Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 10/11/2021

Par : **Massa COULIBALY**

Titre: **Construction de table de maintien en perte d'emploi en assurance emprunteur**

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

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

Membres présents du jury de la filière

Nicolas BARADEL

Entreprise : Société Générale Assurance

Nom : Elsa CHRETIEN
Signature :

Membres présents du jury de l'Institut des Actuaires

Directeur de mémoire en entreprise :

Nom : Marie-Anne PINETTE
Signature :

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

Secrétariat :

Signature du candidat

Bibliothèque :

Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE)
5, avenue Henry Le Chatelier - 91120 PALAISEAU, FRANCE

Remerciements

Je tiens à remercier tout particulièrement ma tutrice de mémoire, Marie-Anne PINETTE, qui a su me conseiller et me faire profiter de son expérience tout au long du mémoire malgré un emploi du temps surchargé, et sans qui ce mémoire aurait difficilement vu le jour.

Je remercie également Jérôme BEAUVIR, Responsable technique Actuariat Prévoyance, et Nicoleta NOKO, Responsable Pôle Valeur, ainsi que tous les membre du pôle Valeur, pour la qualité de leur accueil au sein de l'équipe Actuariat Prévoyance, pour l'atmosphère positive et le dynamisme qui y règnent, et qui ont contribué à élaborer et enrichir mon mémoire au fil de cette expérience professionnelle.

Résumé

Mots-clés : Table de maintien en perte d'emploi, Assurance des emprunteurs, Kaplan-Meier, Censure, Whittaker-Henderson, Modèle de Cox, Chi-deux

Ce mémoire a pour but de retracer les différentes étapes de construction d'une table d'expérience de maintien en perte d'emploi sur le portefeuille du produit Expresso, produit d'assurance emprunteur de la Société Générale Assurance, ainsi que l'impact de son utilisation en provisionnement.

Le travail a été divisé en cinq parties. Dans la première, il a été question de poser le cadre général de l'étude en revenant sur certaines notions importantes et en présentant le produit Expresso. La deuxième partie s'est focalisée sur la création et la fiabilisation de la base de données ainsi que la réalisation de statistiques descriptives sur le portefeuille.

Après avoir présenté théoriquement la méthode de Kaplan-Meier pour l'estimation des taux bruts et la méthode de Whittaker-Henderson pour le lissage de ces derniers dans la troisième partie, nous les avons appliquées au portefeuille d'Expresso dans la quatrième partie. Pour le choix de la courbe de taux lissés finale, nous avons retenu la courbe suffisamment lisse qui avait les meilleures performances en termes de prédiction du nombre de sorties en 2019.

En outre, nous avons également étudié l'influence des covariables comme le sexe, le montant initial du prêt, le montant des échéances du prêt, l'âge et l'ancienneté lors de la survenance du sinistre sur les taux de sortie à travers le test du LogRank et les modèles de Cox (simple et étendu). Par ailleurs, l'effet de la crise sanitaire du Covid-19 sur la durée passée au chômage a également fait l'objet d'un modèle de Cox étendu.

Enfin, dans la cinquième partie, nous avons montré comment est utilisée la table de maintien retenue pour le calcul des provisions et avons étudié l'impact de l'utilisation de cette dernière sur les provisions en mars 2021.

Il est toutefois important de garder à l'esprit que les hypothèses faites lors du traitement des données (identification des sinistres et conditions de censures) peuvent avoir un impact plus ou moins important sur les taux. Par ailleurs, l'insuffisance des observations dans la base de données utilisée peut causer des instabilités de la table dans le temps et par conséquent la table doit toujours faire l'objet de suivi surtout lorsqu'on observe des changements sur la composition du portefeuille.

Abstract

Keywords: Job Loss Maintenance Table, Creditor Insurance, Kaplan-Meier, Censorship, Whittaker-Henderson, Cox Model, Chi-square

The purpose of this thesis is to show the different stages in the construction of a job loss maintenance table on the portfolio of the product Expresso, a Société Générale Assurance's credit insurance product, as well as the impact of its use in reserving.

The work was divided into five parts. The first part sets the general framework of the study by defining several important concepts and by presenting the insurance product Expresso. The second part aims to create and improve the reliability of the database and to produce descriptive statistics on the portfolio.

After presenting theoretically the Kaplan-Meier methods for estimating crude rates and the Whittaker-Henderson method for smoothing the latter, in part three, we applied them to Expresso's portfolio, in part four. For the choice of the final smoothed rate curve, we retained the sufficiently smooth curve which predicted more accurately the number of exits of unemployment in 2019.

In addition, we also investigated the effect of covariates such as the sex, the initial loan amount, the amount of monthly financial commitment of the loan, the insured's age and the seniority when the loss occurs on the probability of exiting unemployment through LogRank test and Cox models (simple and extended). Moreover, the effect of the Covid-19 health crisis on duration of job loss was also the subject of an extended Cox model.

Finally, in the last part, we have shown how the maintenance table is used for the calculation of reserves and we studied the impact of the use of the latter on reserves in March 2021.

However, it is very important to remember that the assumptions made during data processing (identification of claims and censorship conditions) can have a significant impact on the rates. In addition, the insufficiency of observations in the database used can cause instabilities in the time of the table and therefore the table should always be followed up especially if the composition of the portfolio changes.

Table des matières

Table des figures

Liste des tableaux

Introduction

Dans un contexte réglementaire en pleine mutation, avec la directive solvabilité II et la norme IFRS 17 qui entrera en vigueur à partir du 1er janvier 2023, les compagnies d'assurance sont plus que jamais contraintes à calculer le *best estimate* de leurs provisions. En effet, ces nouvelles normes encouragent les assureurs à évaluer les risques en se fondant sur des hypothèses basées sur la « réalité » (avec une forte composante aléatoire).

Dans des garanties comme le décès, l'incapacité ou l'invalidité présentes dans les contrats d'assurance emprunteur, cela se traduit par l'utilisation de tables réglementaires ou de tables d'expérience du portefeuille. Contrairement aux autres garanties, la garantie perte d'emploi ne dispose pas de tables réglementaires (ni pour l'entrée, ni pour le maintien en perte d'emploi). Ainsi, le pôle Valeur de l'équipe Actuariat Prévoyance de la Société Générale Assurance, est obligé de calculer des provisions très prudentes pour les garanties perte d'emploi par manque de table de loi de maintien. Par conséquent, pour répondre aux exigences réglementaires avec le calcul du *best estimate* et d'éviter de sur-provisionner le risque perte d'emploi, il a été décidé de mettre en place un ensemble d'études visant à construire des tables d'expériences de maintien sur la perte d'emploi.

Ce mémoire relate nos travaux réalisés sur ce sujet pour un des produits d'assurance emprunteur de la Société Générale Assurance. Il a pour but d'exposer les différentes étapes de construction d'une table de maintien d'expérience sur le portefeuille du produit Expresso ainsi que son utilisation dans le calcul des provisions.

Afin d'atteindre cet objectif, ce mémoire sera articulé autour de cinq parties. La première posera le cadre général de l'étude en se focalisant sur la définition de certains concepts et sur la présentation du produit Expresso, notamment sa garantie perte d'emploi. La deuxième partie s'attellera d'abord de montrer en détail la construction et la fiabilisation de la base de données avant de présenter des statistiques descriptives sur le portefeuille. Dans la troisième partie, il s'agira de présenter théoriquement les outils utilisés pour la construction de table de loi (sans et avec covariables). Les résultats obtenus en appliquant ces techniques sur le portefeuille étudié seront présentés dans la quatrième partie. Enfin, la cinquième partie mettra en évidence l'utilisation de la table de maintien retenue dans le calcul des provisions.

Partie I. Cadre général de l'étude

I.1 L'assurance des emprunteurs (ADE)

I.1.1 Définition et caractéristiques de l'assurance des emprunteurs

L'assurance emprunteur est une assurance qui garantit la prise en charge de tout ou partie des échéances de remboursement ou du capital restant dû d'un crédit en cas de survenance de certains événements. Ces événements sont le plus souvent le décès, la perte totale et irréversible d'autonomie (PTIA), l'invalidité permanente, l'incapacité temporaire de travail (ITT) et la perte d'emploi. D'après la FFA[1], en 2019, le marché de l'assurance emprunteur avait fait un chiffre d'affaire de 9,8 milliards d'euros.

I.1.1.1 *Type de crédits*

Il s'agit de crédits bancaires plus précisément les crédits immobiliers, crédits à la consommation et les crédits bail.

Le crédit immobilier concerne les opérations d'achat d'un bien immobilier à usage d'habitation (ou mixte professionnel et d'habitation) ou d'un terrain destiné à sa construction. La valeur des encours des crédits immobiliers en France s'élève à 1.270,2 milliards d'euros et presque un ménage sur trois a au moins un crédit immobilier. Du point de vue de l'assurance des emprunteurs, selon la FFA, les primes reçues sur les crédits immobiliers représentaient 71,43% des primes en assurance emprunteur en 2019.

Le crédit à la consommation est une opération de crédit pouvant prendre la forme d'un prêt (amortissable, renouvelable, etc.), d'un découvert en compte ou d'un délai de paiement, destiné à financer des besoins personnels, tels que l'achat de biens mobiliers (voiture, électro-ménager, etc.) ou la fourniture de prestation de services (travaux, voyages, etc.). Sa durée doit être supérieure à 1 mois, son montant au minimum de 200 euros et au maximum de 75 000 euros. Les crédits destinés à financer des besoins professionnels ne sont pas des crédits à la consommation. Par ailleurs, depuis le 1er juillet 2016, sont considérés comme crédit à la consommation, les crédits accordés pour financer exclusivement les dépenses de réparation, amélioration ou entretien d'un immeuble d'habitation quand ils ne sont pas garantis par une hypothèque ou une autre sûreté comparable. Du point de vu de l'assurance des emprunteurs, les

[1] Fédération Française de l'Assurance

primes reçues sur les crédits consommations représentaient 20,41% des primes en assurance emprunteur en 2019 selon la FFA.

Le crédit-bail est un contrat de location d'une durée déterminée, avec option d'achat à terme. Pendant la durée du contrat, l'utilisateur n'est pas juridiquement le propriétaire du bien. Il peut néanmoins le devenir, mais à l'échéance du contrat.

I.1.1.2 *Types de garanties*

Il n'existe pas de disposition légale imposant à un emprunteur de souscrire à une assurance emprunteur, ce qui rend cette dernière non obligatoire. Toutefois, certains établissements prêteurs peuvent la considérer indispensable pour bénéficier d'un crédit. Elle devient ainsi une condition d'octroi du prêt. C'est généralement le cas pour les prêts immobiliers où l'assurance emprunteur est exigée quasi-systématiquement.

Les contrats d'assurance emprunteurs peuvent comporter plusieurs garanties :

- **La garantie décès** : elle est toujours présente dans un contrat d'assurance emprunteur et en cas de décès, elle prévoit de verser à l'établissement prêteur le capital restant dû au jour du décès de l'emprunteur.
- **La garantie Perte Totale et Irréversible d'Autonomie (PTIA)** : une personne est en situation de PTIA lorsqu'elle répond aux trois conditions ci-dessous simultanément :
    - Se trouver dans l'impossibilité totale et définitive de se livrer à une quelconque activité rémunérée pouvant lui procurer gains ou profits ;
    - Être dans l'obligation absolue et présumée définitive d'avoir recours à l'assistance totale et constante d'une tierce personne pour effectuer 3 ou 4 actes ordinaires de la vie courante (se déplacer, se nourrir, faire sa toilette, s'habiller) ;
    - Ne pas avoir atteint l'âge limite prévu au contrat (en général 60 ou 65 ans ou bien l'âge de départ en retraite).

  En revanche, en pratique, la garantie se déclenche dès que l'assuré a droit à une pension d'invalidité de 3ème catégorie d'un régime obligatoire d'assurance maladie (sécurité sociale ou organismes assimilés) sans que celle-ci soit suffisante. Elle prévoit de verser à l'établissement prêteur le capital restant dû en cas de survenance de sinistre.
- **La garantie invalidité** : elle comprend une garantie invalidité permanente totale (IPT) et une garantie invalidité permanente partielle (IPP) qui prévoient de verser à l'organisme prêteur les échéances du prêt durant la période d'invalidité.

- **Invalidité permanente totale (IPT)** : c'est une situation d'inaptitude permanente et totale d'exercer une ou plusieurs activités professionnelles lui procurant gains ou profits en raison d'un handicap physique ou psychique résultant d'une maladie ou d'un accident. En pratique, la garantie est mise en jeu lorsque l'assuré bénéficie d'une pension d'invalidité de 2ème catégorie de la sécurité sociale. L'incapacité permanente totale est appréciée par une date de consolidation fixée au plus tard le 36ème mois d'incapacité temporaire totale continue et d'un taux d'invalidité (supérieur à 66%) déterminé en fonction du taux d'incapacité fonctionnelle et du taux d'incapacité professionnelle.
- **Invalidité Permanente Partielle (IPP)** : c'est une situation d'inaptitude permanente partielle de l'assuré à exercer une ou plusieurs activités professionnelles procurant gains ou profits, en raison d'un handicap physique ou psychique résultant d'une maladie ou d'un accident. En pratique, la garantie est mise en jeu dès lors que l'assuré bénéficie d'une pension d'invalidité de 1ère catégorie par la sécurité sociale. L'incapacité permanente partielle est appréciée par une date de consolidation fixée au plus tard le 36ème mois d'incapacité temporaire totale continue et d'un taux d'invalidité (supérieur à 33%) déterminé en fonction du taux d'incapacité fonctionnelle et du taux d'incapacité professionnelle.

- **La garantie incapacité** : elle comprend une garantie incapacité temporaire de travail totale (ITT) et une garantie incapacité temporaire partielle (ITP) qui prévoient de verser à l'organisme prêteur les échéances du prêt durant la période d'incapacité.
  - **Incapacité temporaire de travail (ITT)** : Il s'agit d'une garantie qui s'applique dans deux cas de figure distincts. Le premier, c'est lorsque l'assuré est en activité professionnelle ou au chômage (fait partie de la population active), il doit être dans une situation d'inaptitude temporaire et totale d'exercer son activité professionnelle lui procurant gains ou profits, en raison d'un handicap physique ou psychique résultant d'une maladie ou d'un accident. Quant au second, il concerne un assuré inactif avec une contrainte temporaire, suite à un accident ou une maladie, d'observer un repos complet et continu à son domicile, l'obligeant à interrompre toutes ses occupations habituelles.
  - **Incapacité temporaire partielle de travail (ITP)** : elle s'applique en cas d'inaptitude temporaire partielle de l'assuré à exercer son activité

professionnelle procurant gains ou profits, en raison d'un handicap physique ou psychique résultant d'une maladie ou d'un accident. En pratique, elle correspond à une reprise de travail à temps partiel.

- **La garantie perte emploi (PE)** : c'est la garantie pour les salariés en contrat de travail à durée indéterminée ayant fait l'objet d'un licenciement, ne disposant d'aucun autre contrat de travail à durée indéterminée en cours de validité et bénéficiant en outre des revenus de remplacement prévus aux articles L.5421-1 à L.5427-10 du Code du travail. Les autres allocations susceptibles d'être versées aux personnes privées d'emploi par le pôle emploi ou tout autre organisme n'ont pas la nature d'allocations d'assurance chômage. Comme pour les garanties invalidités et incapacités, la garantie PE verse les échéances du prêt à l'établissement prêteur durant la période de PE.

I.1.1.3 *Types de contrats*

Les contrats d'assurance emprunteur prennent la forme de contrats collectifs ou de contrats individuels. Cette distinction n'est pas d'origine juridique mais prend son origine de la nature du distributeur et de la segmentation tarifaire pratiquée.

- ✓ **Contrats « groupe »**

Le contrat groupe se définit comme une assurance souscrite par une personne au bénéfice de l'ensemble des membres d'un groupe. La logique de ce type de contrat est de garantir pour tous les emprunteurs le remboursement du crédit pendant toute la durée du prêt, quelle que soit leur évolution (vieillissement, dégradation de l'état de santé), et dans des conditions fixes tant en termes de tarif que de garanties, pour la majeure partie des contrats groupe.

Les tarifs des contrats groupe sont établis à partir de la mutualisation des risques de tous les assurés ; ces derniers payant donc des primes lissées en fonction de leurs profils. De ce fait, les assurés les plus jeunes payent plus que le risque qu'ils représentent, contrairement aux plus âgés. Les fumeurs bénéficient de conditions tarifaires identiques aux autres, de même que les non-cadres.

L'assureur, disposant d'une faible segmentation tarifaire et d'un volume important d'assurés, conserve une bonne mutualisation des risques. De plus, étant donné qu'il n'a pas de démarchage à effectuer, il bénéficie d'une réduction des coûts de gestion et cède en contrepartie des commissions à la banque.

En résumé, le principe du contrat groupe est de couvrir l'ensemble d'une population d'emprunteurs d'une même banque, de bénéficier d'une forte mutualisation. Les adhésions sont considérées dans leur ensemble et non de façon individuelle.

- ✓ **Contrats « individuel »**

Un contrat individuel est un accord entre deux parties, qui ne comprend que deux signataires : l'assuré et l'assureur. Il est destiné à des emprunteurs d'origines diverses, ce qui conduit l'assureur à proposer un tarif adapté au profil de risque de l'assuré, en fonction de différents critères comme son âge, son état de santé, sa catégorie socio-professionnelle, son rapport au tabagisme, sa durée de financement ou encore son capital emprunté.

Les contrats individuels sont construits pour garantir un prêt de façon annuelle, tout en sachant que les conditions tarifaires et de garanties sont susceptibles d'évoluer. Les cotisations d'assurance ne sont pas constantes sur toute la durée du prêt mais correspondent chaque année au risque que représente l'assuré, qui dépend notamment de son âge et de son capital restant dû assuré.

La diversité des emprunteurs couverts par le contrat individuel est à l'origine de la variabilité tarifaire observée. Les profils susceptibles de préférer les contrats individuels aux contrats groupe sont ceux qui présentent le moins de risques ; c'est le cas d'un jeune cadre non-fumeur, par exemple.

I.1.2 Garantie perte d'emploi

Puisque dans ce mémoire les travaux réalisés concernent la garantie perte d'emploi, nous allons d'abord approfondir davantage les caractéristiques de la garantie perte d'emploi. Ensuite, il sera question de mesurer l'importance de la garantie perte d'emploi à travers quelques statistiques. Enfin, nous reviendrons sur l'intérêt des tables d'expérience.

I.1.2.1 *Caractéristiques de la garantie perte d'emploi*

- ✓ **Carence**

La Carence représente la période comprise entre la signature du contrat et l'entrée en vigueur d'une garantie donnée. Durant cette période, l'emprunteur doit honorer ses mensualités (même en cas de perte d'emploi) et ne peut pas bénéficier d'indemnisation. Elle est fréquemment appliquée pour la garantie perte d'emploi, afin d'éviter notamment des adhésions au cours d'une période de préavis de licenciement. En fonction des contrats, la carence peut durer entre 3 et 18 mois.

- ✓ **Franchise**

La franchise est la durée minimale de la perte d'emploi pour que celui-ci ouvre droit à des prestations. Il faut faire la distinction entre franchise absolue et franchise relative. Dans la première, l'assureur ne commence à verser les prestations qu'à la fin de la franchise. Dans le cas de franchise relative, l'assureur verse les prestations correspondant à la durée totale du sinistre lorsque la durée de celui-ci dépasse la franchise. Par exemple, si la franchise est de 3 mois et la perte d'emploi a duré 4 mois, l'assureur verse 1 mois en franchise absolue et 4 mois dans le cas d'une franchise relative.

- ✓ **Limitation de la prise en charge**

Il y a deux types de limitation de prise en charge :

- Limitation du montant : en ADE, la prise en charge mensuelle de la perte d'emploi est plafonnée à la mensualité du prêt ;
- Limitation de la durée : le nombre de prise en charge pendant la perte d'emploi est plafonnée à un certain nombre de mois.

- ✓ **Limite d'âge de couverture**

On y trouve deux limites d'âge différentes :

- Limites d'âge à l'adhésion qui correspond à la tranche d'âges pour laquelle le candidat à l'assurance peut se voir accorder un contrat d'assurance ;
- Limites d'âges de cessation des garanties et des prestations correspondant à l'âge au-delà duquel, l'assuré ne bénéficie plus de la garantie et des prestations.

I.1.2.2 *Intérêt des tables d'expérience*

Une table (réglementaire ou d'expérience) a pour but de retracer la survie d'une population donnée dans un état spécifique au cours du temps. Autrement dit, il s'agit d'un tableau qui donne le nombre de survivants de la population dans l'état pour chaque unité de temps.

