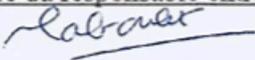


Mémoire présenté le :  
pour l'obtention du diplôme  
de Statisticien Mention Actuariat  
et l'admission à l'Institut des Actuares

Par : Monsieur Shai AZOULAY	
<b>Titre du mémoire : Etude d'un besoin en réassurance pour le risque arrêt de travail d'un portefeuille emprunteur.</b>	
Confidentialité : <input type="checkbox"/> NON <input type="checkbox"/> OUI (Durée : <input type="checkbox"/> 1 an <input checked="" type="checkbox"/> 2 ans)	
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.	
<u>Membres présents du jury de la filière :</u>	Signature : <u>Entreprise :</u>  Nom : CNP ASSURANCES  Signature :
<u>Membres présents du jury de l'Institut des Actuares :</u>	<u>Directeur de mémoire en entreprise</u> Nom : Taboulet Signature : Jacqueline
	<u>Invité :</u> Nom : Signature :
	<b>Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)</b>  <u>Signature du responsable entreprise :</u> 
	<u>Signature du candidat :</u> 

# Remerciements

Je remercie toutes les personnes, qui ont contribué de près ou de loin à la rédaction de ce mémoire.

Bien évidemment, je tiens tout d'abord à remercier ma tutrice de stage, Madame Jacqueline TABOULET, pour la confiance et l'amitié qu'elle m'a témoignées sur ce projet, ses nombreux conseils et encouragements pendant toute la durée de mon alternance ainsi que pour son aide précieuse dans la rédaction de ce mémoire.

Je remercie Madame Khadidiatou DIENG, Responsable du Département Suivi du Risques Techniques France et International d'Assurance de m'avoir accueilli dans son département et pour les nombreux conseils qu'elle m'a prodigués.

J'adresse mes remerciements à Madame Ekaterina NERARD, Responsable du Département Risques Assurance ainsi qu'à l'ensemble de mes collègues du département, pour leur accueil, leur gentillesse et leur disponibilité.

Je tiens également à manifester ma reconnaissance envers mon tuteur académique : Monsieur Nicolas BOUSQUET pour sa disponibilité, ses remarques et ses conseils ainsi qu'à l'ensemble du corps professoral de l'ISUP pour la qualité de leurs enseignements.

Je remercie ma famille, mes amis et Yona pour leur soutien sans faille et leurs encouragements de tous les instants.

## Table des matières

Remerciements	2
Résumé	7
Abstract	8
Note de synthèse	9
Synthesis note	14
Introduction	19
<b>1. Caractéristiques des contrats d'assurance emprunteur</b>	<b>20</b>
1.1. Définition de l'assurance emprunteur	20
1.2. Types de prêts	20
1.2.1. Le prêt immobilier	21
1.2.2. Le prêt professionnel	21
1.2.3. Le prêt à la consommation	22
1.3. Modes de remboursement	22
1.3.1. Le remboursement à échéance constante	22
1.3.2. Le remboursement IN FINE	22
1.3.3. Le différé d'amortissement	23
1.4. Types de contrat	23
1.4.1. Le contrat de groupe	23
1.4.2. Le contrat individuel	23
1.5. Les garanties	23
1.5.1. La garantie décès	24
1.5.2. La garantie PTIA	24
1.5.3. La garantie IPT	24
1.5.4. La garantie IPP	24
1.5.5. La garantie ITP	25
1.5.6. La garantie ITT	25
1.5.7. La garantie perte d'emploi	25
1.6. Limites des garanties	26
1.6.1. Exclusions générales de garantie (ou exclusions légales)	26
1.6.2. Exclusions particulières de garantie (ou exclusions contractuelles)	26
1.7. Analyse du marché de l'assurance emprunteur	27
1.8. Aspects réglementaires	29
1.8.1. Loi Lagarde	29
1.8.2. Loi Hamon	29

1.8.3.	Amendement Bourquin	29
1.8.4.	Loi Lemoine	30
2.	Construction des bases de données	31
2.1.	Base de données des assurés	31
2.1.1.	Descriptif des données brutes mises à disposition	31
2.1.2.	Variables de chaque base	32
2.2.	Base de données des sinistres	34
2.2.1.	Descriptif des données brutes mises à disposition	34
2.2.2.	Variables de chaque base	34
2.3.	Traitement des données	37
2.3.1.	Base de données des assurés	37
2.3.2.	Traitement de la base de données des assurés	38
2.3.3.	Base de données finale des assurés	39
2.3.4.	Base de données des sinistres	39
2.3.5.	Traitement de la base de données des sinistres	39
2.3.6.	Base de données finale des sinistres	39
3.	Étude statistique de l'incidence en arrêt de travail	41
3.1.	Approche envisagée	41
3.2.	Taux d'incidence par âge	41
3.3.	Taux d'incidence par année	43
3.4.	Taux d'incidence moyen par âge sur la période	45
3.5.	Les différentes régressions	47
3.5.1.	La régression linéaire simple	47
3.5.2.	Analyse de la régression linéaire simple	48
3.5.3.	La régression linéaire par morceaux	49
3.5.4.	Vérification numérique de l'efficacité de la régression linéaire par morceaux	53
3.6.	Simulation de taux d'incidence par âge	54
3.6.1.	La théorie des copules	54
3.6.2.	Les différentes copules	55
3.6.3.	Les copules-vines	57
3.6.4.	La Canonical vine (C-vine)	57
3.6.5.	Problématique et méthode	57
3.6.6.	Le critère AIC	58
3.6.7.	Choix de la famille et de la structure de copule pour les deux partenaires	59

3.6.8.	Explications	59
3.7.	Résultats	62
4.	Le maintien en arrêt de travail	64
4.1.	Les tables de maintien en arrêt de travail	65
4.1.1.	Les tables du BCAC	65
4.1.2.	Les tables d'expérience	66
4.1.3.	Concept de mortalité	67
4.1.4.	Définitions	67
4.1.5.	Censure et troncature	68
4.2.	Application	70
4.2.1.	Âge d'entrée en arrêt de travail	70
4.2.2.	Durée de prestation	71
4.2.2.1.	Lois de maintien	71
4.2.2.2.	Analyse	72
4.2.2.3.	Durées résiduelles	73
4.2.2.4.	Comparaison avec les lois chaînées du BCAC	74
5.	La notion de rente	76
5.1.	L'assurance de prêt en cas de sinistre	76
5.1.1.	Déclaration du sinistre dans le cadre d'une assurance emprunteur	76
5.1.2.	Les différentes pièces à fournir	76
5.1.3.	La décision de l'assureur	76
5.2.	Application	77
6.	Résultats	78
6.1.	Méthode d'estimation des charges de sinistres	78
6.2.	Définition du résultat brut d'exploitation	79
6.3.	Distribution de la charge de sinistres par âge	79
6.4.	Comparaison des différents montants	81
6.5.	Analyse	82
7.	Conclusion	83
	Bibliographie	85
	Bibliothèque de R VineCopula: Statistical Inference of Vine Copulas	86
	ANNEXES	87
1.	Les copules-vines	87
2.	La Canonical vine (C-vine)	88

3. Code R	89
4. Étude complémentaire : Le Covid-19	92
Liste des figures	98
Liste des tableaux	99

# Résumé

L'assurance emprunteur est une assurance qui vise à protéger les emprunteurs contre les risques liés à un crédit immobilier, tels que l'invalidité, l'incapacité de travail ou le décès, afin de leur offrir une protection financière et de garantir le remboursement de leur prêt.

Les assureurs, comme CNP Assurances, doivent évaluer les risques liés à ce type d'assurance pour déterminer les primes à proposer et se prémunir contre les pertes éventuelles, en anticipant les dérives de la sinistralité.

Dans ce contexte, les assureurs utilisent des méthodes prospectives pour estimer la charge future de sinistres et peuvent avoir recours à des solutions de réassurance pour protéger leur portefeuille.

Ce mémoire se concentre sur le risque d'arrêt de travail dans le cadre de l'assurance emprunteur.

Le principal objectif de ce mémoire est de proposer une méthode de simulation basée sur la théorie des copules appliquée aux taux d'incidence en arrêt de travail, afin d'analyser la robustesse du portefeuille des assureurs dans le cas de scénarios de stress.

Cette étude permet de déterminer s'il est opportun de mettre en place une solution de réassurance pour protéger les résultats des assureurs.

*Mots-clés : Copule, simulation, régression linéaire, régression linéaire par morceaux, réassurance, Assurance emprunteur, risque arrêt de travail.*

# Abstract

Loan insurance is a type of insurance that aims to protect borrowers against risks associated with a real estate loan, such as disability, inability to work, or death, in order to provide financial protection and ensure the repayment of their loan.

Insurers, such as CNP Assurances, must evaluate the risks associated with this type of insurance to determine the premiums to offer and protect against potential losses by anticipating trends in claims.

In this context, insurers use prospective methods to estimate the future burden of claims and may use reinsurance solutions to protect their portfolio.

This paper focuses on the risk of work stoppage within mortgage insurance.

The main objective of this paper is to propose a simulation method based on copula theory applied to incidence rate and analyze the resilience of insurers' portfolios in the event of a significant increase in this risk.

This study can help determine whether it is appropriate to implement a reinsurance solution to protect insurers' results.

# Note de synthèse

Face à une dérive de la sinistralité des garanties arrêt de travail observée depuis quelques années, les assureurs ont besoin de déterminer les impacts financiers et d'anticiper au mieux ces dérives.

La plupart du temps ils font appel à des solutions de réassurance pour protéger les portefeuilles.

Afin de déterminer les futurs impacts, l'assureur doit effectuer des calculs prospectifs afin d'estimer la charge future de sinistres et effectuer des tests de sensibilité sur cette charge.

Ce mémoire traite des garanties arrêt de travail souscrites dans le cadre d'un contrat assurance emprunteur et propose une méthode d'estimation de la charge de sinistres ainsi qu'une méthode basée sur la théorie des copules permettant d'effectuer des tests de sensibilité.

L'objectif du mémoire est de proposer une méthode permettant de générer des scénarios de stress au regard des données d'expérience.

Le périmètre d'étude de ce mémoire porte sur le portefeuille emprunteur de CNP Assurances et plus particulièrement sur les emprunts immobiliers pour deux partenaires importants. La période d'étude est comprise entre 2015 et 2020.

L'assurance emprunteur est un des domaines de l'assurance de personnes les plus importants sur le marché français et CNP Assurances est le leader de l'assurance emprunteur en France.

Pour tester la robustesse du portefeuille face à une dérive importante du risque arrêt de travail, on compare sa capacité à absorber un choc sur la sinistralité induit par des scénarios défavorables avec sa capacité à créer de la valeur.

La charge de sinistre en arrêt de travail a été estimée à partir de trois composantes : l'incidence par âge, la durée de prestation de la rente et le montant de la rente, la charge finale étant la multiplication de ces trois composantes.

Dans le cadre de ce mémoire, nous avons choisi d'effectuer les scénarios de stress sur les taux d'incidence.

En particulier, la théorie des copules a été utilisée afin de générer les nouveaux couples (âge, taux d'incidence) permettant de tenir compte de la structure bidimensionnelle du vecteur.

L'intérêt de cette étude est de déterminer s'il est opportun de mettre en place une solution de réassurance pour protéger les résultats du portefeuille face à une forte déviation de la charge de sinistres.

## **Caractéristiques des contrats emprunteur**

La première partie de ce mémoire rappelle les caractéristiques des contrats emprunteur : description des garanties proposées, chiffres clés du marché et rappel des dernières évolutions réglementaires.

## Construction des bases de données

La deuxième partie du mémoire détaille la construction de bases de données retraitées à partir des bases de données brutes fournies par CNP Assurances, afin d'obtenir les taux d'incidence.

Les taux d'incidence ont été calculés pour deux partenaires représentatifs du portefeuille emprunteur à partir d'une base de données de sinistres et d'une base de données d'assurés.

## Étude statistique de l'incidence en arrêt de travail

Une fois les bases de données par âge des assurés établies, les taux d'incidence moyens du risque arrêt de travail ont été déterminés par âge, sur le portefeuille emprunteur de CNP pour les crédits immobiliers :

$$T_{i,x} = \frac{\sum S_{i,x}}{\sum A_{i,x}}$$

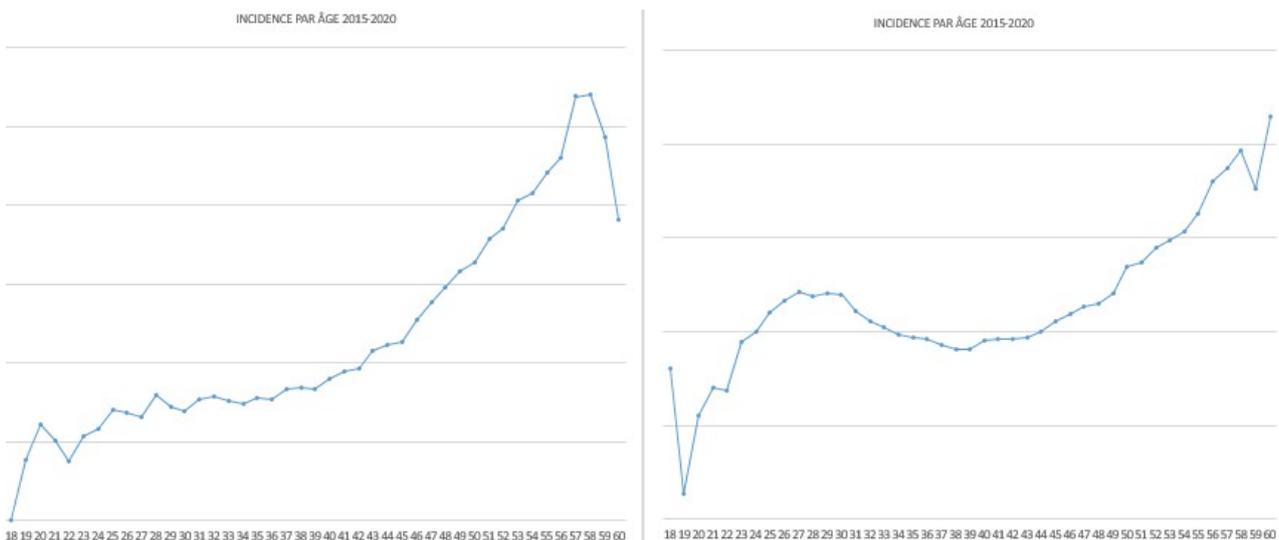
Avec :

$S_{i,x}$  : Nombre de sinistres des personnes assurées sur le portefeuille couvert pour le risque arrêt de travail à l'âge  $i$  pendant l'année de survenance  $x$ .

$A_{i,x}$  : Nombre d'assurés sur le portefeuille couvert pour le risque arrêt de travail à l'âge  $i$  pendant l'année d'inventaire  $x$ .

$i$  : Âge de l'assuré, entre 18 et 65 ans.       $x$  : Année, de 2015 à 2020.

Les graphiques indiquent les taux d'incidence par âge pour chaque partenaire.



Dans un premier temps, nous cherchons une méthode adaptée afin de simuler de nouveaux couples (X, Y) avec X l'âge et Y l'incidence. L'objectif est de simuler l'incidence maximum associée à chaque âge.

Dans le but de simuler de nouveaux taux d'incidence par âge, nous allons chercher à vérifier qu'il

existe une relation linéaire entre l'âge et l'incidence et pour cela la régression linéaire n'est pas suffisante, il faut utiliser la méthode de la régression linéaire par morceaux pour prouver qu'il existe bien une relation entre l'âge et l'incidence.

Ces relations linéaires par morceaux permettront ensuite d'appliquer les simulations sur certains segments d'âge à l'aide de la théorie des copules.

La théorie des copules est une méthode flexible pour la modélisation de la dépendance entre les variables, elle permet de prendre en compte l'incertitude liée à la dépendance entre les variables aléatoires. En effet, on ne peut pas toujours prédire ou anticiper le type de dépendance qui va lier les variables aléatoires et la théorie des copules est fortement utile dans ce type de situation.

On simulera ainsi, 10 000 échantillons bivariés par segment d'âge à l'aide des copules-*vine*. C'est une méthode d'arbre hiérarchique permettant de modéliser les dépendances entre plusieurs variables aléatoires en utilisant les familles de copule de base.

Ci-dessous la comparaison des taux d'incidence maximums simulés (en vert) et des taux d'incidence moyens par âge pour les deux partenaires.

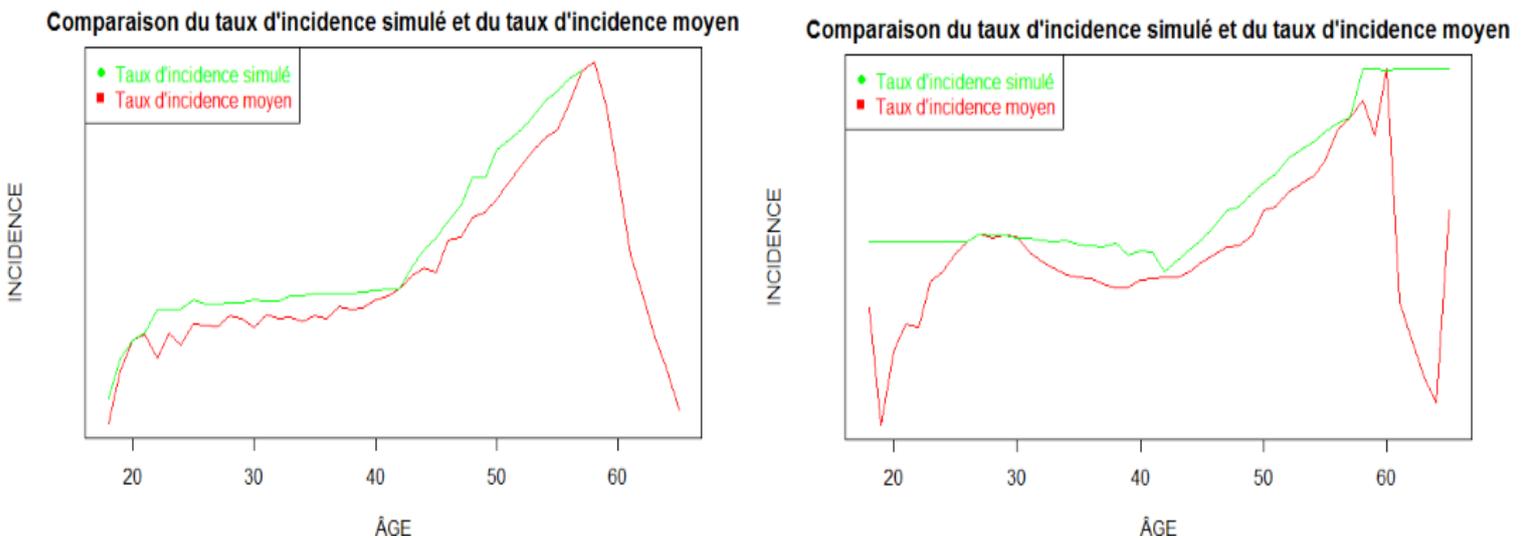


Figure 2 : Graphique des taux d'incidence simulés pour les deux partenaires.

### Le maintien en arrêt de travail

Par ailleurs, afin d'obtenir les durées résiduelles de chaque partenaire, les données des lois de maintien doivent être récupérées en prenant en compte l'âge d'entrée en incapacité et en appliquant la formule ci-dessous :

$$DR_k = \frac{\sum_{j=i+1}^n (S_j)}{(S_i)}$$

Avec :

$DR_k$  : Durée résiduelle (ou durée restante).  
 $S_i$  : Probabilité de survie au-delà de  $i$  mois passé en arrêt de travail.

$i$  : nombre de mois en arrêt de travail.  
 $k$  : le partenaire 1 ou 2.

Voici les lois de maintien des deux partenaires en fonction de leur âge d'entrée en incapacité :

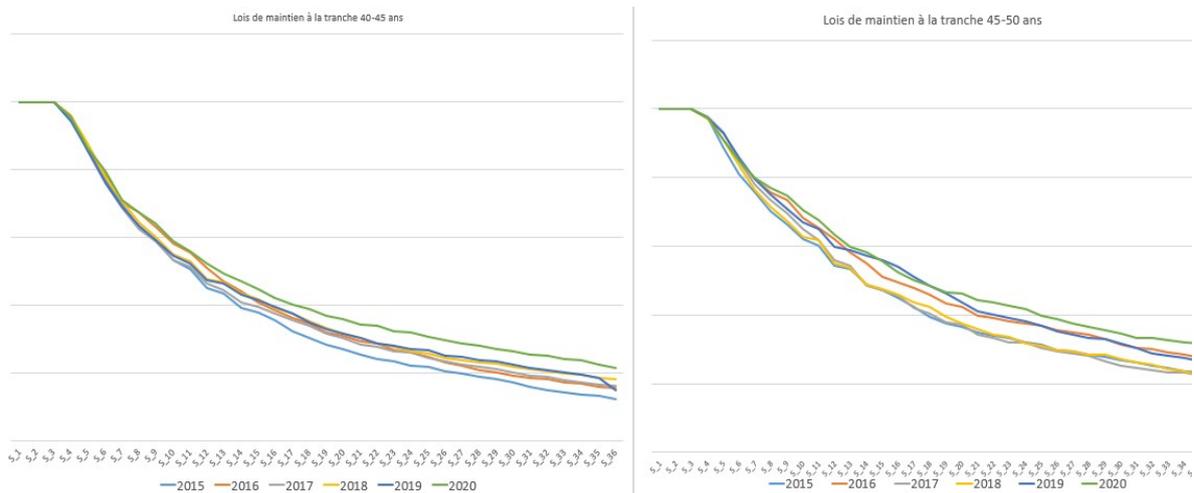


Figure 3 : Lois de maintien en arrêt de travail des deux partenaires

Ainsi, nous obtenons les durées résiduelles suivantes :

Partenaire	Durées résiduelles, en mois
Partenaire n°1	$DR_1$
Partenaire n°2	$DR_2$

Tableau 1 : Les différentes durées résiduelles de chaque partenaire

### Statistiques sur les rentes moyennes

Le montant moyen payé par CNP Assurances en cas de sinistres sur le périmètre a été calculé à partir des bases de données de sinistres :

Partenaire	2015	2016	2017	2018	2019	2020
Partenaire n°1	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$
Partenaire n°2	$R_7$	$R_8$	$R_9$	$R_{10}$	$R_{11}$	$R_{12}$

Tableau 2 : Les montants de rentes obtenus

Toujours dans le cas le plus défavorable, on retient deux montants de sinistres pour chaque partenaire : le montant maximum entre 2015 et 2020 et le montant de sinistre moyen.

Le montant maximum sera utilisé afin de calculer le montant de sinistres supposé maximum que CNP puisse envisager de payer pour le risque arrêt de travail.

Le montant moyen sera utilisé afin de calculer le montant de sinistre supposé moyen que CNP puisse envisager de payer pour le risque arrêt de travail.

### Résultats et conclusion

La stratégie de réassurance de CNP Assurances a pour but de mettre en œuvre une réassurance précise et ciblée afin de se conformer aux limites de risques locaux, en particulier pour garantir une protection en montant, couvrant 50% du résultat brut d'exploitation (RBE).

Afin de déterminer si cette limite a été atteinte, nous avons effectué un calcul pour évaluer le montant en question.

Ainsi, en comparant le montant représentant les 50% du RBE avec notre montant calculé à l'aide de la formule ci-dessous, nous pourrions analyser la pertinence de la stratégie mis en place par CNP ASSURANCES.

Ce montant représente la différence entre le montant de sinistres maximum supposé (calculé à l'aide nos simulation) et le montant de sinistres moyen.

$$S_{dif} = S_{choqué} - S_{moyen}$$

$$= \sum_{i=1}^n A_{x,i} * T_{x,i_{choqué}} * DR_k * Rente_{annuelle} - \sum_{i=1}^n A_{x,i} * T_{x,i} * DR_k * Rente_{annuelle}$$

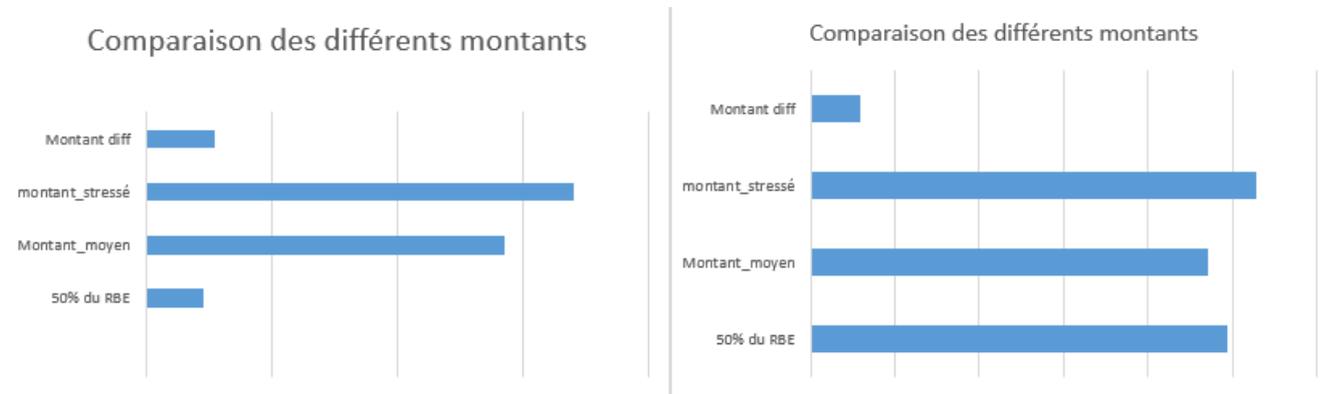


Figure 4 : Comparaison des différents montants pour les deux partenaires.

En examinant ce graphique, on observe que le partenaire n°2 (à gauche) a atteint la limite de la protection prévue, tandis que le partenaire n°1 est loin de l'avoir atteinte.

Le partenaire n°2 est donc plus exposé au risque que l'autre.

Par conséquent, il peut être déduit simplement que pour le partenaire n°2, l'utilisation de la réassurance devrait être envisagée, contrairement au partenaire n°1 pour lequel le montant des sinistres est nettement inférieur à la protection prévue représentant les 50% du RBE.

Néanmoins, la question du recours à la réassurance est beaucoup plus complexe, car il faut prendre en compte la mutualisation de ce risque avec tous les partenaires, et nous avons effectué les calculs seulement pour deux d'entre eux.

## Synthesis note

Faced with an increasing trend in the incidence of work stoppage claims observed over the past few years, insurers need to determine the financial impacts and anticipate these trends as best they can. Most of the time, they rely on reinsurance solutions to protect their portfolios. In order to determine future impacts, insurers must carry out prospective calculations to estimate the future claims costs and perform sensitivity tests on these costs.

This paper deals with work stoppage guarantees taken out as part of a loan insurance contract and proposes a method for estimating claims costs, as well as a method based on copula theory to perform sensitivity tests.

The objective of the paper is to propose a method for generating stress scenarios based on experience data.

The scope of this paper focuses on the loan insurance portfolio of CNP Assurances, particularly on real estate loans for two important partners. The study period covers the years between 2015 and 2020.

Loan insurance is one of the most important areas of personal insurance in the French market, and CNP Assurances is the leading provider of loan insurance in France.

