance nous fournit des informations sur la cadence de règlement des prestations.

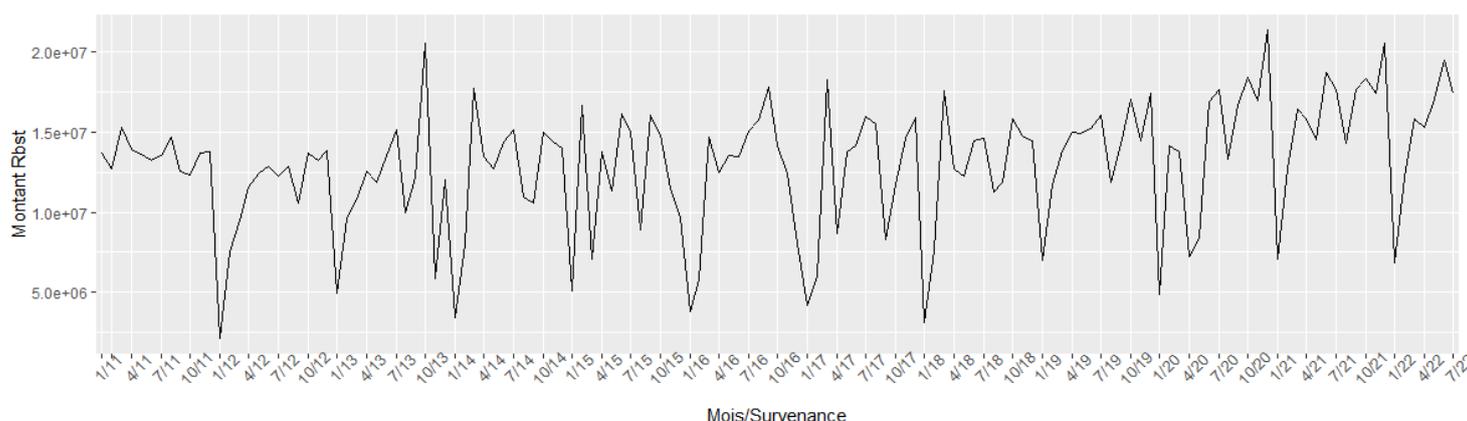


Figure 7 : Remboursements santé de Klesia Prévoyance entre 2011 et 2022

Le premier constat que nous pouvons faire est la saisonnalité importante dans le cadencement des prestations. Nous pouvons décomposer ce graphique en 3 principales périodes :

- Avant 2014, le groupe Klesia était en pleine mutation à la suite de la fusion de deux groupes : Mornay et D.O. Ce rapprochement a engendré des regroupements de plusieurs mutuelles entraînant ainsi des cadences plus ou moins aléatoires.
- Entre 2014 et 2017, on commence à observer une certaine stabilité
- A partir de 2017, les flux sont plus stables et les cadences de règlement sont similaires d'une survenance à l'autre. Une sous-consommation est observable pour chaque survenance sur la période d'avril, de mai et d'août.

Lors de la construction de notre modèle, nous conserverons tout de même les 11 années d'historiques afin d'avoir une profondeur d'historique satisfaisante pour notre étude.

2.3. Statistiques descriptives et analyses

Suite à cette brève présentation, nous réalisons désormais une analyse de la base des prestations santé pour la comprendre. Il est à noter que cette base est différente de celle qui sera utilisée pour

nos modèles. Elle ne contient pas toutes nos données, mais représente une grande partie d'entre elles, ce qui en fait une bonne représentation.

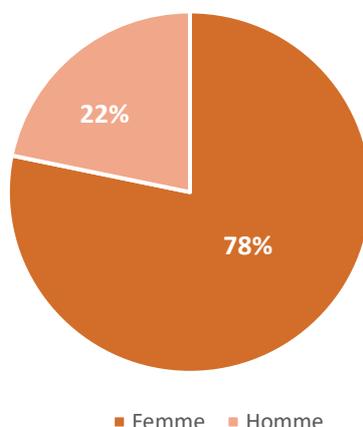


Figure 8 : Répartition par sexe

La composition majoritaire de femmes est le résultat de la présence de secteurs d'activités très représentés dans notre base tels que les pharmacies, les laboratoires ou encore les fleuristes. A titre d'exemple, en 2021, la part des femmes dans le secteur pharmaceutique était de 68%¹.

Les remboursements par âge témoignent d'une présence conséquente des moins de 15 ans dans notre base. Néanmoins Klesia étant une assurance de groupe, la répartition des assurés est supérieure à 20 ans, il ne s'agit que de personnes ayant une activité salariale.

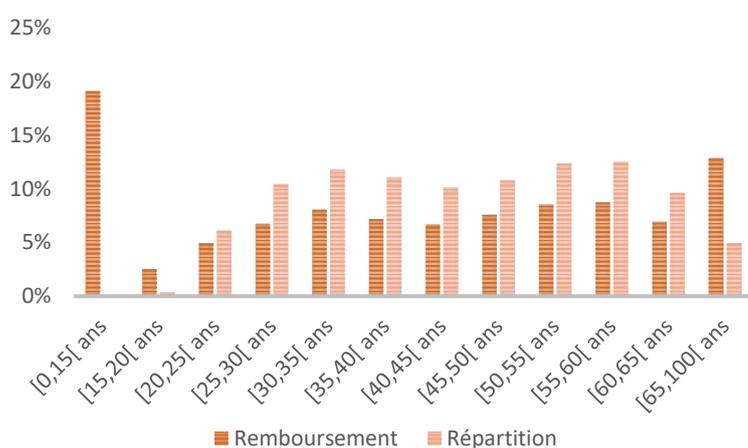


Figure 10 : Remboursements et répartition des soins par âge

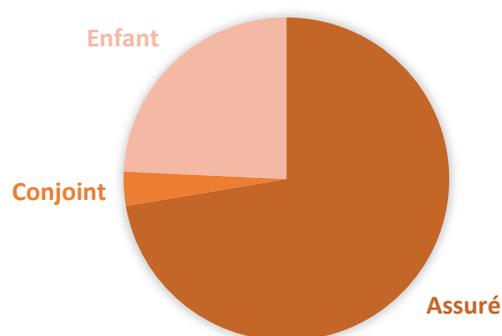


Figure 9 : Répartition par type de bénéficiaire

¹ https://drees.solidarites-sante.gouv.fr/sites/default/files/2021-03/DD76_0.pdf

Concernant les prestations remboursées, les jeunes et les plus de 65 ans sont les plus présents dans la base.

Ces observations sur la répartition par âge et la répartition des remboursements témoignent que les plus jeunes et les séniors ne représentent pas une part importante des effectifs mais contribuent fortement à la sinistralité du portefeuille car ces populations sont les plus sensibles et les plus couteuses.

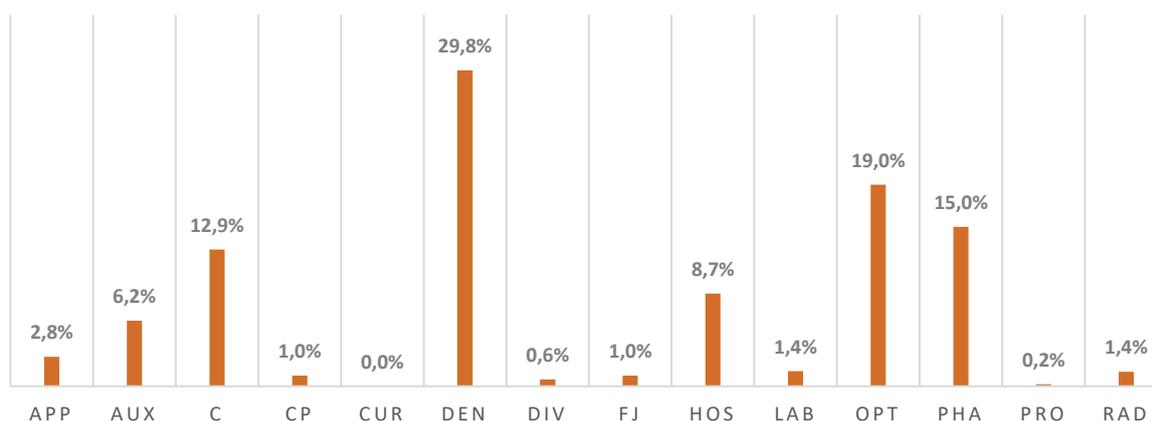


Figure 11 : Remboursements des moins de 15 ans par famille d'acte de soins

La répartition par famille d'acte témoigne d'une concentration des remboursements en optique et dentaire. Aujourd'hui, on considère qu'une personne sur deux a besoin d'un appareil dentaire et ces soins sont majoritairement réalisés pendant l'enfance. A cela s'ajoutent également d'éventuels besoins en optique.

Grâce aux contrats responsables, depuis 2020, les montures de lunettes sont soumises à un plafonnement de 100 euros (anciennement 150 euros entre 2015 et 2020) ce qui permet ainsi de limiter les coûts et la surconsommation sachant que le secteur de l'optique est fortement touché par l'aléa moral.

La pharmacie représente la plus grande part en termes d'effectifs de remboursement. Néanmoins, cette sur-représentation doit être mise en perspective par rapport au montant de remboursement.

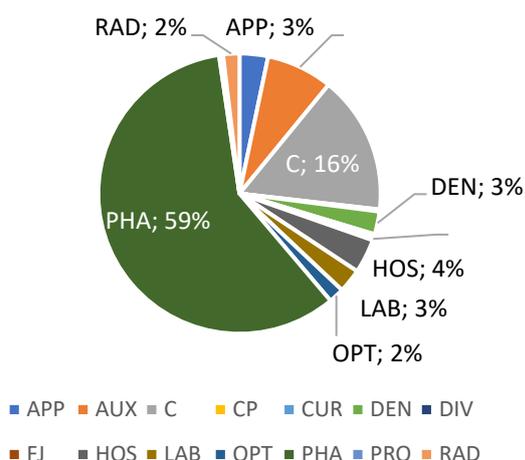


Figure 13 : Répartition en nombre par famille d'acte de soins

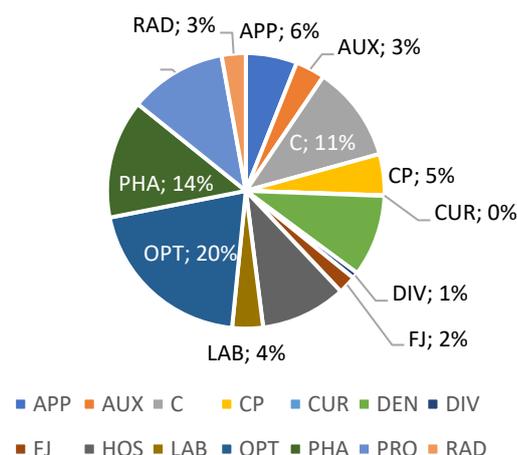


Figure 12 : Remboursements par famille d'acte de soins

En effet, dans ce cas-là, la pharmacie ne représente plus que 14% des montants remboursés, contre 20% en optique, qui est l'acte le plus coûteux dans notre base. Depuis l'instauration du contrat responsable (présenté précédemment), les lunettes ne peuvent être remboursées que tous les deux ans. Cela engendre donc moins d'actes alors que les montants restent conséquents. Ce fonctionnement est inversé pour les médicaments, puisque leur montant, dans la majorité des cas, reste raisonnable, tandis que le nombre est assez conséquent.

2.4. Présentation des indicateurs de prestations

Aujourd'hui, Klesia dispose d'une vision historique des prestations réglées :

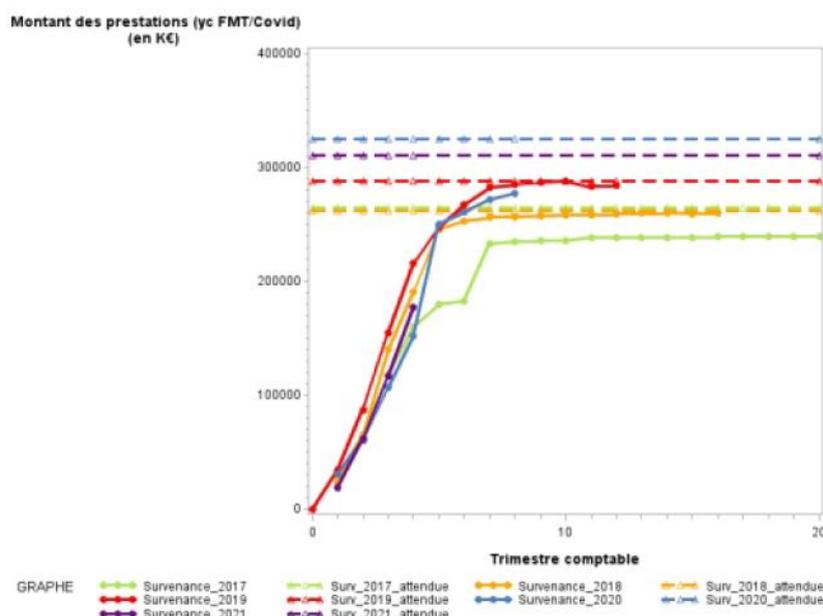


Figure 14 : Graphique historique des prestations santé observées et attendues par survenance sur l'entité Klésia Prévoyance

Les courbes représentent la charge de sinistralité par survenance au cours du temps tandis que les tracés horizontaux indiquent le niveau de sinistralité attendu.

Ce graphique est mis à jour et utilisé trimestriellement afin d'avoir un « état des lieux » de nos prestations réelles cumulées face à notre estimation réalisée lors de l'inventaire. Il est également possible d'obtenir rapidement et visuellement une comparaison entre les différentes survenances afin de constater une potentielle dérive.

De plus, une analyse plus précise est réalisée afin d'estimer les montants de boni-mali (BM) constatés. Pour cela, les provisions, les prestations réelles ainsi que la tendance restante sont prises en compte dans le calcul des BM. Aujourd'hui, la tendance est simplement obtenue par une moyenne des trois voire deux dernières années lorsque des événements atypiques viennent « polluer » le calcul de la tendance. C'est notamment le cas de l'année 2020 marquée par la Covid.

Voici un tableau récapitulatif des études réalisées au 2e trimestre 2022 afin d'estimer les boni-mali de prestations sur l'entité du groupe :

Entité	Estimation inventaire (yc PSI)	Prestations réelles au 2T	BM	Tendance restante	BM Finaux
KP	195,4	186,4	-8,9	11,1	-0,7M€

Figure 15 : Estimation des Boni-Mali en 2022

Ces résultats sont utilisés dans le cadre du Plan prévisionnel d'Activité qui consiste à estimer/déterminer le résultat de la survenance en cours ainsi que les prochaines survenances.

Les résultats de ce mémoire pourraient notamment permettre d'avoir une estimation plus précise des futures tendances pour le risque santé.

Ils pourront également être utilisés à d'autres fins, nous allons maintenant détailler les enjeux de ce mémoire.

2.5. Limite de l'outil

L'enjeu est donc de mettre en place un modèle permettant de projeter les flux relatifs au risque santé pour avoir une vision future de la rentabilité du groupe notamment à travers la charge de sinistralité.

Ce modèle pourra apporter une aide à la décision lors des prochaines clôtures, notamment pour l'atterrissage qui est effectué au 3e trimestre afin d'estimer le résultat de fin d'année. Cet outil pourra être utile dans le cadre de l'estimation des comptes de fin d'année (Solvabilité 1).

Dans un premier temps, il est nécessaire d'estimer les prestations du 4e trimestre compte tenu du fait que la clôture s'effectue en fast-close avec une date d'arrêtée au 30/09/N. Avec ce nouvel outil, une validation de ces estimations pourra être réalisée.

D'autre part, lors de l'estimation du budget, un S/P par survenance est obtenu à partir de celui de l'an passé auquel on applique un taux d'indexation ainsi que des hypothèses de dérive ou de vieillissement. Cette méthode classique pourra être remplacé par les prévisions réalisées dans ce mémoire. Grâce à ces dernières, il sera possible d'estimer la charge de sinistralité par survenance et ainsi obtenir les indicateurs S/P nécessaires pour l'estimation du plan prévisionnel d'activité.

Ces résultats seront également utilisés dans le cadre de travaux de Solvabilité 2. L'objectif sera d'évaluer l'impact sur le ratio S2. La dernière partie de ce mémoire consistera à évaluer les différents impacts de ces scénarios via le ratio Solvabilité 2.

Le mémoire s'articule donc selon deux axes : la première étape consistera en la création du modèle en appliquant les modèles de série temporelles et de lissage exponentiel, une fois les modèles validés, nous serons en mesure d'estimer la charge de sinistralité par survenance et par conséquent les S/P associés. Dans un second temps, une étude sous Solvabilité 2 sera appliquée afin d'établir les nouveaux ratios S2 obtenus par la méthode développée dans ce mémoire et ainsi vérifier les impacts que ces travaux pourraient avoir.

Dans un premier temps, nous allons présenter et décrire les différents modèles mathématiques.

3. Modèles utilisés dans le domaine de l'actuariat

La simulation des montants de prestations en santé à partir d'une base de données historique peut être abordée à l'aide de divers modèles mathématiques et statistiques, en fonction de la complexité des données et des objectifs spécifiques de la simulation.

Plusieurs méthodes de prédiction existent, classées selon le type :

- Méthodes endogènes : série temporelle
- Méthodes exogènes : GLM, arbre de décision, réseau de neurone ...

La **méthode endogène** se base uniquement sur l'historique de nos données afin d'estimer notre futur. Cette méthode ne fonctionne que si le passé est assez complet.

Les modèles SARIMA, les lissages exponentiels, ... entrent dans cette catégorie.

Dans ces modèles, on suppose souvent que notre jeu de données présente une tendance, une saisonnalité, et une composante aléatoire.

A la différence de la méthode endogène, la **méthode exogène** s'appuie sur des variables explicatives qui influent sur le passé de notre jeu de données. L'objectif étant de sélectionner les bonnes variables qui expliquent au mieux notre variable à prédire.

Les régressions linéaires ou GLM mais aussi les modèles de Machine Learning sont des méthodes de prévision exogène.

Nous allons décrire brièvement les méthodes exogènes ci-dessous.

3.1. Les méthodes exogènes

GLM

Les modèles de régression linéaire généralisée (GLM) sont fréquemment employés pour analyser les coûts des soins de santé. Ils constituent une extension des modèles de régression linéaire traditionnels en permettant d'utiliser des distributions de probabilité variées pour la variable que l'on souhaite étudier, ce qui les rend adaptés à des données qui ne se conforment pas à une distribution normale.

Ces modèles permettent de créer une modélisation de la variable à étudier, Y (appelée "variable réponse"), en relation avec plusieurs variables explicatives. Autrement dit, ils nous aident à comprendre comment les facteurs explicatifs influent sur les coûts des soins de santé.

Les variables explicatives sont soigneusement choisies de manière à ce qu'elles aient un impact sur notre variable à étudier, en l'occurrence le montant des prestations.

Après avoir passé en revue les mémoires² d'actuaire traitant de ce sujet, il est observé que les variables explicatives qui sont généralement susceptibles d'influencer les coûts de santé comprennent :

- Le type de bénéficiaire : assuré, bénéficiaire, enfant
- L'ancienneté
- La catégorie Socio-professionnelle
- L'âge
- Le sexe

Les GLM offrent divers avantages :

- La capacité à modéliser différents types de données, qu'il s'agisse de données continues, binaires ou de comptage.
- La possibilité d'obtenir des informations sur l'impact de chaque variable explicative sur la variable à étudier grâce à des coefficients interprétables.

Néanmoins, ces modèles présentent quelques limites notamment :

² L.MARCIANO. (2018). Modélisation de la dérive des soins de santé à court terme.

T.LE HO (2018). Reporting santé : comment le perfectionner à l'aide de la Data Science ?

- Une certaine sensibilité aux valeurs aberrantes
- Si la distribution de la variable à étudier n'est pas correctement connue ou spécifiée, les résultats peuvent être biaisés
- Une base de données substantielle et fiable est également requise

Machine Learning

Les arbres jouent un rôle essentiel dans l'analyse de données et la prise de décision en offrant des méthodes d'exploration et de gestion. Ils sont utilisés pour modéliser et anticiper une variable quantitative ou qualitative en se basant sur diverses variables explicatives.

Plusieurs approches sont disponibles pour élaborer des arbres. Nous nous focalisons ici sur la première méthode, à savoir l'algorithme Classification And Regression Trees (CART), reconnu comme la référence en matière d'apprentissage automatique. Cet algorithme permet de créer des classes d'individus en utilisant les variables explicatives de manière à maximiser l'homogénéité des individus au sein d'une même classe par rapport à la variable d'intérêt. Par le biais de divisions successives, l'algorithme CART va créer des sous-groupes plus homogènes que les précédents au sens d'un critère dépendant du type de la variable Y.

La seconde méthode présentée : Random Forest, représente une technique d'apprentissage utile pour aborder des problèmes de classification et de régression. Cette approche est basée sur le concept de Bagging, qui consiste à combiner plusieurs modèles pour produire une seule prédiction. Le modèle Random Forest introduit une composante aléatoire. L'objectif est d'accroître l'indépendance entre les arbres en introduisant de l'aléa dans le choix des variables utilisées pour construire chaque arbre.

Ainsi, une forêt aléatoire est constituée d'un ensemble d'arbres de décision, chaque arbre étant construit à partir d'un échantillon aléatoire de données d'entraînement.

En comparaison avec le modèle GLM, ces modèles comportent certains avantages :

- Une approche paramétrique : ils n'imposent aucune distribution de probabilité spécifique à la variable à expliquer Y
- La prise en compte des interactions non linéaires entre les variables explicatives et la variable réponse

Cependant, l'algorithme CART présente des lacunes, il est particulièrement instable et réagit fortement aux variations de l'échantillon ayant servi à construire l'arbre. Introduire de l'aléa dans les modèles de Random Forest permet de remédier à cette fragilité, qui peut se manifester lorsque l'on se fie à un seul arbre de régression pour effectuer des prévisions.