Afin de tarifer ou de calculer le provisionnement de certaines garanties, les compagnies d'assurance ont la possibilité d'utiliser les tables réglementaires du Bureau Commun des Assurances Collectives (BCAC). Ces tables fournissent aux assureurs des tables certifiées et reconnues par le régulateur. Toutefois, ces tables présentent beaucoup d'inconvénients.

D'abord, les tables sont construites à partir d'observations sur une population assez spécifique présentant des caractéristiques précises en termes de sur-représentativité de certaines tranches

d'âge, de répartition homme/femme, etc. Par exemple, la garantie a pu être proposée majoritairement à des personnes d'une certaine tranche d'âge. Les caractéristiques de cette population ne coïncident pas forcément avec la population d'assurés de l'assureur ce qui introduit un biais dans le calcul des provisions.

Ensuite, le niveau de sinistralité (durée des sinistres) observé sur la population de la table réglementaire peut s'écarter de celle de l'assureur du fait d'une mauvaise représentativité de sa population d'assurées. Cela peut conduire à des sur-provisionnements ou sous-provisionnements de l'assureur.

Enfin, la définition de la garantie utilisée dans la construction des tables du BCAC peut ne pas correspondre à la garantie proposée par l'assureur (par exemple documents à fournir pour attester de l'incapacité ou des garanties qui se poursuivent au-delà de l'âge limite des tables du BCAC : 60 ans).

Toutes ces raisons font que les tables du BCAC ne sont pas forcément adaptées aux garanties et à la population des assurés de l'assureur. De ce fait, ces tables ne sont pas directement utilisables pour le calcul des provisions ou des tarifs d'où l'intérêt de construire une table d'expérience à partir d'observations menées sur le portefeuille de l'assureur.

Par ailleurs, il n'existe pas de table du BCAC pour la garantie perte d'emploi ce qui rend crucial la construction de table d'expérience pour la tarification et surtout pour le calcul des provisions dans notre situation.

I.2 Présentation du produit Expresso

Le contrat d'assurance Expresso a pour objet de garantir toute personne physique, qu'elle soit emprunteur, co-emprunteur ou caution contre les risques liés au Décès, à la Perte Totale et Irréversible d'Autonomie (PTIA), à l'Invalidité Permanente Totale (IPT), à l'Incapacité Temporaire Totale (ITT) ou Partielle de travail (ITP), survenant à la suite d'une maladie ou d'un accident et à la Perte d'Emploi (PE) et survenant avant le terme d'un prêt à la consommation. Sur le portefeuille de la Société Générales Assurance, en 2020, le produit Expresso représente 19,1% des primes collectées en assurance emprunteur consommation et 9,8% en assurance emprunteur consommation et immobilier réunies.

Deux types de contrats, en fonction de l'âge des candidats à l'assurance, sont proposés. Les candidats à l'assurance âgés entre 18 et 60 ans à la date de demande d'adhésion peuvent adhérer à trois formules de garanties :

- Formule 1 : Décès, PTIA, IPT, ITT, ITP (DIT) et PE ;
- Formule 2 : Décès, PTIA, IPT, ITT, ITP.
- Formule 3 : PE

Les candidats de plus de 60 ans et de moins de 80 ans à la date de demande d'adhésion peuvent adhérer aux garanties Décès, PTIA, IPP, IPT, ITT.

Il faut noter qu'à partir de 65 ans, seule la garantie décès est couverte quelle que soit l'option retenue.

Puisque nous réalisons une étude sur la perte d'emploi, garantie présente uniquement dans le premier et le dernier contrat, nous nous focaliserons davantage sur ces derniers.

- **Crédits garantis**

Les crédits garantis par les présents contrats sont des crédits à la consommation comportant les principales caractéristiques suivantes :

- Amortissables avec ou sans différé, durée maximale de 120 mois,
- Capital emprunté maximal de 120 000 € pour les assurés de moins de 60 ans et de 30 000 € pour les assurés d'au moins 60 ans à l'adhésion.

En cours de crédit, les évolutions en termes de montant d'échéance, de diminution ou allongement de la durée sont admises à l'assurance sans pouvoir excéder la durée maximale.

- **Condition d'adhésion**

Toute personne physique, emprunteur, co-emprunteur ou caution peut bénéficier des garanties sous réserve de respecter les conditions d'âge à l'adhésion, d'avoir rempli une demande d'adhésion et s'être soumise aux formalités demandées par l'assureur.

Pour les garanties Décès, PTIA, IPT, ITT, ITP (DIT) :

Capital emprunté	Formalités médicales à réaliser
- Inférieur ou égal à 20 000 €	Aucune formalité médicale
- Supérieur à 20 000 € et inférieur ou égal à 50 000 €	Questionnaire de Santé Simplifié
- Supérieur à 50 000 €	Questionnaire de Santé

Tableau 1: Formalités médicales à réaliser en fonction du capital emprunté

Pour la garantie PE, en complément des formalités requises pour les garanties DIT, toute personne physique, emprunteur, co-emprunteur ou caution âgée de moins de 60 ans peut

demander à en bénéficier sous réserve de ne pas être en retraite ou pré-retraite à la date de demande d'adhésion.

Toute réticence ou fausse déclaration intentionnelle de la part de l'assuré entraîne la nullité de l'adhésion, dans les conditions de l'article L 113-8 du Code des assurances.

Après examen des formalités d'adhésion contractuelles et complémentaires d'adhésion, l'assureur peut : accepter l'adhésion aux conditions normales, accepter l'adhésion aux conditions spéciales moyennant une cotisation majorée et/ou une restriction de garantie, refuser ou ajourner l'adhésion.

- **Acquisition de droits**

Les droits de l'assuré sont calculés en fonction de sa durée d'activité en contrat de travail à durée indéterminée au cours de la période de référence.

Le début de la période de référence est :

- La date de prise d'effet des garanties si le crédit assuré n'a jamais donné lieu à indemnisation par l'assureur au titre de la garantie PE,
- Le lendemain du dernier jour indemnisé par l'assureur dans le cas contraire.

La fin de la période de référence est la date de fin de contrat de travail à durée indéterminée rompu par un licenciement.

L'assuré peut bénéficier de droits à indemnisation si, au cours de la période de référence, il justifie d'une durée d'activité en contrats de travail à durée indéterminée d'au moins 6 mois continus chez un ou plusieurs employeurs.

La durée maximale d'indemnisation est calculée comme suit :

Durée d'activité en contrats de travail à durée indéterminée cours de la période de au référence	Durée maximale d'indemnisation au cours de la période de référence
moins de 6 mois	Pas de droits
de 6 mois à moins de 12 mois	180 jours
supérieur à 12 mois	360 jours

Tableau 2: Durée maximale d'indemnisation en fonction de la durée de la période de référence

Il convient de noter que le produit Expresso ne détient pas de franchise mais détient plutôt une carence de 6 mois.

- **Reprise d'activité professionnelle suivie d'une nouvelle Perte d'Emploi**

En cas de reprise d'activité et de nouvelle période de chômage suite à un licenciement, l'assureur verse :

- Le reliquat des droits acquis au moment du licenciement ayant donné lieu à l'indemnisation précédente, dans la mesure où la durée de la reprise d'activité n'ouvre aucun nouveau droit ou si la nouvelle Perte d'Emploi n'est pas garantie,
- Le nombre d'indemnités le plus favorable entre le reliquat et la nouvelle durée maximale acquise si la Perte d'Emploi est garantie et si la reprise d'activité a ouvert de nouveaux droits.

La nouvelle durée maximale d'indemnisation annule le reliquat des droits acquis au moment du licenciement ayant donné lieu à l'indemnisation précédente.

- **Montant des prestations**

L'assureur verse 100% du montant des mensualités du Crédit consenti par le Prêteur, venant à échéance à compter du premier jour indemnisé au titre du revenu de remplacement

- **Exclusions relatives au risque perte emploi**

- La retraite ou la pré-retraite, quelle qu'en soit la cause, y compris pour inaptitude au travail ;
- La rupture conventionnelle du contrat de travail à durée indéterminée ;
- La démission, même prise en charge par le Pôle Emploi ou organismes assimilés ;
- Toute cessation d'activité dont la réglementation implique la non-recherche d'un nouvel emploi ;
- Le licenciement pour faute grave ou lourde ;
- Le licenciement si vous êtes salarié :
    - de votre conjoint, d'un de vos ascendants, collatéraux ou descendants
    - d'une personne morale emprunteuse contrôlée ou dirigée par votre conjoint, l'un de vos ascendants, collatéraux ou descendants, sauf si ce licenciement est concomitant à la liquidation judiciaire de l'entreprise.

Partie II. Construction de la base de données et statistiques descriptives

II.1 Construction et fiabilisation de la base de données

Puisque l'objectif de cette étude est de construire une table de maintien en perte emploi, nous avons eu recours à la base de données des sinistres du produit en question. Ainsi, la base de données initiale à notre disposition regroupait les sinistres de perte d'emploi d'autres produits en plus de ceux du produit Expresso. Par ailleurs, chaque ligne de la base correspond à une prestation de l'assureur au profit de l'assuré. Autrement dit il y a autant de lignes qu'il y a de prestations concernant un assuré, un prêt et un sinistre donné. De même, certaines variables pouvant permettre une meilleure segmentation des sous-populations telles que l'ancienneté avant la perte d'emploi ne sont pas présentes dans la base initiale et seront donc à créer.

Le paragraphe précédent a montré que la base initiale ne peut être utilisée directement pour la construction de table de maintien. Par conséquent, le but de cette partie sera d'expliquer les différentes étapes ayant permis à l'aboutissement de la base de données finale utilisée dans l'étude. La première sous-partie traitera du passage de la base à plusieurs lignes par sinistre à la base agrégée avec une seule ligne par sinistre. Ensuite, la deuxième mettra en évidence le processus de création de nouvelles variables telles que l'âge de l'assuré, l'ancienneté avant la perte d'emploi, etc. Enfin, nous exposerons dans la dernière sous-partie le processus de création de la variable indicatrice de censure des sinistres.

II.1.1 Identification des sinistres et agrégation des lignes d'un même sinistre

La base initiale est au format « *.sas7bdat* », extension exploitable uniquement sous SAS, et tous les traitements décrits dans cette sous-partie sont effectuées sur SAS. En appliquant un filtre permettant de se limiter uniquement au produit Expresso nous obtenons une base de données avec 17 378 lignes. Comme nous l'avons dit précédemment, cela ne signifie pas qu'il y a autant de sinistres puisqu'il y a au moins autant de lignes que d'indemnisations concernant le sinistre d'un assuré sur un prêt donné. En effet, il peut y avoir plus de lignes que de prestations pour un sinistre car assez souvent, l'ouverture du dossier de sinistre engendre la création d'une première ligne même en l'absence de versement d'une quelconque prestation à l'assuré.

Par ailleurs, la base de données comporte 32 variables :

- ID_INDIVIDU : Identifiant de l'individu

- ID_Sexe : Sexe de l'assuré
- DTE_NAISSANCE : Date de naissance de l'assuré
- CODE_POSTAL : Code postale de la ville de résidence de l'assuré
- PAYS : Pays de l'assuré
- Id_CSP : Code catégorie socio-professionnelle de l'assuré
- ID_TYPE_RISQUE : Identifiant du type de risque couvert (Décès, Perte d'emploi, etc.)
- FAIT_GENERATEUR : Fait générateur
- DTE_SURVENANCE : Date de survenance de la perte d'emploi
- DTE_DECLARATION : Date de déclaration de la perte d'emploi auprès de l'assureur
- DTE_CREATION : Date de création du dossier sinistre
- COD_AGENCE_DECLARATION :
- ID_PRET : Identifiant du prêt
- DTE_SIGNATURE : Date de signature du contrat par l'assuré
- NB_DUREE_INITIALE : Durée initial du prêt (en mois)
- MNT_NOMINAL_INITIAL : Montant nominal initial du prêt
- MNT_AUTORISE :
- MNT_RESTANT_DU : Montant restant dû
- NB_ECHEANCES_RESTANTE : Nombre d'échéances restantes du prêt
- MNT_ECHEANCE : Montant des échéances du prêt
- ID_SS_PRD_ASS : Code police du produit
- ID_STATUT_PRET : Identifiant du statut du prêt
- DT_STATUT_PRET : Date de statut du prêt
- ID_STATUT_SINISTRE : Identifiant du sinistre
- DTE_STATUT_SINISTRE : Date statut du sinistre
- TX_PEC : Taux de prise en charge de l'échéance (100%
- DTE_DEBUT_VALIDITE : Date de début de la période couverte par la prestation
- DTE_EMISSION_PRESTATION : Date de paiement de la prestation
- MNT_INDEMNISE_Sum : Somme ou cumul des montants indemnisés depuis le début de la perte d'emploi
- NB_ECHEANCE_Sum : Somme ou nombre cumulé des échéances depuis le début de la perte d'emploi

Comme on peut le remarquer, il n'y a pas de variable permettant d'identifier correctement un sinistre dans la base initiale. Ainsi, il est nécessaire de la créer avant de pouvoir agréger les lignes d'un même sinistre.

II.1.1.1 *Identification des sinistres*

L'absence d'identifiant nous pousse donc à utiliser plusieurs variables pour remplir ce rôle. Dans notre cas, nous avons utilisé trois variables : l'identifiant de l'individu, l'identifiant du prêt et le numéro de sinistre (**ID_INDIVIDU, ID_PRET et num_sin2**). Les deux premières variables étaient déjà présentes dans la base initiale mais la dernière a été calculée en utilisant la date de survenance des sinistres ainsi que la durée entre les sinistres pour identifier chaque sinistre.

L'objet de cette section est d'expliquer en détail chacune des 3 étapes de la création de la variable numéro de sinistre. La première étape est relative au traitement des valeurs manquantes de la date de survenance des sinistres alors que la deuxième s'attelle à calculer la durée entre les sinistres et le nombre de prises en charge par sinistre. Ceci permettra, lors de la dernière étape, d'identifier avec plus de certitude les sinistres.

- **Traitement de la date de survenance**

La date de survenance des sinistres offre un premier niveau d'identification des sinistres mais le problème est qu'il y a certaines lignes pour lesquelles elle n'est pas renseignée (valeurs manquantes). En effet, dans la base initiale, un sinistre est représenté sur plusieurs lignes et généralement, pour certaines lignes (les premières le plus souvent) la date de survenance est bien mentionnée mais ne l'est pas pour d'autres lignes (voir figure 1[2]). Il est donc crucial de pouvoir indiquer si ce sont des sinistres différents ou bien la continuation du sinistre initial.

ID_INDIVIDU	ID_PRET	DTE_SURVENANCE	DTE_DECLARATION	DTE_CREATION
000193738	38195270756	.	.	20191028
000193738	38195270756	.	.	20191028
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	20150910	20150921	20150921
000319292	35196636522	.	.	20150921
000346600	35197010503	.	.	20170503
000346600	35197010503	.	.	20170503
000360972	38195359187	.	.	20200331
000360972	38195359187	.	.	20200331
000376672	36198890687	20180101	20180228	20180228
000376672	36198890687	20180101	20180228	20180228
000376672	36198890687	20180101	20180228	20180228
000376672	36198890687	20180101	20180228	20180228
000376672	36198890687	20180101	20180228	20180228
000376672	36198890687	.	.	20180228

Figure 1: Illustration des cas de *valeurs manquantes sur la variable date de survenance*

Le procédé utilisé pour résoudre ce problème se base sur la variable date de création qui est presque toujours renseignée dans la base de données. On pourrait se demander pourquoi ne pas utiliser la variable date de création directement à la place de la date de survenance. Il s'agit

[2] Les illustrations utilisées dans cette partie montrent des données anonymisées.

d'une interrogation tout à fait logique mais le problème est que la variable date de création peut changer plusieurs fois pour un même sinistre. En effet, lorsqu'un sinistre change de statut d'un mois à l'autre (par exemple de statut D : début prise en charge à C : prise en charge en cours), la date de création peut changer (voir figure 2). Ainsi, on ne peut pas utiliser la date de création comme identifiant mais on peut s'en servir pour déterminer le bon identifiant.

ID_INDIVIDU	ID_PRET	DTE_SURVENANCE	DTE_DECLARATION	DTE_CREATION	ID_STATUT_SINISTRE
045726247	34199015115	20140227	20150403	20150403	D
045726247	34199015115	.	.	20150403	J
045726247	34199015115	20140227	20150403	20150821	D
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	C
045726247	34199015115	20140227	20150403	20150921	F
045726247	34199015115	.	.	20150921	J
045726247	34199015115	20140227	20150403	20160705	D
045726247	34199015115	20140227	20150403	20160705	C
045726247	34199015115	20140227	20150403	20160705	F
045726247	34199015115	.	.	20160705	J

Figure 2 : Cas où la date de création change lorsque le statut du sinistre change

Pour imputer la date de survenance là où ce n'est pas renseigné, l'algorithme fixe une date de création et vérifie s'il existe une autre ligne avec la même date de création et une date de survenance renseignée. S'il en existe, la date de survenance manquante sera remplacée par celle trouvée lors de la recherche. En revanche, dans les rares cas où on n'en trouve pas, on a préféré les laisser comme valeur manquante dans la mesure où cela signifie généralement qu'il n'y a pas eu de paiement et que la déclaration du sinistre ne s'est pas bien faite. Il faut noter qu'on pouvait aussi considérer la date de création comme date de survenance.

ID_INDIVIDU	ID_PRET	DTE_SURVENANCE	DTE_DECLARATION	DTE_CREATION	DATE_SURV_SIN	DATE_CREATION_SIN
000193738	38195270756	.	.	20191028	.	28/10/2019
000193738	38195270756	.	.	20191028	.	29/10/2019
000319292	35196636522	20150910	20150921	20150921	10/09/2015	21/09/2015
000319292	35196636522	20150910	20150921	20150921	11/09/2015	22/09/2015
000319292	35196636522	20150910	20150921	20150921	12/09/2015	23/09/2015
000319292	35196636522	20150910	20150921	20150921	13/09/2015	24/09/2015
000319292	35196636522	20150910	20150921	20150921	14/09/2015	25/09/2015
000319292	35196636522	20150910	20150921	20150921	15/09/2015	26/09/2015
000319292	35196636522	20150910	20150921	20150921	16/09/2015	27/09/2015
000319292	35196636522	20150910	20150921	20150921	17/09/2015	28/09/2015
000319292	35196636522	20150910	20150921	20150921	18/09/2015	29/09/2015
000319292	35196636522	20150910	20150921	20150921	19/09/2015	30/09/2015
000319292	35196636522	20150910	20150921	20150921	20/09/2015	01/10/2015
000319292	35196636522	20150910	20150921	20150921	21/09/2015	02/10/2015
000319292	35196636522	.	.	20150921	22/09/2015	03/10/2015
000346600	35197010503	.	.	20170503	.	03/05/2017
000346600	35197010503	.	.	20170503	.	04/05/2017
000360972	38195359187	.	.	20200331	.	31/03/2020
000360972	38195359187	.	.	20200331	.	01/04/2020
000376672	36198890687	20180101	20180228	20180228	01/01/2018	28/02/2018
000376672	36198890687	20180101	20180228	20180228	01/01/2018	01/03/2018
000376672	36198890687	20180101	20180228	20180228	01/01/2018	02/03/2018
000376672	36198890687	20180101	20180228	20180228	01/01/2018	03/03/2018
000376672	36198890687	20180101	20180228	20180228	01/01/2018	04/03/2018
000376672	36198890687	.	.	20180228	01/01/2018	05/03/2018

Figure 3: Après création de la nouvelle variable « date de survenance »

- **Calcul du nombre de prises en charge et création d'identifiant par sinistre**

L'approche générale pour le calcul du nombre de prises en charge est que pour chaque individu et un prêt donné, on fait le cumul du nombre de prises en charge en ignorant le fait que cela soit le même sinistre ou pas. Autrement dit, c'est comme si chaque ligne où il y a une indemnisation effective était un sinistre et on ordonne les lignes par date de paiement (date de fin de sinistre dans la base) ce qui permet de faire le cumul des prises en charge. Ainsi, on donne un rang à chaque ligne en se basant sur le nombre de prises en charge effectuées pour un individu et un prêt donné (voir figure 4).