To test the portfolio's resilience to a significant trend in work stoppage risk, its ability to absorb a shock to claims incurred by unfavorable scenarios is compared to its ability to create value.

The claims cost for work stoppage was estimated based on three components: the incidence by age, the duration of the benefit, and the amount of the benefit, with the final cost being the multiplication of these three components.

In this paper, we chose to perform stress scenarios on incidence rates. Specifically, copula theory was used to generate new pairs (age, incidence rate) to take into account the two-dimensional structure of the vector.

The purpose of this study is to determine the effectiveness of a reinsurance solution to protect the portfolio's results in the face of a significant deviation in claims costs.

### **Characteristics of borrower contracts**

The first part of this paper reviews the characteristics of borrower contracts: description of the guarantees offered, key market figures, and a reminder of recent regulatory developments.

## Construction of databases

The second part of the paper details the construction of reprocessed databases from raw databases provided by CNP Assurances to obtain incidence rates.

The incidence rates were calculated for two representative partners of the borrower portfolio from a claims database and an insureds database.

## Statistical study of incidence rates in work stoppage

Once the age-based insured databases were established, the average incidence rates for work stoppage risk were determined by age, on the CNP borrower portfolio for real estate loans:

$$T_{i,x} = \frac{\sum S_{i,x}}{\sum A_{i,x}}$$

With:

$S_{i,x}$  : Number of claims for insured individuals on the portfolio covered for work stoppage risk at age  $i$  during the occurrence year  $x$ .

$A_{i,x}$  : Number of insured individuals on the portfolio covered for work stoppage risk at age  $i$  during the inventory year  $x$ .

$i$  : Age of the insured, between 18 and 65 years old.       $x$  : Year, from 2015 to 2020.

The graphs show the incidence rates by age for each partner.

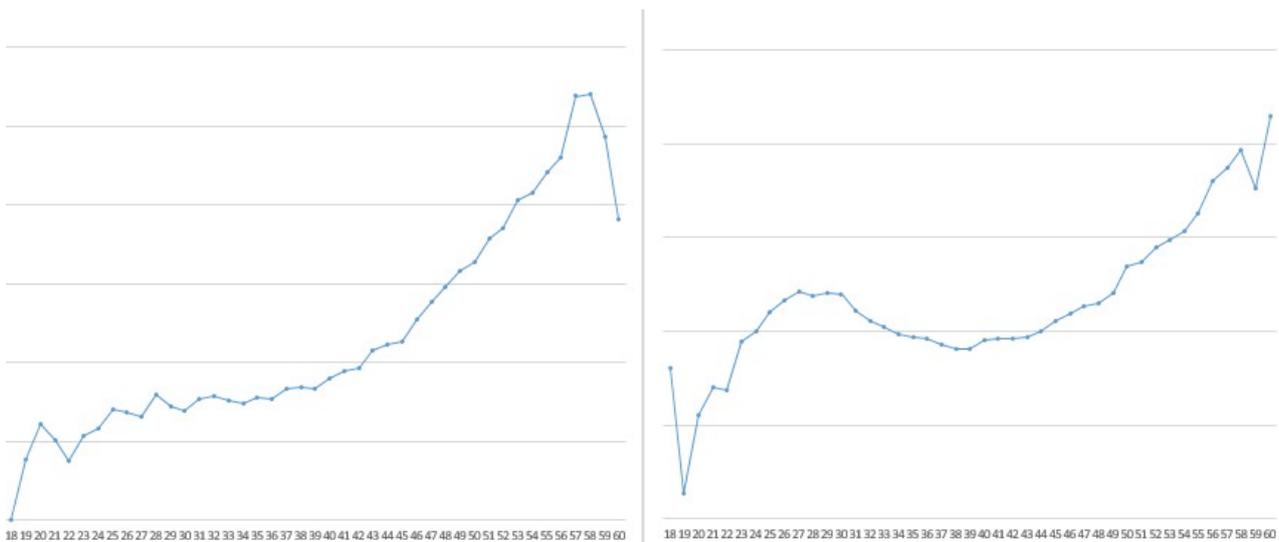


Figure 1: Graphs of average incidence rates by age for the two partners.

Initially, we are looking for a suitable method to simulate new pairs  $(X, Y)$  with  $X$  being age and  $Y$  being incidence. The goal is to simulate the maximum incidence associated with each age.

To simulate new incidence rates by age, we need to verify that there is a linear relationship between age and incidence. Linear regression is not sufficient for this, and we need to use the piecewise linear regression method to prove that there is indeed a relationship between age and incidence.

These piecewise linear relationships will then allow us to apply simulations to certain age segments using copula theory.

Copula theory is a flexible method for modeling the dependence between variables, and it allows for taking into account the uncertainty associated with the dependence between random variables. Indeed, we cannot always predict or anticipate the type of dependence that will link the random variables, and copula theory is particularly useful in this type of situation.

Thus, we will simulate 10,000 bivariate samples per age segment using vine copulas.

This is a hierarchical tree method used to model the dependencies between several random variables using families of basic copulas.

Below is a comparison of the simulated maximum incidence rates (in green) and the average incidence rates by age for the two partners.

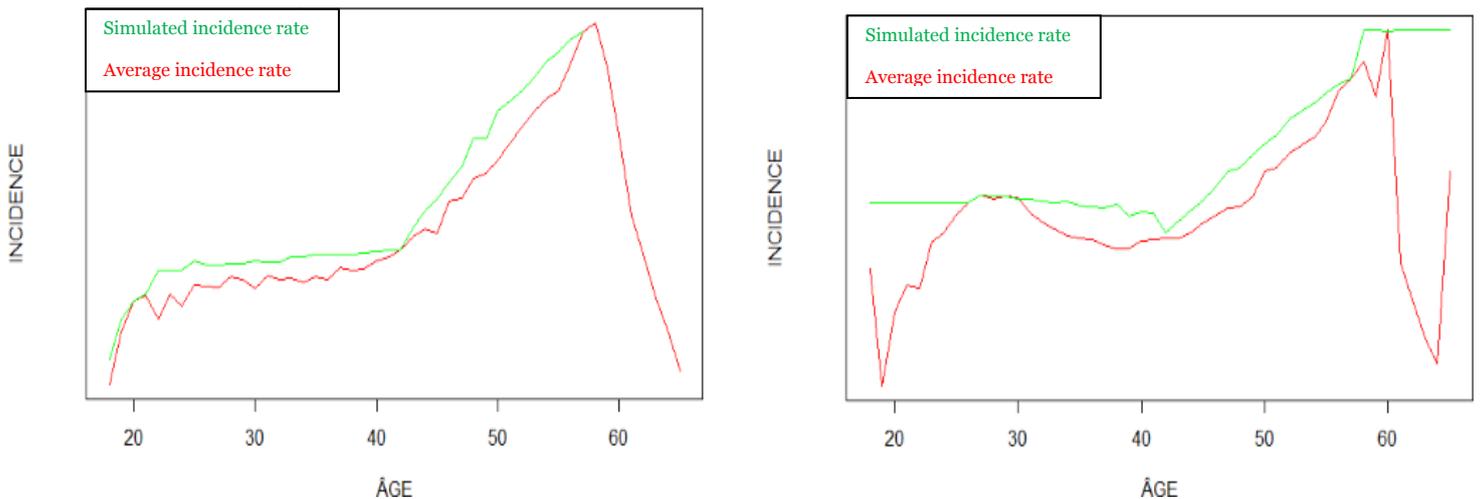


Figure 2: Graph of simulated incidence rates for the two partners

### Work stoppage maintenance.

Furthermore, to obtain the residual durations of each partner, the maintenance law data must be retrieved, taking into account the age of onset of incapacity and applying the formula below :

$$DR_k = \frac{\sum_{j=i+1}^n (S_j)}{(S_i)}$$

With:

$DR_k$  : Residual duration (or remaining duration).  
 $S_i$  : Probability of survival beyond  $i$  months spent on work stoppage.  
 $i$  : Number of months on work stoppage.  
 $k$  : Partner 1 or 2.

Here are the maintenance laws of the two partners based on their age at onset of incapacity:

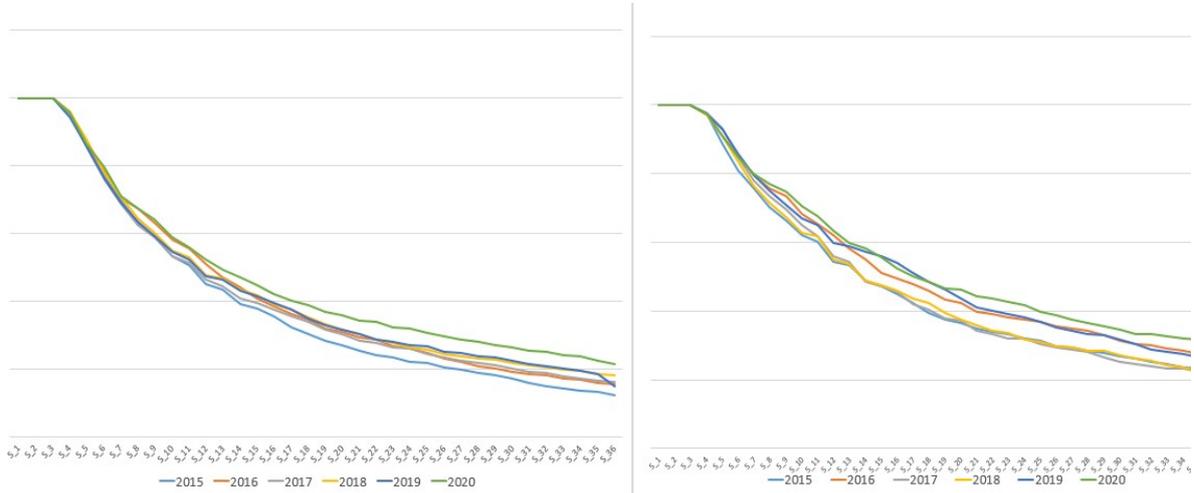


Figure 3: Maintenance laws for work stoppage of the two partners.

Thus, we obtain the following residual durations:

<i>Partner</i>	<i>Residual durations, in months</i>
Partner 1	$DR_1$
Partner 2	$DR_2$

Table 1: Residual durations for each partner.

### Average annuities statistics

The average amount paid by CNP Assurances in the event of claims within the scope was calculated based on claims databases:

<i>Partner</i>	2015	2016	2017	2018	2019	2020
Partner 1	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$
Partner 2	$R_7$	$R_8$	$R_9$	$R_{10}$	$R_{11}$	$R_{12}$

Table 2: Rent amounts obtained.

Continuing with the worst-case scenario, two claim amounts are selected for each partner: the maximum amount between 2015 and 2020 and the average claim amount.

The maximum amount will be used to calculate the maximum claim amount that CNP may be required to pay for the work stoppage risk.

The average amount will be used to calculate the average claim amount that CNP may be required to pay for the work stoppage risk.

### Results and Conclusion

CNP Assurances' reinsurance strategy aims to implement precise and targeted reinsurance in order to comply with local risk limits, particularly to ensure protection in amount, covering 50% of the gross operating result (GOR). To determine if this limit has been reached, we performed a calculation to evaluate the amount in question.

Thus, by comparing the amount representing 50% of the GOR with our calculated amount using the formula below, we can analyze the relevance of the strategy implemented by CNP Assurances.

This amount represents the difference between the supposed maximum amount of claims (calculated using our simulations) and the average amount of claims.

$$S_{dif} = S_{choqué} - S_{moyen}$$

$$= \sum_{i=1}^n A_{x,i} * T_{x,i,choqué} * DR_k * Rente_{annuelle} - \sum_{i=1}^n A_{x,i} * T_{x,i} * DR_k * Rente_{annuelle}$$

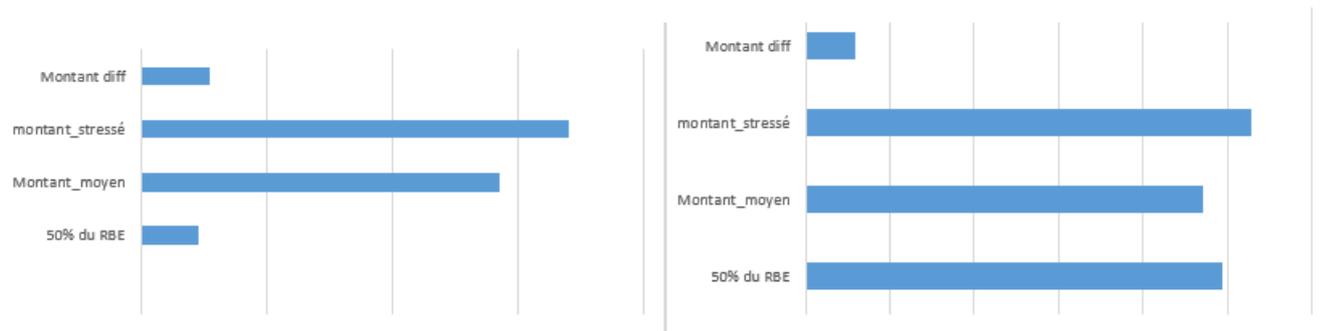


Figure 4: Comparison of different amounts for the two partners.

By examining this graph, we observe that partner 2 (on the left) has reached the limit of the intended protection, while partner 1 is far from reaching it. Partner 2 is therefore more exposed to the risk than the other.

Therefore, it can be simply deduced that for partner 2, the use of reinsurance should be considered, unlike partner 1 for which the amount of claims is significantly lower than the intended protection representing 50% of the RBE. However, the question of whether to use reinsurance is much more complex, as it is necessary to take into account the mutualization of this risk with all partners, and we have only performed calculations for two of them.

# Introduction

L'obtention d'un prêt est conditionnée à la souscription d'un contrat d'assurance emprunteur. L'assurance emprunteur est un des domaines de l'assurance de personnes les plus importants sur le marché français.

CNP Assurances est le leader de l'assurance emprunteur en France.

Dans le but de tester la robustesse du portefeuille face à une dérive importante du risque arrêt de travail, nous allons chercher à calculer la différence entre le montant issu du scénario le plus défavorable et celui issu du montant moyen.

Autrement dit, ce montant à estimer représente la différence entre un montant de sinistre supposé maximal que la société peut envisager de payer et un montant de sinistre moyen.

Dans ce mémoire, nous nous sommes intéressés au cas précis du choc de sinistralité impactant le risque arrêt de travail. Le but de ce mémoire est de mesurer les effets qu'un tel stress peut avoir sur le portefeuille à l'aide de différents indicateurs à CNP.

La première partie de ce mémoire présente les notions de l'assurance emprunteur.

La deuxième partie, présente les données mises à notre disposition et les différentes étapes de retraitement afin de construire la base de données de sinistres et celle des assurés.

Dans les trois parties qui suivent, nous avons déterminé les trois composantes (taux d'incidence, durée résiduelle de maintien en arrêt de travail, montant de rente versée en cas de sinistre) permettant d'estimer la charge totale de sinistres :

- ✓ Dans la troisième partie, nous avons déterminé à partir des observations à disposition les taux d'incidence en arrêt de travail en fonction de l'âge. Puis la théorie des copules a été appliquée afin de simuler des nouveaux taux d'incidence pour effectuer par la suite les tests de sensibilité.
- ✓ Dans la quatrième partie, nous avons étudié les lois de maintien en arrêt de travail afin de déterminer les durées résiduelles pour chaque partenaire.
- ✓ Enfin, dans la cinquième partie, nous avons évalué le montant de rente moyen versé dans le cas de la survenance d'un sinistre.

Une fois ces trois indicateurs déterminés, il a été possible d'estimer la charge totale moyenne de sinistres ainsi que la charge totale maximum déterminée à partir des simulations.

Dans la dernière partie du mémoire, l'ensemble des résultats obtenus est analysé pour conclure ou non à un besoin de couverture en réassurance.

*Pour des raisons de confidentialité, les noms des partenaires ainsi que de nombreux résultats ne seront pas transmis dans ce mémoire. Il est donc conseillé au lecteur de se focaliser sur la méthodologie plus que sur les résultats eux-mêmes.*

## Chapitre 1

### 1. Caractéristiques des contrats d'assurance emprunteur

Ce chapitre décrit l'ensemble des caractéristiques des contrats d'assurance emprunteur : définition des contrats d'assurance emprunteur, description des garanties et rappel des dernières évolutions réglementaires.

#### 1.1. Définition de l'assurance emprunteur

Lorsqu'un particulier ou un professionnel souscrit une demande d'emprunt bancaire à une institution financière, cette dernière sollicite le recours à une assurance emprunteur.

Il s'agit d'un contrat d'assurance qui garantit la prise en charge partielle ou totale du remboursement du prêt en cas de survenance d'événements limitant ou empêchant l'emprunteur de respecter son engagement (décès, accident, chômage etc.).

Bien que l'assurance emprunteur ne soit pas une obligation légale, les établissements de crédit considèrent une telle assurance comme indispensable car elle constitue pour eux une garantie de bonne fin.

Par conséquent, l'assurance emprunteur préserve à la fois l'emprunteur et les établissements de crédit.

Cette assurance s'adresse à toute personne souscrivant un crédit, et concerne tous les types de crédit : immobilier, professionnel et de consommation. Ainsi, nous allons détailler les différents aspects de l'assurance emprunteur :

- Les garanties.
- Les types de souscription.
- La distribution.
- L'évolution du cadre juridique.

#### 1.2. Types de prêts

Il existe différents types de prêts selon le projet que l'emprunteur cherche à financer.

### 1.2.1. Le prêt immobilier

Le prêt immobilier désigne un crédit permettant le financement d'un bien immobilier (logement, immeuble ou terrain) ainsi que les travaux et/ou l'opération de construction qui y sont associés.

Ce bien peut être à usage d'habitation (dans le cas d'une résidence principale ou secondaire) ou à usage mixte, c'est-à-dire professionnel et d'habitation (comme souvent dans le cas d'un investissement locatif). Il s'agit d'un prêt à long terme, ainsi la durée du remboursement varie généralement entre 10 et 25 ans.

Cet emprunt, destiné à couvrir tout ou une partie d'un projet immobilier, est cependant soumis à des conditions imposées par les banques : la solvabilité du demandeur de crédit, c'est-à-dire l'aptitude de celui-ci à honorer ses dettes, et ce, quelle que soit l'échéance de ces dernières (court, moyen et long terme), est systématiquement contrôlée par l'institution financière et est indispensable à l'octroi du crédit immobilier.

Afin de s'assurer de cette solvabilité, l'institution financière va s'appuyer sur l'évaluation de différents éléments, notamment l'apport personnel et le taux d'endettement.

L'apport personnel représente une garantie de la solvabilité de l'emprunteur : il s'agit d'une somme d'argent que l'emprunteur doit posséder avant même de contracter l'emprunt et qui a pour but de compléter l'emprunt afin de soulager la banque d'une partie de l'investissement. Généralement, les établissements prêteurs exigent un apport personnel d'au moins 10% du montant total du projet immobilier.

La solvabilité de l'emprunteur repose sur le calcul du taux d'endettement, qui est un indicateur financier déterminant le montant pouvant être consacré au remboursement des charges financières sans être surendetté.

### 1.2.2. Le prêt professionnel

Le prêt professionnel désigne un crédit permettant le financement de biens, à usage professionnel exclusivement, qu'ils soient matériels, immatériels ou financiers.

Par conséquent, ce crédit permet la réalisation de certains investissements nécessaires au démarrage ou au développement de l'entreprise et est strictement réservé aux entreprises et aux entrepreneurs (ou entreprises individuelles).

Les types de besoin d'un prêt professionnel sont :

- Besoin en trésorerie (pour une entreprise ne possédant pas assez de liquidités).
- Besoin en équipement (matériel informatique, fourniture de bureau, etc...)
- Besoin immobilier (locaux, entrepôts ...)

Les établissements prêteurs exigent cependant des garanties lorsqu'elles octroient des prêts professionnels.

Les principales garanties sont l'hypothèque et le nantissement.

L'hypothèque stipule qu'en cas de défaillance de paiement, l'organisme prêteur devient le propriétaire du bien immobilier.

Le nantissement permet à la banque de récupérer son argent en cas de défaillance de l'emprunteur en vendant des biens incorporels (droit de propriété, brevet, marque).

### 1.2.3. Le prêt à la consommation

Le prêt à la consommation (ou crédit consommation) désigne un crédit destiné au financement de biens de consommation et services, hors activité professionnelle ou achat de biens immobiliers.

Il s'agit donc d'un prêt réservé aux particuliers.

Il existe trois types de crédits à la consommation :

- le crédit affecté, contracté par l'achat d'un bien défini uniquement (une voiture, par exemple)
- besoin en équipement (matériel informatique, fourniture de bureau . . .),
- le crédit non affecté dont le montant est à la libre utilisation de l'emprunteur (mariage, voyage).

Le montant du prêt est compris entre 200 et 75000€.

### 1.3. Modes de remboursement

Les modalités de remboursement du crédit sont déterminées à l'avance. Pour rembourser un emprunt, il faut rembourser le capital mis à disposition par le prêteur ainsi que les intérêts du prêt. Pour cela, différents modes de remboursement existent.

#### 1.3.1. Le remboursement à échéance constante

Sous cette modalité, le remboursement du prêt s'effectue par des échéances identiques pendant toute la durée de l'emprunt.

#### 1.3.2. Le remboursement IN FINE

Sous cette modalité, le remboursement du capital emprunté s'effectue une seule fois, à la fin du contrat. Pendant la durée du contrat uniquement les intérêts et l'assurance sont remboursés.

### 1.3.3. Le différé d'amortissement

Sous cette modalité, le remboursement du prêt s'effectue en plusieurs étapes : pendant une durée déterminée, l'emprunteur paie seulement les intérêts (ainsi que les frais relatifs à son prêt). Le remboursement du capital sera lui, effectué plus tard, à la date définie au contrat.

## 1.4. Types de contrat

Dans le cadre de la souscription d'une assurance emprunteur, deux contrats peuvent être proposés : un contrat de groupe ou un contrat individuel.

### 1.4.1. Le contrat de groupe

Le contrat de groupe est une assurance emprunteur collective, systématiquement proposée par l'établissement de crédit à tous ses clients lors de la souscription du contrat de prêt.

Il s'agit d'un contrat s'appuyant sur la mutualisation des risques de tous les emprunteurs de la banque, quels qu'ils soient (jeunes ou séniors, fumeurs ou non-fumeurs, exerçant un métier à risque ou non).

Par conséquent, sa prime est fixe et pénalise un grand nombre de souscripteurs.

Son adhésion reste cependant simplifiée car elle ne nécessite aucune démarche extérieure.

### 1.4.2. Le contrat individuel

Le contrat individuel est une assurance emprunteur personnalisée, externe à l'établissement de crédit, souscrite auprès d'un établissement indépendant.

Il est autorisé depuis 2010 par la mise en place de la loi Lagarde, permettant la délégation d'assurance et présente l'avantage d'une prime sur-mesure, adaptée à la situation personnelle de l'emprunteur.

Ce type de contrat semble être la solution la plus intéressante pour les souscripteurs de jeunes âges en bonne santé, les non-fumeurs ainsi que ceux ne pratiquant pas de métiers à risques, dont le tarif sera par conséquent adapté.

## 1.5. Les garanties

Lors de la mise en place d'un contrat d'assurance de prêt, l'assurance va s'intéresser à l'assuré d'un point de vue professionnel, mais également personnel, afin de lui proposer une couverture optimale par le biais de garanties.

Si l'un des différents risques présentés ci-dessous survient, alors en fonction des garanties souscrites, l'assureur prend à sa charge en partie, voire en totalité, le remboursement des échéances du prêt ou du capital restant dû.

#### 1.5.1. La garantie décès

L'assurance de prêt comprend toujours une garantie décès. Cependant, elle fait l'objet d'un âge limite de garantie, c'est-à-dire que le risque décès n'est pas toujours couvert jusqu'au terme d'un prêt.

Cette garantie intervient lors du décès de l'assuré, afin de préserver le niveau de vie de sa famille et de leur éviter une éventuelle dette qui pourrait mettre cette famille en difficulté.

#### 1.5.2. La garantie PTIA

Lors de la perte totale et irréversible d'autonomie (PTIA), c'est-à-dire lorsqu'un individu est déclaré inapte (mentalement ou physiquement) à exercer un travail à la suite d'un accident ou d'une maladie, alors l'assureur remboursera la totalité du capital restant dû à la banque.

Pour que la perte totale et irréversible d'autonomie soit reconnue, il faut nécessiter le recours permanent à une tierce personne pour exercer les actes de la vie courante (se lever, se déplacer, se nourrir et se laver). La garantie PTIA concerne le taux d'invalidité maximal : 100% .

#### 1.5.3. La garantie IPT

L'invalidité permanente et totale (IPT) d'un assuré à la suite d'un accident ou à une maladie est le cas où la personne est dans un état physique ou mental le mettant dans l'impossibilité totale/permanente et présumée définitive de produire un travail susceptible de lui procurer un gain ou un profit. Dans ce cas, l'assureur remboursera le capital restant dû de l'emprunt immobilier de l'assuré, à hauteur du pourcentage de garantie souscrite, en cas d'invalidité permanente totale. La garantie IPT concerne le taux d'invalidité entre 66% et 100%.

La garantie IPT prendra en charge la mensualité de l'emprunt de l'assuré, selon la quotité choisie.

#### 1.5.4. La garantie IPP

La garantie invalidité permanente partielle (IPP) permet de couvrir l'emprunt de l'assuré dans le cas où l'assuré se trouve en invalidité, ce qui l'empêcherait d'exercer sa profession et ainsi de ne plus percevoir de revenus pour rembourser son emprunt.

La garantie IPP concerne le taux d'invalidité entre 33% et 66%.

La garantie IPT prendra en charge la mensualité de l'emprunt de l'assuré, selon la quotité choisie.

#### 1.5.5. La garantie ITP

La garantie arrêt de travail partielle (ITP) concerne les personnes qui sont en arrêt de travail partiel.

Elle permet le remboursement des échéances par l'assureur jusqu'à 50% en cas de reprise thérapeutique du travail à temps partiel de l'emprunteur.

#### 1.5.6. La garantie ITT

La garantie incapacité temporaire totale de travail permet de couvrir l'assuré en cas d'accident ou de maladie, lorsqu'il est dans l'incapacité d'exercer son activité professionnelle suite à un arrêt de travail.

En matière d'indemnisation, l'assurance prend en charge les mensualités du prêt sur la période d'incapacité de travailler.

La prise en charge de l'assureur s'effectue sur une durée maximale de 3 ans et cesse dès la reprise de l'activité.

#### 1.5.7. La garantie perte d'emploi

La garantie perte d'emploi ou chômage concerne les personnes en cas de licenciement et lorsqu'elles perçoivent une indemnité chômage. L'assureur prend alors en charge les échéances de remboursement de son prêt.

Cette garantie est disponible seulement pour les salariés en CDI. De plus en cas de démission, de chômage partiel ou de rupture conventionnelle de contrat, cette garantie ne peut pas s'appliquer.

Elle s'applique dans le cadre d'un licenciement ouvrant droit aux allocations de Pôle Emploi.