3.2. Adéquation avec nos enjeux

L'objectif des modèles CART, Random Forest et GLM est d'effectuer des prédictions à une granularité fine, c'est-à-dire au niveau individuel. Cependant, dans le cadre de notre étude, une telle approche détaillée n'est tout simplement pas requise. Nous nous appuyons sur les résultats de l'Estimé Budget, qui sont calculés à une maille plus agrégée, au niveau d'un portefeuille voire d'une entité. Par conséquent, réaliser ces estimations avec une telle précision n'est pas adapté à notre étude.



D'autre part, pour appliquer ces modèles, il serait indispensable de disposer de deux bases de données contenant des variables explicatives : une première base comprenant l'ensemble des cotisants et une seconde basée sur les prestations effectivement réglées. Dans ce contexte, il ne suffit pas de se limiter aux prestations car il est crucial d'inclure les cotisants qui n'ont pas bénéficié de remboursements. Pour ce faire il serait nécessaire de recueillir toutes les données auprès de nos délégataires. Etant donné que la période de l'estimé budget est limitée dans le temps, l'utilisation de ces modèles exogènes ne serait ni réalisable, ni pertinente pour l'application envisagée dans notre étude.

Ainsi, ce mémoire privilégiera les approches endogènes, car elles s'adaptent parfaitement à notre contexte, offrant plusieurs avantages, notamment :

- La capacité de modéliser les tendances et les variations saisonnières
- Une flexibilité qui permet de s'adapter à diverses situations
- La possibilité de réaliser des prévisions à court et long terme
- Une applicabilité dans une variété de domaines tels que l'actuariat, la finance, ...

Notre choix de modèle est conforté par les analyses et études faites sur le marché dans le cadre de différents mémoires réalisés dans les entreprises telles que Galea, April ou encore Groupama³.

Les modèles SARIMA et lissage exponentiel seront donc appliqués à nos données. Nous commencerons dans un premier temps par décrire les modèles SARIMA.

³ (E.BONIN, 2010) (M.BRIEC, 2013) (G. STASINSKI, 2022)

Partie 2 : Théorie sur la projection des prestations santé

1. Les Séries Temporelles

1.1. Généralités

Dans la suite, on définira $(\Omega, \mathcal{F}, \mathbb{P})$ comme un espace probabilisé avec Ω l'espace des événements, \mathcal{F} une tribu adaptée à Ω et \mathbb{P} un espace probabilisé.

Processus du 2nd ordre

On appelle processus aléatoire du 2nd ordre à temps discret, toute suite réelle $X_t : \Omega \rightarrow \mathbb{R}$ définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ telle que :

$$E(|X_t|^2) < \infty \text{ pour tout } t$$

1. Moyenne : $\mu(t) = E(X_t), \forall t \text{ dans } Z$
2. Auto-covariance : $K : Z \times Z \rightarrow \mathbb{R}, K(t, s) = E\left((X_t - \mu(t))(X_s - \mu(s))\right), \forall t, s \in Z$

Bruit blanc

On définit deux types de bruit blanc :

- Bruit Blanc Fort : il s'agit d'une suite de variables indépendantes et identiquement distribuées (ε_t) centrées et de variance σ^2 . On notera $\varepsilon_t \sim \text{BBF}(0, \sigma^2)$
- Un Bruit Blanc Faible est une suite de variables centrées avec :
$$E(\varepsilon_t) = 0, V(\varepsilon_t) = \sigma^2 \text{ et } \text{Cov}(\varepsilon_t, \varepsilon_s) = 0 \text{ avec } t \neq s$$

On notera $\varepsilon_t \sim \text{BB}(0, \sigma^2)$

Dans la suite, nous ferons l'hypothèse que tous les bruits blancs sont faibles.

Opérateur Retard et Avance

On note B l'opérateur retard tel que $BX_t = X_{t-1}$ avec X_t un processus stationnaire et F l'opérateur avance avec $FX_t = X_{t+1}$

On en déduit les propriétés suivantes :

- $B^k X_t = X_{t-k}$
- $B^0 = Id$
- $B^{-1} = F$

1.2. Processus stationnaire : modèles ARMA

Stationnarité

La stationnarité implique que la série ne dépende pas du temps et plus particulièrement qu'elle ne comporte aucune tendance saisonnière ni aucune tendance à la hausse ou à la baisse. On distinguera par la suite 2 types de stationnarité : forte et faible.

Un processus est dit **strictement** stationnaire (stationnarité forte) si pour toute suite finie d'instants $t_1, t_2, \dots, t_k \in \mathbb{Z}$ et pour tout $r \in \mathbb{Z}$, les lois jointes de $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ et de $(X_{r+t_1}, X_{r+t_2}, \dots, X_{r+t_k})$ sont les mêmes i.e. les lois jointes sont invariantes par translation dans le temps.

Cette définition impose que les lois des X_t soient identiques ce qui peut être particulièrement contraignant. La stationnarité au 2nd ordre ou la « faible » stationnarité » sera privilégiée.

Un processus est stationnaire au 2nd ordre (faiblement stationnaire) si la moyenne et la covariance des X_t sont invariantes par translation dans le temps. On a donc :

- $E(X_t) = \mu$, indépendante de t
- $\gamma_X(s, t) = Cov(X_s, X_t) = E[(X_s - E(X_s))(X_t - E(X_t))]$
La fonction d'autocovariance ne dépend que de $(t-s)$ donc $K(t,s) = \gamma(t-s)$

Modèles ARMA

Un processus stationnaire au 2nd ordre X est un processus ARMA(p,q) s'il existe un bruit blanc ε_t défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ et des nombres réels $a_1, \dots, a_p, c_1, \dots, c_q$ tels que :

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=1}^q c_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

En particulier, on a X un processus de moyenne mobile MA(q) tel que :

$$X_t = \varepsilon_t + \sum_{j=1}^q c_j \varepsilon_{t-j} \quad (2)$$

Et X un processus Autorégressif AR(p) avec :

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \varepsilon_t \quad (3)$$



Cette équation peut se réécrire de cette façon : $A(B)X_t = \varepsilon_t$

Avec, A le polynôme unitaire ($A(0)=1$) et $A(x) = 1 - \sum_{k=1}^p a_k x^k$

1.3. Processus non stationnaires

Dans certains cas, le processus peut ne pas être stationnaire, d'autres modèles peuvent alors s'avérer utiles. Seuls les modèles ARIMA et SARIMA seront présentés dans ce mémoire.

1.3.1. Modèles ARIMA et SARIMA

Modèle ARIMA

$X = (X_t)$ est un ARIMA lorsque le processus ne satisfait pas la condition de la stationnarité.

Si $Y_t = (1 - B)^d X_t$ (4) est un processus ARMA(p,q) alors X_t est un processus ARIMA(p,d,q). En effet $1 - B$ correspond à l'opérateur de différentiation qui permet de corriger la non-stationnarité.

Pour identifier le rang de d, on étudie la série brute puis résiduelle jusqu'à ce que l'on puisse accepter la stationnarité. « d » correspondra alors au nombre de fois où la stationnarité a été rejetée.

On obtient donc d'après l'équation (4) :

$$(1 - B)^d A(B)X_t = C(B)\varepsilon_t \quad (5)$$

Modèle SARIMA

$X = (X_n)$ est un SARIMA si comme précédemment le processus n'est pas stationnaire. La différence entre ces deux modèles tient du fait que, pour ce dernier, la série présente une saisonnalité à retirer.

Dans ce cas-là, X_n est un SARIMA(p,d,s,q) si on a :

$$(1 - B)^d (1 - B^s)^D A(B)X_n = C(B)\varepsilon_n \quad (6)$$

La partie $(1 - B^s)^D$ permet de désaisonnaliser la série dont la saison est « s ». L'estimateur s est en général connu, on peut donc se ramener à un ARIMA avec : $(1 - B^s)X_t$

1.3.2. Etapes de sélection du modèle

Une fois la stationnarité vérifiée, on peut décomposer la modélisation en 5 étapes :

- 1- Identification des degrés p et q
- 2- Estimation des paramètres a_k, c_j
- 3- Réduction du modèle avec la significativité des paramètres
- 4- Validation du modèle avec les différents tests d'adéquation
- 5- Sélection du modèle le plus optimal

Les étapes de modélisation d'une série par un ARIMA :

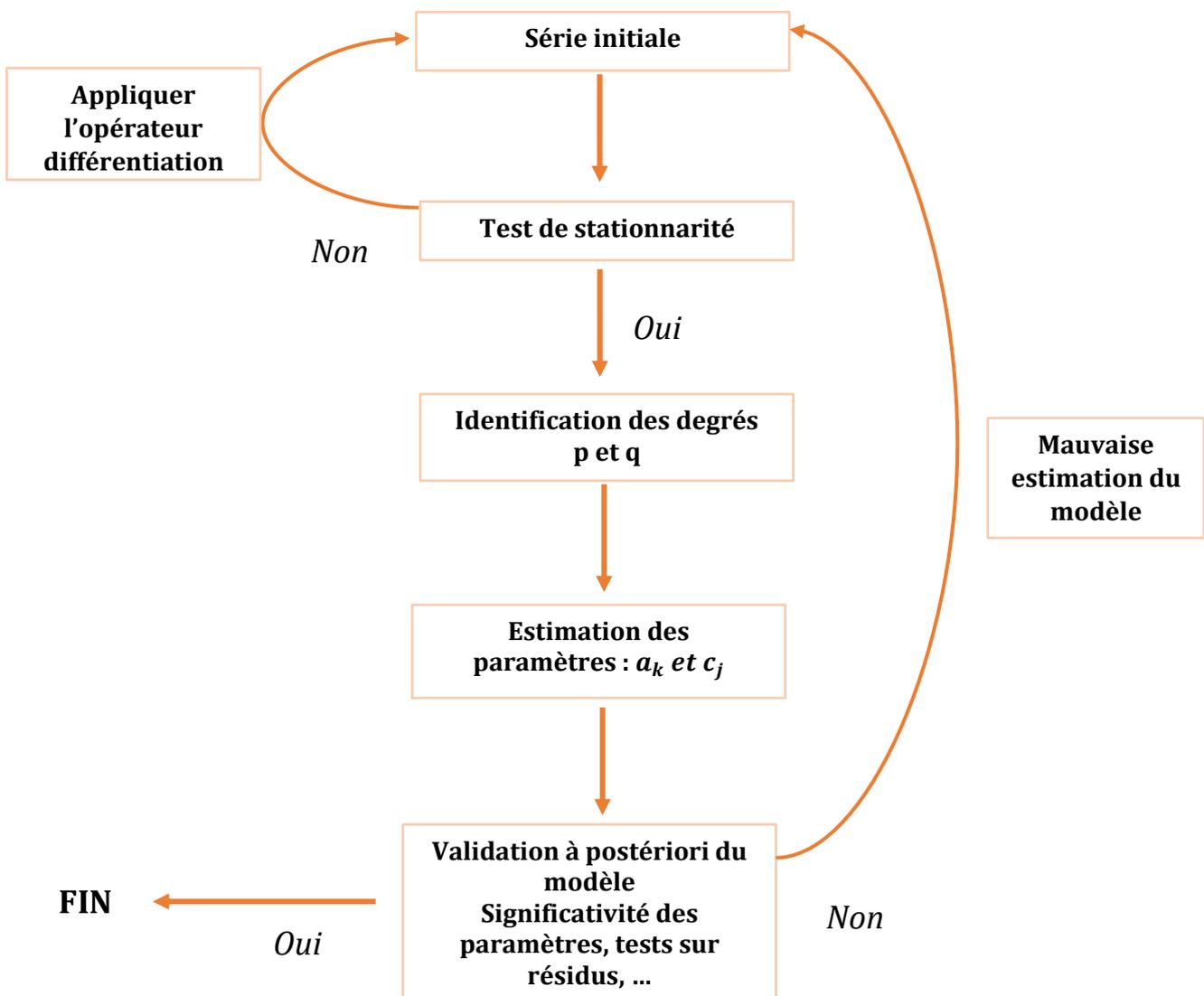


Figure 16 : Etapes de sélection du modèle

Suppression de la stationnarité

Avant tout, la 1er étape est visuelle, elle consiste à observer graphiquement la série. C'est lors de cette étape que l'on pourra détecter en amont la présence de valeurs extrêmes mais également la non-stationnarité avec des tendances à la hausse (resp. baisse) ou des tendances saisonnières.

Ces observations pourront ensuite être validé à l'aide de test plus précis, notamment :

- ❖ Tests de non-stationnarité :
 - Dickey Fuller
 - Phillips Perron

- ❖ Test de stationnarité :
 - Test KPSS

Pour les tests de non-stationnarité, on part du modèle suivant :

$$X_t = \rho X_{t-1} + \epsilon_t \quad (7)$$

Et on teste l'hypothèse : $H_0 : \rho = 1$ contre $H_1 : \rho < 1$. Si $\rho = 1$ alors l'équation (7) est une marche aléatoire et n'est donc pas stationnaire.

Dans le cas où la condition de stationnarité ne serait pas vérifiée, la différentiation est la méthode qui sera appliquée dans ce mémoire. Dans le cadre d'une différentiation d'ordre 1 on aura donc :

$$Y_t = X_t - X_{t-1} = (1 - B)X_t \quad (8)$$

On différenciera la série autant que nécessaire jusqu'à ce qu'elle devienne stationnaire, on obtiendra donc :

$$Y_t = (1 - B)^k X_t, \text{ avec } Y_t \text{ la série obtenue après } k \text{ différenciations}$$

Estimation des degrés

Afin d'estimer les degrés p et q de notre modèle ARMA, la méthode la plus utilisée consiste à étudier les fonctions d'autocorrélation (ACF) et fonction d'autocorrélation partielle (PACF) avec :

❖ Fonction d'autocorrélation

- $\rho(j) = \frac{\gamma(j)}{\gamma(0)} \forall j \in \mathbb{Z}$
- $\gamma(0) = \mathbb{V}(X_t) \forall t \in \mathbb{Z}$

La fonction d'auto-corrélation d'un processus MA(q) vérifie $\gamma(j) = 0 \forall j > q$

❖ Fonction d'autocorrélation partielle

$$r(j) = \text{corr}(X_1 - \mathbb{P}(X_1 | X_2, \dots, X_j), X_{j+1} - \mathbb{P}(X_{j+1} | X_j, \dots, X_2))$$

Si X_t est un processus AR(p) alors la fonction d'auto-corrélation partielle $r(j) = 0 \forall j > p$.

Le processus ARMA est choisi de telle sorte que la série temporelle s'adapte le mieux possible à nos données. L'étude de l'ACF et du PACF permet de mettre en évidence les degrés qui seront utilisés dans la suite de la modélisation.

Les fonctions d'autocorrélations sont des processus du 2nd ordre, ils sont donc adaptés dans l'étude des séries temporelles puisque les processus sont caractérisés par leur moment.

1.3.3. Validation du modèle

Tests Statistiques

Pour tester la significativité de notre modèle, l'analyse des résidus est la première étape. Si ces derniers sont non-corrélés et gaussiens alors le modèle pourra être validé puisqu'il s'ajustera au mieux à nos données. Les résidus seront donc des bruits blancs gaussiens.

❖ Test de normalité

La normalité des résidus peut être vérifiée soit par une analyse descriptive avec le graphique QQ Plot et la droite de Henry soit par des tests statistiques.

Concernant l'analyse descriptive, le QQ Plot permet de comparer nos données à la distribution d'une loi normale. Cependant, l'analyse descriptive n'est pas la solution la plus rigoureuse c'est pourquoi, les tests statistiques seront privilégiés.

Jarque - Bera

Le test de Jarque-Bera est le 1er test statistique que nous allons étudier dans cette section. Il est basé sur les coefficients d'asymétrie et d'aplatissement. L'objectif étant de calculer les écarts entre les coefficients issus de nos données et ceux de la loi normale. On a donc la statistique de test suivante :

$$T = n * \left(\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right) \sim \chi^2(2)$$

S et K étant respectivement l'asymétrie et l'aplatissement de notre jeu de donnée :

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}} \quad \text{et} \quad K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

La région critique est décrite de cette façon :

$$T > \chi^2_{1-\alpha}(2)$$

Par exemple, si le risque acceptable est de 5% alors le seuil critique est 5,991.

Lilliefors

Le test de Lilliefors est issu du test de Kolmogorov-Smirnov mais adapté à la loi normale.

Quant à la statistique du test elle est identique à celle du test de Kolmogorov-Smirnov la différence est que l'espérance et la variance de la loi normale sont inconnus.

$$D = \max_{i=1, \dots, n} \left(F_i - \frac{i-1}{n}, \frac{i}{n} - F_i \right)$$

Comme pour les précédents tests, on obtient une région critique pour la statistique du test D :

$$D > D_{crit}$$

Les valeurs critiques de D sont disponibles à partir du tableau présent en annexe.

❖ Test de blancheur

Dans cette partie, le test consiste à vérifier, pour les résidus du modèle, qu'il s'agisse bien de bruit blanc. Pour cela deux tests seront utilisés : Ljung-Box et Box Pierce, l'hypothèse nulle repose sur la blancheur de nos données dans les deux cas.

On obtient la région critique du test à partir des statistiques de tests suivants :

- Ljung-Box :

$$Q_1 = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

- Box-Pierce :

$$Q_2 = n \sum_{k=1}^h \hat{\rho}_k^2$$

Avec, $\hat{\rho}_k$ l'estimateur de la corrélation et C_k l'estimateur de la covariance :

$$\hat{\rho}_k^2 = \frac{C_k}{C_0} \text{ et } C_k = \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})$$

On rejette donc le test à partir de la région critique suivante :

$$Q_1 < \chi^2_{1-\alpha}(k) \quad (\text{respectivement } Q_2)$$

Critère AIC

Une fois les précédents tests réalisés et validés, un choix devra être réalisé entre différents modèles. Il est nécessaire d'utiliser un critère, parmi lesquels nous retrouvons notamment l'AIC, l'AICc, le BIC, ...

Un bon modèle est caractérisé par une adéquation entre un faible nombre de paramètre et une bonne qualité de prédiction. La complexité réside dans le fait que plus le nombre de paramètre augmente et plus la qualité du modèle augmente et inversement. Il est donc nécessaire de trouver un juste équilibre entre une excellente prédiction et un nombre de paramètre raisonnable. En effet, la complexité d'un modèle peut engendrer un sur-ajustement de ce dernier.

Le sur-ajustement entrainera la prise en compte des bruits blancs ce qui empêchera une généralisation du modèle et aboutira donc à des prédictions d'une faible fiabilité. Il suffit donc de trouver un juste équilibre entre le nombre de paramètre et le pouvoir prédictif du modèle.

C'est pourquoi nous utiliserons le critère AIC pour atteindre cet équilibre. Ce dernier est calculé par la formule suivante :

$$AIC = n * \ln \hat{\sigma}^2 + 2 * (p + q)$$

avec :

- n : le nombre de données
- p et q : les paramètres d'un ARMA(p,q)

Compte tenu des données à notre disposition, nous allons utiliser le modèle SARIMA en tant que modèle endogène sur nos données. Nous travaillons sur des données par survenance contenant nécessairement une saisonnalité.

Après les séries temporelles, le lissage exponentiel sera également appliqué, la théorie est décrite dans cette partie.

2. Le Lissage Exponentiel

On peut définir le lissage exponentiel comme une méthode empirique de prédiction basée sur l'historique qui donnera un poids plus important aux données récentes. Le but étant de minimiser ou augmenter le poids des valeurs passées en fonction de leur ancienneté.

Il existe trois méthodes de lissage exponentiel :

- Lissage simple - LES
- Lissage exponentiel double
- Lissage exponentiel de Holt-Winters

2.1. Lissage exponentiel simple

La première méthode consiste à calibrer un modèle sur des données non saisonnières avec une tendance constante. Comme son nom l'indique, cette méthode reste assez simple pour des données telles que les nôtres.

Elle est définie telle que :

$$\hat{Y}_{t+1|t} = \alpha Y_t + (1 - \alpha) \hat{Y}_{t|t-1}$$

avec :

- Y_t : une série temporelle
- $\alpha \in [0,1]$: le facteur de lissage

Plus le facteur de lissage tend vers 1 et plus le modèle donnera un poids important aux anciennes données lors de la prédiction, et inversement lorsqu'il tend vers 0.

Cette méthode est généralement utilisée pour des données ne présentant ni saisonnalité ni tendance.

2.2. Lissage exponentiel Double

Il s'agit d'une extension du premier modèle, puisqu'il permet de modéliser des données présentant une tendance.

La constante h étant la projection on obtient ainsi :

$$\hat{Y}_{t+h|t} = b_t h + l_t$$

avec :

- $l_t = l_{t-1} + b_{t-1} + (1 - (1 - \alpha)^2)(Y_t - \hat{Y}_{t|t-1})$
- $b_t = b_{t-1} + \alpha^2(Y_t - \hat{Y}_{t|t-1})$

2.3. Lissage exponentiel Holt-Winters

Pour finir, le lissage exponentiel de Holt-Winters est une généralisation de ces premières méthodes, elle est la plus aboutie.

Cette méthode nécessite de déterminer au préalable le type de modèle, la présence d'une tendance et/ou d'une saisonnalité.