Pour la création d'identifiant par sinistre, on considère toujours que chaque observation représente un sinistre jusqu'à la preuve du contraire. L'idée générale est qu'un assuré (pour un prêt donné) ne peut pas avoir un nouveau sinistre alors qu'il y en a déjà un en cours. En effet, l'algorithme de création d'identifiant consiste à affecter la valeur 1 à la première observation lorsque la date de survenance est renseignée et 0 sinon. Pour une observation hors première ligne, on considérera qu'elle est issue du même sinistre que celle qui la précède

si le nombre de mois[3] de la ligne précédente + le cumul de prises en charge de la ligne précédente + 6 > le nombre de mois de la ligne courante

[3] $nombre\ de\ mois = numéro\ du\ mois\ de\ la\ date\ de\ survance + (année\ date\ de\ survenance \times 12)$

Cette condition signifie que si la date de survenance de la ligne actuelle est antérieure à la date de survenance de la ligne précédente ajoutée du nombre de mois de prise en charge alors les deux sinistres sont identiques même si les dates de survenance sont différentes. Le **6** dans la condition n'est qu'une marge que nous avons ajoutée puisque lors de la prise en charge d'un sinistre, il est possible qu'il y ait des ruptures qui sont soit liées à une absence de justificatif soit à d'autres raisons que l'on ignore. Ainsi, le décalage induit par les ruptures au cours de la prise en charge est compensé avec cette marge.

ID_INDIVIDU	ID_PRET	DTE_SURVENANCE	DTE_DECLAI	DTE_CREATI	DATE_SURV_SIN	DATE_CREATION_SIN	cumul_PEC2	num_sin2
071766298	35197461367	20160204	20160205	20151103	04/02/2016	03/11/2015	1	1
071766298	35197461367	20160204	20160205	20151103	04/02/2016	03/11/2015	1	1
071766298	35197461367	.	.	20151103	04/02/2016	03/11/2015	1	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	3	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	4	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	5	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	6	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	7	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	8	1
071766298	35197461367	20160204	20160205	20160905	04/02/2016	05/09/2016	9	1
071766298	35197461367	20170701	20170702	20160905	01/07/2017	05/09/2016	10	2
071766298	35197461367	20170701	20170702	20160905	02/07/2017	05/09/2016	11	2
071766298	35197461367	20170701	20170702	20160905	03/07/2017	05/09/2016	11	2
071766298	35197461367	20180301	20180302	20180411	01/03/2018	11/04/2018	12	2
071768005	35199825056	.	.	20160725	.	25/07/2016	0	0
071768005	35199825056	.	.	20160725	.	25/07/2016	0	0
071768005	35199825056	20160707	20160725	20161202	07/07/2016	02/12/2016	1	1
071768005	35199825056	20160707	20160725	20161202	07/07/2016	02/12/2016	2	1
071768005	35199825056	.	.	20161202	07/07/2016	02/12/2016	2	1
071768005	35199825056	.	.	20161202	07/07/2016	02/12/2016	2	1
071768005	35199825056	20161130	20161202	.	07/07/2016	02/12/2016	5	1
071768005	35199825056	20161130	20161202	.	07/07/2016	02/12/2016	7	1
071768005	35199825056	20161130	20161202	.	07/07/2016	02/12/2016	8	1
071768005	35199825056	20161130	20161202	20161202	07/07/2016	02/12/2016	10	1

Figure 4: Illustration création des variables cumul PEC et d'identifiant des sinistres

II.1.1.2 *Agrégation des lignes d'un même sinistre*

L'étape d'agrégation permet de regrouper et de synthétiser les informations sur les sinistres par individu et par prêt. Au lieu d'avoir plusieurs lignes pour un sinistre concernant le prêt d'un individu donné, il s'agira donc d'agréger les informations de plusieurs observations en une seule ligne. Ainsi, on passe de la base de la figure 4 à la base de la figure 5 en utilisant les identifiants des individus, des prêts et ceux des sinistres comme variables d'agrégation.

ID_INDIVIDU	ID_PRET	DTE_SURVENANCE	DTE_DECLARATION	DTE_CREATION	DATE_SURV_SIN	DATE_FIN_SIN	DATE_CREATION_SIN	cumul_PEC2	num_sin2
071766298	35197461367	20160204	20160205	20160905	04/02/2016	10/02/2017	05/09/2016	9	1
071766298	35197461367	20180301	20180302	20180411	01/03/2018	10/03/2018	11/04/2018	12	2
071768005	35199825056	20161130	20161202	20161202	07/07/2016	30/05/2017	02/12/2016	10	1

Figure 5: Base agrégée par individu, par prêt et par sinistre

Comme constaté dans la figure 5, la base agrégée permet d'avoir de façon agrégée le nombre d'échéances (en d'autres termes la durée du sinistre) et le montant total de la somme indemnisée

en plus des variables relatives à l'assuré (le sexe, la date de naissance, etc.) et à son prêt (montant prêt, montant échéance, etc.).

Cette base agrégée contenant toutes les informations sur les sinistres va servir à la création de la loi de maintien. Elle comporte 1 379 lignes, concernant les sinistres sur la période du 01/01/2012 au 31/12/2018. Toutefois, avant que la base ne soit entièrement opérationnelle, il est primordial de définir la variable de censure sur les sinistres puisque les durées de sinistres observées résultent d'observations incomplètes.

II.1.2 Censure des observations

De façon générale, la collecte de variables de durée est plus contraignante et plus sujette à des erreurs de mesure que les autres types de variable. Il est encore d'autant plus complexe de les observer sur une population d'assurés. En effet, plusieurs phénomènes comme les termes du contrat, résiliation du contrat, décès, etc. peuvent empêcher l'observation complète des durées auprès des assurés. Ignorer ces phénomènes de censure de la variable durée de chômage conduirait à une sous-estimation des probabilités de maintien au chômage. Avant de montrer comment le phénomène de censure est pris en compte dans notre cas, nous allons présenter la notion de censure de façon plus générale.

II.1.2.1 *Notion de censure*

Contrairement à la plupart des variables mesurables instantanément, comme l'âge, la taille, le revenu, les variables de durée se collectent dans le temps. Bon nombre d'événements peuvent se produire et rendre incomplète la collecte de telles variables. Ces événements peuvent se présenter sous la forme de censure ou de troncature. Ces phénomènes sont plus faciles à appréhender à travers des exemples.

Exemple

On cherche à recueillir T, le nombre de mois passés en perte d'emploi avant de retrouver un nouvel emploi auprès des assurés du portefeuille entre le 01/01/2013 et 31/12/2019. Soit C la variable de censure ($C > 0$).

Dans la figure 6, l'assuré est entré et est sortie de la perte d'emploi avant la date de début d'observation (01/01/2013), on parle de **censure à gauche**. Dans ce cas, nous savons juste que la perte d'emploi a eu lieu mais nous ignorons sa durée.

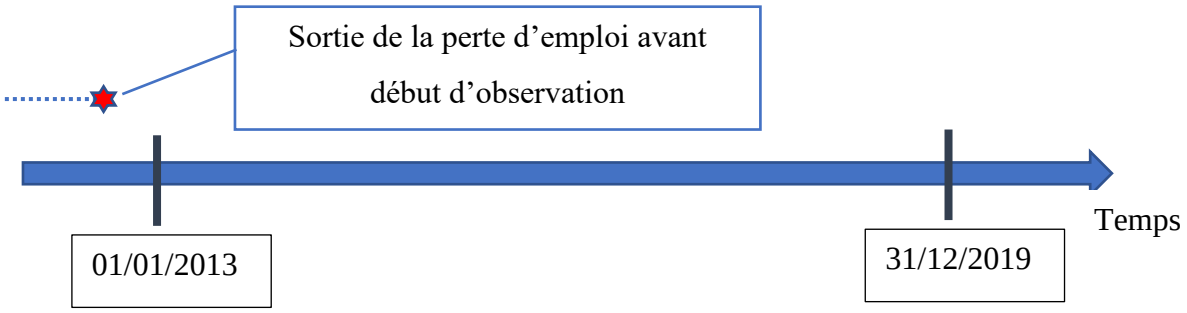


Figure 6: Illustration censure à gauche

La figure 7 illustre le cas où l'assuré est entré en perte d'emploi avant le début de la période d'observation mais est sorti de la perte d'emploi avant la fin de la période d'observation. Cette situation ne présente de censure car l'information peut-être totalement reconstituée.

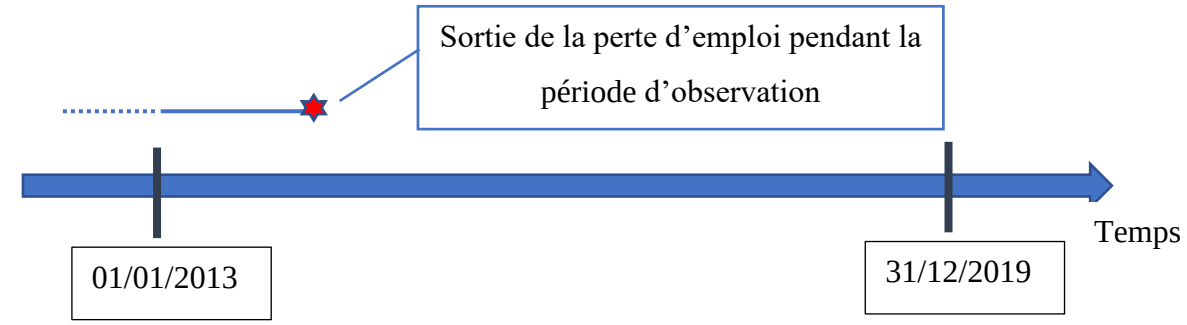


Figure 7: Illustration d'une absence de censure 1

Les situations dans lesquelles la date d'entrée et de sortie se trouvent dans la période d'observation (cf. figure 8), l'observation est complète et donc il n'y a pas de censure.

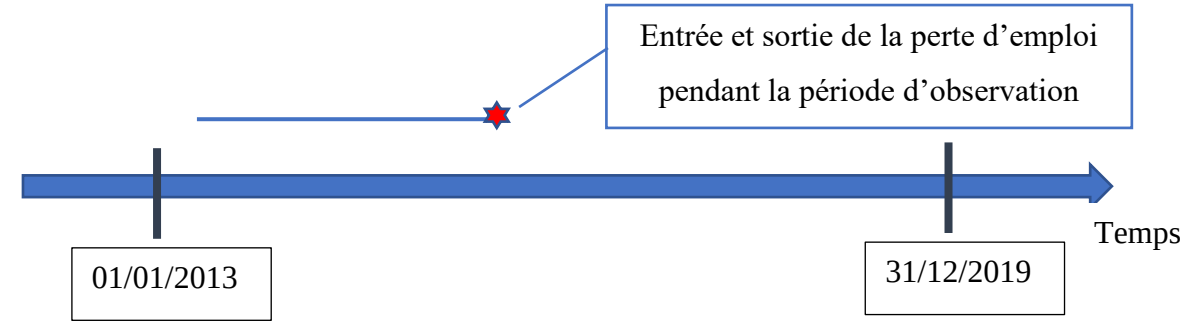


Figure 8: Illustration d'une absence de censure 2

Les deux figures ci-dessous illustrent le phénomène de **censure à droite**. On parle de censure à droite lorsque $C < T$. Autrement dit, lorsqu'on n'observe pas la date de sortie de la perte d'emploi soit parce que la sortie se réalise après la date de fin d'observation, soit parce que l'assuré ne fait plus partie de la population suivie.

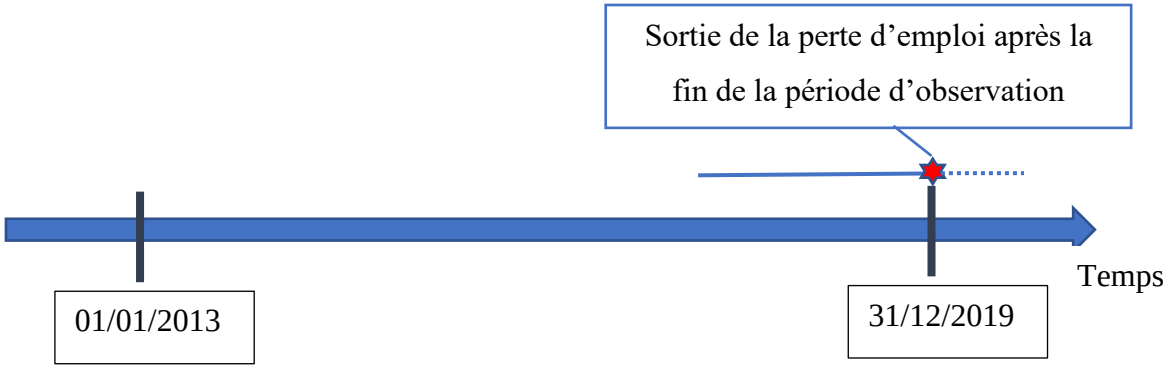


Figure 9: Illustration d'une censure à droite 1

Plusieurs raisons peuvent conduire à une interruption du suivi de la durée de perte d'emploi dans le cadre d'un contrat d'assurance. Les plus fréquentes pour les garanties perte d'emploi sont l'atteinte du nombre maximum de prestations dont peut bénéficier l'assuré. Par exemple, si un assuré passe plus de 12 mois en perte d'emploi dans le cadre de la garantie fournie par le produit Expresso, nous ne pouvons observer que les 12 mois puisque le nombre de prises en charge se limite à 12 échéances maximum. D'autres motifs tels que le décès ou le passage à une situation d'incapacité, d'invalidité ou de décès peuvent être à l'origine de la censure.

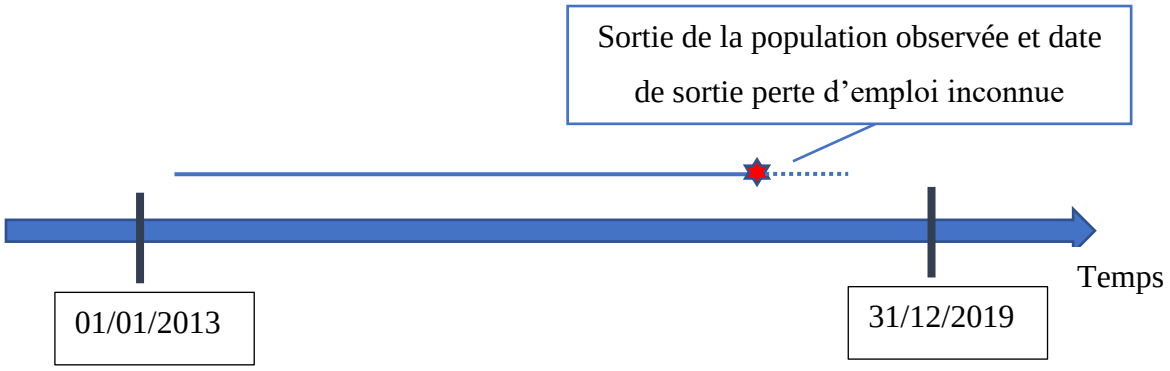


Figure 10: Illustration d'une censure à droite 2

Dans le cadre de notre étude, nous serons davantage concernés par les censures à droite. Pour le définir formellement, nous allons introduire quelques notations :

$(T_1, \dots, T_n)$ un échantillon de durée de vie

$(C_1, \dots, C_n)$ un échantillon de censure strictement positif

On dit qu'il y a censure à droite lorsqu'on observe l'échantillon $((Y_1, D_1), \dots, (Y_n, D_n))$ au lieu de l'échantillon $(T_1, \dots, T_n)$

Avec $Y_i = \min(T_i, C_i)$ et $D_i = \begin{cases} 1 \; si \; T_i \leq C_i \\ 0 \; sinon \end{cases}$

D_i représente l'indicatrice de censure.

II.1.2.2 *Prise en compte de la censure dans la base de données*

La prise en compte des phénomènes de censure se fait à travers une variable indicatrice qui vaut 1 s'il y a censure et 0 sinon. Vu la nature du contrat du produit Expresso PE, il y a beaucoup de sources de censure. Pour renseigner la variable de censure par sinistre, on se base sur un ensemble de conditions généralement mentionnées dans le contrat d'adhésion. Ainsi, nous allons d'abord revenir sur les variables qu'il a fallu créer pour vérifier ces conditions.

II.1.2.2.1 Calcul de nouvelles variables

- **Durée de prise en charge**

Par défaut, la durée de prise en charge est donnée par le rapport entre le montant total indemnisé et le montant d'une indemnisation (ou échéance). Dans certains cas, l'une des variables mentionnées précédemment n'est pas disponible, alors la durée de prise en charge est égale à la durée depuis la déclaration du sinistre. Si la date de déclaration n'est pas renseignée alors on utilise la durée depuis la date de survenance.

- **Durée entre deux sinistres**

Pour les individus avec plusieurs sinistres pour un prêt donné, la durée entre deux sinistres est la différence entre les deux dates de survenance. Par exemple, pour un individu avec deux sinistres, la variable durée entre deux sinistres vaut zéro pour le premier sinistre et vaut la différence des dates de survenance pour le deuxième sinistre.

- **Durée entre date de survenance d'un sinistre et date d'adhésion ou date précédent sinistre**

Lorsqu'un individu n'a qu'un seul sinistre alors, seule la durée écoulée entre la date d'adhésion et la date de survenance est calculée. En revanche, quand un assuré a plusieurs sinistres, alors pour le premier sinistre, c'est la durée écoulée entre la date d'adhésion et la date de survenance qui est calculée tandis que pour les autres sinistres on calculera plutôt la durée par rapport au précédent sinistre.

- **Ancienneté**

Il s'agit de la durée écoulée en mois entre l'adhésion et la date de survenance du sinistre.

- **Age lors de la survenance du sinistre**

Cette variable correspond à la partie entière de l'âge exacte de l'assuré à la date de survenance de la perte d'emploi. Dans la base, il s'agit de la différence entre la date de naissance et la date de survenance du sinistre.

II.1.2.2.2 Application des conditions de censure

- **Conditions de fin d'observation**

Puisque la base utilisée contient des observations qui s'arrêtent en décembre 2018 et que le nombre maximum de prise en charge est 12 mois (donc la durée maximum d'observation d'un sinistre est de 12 mois), tous les sinistres en cours dont la date de début d'indemnisation est antérieure à décembre 2017 sont censurés.

- **Condition liée au montant restant dû**

Les sinistres pour lesquels le montant restant dû est nul sont également automatiquement censurés.

- **Condition liée à l'âge de l'assuré**

Dans le contrat d'adhésion d'Expresso PE, plus précisément dans les conditions de cessation de garanties et de prestation, il est stipulé que la garantie cesse au plus tard au 65ème anniversaire de l'assuré. Par conséquent, si la variable « âge lors de la dernière indemnisation » est supérieure à 65 alors il y a censure.

- **Conditions sur les assurés avec 6 échéances**

L'un des termes dans les conditions d'acquisition de droits d'un assuré est qu'il doit justifier d'une durée d'activité continue d'au moins 6 mois pour bénéficier de 180 jours (6 mois) d'indemnisation. Ainsi, ces conditions sont implémentées comme suit :

Conditions	Censure (Oui / Non)
Si nombre échéances ≤ 6 et **Durée entre date de survenance d'un sinistre et date d'adhésion ou date précédent sinistre** $\in [\mathbf{6-12}[$ et **sinistre Fermé**	Oui
Si nombre échéance $= 6$ et **Durée entre date de survenance d'un sinistre et date d'adhésion ou date précédent sinistre** $\in [\mathbf{6-12}[$	Oui
Si nombre échéance $= 6$ et **Durée prise en charge** $\in [\mathbf{5,5-6,5}[$	Oui

- **Conditions sur les assurés avec 12 échéances**

Ces conditions sont similaires à celles précédentes mais ici il faut au moins 12 mois d'activité pour bénéficier de 360 (12 mois) jours d'indemnisation. Nous résumons l'implémentation des conditions dans le tableau ci-dessous :

Conditions	Censure (Oui / Non)
Si nombre échéance $= 12$ et **Durée entre date de survenance d'un sinistre et date d'adhésion ou date précédent sinistre ≥ 12**	Oui
Si nombre échéance $= 12$ et **sinistre Fermé**	Oui
Si nombre échéance $= 12$ et **Durée prise en charge $\geq 11,5$**	Oui

- **Condition pour individu avec plusieurs sinistres**

Pour un assuré qui a par exemple deux sinistres, nous vérifions au niveau du deuxième sinistre si la durée par rapport au sinistre précédent est inférieure à 12 mois et si la somme du nombre de prises en charge pour ces deux sinistres est égale à 12 alors on censure le deuxième sinistre.

En résumé, l'ensemble de ces traitements ont permis d'obtenir la base finale constituée de 1.379 lignes et de 15 variables. Parmi ces variables, les plus importantes sont la durée avant la sortie de perte d'emploi (ou nombre de prise en charge des échéances de l'assuré par l'assureur) en mois, l'indicatrice de censure et les variables explicatives ou covariables (sexe, âge, ancienneté, montant initial du prêt et montant de l'échéance du prêt).