## 1.6. Limites des garanties

### 1.6.1. Exclusions générales de garantie (ou exclusions légales)

Dans le cadre du contrat d'assurance, les conditions d'exclusion générale sont communes à toutes les compagnies d'assurance et s'appliquent quel que soit le profil du souscripteur et le type de contrat emprunteur. Elles sont fixées par le Code des Assurances.

Les exclusions générales concernent les conséquences :

- de faits exceptionnels, tels qu'une guerre, une émeute ou encore des actes de terrorisme,
- de faits volontaires comme le suicide de l'assuré, son décès suite à une maladie connue non déclarée dans le questionnaire de santé ou encore des délits ou tentatives d'escroquerie et de fraudes à l'assurance.

### 1.6.2. Exclusions particulières de garantie (ou exclusions contractuelles)

Avant toute souscription, le profil de l'assuré est évalué : les informations personnelles telles que l'âge, l'état de santé ou encore ses activités professionnelles ou de loisirs afin d'évaluer le risque qu'encourt l'assureur.

Les conditions d'exclusions particulières propres à chaque compagnie d'assurances peuvent différer d'un contrat à l'autre car elles dépendent du profil de risque et de l'appétence au risque de l'assureur.

Dans certains cas, ces clauses peuvent donner lieu au paiement d'une surprime. Parmi les exclusions contractuelles, on retrouve :

- L'âge maximal d'adhésion, généralement compris entre 65 et 79 ans selon les assurances.
- L'atteinte par une maladie grave ou systémique telle que le cancer, le diabète, le sida ou les maladies cardiaques.
- Les activités professionnelles à risques (avec maniement d'une arme à feu, manipulation de substances dangereuses, conditions extrêmes) : sapeurs-pompiers, guides de haute montagne, policiers...
- La pratique d'un sport dangereux, comme le saut à l'élastique, le parachutisme, la plongée sous-marine, les sports de combat ou même un sport exercé à haut niveau (en tant que professionnel).
- Les sinistres consécutifs à un séjour à l'étranger.

## 1.7. Analyse du marché de l'assurance emprunteur

L'assurance emprunteur a un chiffre d'affaires de 10,284 Md€ d'euros en 2020 en hausse de 14,9% depuis 2016.

L'assurance emprunteur représente donc un des domaines de l'assurance de personne les plus importants sur le marché français.

CNP Assurances est le leader de l'assurance emprunteur en France. Ce marché représente 9,8 Md€ de chiffre d'affaires en 2019 et le chiffre d'affaires de CNP Assurances est de 2,7 Md€.

Sur les 9,8 Md€, près de 7 milliards d'euros correspondent à la couverture des prêts immobiliers, marché en croissance régulière depuis de nombreuses années (+5,4% en 2019) à l'image de l'encours des crédits immobiliers.

CNP Assurances, qui détient 28% des parts de marché, a confirmé sa position de leader de l'assurance emprunteur en 2019 avec une progression de 4,6% reflétant le succès des alliances avec les réseaux bancaires, dont la nouvelle convention de partenariat mise en place avec le groupe BPCE.

En 2019, les contrats de groupe représentaient 91,2% du chiffre d'affaires de CNP Assurances sur cette activité, contre 8,8% pour les contrats individuels.

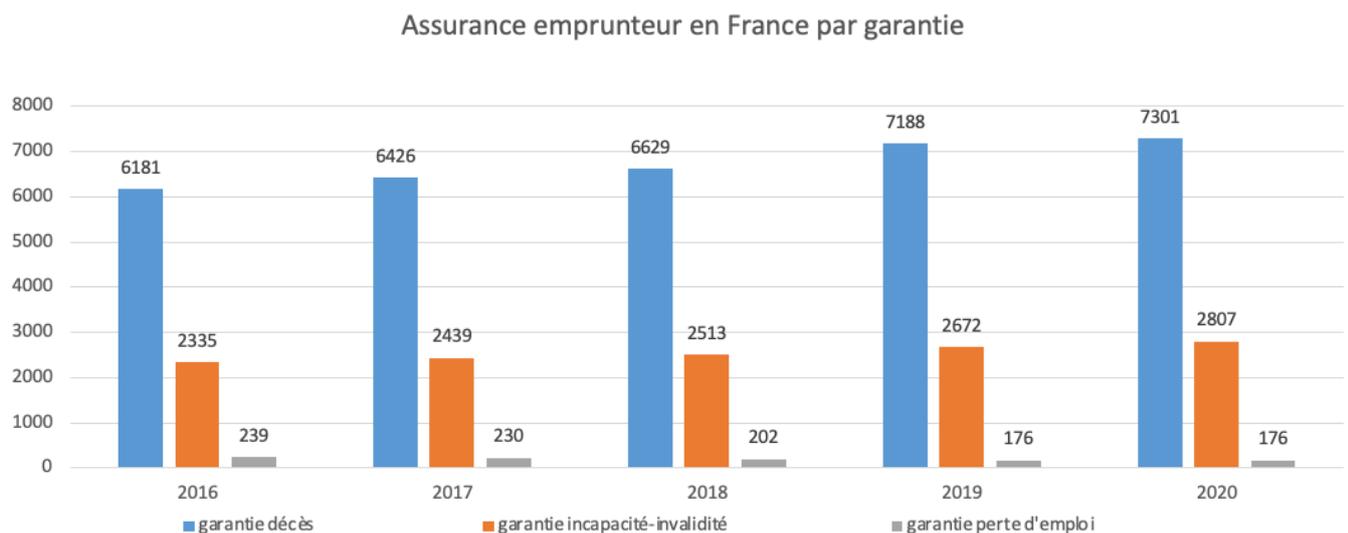


Figure 5 : Chiffres d'affaires de l'assurance emprunteur en France par garantie en milliards d'euros

Source : FRANCE ASSUREURS

Les cotisations d'assurance emprunteur, se répartissent de la façon suivante selon le type de prêt :

- 71% pour les prêts immobiliers (soit 7,1 Md€),
- 22% pour les prêts à la consommation (soit 2,2 Md€),
- 7% pour les prêts professionnels (soit 0,7 Md€).

Le marché des prêts immobiliers a connu une croissance régulière depuis de nombreuses années.

Selon les données de la Banque de France, en 2019, le volume des nouveaux crédits à l'habitat accordés aux ménages en France a augmenté de 5,4% par rapport à l'année précédente, pour atteindre près de 258 Md€.

Plus précisément, la croissance des prêts immobiliers en France a été de 5,4% en 2019 par rapport à l'année précédente, de 4,4% en 2018, de 4,6% en 2017 et de 3,9% en 2016.

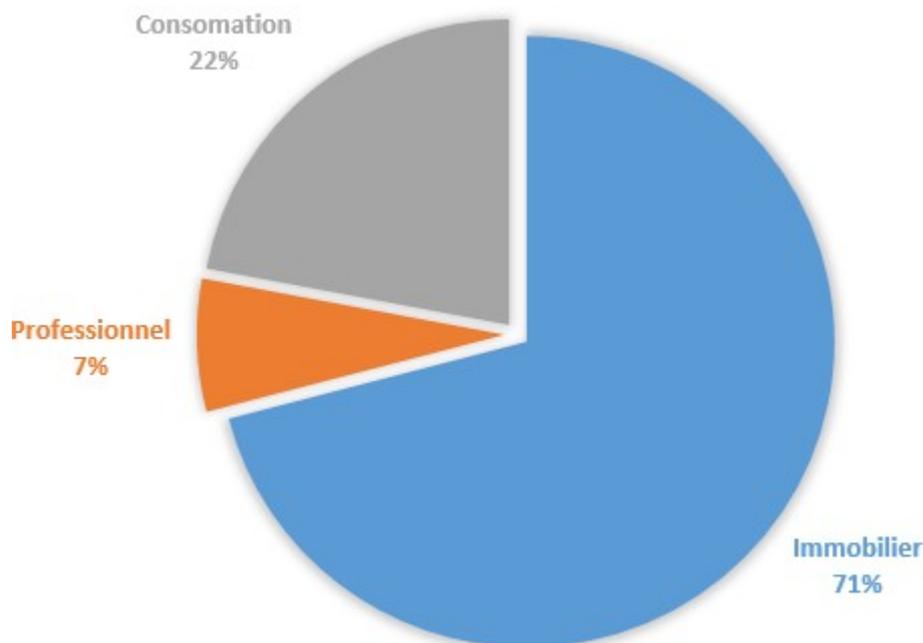


Figure 6 : Répartition moyenne (2016-2020) des cotisations d'assurance emprunteur par type de prêt

Source : FRANCE ASSUREURS

### 1.8. Aspects réglementaires

Au cours des quinze dernières années, l'assurance emprunteur a beaucoup évolué notamment par la mise en place de lois successives, qui ont impacté plus particulièrement les assurances de crédits immobiliers en renforçant le pouvoir du consommateur.

#### 1.8.1. Loi Lagarde

En juillet 2010, la mise en place de la loi Lagarde permet la délégation d'assurance. Par conséquent, elle permet au consommateur de choisir librement son assurance emprunteur, qui peut être externe à celle de la banque, sans que le taux du prêt ne soit impacté.

#### 1.8.2. Loi Hamon

En juillet 2014, la résiliation de l'assurance emprunteur est désormais possible au cours des 12 premiers mois qui succèdent à sa souscription grâce à la loi Hamon.

#### 1.8.3. Amendement Bourquin

En janvier 2018, l'amendement Bourquin va renforcer la loi précédente puisqu'elle va étendre le droit de résiliation : il est désormais possible de résilier le contrat d'assurance non seulement au cours de la première année de souscription mais également tous les ans, à chaque date d'anniversaire de cette dernière.

#### 1.8.4. Loi Lemoine

Enfin, la Loi Lemoine, promulguée le 28 février 2022, va considérablement réformer le marché de l'assurance emprunteur en favorisant son accessibilité au grand public.

Pour cela, elle a recours à trois dispositions :

- Le droit de résiliation du contrat d'assurance est désormais possible à tout moment et sans frais.
- Aucun questionnaire de santé, ni examen médical n'est demandé à l'assuré (pour certains prêts immobiliers notamment si le crédit s'élève à 200 000€ maximum) : cela permet, sous certaines conditions, à tout consommateur d'avoir recours à un prêt immobilier et ce quel que soit son état de santé.
- Renforcement de la convention : toute personne ayant ou ayant eu des problèmes de santé a la possibilité d'emprunter aux institutions financières, sous certaines conditions en faisant appel à la convention d'AERAS.
- Le droit à l'oubli est renforcé : l'assuré n'a plus l'obligation de déclarer à l'assureur une pathologie antérieure.

Actuellement, lorsque le cancer a été diagnostiqué avant 21 ans, le droit à l'oubli s'applique 5 ans à compter de la fin du protocole thérapeutique. Lorsque le cancer a été diagnostiqué après 21 ans, le droit à l'oubli s'applique 10 ans à compter de la fin du protocole thérapeutique.

Le droit à l'oubli s'appuie sur la convention d'AERAS.



Figure 7 : Frise chronologique des récentes lois en assurance emprunteur.

## Chapitre 2

### 2. Construction des bases de données

Dans ce chapitre, nous allons tout d'abord rappeler la structure des données brutes qui nous ont été fournies, puis nous allons décrire les différents traitements utilisés afin d'obtenir deux bases de données cohérentes et exploitables pour notre étude.

L'objectif de ce chapitre est d'avoir une base de données des assurés et une base de données des sinistres pour les deux partenaires retenus dans le cadre de ce mémoire.

Nous disposons de plusieurs tables de données de sinistres et d'assurés dite "tête par tête". Les bases de données ont été construites à l'aide du logiciel SAS.

#### 2.1. Base de données des assurés

##### 2.1.1. Descriptif des données brutes mises à disposition

Les données relatives aux assurés possédant des contrats emprunteur sont contenues dans les tables suivantes :

<i>Base de données</i>	<i>Nombre d'observations</i>
MIP.BUR	254 325 000
PPMUV55_VISION_ASSURE	646 639 200

*Tableau 3 : nombre d'observations par base de données d'assurés*

## 2.1.2. Variables de chaque base

La base MIP\_BUR contient 95 variables. Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
MIP_ID_ASSURE	Numérique	Identifiant de l'assuré
MIP_NAT_PRET	Numérique	Nature du prêt
MIP_MNT_PRIM_AN_GLOB	Numérique	Montant de la prime annuelle globale
MIP_DT_NAISS	Date	Date de naissance de l'assuré
MIP_DT_ARRETE	Date	Date d'inventaire
MIP_CAP_INIT	Numérique	Montant du Capital initial
MIP_CRD	Numérique	Montant capital restant dû à la dernière échéance
MIP_DUR_TOTALE_PRET	Numérique	Durée totale du prêt en mois
MIP_DUR_RESIDU	Numérique	Durée restante du prêt en mois
MIP_QUOTITE_GLOBALE	Numérique	Part assurée (par garantie)

Tableau 4 : Descriptif des variables retenues issues de la table MIP\_BU

La base PMUV55\_VISION\_ASSURE contient 309 variables. Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
ID_ASSURE	Numérique	Identifiant de l'assuré
NAT_PRET	Numérique	Nature du prêt
MNT_PRIM_AN_GLOB	Numérique	Montant de la prime annuelle globale
A_DT_NAISS	Date	Date de naissance de l'assuré
DATE_ARRETE	Date	Date d'inventaire
MNT_CAP_INIT	Numérique	Montant du Capital Initial
MNT_CRD_AP_DERN_ECH	Numérique	Montant du capital restant dû à la dernière échéance
DUR_TOTALE_PRET	Numérique	Durée totale du prêt en mois
DUREE_RESIDUELLE	Numérique	Durée restante du prêt en mois
QUOTITE_GLOBALE	Numérique	Part assurée (par garantie)

Tableau 5 : Descriptif des variables retenues issues de la table PMUV55\_VISION\_ASSURE

## 2.2. Base de données des sinistres

### 2.2.1. Descriptif des données brutes mises à disposition

Les données relatives aux sinistres emprunteur sont contenues dans les tables suivantes :

Base de données	Nombre d'observations
C7SBASE.C7s_BUR	33 632 070
TDO_R_SIACC_CONT_CATALOG	10 121
TDO_D_SIACS_SINIS_PRET	11 965 002
TDO_D_SIACS_COMMUN_SINIS_PRET	4 129 122

Tableau 6 : Nombre d'observations par base de données de sinistres

### 2.2.2. Variables de chaque base

La base C7SBASE.C7s\_BUR contient 71 variables.

Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

Variable	Type	Descriptifs
ID_SYNT_BUR	Numérique	Identifiant du sinistre
DT_ARRETE	Date	Date d'arrêté
CD_PDT	Numérique	Code du partenaire
LIBELLE_RGP_PDT	Chaîne de caractère	Code déterminant la garantie
NUM_CONTRAT	Chaîne de caractère	Code déterminant le numéro du contrat
NAt_PRETT	Chaîne de caractère	Code déterminant la nature du prêt

Tableau 7 : Descriptif des variables issues de la base C7SBASE.C7s\_BUR

La base TDO\_R\_SIACC\_CONT\_CATALOG contient 21 variables.

Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
SIACC_CC_NUM_CONTRAT	Chaîne de caractère	Code déterminant le numéro du contrat
SIACC_CC_CD_PROD	Numérique	Code déterminant la garantie

Tableau 8 : Descriptif des variables retenues issues de la table TDO\_R\_SIACC\_CONT\_CATALOG

La base TDO\_D\_SIACS\_COMMUN\_SINIS\_PRET contient 16 variables. Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
CPR_NUM_SINIS	Numérique	Numéro du sinistre
CPR_NUM_DOSSIER	Numérique	Numéro du dossier
CPR_ID_SINIS	Chaîne de caractère	Identifiant du sinistre
CPR_DT_FIN_GART	Date	Date fin de garantie
CPR_NUM_CONTR	Chaîne de caractère	Code déterminant le numéro du contrat
CPR_DT_NAISS_BENEF	Date	Date de naissance de l'assuré
CPR_DT_SURVENANCE	Date	Date de survenance du sinistre
CPR_NUM_RGP	Numérique	Code déterminant la garantie
CPR_CD_SYST		

Tableau 9 : Descriptif des variables retenues issues de la base TDO\_D\_SIACS\_COMMUN\_SINIS\_PRET

La base SIACS.TDO\_D\_SIACS\_SINIS\_PRET contient 18 variables.

Nous avons seulement retenu les variables suivantes pour les besoins de l'étude :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
PRE_DT_ARRETE	Date	Date d'arrêté
PRE_NUM_RGP	Numérique	Code déterminant la garantie
PRE_CD_CLOTUR	Numérique	Indique si le sinistre est clôturé ou non
PRE_NUM_SINIS	Chaîne de caractère	Numéro du sinistre
PRE_MNT_SINIS_RGP	Numérique	Code déterminant la garantie
PRE_DT_CLOTUR	Date	Date de clôture du sinistre
PRE_DT_DERN_PAIEMT	Date	Date de dernier paiement
PRE_DT_FIN_DERN_PERIO_IND	Date	Date fin de période d'indemnisation
PRE_MNT_PM_PART1	Numérique	Montant provision mathématique 1
PRE_MNT_PM_PART2	Numérique	Montant provision mathématique 2

Tableau 10 : Descriptif des variables retenues issues de la base SIACS.TDO\_D\_SIACS\_SINIS\_PRET

### 2.3. Traitement des données

La période d'étude est comprise entre 2015 et 2020.

Le périmètre d'étude concerne les données emprunteur, plus particulièrement les emprunts immobiliers de deux partenaires importants.

#### 2.3.1. Base de données des assurés

On cherche à calculer de nouvelles variables afin d'obtenir la base finale des données des assurés.

Les nouvelles variables calculées sont :

- L'âge de l'assuré à l'inventaire : compris entre 20 à 65 ans,
- Le nombre d'assurés,
- La moyenne des primes payées par les assurés,
- La moyenne du capital initial,
- La moyenne du capital restant dû,
- La moyenne des durées des prêts,
- La moyenne des durées restantes des prêts.

Voici le descriptif de ces variables :

<i>Variable</i>	<i>Type</i>	<i>Descriptifs</i>
age_arrete	Numérique	Age à l'arrêté
date_arrete	Numérique	Date à l'arrêté
nb_assure	Numérique	Nombre d'assurés
moy_prime	Numérique	Moyenne des primes
moy_CI	Numérique	Moyenne du capital Initial
moy_CRD_YC_QUOT	Numérique	Moyenne du Capital restant dû
moy_tot	Numérique	Moyenne de la durée du prêt
dur_res	Numérique	Moyenne de la durée restante du prêt

Tableau 11 : Descriptif des variables supplémentaires

Les données concernant les assurés du partenaire n°1 sont stockées dans PMUV55\_VISION\_ASSURE pour toutes les années.

Ainsi, on sélectionne la fenêtre d'observation, le partenaire et la nature du prêt.

Les données des assurés du partenaire n°2 sont stockées dans MIPBASE.MIP\_BUR pour l'année 2015 puis dans PMUV55\_VISION\_ASSURE pour les autres années. Il faut traiter deux cas pour obtenir une base de données unique pour le partenaire n°2.

Dans un premier temps, on traite l'année 2015, il faut donc extraire les données de MIP-BASE.MIP\_BUR.

Puis, on traite les années allant de 2016 à 2020, il faut donc extraire les données de PMUV55\_VISION\_ASSURE.

Pour finir, après avoir sélectionné la période d'observation, la nature de l'emprunt et le partenaire, il faut concaténer ces deux bases de données. On obtient ainsi une base de données unique pour le partenaire n°2.

### 2.3.2. Traitement de la base de données des assurés

Après avoir obtenu une base de données unique pour chaque partenaire, des vérifications sont effectuées.

La première vérification consiste à détecter les doublons dans les données d'inventaire. Si une ligne présente une correspondance parfaite avec une autre ligne (ou plusieurs autres lignes) à la même date d'inventaire, elle est exclue de la base de données finale.

La deuxième étape de vérification porte sur les données manquantes. Si certaines données, telles que le montant du capital (initial ou restant dû), ne sont pas renseignées, elles sont remplacées par la moyenne de tous les montants du capital (initial ou restant dû). Les données manquantes ne seront pas prises en compte dans le calcul ultérieur du taux d'incidence lors des prochaines étapes de notre étude.

Si la date de naissance de l'assuré n'était pas renseignée, il a été décidé de supprimer cette observation. Pour rappel, l'âge de l'assuré à l'inventaire a été calculé en soustrayant la date de naissance de l'assuré à la date d'inventaire.

Enfin, la dernière vérification concerne les valeurs aberrantes. Dans ce cas, nous pouvons rencontrer deux types de valeurs aberrantes : lorsque le montant du capital initial est supérieur au montant du capital restant dû et lorsque la durée restante du prêt est supérieure à la durée totale du prêt. Dans de tels cas, la moyenne des montants du capital initial et du capital restant dû est utilisée pour remplacer les valeurs aberrantes.

La sélection du risque lié aux arrêts de travail dans la section des prêts immobiliers, couplée au traitement des données, a conduit à l'élimination de plus de 21% des informations relatives au partenaire n°1 et 26% pour le partenaire n°2. De ce fait, nous bénéficions maintenant de bases de données plus robustes pour chacun de nos partenaires, ce qui est essentiel pour la suite de l'étude.

### 2.3.3. Base de données finale des assurés

On obtient les bases de données finales des assurés pour chacun des deux partenaires.

Dans chaque base de données finale, on peut retrouver les variables :

- Périmètre : nom du partenaire,
- Risque\_IT , Risque\_DC, Risque\_Chommage : fonction indicatrice,
- Nature du prêt : on choisit de regarder le prêt immobilier,
- Version : dernière version des données,
- Nombre de prêts et nombre d'assurés.

### 2.3.4. Base de données des sinistres

Dans un premier temps, afin d'obtenir un maximum d'informations sur chaque sinistre survenu, il faut faire une jointure entre les bases. Elle regroupera les quatre bases de données de sinistres en fonction du numéro de chaque sinistre.

Une fois cette nouvelle base de données créée, on sélectionne la fenêtre d'observation souhaitée, puis le partenaire souhaité, ensuite les sinistres représentant l'arrêt de travail et enfin l'âge de l'assuré ayant eu un sinistre.

### 2.3.5. Traitement de la base de données des sinistres

Une fois que nous avons obtenu une base de données unique pour chaque partenaire en ce qui concerne les sinistres, plusieurs vérifications doivent être effectuées.

La première vérification consiste à détecter les doublons dans les données de sinistre. Si une ligne présente une correspondance parfaite avec une autre ligne (ou plusieurs autres lignes) à la même date d'arrêt, elle est considérée comme un doublon et exclue de la base de données.

La deuxième étape de vérification concerne les données manquantes. Si certaines informations, telles que le montant du sinistre, sont manquantes, elles sont remplacées par la moyenne de tous les montants de sinistres. Ces données manquantes ne sont pas essentielles pour le calcul des taux d'incidence qui seront utilisés dans notre étude ultérieure.

Cependant, si des informations cruciales telles que la date de naissance de l'assuré ne sont pas renseignées, nous avons pris la décision de ne pas inclure la ligne correspondante.

Les retraitements effectués précédemment ont supprimé 2 % des observations de la base de données brutes du partenaire numéro 1 et 9% des observations de la base de données brutes du partenaire numéro 2.

### 2.3.6. Base de données finale des sinistres

On obtient les bases de données finales des sinistres pour chacun des deux partenaires.

Dans chaque base de données finale, on peut retrouver les variables :

- Périmètre : nom du partenaire,
- Année de survenance : entre 2015 et 2020,
- Année de naissance,
- Nombre de sinistres,
- Montant total du sinistre : eu euros,
- Âge à la survenance.

Ces différentes bases de données vont permettre d'étudier l'incidence en arrêt de travail.

## Chapitre 3

### 3. Étude statistique de l'incidence en arrêt de travail

L'objectif de cette partie est de présenter la méthode retenue afin de calculer les taux d'incidence pour le risque arrêt de travail.

Le périmètre étudié correspond aux prêts immobiliers.

Dans un premier temps, nous présenterons les taux d'incidence obtenus par partenaire et par âge. Puis, nous analyserons la relation existante entre l'âge de l'assuré et l'incidence du risque arrêt de travail.

Enfin, de nouveaux taux d'incidence seront simulés à l'aide de la théorie des copules afin de tester la sensibilité des indicateurs économiques retenus.

#### 3.1. Approche envisagée

En premier lieu, nous allons calculer le taux d'incidence en arrêt de travail selon l'âge à partir des observations issues du portefeuille de prêts immobiliers de deux partenaires de poids statistiquement représentatifs.

La période d'observation se situe entre les années 2015 à 2020 et nous serons amenés à calculer également un taux d'incidence moyen par âge sur toute cette période.

Dans un premier temps, l'objectif sera d'étudier les différentes façons de modéliser le lien existant entre l'âge et l'incidence.

Dans un deuxième temps, la théorie des copules sera utilisée pour simuler de nouveaux couples : « (Age, Taux d'incidence) ».

#### 3.2. Taux d'incidence par âge

Le taux d'incidence par âge correspond au ratio entre le nombre de personnes entrant en arrêt de travail dans l'année d'observation parmi les assurés d'un âge donné observé au sein d'un portefeuille.

Le taux d'incidence s'exprime en pourcentage.

$$T_{i,x} = \frac{S_{i,x}}{A_{i,x}}$$

Avec :

$S_{i,x}$  : Nombre de sinistres survenus à l'âge  $i$  pendant l'année  $x$ .

$A_{i,x}$  : Nombre d'assurés d'âge  $i$  couverts pour le risque arrêt de travail pendant l'année  $x$ .

$i$  : Âge de l'assuré.

$x$  : Année de survenance.

Les graphiques présentent pour chaque partenaire les taux d'incidence obtenus par année en fonction de l'âge à l'incidence.

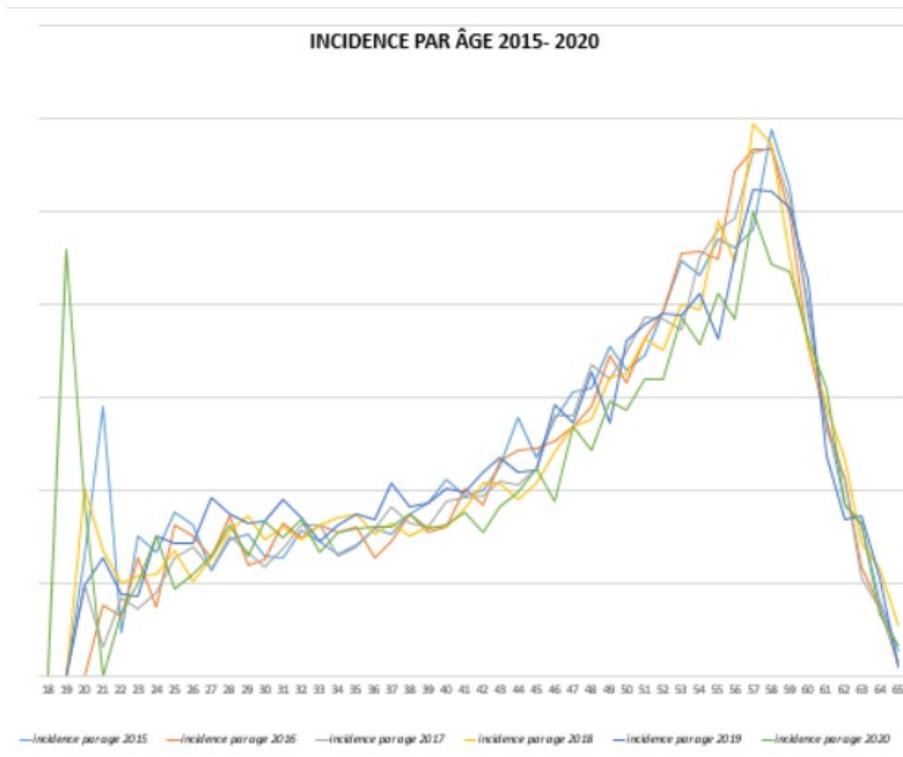


Figure 8 : Graphique d'incidence par âge pour le partenaire n°1.