Grâce à cela, il est possible d'obtenir 3 modèles différents :

- 1- $Y_t = T_t$
- 2- $Y_t = T_t + S_t$
- 3- $Y_t = T_t * S_t$

Avec :

- T_t : tendance
- S_t : saisonnalité

Le lissage exponentiel de Holt-Winters diffère du précédent puisqu'il est défini par deux paramètres de lissage, α et β avec :

$$l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

De plus, il permet d'ajuster une composante saisonnière :

$$\hat{Y}_{t+h} = l_t + hb_t + s_t$$

Avec, $s_t = \delta(Y_t - l_t) + (1 - \delta)s_{t-T}$, $\delta \in [0,1]$

Il s'agit ici du cas d'un modèle additif, pour le cas multiplicatif, on a :

$$\hat{Y}_{t+h} = (l_t + hb_t)s_t$$

Avec :

- $l_t = \alpha \frac{Y_t}{s_{t-T}} + (1 - \alpha)(l_{t-1} + b_{t-1})$
- $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$
- $s_t = \delta \frac{Y_t}{l_t} + (1 - \delta)s_{t-T}$
- $\alpha, \beta, \delta \in [0,1]$

Les modèles de lissage exponentiel simple et double ne sont pas assez performants pour pouvoir être appliqués à nos données. Ils ne permettent pas de capter la saisonnalité ce qui les rend inutilisable dans notre cas pratique. Notre base est construite par survenance ce qui engendre naturellement une saisonnalité.



Cependant le lissage exponentiel de Holt-Winters sera appliqué dans les prochaines parties puisque contrairement aux autres il présente une composante saisonnière mais également une tendance.

Nous venons de décrire les différents modèles que nous allons appliquer à nos données dans le cadre de ce mémoire, le premier appliqué concerne les séries temporelles.

Partie 3 : Mise en œuvre et application des modèles

1. Modèles SARIMA

1.1. Type de modèle : additive ou multiplicative

Modèle additif

Le modèle ARMA additif se décrit à l'aide de l'équation suivante :

$$Y_T = S_T + T_T + \varepsilon_T$$

Graphiquement, dans le modèle additif, l'amplitude des observations est constante dans le temps. Autrement dit, l'écart entre les maximas et les minimas ne varie que très peu au cours du temps.

Dans ce modèle, les composantes saisonnières et résiduelles sont indépendantes de la tendance.

Avant d'étudier notre série temporelle il faut avant tout identifier le type de modèle à savoir additif ou multiplicatif. Deux méthodes permettent cela : méthode de la bande et la méthode analytique de Buys-Ballot.

❖ Méthode de la bande

Dans notre étude, les variations correspondant à la première année comptable des survenances de 2011 à 2020 forment bien un modèle additif.

La méthode de la bande permet de tracer les maximas et minimas de nos observations, si les deux droites sont parallèles il s'agit d'un modèle additif.

Dans notre cas, les deux droites sont parallèles :

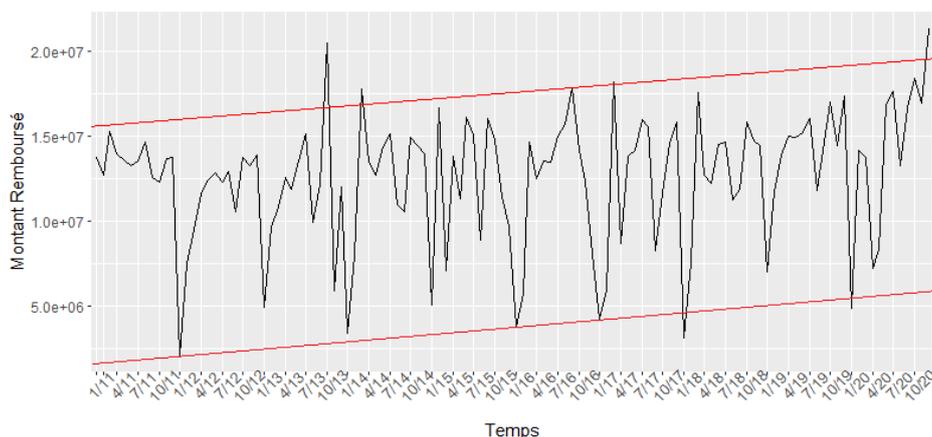


Figure 17: Méthode de la bande sur les prestations santé KP par survenance sur l'année comptable N entre 2011 et 2020

❖ Méthode analytique : Buys-Ballot

La méthode analytique consiste à calculer la moyenne et les écarts types des observations par année de survenance. Le tracé de la droite des moindres carrés : $\sigma = \alpha x + \beta$ permet de déterminer le type de modèle, si α est proche de 0 il s'agira d'un modèle additif et inversement pour le modèle multiplicatif.

Nos résultats sont condensés dans ce tableau :

Survenance	2013	2014	2015	2016	2017	2018	2019	2020
Ecart-type	11586955	12483123	12180546	12249460	12255987	12547553	14059650	14132890
Moyenne	4066284	3820247	3820247	4210942	4455716	3907422	2825222	4984338

Figure 18 : Ecart-type et moyenne par année de survenance

Les paramètres α et β de la droite des moindres carrés sont donnés dans la figure 20. Le coefficient « alpha » est estimé à -0,24. L'intervalle $[(\alpha \pm 2) * StdError] = [-0.9; 0.7]$ conduit à accepter la nullité du coefficient alpha.

	Estimate	Std. Error	T value	Pr(>t)
β	6767168	5.6e+06	1,208	0,266
α	-0,2499	0,4388	-0,57	0,587

Figure 19 : Résultats de la méthode Buys-Ballot

En prenant en compte ces deux méthodes, on accepte le fait que nos données forment un modèle additif.

Modèle Multiplicatif

Le modèle multiplicatif se décrit à l'aide de l'équation suivante :

$$Y_T = S_T * T_T * \varepsilon_T$$

A l'inverse du modèle additif, la composante multiplicative implique que les composantes saisonnières et résiduelles soient dépendantes de la tendance.

Graphiquement, on constate à la fois une augmentation des maxima et une diminution des minima au cours du temps : l'amplitude des variations augmente.

On peut facilement se ramener à un modèle additif en appliquant le logarithme :

$$\ln Y_T = \ln S_T + \ln T_T + \ln \varepsilon_T$$

❖ Méthode de la bande

Contrairement au modèle lié aux remboursements de la première année comptable pour chaque survénance, les prestations liées à la deuxième année comptable semblent former un modèle multiplicatif compte tenu du fait que les droites formées par les maxima et les minima se croiseront obligatoirement.

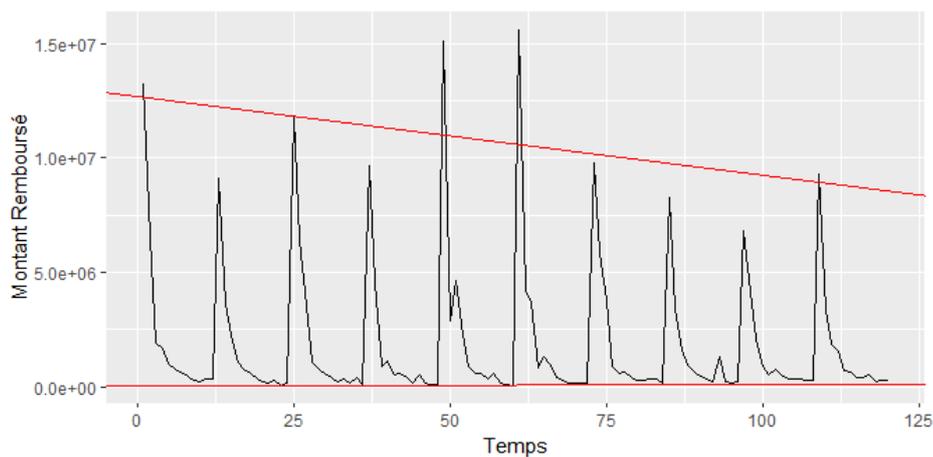


Figure 20 : Graphique des prestations santé KP par survénance sur l'année comptable N+1

La 2nde méthode sera tout de même appliquée afin de confirmer nos hypothèses.

❖ Méthode analytique : Buys-Ballot

La droite des moindres carrés est obtenue avec : $\sigma = 2.13\bar{x} - 837865.4$

Avec une erreur de 0.1, il semble évident que le coefficient alpha est bien différent de 0, ce qui confirme notre précédente hypothèse.

Lors de l'estimation de notre modèle pour ces données, il suffira d'appliquer le logarithme à notre série de données afin de nous ramener à un modèle additif.

Cette étude a donc permis de mettre en évidence le fait que les prestations, remboursées la première année, forment un modèle additif tandis que celles liées à la seconde année peuvent être décomposées comme un modèle multiplicatif.

Cette étape faite, nous pouvons aborder le sujet de la stationnarité et de la saisonnalité de nos modèles.

1.2. Sélection du modèle

✚ Stationnarité et saisonnalité

Graphiquement, on observe une saisonnalité confirmée lorsque l'on procède à une décomposition de notre série :

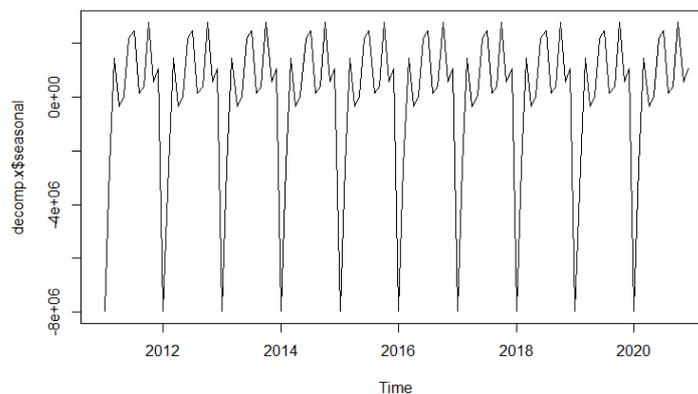


Figure 21 : Saisonnalité de la série

La saisonnalité de douze mois est cohérente avec la construction de notre base. En effet, les données présentées ici correspondent à la première année comptable d'une survenance N, à savoir Année Survenance = Année Comptable. Il est donc normal d'observer un phénomène qui se répète tous les 12 mois.

L'opérateur $\nabla = 1 - L^{12}$ est donc appliqué à notre jeu de données.

Une fois la saisonnalité retirée, des tests de stationnarité sont effectués afin de confirmer ou infirmer notre hypothèse de stationnarité basée sur les données graphiques:

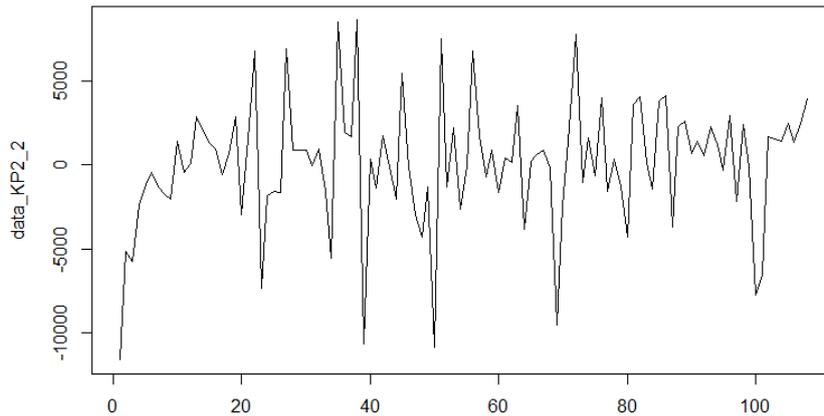


Figure 22 : Série désaisonnalisée

Les différents tests conduisent aux résultats présentés dans ce tableau :

Test	Valeur	P-Value
Dickey-Fuller	-5,267	<0,01
Phillips-Perron	-101,87	<0,01
KPSS	0,27	>0,1

Figure 23 : Résultats des tests de stationnarité

Notre hypothèse initiale est donc validée par ces résultats. Les données sont non-stationnaires. Dickey-Fuller et Phillips-Perron étant des tests de non-stationnarité, les p-value inférieures à 0.01 permettent de rejeter l'hypothèse (H_0 : la série est non stationnaire) et inversement pour le test KPSS.

Ces différents tests permettent de retenir le modèle SARIMA présentant une saisonnalité de 12 mois ne nécessitant pas une différenciation pour atteindre la stationnarité.

Une fois la saisonnalité retirée et la stationnarité validée, les paramètres seront sélectionnés dans la prochaine partie.

Sélection des paramètres

L'estimation du modèle débute par l'étude graphique des autocorrélogrammes :

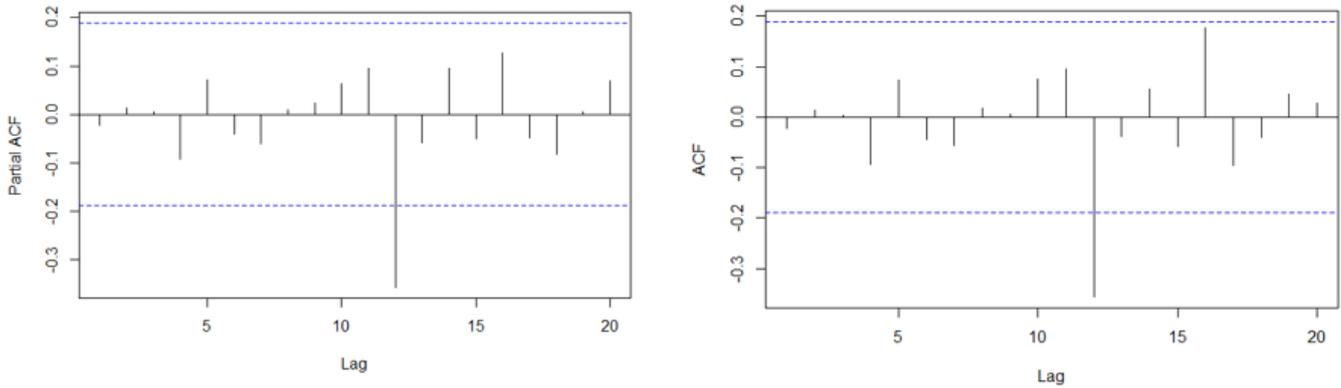


Figure 24 : ACF et PACF de la série désaisonnalisée et stationnaire

A première vue, le modèle semble se diriger vers un SARIMA avec un AR(12) et un MA(12). On peut clairement observer que les fonctions d'auto-corrélation et d'auto-corrélation partielle sont supérieures à 0 lorsque $j=12$. Le nombre de paramètres étant particulièrement élevé, l'objectif est de réduire l'ordre. Pour cela, on peut tester la valeur des coefficients du modèle, en particulier les derniers coefficients, grâce au « t-test » dans le but de vérifier l'hypothèse (H_0) : « le coefficient a_k et/ou b_k est nul ». Lorsque cette hypothèse est confirmée on réduit itérativement l'ordre jusqu'à obtenir une p-value supérieure à 5% pour les derniers paramètres : a_p et b_q .

❖ **Modèle 1 initial : ARMA(8,8)**

Paramètres	a_{12}	b_{12}
P-value	0,25	0,12

Figure 25 : P-value des coefficients maximaux de la série

Au vu des résultats, on rejette ce modèle puisque les p-values, des derniers paramètres, sont nettement supérieures au seuil de 5%.

Après la réduction du modèle, on retient un ARMA(8,8) pour lequel les p-values des paramètres des ordres maximaux sont inférieures à 5%.

Paramètres	a_8	b_8
P-value	<0,001	<0,001

Figure 26 : P-value des coefficients maximaux de la série

Précédemment, le modèle a été obtenu par observation des graphiques ACF et PACF.

La 2e méthode pour l'obtention d'un modèle est la sélection par le critère de l'AIC expliqué précédemment. L'objectif est de minimiser ce dernier afin de conserver le modèle qui allie le mieux pouvoir prédictif et faible nombre de paramètre. Il s'agit ici d'un ARMA(0,1) :

❖ **Modèle 2 : ARMA(0,1)**

Dans un premier temps, le modèle ARMA(8,8) semble être plus approprié puisqu'il minimise le critère AIC, avec 2089 contre 2319 pour le second : l'ARMA(0,1).

La comparaison en termes de résidus permettra de confirmer ce point.

L'idéal serait d'obtenir des résidus non corrélés et gaussiens autrement dit des bruits blancs gaussiens ; cela témoignera d'un bon ajustement à nos données.

Comme présenté précédemment, la non-corrélation des résidus sera vérifiée par le test de Box-Pierce tandis que la normalité par le test de Lilliefors.

Les résultats rejoignent ceux constatés avec le critère AIC puisque les résidus du premier modèle sont effectivement des bruits blancs gaussien : les deux tests sont vérifiés. A l'inverse pour le second modèle, les résidus sont bien corrélés mais l'hypothèse de normalité n'étant pas vérifiée, le premier modèle ARMA(8,8) est privilégié.

Modèles	Box-Pierce	Lilliefors
SARIMA(8,0,8)(1,1,1)[12]	0,9	0,12
SARIMA(0,0,1)(0,1,1)[12]	0,9	0,04

Figure 27 : Résultats des tests Box-Pierce et Lilliefors

Toutefois, nous appliquerons les deux modèles afin de vérifier que notre choix s'est porté sur le meilleur modèle. Nous pourrons ainsi comparer les résultats pour voir si notre hypothèse se confirme, le meilleur modèle sera celui qui minimisera l'erreur de prédiction. Cette erreur sera analysée mensuellement mais également annuellement. Nous allons donc comparer avec les données réelles, les cadences de remboursement par mois mais également les ratios de S/P obtenus par survenance, il sera donc nécessaire d'estimer deux années comptables pour obtenir une survenance complète.

Les prochaines sections seront consacrées à nos applications numériques. Pour plus de clarté, les prédictions concerneront dans un premier temps les prestations liées à la première année de remboursement de la survenance 2021, et ensuite celles remboursées la seconde année.

1.3. Mise en application

Année comptable N pour des remboursements de santé survenus l'année N

Les résultats de 2021, corroborent notre précédente intuition, le 2e modèle n'est pas le plus adapté :

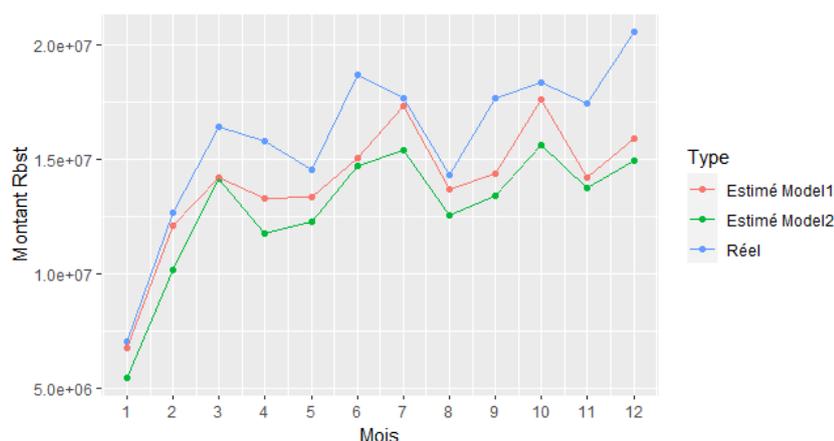


Figure 28 : Comparaison des prédictions de l'année 2021 pour la survenance 2021 entre modèle 1 – $SARIMA(8,0,8)(1,1,1)[12]$, le modèle 2 – $SARIMA(0,0,1)(0,1,1)[12]$ et les prestations réelles

En vert, il sous-estime la charge de sinistralité de l'année 2021. Cependant, le même constat est réalisé sur le 1e modèle, même si l'estimation est légèrement meilleure.

Dans les deux cas, nos résidus sont tout de même des bruits blancs gaussiens puisque nos tests de Box-Pierce et Lilliefors sont vérifiés.

En comparant avec le réel, des taux d'erreur de 19% et de 12% sont constatés pour respectivement le 2e et 1e modèle, ce qui correspond à un écart de 36 millions d'euros et 23 millions d'euros. Même si le modèle avec le plus de paramètres est meilleur, on ne peut pas accepter un écart aussi important avec la réalité.

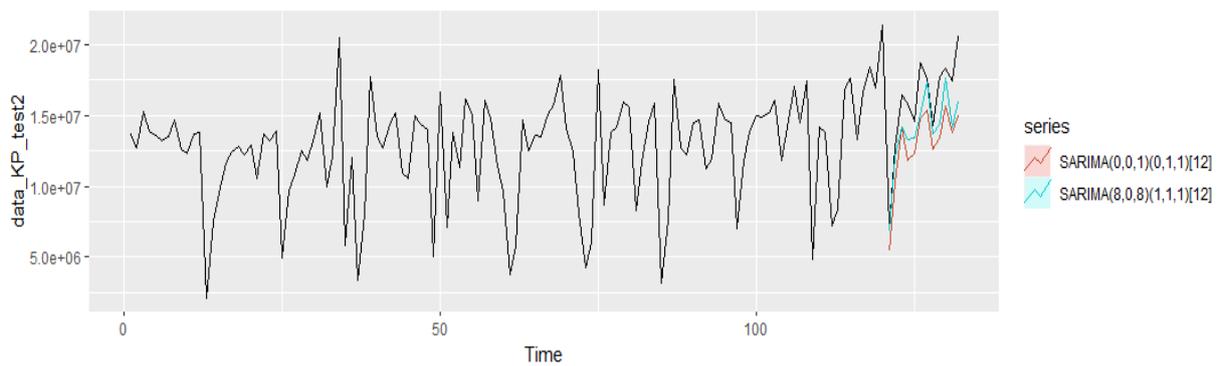


Figure 29 : Prestations santé estimées (selon 2 modèles) en survenance 2021 et réelles entre les survenances 2011 et 2021 pour l'année comptable N

L'année 2021 intervient après l'année Covid qui a marquée 2020, notamment sur les cinq premiers mois. En effet une sous-consommation est nettement constatée sur ce graphique entre février et mai 2020, notamment en avril et mai.

La COVID-19 a engendré un premier confinement drastique entre le 17 mars et le 11 mai 2020 ce qui a naturellement entraîné une nette diminution des soins de santé habituels. Cette année étant atypique il serait normal de penser qu'elle ne se reproduira pas et que nous pouvons alors la retraiter afin d'améliorer notre prédiction. Nous décidons ainsi d'appliquer un retraitement nécessaire pour notre étude, qui engendrera une meilleure cohérence de nos résultats et évitera ainsi une mauvaise estimation.