II.2 Statistiques descriptives sur le portefeuille

Cette section présente les statistiques descriptives sur le portefeuille afin de mieux comprendre la population sur laquelle il sera question de construire une table de maintien.

II.2.1 Statistiques sur les sinistres

- **Durée en perte d'emploi**

L'analyse de la figure 11 montre que la majeure partie des sorties de la perte emploi se fait au bout de 6 ou 12 mois. En effet, 45% des sorties observées se sont produites au 12ème mois et presque 25% au 6ème mois. La proportion de sorties hors 6ème et 12ème mois reste inférieure à 5%. Il est important de noter que cette proportion des sorties au 6ème et 12ème pourrait s'expliquer par les termes du contrat qui limitent le nombre de prestations à 12 mois si l'assurée a cotisé pendant au moins 360 jours et à 6 mois si l'assuré n'a cotisé qu'entre 180 et 360 jours. Ainsi, beaucoup de ces sorties sont en réalité censurées à droite puisqu'on ne peut plus les suivre dans la base de données des sinistres.

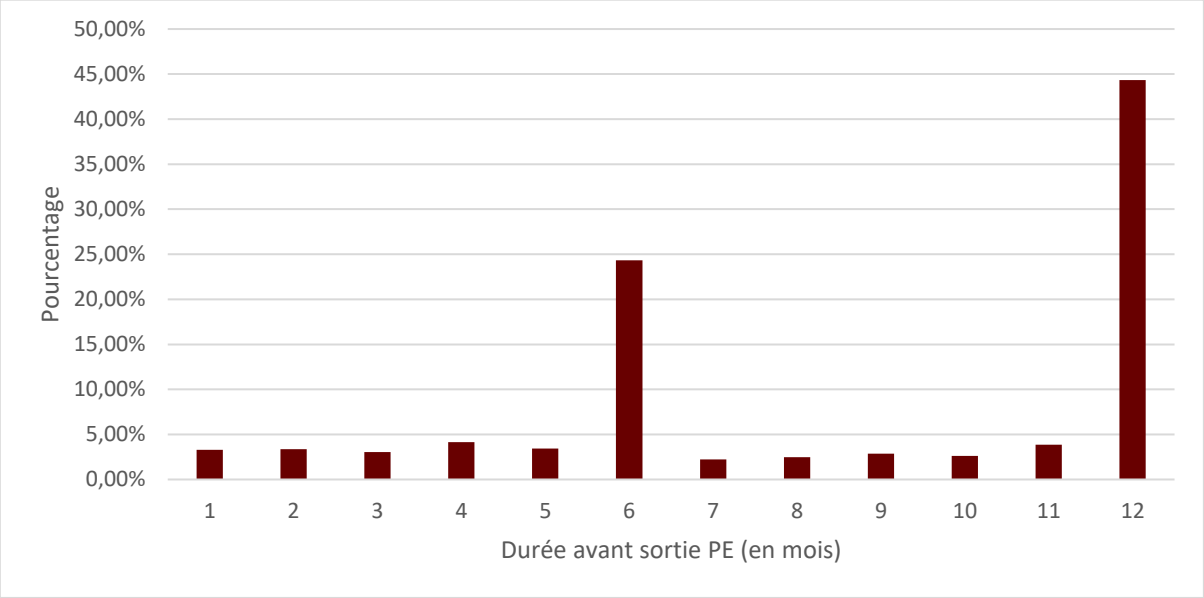


Figure 11: Répartition des sinistres en fonction de la durée passée en perte d'emploi

- **Mois de sortie**

En termes de saisonnalité, les sorties sont réparties quasi uniformément suivant les 12 mois de l'année. A l'exception des mois de mars, septembre et novembre qui enregistrent chacun plus de 9% des sorties, la proportion de sorties survenues pendant les autres mois est comprise entre 7% et 9%.

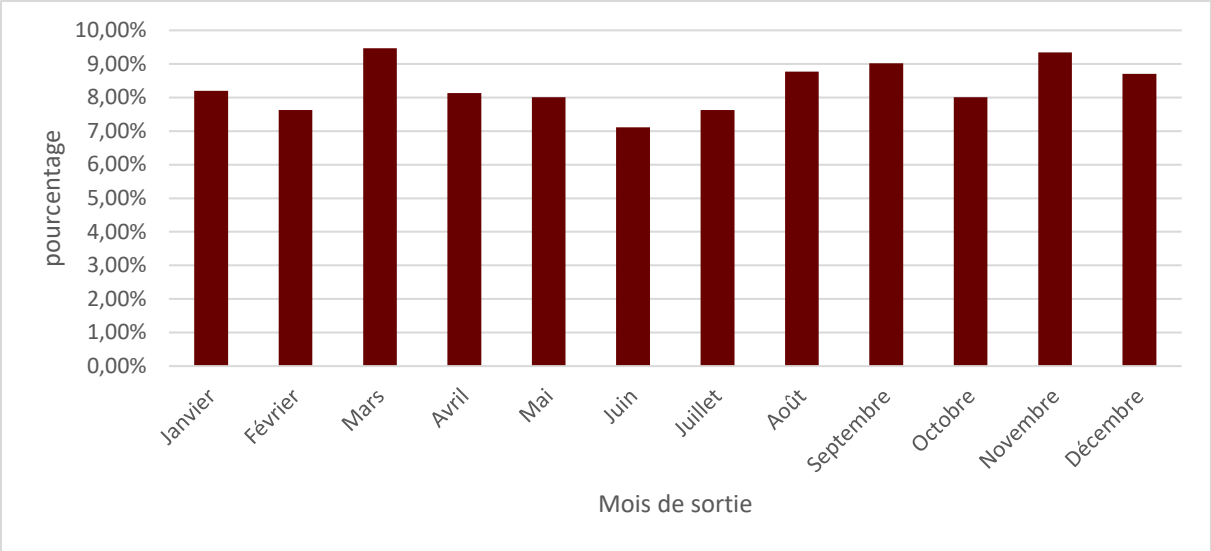


Figure 12: Répartition des sinistres en fonction du mois de sortie de la perte d'emploi

- **Année de sortie**

Contrairement aux mois, les sorties ne sont pas réparties uniformément. En effet, c'est en 2016 (resp. 2012) que le plus grand (resp. petit) nombre de sorties est enregistré avec 21,9% (resp. 0,01). Il convient de noter que depuis 2016, la proportion des sorties diminue alors qu'on observe le contraire de 2012 à 2016.

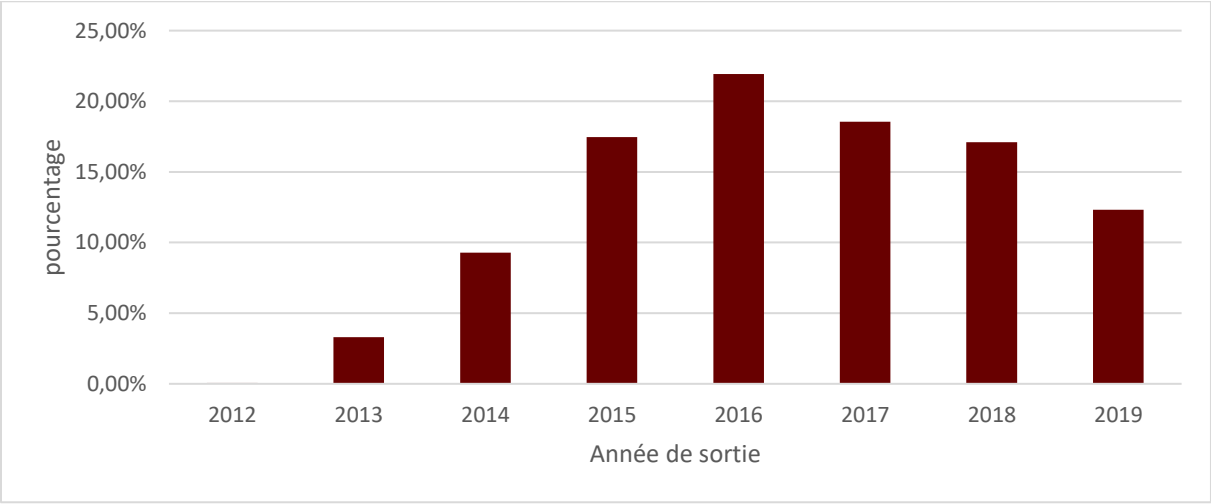


Figure 13: Répartition des sinistres en fonction de l'année de sortie de la perte d'emploi

II.2.2 Statistiques sur les assurés

- **Sexe**

La figure 14 montre la répartition des assurés sinistrés suivant le sexe et on constate qu'il y a plus d'hommes, avec 58,45% des assurés, que de femmes.

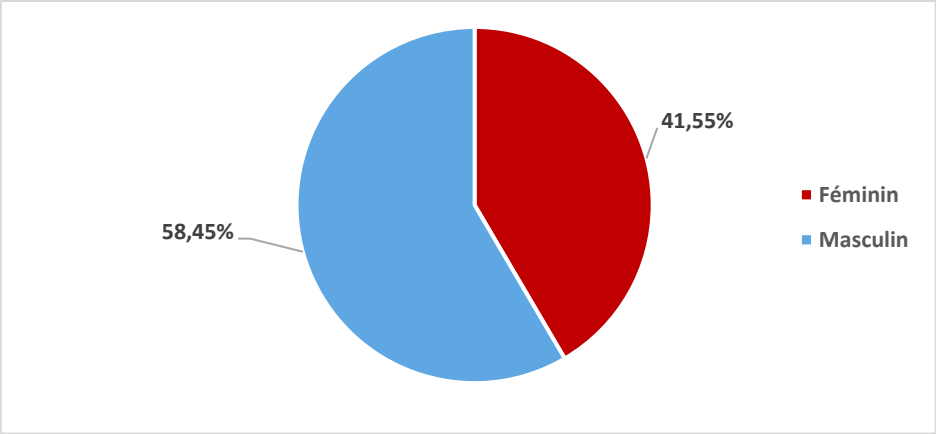


Figure 14: Répartition des assurés en perte d'emploi en fonction du sexe

- **Age lors de la survenance de la perte d'emploi**

L'âge lors de la survenance de la perte d'emploi des assurés sinistrés est compris entre 18 et 63 ans. L'âge moyen et l'âge médian sont très proches et valent respectivement 41,19 ans et 41,00 ans.

Par ailleurs, afin de pouvoir construire des tables de maintien en fonction de l'âge, nous avons fait un regroupement de la variable en deux classes : moins de 40 ans (47,90% des sorties) et plus de 40 ans (avec 52,10% des sorties).

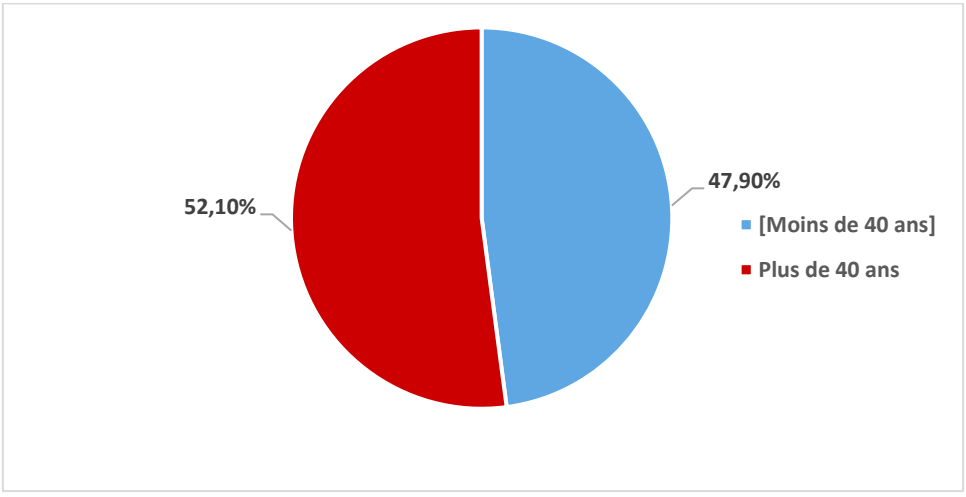


Figure 15: Répartition des assurés en perte d'emploi en fonction de la classe d'âge

- **Ancienneté lors de la survenance de la perte d'emploi**

L'ancienneté lors de la perte d'emploi des assurés sinistrés varie entre 3,13 et 70,07 mois. L'ancienneté médiane (18,40 mois) est inférieure à la moyenne qui vaut 21,60 mois ce qui dénote une certaine asymétrie de sa distribution.

Comme pour l'âge, l'ancienneté a été regroupée en classes. La première classe, constituée des assurés de moins de 18 mois d'ancienneté, regroupe le plus d'assurés (46,76%) comparé aux deux autres classes entre 18 et 30 mois (28,02%) et plus de 30 mois (25,00%).

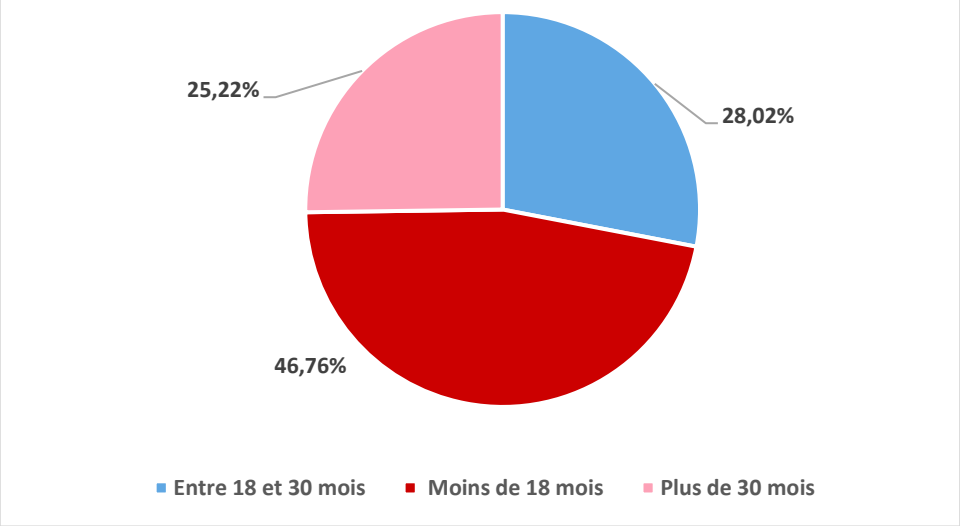


Figure 16: Répartition des assurés en perte d'emploi en fonction de l'ancienneté avant la perte d'emploi

II.2.3 Statistiques sur les prêts

- **Montant initial**

Le plus petit capital assuré parmi les assurés en perte d'emploi s'élève à 1.500 euros alors que le grand atteint 56.000 euros. Comme pour l'ancienneté, le montant initial du prêt moyen (16.161 euros) est supérieur au montant médian (14.000). Nous avons également regroupé la variable en quatre classes avec des proportions presque uniformes même si la classe des moins de 9.000 euros regroupe le plus d'assurés (26,50%) et la classe entre 9.000 et 14.000 euros le moins (23,82%).

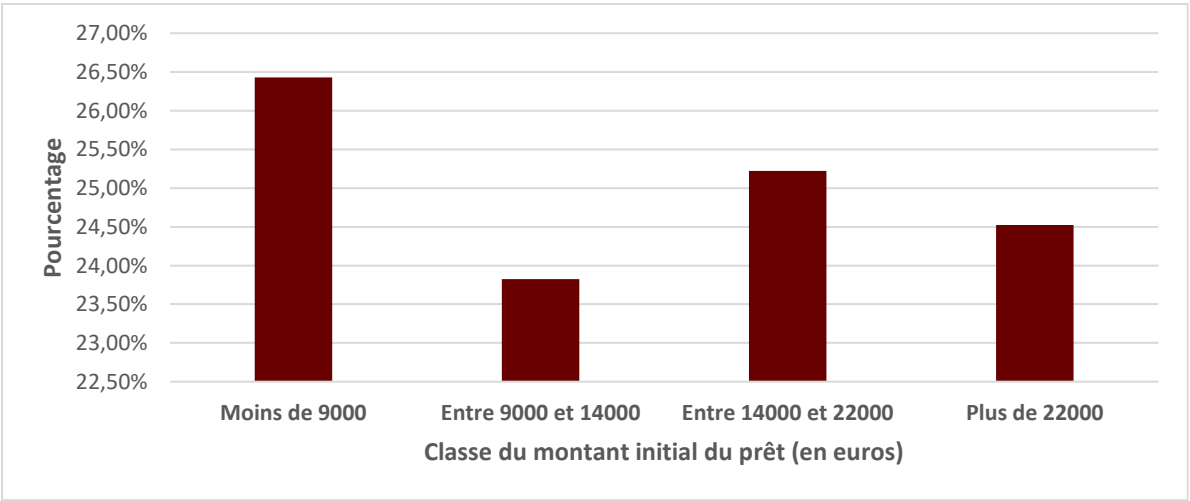


Figure 17: Répartition des sinistres en fonction du montant initial du prêt

- **Montant de l'échéance du prêt**

Le montant de l'échéance du prêt des assurés sinistrés varie entre 2,48 et 1.326,04 euros avec une échéance moyenne de 298,78 euros. Par ailleurs, comme le montant initial, nous avons regroupé cette variable en quatre classes avec chacune d'elle couvrant environ 25% de l'échantillon.

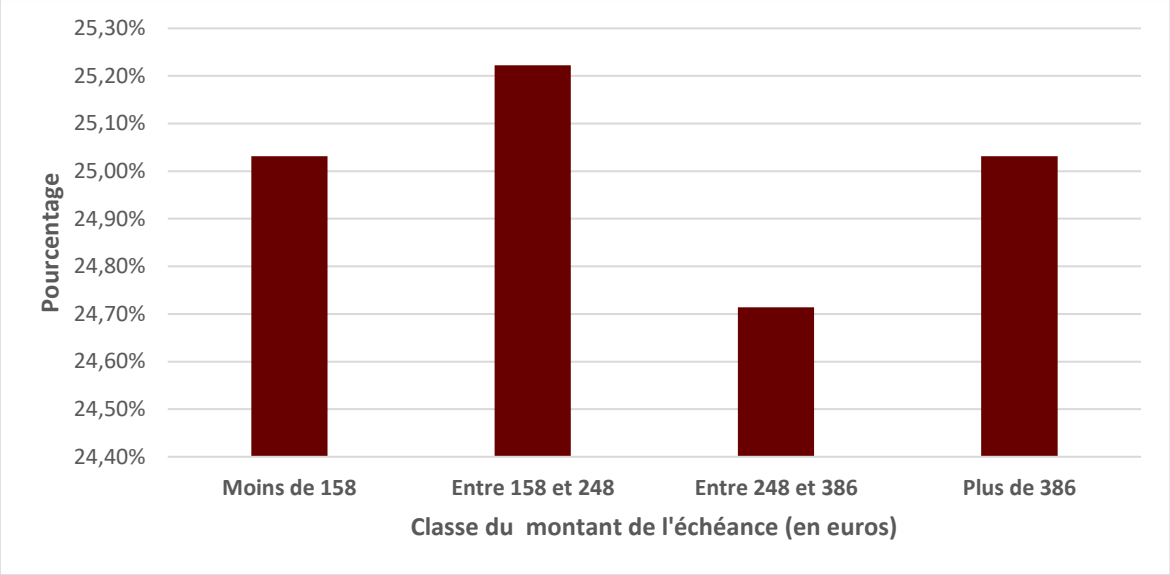


Figure 18:Répartition des sinistres en fonction du montant de l'échéance

Partie III. Méthodologie de construction de la table de loi de maintien

III.1 Construction de table sans variables explicatives

III.1.1 Estimateurs des taux bruts par la méthode de Kaplan-Meier

La méthode de Kaplan - Meier a été introduite en 1958 par Edward L. KAPLAN et Paul MEIER dans un article du journal *the American Statistical Association* nommé « Nonparametric estimation from incomplete observations ». Elle permet de prendre en compte les censures et les troncatures. Il s'agit d'un modèle très répandu autant en actuariat pour estimer les lois de mortalité ou de maintien en incapacité/invalidité/perte d'emploi, qu'en démographie et médecine pour évaluer la durée de vie de patients suivant un traitement par rapport à ceux qui ne le suivent pas. Ce modèle est assez simple à mettre en place d'où sa popularité. Le principe de cette méthode est d'estimer la fonction de survie des vies observées par un estimateur non paramétrique et est basé sur le principe qu'être en vie à l'instant t c'est être en vie juste avant et ne pas décéder à cet instant. Il fait donc intervenir les probabilités de survie en t conditionnellement au fait d'être en vie juste avant.