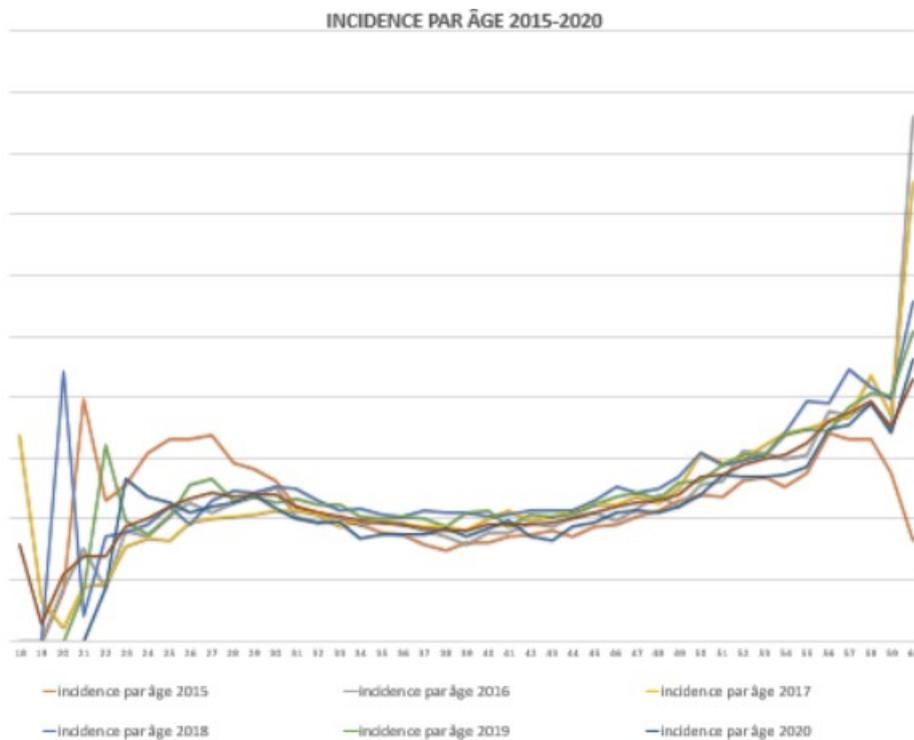


Figure 9 : Graphique d'incidence par âge pour le partenaire n°2.

On constate pour les deux partenaires étudiés que les taux d'incidence augmentent avec l'âge. L'augmentation devient « exponentielle » pour le partenaire n°2 au-delà de 58 ans. Ce constat ne peut être effectué pour le partenaire n°1 du fait sans doute d'un manque de données au-delà de 58 ans.

Avant 30 ans et après 55 ans, selon le partenaire, la volatilité est liée à la robustesse des valeurs du taux d'incidence expliquée probablement par le manque d'observations.

### 3.3. Taux d'incidence par année

Le taux d'incidence pour une année correspond au ratio du nombre total d'entrées en arrêt de travail quel que soit l'âge par le nombre total d'assurés présent cette même année.

Pour  $x$ , une année fixée, nous obtenons pour le taux d'incidence calculé tout âge confondu.

$$T_x = \frac{\sum_{i=1}^n S_{i,x}}{\sum_{i=1}^n A_{i,x}}$$

Ci-dessous, le tableau d'incidence par année pour chaque partenaire :

Année	2015	2016	2017	2018	2019	2020
Incidence	$T_{2015}$	$T_{2016}$	$T_{2017}$	$T_{2018}$	$T_{2019}$	$T_{2020}$

Tableau 12 : Taux d'incidence par année

*Pour des raisons de confidentialité, les taux ainsi obtenus ne sont pas communiqués.*

On positionne ces données sur un graphique avec en abscisse l'année et en ordonnée le taux d'incidence par année.

Voici les taux d'incidence annuels obtenus pour chaque partenaire

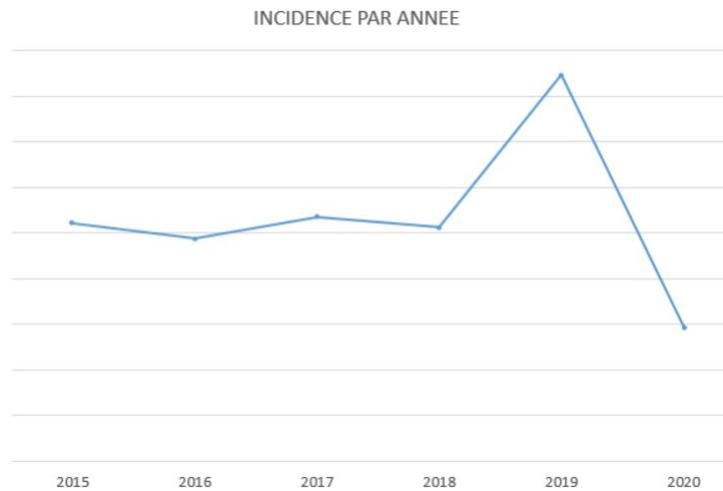


Figure 10 : Graphique d'incidence par année pour le partenaire n°1.

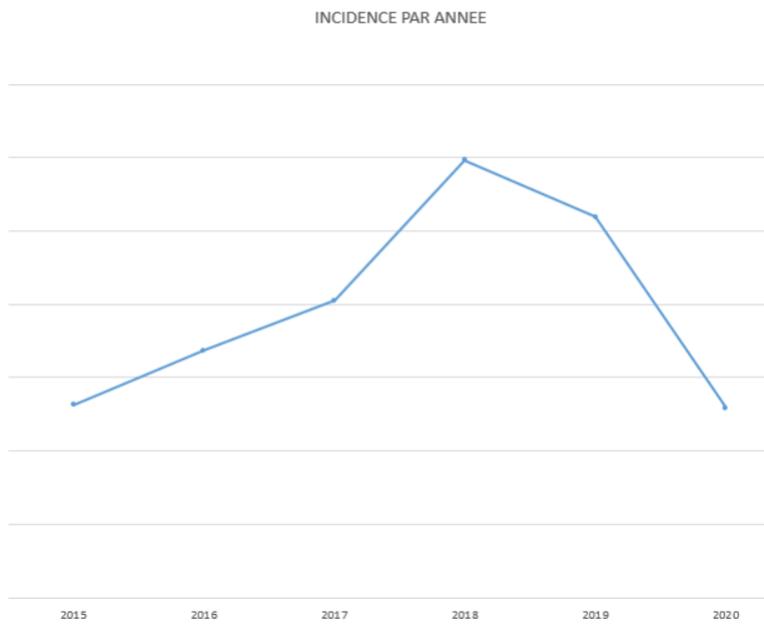


Figure 11 : Graphique d'incidence par année pour le partenaire n°2.

Pour le premier partenaire, le taux d'incidence par année augmente pour atteindre un maximum en 2019. Pour le deuxième partenaire, le maximum est atteint en 2018.

En 2020, les taux d'incidence pour les deux partenaires sont très bas en comparaison avec les années précédentes, en raison du fait que cette année est la plus récente et que tous les sinistres ne sont pas encore connus.

3.4. Taux d'incidence moyen par âge sur la période

La formule définissant l'incidence moyenne pour un âge donné  $i$  de 2015 à 2020 est:

$$T_{i, [2015:2020]} = \frac{\sum_{i=18}^n S_{i,2015} + \sum_{i=18}^n S_{i,2016} + \sum_{i=18}^n S_{i,2017} + \sum_{i=18}^n S_{i,2018} + \sum_{i=18}^n S_{i,2019} + \sum_{i=18}^n S_{i,2020}}{\sum_{i=18}^n A_{i,2015} + \sum_{i=18}^n A_{i,2016} + \sum_{i=18}^n A_{i,2017} + \sum_{i=18}^n A_{i,2018} + \sum_{i=18}^n A_{i,2019} + \sum_{i=18}^n A_{i,2020}}$$

Avec  $n=65$  pour le partenaire n°1 et  $n=60$  pour le partenaire n°2.

Les représentations graphiques de ces valeurs sont présentées ci-dessous.

*Pour des raisons de confidentialité, les taux ainsi obtenus ne sont pas communiqués*

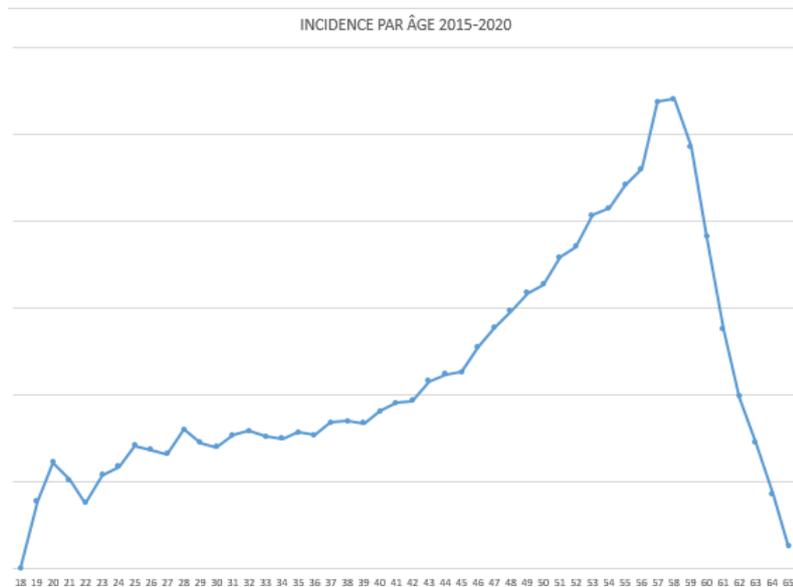


Figure 12 : Graphique d'incidence moyen par âge pour le partenaire n°1.

On constate globalement une forte croissance de l'incidence sur l'intervalle [18 ;57], suivie d'une très forte décroissance entre [58 ;65].

Pour le partenaire n°2, le graphique suivant :

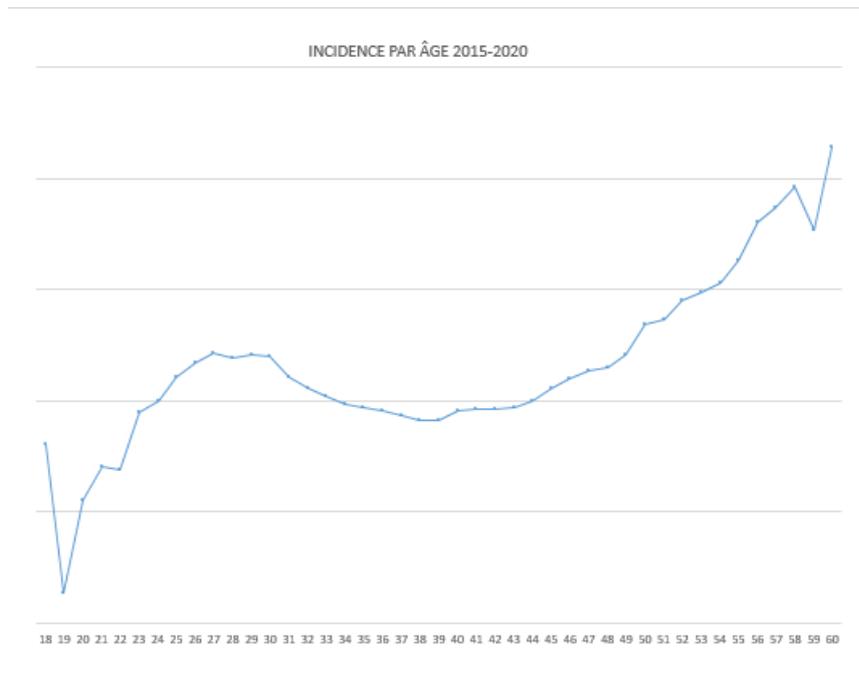


Figure 13 : Graphique d'incidence moyen par âge pour le partenaire n°1.

Les taux d'incidence sont particulièrement volatils avant l'âge de 20 ans et après 57 ans en raison du manque de données. Entre 20 et 27 ans, on observe une tendance à la hausse des taux d'incidence, suivie d'une décroissance entre 27 et 40 ans, puis d'une reprise de la croissance entre 40 et 57 ans.

### 3.5. Les différentes régressions

Dans cette partie, nous cherchons à vérifier s'il existe une relation linéaire entre l'âge et l'incidence.

De plus, il est usuel de tenter des régressions linéaires, et que les formes des figures 3.5 et 3.6 nous y incite.

#### 3.5.1. La régression linéaire simple

L'analyse de régression linéaire vise à établir une relation entre deux variables, où une variable (appelée variable dépendante ou variable à expliquer) est prédite en fonction de l'autre variable (appelée variable indépendante ou variable explicative). Le but de cette analyse est de comprendre comment la variable dépendante varie en fonction de la variable indépendante.

Soit  $n$  couples  $(x_i, y_i)$  constituant un  $n$ -échantillon d'observations indépendantes. On suppose que la régression est linéaire c'est dire que quel que soit  $i$  :

$$y_i = ax_i + b + \epsilon_i$$

Avec :

$y_i$  : variable aléatoire observée appelée variable à expliquer.

$x_i$  : variable aléatoire observée appelée variable explicative.

$\epsilon_i$  : l'erreur aléatoire du modèle.

$a$  : représente la pente de la droite de régression.

$b$  : représente le point  $x$  de la droite de régression avec l'ordonnée à l'origine.

Le modèle repose sur les hypothèses suivantes :

- 1-  $E[\epsilon_i] = 0$ , l'erreur est centrée.
- 2-  $E[\epsilon_i^2] = \sigma^2$ , la variance de l'erreur est constante.
- 3-  $\text{Cov}[\epsilon_i, \epsilon_j] = 0$ , avec  $i \neq j$ , les erreurs ne sont pas auto corrélées.
- 4-  $\text{Cov}[x_i, \epsilon_j] = 0$ , l'erreur n'est pas corrélée avec la variable exogène.
- 5- La variable exogène  $X_i$  n'est pas aléatoire.
- 6- Le modèle est linéaire en  $X$  par rapport aux paramètres.

Les six hypothèses ci-dessus sont les conditions requises pour que la régression linéaire soit valide et que les estimations des paramètres soient fiables.

L'existence d'une relation linéaire entre deux variables peut être évaluée en utilisant le coefficient de détermination ( $R^2$ ) et la  $p$ -value associée à la régression linéaire.

## 3.5.2. Analyse de la régression linéaire simple

Dans un premier temps, on regarde la significativité de la pente de la régression linéaire : la *p-value*.

La *p-value* est utilisée pour évaluer la significativité statistique d'une régression linéaire.

Une *p-value* inférieure à 0,05 permet de conclure à un lien linéaire significatif entre la variable réponse et la variable explicative.

On effectue une régression linéaire pour chaque partenaire et si on obtient des *p-value* < 0,05, on en conclut qu'il existe une relation linéaire reliant l'âge et l'incidence.

Après avoir obtenu la régression linéaire, il est nécessaire de mesurer sa qualité en utilisant le coefficient de détermination  $R^2$ . Cet indicateur permet d'évaluer l'adéquation entre le modèle et les données observées dans une régression linéaire simple.

Ce coefficient est calculé ainsi :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

Avec :

$y_i$  : La valeur de la mesure n°i.

$\hat{y}_i$  : La valeur prédite correspondante.

$\bar{y}_i$  : La moyenne des mesures.

$n$  : Le nombre de mesures.

Le coefficient de détermination pour le partenaire n°1 est de 35%, ce qui signifie que le modèle de régression explique 35% de la variation totale.

Le coefficient de détermination pour le partenaire n°2 est de 10%, ce qui signifie que le modèle de régression explique 10% de la variation totale.

Les résultats précédents présentent un  $R^2$  insuffisant surtout pour le partenaire n°2, ce qui indique que la régression linéaire simple n'est sans doute pas le modèle le plus adapté à nos données.

Cela implique que d'autres facteurs influencent la variable dépendante, mais ne sont pas inclus dans le modèle.

Dans notre cas, d'autres facteurs tels que la charge de travail, le type d'emploi, le niveau de stress ou encore les conditions de travail ne sont pas pris en compte dans le modèle de régression linéaire simple mais peuvent influencer la variable dépendante.

Par conséquent, nous allons nous intéresser à la régression linéaire segmentée (ou "par morceaux").

### 3.5.3. La régression linéaire par morceaux

La régression linéaire par morceaux, également appelée régression linéaire « segmentée », est une méthode de régression qui consiste à ajuster plusieurs segments de droite successifs à des parties distinctes d'un ensemble de données.

Soit  $Y$  une variable dépendante et  $X$  une variable indépendante.

Supposons que les données sont divisées en  $k$  segments, et que pour chaque segment  $i$ , la relation entre  $Y$  et  $X$  peut être approximée par une droite de la forme :

$$Y = a_i X + b_i$$

Pour  $x_{\min} < X < x_{\max}$  où  $a_i$  et  $b_i$  sont les coefficients de la droite pour le segment  $i$ , et  $x_{\min}$  et  $x_{\max}$  sont les valeurs minimales et maximales de  $X$  pour ce segment.

Le but de la régression linéaire par morceaux est d'estimer les coefficients  $a_i$  et  $b_i$  pour chaque segment de droite.

Ainsi, la régression linéaire par morceaux est une méthode de régression que nous utilisons lorsqu'il existe des "points de rupture" dans un jeu de données.

Nous cherchons à décrire la relation entre l'incidence moyenne (variable à expliquer) et l'âge (variable explicative).

Dans un premier temps, pour effectuer la régression linéaire par morceaux, nous allons commencer par visualiser les jeux de données.

Puis, pour ajuster le modèle de régression par morceaux, nous utiliserons la fonction `segmented()` du package 'segmented' dans R afin d'adapter un modèle de régression à nos jeux de données.

La fonction du package "segmented" permet d'estimer les paramètres de régression pour une relation linéaire par morceaux.

Une relation segmentée est définie par les paramètres de pentes et les points de ruptures où la relation linéaire change.

Cette fonction va tenter d'estimer un modèle de régression en déterminant une ou plusieurs relations segmentées.

Ainsi, pour visualiser le modèle final de régression par morceaux, nous serons amenés à utiliser les packages `plot()` dans R qui nous permettent d'afficher sous forme de graphique les résultats obtenus. On dispose de valeurs calculées sur des plages différentes : pour le partenaire n°1  $n=65-18 = 47$  et pour le partenaire n°2  $n=60-18=42$ .

Partenaire n°1

En effectuant la régression linéaire par morceaux sur le partenaire n°1, nous obtenons le graphique suivant :

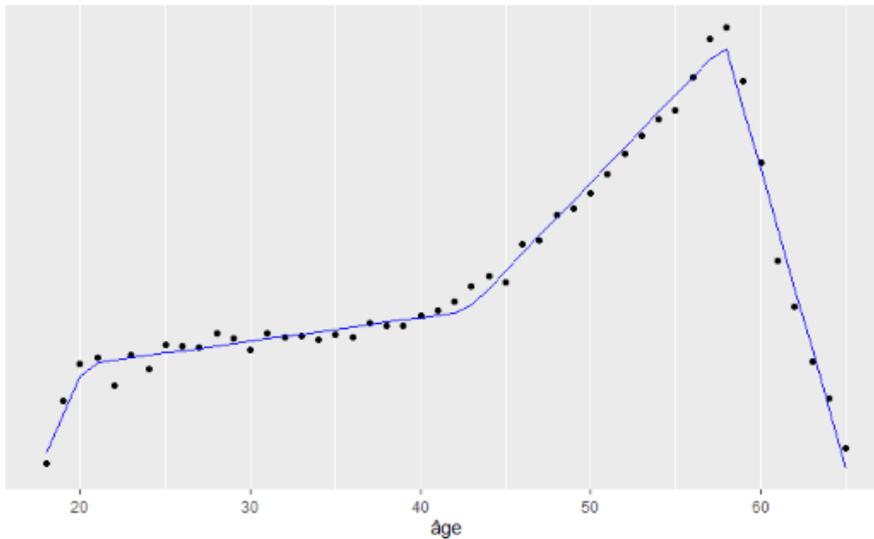


Figure 14 : La régression linéaire par morceaux sur le partenaire n°1

La fonction segmented() détecte trois points d'arrêt à  $x_1 = 20.350$  ,  $x_2 = 42.488$  et  $x_3 = 57.610$ . On cherche donc les différentes fonctions linéaires sur ces différents intervalles.

Ainsi nous obtenons le tableau suivant :

Intervalle	$f(x)$
$x \in [0, 20]$	$= -1,659 * 10^{-2} + 9,355 * 10^{-4}(x)$
$x \in [21, 42]$	$= 0.002447425 + (x - 20.350) * (5,63 * 10^{-5})$
$x \in [43; 57]$	$= 0.003693794 + (x - 42.488) * (0.0004185)$
$x \in [58, 65]$	$= 0.01020649 + (x - 58.050) * (-0.0014925)$

Tableau 13 : Les différentes fonctions linéaires par morceaux du partenaire n°1

La représentation graphique de ces fonctions linéaires semble indiquer une concordance forte avec les valeurs observées ainsi la régression linéaire par morceaux est une approche plus appropriée que la régression linéaire simple.

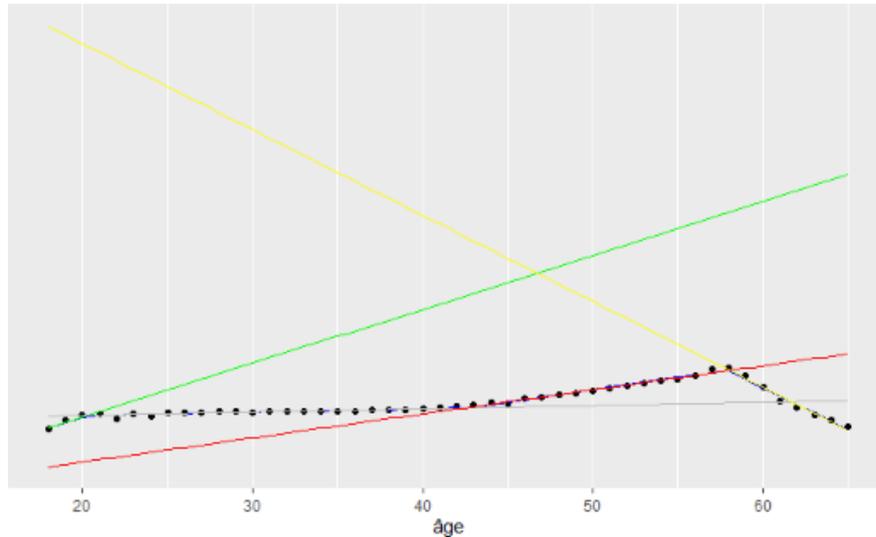


Figure 15 : Vérification des fonctions linéaires pour le partenaire n°1

Partenaire n°2

En effectuant la régression linéaire par morceaux sur le partenaire n°2, nous obtenons le graphique suivant :

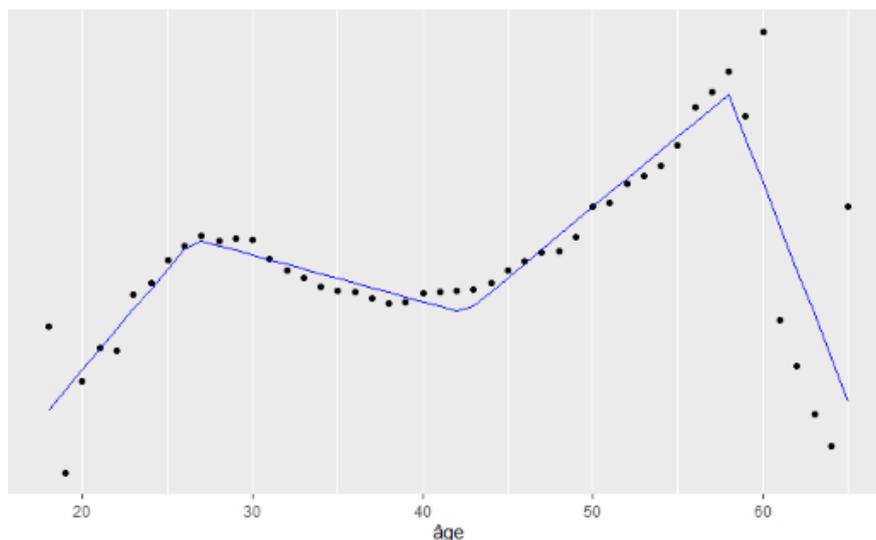


Figure 16 : La régression linéaire par morceaux sur le partenaire n°2

La fonction segmented() détecte trois points d'arrêt à  $x_1 = 26.488$ ,  $x_2 = 42.449$  et  $x_3 = 58.000$ .

On cherche donc les différentes fonctions linéaires sur ces intervalles.

Ainsi nous obtenons le tableau suivant :

<i>Intervalle</i>	<i>f(x)</i>
$x \in [0, 26]$	$= -0,012\ 409\ 2 + 0,000\ 922\ 2 * (x)$
$x \in [26, 42]$	$= 0.01201803 + (x - 26.488) * (-0,000\ 2135)$
$x \in [42; 58]$	$= 0.00861036 + (x-42.449) *(0.000641)$
$x \in [58, 65]$	$= 0.01857855 + (x - 58.00) * (-0.0020021)$

Tableau 14 : Les différentes fonctions linéaires par morceaux du partenaire n°2

La représentation graphique de ces fonctions linéaires semble indiquer encore une fois une concordance forte avec les valeurs observées, ainsi la régression linéaire par morceaux est une approche plus appropriée que la régression linéaire simple.

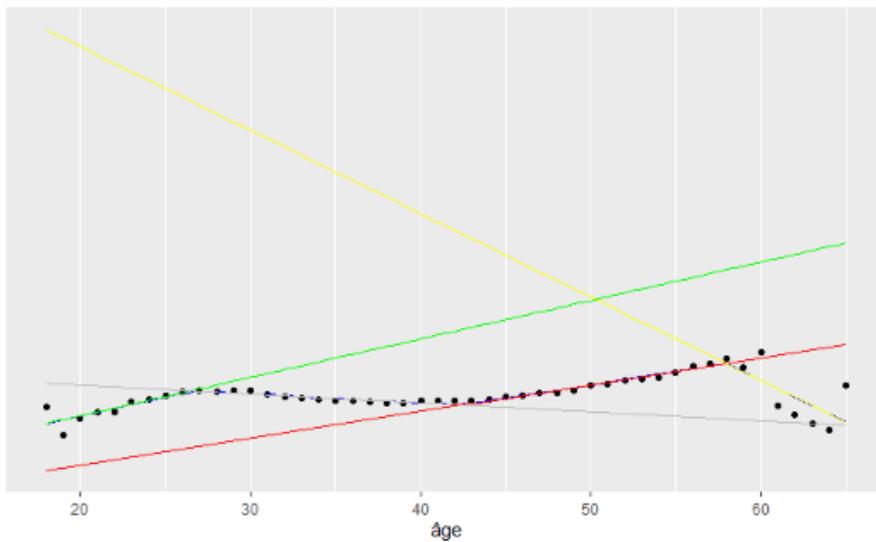


Figure 17 : Vérification des fonctions linéaires pour le partenaire n°2

### 3.5.4. Vérification numérique de l'efficacité de la régression linéaire par morceaux

En utilisant la méthode de régression linéaire par morceaux, nous avons identifié quatre sous-segments d'âge pour lesquels une relation linéaire simple peut expliquer l'incidence, contrairement à la méthode de régression linéaire simple qui a donné des coefficients de détermination faibles (10% et 36%).