Avant d'appliquer un quelconque retraitement, nous allons terminer notre application en prévisionnant la 2e année comptable.

Année comptable N+1 pour des frais de santé survenus l'année N

Il est inutile de réexpliquer les étapes permettant d'obtenir notre modèle pour l'année N+1 (cf. Annexe A). La seule différence avec le dernier modèle, c'est qu'il nous a fallu appliquer un logarithme à nos données lors de la sélection de notre modèle puisque comme expliqué précédemment, ce dernier est multiplicatif.

Nous obtenons deux modèles qui pourraient correspondre à nos données. En effet, lors de l'estimation, les résidus forment un bruit blanc dans les deux cas :

- Modèle 1 obtenu par minimisation de l'AIC : SARIMA(1,0,0)(1,1,0)[12]
- Modèle 2 obtenu par observation des graphiques des autocorrélogrammes puis en testant la significativité des coefficients : SARIMA(12,0,12)(1,1,1)[12]

Le backtesting sur l'année 2021 permettra de choisir le meilleur modèle pour nos prédictions futures.

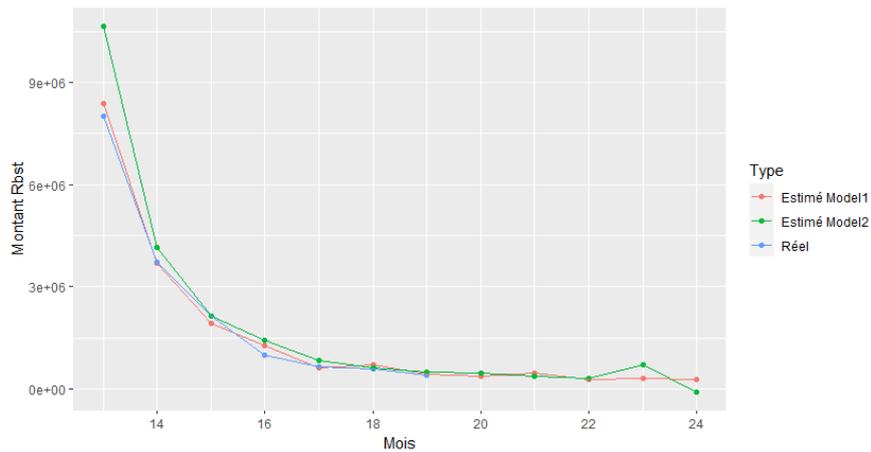


Figure 30 : Comparaison des prédictions de l'année 2022 pour la survenance 2021 entre modèle 1 – SARIMA(1,0,0)(1,1,0)[12] , le modèle 2 – SARIMA(12,0,12)(1,1,1)[12] et les prestations réelles

Les données 2021 s'arrêtant au 19e mois, la comparaison est faite sur ces premiers remboursements. L'absence des derniers n'a aucune incidence étant donné le faible volume de prestations réglées en fin d'année.

Dans ce laps de temps, le modèle SARIMA ayant le moins de paramètres est le plus adapté, tandis que le 2nd modèle surestime la charge de sinistralité.

Si on regarde les données dans leur ensemble, la courbe bleue est fortement impactée par la hausse de 2020. Elle est sans doute la conséquence de la sous-consommation observée précédemment sur les remboursements pendant l'année Covid. Un effet de « décalage » dans le temps a été constaté, les soins annulés en 2020 ont été reportés et/ou réglés en 2021.

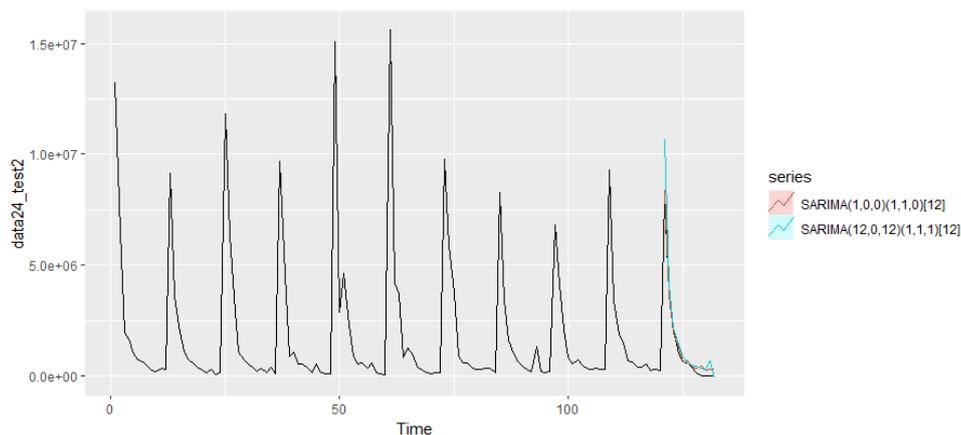


Figure 31 : Prestations santé estimées (selon 2 modèles) en survenance 2021 et réelles entre les survenances 2011 et 2021 pour l'année comptable N+1

Contrairement à nos premières données l'effet Covid n'est pas clairement visible sur l'année comptable N+1. Sur les précédentes analyses, la sous-consommation ne pouvait s'expliquer que



par la mise en place d'un confinement. Ici, on peut observer une hausse des remboursements qui pourrait traduire un décalage dans le temps des remboursements. Cependant, étant donné nos données, il nous est difficilement possible de conclure à cela sans émettre aucune réserve. C'est pourquoi, un retraitement ne sera pas réalisé d'autant plus que cette hausse ne semble pas impacter nos résultats.

En conclusion de cette partie, le modèle retenu est un SARIMA(1,0,0)(1,1,0)[12]. Celui-ci obtient un AIC inférieur au 2nd modèle, sa prédiction 2021 est ainsi nettement meilleure avec une marge d'erreur à 2,8%.

Maintenant que la survenance 2021 a été entièrement prédite, nous pouvons expliciter par la suite deux méthodes de retraitements qui viendront corriger l'effet covid de nos prédictions.

1.4. Retraitement Covid

1.4.1. Impact du covid sur les institutions de prévoyance

L'arrivée de la Covid en 2020 a impacté l'ensemble des Institutions de Prévoyance qui ont dû s'adapter pour offrir les meilleures garanties à leurs assurés et de nombreuses initiatives prises en ce sens.

En effet, de nombreux secteurs ont particulièrement été touchés par la crise notamment la branche Hôtel-Café-Restaurant qui a vu son activité fortement diminuée en 2020 et 2021. Cet arrêt brutal de l'activité a engendré la mise en place d'une activité partielle pour certains salariés, voire un chômage partiel pour d'autres. Ce dernier a eu pour effet une baisse du chiffre d'affaires annuel des IP, de l'ordre de 500 M€, puisque de nombreuses entreprises y ont eu recours avec plus de 13 millions de salariés au chômage partiel au plus fort de la crise.

Au chômage partiel s'ajoutent d'éventuelles exonérations réalisées sur un trimestre, comme ce fut le cas pour la branche Hôtel Café Restaurant qui a été exonérée d'un trimestre de cotisations en 2020.

A cela s'ajoute de potentiels impayés de la part de certaines entreprises, il en ressort une baisse de 3,4% des cotisations en santé et prévoyance en 2020.

Les salariés en chômage partiel ou dont l'activité a été fortement réduite ont bénéficié d'un maintien de leurs droits à titre onéreux.

Du côté des prestations, Klesia Prévoyance a vu une nette diminution de ces remboursements santé. En effet, l'interdiction de tout déplacement lors du premier confinement a directement impacté la charge de sinistralité de cette période. On chiffre cette baisse mensuelle à 50% sur ces deux mois.

Pendant ces mois-ci, la Covid a monopolisé l'ensemble des personnels de santé ce qui a naturellement entraîné un report de l'ensemble des opérations et/ou des rendez-vous médicaux. Cependant une fois la vague de covid passée, l'ensemble des soins annulés ont dû être réalisés ce qui a engendré un rattrapage des dépenses et donc des remboursements par Klesia Prévoyance. Sur les 6e mois de 2020, une baisse de 16% est observable tandis qu'une augmentation de 15% sur la fin de l'année permet une compensation de l'effet Covid. Cette hausse est notamment visible graphiquement sur la première année comptable 2020 (figure 31).

Nous allons présenter dans ce qui suit les deux méthodes de retraitements des données Covid.

1.4.2. Retraitement avec utilisation de l'historique

Pour cette méthode, le coefficient de passage d'un mois à l'autre est calculé par année :

$$\text{coefficient passage}_i^n = \frac{x_i^n}{x_{i+1}^n}$$

Avec x_i^n , les prestations remboursées le mois i de l'année comptable n pour l'année de survenance n .

Année survenance	2019	2018	2017
Prestations janvier	6 980 701	3 143 601	4 202 295
Coef. Passage février	1,677	2,412	1,423
Coef. Passage mars	1,185	2,318	3,05
Coef. Passage avril	1,080	0,722	0,476
Coef. Passage mai	0,995	0,965	1,592
Coef. Passage juin	1,020	1,183	1,025
Coef. Passage juillet	1,057	1,010	1,128
Coef. Passage aout	0,737	0,770	0,975
Coef. Passage septembre	1,2	1,053	0,532
Coef. Passage octobre	1,2	1,332	1,414
Coef. Passage novembre	0,849	0,931	1,247
Coef. Passage décembre	1,203	0,981	1,088

Figure 32 : coefficients de passage des survenances 2017 à 2019

On constate une déformation des règlements santé (entre 2018 et 2019) à la suite d'un changement dans notre gestion interne. A partir de 2019, un nouveau gestionnaire a repris la gestion des prestations santé de certains portefeuilles. Cela se traduit par une augmentation des prestations réglées en janvier 2019 par rapport aux précédentes années. En effet, les prestations sont désormais réglées plus rapidement. Ce constat est visible sur le graphique suivant.

Pour estimer janvier 2020, nous décidons donc de ne retenir que janvier 2019 compte tenu des constatations précédentes.

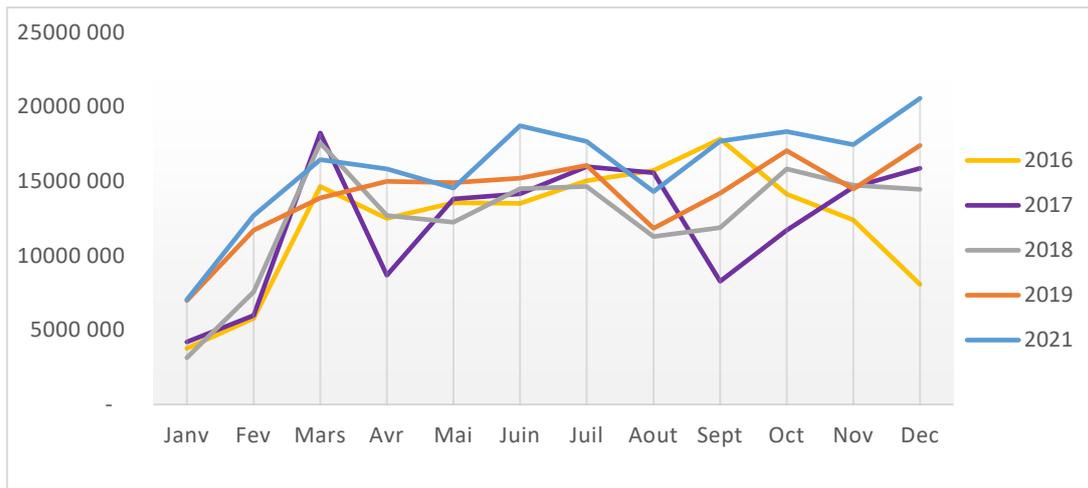


Figure 33 : Remboursements de janvier à décembre pour l'année de survenance et comptable N

Malgré une hausse à partir de juin 2020 (Figure 7 et 29), le retraitement ne sera réalisé qu'entre janvier et mai, correspondant au début de la pandémie et au premier confinement. En effet, la hausse en fin d'année perdue en 2021. Il pourrait s'agir ici d'un effet lié à la réforme 100% santé et non pas à la pandémie. Cette partie ne sera donc pas retraitée.

Ici, nous optons pour appliquer la méthode utilisant les coefficients de passage d'un mois à l'autre selon différents paramètres. Il s'agit de prendre la moyenne de ces derniers sur plusieurs années.

En résumé, trois cas seront étudiés. Pour commencer seul les coefficients de passage de l'année 2019 seront appliqués itérativement aux prestations de 2020, en commençant par le mois de janvier obtenu par moyenne de 2021 et 2019. Lors du 2nd cas, nous prendrons la moyenne des coefficients de 2019 et 2018, puis nous étendrons cette moyenne jusqu'à l'année 2017.

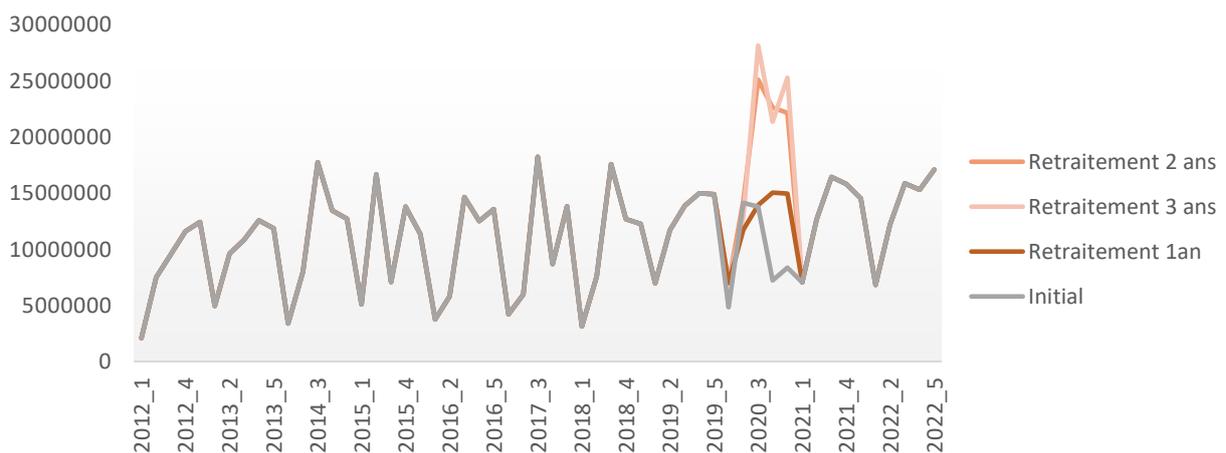


Figure 34 : Comparaison des retraitements réalisés avec une profondeur de 1 an à 3 ans.

Ce graphique illustre les trois cas de retraitement. Dans la figure 32, on constate que les cadences de règlement ont subi une évolution à partir de 2019. Une stabilité est constatée dans les montants remboursés d'un mois à l'autre sur la figure 33. Par exemple, entre février et mars, une hausse importante était systématiquement observable avant 2019, ce qui se répercute donc dans les

coefficients. Ainsi, 2019 apparaît comme l'année la plus semblable aux nouvelles survenances. Comme expliqué précédemment, cela est dû à un changement de gestionnaire chez Klesia.

Cela implique que les cadences de règlements antérieures à cette année augmente considérablement la charge de sinistralité. C'est pourquoi via ce graphique, une hausse significative est constatée pour les retraitements ayant un historique supérieur à une année. Le choix d'appliquer les coefficients de cadences de l'année 2019 est donc tout à fait justifié ici.

Cette méthode de retraitement semble donner des résultats concluants néanmoins seul une nouvelle prédiction à partir de cette nouvelle base de données retraitée pourra attester de cela.

Nous allons désormais retraiter nos données non pas à partir de l'historique cette fois ci mais en retirant la composante résiduelle de notre base pour une période donnée.

1.4.3. Retraitement des résidus

La deuxième méthode de retraitement consiste à se focaliser sur la partie résiduelle de la série temporelle. Notre série est structurée de la manière suivante :

$$Y_t = S_t + T_t + \varepsilon_t \text{ avec } \varepsilon_t, \text{ la partie résiduelle de nos données.}$$

La partie résiduelle modélise des événements accidentels et engendre des fluctuations irrégulières. La pandémie constitue une circonstance imprévisible captée par cette composante des séries temporelles. Pour atténuer l'effet de la Covid-19, nous allons retirer de nos données les résidus des six premiers mois de 2020. Comme expliqué précédemment, il s'agit de la période la plus impactée par la pandémie.

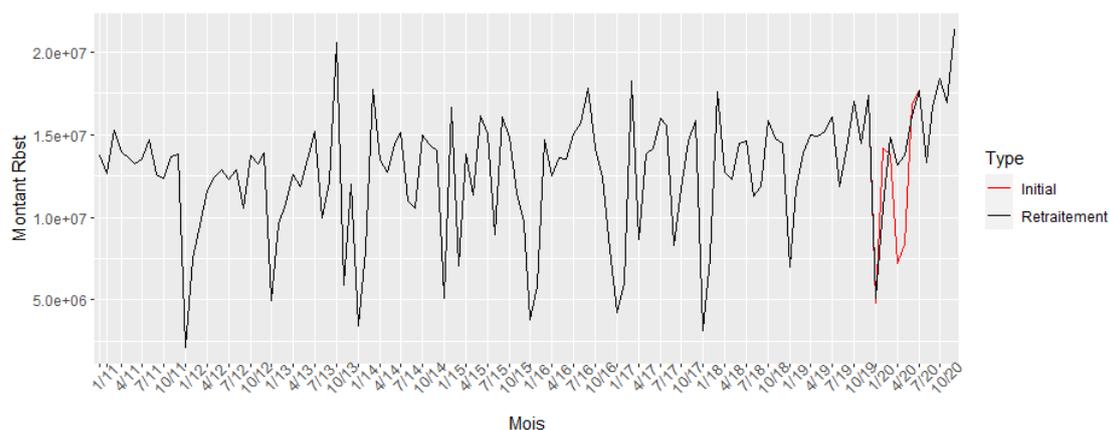


Figure 35 : Résultat graphique du retraitement des résidus

Le graphique présente le retraitement réalisé grâce à cette méthode : le delta entre la courbe rouge et la courbe noire représente les résidus. La sous-consommation a donc pu être corrigée ce qui améliorera nos futurs résultats.

1.4.4. Mise en application

Les résultats sont présentés graphiquement :

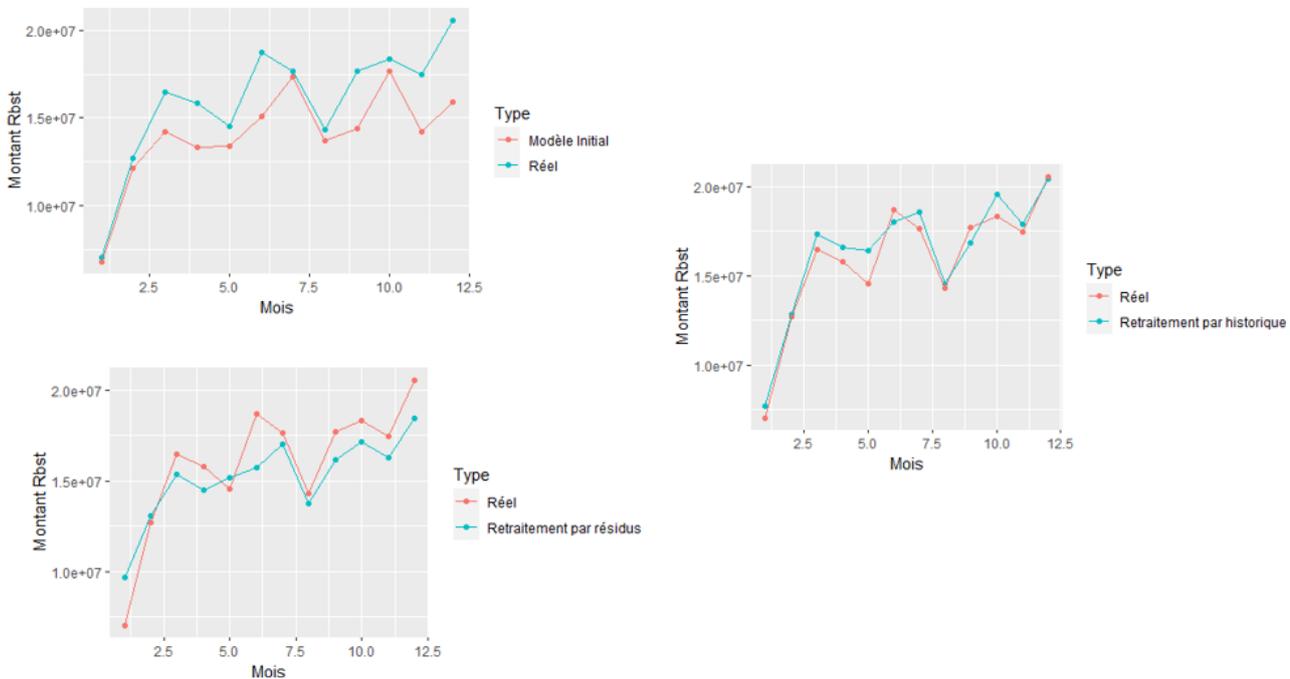


Figure 36 : résultat graphique du modèle initial et des deux méthodes de retraitement

Sur ces graphiques sont présentés les résultats des trois modèles, le modèle initial à savoir celui retenu avant les différents retraitements (cf. partie 3.1.4). Les deux autres modèles sont les résultats des deux retraitements : utilisation de l'historique via les coefficients de cadences et retrait de la composante résiduelle.

Une fois les retraitements réalisés sur 2020, un nouveau calibrage du modèle est réalisé. Toutes les étapes explicitées précédemment ont été réalisées à savoir : suppression de la saisonnalité, de la tendance, estimation des paramètres et sélection du meilleur modèle.