- ✓ **Notations**

T : variable aléatoire de durée avant sortie de perte d'emploi pour un individu

$S(t)$: fonction de survie à l'âge t, définie par $S(t) = P(T > t)$

q_i : probabilité de sortir de la perte d'emploi le mois i

n_i : nombre d'assurés exposés au risque au début du mois i

d_i : nombre de personnes sortant de la perte d'emploi le mois i

c_i : nombre de personnes censurées à droite le mois i (sorties de la période d'observation et toujours en perte d'emploi)

t_i : nombre de personnes tronquées à gauche le mois i (arrivées en cours d'observation)

Les n_i sont calculés par la formule de récurrence suivante, pour $i > i_{min}$ (i_{min} correspondant à la date associée)

$$n_{i+1} = n_i - d_i - c_i + t_i$$

III.1.1.1 *Estimation*

L'estimateur de Kaplan-Meier repose sur la possibilité d'exprimer $S(t)$ à l'aide de $S(s)$, pour $t > s$:

$$S(t) = P(T > t/T > s) \times S(s)$$

Par récurrence, on peut alors écrire la fonction de survie à une date t (exprimée en mois) :

$$S(t) = \prod_{i=i_{min}}^{i=t-1} P(T > i+1 \,/\, T > i)$$

Or, un estimateur naturel de :

$$q_i = 1 - P(T > i+1 \,/\, T > i)$$

C'est-à-dire, de la probabilité de sortie de la perte d'emploi le mois i est :

$$\hat{q}_i = \frac{d_i}{n_i}$$

Par conséquent, on obtient l'estimateur suivant pour la fonction de survie à la date t :

$$\hat{S}(t) = \prod_{i=i_{min}}^{i=t-1} \left(1 - \frac{d_i}{n_i}\right)$$

Par suite, un estimateur de la probabilité de sortir de la perte d'emploi au bout de x mois, c'est-à-dire sur l'intervalle $[x, x+1[$ est :

$$\hat{q}_x = 1 - \frac{\hat{S}(x+1)}{\hat{S}(x)}$$

- ✓ **Exemple d'estimation des taux bruts $\widehat{q_x}$**

Les données utilisées dans cet exemple sont fictives. Les colonnes en bleu sont les données issues de la base et les colonnes avec écriture en rouge sont les résultats de l'application de la méthode de calcul de Kaplan-Meier.

Durée avant sortie de la PE (en mois)	d_i	c_i	t_i	n_i	$1-\frac{d_i}{n_i}$	Probabilité de survie de PE ($S(x)$)	Probabilité de sortie de PE ($\widehat{q_x}$)
0	0	0	0	1846	1	100,00%	0,00%
1	68	13	0	1846	0,963	96,32%	3,68%
2	75	19	0	1765	0,958	92,22%	4,25%
3	61	15	0	1671	0,963	88,86%	3,65%
4	68	29	0	1595	0,957	85,07%	4,26%
5	64	13	0	1498	0,957	81,43%	4,27%
6	384	17	0	1421	0,730	59,43%	27,02%
7	42	22	0	1020	0,959	56,98%	4,12%
8	35	15	0	956	0,963	54,89%	3,66%
9	43	13	0	906	0,953	52,29%	4,75%
10	41	16	0	850	0,952	49,77%	4,82%
11	61	12	0	793	0,923	45,94%	7,69%
12	0	720	0	720	1		0,00%

Tableau 3: Exemple d'application de la méthode Kaplan-Meier

- ✓ **Variance de Greenwood et intervalle de confiance**

Pour mesurer le niveau de robustesse de l'estimateur de Kaplan-Meier, on calcule sa variance plus connue sous le nom de Variance de Greenwood. Elle est donnée par :

$$Var\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \sum_{i=i\_min} \frac{d_i}{n_i \times (n_i - d_i)}$$

A partir de cette variance, on déduit le calcul des intervalles de confiance de niveau $1-\alpha$ de l'estimateur de Kaplan-Meier et les bornes sont définies comme suit :

$$\hat{S}(t) \pm \phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{Var(\hat{S}(t))}$$

Avec

$\hat{S}(t)$: l'estimateur de Kaplan-Meier à t

$\phi^{-1}\left(\frac{\alpha}{2}\right)$: quantile d'ordre $\frac{\alpha}{2}$ d'une loi normale centrée et réduite

$Var(\hat{S}(t))$: la variance de Greenwood à t

III.1.1.2 *Remarques sur l'estimateur de Kaplan-Meier*

La construction de l'estimateur de Kaplan-Meier repose sur deux hypothèses importantes : l'hypothèse de censure non informative et l'hypothèse d'homogénéité de la population étudiée.

- ✓ **Hypothèse de censure non informative**

Comme l'explique (COLLETAZ, 2020), elle correspond à l'hypothèse d'indépendance entre le processus déterminant le temps de survenance de l'événement T et celui déterminant le temps de censure C_i. Cette hypothèse est fondamentale car lorsqu'elle n'est pas vérifiée, l'estimateur de Kaplan-Meier est biaisé. Toutefois, il n'y a pas de tests statistiques pour la vérifier.

- ✓ **Hypothèse d'homogénéité de la population étudiée**

Comme son nom l'indique, cette hypothèse signifie que la distribution ou la loi du temps de survenance de l'événement étudié (sortie de la perte d'emploi) est identique pour tous les individus de la population. Lorsque cette hypothèse n'est pas vérifiée, les estimations réalisées correspondent à des mélanges de distributions difficilement interprétables. Ainsi, il est généralement recommandé (COLLETAZ, 2020) d'effectuer l'estimation de Kaplan-Meier au sein de sous-échantillons plus homogènes (homme/femme, tranche d'âge lors de la perte d'emploi, etc.). Contrairement à l'hypothèse de censure non informative, l'hypothèse d'homogénéité peut être vérifiée à l'aide de tests statistiques de comparaison de fonctions de survie de différentes sous-populations.

III.1.2 Lissage des taux bruts

III.1.2.1 *Idée générale d'un lissage*

L'ensemble des taux bruts par mois estimés forme une courbe de sortie qui est généralement assez irrégulière. Les irrégularités proviennent généralement de grandes fluctuations des taux bruts de sortie d'un mois à l'autre. Par ailleurs, la courbe de taux bruts ne reproduit pas forcément les a priori formés sur le phénomène. Ainsi, les méthodes de lissage permettent de corriger les manquements des taux bruts.

Formellement, les estimations des taux bruts $\widehat{q_i}$ du vrai taux q_i a généré une erreur $e_i = \widehat{q_i} - q_i$. Le lissage a pour objectif de réduire cette erreur tout en rendant la courbe de taux aussi lisse que possible. Par conséquent, le choix du lissage conduit à un arbitrage entre deux critères : la fidélité (minimiser l'erreur $e_i = \widehat{q_i} - q_i$) et la régularité (smoothness).

Plusieurs méthodes de lissage sont présentes dans la littérature mais nous retiendrons la méthode de Whittaker-Henderson, méthode très utilisée en construction de table d'expérience.

III.1.2.2 *Méthode de Whittaker-Anderson*

Le lissage par la méthode de Whittaker-Henderson, est une méthode reposant sur la minimisation des critères de fidélité et de régularité. En appelant $(w)_{i=1,\dots,n}$ un vecteur de poids, le critère s'exprime ainsi :

$$F = \sum_{i=1}^{n} w_i \times (q_i - \hat{q}_i)^2$$

De plus, en notant $\triangle^r$ la différence avant d'ordre r, c'est-à-dire que :

$$si \ \triangle u(x) = u(x+1) - u(x) \ alors \ \triangle^r u(x) = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} u(x+j)$$

En notant z le nombre de paramètres du modèle, le critère de régularité s'exprime ainsi:

$$S = \sum_{i=1}^{n-z} (\triangle^r q_i)^2$$

En notant également h un paramètre de poids réel positif, la méthode de Whittaker-Henderson minimise une combinaison linéaire de deux critères :

$$M = F + h \times S$$

Le réel h permet de contrôler l'influence que l'on souhaite donner à la fidélité par rapport à la régularité. Si h est grand, la régularité est privilégiée, sinon c'est la fidélité qui est privilégiée.

III.1.2.3 *Critères de validation des taux lissés*

Puisqu'il n'existe pas de solution miracle pour le choix du paramètre de lissage et de l'ordre de différence qui permettrait d'obtenir une courbe de taux fidèle et suffisamment lisse, il faut utiliser des critères indiquant si le niveau de fidélité et de régularité de la courbe obtenue après lissage est acceptable. Ainsi, nous utilisons, d'une part, le nombre de signes positifs et négatifs des résidus de la réponse comme critère évaluant la régularité. De l'autre, les critères du chi-deux (χ^2) et le coefficient de détermination (R^2) seront utilisés pour évaluer la fidélité comme préconisé dans (TOMAS & PLANCHET, 2016).

III.1.2.3.1Test des signes

Le test des signe est un test non-paramétrique qui étudie la fréquence des changements de signes de la différence entre les taux de sortie bruts et lissés. Sous l'hypothèse nulle du test, la médiane des différences entre les taux est nulle. Par conséquent, l'hypothèse nulle de ce test signifie que la courbe des taux lissés est assez régulière. La statistique du test est donnée par :

$$\xi^{SIG} = \frac{|n_+ - n_-| - 1}{\sqrt{n}}$$

Avec

n_+: le nombre de signes positifs,

n_-: le nombre de signes négatifs

et $n = n_+ + n_-$.

Sous l'hypothèse nulle H_0, la statistique ξ^{SIG} suit une loi normale centrée réduite, $\xi^{SIG} \sim \mathcal{N}(0,1)$. L'hypothèse nulle H_0 est rejetée si

$$|\xi^{SIG}| > \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$\phi^{-1}\left(1 - \frac{\alpha}{2}\right)$: quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée et réduite.

III.1.2.3.2Test du χ^2

Il s'agit d'une distance permettant de mesurer l'écart entre les taux bruts et ceux lissés. La statistique du test s'écrit :

$$\chi^2 = \sum_{i=i_{min}}^{i_{max}} \frac{n \times \left(\widehat{q_i} - \widehat{q_i^l}\right)^2}{\widehat{q_i} \times (1 - \widehat{q_i^l})}$$

i_{min} : durée minimum

i_{max} : durée maximum

$\widehat{q_i}$: probabilité brute (taux brut) de sortir de la perte d'emploi le mois i

$\widehat{q_i^l}$: probabilité lissée (taux lissé) de sortir de la perte d'emploi le mois i

Sous l'hypothèse nulle H_0, stipulant que les deux taux bruts et lissés sont statistiquement identiques, cette statistique suit une loi du chi-deux à $i_{max} - 1$ degrés de liberté : $\chi^2 \sim \chi(i_{max} - 1)$. Plus petite sera cette distance, mieux est l'adéquation des taux lissés.

Ainsi, on rejette l'hypothèse nulle, si

$$\chi^2 > \chi_{1-\alpha}(i_{max} - 1)$$

avec $\chi_{1-\alpha}(i_{max} - 1)$: quantile d'ordre $1 - \alpha$ d'une loi du chi-deux à $i_{max} - 1$ degrés de liberté.

III.1.2.3.3Coefficient de détermination R^2

Généralement utilisé pour mesurer l'adéquation entre un modèle et les données observées, le coefficient de détermination est défini comme le rapport entre la variance expliquée et la variance totale. En adaptant cette définition au présent cas de figure, le coefficient de détermination représentera la part de la variance expliquée par les taux lissés sur la variance des totale des taux bruts :

$$R^2 = 1 - \left(\frac{\sum_{i=\, i_{min}}^{i\_max} (\widehat{q_i} - \widehat{q_i^l})^2}{\sum_{i=i_{min}}^{i_{max}} (\widehat{q_i} - \frac{1}{i_{max}} \sum_{i=i_{min}}^{i_{max}} \widehat{q_i})^2} \right)$$

Il a une valeur comprise entre 0 et 1 et plus il est proche de 1, plus les taux lissés sont fidèles aux taux bruts. Pour un ordre de différence r fixé, le R^2 a tendance à diminuer systématiquement lorsque le paramètre de lissage h augmente.

III.1.3 **Backtesting**

Le backtesting consiste à comparer le nombre de sorties de perte d'emploi prédites par la table retenue avec le nombre de sorties réellement observées sur des observations qui n'ont pas servi à la construction de la table. Généralement, ces observations sont constituées des sorties de perte d'emploi sur une période différente de la période d'observation de l'échantillon initial. L'objectif est de voir si les taux de la table reflètent bien la réalité et s'ils sont suffisamment stables également dans le temps.

Pour chaque durée en perte d'emploi, le nombre de sorties prédites est calculé, compte tenu du nombre d'assurés à risque :

$$\widehat{d_i} = sorties\ de\ la\ perte\ d'emploi\ prédites\ au\ mois\ i = n_i \times \widehat{q_i^l}$$

avec $\widehat{q_i^l}$: probabilité lissée (taux lissé) de sortir de la perte d'emploi le mois i et n_i : nombre d'assurés exposés au risque au début du mois i.

Par ailleurs, en supposant que $\widehat{q_i^l}$ suit une loi normale de moyenne q_i et de variance $\frac{q_i \times (1-q_i)}{n_i}$, on construit un intervalle de confiance des sorties prédites au mois i :

$$\widehat{d_i} \pm \sqrt{n_i \times \widehat{q_i^l} \times (1 - \widehat{q_i^l})} \times \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$\phi^{-1}\left(1 - \frac{\alpha}{2}\right)$: le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée réduite

Lorsque toutes les sorties observées se trouvent dans les intervalles de confiance, on considère que la table retenue a réussi le backtesting.

III.2 Evaluation de l'influence des covariables et modèles de régression

Afin d'évaluer l'influence et la significativité des covariables, nous allons réaliser un test de comparaison de sous-populations et des modèles régression. Le test de comparaison a l'avantage d'être totalement non-paramétrique alors que les modèles de régression utilisés, modèle à risques proportionnels, ne sont que semi-paramétriques.

III.2.1 Test de comparaison de sous-populations : test de LogRank

Les tests de comparaison de sous-populations sont des problèmes assez connus en statistique. On cherche généralement à comparer les moyennes de deux ou plusieurs populations avec une hypothèse nulle stipulant l'égalité des deux moyennes. Le test le plus connu pour ce cas de figure est le t-test de Student, utilisé pour comparer par exemple la taille moyenne chez des hommes et femmes d'une population. Toutefois, ce genre de test paramétrique nécessite de faire une hypothèse sur la distribution des données en l'occurrence la normalité de la variable étudiée. La nature des données de durée (distribution non normale) et la présence des censures dans les observations rendent inadéquate l'utilisation des tests classiques de comparaison de sous-population (MOORE, 2016). Par conséquent, il faut faire appel à des tests non-paramétriques basés sur les rangs pour évaluer la significativité de variables explicatives. Lorsqu'on adapte ces tests dans le cadre des données de durée, pour une variable explicative qualitative à deux modalités c'est-à-dire séparant la population en deux sous-populations ou groupes, on a la spécification suivante :

$$H_0 : S_1(t) = S_2(t) \text{ contre } H_1 : S_1(t) \neq S_2(t)$$

Où S_1est la fonction de survie de la première sous-population et S_2 celle de la deuxième.

Nous allons présenter le test de LogRank, test basé sur les rangs et adapté aux données de durée.

Connu également sous le nom de test de Mantel-Haenzel, le test de LogRank fait partie des tests de rang les plus utilisés. Nous présenterons d'abord le principe et le fonctionnement du test pour deux sous-populations avant de traiter l'extension à plus de deux sous-populations.

- ✓ **Cas avec deux sous-populations**

Après avoir ordonné les durées de sorties, on construit le tableau de contingence suivant pour chacune d'elle :

Groupe	1	2	**Total**
Sorties censurées	c_{1i}	c_{2i}	c_i
Sorties non-censurées	d_{1i}	d_{2i}	d_i
Total	n_{1i}	n_{2i}	n_i

Avec n_{ji} : nombre de personne du groupe j exposés au risque au début du mois i ($j = 1, 2$)

d_{ji} : nombre de personnes du groupe j sortant de la perte d'emploi le mois i

c_{ji} : nombre de personnes du groupe j censurées à droite le mois i (sorties de la période d'observation et toujours en perte d'emploi)

Sous l'hypothèse nulle (H_0: $S_1 = S_2$), la proportion attendue de sorties non-censurées au mois i est donnée par $\frac{d_i}{n_i}$. Le nombre de sorties non-censurées du groupe j au mois i, d_{ji}, suit une loi hypergéométrique (voir (MOORE, 2016) pour plus de détail) de moyenne et de variance donnée par :

$$\boldsymbol{e_{ji}} = \mathbb{E}(d_{ji}) = \frac{n_{ji}d_i}{n_i} \text{ et } \boldsymbol{v_{ji}} = \mathbb{V}ar(d_{ji}) = \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}$$

Par ailleurs, quatre quantités sont calculées :

- O_1 : le nombre de total de sorties observées du groupe 1, somme des d_{1i}
- O_2 : le nombre de total de sorties observées du groupe 2, somme des d_{2i}
- E_1 : le nombre de total de sorties espérées du groupe 1, somme des e_{1i}
- E_2 : le nombre de total de sorties espérées du groupe 2, somme des e_{2i}

Ainsi, la statistique du test (qui suit une loi du chi-deux à 1 degré de liberté) est :

$$\frac{{U_1}^2}{V_1} \sim \chi_1^2$$

Avec $U_1 = O_1 - E_1$ et $V_1 = \mathbb{V}ar(U_1) = \sum_{i=1}^{i_{max}} v_{1i}$

On rejette l'hypothèse nulle si la statistique est supérieure au quantile d'ordre $1 - \alpha$ d'une loi du chi-deux à 1 degré de liberté ou si la $p - valeur$[4] $= P(\chi_1^2 > \frac{U_1^2}{V_1})$ est inférieure à α.

[4] La p-valeur est la plus petite valeur du risque de première espèce α pour laquelle on rejette l'hypothèse nulle du test.

- ✓ **Extension avec k sous-populations, $k \geq 2$**

L'extension du test à k sous-populations est automatique. En effet, toutes les étapes décrites précédemment restent valides mais dans ce cas-ci, il y a k groupes. La statistique reste la même en revanche la variance V_1 est remplacée par une matrice de variance covariance et U_1 devient la somme de $k-1$ termes $O_j - E_j$ choisis au hasard parmi les k (voir (COLLETAZ, 2020) et (MOORE, 2016) pour plus de détails). Ainsi, la statistique du test suit une loi du chi-deux à $k-1$ degrés de liberté.

III.2.2 Modèle à risques proportionnels : cas du modèle de Cox

Dans cette section, nous allons étudier la significativité des variables explicatives et la constance de leurs effets dans le temps à l'aide de modèles de Cox. Nous présenterons d'abord, le modèle de Cox sous sa forme simple (standard) et sous sa forme étendue (extended Cox model) permettant de résoudre le problème engendré par la violation de l'hypothèse de risques proportionnels.

III.2.2.1 *Généralités sur les modèles de Cox*

Le modèle de Cox fait partie de la famille des modèles à risques proportionnels (proportionnal hazard). Ces derniers sont usuellement spécifiés comme suit :

$$h(t) = h_0(t)r(x)$$

où $h()$ est la fonction de risque définie par $h(t) = \frac{f(t)}{S(t)}$, $f(t)$ la densité et $S(t)$ la fonction de survie ;

$h_0()$ est la fonction de risque de base ;

$r()$ est une fonction indépendante du temps et qui ne dépend que de x, les caractéristiques des individus.

Cette spécification stipule que la fonction de risque de $h()$ est proportionnelle à une autre fonction de risque de base $h_0()$ et le coefficient de proportionnalité dépend uniquement des caractéristiques des individus.

Ainsi, on peut voir que cette spécification implique la nécessité de formuler deux hypothèses : l'une sur la fonction de risque de base $h_0()$ et l'autre sur la forme de $r()$.