Grâce à cette nouvelle méthode, nous avons obtenu des coefficients de détermination élevés, ce qui nous permet de sélectionner des sous-segments d'âge pour simuler de nouveaux taux d'incidence en fonction de l'âge.

Voici les résultats de nouveaux coefficients de détermination obtenus pour nos deux partenaires :

<i>Intervalle</i>	$R^2$
$x \in [18, 20]$	= 95%
$x \in [21, 42]$	= 75%
$x \in [43; 57]$	=97%
$x \in [58, 65]$	=97%

Tableau 15 : Les différents coefficients de déterminations, partenaire n°1.

<i>Intervalle</i>	$R^2$
$x \in [0, 26]$	= 56%
$x \in [26, 42]$	= 83%
$x \in [42; 58]$	=96%
$x \in [58, 65]$	=38%

Tableau 16 : Les différents coefficients de déterminations, partenaire n°2

Nous en déduisons que la régression linéaire par morceaux est une méthode plus adaptée. Par conséquent, cette méthode permettra, pour la suite de l'étude de simuler des taux d'incidence par âge en fonction des différents segments.

### 3.6. Simulation de taux d'incidence par âge

Dans notre étude sur le risque arrêt de travail, nous allons simuler 10 000 occurrences des taux d'incidence par tranche d'âge et ainsi sélectionner le taux d'incidence maximum obtenu pour chaque d'âge.

Ces simulations vont être effectuées à l'aide de la théorie des copules.

L'utilisation de la théorie des copules est une méthode flexible pour la modélisation de la dépendance entre les variables car elle permet de modéliser des dépendances de toutes formes.

De plus, la théorie des copules offre une approche généralisant toute approche de régression car, si la régression a pour principe de s'intéresser au comportement de  $Y|X$  ( $Y$  sachant  $X$ ) et de proposer un prédicteur sous la forme d'une espérance conditionnelle ( $E[Y|X]$ ), la théorie des copules part du principe plus général que  $(X,Y)$  suit une distribution de probabilité jointe  $P$ , que l'on peut déterminer.

Cela sous-entend qu'on tient compte de l'incertitude sur la covariable  $X$ .

On peut alors en dériver la loi conditionnelle de  $Y$  sachant  $X$  et opérer une prévision, tout en raffinant la nature de l'erreur (notée  $\epsilon$  dans l'approche par régression)

Cette partie de l'étude se concentrera sur la présentation de la théorie des copules ainsi que sur les copules-*vines*, qui sont des concepts cruciaux pour la simulation des échantillons.

Tout d'abord, nous allons présenter les notions de la théorie des copules qui seront utilisées dans la suite de notre étude. Ensuite, nous aborderons les différents types de copules, y compris la notion de copule-*vines*. Nous présenterons également la méthode de simulation et enfin, nous exposerons les résultats obtenus grâce à cette méthode.

#### 3.6.1. La théorie des copules

Une copule est une fonction mathématique qui permet de modéliser la dépendance entre deux ou plusieurs variables aléatoires.

**Définition 1.1 :** Les fonctions copules sont des distributions multivariées  $C$  définies sur l'hypercube unité  $[0, 1]^M$  avec des distributions marginales uniformes sur l'intervalle  $[0, 1]$ .

**Définition 1.2 :** Une copule bidimensionnelle est une fonction  $C$  de  $I = [0, 1]^2$  dans  $I$  ayant les propriétés suivantes :

- La copule  $C$  est attachée si:  $C(u, 0) = C(0, v) = 0 \forall (u, v) \in I$
- Les marges sont uniformes si:  $C(u, 1) = u$  et  $C(v, 1) = v \forall (u, v) \in I$
- La copule  $C$  est 2-croissantes si:  $\forall u_1, u_2, v_1, v_2 \in I$  tels que  $u_1 \leq u_2$  et  $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Théorème : Théorème de Sklar (1959)

Soit  $F$  une fonction de répartition  $n$ -dimensionnelle avec des marginales  $F_1, \dots, F_n$ , alors il existe une  $n$ -copule  $C$  telle que pour tout  $x$  de  $R^n$ ,

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

Corollaire (Inverse du théorème de Sklar)

Soit  $H$  une fonction de répartition bidimensionnelle de fonctions de répartition marginale  $F_X$  et  $F_Y$  alors la copule  $C$  associée à  $H$  est donnée par :

$$(u, v) = H(F_X^{-1}(u), F_Y^{-1}(v)) \quad \forall u, v \in I^2$$

La copule  $C$  contient ainsi toute l'information sur la distribution jointe  $F_{X,Y}(x, y)$ , exceptée l'information concernant la structure de distribution marginale.

Les copules représentent donc des exemples de distributions bivariées avec des distributions marginales standardisées.

Cas de distribution bivariée

Dans notre cas où nous avons une distribution bivariée, c'est-à-dire  $F_{X,Y}(x, y)$ , les fonctions de distribution marginale  $F_X(x)$  et  $F_Y(y)$  peuvent être exprimées de la manière suivante :

- Soient  $U = F_X(x) = F_{X,Y}(X, \infty)$  et  $V = F_Y(y) = F_{X,Y}(\infty, Y)$ , alors  $U$  et  $V$  ont des distributions uniformes sur l'intervalle  $[0, 1]$ .
- D'après le théorème de Sklar, il existe une fonction  $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$  satisfaisant :  $F_{X,Y}(x, y) = C(U, V) = C(F_X(x), F_Y(y))$   
Si  $F_X$  et  $F_Y$  sont continues, alors  $C$  est unique.

### 3.6.2. Les différentes copules

Dans la littérature, on peut trouver un grand nombre de copules différentes que nous n'étudierons pas en intégralité.

Nous en présenterons quatre qui seront utilisées par la suite.

#### 1. Copule indépendante :

La copule indépendante, également connue sous le nom de copule produit, est définie simplement comme le produit des marginales de chaque variable :

$$C(u_1, u_2, \dots, u_d) = u_1 u_2 \cdots u_d$$

2. Copule d'Archimède :

La copule d'Archimède est définie par une fonction de générateur appelée fonction de dépendance :

$$C(u_1, u_2) = \phi(\phi^{-1}(u_1) + \phi^{-1}(u_2))$$

Où  $\phi$  est une fonction de générateur qui est une fonction continue, strictement décroissante, concave et satisfait  $\phi(1) = 0$ .

3. Copule de Student :

La copule de Student est une fonction de dépendance multivariée qui permet de relier une distribution multivariée à des marginales de loi de Student. Elle est définie comme suit :

Pour un vecteur aléatoire continu  $(U, V)$  de dimension 2 ayant des marginales de loi de Student  $t$ , la copule de Student de paramètre de forme  $\nu$  est définie par :

$$C_{\nu, \rho}(u, v) = \frac{\nu + \eta}{\nu + \rho + \eta} t_{\nu} \left( \frac{\sqrt{\nu + \eta z}}{\sqrt{1 - \rho^2}} \right) \sqrt{\frac{1 - \rho^2}{\nu + \eta}}$$

Où  $z = t_{\nu}^{-1}(u)$  et  $\eta = t_{\nu}^{-1}(v)$  est le coefficient de corrélation de Pearson.

Les marginales de cette copule sont uniformes sur l'intervalle  $[0, 1]$ .

Cette copule est utilisée pour modéliser la dépendance entre des variables aléatoires ayant des queues de distributions lourdes.

4. Copule variante de la copule de Joe :

La copule variante de la copule de Joe est une copule rotative qui est obtenue en appliquant une rotation de 90 degrés à la copule de Joe standard. Elle est définie par la fonction de densité suivante :

$$c(u_1, u_2) = \frac{\theta(1 - \theta) \sin\left(\frac{\pi}{2} u_1\right) \sin\left(\frac{\pi}{2} u_2\right)}{\left[\theta^2 \sin^2\left(\frac{\pi}{2} u_1\right) + (1 - \theta)^2 \sin^2\left(\frac{\pi}{2} u_2\right)\right]^{1/2}}$$

Où  $\theta$  est un paramètre de corrélation qui peut prendre des valeurs entre -1 et 1.

### 3.6.3. Les copules-vines

La méthode des copules-*vines* utilise différentes familles de copules pour modéliser les relations multivariées (dans notre cas bivariée) complexes, en combinant des copules bivariées dans une structure arborescente.

Cette méthode nous permettra de trouver la copule la plus adaptée pour modéliser la dépendance entre nos variables.

Dans un premier temps, afin d'obtenir une décomposition en copule-*vines*, il faut considérer la décomposition de la densité multivariée en n'utilisant que des copules bivariées conditionnelles.

Dans le but de comprendre la mécanique générale de cette méthode, en annexes, nous prenons l'exemple d'un vecteur à 3 variables ( $X_1, X_2, X_3$ ).

En pratique, il existe trois types de méthodes de décomposition en *vine* : *Regular-vine*, *Canonical-vine* et *Drawable-vine*.

Cependant, une seule sera détaillée car prise en hypothèse pour la suite de ce mémoire.

### 3.6.4. La Canonical vine (C-vine)

La méthode de Canonical *vine* est une technique de modélisation statistique qui permet d'analyser les relations complexes entre plusieurs variables.

La méthode de la copule-*vine* consiste à décomposer la copule de la distribution jointe en une série de copules bivariées en utilisant un arbre de dépendance.

La Canonical *vine* suppose que toutes les branches sont connectées à un unique nœud appelé « pivot ».

En annexe, une décomposition en arbres présentant la méthode Canonical *vine* pour une distribution jointe à trois variables ( $X, Y, Z$ ) est présenté.

### 3.6.5. Problématique et méthode

Cette partie sera consacrée à définir la méthode permettant la simulation de nouveaux taux d'incidence par âge. Cela permettra ensuite de sélectionner le scénario le plus défavorable parmi 10 000 simulations.

Il représente le taux d'incidence maximum pour chaque âge.

Soit  $(X_i, Y_i), i = 1, \dots, n$  un échantillon de variables aléatoires bivariées représentant l'âge  $X$  et l'incidence  $Y$  d'un risque, pour  $n$  individus donnés.

On cherche à simuler 10 000 nouvelles valeurs pour chaque échantillons  $(X_i, Y_i), i = 1, \dots, n$ .

Ainsi :  $C = (X_i, Y_i)$

$X_i$  représente l'âge.

$Y_i$  représente les taux d'incidence.

$F_{Y_i}$  et  $F_{X_i}$  les fonctions de répartition marginales.  
 FC la fonction de répartition jointe.

La première étape de la méthode consiste à appliquer une transformation linéaire à chaque élément de notre paire de vecteurs aléatoires, afin de les convertir en variables uniformément distribuées, puisque les copules utilisent des rangs plutôt que des valeurs numériques pour l'analyse statistique.

Ainsi, nous obtenons deux nouveaux vecteurs :  $M_i = \frac{\text{rang}(X_i)}{n+1}$  et  $N_i = \frac{\text{rang}(Y_i)}{n+1}$

pour tout  $i \in [1 ; n]$  avec  $n$  la longueur de l'échantillon.

La deuxième étape consiste à choisir la structure de copule de *vine* (R-*vine*, C-*vine* et D-*vine*) optimale pour les données d'entrée.

De plus, la sélection de la structure de copule à utiliser est importante. On cherche à déterminer la structure de copule la plus appropriée pour nos données. Pour cela, nous allons effectuer pour chaque simulation un test d'indépendance qui nous permettra ainsi de la choisir.

La troisième étape consiste à choisir la famille de copules qui modélisera la dépendance entre les deux variables de notre échantillon (bivarié ou multivarié).

Une fois la structure de copule-*vine* et la famille de copules déterminées, il nous reste à simuler l'échantillon en utilisant la copule-*vine* spécifiée. Le résultat est un ensemble de points uniformément répartis entre 0 et 1 dans l'espace multivarié.

La dernière étape consiste à ramener cet ensemble de points « à la réalité » et donc d'utiliser la fonction inverse de la copule c'est-à-dire :  $F_n^{-1}(M_i)$  et  $F_n^{-1}(N_i)$ .

### 3.6.6. Le critère AIC

Le critère d'information d'Akaike (AIC) est une mesure statistique qui tente de trouver un compromis entre la complexité du modèle et sa qualité d'ajustement aux données. Il est calculé en tenant compte du nombre de paramètres du modèle et de la qualité d'ajustement du modèle aux données. Le modèle avec le plus petit AIC est considéré comme le meilleur car il offre le meilleur équilibre entre la qualité d'ajustement et la complexité du modèle.

Le critère AIC est largement utilisé dans les modèles de copules-*vines* pour sélectionner le meilleur modèle de copule parmi un ensemble de modèles candidats, en tenant compte de la complexité de la structure de dépendance et de la qualité d'ajustement du modèle aux données.

Les copules-*vines* sélectionnées sont choisies en utilisant le critère AIC, car il n'y a pas de garantie d'unicité des résultats. Il convient de souligner l'importance de ce critère et de préciser que les copules proposées pour chaque classe d'âge ont été préalablement sélectionnées parmi les familles de copules paramétriques.

## 3.6.7. Choix de la famille et de la structure de copule pour les deux partenaires

Dans le but de simuler les taux d'incidence par âge, nous utilisons les copules qui permettent de modéliser la dépendance entre deux variables aléatoires en capturant la relation statistique entre ces variables, qui peuvent être mesurées sous forme de rangs.

En sélectionnant la structure de copule appropriée, on peut donc modéliser la nature de la dépendance entre les variables aléatoires, ce qui permet de simuler des données cohérentes avec cette dépendance.

Ce sont les fonctions `RVineStructureSelect()` et `RVineCopSelect()` qui nous ont permis de choisir la structure ainsi que les familles de copules à utiliser.

Ces fonctions utilisent une méthode de sélection basée sur le critère d'information d'Akaike (AIC).

De plus, une représentation graphique nous permettra de justifier davantage (particulièrement le critère AIC) le choix effectué par ces fonctions.

Ainsi, voici les familles de copules les plus adaptées à chaque jeu de données.

Partenaire n°1 :

Variable	Famille de copule	Structure de copule
[0 ;20]	Indépendante	C-vine
[21 ;42]	Archimédienne	C-vine
[43 ;57]	Student	C-vine
[58 ; 65]	Indépendante	C-vine

Tableau 17 : Les différents familles et structures de copules, partenaire n°1.

Partenaire n°2 :

Variable	Famille de copule	Structure de copule
[18 ;26]	Indépendante	C-vine
[27 ;42]	Variante de la copule de Joe	C-vine
[42 ;58]	Student	C-vine
[58 ; 65]	Indépendante	C-vine

Tableau 18 : Les différents familles et structures de copules, partenaire n°2.

## 3.6.8. Explications

Le choix des familles de copules utilisées ci-dessus peut être expliqué par des graphiques de dispersion, qui permettent de justifier la dépendance entre l'âge et l'incidence.

Partenaire n°1 :

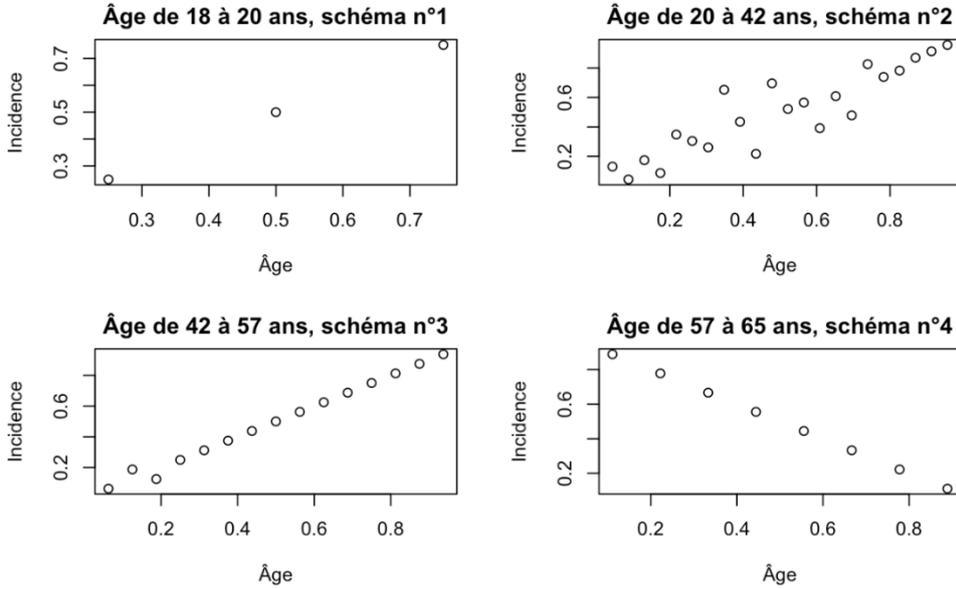


Figure 18 : Graphique de dispersion pour le partenaire n°1

Partenaire n°2 :

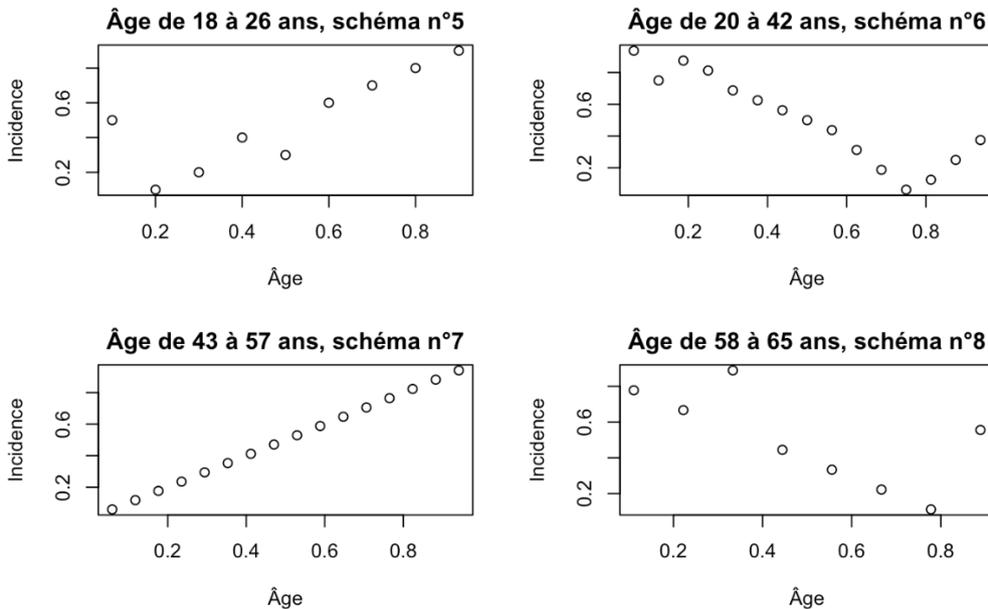


Figure 19 : Graphique de dispersion pour le partenaire n°2

Dans un premier temps, on constate que les points représentés dans les schémas n°1, 4, 5 et 8 sont répartis de manière assez aléatoire et uniforme. On peut donc suggérer de manière

graphique qu'il n'y a pas de relation significative entre les deux variables, et donc que les copules à utiliser sont des copules indépendantes.

Les schémas n°3 et n°7 montrent une relation croissante et linéaire entre les deux variables dans le graphique de dispersion.

Dans ce cas, une copule de Student peut être utilisée pour modéliser cette relation. La copule de Student est souvent préférée pour modéliser des relations linéaires, car elle permet de prendre en compte les queues épaisses de la distribution des données et de mieux modéliser les relations non linéaires que la copule Gaussienne.

Dans le schéma n°2, les valeurs des deux variables sont dispersées sur tout le graphique, sans qu'une relation clairement identifiable ne se dégage. Dans ce cas, l'utilisation d'une copule d'Archimède peut être justifiée car elle est souvent utilisée pour modéliser des relations non linéaires ou asymétriques entre les variables, ou encore lorsque les valeurs sont dispersées sur le graphique.

Pour finir, sur le schéma n° 6, on peut observer une relation initialement décroissante entre les deux variables puis croissante. Dans ce cas, il peut être justifié d'utiliser une variante de la copule de Joe, qui est adapté pour modéliser des relations non monotones entre les variables.

Les résultats obtenus par cette méthode de simulation sont présentés dans ce qui suit.

### 3.7. Résultats

En utilisant la méthode décrite précédemment et en appliquant la structure et la famille de copules adaptées pour chaque tranche d'âge, nous obtenons pour chaque partenaire les graphiques qui suivent.

Ces nouveaux taux d'incidence simulés représentent le scénario le plus défavorable pour chaque âge c'est-à-dire : le taux d'incidence maximum simulé pour chaque âge au sein de 40 000 simulations (10 000 simulations pour chaque tranche d'âge).

Le graphique présente en vert la courbe des taux d'incidence simulés et en rouge le taux d'incidence moyen observé.

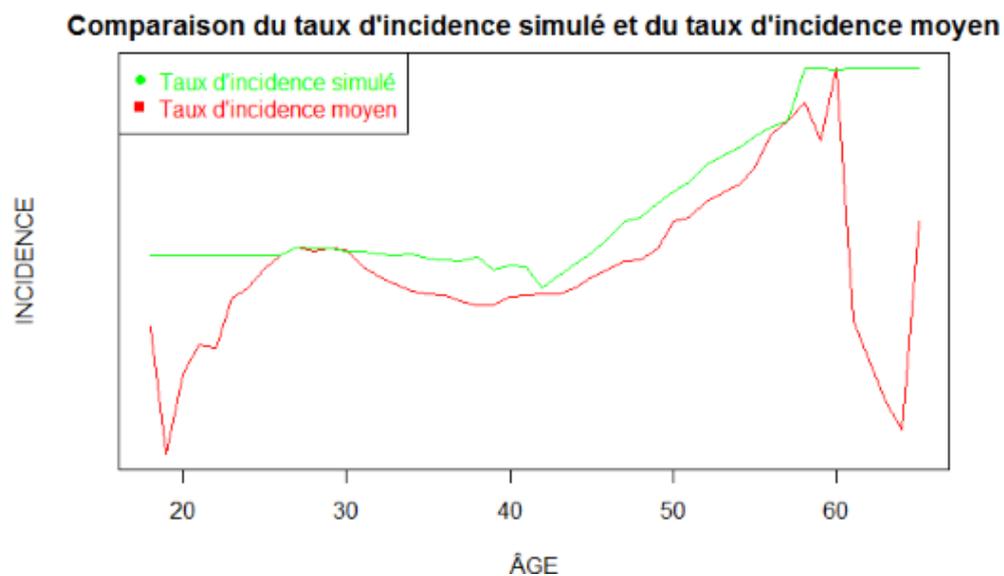
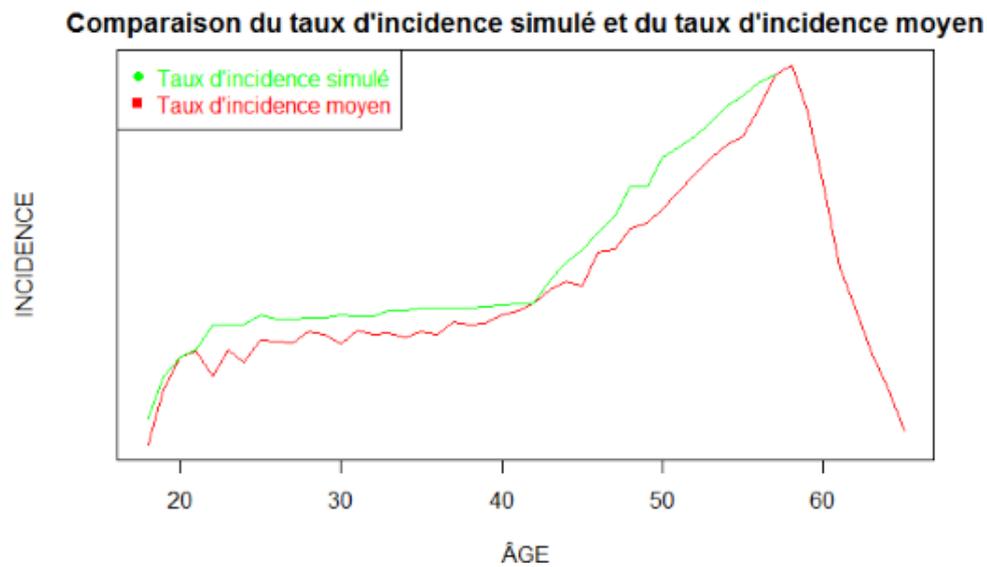


Figure 20 : Graphique des taux d'incidence simulés pour les deux partenaires

Dans un premier temps, en effectuant une simulation de 10 000 taux d'incidence par segment d'âge (et donc 40 000 simulations par partenaire), nous avons constaté que, les taux simulés étaient toujours supérieurs ou égaux aux taux moyens observés.

En analysant les résultats pour les deux partenaires retenus, nous avons noté :

- ✓ Pour le partenaire n°1, il y a un écart de 0,14% entre le taux d'incidence maximum simulé et le taux d'incidence moyen observé au même âge (52 ans).
- ✓ Pour le partenaire n°2, il y a un écart de 0,25% à l'âge de 38 ans (c'est-à-dire la différence entre le taux d'incidence maximum simulé et le taux moyen observé au même âge).

Il convient également de souligner certaines limites à cette méthode de simulation.

La première limite de cette méthode est liée à la nécessité de fournir les informations relatives aux taux d'incidence minimum et maximum pour pouvoir simuler les données. Cette méthode est très dépendante de la structure des données fournies. Les nouveaux taux d'incidences simulés dépendent des informations sur les taux d'incidence minimum et maximum et sont donc bornés.

La deuxième limite de cette méthode concerne l'âge simulé. En effet, l'âge simulé n'est pas un nombre entier (par exemple 43,2 ans), et seule la partie entière de l'âge est prise en compte (c'est-à-dire 43 ans).

Néanmoins, cette méthode de simulation reste intéressante, car elle permet de générer 10 000 nouveaux taux d'incidence pour chaque segment d'âge. Ainsi, en lui imposant des taux d'incidence minimum et maximum, cette méthode de simulation reste toujours pertinente mais demeure tributaire de ces informations.

## Chapitre 4

### 4. Le maintien en arrêt de travail

L'objectif de cette partie est de présenter la méthode retenue afin de déterminer une durée résiduelle de maintien en arrêt de travail pour chaque partenaire.

Le périmètre étudié correspond aux prêts immobiliers.

Dans un premier temps, nous présenterons la méthode afin de calculer des durées résiduelles de maintien en arrêt de travail.

Puis, nous appliquerons cette méthode à nos données.