Il est ainsi obtenu deux nouveaux modèles :

- un SARIMA(2,1,1)(2,1,0)[12] après retraitement via l'historique
- un SARIMA(3,1,2)(2,0,0)[12] après retraitement des résidus

Graphiquement, ces résultats semblent meilleurs ; en effet, avant retraitement, l'écart au réel était plus important que nos estimations obtenues après retraitement. Cependant, la première méthode (utilisation des cadences) sera privilégiée à la seconde (retraitement des résidus) puisqu'elle permet d'obtenir des résultats plus concluants.

Dans ce tableau, les erreurs de prévisions sont résumées par mois :

Mois	Modèle SARIMA Initial	Modèle SARIMA avec retraitement Cadence	Modèle SARIMA avec retraitement Résidu
1	-3%	9,5%	37%
2	-5%	1%	3%
3	-14%	5,6%	-7%
4	-16%	4,9%	-8%
5	-8%	12,7%	4%
6	-19%	-3,9%	-16%
7	-2%	5,3%	-4%
8	-4%	2%	-4%
9	-19%	-4,8%	-9%
10	-4%	6,5%	-7%
11	-19%	2,7%	-7%
12	-23%	-0,5%	-10%
GLOBAL	-12,1%	2,9%	-4,7%

Figure 37 : Taux d'erreur sur les prévisions mensuelles entre le modèle SARIMA sans retraitement et avec retraitement

Le modèle initial n'est pas retenu ; à chaque mois, les prévisions sont éloignées de la réalité et parfois nettement inférieures. De ce fait, le montant global de prestations prévisionnelles engendre un taux d'erreur négatif de 12,1%. Cette valeur est trop importante pour accepter ces résultats.

Le 3e modèle, après le retraitement des résidus, améliore les résultats au global. Le taux d'erreur est presque divisé par trois par le simple fait de retirer la composante résiduelle sur les six premiers mois de l'année 2020. Cependant, en comparant avec le 2e modèle, le constat au global nous conduit à rejeter le modèle par comparaison des écarts en valeur. De plus, ces écarts mensuels sont plus faibles chaque mois, excepté pour le mois de mai. Le retraitement via l'utilisation de coefficients de passage permet de réduire nettement le taux d'erreur global, puisque celui-ci est à moins de 3%. Ce modèle n'est pourtant pas parfait mensuellement. Il s'agit tout de même du meilleur modèle au regard des deux autres. De plus, grâce à ce dernier, il est possible d'avoir une légère marge de prudence. La compensation au fur et à mesure des mois permet d'obtenir, au global, des résultats acceptables, ce qui sera très efficace pour le calcul du ratio S/P (Sinistres/Primes) lors du plan prévisionnel d'activité. Le taux d'erreur global est donc à mettre en perspective, puisque les résultats ne sont pas parfaits.

En milieu d'année, l'institution de prévoyance réalise un PPA (Plan Prévisionnel d'Activité) pour l'année à venir. Pour cela, il est nécessaire d'actualiser les ratios « S/P » (Sinistres/Primes) de l'année en cours pour l'année suivante. Actuellement, les S/P de l'année N sont repris avec l'application ou non d'une hypothèse de dérive de la santé.

L'objectif est d'améliorer cette méthode afin d'obtenir des « S/P » beaucoup plus proches de la réalité. C'est pourquoi, les prestations santé ont été estimées sur deux années puisque la loi indique que les assurés disposent d'un délai maximal de 2 ans pour se faire rembourser.

Avant la comparaison avec le plan prévisionnel d'activité (PPA), nous allons appliquer le modèle de Holt-Winters.

2. Lissage Holt-Winters

2.1. Mise en application

Comme expliqué précédemment, le modèle de Holt-Winters sera appliqué dans cette partie puisqu'il s'agit du plus adapté lorsque les données présentent une saisonnalité ainsi qu'une tendance.

Tout comme pour les séries temporelles, le modèle nécessite de connaître au préalable la nature de la série : additive ou multiplicative. Ici, nous savons d'après les études réalisées dans la partie 3.1.1 que notre série vérifie ces deux cas puisque la première année comptable est additive tandis que la seconde est multiplicative.

Les étapes seront similaires à celles des dernières parties : prévision 2021 sans retraitement, puis avec retraitement pour la première année comptable, et enfin, on réalise le prévisionnel comptable 2022 et la survenance 2021 sans retraitement.

Après le premier lancement du modèle sous R, relatif au prévisionnel comptable 2021 sans retraitement, il est obtenu le modèle suivant pour la 1^e année comptable :

$$\hat{Y}_{t+12} = 14\,235\,501 + 12 * 95\,137 + s_t$$

Avec les paramètres de lissage suivants :

- $\alpha = 0.076$
- $\beta = 0.088$
- $\gamma = 0.179$

On peut déjà constater dans un premier temps que la mise en œuvre du modèle de lissage exponentiel est beaucoup plus simple que celle des séries temporelles. Là où, dans ce dernier, de nombreuses phases sont requises pour déterminer le modèle, dans celui-ci, la prévision est quasi-immédiate. Il s'agit d'une donnée à prendre en compte après la présentation des résultats.

Mois	Modèle Lissage Initial	Modèle Lissage avec retraitement sur les cadences
1	-8%	15%
2	-4%	3%
3	-3%	5%
4	-11%	7%
5	3%	20%
6	-7%	1%
7	1%	9%
8	9%	20%
9	-7%	2%
10	0%	9%
11	-2%	8%
12	-12%	-3%
GLOBAL	-4%	7%

Figure 38 : Taux d'erreur sur les prévisions mensuelles entre le modèle Holt-Winters sans retraitement et avec retraitement

Le premier modèle est biaisé en raison de la présence de l'effet Covid. Une prévision repose principalement sur un historique stable, ce qui n'est pas le cas compte tenu des effets de la pandémie. Se baser sur un historique biaisé ne permettra pas d'obtenir un modèle cohérent.

C'est pourquoi le modèle avec un taux d'erreur de 7% est retenu. Nous comparerons plus tard les résultats obtenus précédemment avec le modèle des séries temporelles.

Contrairement au modèle des séries temporelles, le pouvoir prédictif de celui lié à l'année comptable 2022 est faible. Avec une erreur de 52%, l'estimation des remboursements santé est nettement supérieure à la réalité.

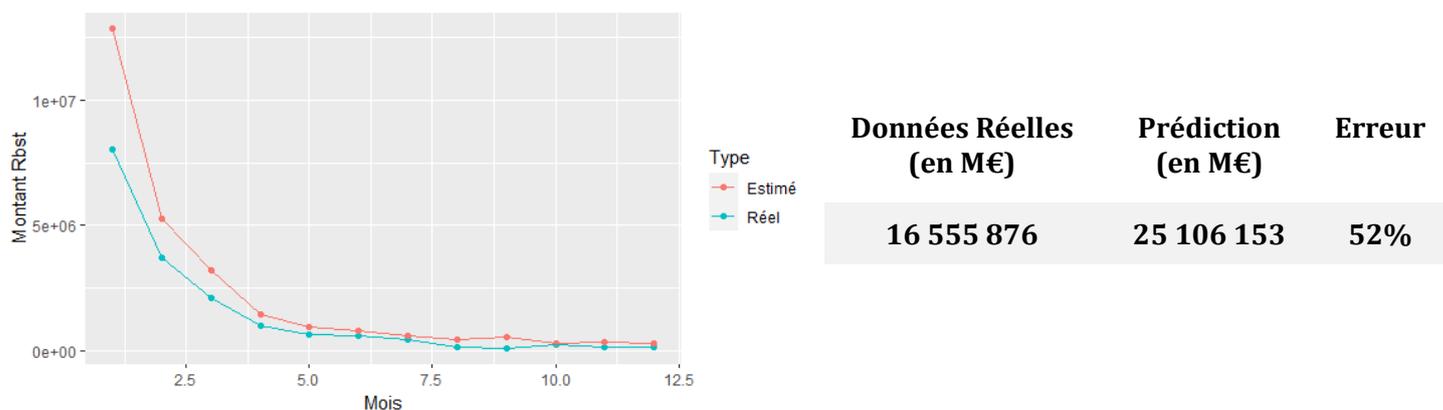


Figure 39 : Résultats du prévisionnel comptable 2022 et survenance 2021

Cette erreur provient principalement des trois premiers mois. Le modèle surestime la consommation de plus de 7 millions d'euros. Comme nous l'avons constaté dans les parties précédentes et sur le graphique suivant, la hausse de l'année 2020 a impacté notre prévisionnel. Le modèle a estimé que cette augmentation se prolongerait sur l'année 2021, ce qui n'a pas été le cas ici.

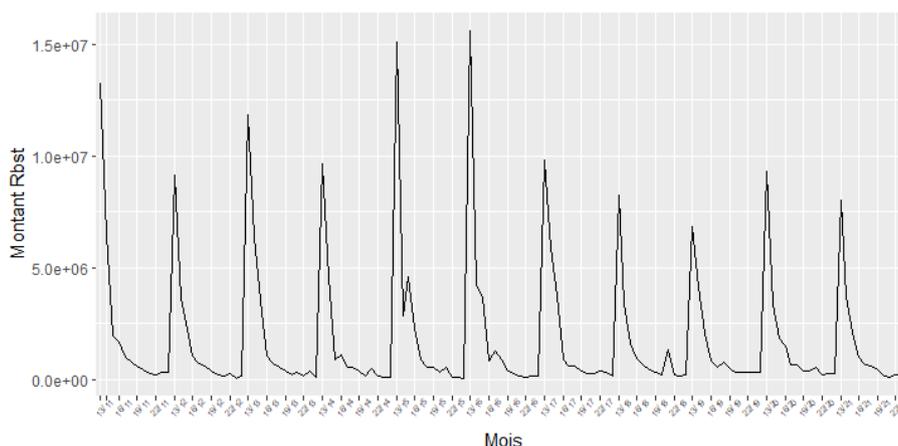


Figure 40 : Remboursements des soins santé en 2022 survenus en 2021

Contrairement à l'année comptable 2021, un retraitement sur 2022 n'est pas idéal. Sur la figure 40, la hausse est stoppée en 2020, mais il existe tout de même une évolution en 2021 en comparaison des années antérieures. De plus, sur l'année comptable 2021, l'impact du Covid est clairement visible, ce qui n'est pas le cas ici. Nos données ne nous permettent pas d'obtenir des résultats approfondis pour détecter la présence de l'impact Covid, qui se traduirait par un rattrapage sur les postes de soins sous-consommés l'année passée.

D'après tous ces résultats, il est préférable de retenir le modèle SARIMA.

Nous pouvons désormais comparer nos résultats avec ceux du plan prévisionnel d'activité de l'année précédente. Cela permettra à la fois de contrôler si nos modèles sont plus performants que la méthode actuelle du PPA, mais également de comparer nos résultats entre eux.

Si les conclusions sont positives, le modèle sera définitivement validé.

3. Comparaison avec le Plan Prévisionnel d'Activité

En santé, le Comité des Risques et de Solvabilité (CRS) fixe le niveau de dérive attendu pour l'année à venir. Cette dérive s'élève à 2% en santé pour l'année 2021. Lors de l'estimation du budget, cette dérive est donc appliquée aux S/P de l'année précédente. Cependant, cette méthode d'estimation des indicateurs se fait à une maille d'analyse plus précise. C'est pourquoi, dans la prochaine partie, nous allons réaliser des prédictions sur certains des plus gros portefeuilles qui composent l'entité.

Pour commencer, nous allons comparer nos résultats avec les données observées.



Modèle	Chiffre d'affaires	Charge Sinistralité théorique	Charge Sinistralité Réelle	S/P prévisionnel	S/P Réel	Erreur Prévision (en M€)
SARIMA après retraitement	219 540 842	208 574 927	207 898 825	95,01%	94,7%	0,7
SARIMA sans retraitement	219 540 842	185 154 360	207 898 825	84,34%	94,7%	-22,7
Holt-Winters après retraitement	219 540 842	230 154 040	207 898 825	104,8%	94,7%	22,3
Holt-Winters avant retraitement	219 540 842	209 551 886	207 898 825	95,45%	94,7%	1,6

Figure 41 : Résultats numériques du CA, de la charge de sinistralité et des SP obtenus avec les modèles SARIMA et Holt-Winters

Le retraitement de l'année 2020 améliore significativement nos modèles de série temporelle. L'écart entre le S/P réel et celui estimé par notre modèle est réduit, évoluant à la hausse par rapport à celui obtenu avec les données brutes. Avec un écart inférieur à 1 million d'euros, la charge de sinistralité estimée surestime nos données, ce qui permet une légère prudence.

Pour le modèle Holt-Winters, l'efficacité du retraitement est modérée. Comme constaté et retenu dans la dernière partie, il y a une surestimation lors de la prédiction pour l'année comptable 2021. Cependant, étant donné la hausse de l'année comptable 2022, la charge de sinistralité finale est nettement éloignée de la réalité, avec un écart de 22,3 millions d'euros. Par ailleurs, le S/P pour ce modèle est supérieur à 100%, ce qui pourrait indiquer, dans ce cas, que l'entité Klesia Prévoyance ne serait pas rentable. Cependant la réalité est différente. Une prudence est acceptable, tant qu'elle reste proche de la réalité et ne conduit pas à considérer un portefeuille comme non rentable lorsque ce n'est pas le cas. C'est pourquoi le second modèle Holt-Winters (avant retraitement) est favorisé.

Dans la suite, et pour plus de clarté, le modèle 1 sera le modèle SARIMA après retraitement et le modèle 2 sera le Holt-Winters avant retraitement.

La question que l'on doit maintenant se poser est le niveau de prudence que l'on souhaite. Dans le modèle 1, la prudence est moins élevée que dans le second modèle. D'autre part, un deuxième paramètre à prendre en compte est la rapidité d'exécution de ce dernier. De plus, le modèle de série temporelle est plus « adaptable », alors que le modèle de lissage exponentiel beaucoup plus une « boîte noire ». Pour toutes ces raisons, le modèle 1 est retenu.

Année survenance	Chiffre d'affaires	Charge Sinistralité théorique	Charge Sinistralité Réelle	S/P théorique	S/P Réel	Erreur Prévision (en M€)
2021 - Budget	219 540 842	205 709 769	207 898 825	93,7%	94,7%	-2,2

Figure 42 : Résultats numériques du CA, de la charge de sinistralité et des SP obtenus au budget

Les résultats du budget mettent en évidence une sous-estimation de la consommation de l'ordre de 2 millions d'euros, tandis que les résultats de notre modèle (modèle 1) surestiment la consommation, offrant ainsi une certaine marge de prudence. Les résultats obtenus dans ce mémoire sont plus prudents que ceux du plan prévisionnel d'activité.

D'autre part, les résultats développés ici pourraient renforcer la méthode actuellement utilisée chez Klesia. Avec cette dernière, seules des hypothèses de dérivabilité sont prises en compte, alors qu'avec la méthode développée dans ce mémoire, nous pouvons combiner à la fois une prise en compte de l'historique de l'entité et ajouter des hypothèses de dérives si les prédictions semblent sous-estimer le futur.

Il est maintenant possible d'affirmer que nos résultats ont un sens et pourraient permettre d'améliorer les estimations futures réalisées chez Klesia.

Désormais et afin d'affiner nos conclusions, un modèle sera réalisé sur les portefeuilles composant l'entité Klesia Prévoyance : IMMO, PHIE, PHAC. De plus, une cinquième analyse sera réalisée sur un regroupement de deux portefeuilles « Branches » composé de CNVS (vins et spiritueux) et FLEU (Fleuriste).

4. Application du modèle aux branches de Klesia Prévoyance

4.1. Etude des portefeuilles

Etude des portefeuilles pharmaceutiques

Comme expliqué précédemment, l'entité Klesia Prévoyance couvre de nombreuses branches professionnelles telles que les métiers de la pharmacie. Cette branche est représentée par deux portefeuilles : PHIE et PHAC. Le premier concerne les salariés non-cadres, tandis que le second est réservé aux pharmaciens cadres.

Ces portefeuilles, étant particuliers en santé, puisque déficitaires sur l'historique, une étude approfondie ainsi qu'une projection sont réalisées sur ces segments.

Avant 2019, le ratio Sinistres à Primes était supérieur à 100%, témoignant de résultats déficitaires. A partir de 2020, un phénomène d'inversion se produit avec des cotisations en hausse, résultant d'indexation et d'affaires nouvelles venant améliorer ces ratios. En 2022, le plan prévisionnel d'activité prévoit une indexation de 1,5%, ainsi que des affaires nouvelles venant augmenter nos cotisations à hauteur d'un million d'euros. A cela s'ajoutent les résiliations, pour finalement avoir une augmentation de 1,3 millions d'euros entre les cotisations de 2022 et 2021.

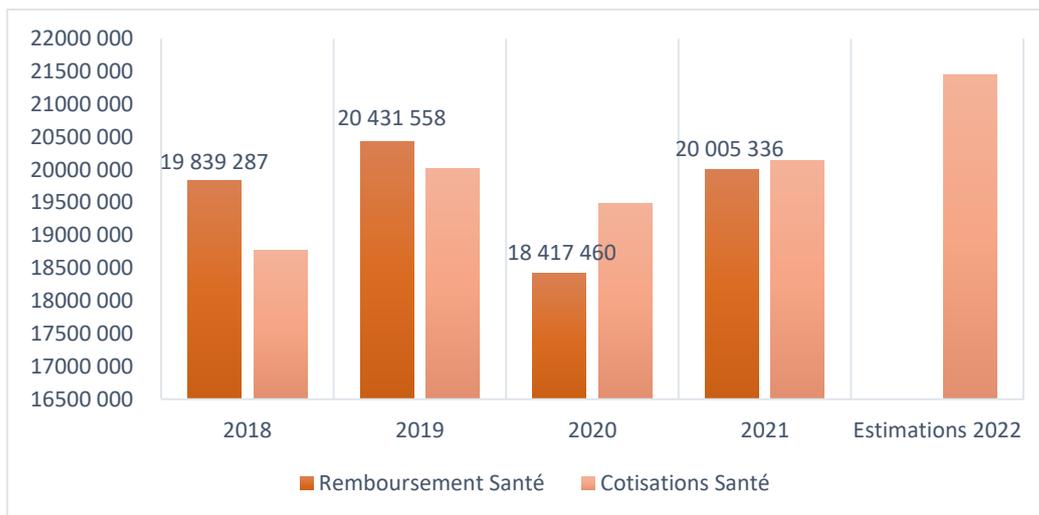


Figure 43 : Prestations et cotisations du portefeuille PHAC entre 2018 et 2022

Une analyse de 2020 n'est pas nécessaire et non constructive. L'année de la Covid a non seulement engendré une diminution nettement visible des prestations, avec une baisse de 2 millions d'euros sur le portefeuille PHAC, mais également une diminution des cotisations. Cette baisse est principalement due à l'impact du chômage partiel sur les salariés pendant la crise.

Du côté du portefeuille PHIE, le phénomène d'inversion du niveau des cotisations et des prestations n'est pas observable. Les cotisations sont tout de même en hausse depuis 2018, avec une augmentation plus marquée en 2022. Comme dans la précédente étude réalisée sur les salariés cadres, une indexation de 1,5% sur ce périmètre, ainsi que des affaires nouvelles à hauteur de 1,9 millions d'euros, viennent accroître le chiffre d'affaires du portefeuille. Une hausse importante est donc prévue entre 2021 et 2022, estimée à 2,5 millions d'euros. Cependant, une dégradation significative de la sinistralité en 2021 pourrait ne pas améliorer le ratio S/P.

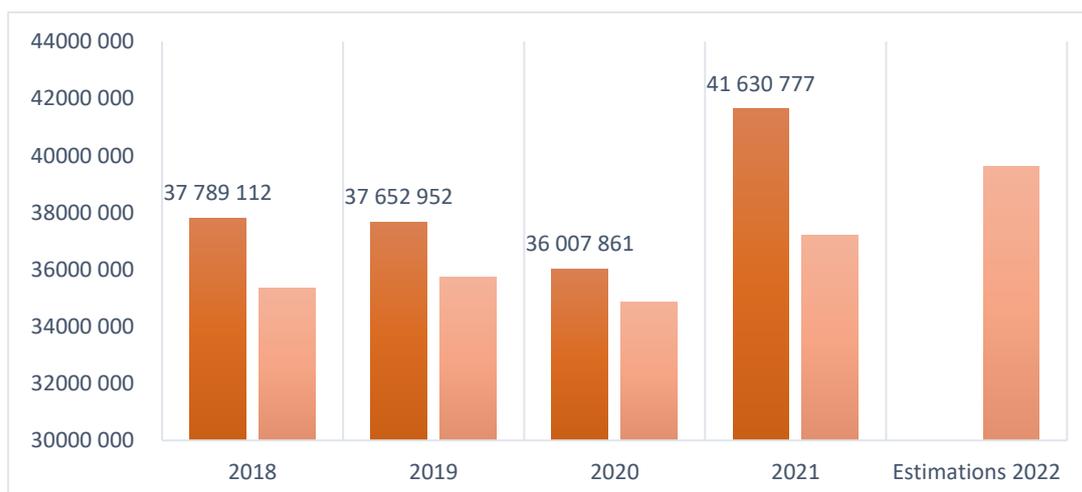


Figure 44 : Prestations et cotisations du portefeuille PHIE entre 2018 et 2022

La hausse de près de 5M d'euros entre 2021 et les survenances antérieures pourrait s'expliquer par la mise en place du 100% santé. C'est à partir de 2021 que la réforme est totalement déployée en France. Une hausse significative est présente sur les postes concernés par cette réforme, notamment le dentaire, aussi bien sur les consultations classiques que sur la pose d'implant dentaire. Ces deux actes médicaux ont connu une augmentation de 35%, représentant une hausse de 2,5 millions d'euros, tandis que l'optique enregistre une hausse de 9%, soit 700 000 euros.