III.2.2.2 *Modèle de Cox simple*

III.2.2.2.1 Présentation et estimation

Le modèle de Cox ne fait aucune hypothèse sur $h_0()$ à part le fait qu'elle soit positive ($h_0(t) > 0$) mais donne une forme paramétrique à $r(\boldsymbol{x})$:

$$r(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\beta) = \exp\left(x_1\beta_1 + x_2\beta_2 + \cdots\right)$$

En effet, cette spécification fait que le modèle de Cox est dit semi-paramétrique. Seuls les coefficients β sont à estimer par la méthode du maximum de vraisemblance. Ainsi, la vraisemblance de l'échantillon est donnée par :

$$L = \prod_{i=1}^{n}[f_i(t_i)\,]^{\delta_i}[S_i\,(t_i)\,]^{1-\delta_i}$$

Avec $f_i(t_i)$ la fonction de densité de l'individu i au temps d'événement t_i et on peut réécrire : $f_i(t_i) = h_i(t_i)S_i(t_i)$;

$S_i(t_i)$ la fonction de survie de l'individu au temps d'événement t_i ;

δ_i l'indicatrice de censure telle que $\delta_i = 1$ si l'événement est entièrement observé pour l'individu et $\delta_i = 0$ sinon ;

n : la taille de l'échantillon.

En remplaçant $f_i(t_i)$ par $h_i(t_i)S_i(t_i)$, la vraisemblance devient :

$$L = \prod_{i=1}^{n}\left[\frac{h_i(t_i)}{\sum_{j=1}^{n_{t_i}} h_j(t_i)}\right]^{\delta_i}\left[\sum_{j=1}^{n_{t_i}} h_j(t_i)\right]^{\delta_i} S_i\,(t_i)$$

Cox introduit la vraisemblance partielle, expression tirée de la vraisemblance L, pour éliminer la référence de base dans son expression ($h_0\ et\ S_0$) :

$$PL = \prod_{i=1}^{n}\left[\frac{h_i(t_i)}{\sum_{j=1}^{n_{t_i}} h_j(t_i)}\right]^{\delta_i} = \prod_{i=1}^{n}\left[\frac{h_0(t_i)\times exp(\boldsymbol{x_i}^T\beta)}{\sum_{j=1}^{n_{t_i}} h_0(t_i)\times exp(\boldsymbol{x_j}^T\beta)}\right]^{\delta_i} = \prod_{i=1}^{n}\left[\frac{exp(\boldsymbol{x_i}^T\beta)}{\sum_{j=1}^{n_{t_i}} exp(\boldsymbol{x_j}^T\beta)}\right]^{\delta_i}$$

Cox propose d'utiliser la vraisemblance partielle pour estimer les coefficients β avec un algorithme du type Newton-Raphson.

III.2.2.2.2Diagnostic du modèle

- ✓ **Tests de significativité de l'ensemble des coefficients**

Les tests de significativité sont des tests basés sur la vraisemblance qui cherchent à vérifier si tous les coefficients du modèle sont simultanément nuls. Ainsi, l'hypothèse nulle est : H_0: $\hat{\beta}_1 = \hat{\beta}_2 = \cdots = 0$ et l'alternative est qu'il existe au moins un des coefficients β_i non nul.

- ○ **Likelihood Ratio (test du rapport de vraisemblance)**

La statistique du Likelihood Ratio est :

$$LR = -2\ln\left(\frac{L(\hat{\beta})}{L(0)}\right) = -2\big(lnL(\hat{\beta}) - lnL(0)\big) \sim \chi^2 \text{ à } p \text{ degrés de liberté}$$

Où $L(\hat{\beta})$ est la vraisemblance calculée sur les observations de l'échantillon et $L(0)$ la vraisemblance sous l'hypothèse nulle c'est-à-dire tous les coefficients du modèle sont nuls ; p étant le nombre de paramètres $\hat{\beta}$ à estimer.

- ○ **Wald**

$$\big(\hat{\beta} - 0\big)^T I\big(\hat{\beta}\big)^{-1}\big(\hat{\beta} - 0\big) \sim \chi^2 \text{ à } p \text{ degrés de liberté}$$

Avec $\hat{\beta}$ vecteur de paramètre de dimension p, I la matrice d'information de Fisher

- ○ **Score**

$$U^T(0)I(0)^{-1}U(0) \sim \chi^2 \text{à } p \text{ degrés de liberté}$$

Avec $U(0) = \nabla_\beta L(0) = 0$: la valeur du gradient de la vraisemblance ou la fonction de score au point $\beta = 0$.

Puisque les trois statistiques suivent une loi du chi-deux alors on peut en déduire la p-valeur associée à chacune d'elle.

- ✓ **Tests de significativité individuelle des coefficients**

La significativité individuelle des coefficients est testée avec le classique t-test de Student. En effet, la statistique calculée pour chaque coefficient $\widehat{\beta_\iota}$ estimé est la suivante :

$$T = \frac{\widehat{\beta_\iota} - 0}{\sqrt{\widehat{Var(\beta_\iota)}}}$$

La statistique T suit une loi de Student à $n-1$ degré de liberté (n étant la taille de l'échantillon)

- ✓ **Diagnostic des résidus de Schoenfeld et de l'hypothèse de risques proportionnels**

Les résidus de Schoenfeld fournissent un moyen de tester la robustesse de l'hypothèse de risques proportionnels. Ils sont calculés pour chaque variable k chaque individu i et pour chaque durée t_i avant sortie. En choisissant une variable x_k du vecteur de variables explicatives (covariables) x pour l'individu i et en posant $p(\widehat{\beta_k}, x_k) = \frac{exp(x_{ki}\beta_k)}{\sum_{j=1}^{n_{t_i}} exp(x_{kj}\beta_k)}$, les résidus de Schoenfeld sont définis par :

$$\widehat{r_{ki}} = x_{ki} - \sum_{j=1}^{n_{t_i}} x_{ki} \times p(\widehat{\beta_k}, x_{kj}) = x_{ki} - \overline{x_k}(t_i)$$

(GRAMBSCH & THERNEAU, 1994) proposent de normaliser les résidus en les divisant par la variance de $\widehat{\beta_k}$ et en multipliant par n_{t_i}. Ainsi les résidus normalisés peuvent être approchés par :

$$r_{ki}^* = \frac{r_{ki} \times n_{t_i}}{var(\widehat{\beta_k})}$$

Par ailleurs, ces mêmes auteurs montrent que si le ratio des taux de risque est une fonction de $t, \beta(t)$, alors

$$E(r_{ki}^*) \approx \beta_k + \beta_{ki}(t_i)$$

Ainsi, on peut obtenir une approximation de $\beta_{ki}(t_i)$ en soustrayant des résidus normalisés $\widehat{\beta_k}$, estimation issue du modèle de Cox sous l'hypothèse de risques proportionnels. Il est possible de construire un test statistique pour vérifier l'hypothèse de risques proportionnels. Ce test consiste à effectuer une régression linéaire des résidus normalisés sur le temps et de tester sa significativité. En effet, il s'agit de la régression suivante :

$$\beta_k + \beta_{ki}(t_i) = a + bt + \epsilon$$

Lorsque le coefficient b est significativement nulle alors l'hypothèse de risques proportionnels ne pourrait être rejetée. Il convient de noter que ce test ne se limite qu'à une spécification linéaire et n'est pas en mesure de détecter une relation d'ordre supérieur (quadratique par exemple).

Par ailleurs, on est en mesure de représenter graphiquement le nuage de points des résidus normalisés et donc des $\beta_{ki}(t_i)$. En effet, avec une régression de LOESS (Locally Estimated Scatterplot Smooting), il est possible de tracer la courbe de $\beta_k(t)$ en fonction du temps comme le permet la fonction ***plot.cox.zph*** du package *survival* de R. La régression de LOESS est une régression non-paramétrique permettant de fournir des courbes lissées et bien ajustées à un nuage de points (voir (CLEVELAND, 1974) pour plus de détails).

III.2.2.3 *Modèles de Cox étendus*

L'hypothèse de risques proportionnels et indépendants du temps est une hypothèse très forte, qui signifie que les coefficients β sont indépendants du temps (dans notre cas, la durée avant sortie de perte d'emploi). Dans certaines situations, les tests montrent une dépendance des coefficients β au temps. De plus, une ou plusieurs variables explicatives (caractéristiques des individus) peuvent dépendre du temps. Dans ces deux cas, l'hypothèse de risques proportionnels n'est pas vérifiée et il faut donc utiliser d'autres spécifications, plus connues sous le nom de modèles de Cox étendus, pour résoudre ce problème ((HARELL, 2006); (ALLISON, 2010)). L'une des alternatives est d'introduire une interaction avec une fonction déterministe du temps dans l'expression du taux de risque instantané comme dans les équations ci-dessous :

$$h(t_i) = h_0(t_i)\exp(\boldsymbol{x^T} g(\beta, t_i))$$
$$h(t_i) = h_0(t_i)\exp(\boldsymbol{x^T}(t_i)\beta)$$

avec $g(\beta, t_i)$ une fonction déterministe du temps et de β.

La première équation est adaptée lorsque la variable explicative est constante dans le temps mais le coefficient ne l'est pas et la seconde équation traite le cas des variables explicatives dépendant du temps (THOMAS & REYES, 2014).

Généralement $g(\beta, t_i)$ est une fonction simple et facile à interpréter telle que $g(\beta, t_i) = G(t_i)^T\beta$ où $G(t_i)^T = (g_1(t_i), g_2(t_i), \dots,)$. Par exemple, si on définit t'un temps fixé $\epsilon\ [t_1, t_n]$ et $\mathbb{I}$ la fonction indicatrice, une forme facilement interprétable de $G(t_i)$ est la suivante :

$$G(t_i) = (\mathbb{I}_{(t_i<t')}, \mathbb{I}_{(t_i\geq t')})$$

Cette forme permet de définir un modèle où les coefficients sont constants par intervalle. Dans cet exemple, cela signifie qu'au lieu d'estimer uniquement un seul coefficient pour la

covariable, on estimera un coefficient β_1 sur l'intervalle $[t_1, t'[$ et un autre coefficient β_2 sur l'intervalle $[t', t_n]$. $G(t_i)$ est donc une fonction constante par morceaux.

On peut réécrire la fonction de risque proportionnel comme suit :

$$h(t_i) = h_0(t_i) \exp(\boldsymbol{x^T} G(t_i)^T \beta)$$
$$h(t_i) = h_0(t_i) \exp(\boldsymbol{x^T}(t_i)\beta)$$

où $\boldsymbol{x^T}(t_i) = \{\boldsymbol{x}G(t_i)\}^T$. Cette formulation montre que les modèles à coefficient dépendant du temps peuvent être estimés comme les modèles à variables explicatives dépendant du temps. Par exemple, en revenant sur cas où $G(t_i) = (\mathbb{I}_{(t_i<t')}, \mathbb{I}_{(t_i\geq t')})$, il s'agira de passer d'un vecteur x constant pour tout t_i à un vecteur dont chacune des variables x_k aura deux valeurs (distinctes ou non) $x_k(t_i|t_i < t')$ et $x_k(t_i|t_i \geq t')$.

Dans les modèles à coefficient dépendant du temps ou à variables explicatives dépendant du temps, l'estimation des coefficients β se fait comme dans le modèle de Cox simple puisqu'ils sont redevenus indépendants du temps par conséquent l'hypothèse de risques proportionnels redevient valide.

Partie IV. Résultats

Dans cette section, il s'agit de présenter d'abord la loi brute obtenue avec l'estimateur de Kaplan-Meier, ensuite de montrer les lois lissées avec la méthode de Whittaker-Anderson et enfin, d'analyser l'influence des covariables (sexe, âge, ancienneté, etc.) et l'effet de la crise sanitaire sur les taux de sortie.

IV.1 Construction de la table de maintien sur le portefeuille

IV.1.1 Tables brutes

La table de taux bruts est obtenue à partir de la fonction de survie estimée avec l'estimateur de Kaplan-Meier. Pour mettre en œuvre l'estimation des taux bruts, nous avons utilisé la fonction ***survfit*** du package ***Survival*** de R.

Les probabilités de survie estimées semblent assez précises compte tenu des intervalles de confiance assez restreintes dans lesquels elles se trouvent. Il faut toutefois noter que l'amplitude des intervalles de confiance augmente avec la durée passée en perte emploi (cf figure 19). Cette hausse est logique puisqu'il y a moins d'observations pour les durées élevées ce qui introduit de l'incertitude et donc une augmentation de la variance.

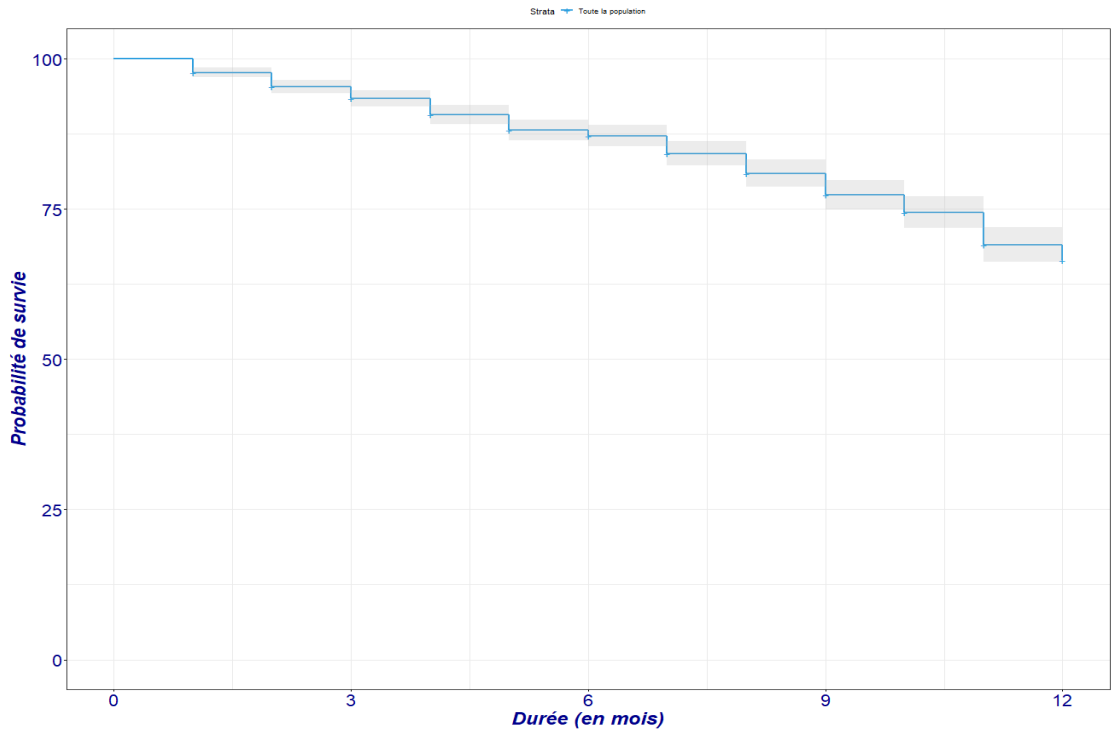


Figure 19: Fonction de survie de la loi brute

Les taux bruts (q_x) représentent la hauteur du saut de la fonction de survie entre deux durées. Ce calcul permet d'obtenir la table des taux bruts ci-dessous :

Temps depuis survenance PE (en mois)	Probabilité de survie de PE %)	Taux bruts de sortie de PE (q_x) (%)
0	100,00	2,31
1	97,69	2,39
2	95,36	2,09
3	93,37	2,88
4	90,69	2,84
5	88,11	1,05
6	87,19	3,38
7	84,24	3,91
8	80,95	4,50
9	77,30	3,74
10	74,42	7,22
11	69,05	3,76
12	66,45	

Tableau 4: Taux bruts de sortie

A partir de ces taux bruts de sorties, on parvient à tracer la courbe de taux bruts de sorties :

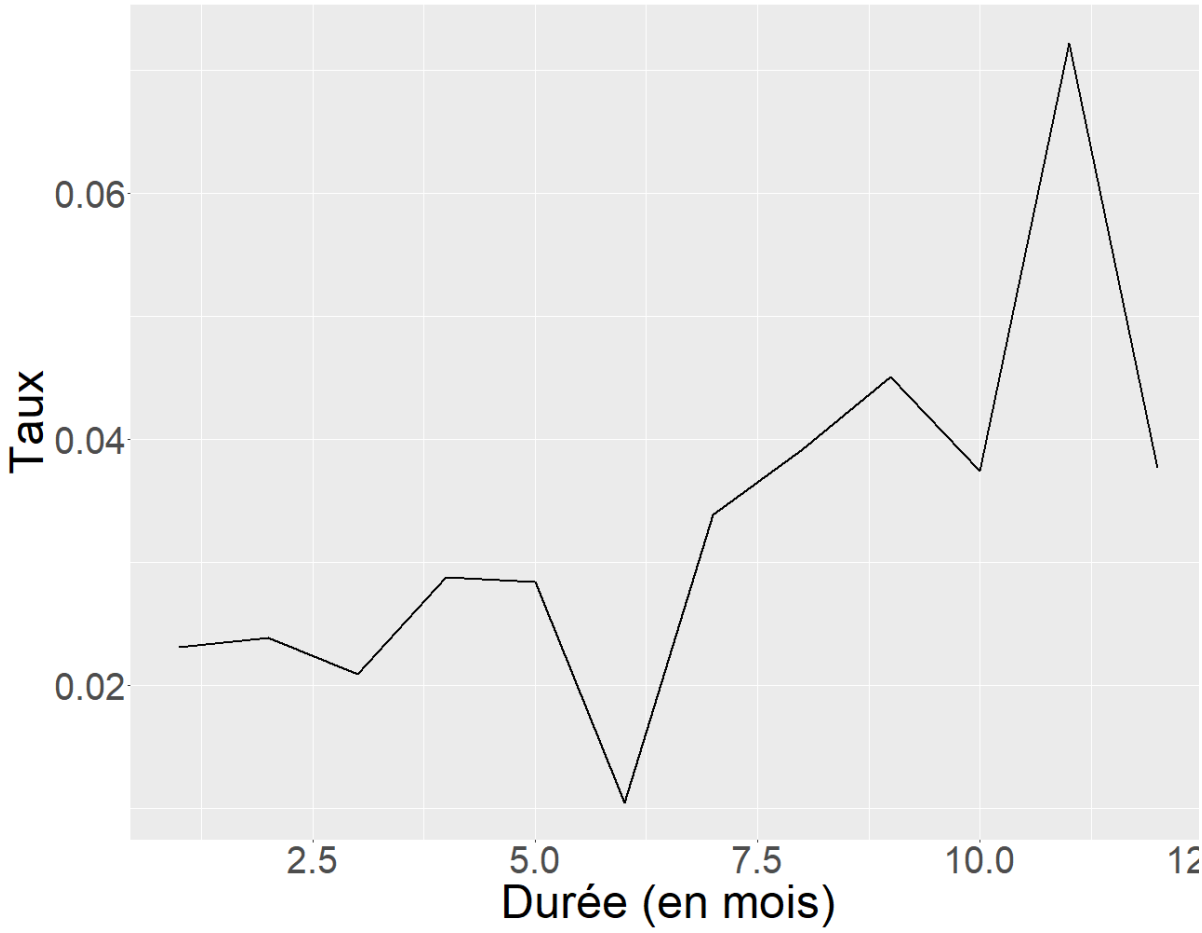


Figure 20: Courbe des taux bruts de sortie de PE

La courbe de taux bruts est très erratique, il est donc nécessaire de la lisser afin d'obtenir la courbe de taux lissées finale.

IV.1.2 **Taux lissés**

Pour obtenir la table des taux lissés, nous avons appliqué la méthode de lissage de Whittaker-Henderson sur les taux bruts. Plusieurs couples de paramètres (z, h) ont été testés mais nous n'avons retenu que cinq couples : $(2{,}1)$, $(3, 200)$, $(3, 500)$, $(5, 500)$ et $(4, 10.000)$.

Le graphique ci-dessous compare les courbes de taux lissés et la courbe de taux bruts.