Enfin, nous déterminerons une durée résiduelle pour chacun des partenaires.

## 4.1. Les tables de maintien en arrêt de travail

## 4.1.1. Les tables du BCAC

Pour le provisionnement des prestations servies en cas d'incapacité et d'invalidité, on utilise les lois produites par le BCAC<sup>1</sup>. À l'origine, ces tables ont été construites en 1993 puis modifiées en 2013.

Les tables du BCAC sont constituées de trois tables différentes :

- Une table de maintien en incapacité pour des âges allant de 20 à 65 ans et avec une ancienneté en incapacité allant de 0 à 36 mois.
- Une table de passage de l'état d'incapacité à l'état d'invalidité pour des âges allant de 20 à 62 ans, dont les anciennetés vont de 0 à 45 ans.
- Enfin une table de maintien en invalidité pour des âges allant de 18 à 65 ans, dont les anciennetés vont de 0 à 45 ans.

C'est la table de maintien en incapacité que nous allons retenir pour la suite de notre étude car cette table fournit les informations sur la durée moyenne de l'incapacité.

Ci-dessous un extrait de la loi de maintien en incapacité du BCAC de 2013. Ce tableau représente le nombre d'individus entrés à l'âge x et toujours présents en incapacité après n mois.

Âge/Ancienneté	0	1	2	3	...	20	...	26	...
20	10 000	4 414	2 676	1 939	...	168	...	80	...
21	10 000	4 638	2 855	2 064	...	184	...	100	...
22	10 000	4 848	3 024	2 173	...	187	...	110	...
23	10 000	5 026	3 169	2 274	...	186	...	118	...
24	10 000	5 159	3 286	2 367	...	186	...	123	...
25	10 000	5 273	3 395	2 456	...	192	...	128	...
26	10 000	5 357	3 478	2 528	...	202	...	134	...
27	10 000	5 427	3 550	2 590	...	217	...	141	...
28	10 000	5 477	3 604	2 641	...	231	...	150	...
29	10 000	5 526	3 654	2 687	...	242	...	156	...
30	10 000	5 567	3 699	2 731	...	250	...	157	...
...	...	...	...	...	...	...	...	...	...
42	10 000	5 606	3 774	2 853	...	418	...	269	...
...	...	...	...	...	...	...	...	...	...
45	10 000	5 666	3 846	2 931	...	461	...	298	...
...	...	...	...	...	...	...	...	...	...

Tableau 19 : Extrait de la loi de maintien en incapacité du BCAC (Bureau commun des assurances collectives)

Effectivement, l'objectif de cette étude n'est pas de construire une nouvelle table de maintien en arrêt de travail, mais plutôt d'utiliser les tables de maintien existantes pour estimer un coefficient moyen de maintien pour chacun des partenaires de CNP.

Pour ce faire, nous utilisons des tables d'expérience maintien qui ont été établies à partir de données collectées auprès de compagnies d'assurance. Ces tables fournissent des informations sur la probabilité de la durée de l'incapacité.

### 4.1.2. Les tables d'expérience

Un organisme d'assureur peut construire ses propres lois d'expérience, lui permettant d'obtenir un provisionnement propre des prestations servies.

Cependant, une certification de table d'expérience est effectuée par des Actuaires.

Ainsi, la certification de ces tables d'expérience sera effectuée par des actuaires sous des conditions très particulières. L'actuaire doit avoir une expérience d'au moins cinq ans dans le domaine de la tarification et du provisionnement avec un niveau en statistiques adéquat.

En aucun cas, l'actuaire ayant construit ces lois d'expérience ne pourra les certifier par lui-même ou par un tiers dépendant de lui. Enfin, la durée de validité des tables d'expérience dépend d'un suivi associé.

Pour les tables d'expérience de maintien en arrêt de travail, la validité maximale est de quatre ans en cas de suivi et de deux ans après certification en cas de non suivi.

Ce suivi vise à vérifier que la stabilité du portefeuille à l'aide d'indicateurs statistiques. La fréquence de ce suivi est d'une fois par an.

### 4.1.3. Concept de mortalité

Dans le cadre de cette étude, c'est-à-dire dans un portefeuille emprunteur, on souhaite déterminer un coefficient de maintien "moyen" pour chacun des partenaires.

En d'autres termes, on souhaite modéliser la durée entre l'adhésion au contrat et sa date de sortie (ou de censure) pour obtenir une probabilité de maintien en arrêt de travail.

Pour cela, on utilise la théorie des modèles de durée.

### 4.1.4. Définitions

Soit  $T$  une variable aléatoire telle que  $T \in [0; +\infty]$  décrivant le temps passé dans un intervalle  $[x; x+t]$ . En effet,  $x$  désigne l'âge de l'assuré.

- La fonction de répartition de  $T$  notée :

$$F(t) = P(T \leq t)$$

- La fonction de survie notée :

$$S(t) = 1 - F(t) = 1 - P(T \leq t) = P(T > t)$$

$S$  est une fonction décroissante telle que  $S(0) = 1$  si  $P(T = 0) = 0$  ce que nous supposons et

$$\lim_{t \rightarrow +\infty} S(t) = 0.$$

- La fonction de hasard  $h$  et la fonction de densité  $f$  (lorsqu'elle existe) :

$$h(t) = \frac{f(t)}{S(t)} = \frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t))$$

$$f(t) = \frac{d}{dt} F(t) = -\frac{d}{dt} S(t)$$

La fonction de survie s'exprime donc par la fonction de hasard telle que :  $S(t) = \exp(-H(t))$  où  $H(t)$  est l'intégrale de  $h(t)$  entre 0 et  $t$ .

#### 4.1.5. Censure et troncature

Lors de l'analyse des données de survie, nous serons confrontés à deux phénomènes importants : la censure et la troncature.

Si ces phénomènes ne sont pas traités, alors nos jeux de données seront de mauvaise qualité à cause de données incomplètes.

##### Censure

La durée est dite censurée si l'individu n'a pas subi l'événement à sa dernière observation.

Il existe deux types de censure : la censure à droite et la censure à gauche.

Soit  $X_1, \dots, X_n$  un n-échantillon de durées.

Un phénomène de censure est observé s'il existe  $C_1, \dots, C_n$  n variables aléatoires i.i.d<sup>1</sup> telles que

$\forall 1 \leq i \leq n$  les variables  $T_i = \min(X_i, C_i)$ .

Ainsi, on peut observer la durée réellement observée  $T_i$  et un indicateur  $\delta_i = \mathbf{1}_{X_i \leq C_i}$ .

Si  $\delta_i = 1$  alors l'événement est observé (durées complètes), sinon l'événement est censuré.

##### Censure à gauche

La censure à gauche se produit lorsque l'on a commencé à étudier les individus après que certains événements se sont déjà produits.

Ci-dessous, un exemple de censure à gauche :



Figure 21 : Censure à gauche.

1. Variables indépendantes et identiquement distribuées

### Censure à droite

La censure à droite est une situation où l'on ne connaît pas la durée de survie pour certaines observations car elles ont été perdues de vue avant la fin de l'étude.

Ci-dessous, un exemple de censure à droite :



Figure 22 : Censure à droite.

Le phénomène de troncature se définit lorsque la variable n'est pas observable.

### Troncature à gauche

La troncature à gauche est une situation où l'on ne dispose pas des informations sur les individus qui sont sortis de l'étude avant le début de l'observation. Cela signifie que leur temps de survie est "tronqué" ou coupé à gauche de la période d'observation.

Ci-dessous, un exemple de troncature à gauche :

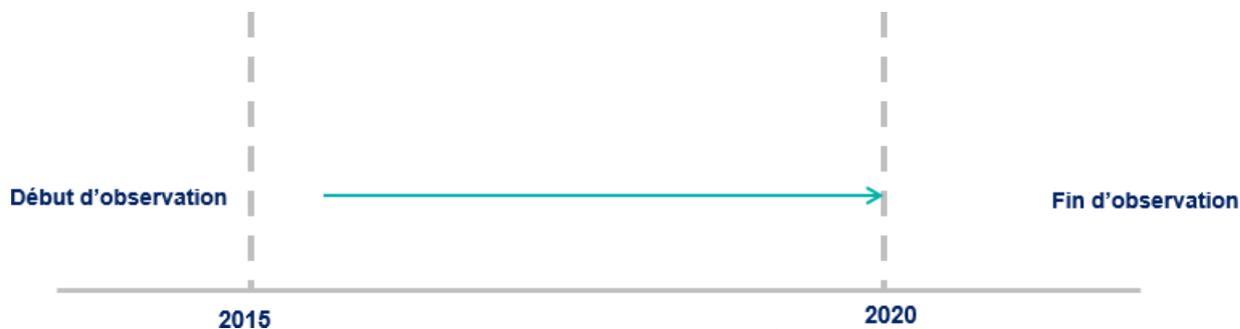


Figure 23 : Troncature à gauche

### Troncature à droite

La troncature à droite est une situation où l'on ne dispose pas de l'information complète sur les individus qui ont survécu jusqu'à la fin de l'étude, car leur durée de survie dépasse la période d'observation.

Ci-dessous, un exemple de troncature à droite, c'est-à-dire lorsque la variable d'intérêt n'est pas observable au-dessus d'un seuil.

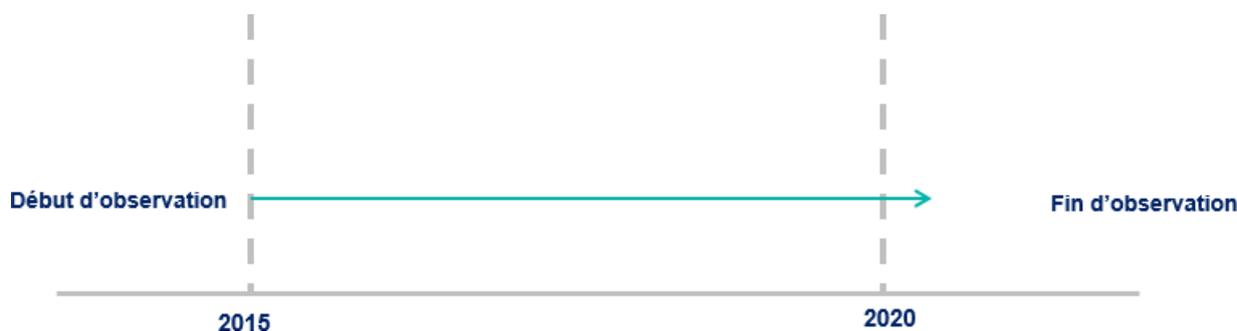


Figure 24 : Troncature à droite

Cette partie a permis de rappeler les concepts de mortalité qui vont être utilisés dans la suite l'étude.

## 4.2. Application

Après avoir rappelé les concepts de mortalité, on souhaite appliquer ces notions afin d'obtenir une durée résiduelle de maintien en arrêt de travail pour chacun de nos partenaires.

### 4.2.1. Âge d'entrée en arrêt de travail

D'après les bases de données de sinistres, nous sommes capables de déterminer les âges moyens à l'entrée en arrêt de travail sur le portefeuille emprunteur pour un prêt immobilier.

Pour le partenaire n°1, l'âge moyen à l'entrée en arrêt de travail est de 45,2 ans. Pour le partenaire n°2, l'âge moyen à l'entrée en arrêt de travail est de 42,6 ans.

Ainsi, nous allons nous baser sur ces âges pour déterminer les durées résiduelles de chaque partenaire.

## 4.2.2. Durée de prestation

### 4.2.2.1. Lois de maintien

À présent, nous allons observer les tables d'expériences de CNP ASSURANCES sur une fenêtre d'observation allant de 2015 à 2020 pour nos deux partenaires.

Les tables d'expérience sont des bases de données statistiques qui regroupent les données de mortalité ou de survie d'une population.

En appliquant les méthodes de censure et de troncature sur les données des deux partenaires, nous limiterons notre fenêtre d'observation pour observer les lois de maintien.

Les durées résiduelles découlent des lois de maintien.

Nous utiliserons les tables d'expérience propres à CNP pour la suite de l'étude.

Ci-dessous, la loi de maintien pour le partenaire n°1 pour la tranche d'âge 45-50 ans.

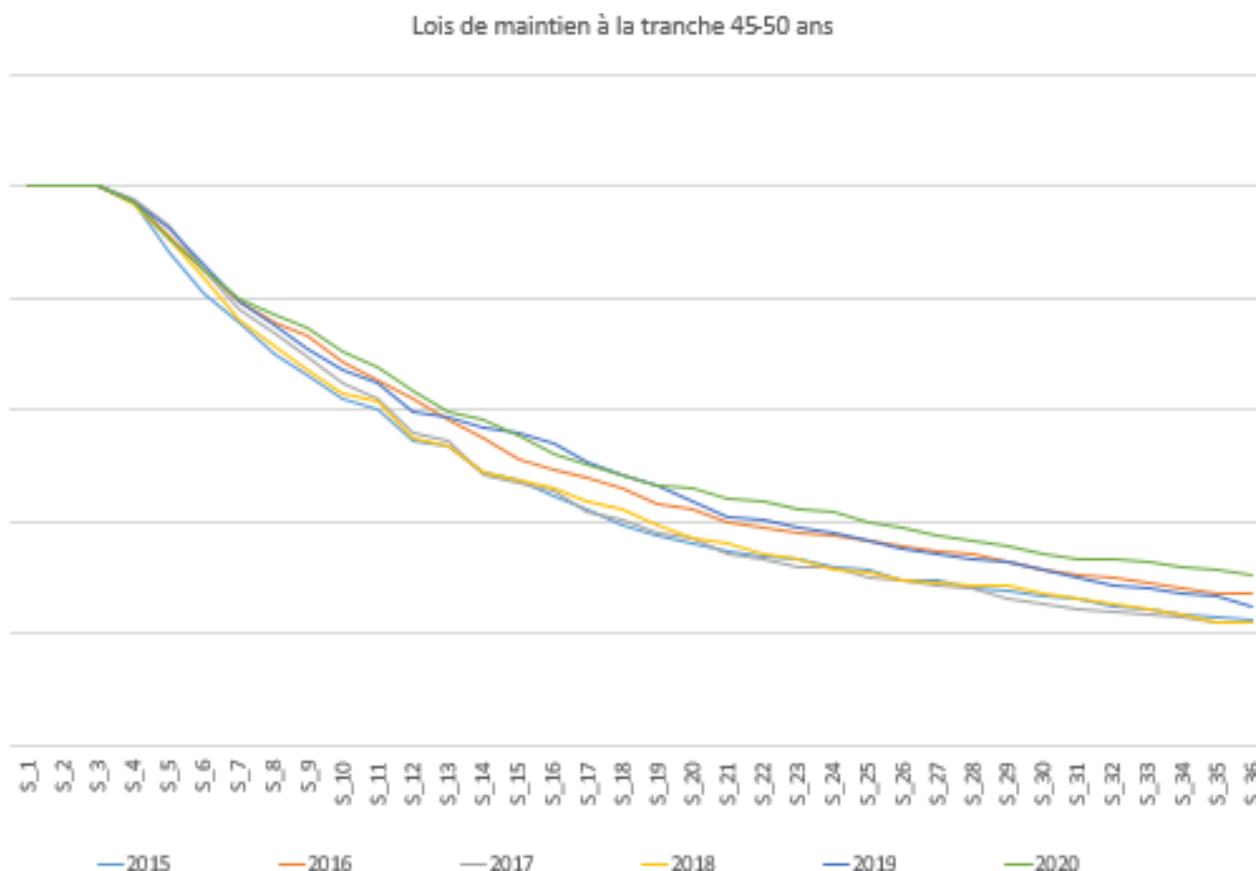


Figure 25 : Lois de maintien pour la tranche 45-50 ans du partenaire n°1

Ci-dessous, la loi de maintien pour le partenaire n°2 pour la tranche d'âge 40-45 ans.

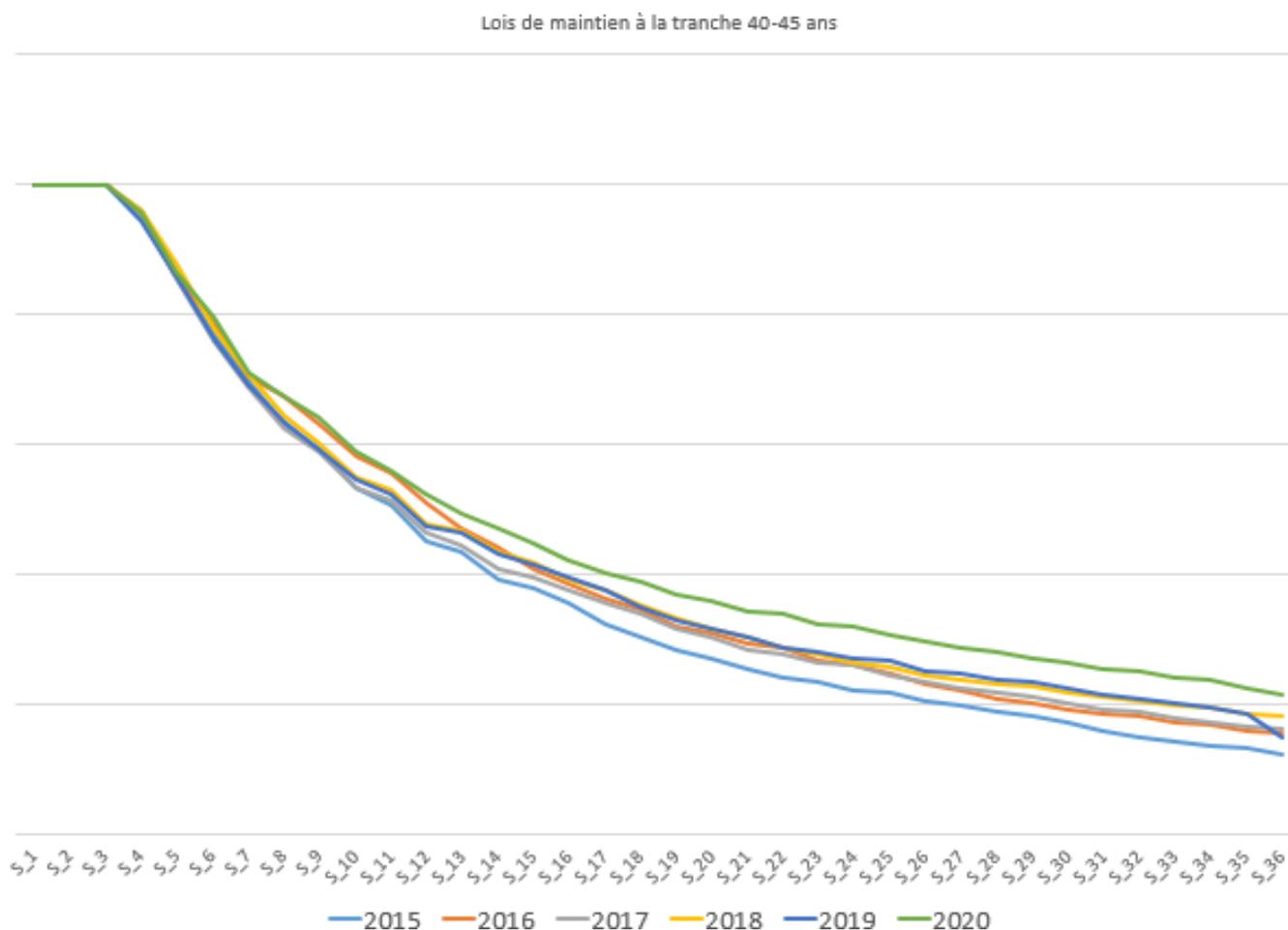


Figure 26 : Lois de maintien pour la tranche 40-45 ans du partenaire n°2

#### 4.2.2.2. Analyse

Selon les graphiques précédents, les périodes de calibrage de 2015 à 2020 montrent une augmentation du taux de maintien au fil des années.

La courbe de la loi du maintien diminue au fil du temps, avec une pente plus forte durant les premiers mois (à partir de trois mois) puis une diminution progressive.

La diminution de la pente s'explique par le fait qu'un assuré a une plus grande probabilité de sortir de l'état d'incapacité en début d'arrêt de travail.

En revanche, il a moins de chances d'en sortir en fin d'arrêt de travail. De plus, l'âge aggrave la probabilité d'incapacité.

Ainsi, plus un assuré est âgé à l'entrée en incapacité et plus sa probabilité d'y rester sera grande.

Et à l'inverse, plus un assuré est jeune à l'entrée en incapacité et plus sa probabilité d'y rester sera faible.

#### 4.2.2.3. Durées résiduelles

Pour obtenir les durées résiduelles, on doit récupérer les données des lois de maintien et appliquer la formule ci-dessous :

$$DR_k = \frac{\sum_{j=i+1}^n (S_j)}{(S_i)}$$

Avec :

$DR_k$  : Durée résiduelle (ou durée restante).

$S_i$  : Probabilité de survie au-delà de  $i$  mois passé en arrêt de travail.

$i$  : nombre de mois en arrêt de travail.

$k$  : le partenaire 1 ou 2.

Ainsi les durées résiduelles sont :

<i>Partenaire</i>	<i>Durées résiduelles, en mois</i>
Partenaire n°1	DR <sub>1</sub>
Partenaire n°2	DR <sub>2</sub>

Tableau 20 : Les différentes durées résiduelles de chaque partenaire.

*Pour des raisons de confidentialité les durées résiduelles obtenues ne sont pas communiquées.*

## 4.2.2.4. Comparaison avec les lois chaînées du BCAC

Les tables actuelles du BCAC considèrent l'incapacité selon les règles de la Sécurité sociale contrairement à la CNP.

Notre base de données de sinistres ne distingue pas les différents états que sont l'incapacité et l'invalidité.

Or, la CNP n'est pas dans l'obligation de s'aligner sur la définition stricte de la Sécurité sociale. Avant la mise en place des lois d'expériences de CNP, les lois chaînées étaient utilisées pour provisionner l'emprunteur.

Elles permettent de prendre en compte les deux états incapacité et invalidité.

Ci-dessous, la loi de maintien provenant des lois chaînées du BCAC pour la tranche d'âge 40-50 ans.

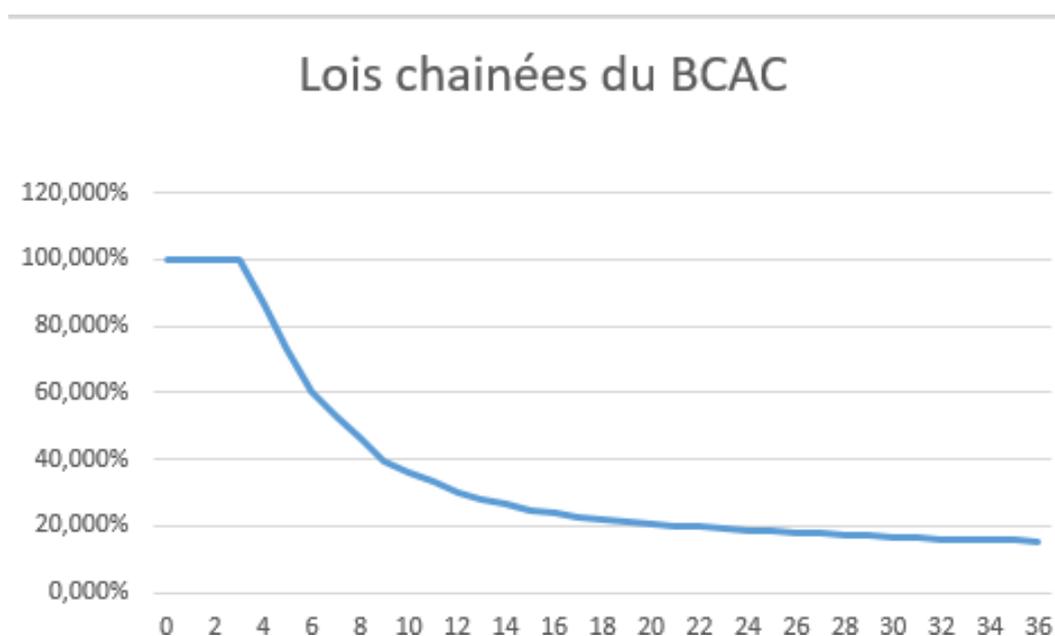


Figure 27 : Loi de maintien chaînée à la tranche 40-50 ans du BCAC

On cherche à comparer la durée résiduelle (provenant des lois chaînées) du BCAC avec les durées résiduelles des deux partenaires de CNP.

Pour une loi de maintien allant jusqu'à 36 mois, nous obtenons les résultats suivants :

<i>Partenaire</i>	<i>Les différentes durées résiduelles (en mois)</i>
Partenaire n°1	DR <sub>1</sub>
Partenaire n°2	DR <sub>2</sub>
BCAC	9.6 mois

*Tableau 21 : Comparaison des durées résiduelles*

*Pour des raisons de confidentialité les durées résiduelles ainsi obtenues ne sont pas communiquées.*

De nombreuses raisons peuvent justifier les différences entre les coefficients de maintien et les durées résiduelles de CNP versus BCAC.

Par exemple, cela peut être expliqué par le fait que pour souscrire à une assurance de prêt en immobilier, il faut passer par une étape en amont : la sélection médicale.

On peut donc en supposer que la sélection médicale a filtré le risque lors de la souscription du contrat d'assurance emprunteur.

C'est la raison pour laquelle le taux d'incidence est plus faible pour CNP, et par conséquent, que la probabilité de se maintenir en arrêt de travail est presque trois fois plus forte pour la CNP que pour lois chaînées du BCAC.

# Chapitre 5

## 5. La notion de rente

Ce chapitre décrit la notion de rente à la survenance d'un sinistre pour donner suite à la souscription d'un contrat d'assurance emprunteur.

### 5.1. L'assurance de prêt en cas de sinistre

L'assurance emprunteur permet de prendre en charge les mensualités d'un emprunt à la banque en cas de défaillance de l'assuré.

Afin de mettre en œuvre les garanties prévues au contrat, le sinistre doit être déclaré.

#### 5.1.1. Déclaration du sinistre dans le cadre d'une assurance emprunteur

En cas de sinistre couvert par le contrat d'assurance emprunteur, il faut avertir au plus vite l'assureur.

Un délai maximum est mis en place par l'assureur dans les conditions générales du contrat. Ce délai doit être respecté par l'assuré afin d'être pris en charge par l'assureur.

Il existe différentes façons de réaliser une déclaration de sinistre.

La plupart du temps il faut envoyer une lettre recommandée avec accusé de réception détaillant les circonstances du sinistre. Cependant, certains assureurs permettent de faire cette déclaration en ligne en se connectant sur l'espace personnel du site internet.

#### 5.1.2. Les différentes pièces à fournir

Dans le but de faire jouer les garanties, il est nécessaire d'envoyer la copie du contrat de prêt immobilier ainsi que celle de son échéancier.

D'autres pièces à fournir sont indispensables afin d'être pris en charge par l'assureur.

Parmi ces pièces, on peut noter ;

- l'acte de décès,
- l'attestation de la Sécurité sociale prouvant le classement en invalidité de l'assuré,
- une copie du contrat de travail, d'une lettre de licenciement et une attestation à pôle emploi en cas de perte d'emploi de l'assuré.

#### 5.1.3. La décision de l'assureur

Une fois la déclaration et les pièces justificatives envoyées par l'assuré, l'assureur va se charger de constituer un dossier de sinistre au nom de l'emprunteur ou de ses ayants droit.