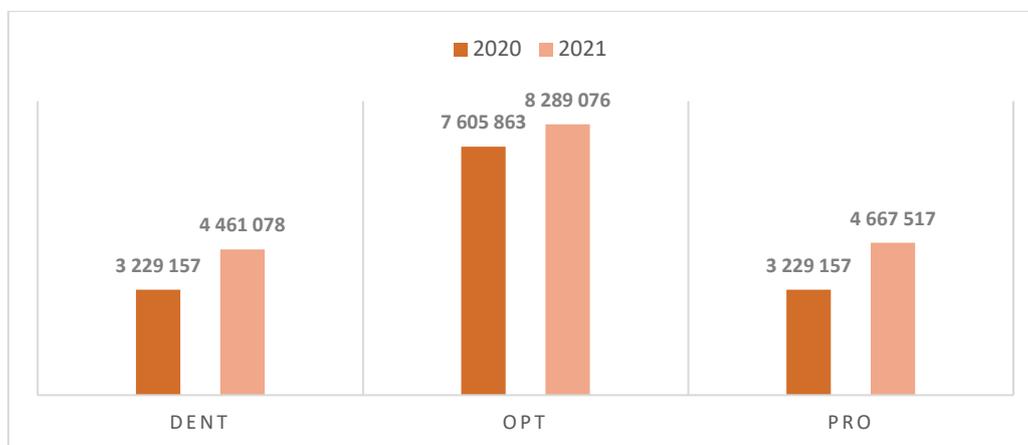


Figure 45 : Remboursements Dentaire, Optique et Prothèse Dentaire entre 2020 et 2021

Cette dégradation étant due à la réforme, elle devra naturellement se maintenir sur l'année 2022, il ne s'agit pas ici d'un phénomène exceptionnel. Ces informations nous donnent des pistes de prédictions. Dans notre cas pratique, la prédiction est réalisée sur l'année 2021 afin de pouvoir back-tester nos résultats. Dans le cas de la prévision 2022, le modèle devrait donc projeter le même niveau de sinistralité que celui de 2021 (sans facteur exceptionnel).

Etude sur les autres branches

Dans notre maille d'analyse sous Solvabilité 2, les autres branches sont regroupées dans un segment du même nom. Dans ce dernier, sont regroupées les entreprises des métiers de l'immobilier, des laboratoires médicaux, des fleuristes, ainsi que la branche « vin et spiritueux ».

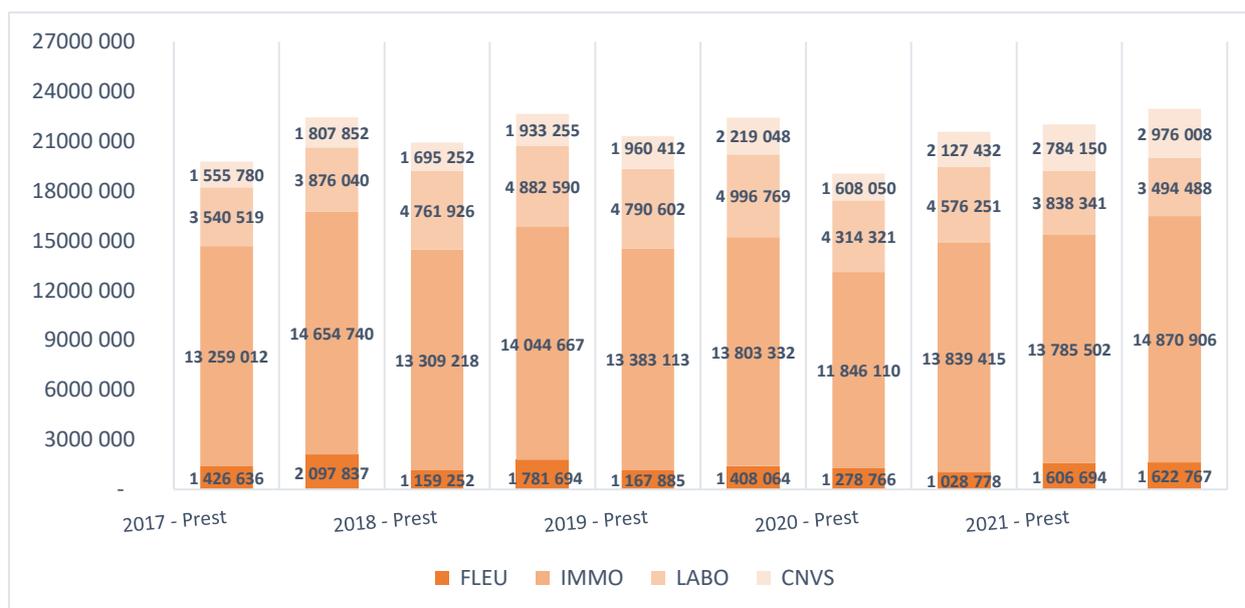


Figure 46 : Prestations des branches de KP entre 2017 et 2018

Ces portefeuilles restent stables d'une année à l'autre depuis 2017. Prévisionnellement, en ne prenant en compte que les années antérieures à 2021, la charge de sinistralité ne devrait pas montrer de nette évolution à la baisse comme à la hausse.

4.2. Résultats de la modélisation

Avant de débiter l'analyse et la comparaison des résultats, il est à noter que les prestations réelles de la survenance 2021 ne sont pas complètes, puisque cinq mois sont manquants. Cependant, ces mois ne représentent qu'un montant de remboursement très faible, avec une moyenne d'environ 110 000 euros au cours des trois dernières années.

Les prochains graphiques résument l'ensemble des résultats, incluant les estimations réalisées par un modèle SARIMA, celles du budget, ainsi que les données réelles, à titre de comparaison.

Le détail des résultats de la modélisation est présenté en annexe B.

Le premier concerne le portefeuille PHAC, représentant les pharmacies cadres et le second, PHIE, destiné aux pharmacies non-cadres.

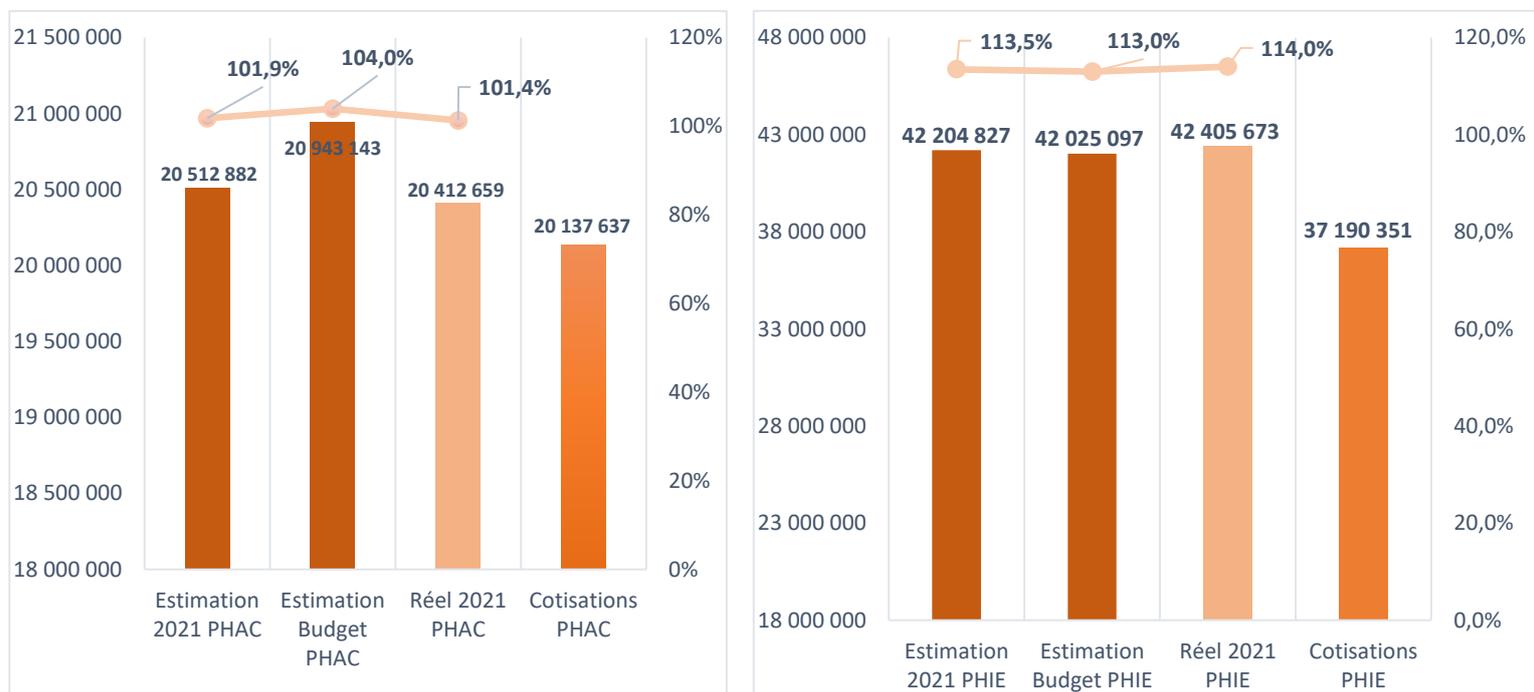


Figure 47 : Prévisionnel 2021 PHIE & PHAC

Si l'on se place sur le périmètre PHAC, l'estimation réalisée par un modèle SARIMA donne une sinistralité de 20,5 millions d'euros, soit un écart de 100 000 euros par rapport au réel.

En examinant le plan prévisionnel d'activité élaboré lors du budget, la sinistralité apparaît légèrement plus dégradée. Cependant, dans la comparaison des ratios, les résultats sont très proches, témoignant tous d'un portefeuille déficitaire. Avec cet écart modeste, l'estimation est considérée comme acceptable.

Le portefeuille PHIE présente une hausse significative lors de la survenance 2021 (figure 44). Malgré cela, le modèle SARIMA a su capter cette hausse du fait de la prise en compte des trois premiers mois lors de la prévision. En pratique, lors de l'établissement du plan prévisionnel d'activité, les mois de janvier à mars sont connus, ce qui fournit au modèle une base à partir de laquelle prédire les mois suivants. Les résultats obtenus par la méthode des séries temporelles sont cohérents avec l'estimation du budget. On constate un S/P presque équivalent entre les deux méthodes, avec un différentiel inférieur à 200 000 euros. De plus, ces résultats sont très proches de ceux observés dans la réalité.

Mensuellement, les données représentent très bien la réalité observée. Les trois premiers mois étant déjà connus au moment du budget, l'estimation débute à partir du quatrième mois. Cette connaissance des premiers mois permet d'améliorer nos cadences.

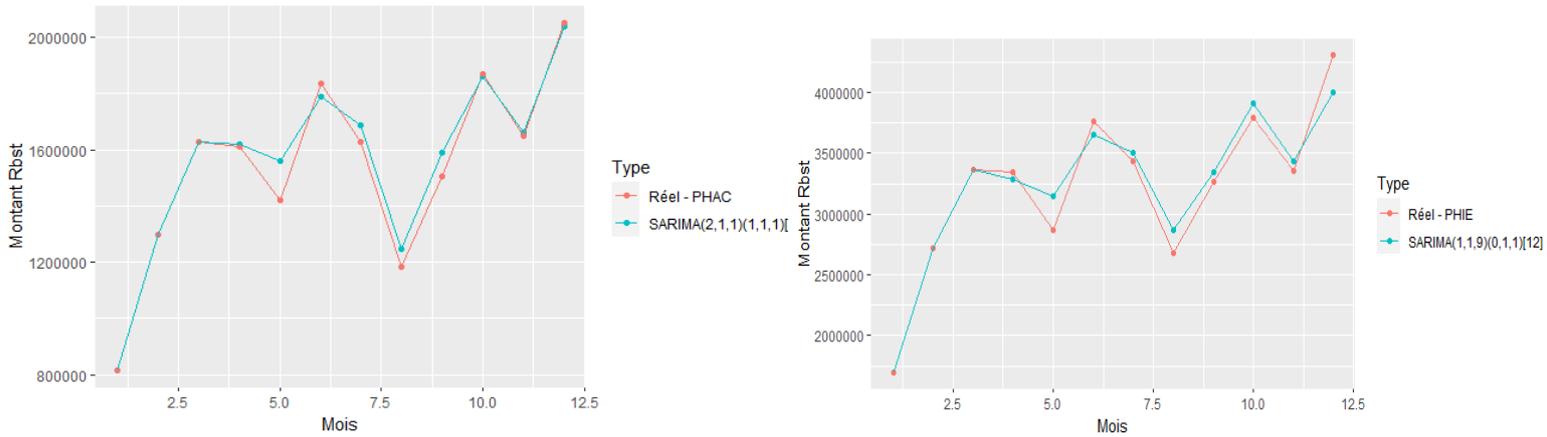


Figure 48 : Résultats graphiques – PHIE et PHAC

Du côté du segment « Branches », des S/P ont été estimés de la même façon que les portefeuilles « pharmaceutiques ».

Sur la survenance complète, nos estimations sont légèrement inférieures à la réalité. La charge de sinistralité atteint 22,3 millions d’euros, soit une estimation très proche de la réalité puisque le delta est inférieur à 200 000 euros, ce qui se reflète sur le ratio. Avec un S/P quasi-identique, l’estimation serait ainsi concluante.

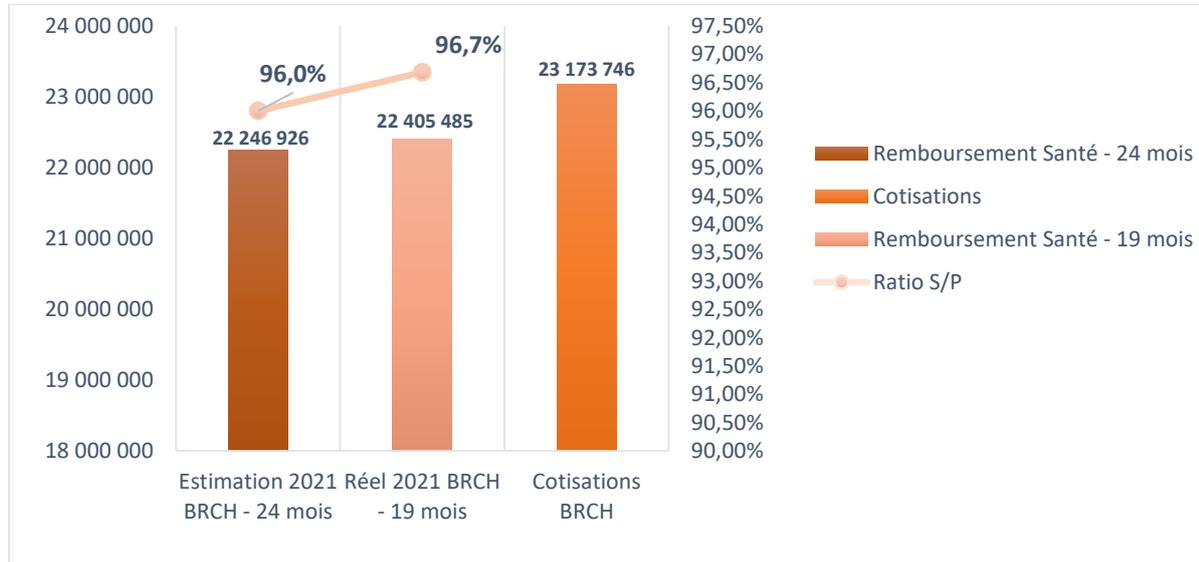


Figure 49 : Prévisionnel 2021 segment « Branches »

Lorsque l’on adopte un point de vue plus détaillé, on remarque que les remboursements de certains mois ne sont pas pris en compte par notre modèle. Il s’agit spécifiquement des sixièmes et douzièmes mois, lesquels s’éloignent le plus de la réalité. Annuellement, cela ne se perçoit pas, en raison de la compensation de ces deux moins bonnes prédictions. Dans l’ensemble, la prévision est donc convaincante.

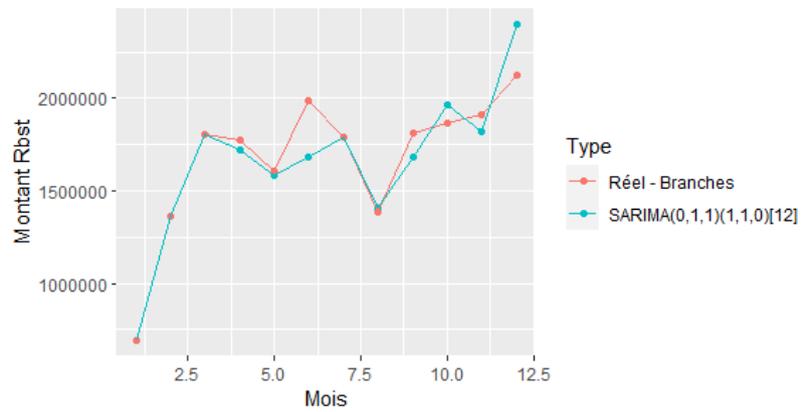


Figure 50 : Résultats graphique - Branches

Nous avons montré précédemment que le modèle SARIMA fonctionnait sur la globalité de l'entité. Dans cette partie, nous démontrons que cette théorie peut également s'appliquer à une maille plus fine, tels que les portefeuilles. Etant donné les résultats présentés précédemment, les S/P obtenus peuvent désormais être appliqués concrètement dans notre modèle S2.

Partie 4 : Impact des prévisions sur Solvabilité 2

1. Présentation Solvabilité 2

Solvabilité 2 est un cadre réglementaire européen qui a été mis en place en 2016 pour régir la solvabilité et la gestion des risques des compagnies d'assurance dans l'Union européenne (UE). Il s'agit d'un ensemble complexe de règles et de normes élaborées pour renforcer la stabilité financière du secteur de l'assurance et protéger les consommateurs.

La mise en œuvre de Solvabilité 2 a entraîné d'importants changements pour les compagnies d'assurance de l'UE, notamment en matière de reporting et de gestion des risques. Ces compagnies doivent désormais fournir régulièrement des rapports aux autorités de réglementation sur leur solvabilité et leur exposition aux risques. De plus, elles doivent adopter des pratiques de gestion des risques plus sophistiquées afin de s'assurer que leur capital est suffisant pour couvrir les risques auxquels elles sont exposées.

Le site de la Commission Européenne définit Solvabilité 2 comme "un cadre réglementaire pour la surveillance et la supervision des entreprises d'assurance". Ce cadre repose sur trois piliers qui seront explicités par la suite. L'objectif de cette section est de développer plus en détail les éléments abordés par le premier pilier.

4.3. Pilier 1 : Exigence Quantitative

Le pilier 1 repose sur une approche axée sur les risques, exigeant des compagnies d'assurance qu'elles évaluent et quantifient les risques auxquels elles sont exposées. Cette exigence introduit la mise en place d'un bilan en vision économique, anciennement comptable, et instaure de nouveaux indicateurs tels que le SCR (niveau de capital de solvabilité requis) et le MCR (niveau de capital minimum réglementaire), ainsi qu'une vision S2 des provisions, comprenant le Best Estimate et la Risk Margin.

Ces nouveaux indicateurs sont calculés par les assureurs à l'aide d'une formule standard et/ou d'un modèle interne. Chez Klesia, le SCR est calculé par la formule standard.

Le **SCR** est utilisé par les compagnies d'assurance pour déterminer le niveau de fonds propres nécessaire afin de couvrir leurs risques et de garantir leur capacité à honorer leurs engagements envers leurs clients. Selon l'Autorité européenne des assurances et des pensions professionnelles (EIOPA), le SCR est défini comme "le montant de capital minimal que la compagnie d'assurance doit détenir pour garantir qu'il y a une probabilité de 99,5% qu'elle dispose d'un capital suffisant pour répondre à ses obligations envers ses clients au cours des douze prochains mois, compte tenu de ses risques".

Il est une agrégation de différents sous-risques : le risque de marché, risque de souscription, risque de contrepartie et le risque opérationnel.

Le **MCR** est une combinaison d'un montant minimal dépendant de l'activité et un pourcentage du SCR. Il s'agit du capital minimum en dessous duquel l'organisme assureur est placé sous tutelle par l'organisme de contrôle.

Le **Best Estimate**, sous-entendu la meilleure estimation des flux futurs, est défini par l'article 77 de la Directive 2009/138/CE comme étant « la moyenne pondérée, par leur probabilité, des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent, estimée sur la base de la courbe des taux sans risque ». Mathématiquement cela se traduit de la façon suivante :

$$BE = \mathbb{E}^{\mathbb{P} \otimes \mathbb{Q}} \left(\sum_{t \geq 0} \frac{F_t}{(1 + R_t)^t} \right)$$

- F_t : les flux correspondant à l'instant t
- R_t : taux d'actualisation à l'instant t
- \mathbb{P} : la probabilité historique
- \mathbb{Q} : la probabilité risque neutre

La **Risk Margin** est calculée de telle manière que la valeur des provisions techniques S2 corresponde au montant qu'une entreprise d'assurance demanderait pour reprendre et honorer les engagements de l'assureur actuel. La Risk Margin, qui s'ajoute au Best Estimate au passif du bilan prudentiel, correspond ainsi au coût d'immobilisation d'un montant de fonds propres égal au SCR. Le taux correspondant au coût du capital est fixé par la réglementation à 6%. Ainsi, nous avons :

$$RM = CoC \sum_{t \geq 0} \frac{SCR_t}{(1 + R(t + 1))^{t+1}}$$

- CoC (coût du capital): 6%
- SCR_t : SCR requis l'année t
- $R(t + 1)$: taux d'intérêt sans risque de maturité $t+1$

Un bilan sous Solvabilité 2 est présenté schématiquement :

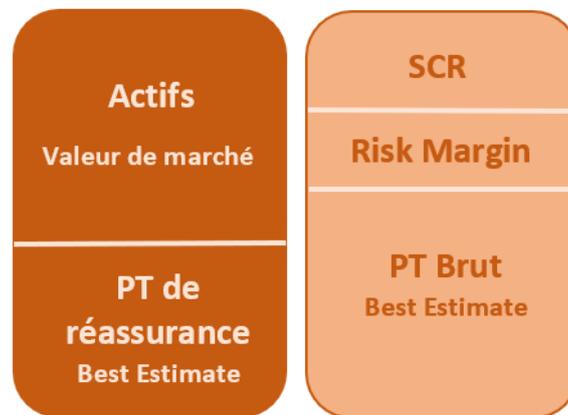


Figure 51 : Bilan Solvabilité 2

4.4. Pilier 2 : Exigence Qualitative & Pilier 3 : Communication

Le pilier 2 impose une bonne gestion aux organismes d'assurance à travers un système de gouvernance. Sous Solvabilité 2, chaque compagnie se doit de définir quatre fonctions clés : la fonction de gestion des risques, la fonction de conformité, la fonction actuarielle et la fonction d'audit interne.