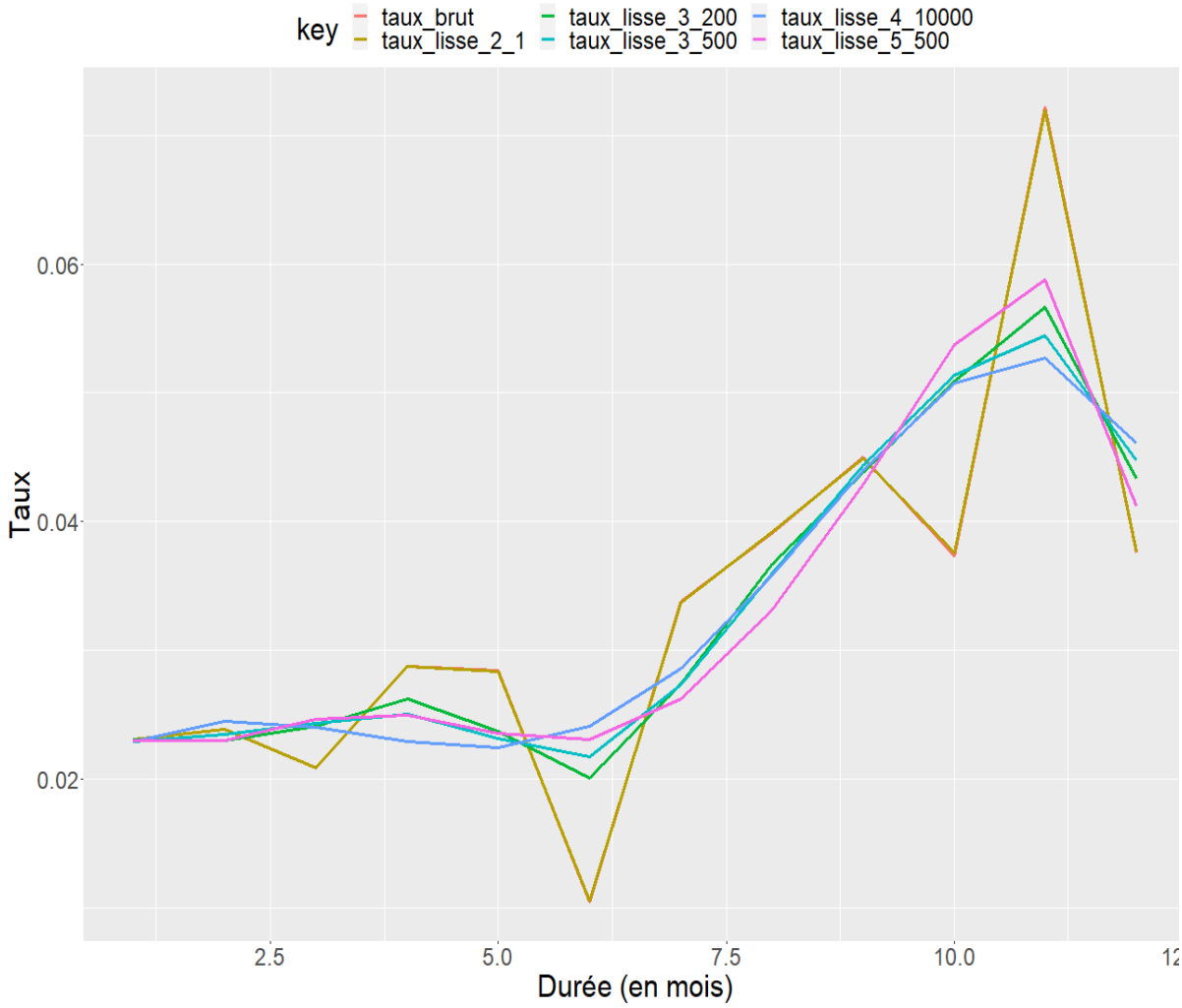


Figure 21: Comparaison taux bruts et des taux lissés

En calculant les statistiques de tests des signes, test du chi-deux et le coefficient de détermination pour les différentes courbes de taux lissés, on obtient le tableau ci-dessous. D'après le tableau 5, aucune des courbes lissées ne rejette l'hypothèse nulle du test des signes puisque la valeur de la statistique est inférieure à 1,96, le quantile d'ordre 97,5% d'une loi normale centrée réduite. En d'autres termes, le test des signes considère que toutes les courbes lissées sont suffisamment lisses. En ce qui concerne le test du chi-deux, on compare la valeur

de la statistique calculée par courbe au quantile d'ordre 95% d'une loi de chi-deux à onze degrés de liberté qui vaut 19,68. Dans ce cas, seule la statistique de la dernière courbe lissée ($z = 4$ et $h = 10\,000$) dépasse ce quantile et donc cette dernière s'écarte significativement de la courbe des taux bruts. Quant au coefficient de détermination (R^2), comme on s'y attendait, il diminue lorsque le paramètre de lissage (paramètre h) augmente c'est-à-dire lorsque la courbe est fortement lissée. Par exemple, la courbe 1, très erratique, a un coefficient de détermination maximal alors que la courbe 5 enregistre le coefficient de détermination la plus petite (0,65).

Numéro de courbe	Z	h	Test des signes	Chi-deux	R^2
1	2	1	0,29	0,00	1,00
2	3	200	0,87	15,03	0,76
3	3	500	0,87	18,77	0,70
4	5	500	0,87	19,68	0,71
5	4	10000	0,29	23,14	0,65

Tableau 5: Tableau de comparaison des statistiques de validation des courbes lissées

On peut constater que les tests purement statistiques à eux seuls ne facilitent pas le choix des paramètres de lissage. Ainsi, afin de choisir la courbe à retenir, nous allons nous baser sur les résultats du backtest sur l'année 2019. La courbe à retenir sera celle qui parviendra à prédire au mieux les sorties sur la période du 01/01/2019 au 31/12/2019.

IV.1.3 **Backtesting**

Dans cette partie, nous présenterons le tableau des résultats du backtesting pour chacune des courbes de taux lissés. Pour chaque durée au chômage, nous comparons d'une part les sorties réellement observées, avec les sorties prédites par les taux bruts et par les taux lissés. Les bornes inférieure et supérieure des intervalles de confiance de sortie (à l'ordre de 95%) sont également estimées.

- ✓ **Courbe 1 : z = 2 et h = 1**

Cette courbe de taux lissés prédit plus de sorties que ce qui est réellement observé. En effet, la courbe 1 prédit (globalement) 54 sorties contre 45 observées, soit un taux d'erreur de 17%. Pour toutes les durées avant sortie de la perte d'emploi, la courbe 1 prédit autant de sorties que les taux bruts (c'était attendu car cette courbe lissée est très similaire à la courbe de taux brut).

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(2)
1	196	4	5	5	0,4	8,64	-0,11
2	189	2	5	5	0,4	8,62	-0,56
3	181	3	4	4	0,01	7,56	-0,21
4	174	7	5	5	0,68	9,32	0,4
5	163	4	5	5	0,47	8,77	-0,13
6	156	2	2	2	0	4,15	0,21
7	122	1	4	4	0,2	8,02	-0,76
8	119	1	5	5	0,51	8,81	-0,79
9	113	6	5	5	0,76	9,39	0,18
10	105	6	4	4	0,13	7,77	0,52
11	95	7	7	7	1,9	11,77	0,02
12	87	2	3	3	0	6,76	-0,39
Total	-	45	54	54	5,46	99,58	-0,17

Tableau 6: Comparaison des sorties observées et prédites de la courbe 1 (z = 2, h = 1)

- ✓ **Courbe 2 : z = 3 et h = 200**

A l'image de la courbe 1, la courbe 2 prédit plus de sorties qu'observées en 2019 (51 contre 45). Bien qu'elle prédise plus de sorties, la courbe 2 est toutefois plus prudente que la courbe de taux brut.

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(2)
1	196	4	5	5	0,4	8,64	-0,11
2	189	2	5	4	0,31	8,39	-0,54
3	181	3	4	4	0,32	8,4	-0,31
4	174	7	5	5	0,43	8,69	0,54
5	163	4	5	4	0,05	7,66	0,04
6	156	2	2	3	0	6,57	-0,36
7	122	1	4	3	0	6,88	-0,7
8	119	1	5	4	0,34	8,38	-0,77
9	113	6	5	5	0,69	9,22	0,21
10	105	6	4	5	0,93	9,76	0,12
11	95	7	7	5	0,97	9,8	0,3
12	87	2	3	4	0,05	7,49	-0,47
Total		45	54	51	4,49	99,88	-0,12

Tableau 7: Comparaison des sorties observées et prédites de la courbe 2 (z = 3, h = 200)

- ✓ **Courbe 3 : z = 3 et h = 500**

Quant à la courbe 3, même si elle prédit plus de sorties (49 sorties) que celles observées, elle semble faire les meilleures prévisions de sorties pour 2019 jusque-là. Cette courbe de taux lissés est, par conséquent, prudente comparé à la courbe de taux bruts.

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(2)
1	196	4	5	4	0,38	8,59	-0,11
2	189	2	5	4	0,36	8,51	-0,55
3	181	3	4	4	0,34	8,47	-0,32
4	174	7	5	4	0,32	8,4	0,61
5	163	4	5	4	0,01	7,54	0,06
6	156	2	2	3	0	6,96	-0,41
7	122	1	4	3	0	6,87	-0,7
8	119	1	5	4	0,3	8,26	-0,77
9	113	6	5	5	0,72	9,3	0,2
10	105	6	4	5	0,96	9,83	0,11
11	95	7	7	5	0,84	9,5	0,35
12	87	2	3	4	0,11	7,67	-0,49
Total		45	54	49	4,34	99,9	-0,08

Tableau 8: Comparaison des sorties observées et prédites de la courbe 3 (z = 3, h = 500)

- ✓ **Courbe 4 : z = 5 et h = 500**

Comme les courbes précédentes, la courbe 4 prédit plus de sorties qu'observées en 2019 (53 contre 45). Bien qu'elle prédise plus de sorties, la courbe 4 est toutefois plus prudente que la courbe de taux bruts.

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(2)
1	196	4	5	5	0,4	8,63	-0,11
2	189	2	5	4	0,3	8,38	-0,54
3	181	3	4	4	0,37	8,55	-0,33
4	174	7	5	4	0,31	8,37	0,61
5	163	4	5	4	0,04	7,63	0,04
6	156	2	2	4	0	7,27	-0,44
7	122	1	4	3	0	6,65	-0,69
8	119	1	5	4	0,11	7,76	-0,75
9	113	6	5	5	0,63	9,07	0,24
10	105	6	4	6	1,12	10,18	0,06
11	95	7	7	6	1,09	10,07	0,25
12	87	2	3	4	0	7,22	-0,44
Total		45	54	53	4,37	99,78	-0,15

Tableau 9 : Comparaison des sorties observées et prédites de la courbe 4 (z = 5, h = 500)

- ✓ **Courbe 5 : z = 4 et h = 10 000**

Comme les courbes précédentes, la courbe 5 prédit également plus de sorties qu'observées en 2019 (51 contre 45) même si elle est plus prudente que la courbe de taux bruts.

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(2)
1	196	4	5	4	0,38	8,59	-0,11
2	189	2	5	5	0,46	8,78	-0,57
3	181	3	4	4	0,31	8,39	-0,31
4	174	7	5	4	0,12	7,85	0,76
5	163	4	5	4	0	7,37	0,09
6	156	2	2	4	0	7,51	-0,47
7	122	1	4	3	0	7,1	-0,71
8	119	1	5	4	0,29	8,23	-0,77
9	113	6	5	5	0,7	9,24	0,21
10	105	6	4	5	0,92	9,74	0,13
11	95	7	7	5	0,74	9,28	0,4
12	87	2	3	4	0,18	7,85	-0,5
Total		45	54	51	4,1	99,93	-0,12

Tableau 10: Comparaison des sorties observées et prédites de la courbe 5 (z = 4, h = 10 000)

Toutes les courbes lissées prédisent plus de sorties que les sorties observées en 2019, même si elles prédisent moins de sorties que les taux bruts. Par ailleurs, la courbe 3 fait le moins d'erreurs de prédiction et pour chaque durée, les sorties observées se situent dans l'intervalle de confiance du nombre de sorties prédites. Ainsi, nous retenons la courbe 3 pour constituer la table finale.

La figure ci-dessous compare les taux bruts et les taux lissés retenus.

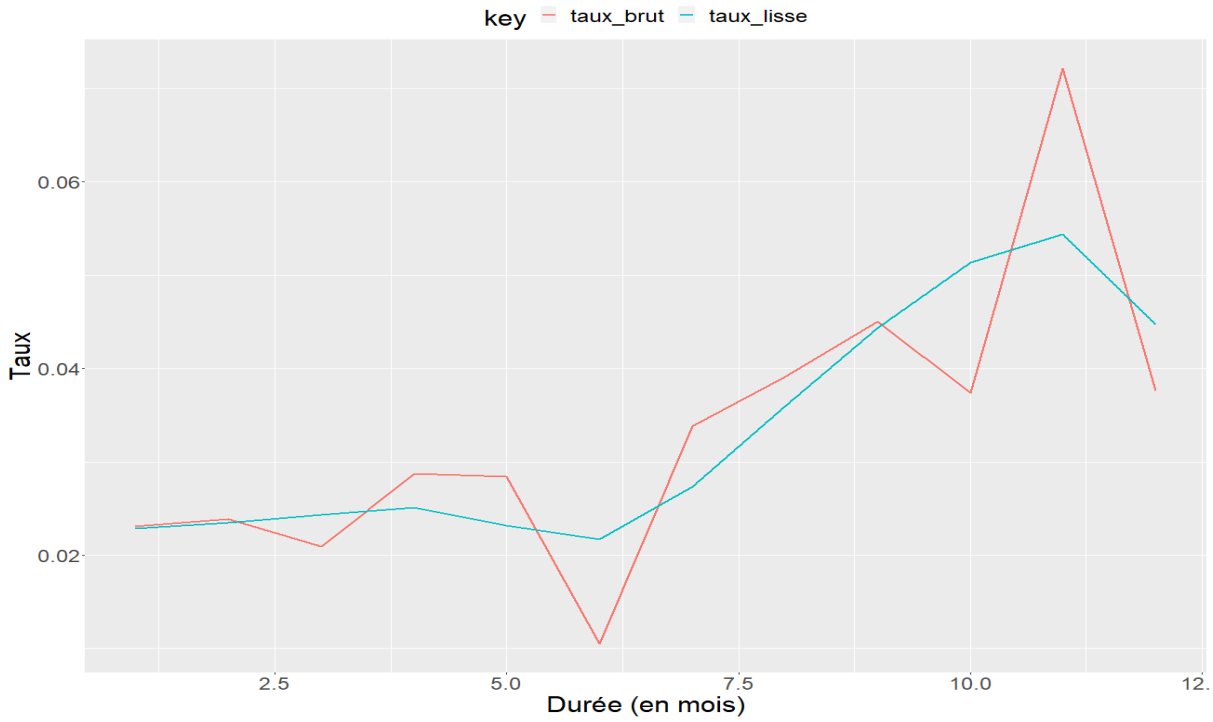


Figure 22: Comparaison des taux bruts et taux lissés retenus

De ces taux lissés, il est possible de reconstruire la fonction de survie lissée qui est représentée dans la figure ci-dessous. La fonction de survie lissée est plus pentue que celle brute.

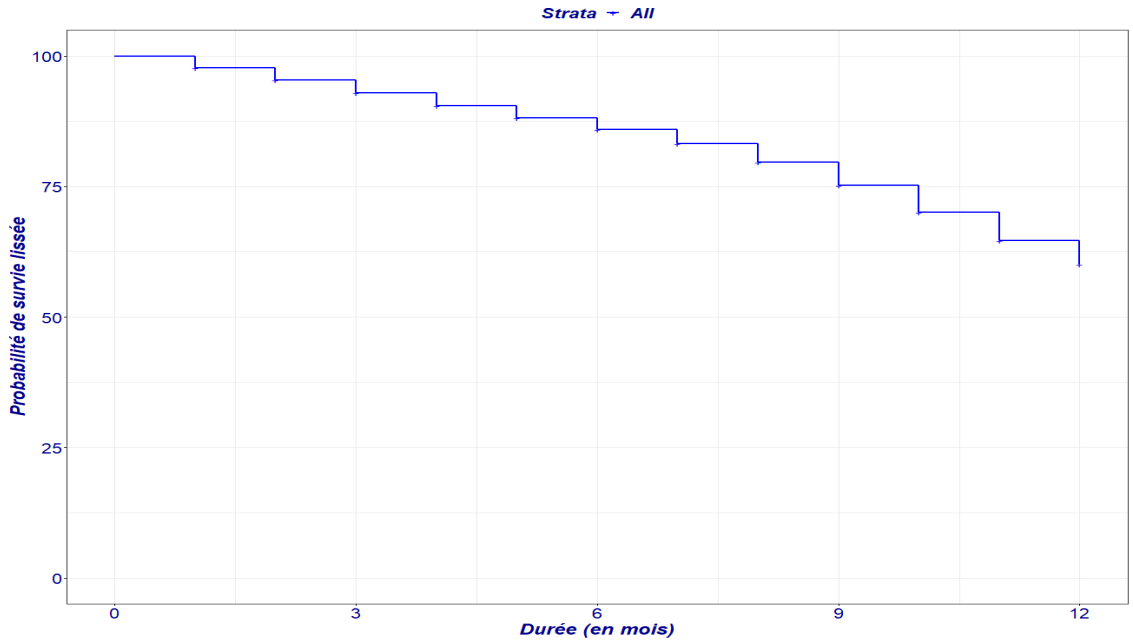


Figure 23: Fonction de survie issue des taux lissés retenus

In fine, on peut comparer l'espérance de maintien résiduelle[5] en perte d'emploi à l'entrée des taux bruts et celle des taux lissés retenus :

[5] L'espérance de maintien résiduelle à l'entrée est donnée par la somme des valeurs de la fonction de survie.

$$E_0(taux\ brut) = 11{,}05 \text{ mois et } E_0(taux\ lissés) = 11{,}06 \text{ mois.}$$

L'espérance de la loi lissées est légèrement supérieure à celle de la loi brute.

IV.2 Influence des covariables

Dans cette sous-section, nous étudions la significativité de l'influence de covariables à travers d'abord les tests de comparaison de fonctions de survie puis à travers le modèle de Cox. Nous étudierons également l'effet de la crise sanitaire sur les taux de sortie du chômage via un modèle de Cox.

IV.2.1 Tests de comparaison de fonctions de survie

Dans le tableau 11 sont consignés les résultats (statistique chi-deux et p-valeur) du test de LogRank pour chacune des covariables. En effet, le test rejette l'hypothèse nulle pour l'âge de survenance et l'ancienneté (p-valeur $> 0{,}05$) mais accepte l'hypothèse nulle pour le sexe, le montant initial et le montant d'échéance. En d'autres termes, les fonctions de survie associées aux sous-populations engendrées par les modalités de chacune de ces trois dernières variables sont identiques. Ainsi, le sexe, le montant initial du prêt et le montant des échéances du prêt n'ont pas d'effet sur les taux de sorties de perte d'emploi des assurés. En revanche, le rejet de l'hypothèse nulle pour l'âge lors de la survenance de la perte d'emploi signifie que la fonction de survie des assurés de moins 40 ans et de celle des assurés de plus de 40 ans ne sont pas les mêmes. Il en est de même pour les fonctions de survie des assurés avec une ancienneté inférieure à 18 mois, comprise entre 18 et 30 mois ou supérieure à 30 mois.

Variables	**LogRank**	
	Statistique chi-deux	**P-valeur**
Sexe	0	0,90
Age de survenance	10,30	0,00
Ancienneté	8,50	0,01
Montant initial	3,30	0,40
Montant échéance	6,90	0,07

Tableau 11: Résultat des tests LogRank de comparaison de fonctions de survie

Bien que les tests non paramétriques soient assez puissants pour évaluer l'impact des covariables, ils ne permettent pas de quantifier la différence entre les fonctions de survies des différentes sous-populations et ne prennent pas en compte les éventuelles interactions pouvant exister entre les différentes covariables. Afin de dépasser ces limites, nous allons utiliser le modèle de Cox dont les résultats sont présentés dans la section suivante.

IV.2.2 **Modèle de Cox**

Comme avec les tests de comparaison de fonctions de survie de sous-populations, nous utilisons les variables quantitatives regroupées en classe (excepté la variable sexe qui est déjà qualitative) afin de simplifier l'interprétation des résultats. Il convient de noter également que la régression de Cox est effectuée en utilisant l'ensemble des cinq variables simultanément. Dans un premier temps, nous montrerons les résultats initiaux produits par le modèle de Cox simple, puis nous montrerons, dans un second temps, les résultats du modèle de Cox étendu.

IV.2.2.1 ***Modèle de Cox de simple***

L'individu de référence du modèle (l'individu avec la fonction de risque h_0) est un assuré : de sexe féminin, âgé de moins de 40 ans, avec une ancienneté comprise entre 18 et 30 mois, contractant un prêt dont le montant initial est compris entre 14.000 et 22.000 euros et dont le montant de l'échéance est entre 158 et 248 euros.

Le modèle de Cox est globalement significatif puisque toutes les trois statistiques de significativité globale du modèle ont des p-valeurs inférieures à 5% (voir tableau 9).