Dans le cas où le dossier est complet, l'assureur va prendre en charge l'assuré en fonction des

garanties souscrites et des conditions par le versement d'une rente ou d'un capital.

## 5.2. Application

À présent, on souhaite savoir quel est le montant moyen payé par CNP en cas de sinistre. Afin d'obtenir ces chiffres, on regarde les bases de données de sinistres.

Après plusieurs manipulations et requêtes sous SAS, on obtient un tableau représentant le montant moyen payé par CNP Assurance en cas de sinistre sur le périmètre souhaité pour les années allant de 2015 à 2020.

On obtient le tableau suivant pour le partenaire n°1 :

<i>Partenaire</i>	2015	2016	2017	2018	2019	2020
Partenaire n°1	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>

*Tableau 22 : Montant de rente moyenne par année de survenance du partenaire n°1*

On obtient le tableau suivant pour le partenaire n°2 :

<i>Partenaire</i>	2015	2016	2017	2018	2019	2020
Partenaire n°2	R <sub>7</sub>	R <sub>8</sub>	R <sub>9</sub>	R <sub>10</sub>	R <sub>11</sub>	R <sub>12</sub>

*Tableau 23 : Montant de rente moyenne par année de survenance du partenaire n°2*

Toujours dans le cadre le plus défavorable, on retient deux montants de sinistres pour chaque partenaire : le montant maximum entre 2015 et 2020 et le montant de sinistre moyen.

Le montant maximum sera utilisé pour calculer le montant de sinistre supposé maximum que CNP puisse envisager de payer pour le risque arrêt de travail.

Le montant moyen sera utilisé pour calculer le montant de sinistre supposé moyen que CNP peut envisager de payer pour le risque arrêt de travail.

*Pour des raisons de confidentialité, les rentes moyennes ne sont pas communiquées.*

## Chapitre 6

### 6. Résultats

Ce chapitre décrit les résultats obtenus à partir de l'incidence en arrêt de travail, dumaintien en incapacité et de la rente moyenne.

#### 6.1. Méthode d'estimation des charges de sinistres

A partir des indicateurs estimés dans les parties précédentes, on cherche à calculer à la fois la charge annuelle moyenne de sinistres et la charge annuelle maximum.

La formule pour obtenir le montant moyen annuel payé (ou Best Estimate) par la CNP :

$$S_{moyen} = NB_{AT} * DR_j * Rente_{annuelle}$$

$$S_{moyen} = \sum_{i=1}^n A_{i,x} * T_{x,i} * DR_k * Rente_{annuelle}$$

La formule pour obtenir du scénario le plus défavorable que CNP risque de payer :

$$S_{choqué} = NB_{AT_{choqué}} * DR_k * Rente_{annuelle}$$

$$S_{choqué} = \sum_{i=1}^n A_{i,x} * T_{x,i_{choqué}} * DR_k * Rente_{annuelle}$$

Avec :

$S_{choqué}$  et  $S_{moyen}$  : Le montant choqué et le montant moyen.

$NB_{AT}$  : Le nombre de personnes entrant en arrêt de travail sur le portefeuille.

$DR_j$  : Durée résiduelle pour le partenaire j (ou k).

$Rente_{annuelle}$  : Rente annuelle pour le partenaire j (ou k).

$A_{i,x}$  : Nombre d'assurés d'âge i couverts pour le risque arrêt de travail pendant l'année x.

$T_{x,i_{choqué}}$  : Taux d'incidence stressé pour un assuré d'âge i pendant l'année x.

$T_{x,i}$  : Taux d'incidence pour un assuré d'âge i pendant l'année x.k : Le partenaire k=1,2 .

Enfin, le montant qui nous intéresse se calcule grâce à la formule suivante :

$$S_{dif} = S_{choqué} - S_{moyen}$$

$$S_{dif} = (DR_{i,j} * Rente_{annuel}) (NB_{AT_{choqué}} - NB_{AT})$$

## 6.2. Définition du résultat brut d'exploitation

Le RBE est un indicateur qui mesure la performance économique de l'exploitation d'une entreprise. Il est calculé en soustrayant les frais du produit net d'exploitation (PNA).

La politique de réassurance de CNP Assurances a pour objectif d'utiliser la réassurance de manière ciblée afin de respecter les limites de risques locaux, notamment la protection de 50% du RBE.

Pour chaque partenaire, nous allons vérifier si le montant de sinistres calculé dépasse cette limite.

## 6.3. Distribution de la charge de sinistres par âge

Le graphique présente la charge totale de sinistres moyenne ainsi que la charge totale de sinistres maximum payées par la CNP.

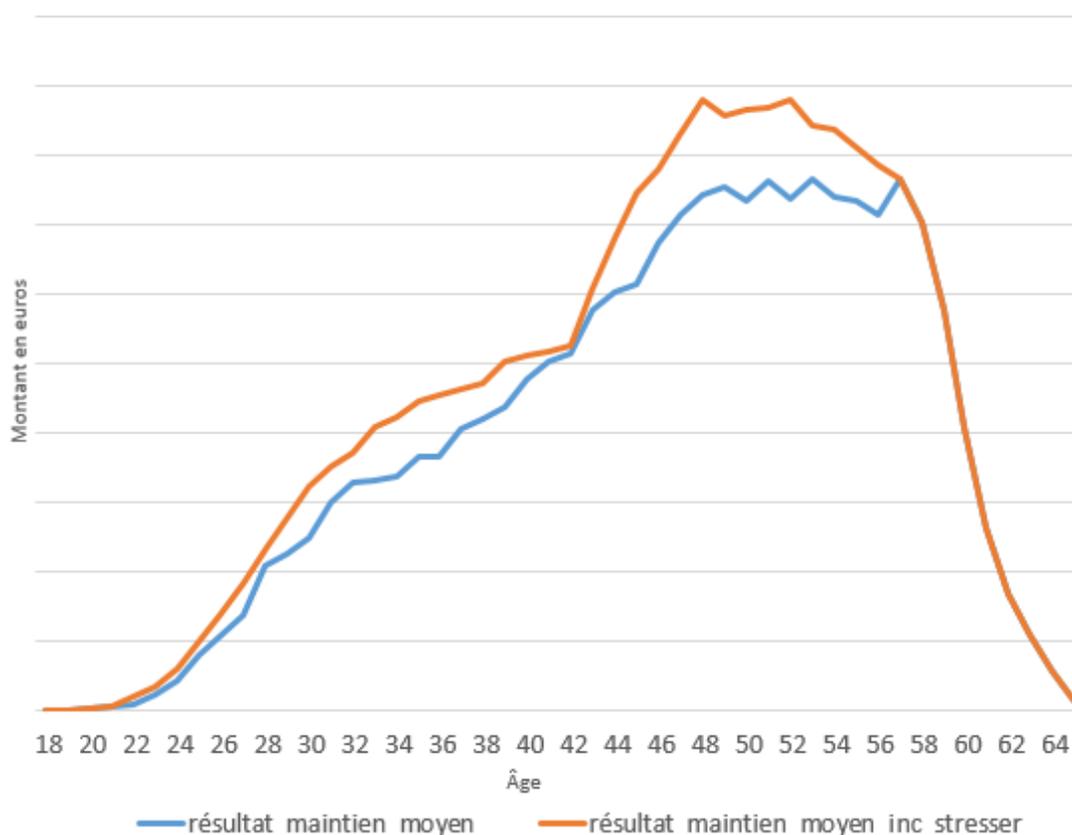


Figure 28 : Montants de sinistres moyen et maximum payés par la CNP pour le partenaire<sup>o1</sup>

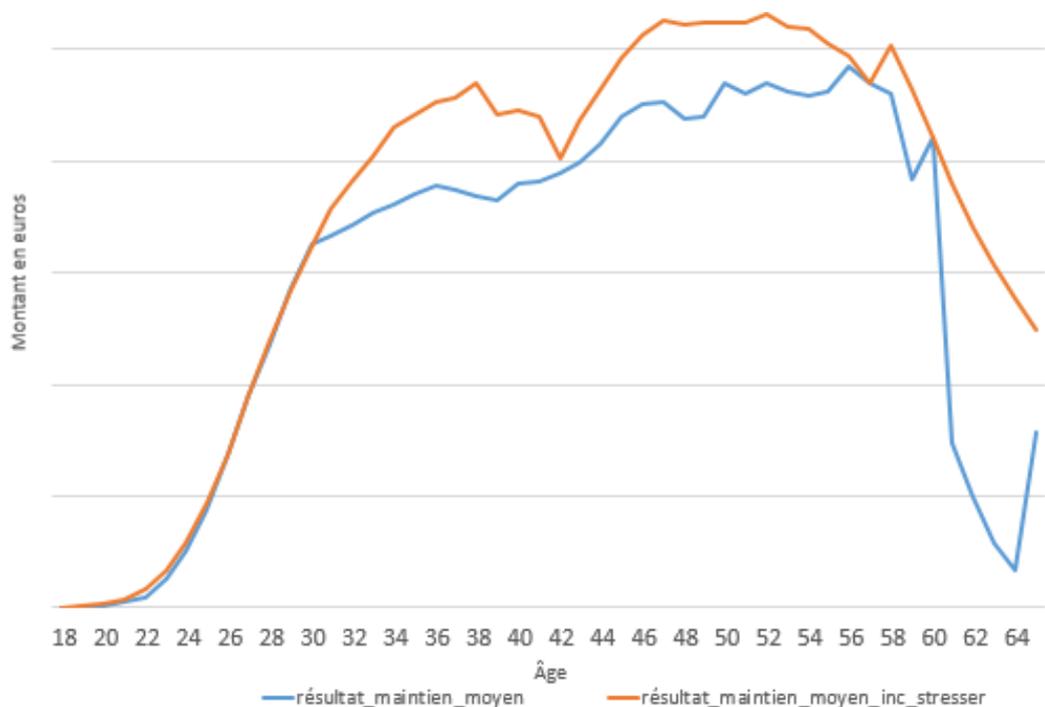


Figure 29 : Montants de sinistres moyen et maximum payés par la CNP pour le partenaire n°2

D'après ces graphiques, nous constatons que le montant de sinistres obtenu en utilisant l'incidence choquée était systématiquement supérieur ou égal à celui obtenu en utilisant l'incidence moyenne pour tous les groupes d'âge.

À présent, nous souhaitons quantifier la différence entre les montants moyens et les montants choqués pour chaque groupe d'âge afin de comparer les résultats bruts d'exploitation de chaque partenaire.

Pour ce faire, nous allons additionner les montants obtenus pour chaque groupe d'âge, en commençant par l'âge minimum de 18 ans et en allant jusqu'à l'âge maximum de 65 ans pour chaque partenaire.

#### 6.4. Comparaison des différents montants

Les graphiques ci-dessous présentent les montants potentiels que CNP ASSURANCES auraient à payer pour les deux partenaires, dans le scénario le plus défavorable parmi les 40 000 simulations effectuées.

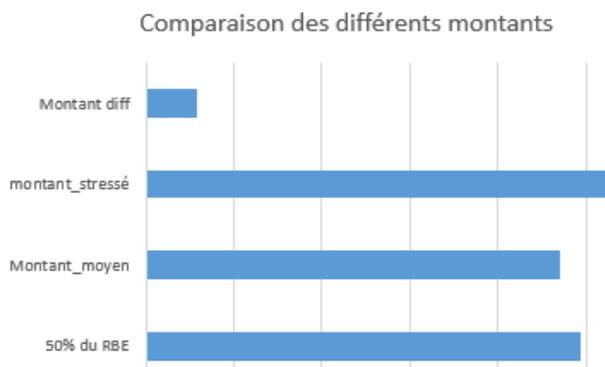


Figure 30 : Résultat du montant moyen payé par CNP pour le partenaire n°1.

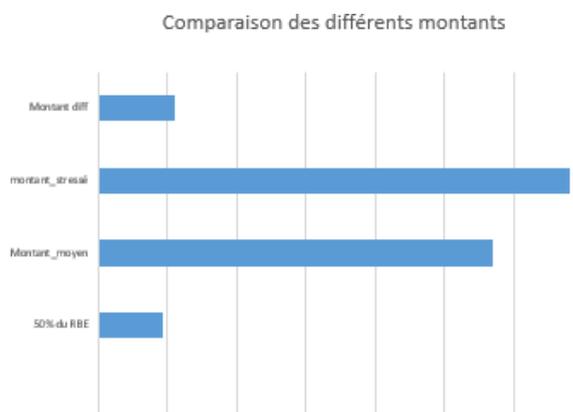


Figure 31 : Résultat du montant moyen payé par CNP pour le partenaire n°2.

Le montant stressé représente le montant obtenu à l'aide de l'incidence stressée.  
Le montant moyen représente le montant obtenu à l'aide de l'incidence moyenne.  
Le montant diff représente la différence entre le montant stressé et le montant moyen.

En comparant les résultats obtenus pour les différents partenaires, on constate que les montants moyens ainsi que les montants stressés sont nettement supérieurs au montant qui représente 50% du RBE.

Cependant, pour le partenaire numéro 2 uniquement, la différence entre le montant choqué et le montant moyen atteint 50% du résultat brut d'exploitation.

Cette analyse a été effectuée sur la base de 40 000 simulations pour chaque partenaire.

Il est donc clair que le stress du risque arrêt de travail a eu un impact significatif sur les montants à payer.

En effet, le montant obtenu pour le partenaire n°1 a augmenté de près de 12% par rapport au montant moyen, tandis que pour le partenaire n°2, l'augmentation a été de l'ordre de 20%.

Ces résultats soulignent l'importance de prendre en compte le risque d'arrêt de travail dans l'analyse actuarielle des portefeuilles d'assurance.

### 6.5. Analyse

La politique de réassurance de CNP Assurances a pour but d'utiliser la réassurance de manière ciblée afin de respecter les limites locales telles que la protection de 50% du RBE, tout en développant les opportunités de souscription, et en encadrant la sélection des réassureurs.

Les graphiques 6.3 et 6.4 ont démontré que le montant représentant 50% du RBE n'a pas été atteint pour le partenaire n°1, tandis que pour le partenaire n°2, le montant  $S_{dif}$  a atteint cette limite.

Cela indique que le risque d'arrêt de travail est plus sensible pour le partenaire n°2 que pour le partenaire n°1, et donc que le partenaire n°1 présente une meilleure capacité à absorber un tel risque.

Pour partager ce risque, une solution serait d'utiliser la réassurance, qui permet de réduire les risques pris par l'assureur en transférant une partie des risques qu'elle supporte à un ou plusieurs réassureurs en échange de compensations financières. Cependant, pour proposer un type de contrat de réassurance, il est nécessaire de connaître tous les partenaires de CNP Assurances afin de pouvoir mutualiser le risque et décider de faire appel ou non à la réassurance.

Ainsi, il serait judicieux d'examiner les autres partenaires de CNP Assurances et d'appliquer la même démarche de stress pour évaluer la nécessité de conclure un contrat de réassurance.

## 7. Conclusion

L'assurance emprunteur répond à un besoin économique réel en permettant aux emprunteurs de se protéger financièrement en cas de décès, d'invalidité ou de perte d'emploi. Cela permet aux emprunteurs de rembourser leur prêt sans mettre en danger leur situation financière personnelle et familiale.

Dans le cas d'une défaillance de l'assuré, les assurances doivent être en mesure de se protéger contre ce risque et ainsi se prémunir contre le risque d'arrêt de travail.

Ce mémoire a pour but de mesurer l'impact potentiel de ce risque sur le portefeuille de la CNP, en évaluant la probabilité qu'une personne couverte par CNP assurances soit en arrêt de travail en moyenne et en simulant le pire des cas.

Dans le but d'obtenir un montant de sinistre supposé maximum, nous avons simulé le couple (âge, incidence).

Les simulations à partir cet échantillon bivarié ont été faites à l'aide de la théorie des copules, et plus particulièrement des copules *-vines*.

De plus, la régression linéaire par morceaux qui nous ont donné la possibilité de simuler selon des segments spécifiques d'âge.

La méthode des copules de *vines* est une méthode flexible permettant de simuler des copules en choisissant les familles de copules les plus adaptées à un jeu de données afin de simuler leurs dépendances.

On peut observer que la méthode employée fonctionne et plus de 40 000 échantillons (âge, incidence) ont été simulés.

La méthode retenue pour choquer le risque arrêt de travail est de sélectionner le taux d'incidence maximum simulé pour chaque âge.

Néanmoins, l'étude menée dans ce mémoire présente des limites.

La première limite concerne les données utilisées. Cela inclut la période pendant laquelle nous avons observé les événements et les ajustements que nous avons effectués pour créer les bases de données des assurés et des sinistres.

En effet, les manipulations de données ont une influence directe sur les résultats que nous avons obtenus, étant donné que des retraitements ont été effectués sur les données. Ces ajustements ont entraîné une réduction significative de l'échantillon, avec une baisse marquée de 26 % des observations relatives aux assurés du portefeuille du partenaire n°1.

Par conséquent, il est recommandé de considérer les résultats avec précaution, en tenant compte des incertitudes liées aux données et de la qualité de ces dernières.

La deuxième limite est liée à la nécessité de fournir les informations relatives aux taux d'incidence minimum et maximum pour pouvoir répliquer les simulations. Ainsi, nos simulations sont dépendantes de ces informations.

Enfin, la dernière limite de cette méthode de simulation concerne la modélisation de l'âge. En effet, l'âge simulé n'est pas un nombre entier et seule la partie entière de l'âge est prise en compte, ce qui peut introduire une certaine imprécision dans les résultats.

D'autre part, faire appel à la réassurance est une décision complexe pour une compagnie d'assurance, car cela implique plusieurs aspects à prendre en compte, tel qu'une vision globale

du portefeuille, la négociation des contrats de réassurance, la gestion et les coûts supplémentaires qui y sont associés.

En ayant une vision partielle de ce portefeuille, c'est-à-dire seulement deux des nombreux partenaires de CNP, nous pouvons répondre de manière assez simpliste que pour le partenaire n° 2 la réassurance peut être une méthode de partage de risque intéressante.

L'analyse des simulations permet de comprendre l'impact de différents scénarios sur les résultats d'une compagnie d'assurance.

Les simulations ont montré que le montant des sinistres pour les deux partenaires a augmenté de 12% et 20%.

Dans ce mémoire, les scénarios de stress ont uniquement été appliqués aux taux d'incidence en arrêt de travail. Il serait opportun également de déformer les lois de maintien en arrêt de travail et de voir les impacts sur la charge de sinistres.

Il est important de noter que dans ce mémoire, le traitement spécifique du Covid-19 et des retraites n'a pas été abordé. Cependant, une étude complémentaire sur l'impact du Covid-19 est disponible en annexe, fournissant des informations supplémentaires sur cette crise sanitaire et son influence potentielle sur les résultats de l'étude.

# Bibliographie

## Méthodes de Monte Carlo

Nicolas BAUD, Thierry Roncalli, Jean-Frédéric Jouanin. "*Gestion des Risques Multiples*".

- <http://www.thierry-roncalli.com/download/Lecture-Notes-Copula-Ensay.pdf>

## Arrêt de travail : lois de provisionnement du BCAC

PLANCHET F. "*Ressources Actuarielles*".

- <http://www.ressources-actuarielles.net/bcac>

## Construction d'une loi de maintien en arrêt de travail et choc de la formule standard

Nicolas GEORGES. "*Mémoire d'actuariat*".

- <https://www.institutdesactulaires.com/docs/mem/f4c9dc2b8a03fdc5905d0e1518752f7c.pdf>

## Construction d'une table de mortalité d'expérience en assurance emprunteur

Manon DAL PONT. "*Mémoire d'actuariat*".

- <https://www.institutdesactulaires.com/docs/mem/1e992efa93786a498553e9a184326e4a.pdf>

## Impact de l'utilisation des tables d'expérience sur le provisionnement en pré-voyance

Nezha Fetoui. "*Mémoire d'actuariat*".

- <https://www.institutdesactulaires.com/docs/mem/e598f78dfd942b13019173a5b9f86282.pdf>

## Comment votre assurance de prêt prend-elle en charge vos échéances en cas de sinistre ?

Cardif. "*Article*".

- <https://www.cardif.fr/assurance-emprunteur/faire-jouer-ses-garanties-assurance-pret-en-cas-de-sinistre>

## L'ASSURANCE FRANÇAISE

France Assureurs. "Article".

- <https://www.franceassureurs.fr/wp-content/uploads/VF-Donnees-cles-2020.pdf>

## Malakoff Humanis présente les résultats 2021 de son Baromètre annuel absentéisme Maladie

Malakoff Humanis, "Article".

- <https://newsroom.malakoffhumanis.com/actualites/malakoff-humanis-presente-les-resultats-2021-de-son-barometre-annuel-absenteisme-maladie-e21d-63a59.html>

## Bibliothèques VineCopula De R

Bibliothèque de R VineCopula: Statistical Inference of Vine Copulas

- <https://cran.r-project.org/web/packages/VineCopula/index.html>

## MATVines : A vine copula package for MATLAB

Article scientifique.

- <https://www.sciencedirect.com/science/article/pii/S2352711021000455>

## Python Package for Sampling from Copulae : clayton

Article scientifique, Alexis Boulin.

- <https://computo.sfds.asso.fr/published-202301-boulin-clayton/>

## Régression : Théorie et applications

Ouvrage, Pierre-André Cornillon et Eric Matzner-Lober.

- <https://www.eyrolles.com/Sciences/Livre/regression-9782287396922/>

# ANNEXES

## 1. Les copules-vines

Dans le but de comprendre la mécanique générale de cette méthode, on prend l'exemple d'un vecteur à 3 variables ( $X_1, X_2, X_3$ ) :

$$f(x_1, x_2, x_3) = f_{3|1,2}(x_3|x_1, x_2) \times f_{2|1}(x_2|x_1) \times f_1(x_1)$$

D'après la formule de Bayes :

$$f_{2|1}(x_2|x_1) = f_2(x_2) \times c_{1,2}(F_1(x_1), F_2(x_2))$$

Avec :  $c_{1,2}(F_1(x_1), F_2(x_2))$  la densité de la copule liant les deux marginales  $F_1(x_1)$  et  $F_2(x_2)$ .

Ainsi :

$$f_{3|1,2}(x_3|x_1, x_2) = c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times f_{3|2}(x_3|x_2)$$

$$f_{3|1,2}(x_3|x_1, x_2) = c_{1,3|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\ \times c_{2|3}(F_2(x_2), F_3(x_3)) \times f_3(x_3)$$

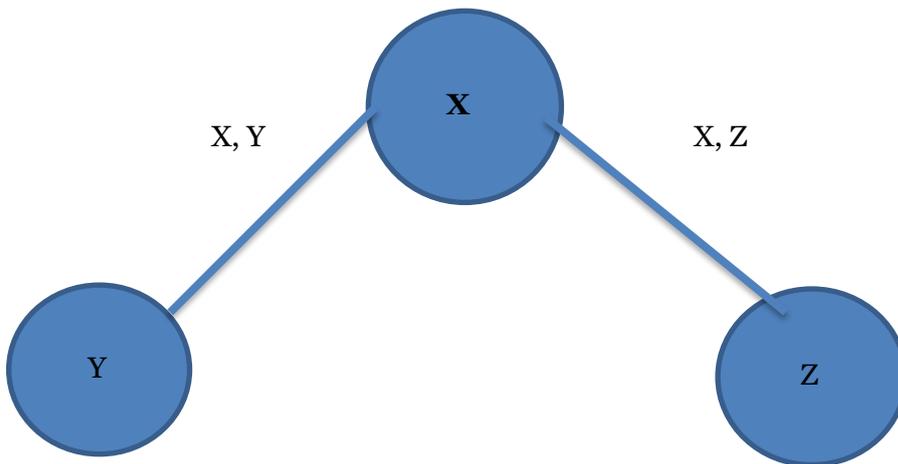
Finalement :

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3) \\ \times c_{1,2}(F_1(x_1), F_2(x_2)) \times c_{2,3}(F_2(x_2), F_3(x_3)) \\ \times c_{1,3|2}(F_1(x_1), F_3(x_3)|F_2(x_2))$$

## 2. La Canonical vine (C-vine)

Voici un exemple de décomposition en arbres de la méthode Canonical vine pour une distribution jointe à trois variables (X, Y, Z).

$T_1$  :



Dans cet arbre, chaque nœud représente une variable et chaque arrête représente une copule bivariée.

$T_2$  :

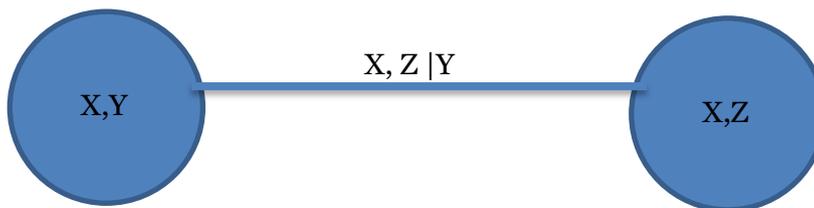


Figure : C-vine à 3 variables.

### 3. Code R

#### 1. Régression linéaire

Voici le code pour obtenir une régression linéaire ainsi eu les résultats de cette même cette régression linéaire qui en ressorte.

```
library("readxl")
base_inc_CE= read_excel("/Users/shaiaazoulay/Desktop/BDD_f_1807/inc/INC_R/INC_CE.xlsx" , sheet="INC_CE")
data_test<- data.frame(X = base_inc_CE`Age à l'inventaire`, Y = base_inc_CE$inc_tot)

x_test=data_test$X
y_test=data_test$Y

x_test
y_test

data_test2<- data.frame(x=x_test, y=y_test)
data_test2
reg1= lm(y ~ x, data = data_test2)
plot(reg1)
summary(reg1)
```

Figure 32 : Code R pour obtenir la régression linéaire.

```
Call:
lm(formula = y ~ x, data = data_test2)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0066942 -0.0005955 -0.0001455  0.0008145  0.0043233

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.787e-04  9.443e-04  -0.189   0.851
x            1.118e-04  2.158e-05   5.180  4.8e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.002072 on 46 degrees of freedom
Multiple R-squared:  0.3684,    Adjusted R-squared:  0.3546
F-statistic: 26.83 on 1 and 46 DF,  p-value: 4.8e-06
```

Figure 33 : Coefficient de détermination du partenaire n°1.

On peut observer une p-value<0,5 et un coefficient de détermination égal à 35%.

## 2. Le package 'segmented' de R

Le package segmented offre des fonctionnalités pour estimer et résumer des modèles linéaires avec relations segmentées.

La fonction "segmented" du logiciel R :

De plus, ci-dessous le code pour obtenir une régression linéaire par morceaux.

```
```{r}
xx_ce <- base_inc_CES`Age à l'inventaire` # récupération de nos données
yy_ce <- base_inc_CES`inc_tot`
dati2_ce <- data.frame(x = xx_ce, y = yy_ce)

out.lm_ce <- lm(y ~ x, data = dati2_ce) #régression linéaire
o_ce <- segmented(out.lm_ce, seg.Z = ~x, psi = list(x = c(21,43,59)), #utilisation de la fonction segmented
  control = seg.control(display = FALSE)
)
dat2_ce = data.frame(x = xx_ce, y = broken.line(o_ce)$fit)

ggplot(dati2_ce, aes(x = x, y = y)) +
  geom_point() +
  geom_line(data = dat2_ce, color = 'blue') # tracer la régression linéaire par morceaux

plot(out.lm_ce )
plot(o_ce)

summary(o_ce)
```
```

Figure 34 : Code R pour obtenir la régression linéaire par morceaux

### 3. La variable rang

Comme indiqué dans notre mémoire, nous avons utilisé la méthode des rangs avec les copules pour transformer notre échantillon bivarié. Ci-dessous, nous présentons le code que nous avons utilisé pour effectuer cette transformation.