Les assureurs doivent identifier leurs propres risques, les mesurer et les évaluer. Tous ces éléments doivent faire l'objet d'un rapport à l'autorité de contrôle. Cela nous conduit au dernier pilier de Solvabilité 2 basé sur la communication.

Dans le cadre de ce pilier, des rapports qualitatifs, quantitatifs ainsi que des rapports narratifs se doivent d'être produits à destination de l'ACPR. Il existe trois reporting obligatoires :

- SFCR (Solvency and Financial Conditions Report) : rédigé annuellement et destiné au public, il reflète la solvabilité de l'organisme
- RSR (Regular Supervisory Reporting) : destiné à l'ACPR pour mener sa revue, les évolutions importantes doivent être mis à jour annuellement.
- Les QRT (Quantitative Reporting Template) : ils fournissent des informations quantitatives à l'ACPR.

Tous ces rapports font état des informations transmises au niveau des piliers 1 et 2.

2. Présentation du modèle

L'entrée en vigueur de la norme Solvabilité 2 oblige les organismes d'assurance à se doter d'un dispositif intégrant des mesures d'évaluation quantitatives et qualitatives de leurs risques et de leur solvabilité. Dans ce sens, le Groupe KLESIA s'est doté de l'outil de modélisation ADDACTIS et a opté pour l'utilisation de la formule standard pour l'évaluation quantitative de l'ensemble de ses risques.

Le modèle ADDACTIS Modeling permet de projeter les cash-flows relatifs aux portefeuilles d'actifs et de passifs d'assurances à partir de scénarios déterministes des facteurs de risque sous-jacents. Le modèle couvre l'ensemble des risques relevant de l'assurance de personne et est commun à toutes les entités du Groupe. Il revêt une importance particulière dans le processus de production des chiffres Solvabilité 2, tels que le Best Estimate des engagements, la Risk Margin, le SCR ou encore le MCR.

2.1. Modélisation de la charge de sinistralité

La modélisation de la charge de sinistralité sera la partie impactante de notre étude. Elle est divisée entre les encours et la nouvelle production ; des hypothèses de liquidation, préalablement renseignées dans les différents inputs, permettent de projeter la charge de sinistralité sur cinq années.

Le règlement de l'encours est calculé à partir des provisions S1, et une cadence de liquidation est établie pour chaque survenance N :

$$\text{Sinistralité}_{\text{encours}}^N = PT_{S1}^N * \text{Cadence liquidation}_{\text{encours}}^N$$

Quant à la nouvelle production elle est obtenue à l'aide des indicateurs S/P remplis préalablement dans les inputs :

$$\text{Sinistralité}_{\text{simulée}}^N = (S/P^N * CA^N) * \text{Cadence liquidation}_{\text{simulée}}^N$$

Notre étude impactera donc la sinistralité de la charge simulée à travers les S/P de la survenance N+1.

L'objectif est d'établir l'impact que les changements du ratio S/P auront sous S2 à travers les Best Estimate de sinistres qui influenceront sur les fonds propres.

3. Impact sur le résultat Solvabilité 2

Les ratios S/P obtenus précédemment pour les portefeuilles PHIE, PHAC, CNVS, LABO, IMMO, FLEU sont rappelés dans le tableau suivant :

Portefeuilles	PHIE	PHAC	BRANCHES	IMMO	LABO
Ratio S/P - SARIMA	113%	102%	93,2%	93,9%	81,83%
Ratio S/P - Budget	113%	104%	88,6%	93,1%	109,3%
Ratio S/P - Réel⁴	112%	100%	93,5%	93,48%	82%

Figure 52 : Résultats des ratios S/P

Pour les besoins de la modélisation, la projection réalisée précédemment concernant le segment « Branches » a été divisée en trois segments :

- Branches : composé du portefeuille CNVS et FLEU
- LABO
- IMMO

D'après ces résultats, nos estimations sont pertinentes puisque très proches des ratios estimés par la méthode historique utilisée à Klesia et/ou proches de la réalité. L'objectif est d'obtenir des indicateurs au plus proche du réel.

3.1. Impact sur les BE et le bilan S2

La première étude comparative portera sur les BE des sinistres. Cette analyse révèle une baisse de 2,6% par rapport aux valeurs observées lors du budget pour le périmètre PHAC en santé et de 4% sur le segment « Branches ».

Le ratio obtenu par le modèle SARIMA pour PHIE améliore notre sinistralité puisque ce dernier est revu à la baisse. Cette diminution est cohérente avec l'impact S2 observable sur ces BE de sinistres bruts. Naturellement, aucun impact n'est constaté sur le portefeuille PHIE. Concernant le segment « branches », bien que nous ayons augmenté la sinistralité, il s'agit de portefeuilles soumis aux provisions d'égalisation. Celles-ci viennent donc absorber une partie de la sinistralité. Il n'est donc pas surprenant de constater une diminution du BE Sinistres.

L'évolution la plus significative concerne le portefeuille LABO, avec une diminution de 26% du BE de sinistres. Cependant, ce taux doit être modéré compte tenu de son montant relativement bas, le BE de cette branche étant inférieur à 2 millions. Cela reste tout de même cohérent face au ratio S/P relativement éloigné de celui estimé lors du budget de 2021.

⁴ Les données réels ne sont pas complètes à la date de ce mémoire 5 mois sont manquants correspondant au 4^e trimestre de l'année 2022 mais négligeable du fait de leur faible montant.

Segment	Produit	BE Sinistres Brut – Hors Frais Modèle SARIMA	BE Sinistres Brut – Hors Frais Budget	Impact (en %)
PHAC	Santé	9 849 709	10 110 190	-2,6%
PHIE	Santé	14 459 842	14 459 345	0%
LABO	Santé	1 453 440	1 976 674	-26,5%
IMMO	Santé	5 532 833	5 492 451	0,7%
Branches	Santé	2 331 937	2 429 900	-4%

Figure 53 : Impact sur les BE Sinistres

Les résultats précédents ont permis d'établir les conséquences des modifications des ratios S/P uniquement sur le périmètre concerné. Cette fois-ci, les tableaux suivants montrent l'impact sur l'ensemble de l'entité Klesia Prévoyance.

Le nouveau bilan S2 de KP pour l'année 2021 est présenté ci-dessous :

ACTIF en k€	2 021	Variations ACTIF en k€	2 021
Placements	1 902 830 463	Placements	- 193 134
Provisions de Réassurance	1 586 423 274	Provisions de Réassurance	- 1 561 304
Créances	185 453 923	Créances	-
Autres actifs	53 260 691	Autres actifs	-
Total	3 727 968 350	Total	- 1 754 437

0

PASSIF en k€	2 021	Variations PASSIF en k€	2 021
Excedent de fonds propres	64 163	Excedent de fonds propres	796 850
SCR	279 466	SCR	- 104 566
Fonds Propres	343 629	Fonds Propres	692 284
Dettes subordonnées	160 000	Dettes subordonnées	-
Best estimate	2 260 677	Best estimate	- 2 021 230
Marge de risque	23 298	Marge de risque	- 4 159
Provisions techniques	2 283 975	Provisions techniques	- 2 025 389
Dettes	940 365	Dettes	- 421 333
Total	3 727 968 350	Total	- 1 754 437

Figure 54 : Impact sur le Bilan S2 (droite : variation entre le nouveau bilan S2 et celui du PPA)

Le tableau de gauche correspond au bilan résultant de la modification des ratios dans les inputs, tandis que le second montre l'évolution du bilan entre le bilan initial et le nouveau.

Les écarts présentés rejoignent ceux observés sur les BE de sinistres. Comme nous l'avons explicité dans la partie précédente sur la modélisation de la charge sinistre, la sinistralité de la nouvelle production sera modifiée par nos nouveaux ratios. Cela se confirme par les résultats à l'actif comme au passif. Les provisions techniques sont réduites, engendrant une réduction

proportionnelle sur l'actif à travers les provisions techniques cédées. En effet, globalement, les ratios S/P ont été diminués, et par conséquent, la charge de sinistralité également.

Par effet de cascade, les placements voient également leur valeur diminuer, puisqu'ils sont intrinsèquement liés aux provisions techniques. A l'inverse des placements, les fonds propres, quant à eux, sont majorés du fait de la baisse de la charge de sinistralité, augmentant ainsi le résultat naturellement.

Concernant le SCR, son évolution sera analysée plus en détail dans la suite.

3.2. Impact sur le SCR

Le détail des SCR est visible ci-après. Les SCR impactés sont : SCR de marché, SCR de souscription – Santé et le SCR Opérationnel.

		Evolution des SCR après modification des ratios	
SCR Actifs incorporels	-	SCR Actifs incorporels	-
SCR de marché	233 732 013	SCR de marché	- 34 857
SCR de contrepartie	23 653 788	SCR de contrepartie	25
SCR souscription - Non Vie	-	SCR souscription - Non Vie	-
SCR souscription - Vie	23 417 494	SCR souscription - Vie	86
SCR souscription - Santé	71 244 201	SCR souscription - Santé	- 20 480
SCR santé (similaire Vie)	10 865 459	SCR santé (similaire Vie)	80
SCR santé (non similaire Vie)	64 574 836	SCR santé (non similaire Vie)	- 20 732
SCR santé - catastrophe	2 159 813	SCR santé - catastrophe	0
BSCR	276 758 131	BSCR	- 43 588
SCR opérationnel	27 865 772	SCR opérationnel	- 60 683
Ajustements (provisions)	- 25 157 860	Ajustements (provisions)	- 295
Ajustements (impôts)	-	Ajustements (impôts)	-
SCR	279 466 043	SCR	- 104 566
MCR	69 866 511	MCR	- 26 141
Fonds Propres de Base	343 628 547	Fonds Propres de Base	692 284
Fonds Propres éligibles au calcul de SCR	483 361 568	Fonds Propres éligibles au calcul de SCR	640 001
Fonds Propres éligibles au calcul de MCR	357 601 849	Fonds Propres éligibles au calcul de MCR	687 056
Ratio de solvabilité	173%	Ratio de solvabilité	0,3%

Figure 55 : Impact sur le SCR 2021 (droite : variation entre les nouveaux SCR et ceux du PPA)

SCR Souscription Santé

Pour commencer, le changement des ratios a un impact sur les SCR Souscription Santé, et plus particulièrement sur le risque de primes et réserves. Cela permet d'évaluer le risque auquel l'assureur s'expose en cas de sous-provisionnement. Cela se traduit par des provisions au bilan insuffisantes pour couvrir le montant des prestations des sinistres déjà connus. Il s'agit du risque de réserve, mesuré en comparant les provisions en début d'année N aux prestations et provisions en vision Best Estimate à la fin de l'année.

Concernant le risque de primes, il correspond à un risque de sous-tarification où les primes ne permettraient pas de faire face aux montants des prestations. Ce dernier est mesuré par le ratio « S/P » (Sinistres/Primes), comparant ainsi le volume des cotisations à la charge à l'ultime.

Dans ce mémoire, nous avons développé un modèle permettant de calculer ce ratio. Comme mentionné précédemment, le ratio du portefeuille PHAC est inférieur à celui du budget. Par conséquent, il serait normal de s'attendre à une diminution du SCR Primes et Réserves. Cependant, cette baisse ne devrait pas avoir un impact significatif, comme nous l'avons déjà observé pour les BE de sinistres et les provisions techniques.

SCR Souscription	S2 - SARIMA	S2 - Budget	Impact
<i>SCR santé (non similaire Vie)</i>	64 574 835	64 595 568	-20,7K -0,03%

SCR Marché

Le risque de marché revêt une grande importance en assurance vie. Le SCR représente le capital nécessaire pour que l'assureur puisse faire face aux variations des marchés financiers. Etant donné la volatilité potentielle des prix du marché, il est essentiel de disposer d'un capital capable d'absorber ces pertes éventuelles.

Le SCR de marché est calculé selon la formule standard :

$$SCR_{\text{marché}} = \text{MAX}(SCR_{\text{Up}}, SCR_{\text{down}})$$

Des chocs sont réalisés par une hausse et une baisse du marché ; le SCR correspond au maximum des deux.

Les fluctuations sur les marchés financiers auront un impact sur les placements effectués pour le compte de Klesia Prévoyance, les chocs sont donc appliqués à ces derniers. Mais ces placements sont directement liés aux provisions techniques. Or nous avons montré précédemment que ces derniers avaient nettement diminués, de l'ordre de 2 millions d'euros. Il est donc tout à fait normal que les SCR_{Up} et SCR_{Down} soient revus à la baisse et par conséquent le SCR de marché également.

SCR Souscription	S2 - SARIMA	S2 - Budget	Impact
SCR Marché	233 732 013	233 766 870	-34,8K -0,03%

SCR Opérationnel

La directive 2009/138/CE du parlement européen et du conseil définit le risque opérationnel comme « le risque de perte résultant de procédures internes, de membres du personnel ou de systèmes inadéquats ou défaillants, ou d'événements extérieurs ».

Un surplus de capital a donc été prévu afin de faire face à ces risques opérationnels.

Le SCR est calculé selon la formule standard :

$$SCR_{Op} = \min(30\% * BSCR; Op) + 25\% * Exp_{ul}$$

Le BSCR désigne le capital minimum requis. Exp_{ul} correspond au montant de dépense annuelle sur les contrats en assurance vie. Dans notre étude, cette composante est nulle.

$$Op = \max(Op_{primes}; Op_{PT})$$

Avec Op_{primes} et Op_{PT} , le capital requis pour le risque opérationnel est calculé sur la base des primes acquises et des provisions techniques, respectivement. Ces deux éléments sont calculés de la façon suivante :

$$Op_{PT} = 3\% * \max(PT_{non-vie}; 0) + 0,45\% * \max(0; PT_{vie} - PT_{UC})$$

$$Op_{primes} = 3\% * P^N_{non-vie} + \max\left(0; 3\% * (P^N_{non-vie} - 120\% * P^{N-1}_{non-vie})\right) + 4\% * (P^N_{vie} - P^N_{UC}) + \max\left(0; 4\% * \left((P^N_{vie} - 120\% * P^{N-1}_{vie}) - (P^N_{UC} - 120\% * P^{N-1}_{UC})\right)\right)$$

$P^N_{non-vie}$ correspond aux primes brutes de réassurance de l'année N pour un risque non-vie.

$PT_{non-vie}$ il s'agit des provisions techniques non-vie.

Le SCR Opérationnel est le SCR le plus impacté par nos modifications ; de ce fait, une analyse des Op permettra d'expliquer cette hausse.

	Op_{primes}	Op_{PT}	Op	$30\% * BSCR$	SCR
<i>Budget - 2021</i>	22 804 413	27 926 455	27 926 455	83 040 515	27 926 454
<i>Nouveau Modèle</i>	22 804 413	27 865 772	27 865 772	83 027 439	27 865 771

Dans les deux cas, on a : $SCR = Op$ mais également $Op = Op_{PT}$

Cette baisse est donc naturellement expliquée et on retrouve les explications initiales, à savoir que la baisse des provisions techniques engendre une baisse du SCR Opérationnel. En effet, la diminution des provisions engendre intuitivement une réduction des erreurs opérationnelles car l'assiette soumise au risque est plus faible.

Le BSCR n'a aucun impact dans ce cas.

En conclusion, le SCR a globalement diminué de 105K, entraînant une augmentation très faible du ratio : +0,3 point. Soit un impact négligeable de 0,03%.

Cette section a permis de mettre en évidence que la diminution du ratio S/P de quelques portefeuilles santé a un impact inégal. Les BE et les fonds propres ont été les plus exposés, avec une baisse respective de 840K et 692K.

Par ailleurs, l'objectif n'était pas d'obtenir des résultats à l'opposé de la méthode actuellement utilisée lors du budget, mais plutôt d'établir une méthode basée sur un modèle mathématique afin d'estimer les ratios de sinistralité et de fiabiliser la méthode déjà existante.

Conclusion

L'objectif de ce mémoire était d'appliquer des modèles de prévision afin d'établir une nouvelle méthode d'estimation des ratios S/P. Actuellement, la méthode consiste à attribuer un taux de dérive aux S/P de l'année précédente.

Les modèles de séries temporelles ont permis de créer une nouvelle méthode dans l'obtention de ces ratios. Lors de leur construction, nous avons été confrontés à l'impact d'événements externes tels que la Covid, ce qui a modifié le rythme de règlement des prestations du portefeuille. Cette analyse a eu pour conséquence le retraitement de nos données sur cette période précise.

Une fois les modèles obtenus et validés, les ratios S/P ont pu être estimés à partir de la charge de sinistralité et du chiffre d'affaires établi lors du budget. Ces ratios ainsi obtenus ont démontré leur fiabilité, étant assez proches de la réalité. Les cadences mensuelles ont également confirmé ce résultat en restant cohérentes avec la réalité, malgré une faible marge d'erreur, tout à fait acceptable. Il en a été de même lors de l'application à la maille portefeuille, un taux d'erreur inférieur à 3% a été observé dans chaque cas étudié, rendant les résultats concluants.

Il en ressort des résultats pertinents. L'objectif étant d'utiliser ces modèles comme outil de support et de comparaison pour les prochains plans prévisionnels d'activité, afin de valider et de fiabiliser définitivement ces résultats.

Pour conclure, une analyse de l'impact sous Solvabilité 2 a été réalisée afin de comparer les résultats obtenus lors du budget. Il en résulte un effet modéré sur les provisions techniques, les BE ainsi que sur les fonds propres et les SCR. Nous avons pu ainsi identifier le poste ayant subi la variation la plus importante. Dans notre cas, à l'exception des provisions techniques qui sont naturellement les plus impactées, les BE de sinistres ont également subi une diminution supérieure aux autres postes.

Comme expliqué précédemment, l'objectif n'était nullement d'observer un changement significatif par rapport au budget, puisque cette méthode fonctionne aujourd'hui. L'intention était davantage d'établir une nouvelle méthode de prévision visant à challenger les résultats et basée sur un modèle mathématique permettant d'utiliser l'historique de l'entité. Dans les années à venir, ce modèle pourra être exploité afin d'améliorer nos prévisions.