Sous le logiciel R :

- La fonction `rank()` est utilisée pour renvoyer les exemples de rangs des valeurs d'un vecteur.
- La fonction `length()` est utilisée pour connaître la longueur d'un vecteur.

Exemple de l'utilisation du rang sur un vecteur  $x_i=(2,4,1,-3,7,3)$  .Alors, le rang de cette échantillon est  $R_i=(3,5,2,1,6,4)$  .

```
# uniform by ranking
U1 <- rank(sigma.ok)/(length(sigma.ok)+1)
U2 <- rank(r.ok)/(length(r.ok)+1)
```

Figure 35 : Code R pour ordonner les données uniformément par leur rang

### 4. Package "VinesCopula"

Le package "VinesCopula" offre des fonctionnalités pour l'analyse statistique des modèles de copules de vigne, y compris l'estimation des paramètres, la sélection de modèles, la simulation, les tests d'ajustement et la visualisation.

Des outils d'estimation, de sélection et d'analyse exploratoire de données de modèles de copules bivariées sont également fournis.

Dans notre cas, c'est la fonction `RVineSim()` qui va nous permettre de simuler un nouvel échantillon de données pour l'âge et l'incidence à partir de notre vecteur  $V = (V_1, V_2)$ .

Pour simuler un échantillon de données pour l'âge et l'incidence, nous utilisons la fonction `RVineSim()` du package `VineCopula` du CRAN.

La structure du modèle et les estimations sont fournies en entrée, et le résultat est une matrice de simulations de taille variable selon le nombre de variables.

Chaque élément de la matrice se situe entre 0 et 1, nécessitant une transformation inverse des résidus standardisés à l'aide de `qunif()`.

En utilisant la fonction `qunif()` de R avec les taux min et max d'incidence, nous obtenons la fonction quantile de la distribution uniforme.

#### 4. Étude complémentaire : Le Covid-19

Le COVID-19, également connu sous le nom de coronavirus, est une maladie infectieuse causée par le virus SARS-CoV-2. Il a été identifié pour la première fois en décembre 2019 dans la ville de Wuhan, en Chine. Depuis lors, il s'est rapidement propagé à travers le monde, provoquant une pandémie.

En raison de la propagation rapide du virus et de ses effets graves sur la santé publique, de nombreuses mesures de confinement et de distanciation sociale ont été mises en place dans le monde entier. Ces mesures comprennent des fermetures d'entreprises, des restrictions de voyage, des ordres de rester à domicile, des quarantaines et des confinements stricts.

La pandémie de Covid-19 a engendré des bouleversements majeurs à l'échelle mondiale, affectant tous les domaines de notre société, y compris la santé, l'économie et les interactions sociales. Cette crise sanitaire sans précédent a nécessité une analyse approfondie de ses impacts afin de mieux comprendre ses conséquences et d'orienter les mesures appropriées.

Dans cette étude, nous nous concentrons sur l'impact du Covid-19 en termes de charge de sinistre pour deux des partenaires de CNP ASSURANCES.

Notre objectif est d'évaluer comment cette pandémie a influencé les coûts des sinistres et de différencier les charges de sinistre maximale et moyenne en distinguant deux périodes d'étude : l'année 2020, marquée par le début de la crise du Covid-19, et les années précédentes de 2015 à 2019.

Pour atteindre cet objectif, nous analyserons les variations des taux d'incidence du risque arrêt de travail du portefeuille emprunteur de CNP ASSURANCES, des sinistres entre ces périodes spécifiques.

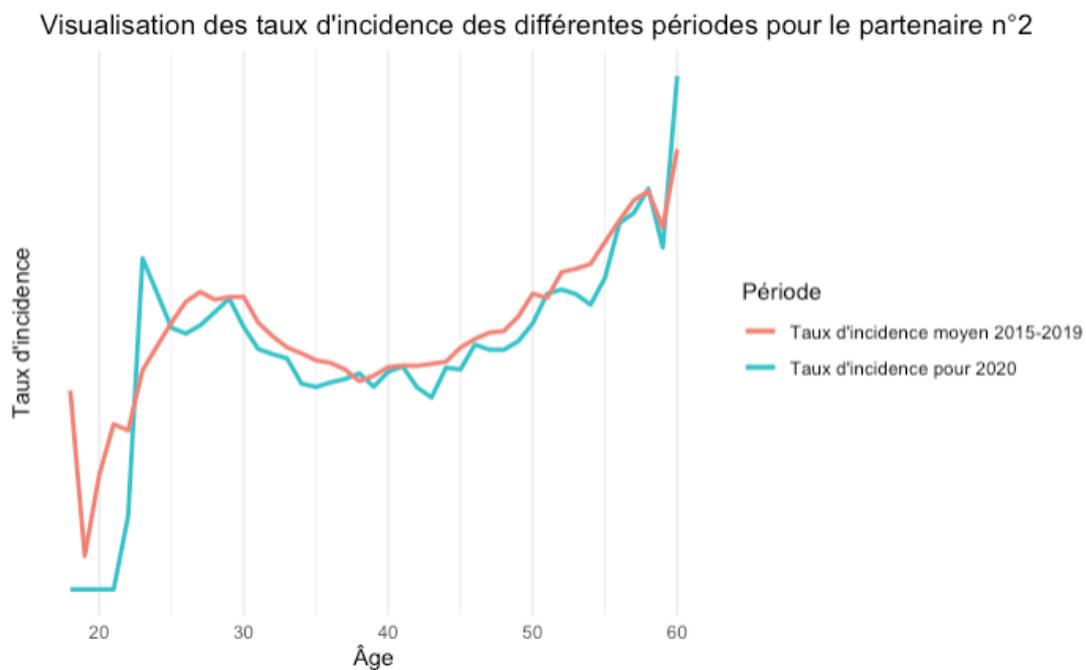
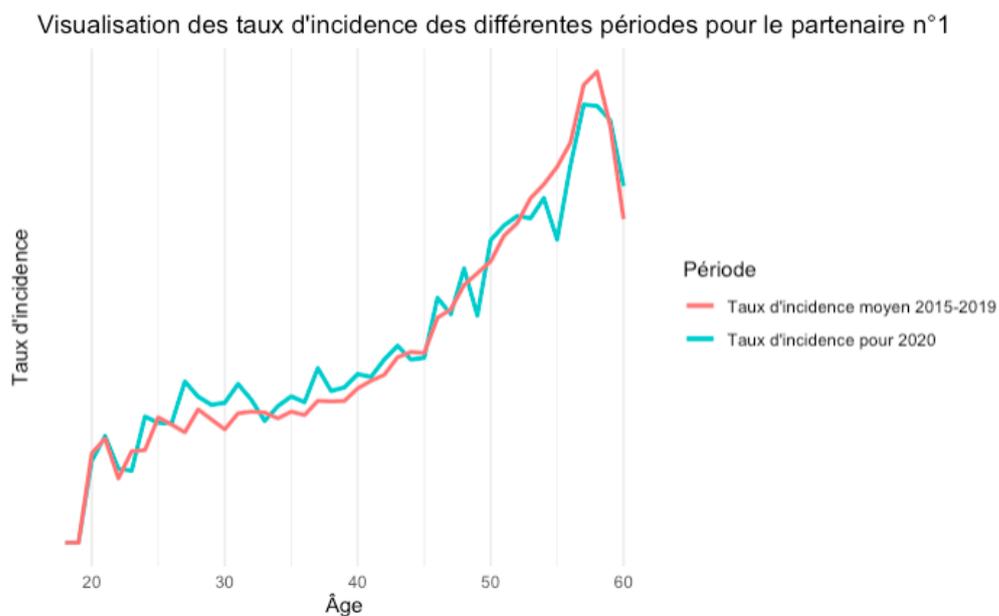
En examinant les données disponibles, nous identifierons les différences significatives dans les taux d'incidence des sinistres entre les périodes étudiées. Cela nous permettra d'estimer une charge de sinistre supposée maximale grâce à nos simulations et moyenne pour chaque période, en prenant en compte les variations observées.

L'étude complémentaire a pris en compte les personnes âgées de 18 à 60 ans pour l'analyse des données.

*Pour des raisons de confidentialité, les noms des partenaires ainsi que de nombreux résultats ne seront pas transmis dans ce mémoire. Il est donc conseillé au lecteur de se focaliser sur la méthodologie plus que sur les résultats eux-mêmes.*

Le chapitre 3 de ce mémoire nous permet de calculer le taux d'incidence en fonction de l'âge. Ainsi, nous sommes en mesure de mettre en évidence ce risque en utilisant une méthode de simulation détaillée décrite dans ce chapitre.

Voici la représentation des taux d'incidence pour chaque période et chaque partenaire :



De manière générale, pour le partenaire n°2, les taux d'incidence moyens de 2015 à 2019 sont supérieurs à ceux calculés pour l'année 2020, sauf dans certaines situations, comme à l'âge de 23 ans. En revanche, pour le partenaire n°1, on observe des fluctuations plus nombreuses entre ces deux périodes.

Par ailleurs, en appliquant la méthode de régression linéaire simple pour chaque partenaire et chaque période, nous avons observé les résultats suivants :

Pour le partenaire n°1 :

- Pour l'année 2020, le coefficient de détermination ( $R^2$ ) est de 87,28%. Cela indique que 87,28% de la variation des taux d'incidence pour cette année peut être expliquée par la régression linéaire.
- Sur la période de 2015 à 2020, le coefficient de détermination ( $R^2$ ) est de 86,44%. Cela signifie que 86,44% de la variation des taux d'incidence sur cette période peut être expliquée par la régression linéaire.

Pour le partenaire n°2 :

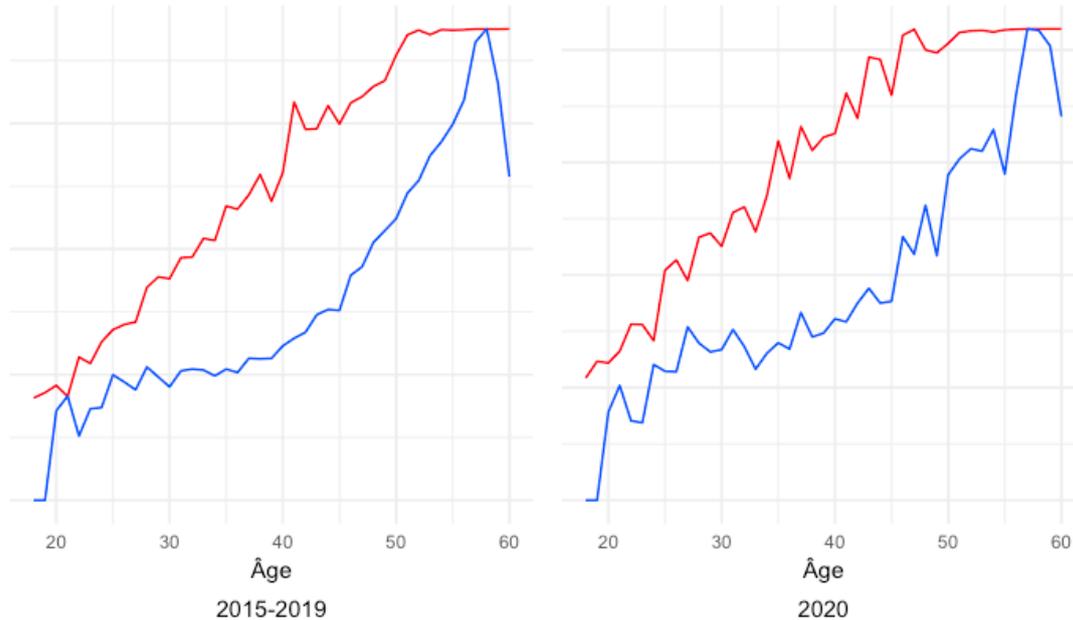
- Pour l'année 2020, le coefficient de détermination ( $R^2$ ) est de 45,11%. Cela indique que 45,11% de la variation des taux d'incidence pour cette année peut être expliquée par la régression linéaire.
- Sur la période de 2015 à 2020, le coefficient de détermination ( $R^2$ ) est de 51,64%. Cela signifie que 51,64% de la variation des taux d'incidence sur cette période peut être expliquée par la régression linéaire.

Ces coefficients de détermination donnent une indication de la qualité de l'ajustement de la régression linéaire aux données observées. Plus le coefficient de détermination est élevé, plus la régression linéaire fournit une explication précise des variations des taux d'incidence.

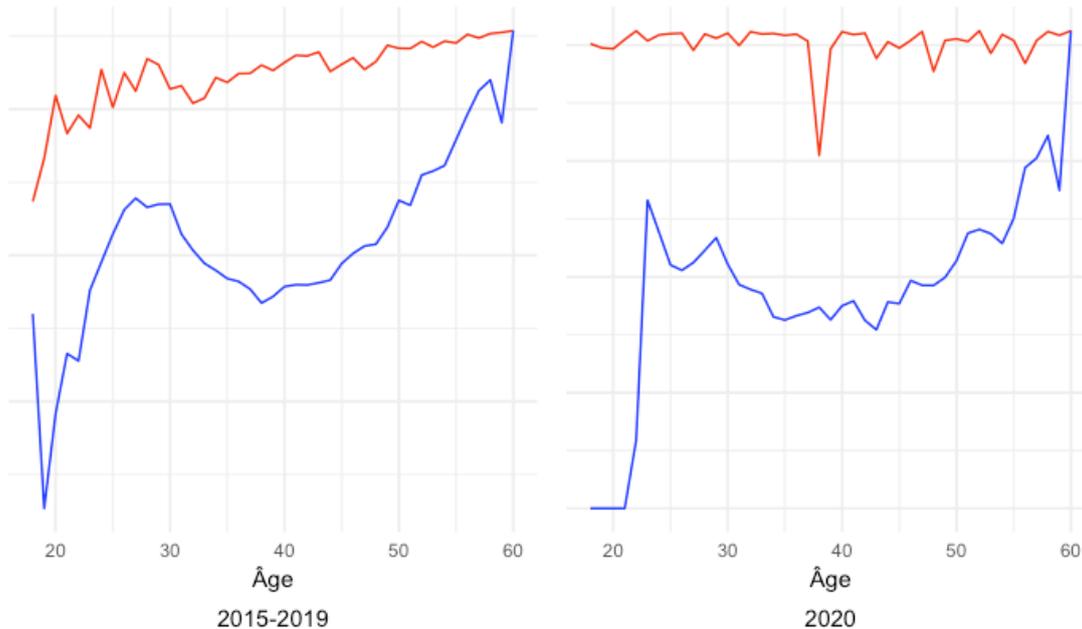
Ces résultats suggèrent que le modèle de régression linéaire simple est adapté pour analyser les données. Par conséquent, nous pouvons procéder aux simulations à l'aide de la théorie des copules.

Grâce à l'application de la méthode de simulation détaillée décrite dans le chapitre 3, nous avons obtenu des résultats simulés pour chaque partenaire et pour chaque période.

Comparaison du taux d'incidence simulé et moyen pour le partenaire n°1



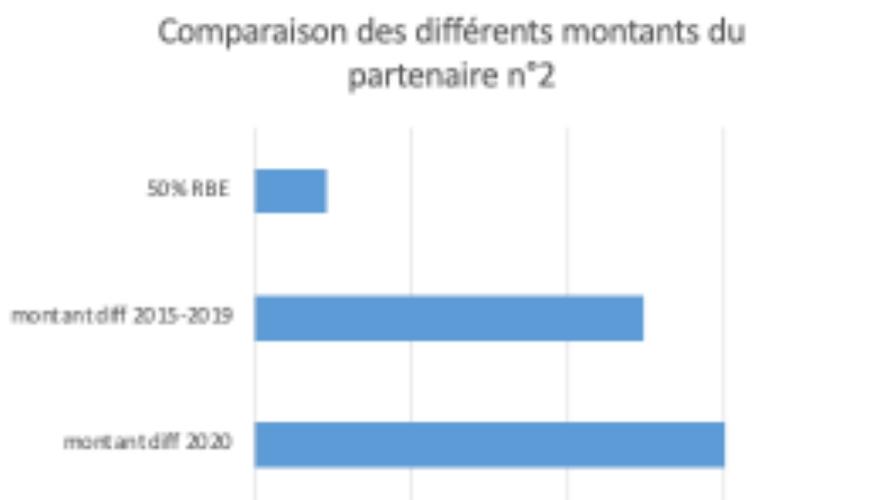
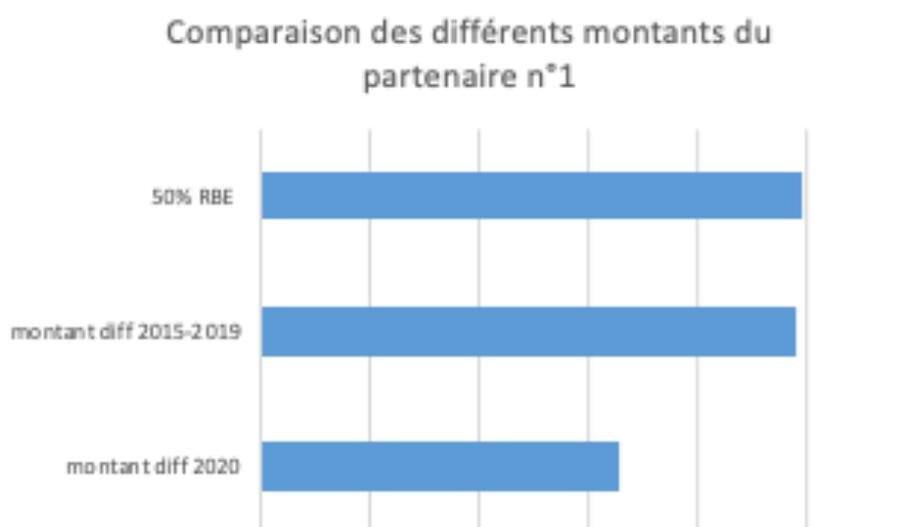
Comparaison du taux d'incidence simulé et moyen pour le partenaire n°2



Les taux d'incidence simulés sont représentés en rouge, tandis que les taux d'incidence moyens sont représentés en bleu.

Pour appliquer cette méthode, nous avons sélectionné une structure de famille C-vine, et la famille de copules utilisée est la famille de copules de Student. Cette famille a été choisie car elle permet de modéliser efficacement les dépendances linéaires croissantes observées dans les données.

En fin de compte, en appliquant la méthode d'estimation de la charge des sinistres décrite au chapitre 6, nous obtenons les résultats suivants :



Par ailleurs, afin d'obtenir des résultats pertinents, nous avons distingué les données de 2020, année impactée par la Covid, de celles des autres années. Cette démarche a été appliquée tant pour le calcul de la rente que pour celui du maintien. Isoler ainsi les données de l'année Covid nous permet d'avoir une vision plus claire des résultats, sans l'influence de cette année atypique.

En analysant ces deux graphiques, on observe que seul le partenaire N°2 atteint le seuil de 50% du RBE, critère fixé par la politique de CNP ASSURANCES pour déclencher l'utilisation ciblée de la réassurance. Le partenaire N°1, quant à lui, frôle ce seuil sans toutefois l'atteindre.

En conclusion de cette étude sur l'analyse de l'impact spécifique du Covid-19 sur la charge des sinistres, en distinguant les périodes pendant la pandémie et celles d'avant, l'évaluation des variations des taux d'incidence des sinistres nous a permis d'estimer les charges maximales et moyennes, afin de mieux appréhender les conséquences financières de cette crise sanitaire mondiale.

De plus, bien que la méthodologie de simulation ait utilisé la théorie des copules, la détermination de la tranche d'âge pour la simulation a été réalisée uniquement à l'aide de la régression linéaire simple.

Il est important de souligner que le taux d'incidence du risque d'arrêt de travail a été globalement plus faible pendant l'année de début de la pandémie par rapport aux années précédentes. Ce constat peut s'expliquer par l'impact des mesures de distanciation sociale, des restrictions de déplacement et de la sensibilisation accrue à la santé et à l'hygiène, qui ont contribué à réduire les accidents du travail. De plus, l'adoption généralisée du télétravail a également joué un rôle dans la limitation des risques liés aux déplacements professionnels et aux accidents sur le trajet domicile-travail.

Finalement, nous confirmons que le risque d'arrêt de travail du portefeuille n°2 se révèle plus sensible que celui du partenaire n°1, étant donné que le seuil de 50% du RBE, critère déterminant pour une utilisation ciblée de la réassurance selon la politique de CNP ASSURANCES, n'est atteint que par le portefeuille n°2. En conséquence, il est crucial de s'interroger sur la pertinence de recourir à la réassurance pour le partenaire n°2, en vue d'une gestion optimale des risques.

## Liste des figures

|   |    |
|---|----|
| Figure 1 : Graphiques des taux d'incidence moyens par âge pour les deux partenaires.....                  | 10 |
| Figure 2 : Graphique des taux d'incidence simulés pour les deux partenaires. ....                         | 11 |
| Figure 3 : Lois de maintien en arrêt de travail des deux partenaires .....                                | 12 |
| Figure 4 : Comparaison des différents montants pour les deux partenaires. ....                            | 13 |
| Figure 5 : Chiffres d'affaires de l'assurance emprunteur en France par garantie en milliards d'euros..... | 27 |
| Figure 6 : Répartition moyenne (2016-2020) des cotisations d'assurance emprunteur par type de prêt.....   | 28 |
| Figure 7 : Frise chronologique des récentes lois en assurance emprunteur. ....                            | 30 |
| Figure 8 : Graphique d'incidence par âge pour le partenaire n°1.....                                      | 42 |
| Figure 9 : Graphique d'incidence par âge pour le partenaire n°2. ....                                     | 42 |
| Figure 10 : Graphique d'incidence par année pour le partenaire n°1. ....                                  | 44 |
| Figure 11 : Graphique d'incidence par année pour le partenaire n°2. ....                                  | 44 |
| Figure 12 : Graphique d'incidence moyen par âge pour le partenaire n°1.....                               | 45 |
| Figure 13 : Graphique d'incidence moyen par âge pour le partenaire n°1.....                               | 46 |
| Figure 14 : La régression linéaire par morceaux sur le partenaire n°1.....                                | 50 |
| Figure 15 : Vérification des fonctions linéaires pour le partenaire n°1.....                              | 51 |
| Figure 16 : La régression linéaire par morceaux sur le partenaire n°2 .....                               | 51 |
| Figure 17 : Vérification des fonctions linéaires pour le partenaire n°2 .....                             | 52 |
| Figure 18 : Graphique de dispersion pour le partenaire n°1.....   | 60 |
| Figure 19 : Graphique de dispersion pour le partenaire n°2 .....  | 60 |
| Figure 20 : Graphique des taux d'incidence simulés pour les deux partenaires .....                        | 62 |
| Figure 21 : Censure à gauche.....   | 68 |
| Figure 22 : Censure à droite. ....  | 69 |
| Figure 23 : Troncature à gauche.....  | 69 |
| Figure 24 : Troncature à droite .....   | 70 |
| Figure 25 : Lois de maintien pour la tranche 45-50 ans du partenaire n°1 .....                            | 71 |
| Figure 26 : Lois de maintien pour la tranche 40-45 ans du partenaire n°2 .....                            | 72 |
| Figure 27 : Loi de maintien chaînée à la tranche 40-50 ans du BCAC .....                                  | 74 |
| Figure 28 : Montants de sinistres moyen et maximum payés par la CNP pour le partenaire n°1 .....          | 79 |
| Figure 29 : Montants de sinistres moyen et maximum payés par la CNP pour le partenaire n°2.....           | 80 |
| Figure 30 : Résultat du montant moyen payé par CNP pour le partenaire n°1. ....                           | 81 |
| Figure 31 : Résultat du montant moyen payé par CNP pour le partenaire n°2. ....                           | 81 |
| Figure 32 : Code R pour obtenir la régression linéaire. ....  | 89 |
| Figure 33 : Coefficient de détermination du partenaire n°1.....   | 89 |
| Figure 34 : Code R pour obtenir la régression linéaire par morceaux.....                                  | 90 |
| Figure 35 : Code R pour ordonner les données uniformément par leur rang.....                              | 91 |

## Liste des tableaux

|  |           |
|--|-----------|
| Tableau 1 : Les différentes durées résiduelles de chaque partenaire.....   | 12        |
| Tableau 2 : Les montants de rentes obtenus .....   | 12        |
| Tableau 3 : nombre d'observations par base de données d'assurés .....  | 31        |
| <i>Tableau 4 : Descriptif des variables retenues issues de la table MIP_BU .....</i>                             | <i>32</i> |
| Tableau 5 : Descriptif des variables retenues issues de la table PMUV55_VISION_ASSURE.....                       | 33        |
| Tableau 6 : Nombre d'observations par base de données de sinistres .....   | 34        |
| <i>Tableau 7 : Descriptif des variables issues de la base C7SBASE.C7s_BUR .....</i>                              | <i>34</i> |
| <i>Tableau 8 : Descriptif des variables retenues issues de la table TDO_R_SIACC_CONT_CATALOG .....</i>           | <i>35</i> |
| <i>Tableau 9 : Descriptif des variables retenues issues de la base TDO_D_SIACS_COMMUN_SINIS_PRET .....</i>       | <i>35</i> |
| <i>Tableau 10 : Descriptif des variables retenues issues de la base SIACS.TDO_D_SIACS_SINIS_PRET.....</i>        | <i>36</i> |
| Tableau 11 : Descriptif des variables supplémentaires .....  | 37        |
| <i>Tableau 12 : Taux d'incidence par année.....</i>  | <i>43</i> |
| Tableau 13 : Les différentes fonctions linéaires par morceaux du partenaire n°1 .....                            | 50        |
| Tableau 14 : Les différentes fonctions linéaires par morceaux du partenaire n°2.....                             | 52        |
| <i>Tableau 15 : Les différents coefficients de déterminations, partenaire n°1.....</i>                           | <i>53</i> |
| Tableau 16 : Les différents coefficients de déterminations, partenaire n°2 .....                                 | 53        |
| <i>Tableau 17 : Les différents familles et structures de copules, partenaire n°1.....</i>                        | <i>59</i> |
| <i>Tableau 18 : Les différents familles et structures de copules, partenaire n°2.....</i>                        | <i>59</i> |
| Tableau 19 : Extrait de la loi de maintien en incapacité du BCAC (Bureau commun des assurances collectives)..... | 65        |
| <i>Tableau 20 : Les différentes durées résiduelles de chaque partenaire. ....</i>                                | <i>73</i> |
| Tableau 21 : Comparaison des durées résiduelles.....   | 75        |
| <i>Tableau 22 : Montant de rente moyenne par année de survenance du partenaire n°1.....</i>                      | <i>77</i> |
| <i>Tableau 23 : Montant de rente moyenne par année de survenance du partenaire n°2.....</i>                      | <i>77</i> |