Bibliographie

Mémoires Actuariat

- E.BONIN. (2010). *Méthodes de projection du risque santé : intérêt des séries temporelles.*
- M.BRIEC. (2013). *Développement et déploiement d'un outil de prévision dans le cadre de la surveillance d'indicateurs.*
- S.HUYNH. (2021). *Open data et assurance santé : l'union fait la force ?*
- A. MEKONTSO FOTSING. (2018). *L'open DAMIR : apport à la maîtrise des dépenses de santé.*
- Y.RASSID. (2020). *Le processus ORSA au service d'une étude sur l'impact du COVID-19.*
- G. STASINSKI. (2022). *Impact des mesures prises contre le Covid-19 sur les dépenses santé en France métropolitaine en 2020 à l'aide de la base Open DAMIR.*
- L.MARCIANO. (2018). *Modélisation de la dérive des soins de santé à court terme.*
- T.LE HO (2018). *Reporting santé : comment le perfectionner à l'aide de la Data Science ?*
- C.NICOLLE (2017). *Tarifification au trajet à l'aide de l'Open Data.*
- N.KHEMLICHE (2017). *Tarifification du produit Propriétaire Non Occupant.*
- N.COULIBALY (2011) *Modélisation de la fréquence et du coût moyen des sinistres en Assurance Automobile du particulier avec une approche temporelle et l'intégration de variables explicatives*

Articles

- E.ALIFANO, E.TITON, V.SOULAS. (2022). *Prédiction des consommations médicales et des indemnités journalières en France pour 2022-2025.*
- DREES. (2021) *Quelle démographie récente et à venir pour les professions médicales et pharmaceutique ?*

Textes

DIRECTIVE 2009/138/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL

Table des Figures

FIGURE 1 : REPARTITION DES REMBOURSEMENTS DE LA SECURITE SOCIALE EN 2018	8
FIGURE 2 : ILLUSTRATION DES TAUX DE REMBOURSEMENT	9
FIGURE 3 : SCHEMA DE REMBOURSEMENT DE LA SECURITE SOCIALE	9
FIGURE 4 : SCHEMA DE REMBOURSEMENT 100% SANTE	10
FIGURE 5 : REMBOURSEMENTS DES SOINS SANTE EN 2021	13
FIGURE 6 : REPARTITION DES REMBOURSEMENTS SANTE DE KLESIA PREVOYANCE PAR PORTEFEUILLE	13
FIGURE 7 : REMBOURSEMENTS SANTE DE KLESIA PREVOYANCE ENTRE 2011 ET 2022.....	14
FIGURE 8 : REPARTITION PAR SEXE	15
FIGURE 9 : REPARTITION PAR TYPE DE BENEFICIAIRE	15
FIGURE 10 : REMBOURSEMENTS ET REPARTITION DES SOINS PAR AGE.....	15
FIGURE 11 : REMBOURSEMENTS DES MOINS DE 15 ANS PAR FAMILLE D'ACTE DE SOINS	16
FIGURE 12 : REMBOURSEMENTS PAR FAMILLE D'ACTE DE SOINS	17
FIGURE 13 : REPARTITION EN NOMBRE PAR FAMILLE D'ACTE DE SOINS	17
FIGURE 14 : GRAPHIQUE HISTORIQUE DES PRESTATIONS SANTE OBSERVEES ET ATTENDUES	18
FIGURE 15 : ESTIMATION DES BONI-MALI EN 2022.....	18
FIGURE 16 : ETAPES DE SELECTION DU MODELE.....	26
FIGURE 17 : METHODE DE LA BANDE SUR LES PRESTATIONS SANTE KP PAR SURVENANCE SUR L'ANNEE COMPTABLE N ENTRE 2011 ET 2020.....	35
FIGURE 18 : ECART-TYPE ET MOYENNE PAR ANNEE DE SURVENANCE	35
FIGURE 19 : RESULTATS DE LA METHODE BUYS-BALLOT	35
FIGURE 20 : GRAPHIQUE DES PRESTATIONS SANTE KP PAR SURVENANCE SUR L'ANNEE COMPTABLE N+1	36
FIGURE 21 : SAISONNALITE DE LA SERIE.....	37
FIGURE 22 : SERIE DESAISONNALISEE	38
FIGURE 23 : RESULTATS DES TESTS DE STATIONNARITE.....	38
FIGURE 24 : ACF ET PACF DE LA SERIE DESAISONNALISEE ET STATIONNAIRE	39
FIGURE 25 : P-VALUE DES COEFFICIENTS MAXIMAUX DE LA SERIE	39
FIGURE 26 : P-VALUE DES COEFFICIENTS MAXIMAUX DE LA SERIE	39
FIGURE 27 : RESULTATS DES TESTS BOX-PIERCE ET LILLIEFORS	40
FIGURE 28 : COMPARAISON DES PREDICTIONS DE L'ANNEE 2021 POUR LA SURVENANCE 2021 ENTRE MODELE 1 – SARIMA(8,0,8)(1,1,1)[12] , LE MODELE 2 – SARIMA(0,0,1)(0,1,1)[12] ET LES PRESTATIONS REELLES	41
FIGURE 29 : PRESTATIONS SANTE ESTIMEES (SELON 2 MODELES) EN SURVENANCE 2021	42
FIGURE 30 : COMPARAISON DES PREDICTIONS DE L'ANNEE 2022 POUR LA SURVENANCE 2021 ENTRE MODELE 1 – SARIMA(1,0,0)(1,1,0)[12] , LE MODELE 2 – SARIMA(12,0,12)(1,1,1)[12] ET LES PRESTATIONS REELLES.....	43
FIGURE 31 : PRESTATIONS SANTE ESTIMEES (SELON 2 MODELES) EN SURVENANCE 2021 ET REELLES ENTRE LES SURVENANCES 2011 ET 2021 POUR L'ANNEE COMPTABLE N+1.....	43
FIGURE 32 : COEFFICIENTS DE PASSAGE DES SURVENANCES 2017 A 2019.....	45
FIGURE 33 : REMBOURSEMENTS DE JANVIER A DECEMBRE POUR L'ANNEE DE SURVENANCE ET COMPTABLE N.....	46
FIGURE 34 : COMPARAISON DES RETRAITEMENTS REALISES AVEC UNE PROFONDEUR DE 1 AN A 3 ANS.....	46
FIGURE 35 : RESULTAT GRAPHIQUE DU RETRAITEMENT DES RESIDUS.....	47
FIGURE 36 : RESULTAT GRAPHIQUE DU MODELE INITIAL ET DES DEUX METHODES DE RETRAITEMENT	48
FIGURE 37 : TAUX D'ERREUR SUR LES PREVISIONS MENSUELLES ENTRE LE MODELE SARIMA	49
FIGURE 38 : TAUX D'ERREUR SUR LES PREVISIONS MENSUELLES ENTRE LE MODELE HOLT-WINTERS	51
FIGURE 39 : RESULTATS DU PREVISIONNEL COMPTABLE 2022 ET SURVENANCE 2021	51
FIGURE 40 : REMBOURSEMENTS DES SOINS SANTE EN 2022 SURVENUS EN 2021	52
FIGURE 41 : RESULTATS NUMERIQUES DU CA, DE LA CHARGE DE SINISTRALITE ET DES SP OBTENUS AVEC LES MODELES SARIMA ET HOLT-WINTERS	53
FIGURE 42 : RESULTATS NUMERIQUES DU CA, DE LA CHARGE DE SINISTRALITE ET DES SP OBTENUS AU BUDGET.....	53

FIGURE 43 : PRESTATIONS ET COTISATIONS DU PORTEFEUILLE PHAC ENTRE 2018 ET 2022.....	55
FIGURE 44 : PRESTATIONS ET COTISATIONS DU PORTEFEUILLE PHIE ENTRE 2018 ET 2022.....	55
FIGURE 45 : REMBOURSEMENTS DENTAIRE, OPTIQUE ET PROTHESE DENTAIRE ENTRE 2020 ET 2021	56
FIGURE 46 : PRESTATIONS DES BRANCHES DE KP ENTRE 2017 ET 2018.....	57
FIGURE 47 : PREVISIONNEL 2021 PHIE & PHAC	58
FIGURE 48 : RESULTATS GRAPHIQUES – PHIE ET PHAC.....	59
FIGURE 49 : PREVISIONNEL 2021 SEGMENT « BRANCHES »	59
FIGURE 50 : RESULTATS GRAPHIQUE - BRANCHES.....	60
FIGURE 51 : BILAN SOLVABILITE 2	63
FIGURE 52 : RESULTATS DES RATIOS S/P	65
FIGURE 53 : IMPACT SUR LES BE SINISTRES.....	66
FIGURE 54 : IMPACT SUR LE BILAN S2 (DROITE : VARIATION ENTRE LE NOUVEAU BILAN S2 ET CELUI DU PPA)	66
FIGURE 55 : IMPACT SUR LE SCR 2021 (DROITE : VARIATION ENTRE LES NOUVEAUX SCR ET CEUX DU PPA)	67
FIGURE 56 : SERIE STATIONNAIRE.....	75
FIGURE 577 : TEST DE STATIONNARITE.....	75
FIGURE 588 : GRAPHIQUES DES REMBOURSEMENTS PHIE	76
FIGURE 599 : GRAPHIQUES DES REMBOURSEMENTS PHAC	76
FIGURE 600 : GRAPHIQUES DES REMBOURSEMENTS BRANCHES.....	77
FIGURE 611 : TEST DE STATIONNARITE - PORTEFEUILLES.....	77
FIGURE 622 : MODELES PAR PORTEFEUILLES	77
FIGURE 633 : TEST STATIONNARITE	78
FIGURE 644 : RESULTAT GRAPHIQUE PHIE – ANNEE COMPTABLE 2022.....	79
FIGURE 655 : RESULTAT GRAPHIQUE PHAC – ANNEE COMPTABLE 2022.....	79
FIGURE 666 : RESULTAT GRAPHIQUE BRANCHES – ANNEE COMPTABLE 2022	80

Annexes

Annexe A : Résultats de la prévision année comptable N+1 pour des frais de santé survenus l'année N

Ces données présentaient comme les précédentes une saisonnalité de 12. Le graphique de la série stationnaire obtenue après les différents retraitements :

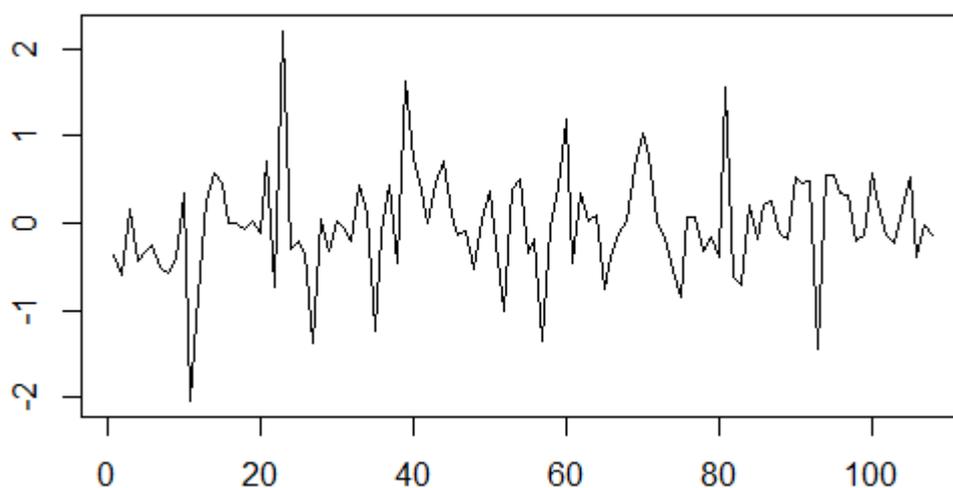


Figure 56 : Série stationnaire

En effet les différents tests, Dickey Fuller, Phillip Perron, ont permis de prouver sa stationnarité. Les résultats sont présentés dans le tableau :

Test	Valeur	P-Value
Dickey-Fuller	-4,7	<0,01
Phillips-Perron	-99,9	<0,01
KPSS	0,21	>0,1

Figure 577 : Test de stationnarité

Annexe B : Modèle par portefeuille

1. Année comptable 2021 pour des frais de santé survenus l'année 2021

➤ Retraitement des effets liés à la Covid

Pour chaque portefeuille, une baisse de la sinistralité est enregistrée entre janvier et mai 2020. De ce fait, le retraitement Covid appliqué sur l'entité Klesia Prévoyance a été mis en œuvre également.

Le graphique suivant représente les remboursements entre 2019 et 2022 avant et après le retraitement pour chaque portefeuille :

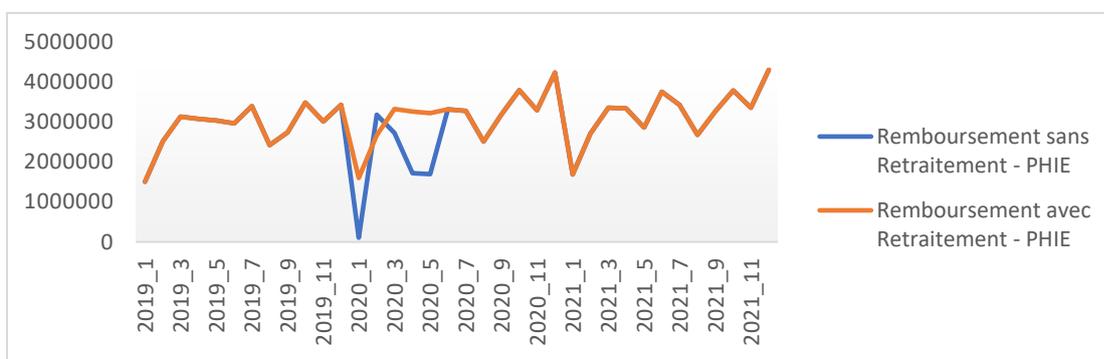


Figure 588 : Graphiques des remboursements PHIE

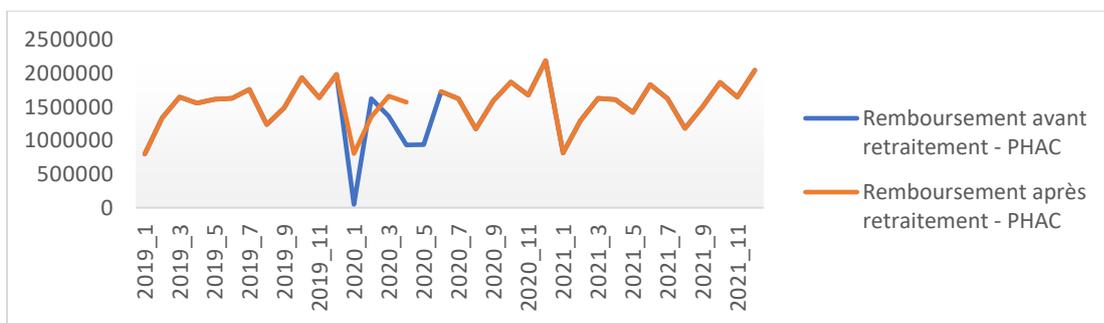


Figure 599 : Graphiques des remboursements PHAC

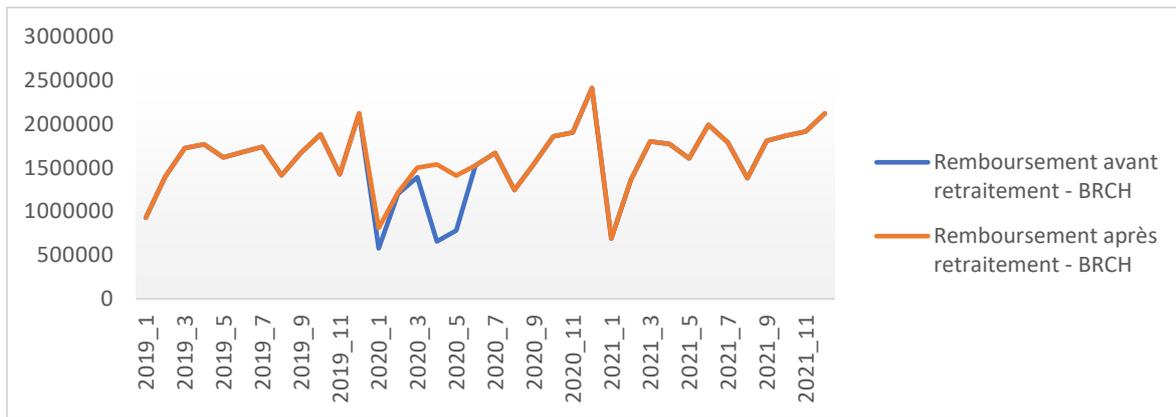


Figure 600 : Graphiques des remboursements BRANCHES

Dans chacun de ces cas, ces retraitements améliorent la sinistralité et corrige un effet Covid.

➤ Etapes de modélisation

Chaque série de donnée (liée à chaque portefeuille) a été désaisonnalisée du fait de la présence d'une saisonnalité de 12. Une fois la série sans saison obtenue des tests de stationnarité ont été réalisés et condensés dans le tableau suivant :

Portefeuilles	Dickey-Fuller	Phillips-Perron
PHIE	0,1	0,12
PHAC	0,49	0,21
BRANCHES	0,34	0,1

Figure 611 : Test de stationnarité - Portefeuilles

Les séries PHIE, PHAC et Branches n'étant pas stationnaires, il est nécessaire d'appliquer un opérateur de différenciation. Dans notre cas une seule différenciation sera requise pour obtenir une stationnarité.

Nos trois séries de données étant désormais stationnaires, il nous est donc possible d'estimer notre modèle SARIMA. En minimisant le critère AIC, on obtient :

Portefeuilles	SARIMA
PHIE	SARIMA(1,1,9)(0,1,1)[12]
PHAC	SARIMA(2,1,1)(1,1,1)[12]
BRANCHES	SARIMA(0,1,1)(1,1,0)[12]

Figure 622 : Modèles par portefeuilles

Les résultats graphiques ont été présentés dans le corps de ce mémoire.

2. Année comptable 2022 pour des frais de santé survenus l'année 2021

De la même façon qu'aucun retraitement n'a été réalisé sur les données de l'entité Klesia Prévoyance, il en sera de même pour les séries relatives aux portefeuilles.

➤ Etapes de la modélisation

Tout comme les données de l'année comptable N, la saisonnalité est retirée sur 2022. D'autre part, la stationnarité est vérifiée grâce aux mêmes tests que ceux utilisés précédemment afin de retirer une potentielle tendance.

Portefeuilles	Dickey-Fuller	Phillips-Perron
PHIE	0,025	0,01
PHAC	0,016	0,01
BRANCHES	0,033	0,5

Figure 633 : Test stationnarité

Excepté pour le périmètre « Branches » la stationnarité est validée. Pour ce dernier, la décision est prise d'appliquer un opérateur de différenciation étant donné que le test de Phillips-Perron ne rejette pas la non-stationnarité.

Les résultats graphiques des prévisions réalisées dans le cadre de l'année comptable 2022 pour la survenance 2021 sont présentés ci-dessous. Les modèles sont calibrés à partir de la minimisation du critère AIC.

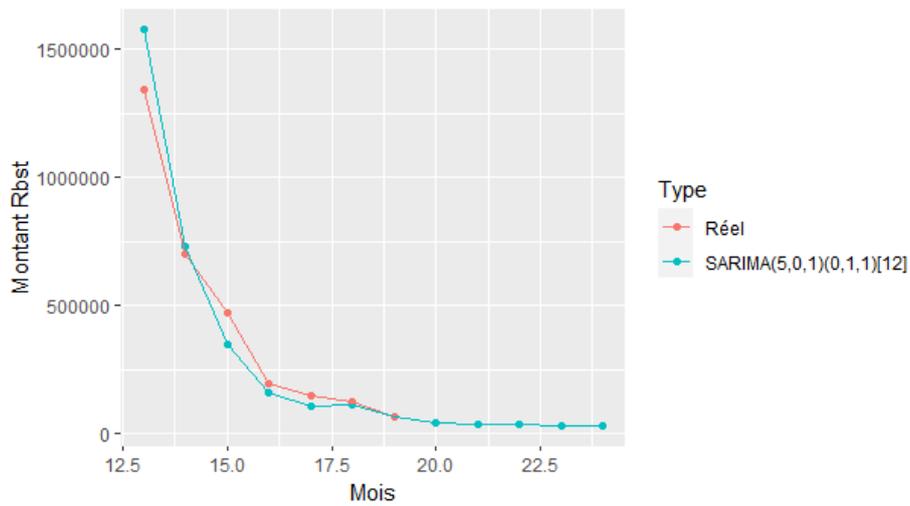


Figure 644 : Résultat Graphique PHIE – Année comptable 2022

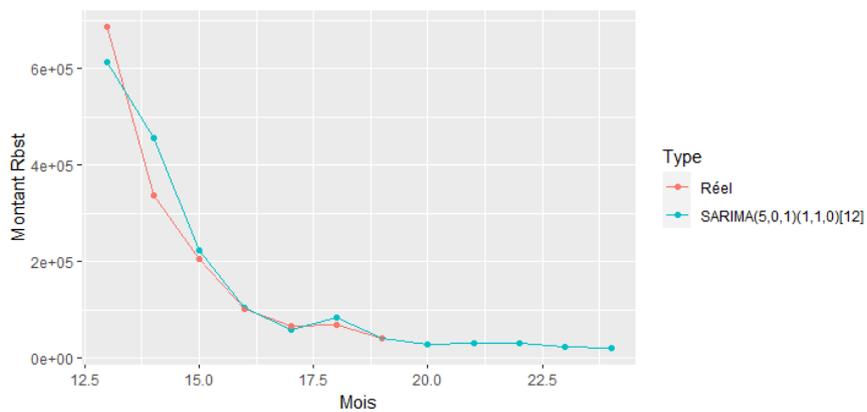


Figure 655 : Résultat Graphique PHAC – Année comptable 2022

Sur ces graphiques, on constate que les prévisions sont plutôt cohérentes avec le réel. Pour PHIE, le delta est inférieur à 40K euros tandis que sur PHAC le taux d'erreur est inférieur à 5%. Cependant, on constate tout de même que le modèle « PHAC » peine à capté les deux premiers mois.

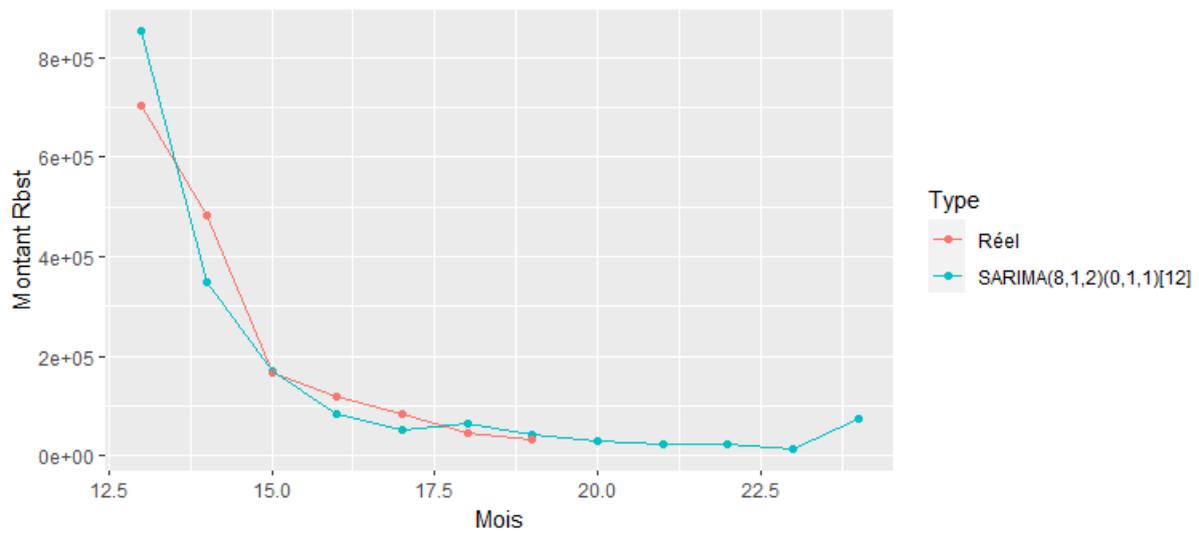


Figure 666 : Résultat Graphique BRANCHES – Année comptable 2022

Tout comme « PHAC », les deux premiers mois sont les moins bien représentés par ce modèle. Cependant au global, on obtient un taux d'erreur très faible. Cela vient de la compensation entre le premier et le deuxième mois. L'un surestime tandis que l'autre sous-estime le montant de prestation réglé.