Nom test	z	df	p
Likelihood ratio test	39,44	10	0
Wald test	38,52	10	0
Score test	38,55	10	0

Tableau 12: Tests de significativité globale du modèle de Cox simple

Le tableau 13 récapitule les valeurs des coefficients estimées et leur niveau de significativité pour chacune des modalités des covariables. On peut constater que toutes les variables sont significatives sauf le sexe. En effet, au moins une des modalités de chacune des covariables significatives a un coefficient significativement différent de zéro (lignes en gras sur la colonne p-valeur du tableau).

Pour le sexe, non seulement le coefficient associé à la modalité masculin est très proche de zéro (-0,01) mais sa p-valeur est supérieure à 5%.

L'âge lors de la survenance de la perte d'emploi a un effet négatif sur les taux de sortie de la perte d'emploi. En effet, les assurés de plus de 40 ans ont 0,69 fois moins de chance de trouver un emploi que ceux de moins de 40 ans. Ce résultat vient confirmer un fait constaté en statistique de l'emploi c'est-à-dire la durée passée au chômage augmente avec l'âge.

L'ancienneté du contrat de l'assuré au moment de la perte d'emploi a une influence positive sur le risque de sortie de la perte d'emploi. Comparés aux assurés avec un contrat âgé entre 18 et

30 mois, les assurés de contrat de moins de 18 mois d'ancienneté ont 0,69 fois moins de chance de sortir du chômage. Tandis que les assurés avec une ancienneté de plus de 30 mois, ont plus 1,14 fois plus de chance de sortir de la perte d'emploi que la même population de référence. Il faut toutefois noter que le coefficient associé à cette modalité n'est pas statistiquement significatif.

Quant au montant initial du prêt, il semble affecter négativement les taux de sortie de la situation de perte d'emploi. Autrement dit, plus le montant initial du prêt est élevé, plus l'assuré a tendance à durer dans sa situation de chômage. Cela s'illustre par le fait que les assurés des deux classes « moins de 9.000 » et « Entre 9.000 et 14.000 » situées avant la classe de référence (« Entre 14.000 et 22.000 ») ont respectivement 1,55 et 1,66 fois plus de chance de sortir de la perte d'emploi que ceux de la classe de référence. En revanche, les assurés de la classe au-dessus de la classe de référence (« Plus de 22.000 ») ont 34% moins de chance de trouver un nouvel emploi, même si le coefficient n'est pas significativement différent de zéro au seuil de 5%.

Contrairement au montant initial du prêt, le montant des échéances du prêt a un effet positif sur les chances de sortir de la perte d'emploi. En effet, les assurés des deux classes « plus de 386 » et « Entre 248 et 386 » situées après la classe de référence « Entre 158 et 248 » ont respectivement 1,79 et 1,26 fois plus de chance de sortir de la perte d'emploi que ceux de la classe de référence. En revanche, les assurés de la classe en-dessous de la classe de référence « moins de 158 » ont 37% moins de chance de trouver un nouvel emploi. Il convient de noter aussi que le coefficient associé à la classe « entre 248 et 386 » n'est pas significatif à 5%.

Modalités	coef	exp(coef)	se(coef)	z	P-valeur
Sexe : Masculin	-0,01	0,99	0,11	-0,12	0,90
Age survenance : Plus de 40 ans	-0,38	0,69	0,11	-3,48	**0,00**
Ancienneté : Moins de 18 mois	-0,37	0,69	0,13	-2,79	**0,01**
Ancienneté : Plus de 30 mois	0,13	1,14	0,13	0,97	0,33
Montant initial : Entre 9000 et 14000	0,44	1,55	0,18	2,41	**0,02**
Montant initial : Moins de 9000	0,51	1,66	0,23	2,26	**0,02**
Montant initial : Plus de 22000	-0,27	0,76	0,21	-1,28	0,20
Montant échéance : Entre 248 et 386	0,23	1,26	0,18	1,28	0,20
Montant échéance : Moins de 158	-0,46	0,63	0,19	-2,44	**0,01**
Montant échéance : Plus de 386	0,58	1,79	0,25	2,31	**0,02**

Tableau 13 : Tests de significativité des coefficients des variables du modèle de Cox simple

Les résultats du test de l'hypothèse de risques proportionnels pour chacune des covariables sont résumés dans le tableau 14. Pour rappel, l'hypothèse nulle, dans ces tests, est la proportionnalité

des risques. Sur cette base, toutes les variables respectent donc l'hypothèse de risque proportionnels à l'exception de l'ancienneté. Autrement dit, afin de bien mesurer l'influence de l'ancienneté sur les taux de sorties, il faudra utiliser un modèle de Cox étendu. Toutefois, nous allons diagnostiquer les résidus de Schoenfeld (surtout la régression LOESS issue de ce nuage de points) pour avoir une idée sur la spécification la plus adéquate à cette situation.

Variables	chisq	df	p
Sexe	0,05	1,00	0,82*
Age de survenance	0,44	1,00	0,51*
Ancienneté	6,31	2,00	**0,04**
Montant initial	4,55	3,00	0,21*
Montant échéance	5,73	3,00	0,13*
GLOBAL	16,53	10,00	0,09*

Tableau 14 : Tests de l'hypothèse de risques proportionnels du modèle de Cox simple

La figure 24 combine les graphiques représentant les résidus de Schoenfeld (nuage de points), la courbe issue de la régression LOESS (courbe en noir avec son intervalle de confiance) et le coefficient β associé (droite horizontale en pointillée rouge) pour chacune des cinq covariables.

L'analyse de la significativité des covariables et l'analyse du respect de l'hypothèse de risques proportionnels sont effectuées plus aisément via la courbe issue de la régression de LOESS. En effet, puisque cette courbe est vue comme la courbe d'évolution du coefficient $\beta(t)$ en fonction de la durée avant sortie de la perte d'emploi, alors elle fournit beaucoup d'informations visuelles sur la significativité et la proportionnalité des risques. Pour une variable dont le coefficient est significativement nul, comme le sexe, la valeur zéro se situe toujours dans l'intervalle de confiance de la courbe issue de la régression de LOESS.

Par ailleurs, pour que l'hypothèse de risques proportionnels soit vérifiée, il faut que la courbe issue de la régression de LOESS soit plus ou moins constante dans le temps (aussi horizontale que possible) et on peut clairement constater que tel n'est pas le cas pour l'ancienneté. On constate sur le graphique que l'actuel coefficient estimé n'est pas comme une sorte de moyenne des valeurs de la courbe LOESS. L'évolution de la courbe LOESS met en évidence deux dynamiques sur la valeur du coefficient β associé à l'ancienneté (la première pour les durées inférieure à 7 et la deuxième pour les durées supérieures à 7).

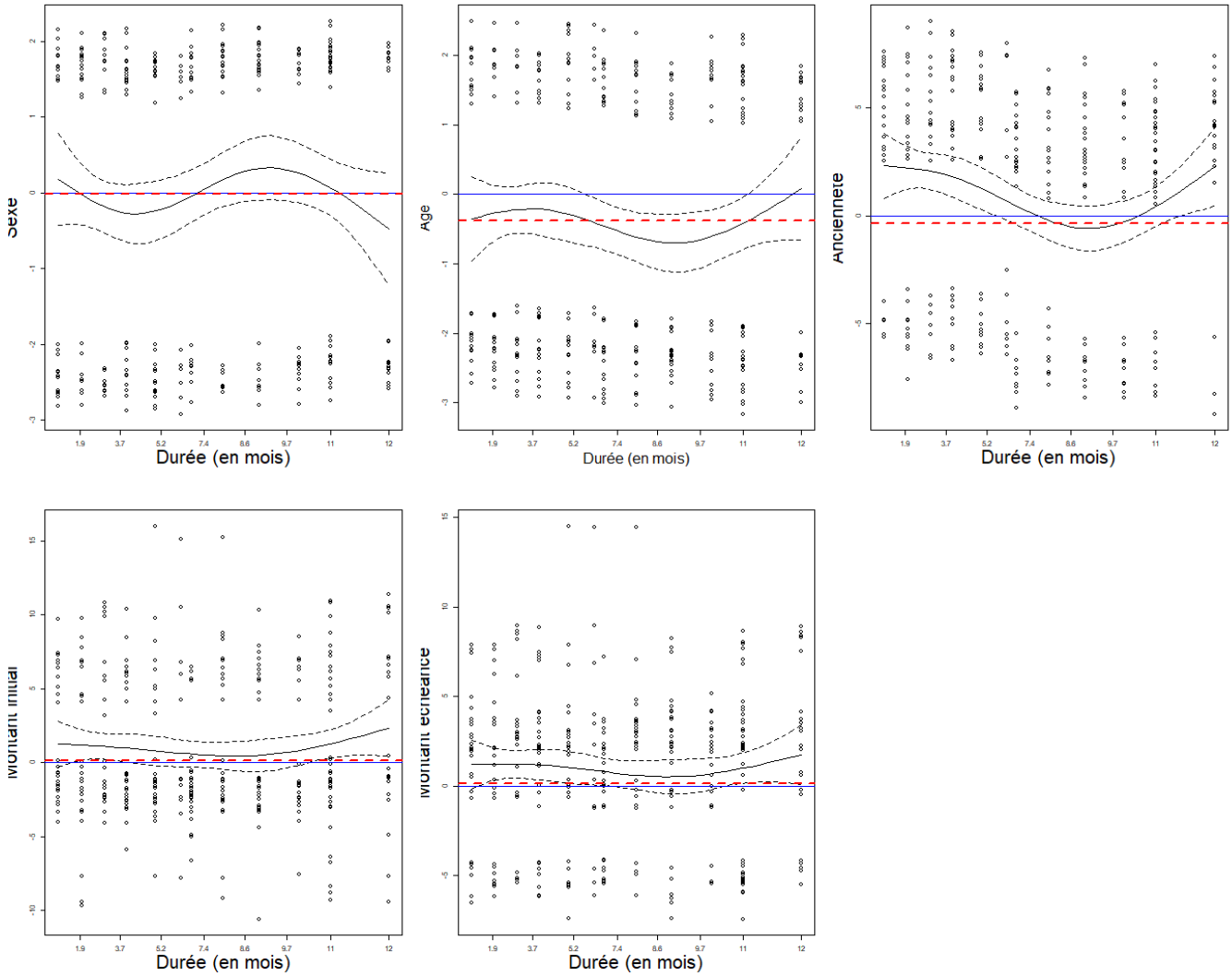


Figure 24:Evolution dans le temps du coefficient beta associé à chaque variable

IV.2.2.2 ***Modèle de Cox étendu***

La partie la plus difficile dans l'estimation d'un modèle de Cox étendu réside dans la spécification de la forme de la fonction G. Dans ce cas-ci, G est une fonction définie par morceaux, pour $t' = 7$:

$$G(t_i) = (\mathbb{I}_{(t_i<t')}, \mathbb{I}_{(t_i\geq t')})$$

En d'autres termes, pour chacune des deux modalités, deux coefficients β seront estimés : un coefficient pour la période avant le mois 7 et un autre après.

A l'image du modèle de Cox simple, le modèle de Cox étendu est globalement significatif puisque toutes les trois statistiques de significativité globale du modèle ont des p-valeurs inférieures à 5% (voir tableau 15).

Nom test	z	df	p-valeur
Likelihood ratio test	35,87	10	0
Wald test	35,29	10	0
Score test	35,24	10	0

Tableau 15 : Tests de significativité globale du modèle de Cox étendu

D'après le tableau récapitulatif des coefficients et des tests de significativité des coefficients des variables (tableau 16), toutes les variables sont significatives sauf le sexe comme avec le modèle de Cox simple. Le seul changement à noter est que l'ancienneté n'est plus significative à 5% mais plutôt à 10%. Par ailleurs, le sens et l'amplitude des effets des covariables sur les taux de sorties n'ont pas connu de fortes modifications.

Modalités	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Sexe : Masculin	-0,01	0,99	0,11	-0,12	0,90
Age survenance : Plus de 40 ans	-0,38	0,68	0,11	-3,51	**0,00**
Montant initial : Entre 9000 et 14000	0,44	1,55	0,18	2,41	**0,02**
Montant initial : Moins de 9000	0,53	1,71	0,23	2,35	**0,02**
Montant initial : Plus de 22000	-0,26	0,77	0,21	-1,26	0,21
Montant échéance : Entre 248 et 386	0,25	1,28	0,18	1,37	0,17
Montant échéance : Moins de 158	-0,47	0,62	0,19	-2,52	**0,01**
Montant échéance : Plus de 386	0,58	1,79	0,25	2,30	**0,02**
Ancienneté : Entre 18 et 30 mois : t ≤ 7	-0,16	0,85	0,18	-0,88	0,38
Ancienneté : Moins de 18 mois : t ≤7	-0,75	0,47	0,18	-4,12	**0,00**
Ancienneté : Plus de 30 mois : t ≤ 7	NA	NA	0,00	NA	NA
Ancienneté : Entre 18 et 30 mois : t > 7	-0,09	0,91	0,19	-0,48	0,63
Ancienneté : Moins de 18 mois : t > 7	-0,15	0,86	0,20	-0,72	0,47
Ancienneté : Plus de 30 mois : t > 7	NA	NA	0,00	NA	NA

Tableau 16 : Tests de significativité des coefficients des variables du modèle de Cox étendu

En ce qui concerne les tests de l'hypothèse de risques proportionnels, l'hypothèse nulle, c'est-à-dire la proportionnalité des risques est encore acceptée pour toutes les variables à l'exception de l'ancienneté. Le modèle de Cox étendu n'a, non seulement, pas résolu le problème de violation de l'hypothèse de risques proportionnels mais a entraîné une violation de l'hypothèse de proportionnalité des risques pour l'ensemble du modèle (cf. la dernière ligne du tableau 17).

Variables	chisq	df	p-valeur
Sexe	0,000	1	0,98*
Age de survenance	0,043	1	0,84*
Montant initial	4,810	3	0,19*
Montant échéance	5,590	3	0,13*
Ancienneté : strata(tgroup)	12,200	4	**0,02**
GLOBAL	22,700	12	**0,03**

Tableau 17 : Tests de l'hypothèse de risques proportionnels du modèle de Cox étendu

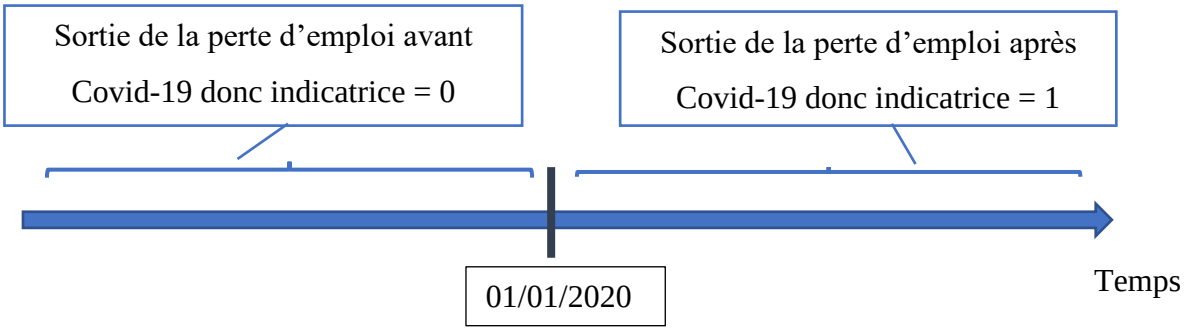
IV.2.3 Effet crise sanitaire sur les taux de sortie

La table de maintien retenue a pour vocation à être utilisée dans le calcul des provisions de la garantie perte d'emploi du produit Expresso. Toutefois, depuis 2020, la crise sanitaire a beaucoup touché l'économie particulièrement le chômage. En combinant le ralentissement de l'activité économique et toutes les mesures sanitaires mise en œuvre pour limiter la propagation de la pandémie de Covid-19, il est logique de penser que le temps passé au chômage devrait s'allonger. Puisqu'il n'y a pas de statistiques publiques permettant de mesurer directement le temps passé au chômage, il est difficile de quantifier l'effet du Covid-19 sur l'allongement de la durée en perte d'emploi. Une certitude est que le taux de chômage a beaucoup fluctué au cours de l'année 2020 (mais sa moyenne annuelle est inférieure à celui de 2019) tandis que la part du chômage partiel a connu de fortes hausses (Yves Jauneau, Joëlle Vidalenc (Insee), 2021).

Ainsi, avant d'appliquer la table dont les taux ont été obtenus avec des données de la période avant 2020, il est important de réajuster les taux en tenant compte de la crise sanitaire. Pour cela, nous allons utiliser le modèle de Cox, afin d'estimer le coefficient d'ajustement.

IV.2.3.1 *Principe de modélisation*

L'étude de l'effet du Covid-19 sur les taux de sortie se fait à travers une variable indicatrice qui vaut 1 si la sortie est survenue après le premier janvier 2020 et 0 sinon. Cette indicatrice constituera l'unique variable explicative du modèle de Cox et l'exponentiel du coefficient qui lui est associé pourra être considéré comme le coefficient d'ajustement des taux de sortie.



Puisque la variable indicatrice Covid-19 dépend du temps, nous appliquerons l'approche du modèle de Cox étendu spécifié comme suit :

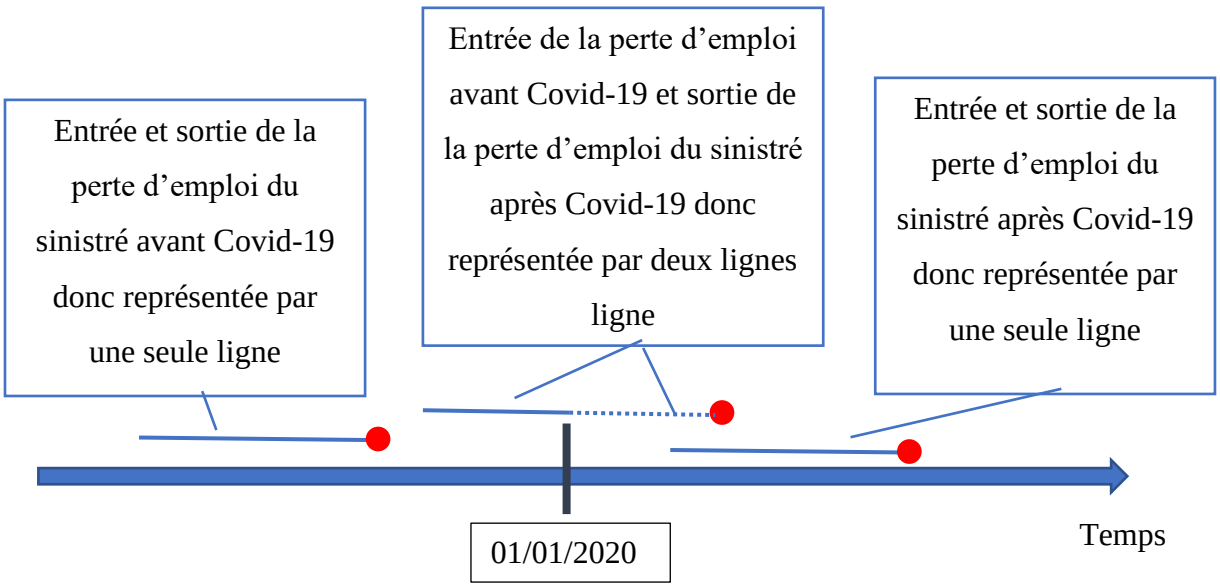
$$h(t_i) = h_0(t_i)\exp(\boldsymbol{I_{Covid}}(t_i)\beta_{Covid})$$

La situation de référence est celle d'avant la crise et donc les taux de sortie d'avant 2020.

Avec cette approche, nous devront reformater la base de données des sinistres et y intégrer les sorties de perte d'emploi de 2020 et 2021.

IV.2.3.2 *Reformatage de la base de données*

Nous avons étendu la période d'observation jusqu'en juin 2021 et avons appliqués des traitements similaires à ceux de la base précédente. Toutefois, pour cette base au lieu d'avoir à chaque fois une ligne par sinistre, on peut avoir jusqu'à deux lignes selon que la sortie de perte d'emploi soit survenu avant ou après 2020. En d'autres termes, tous les sinistres pour lesquels l'assuré est sortie de la perte d'emploi (soit par vraie sortie ou par censure) avant janvier 2020 seront représentés par une seule ligne dans la base et 2 lignes sinon. Il faut noter également que les assurés entrés en perte d'emploi après 2020 sont également représentés par une seule ligne.



Pour les sinistres avec deux lignes, au niveau de la première ligne, nous avons la durée du sinistre jusqu'au 31 décembre 2019, la variable censure qui vaut 1 et la variable indicatrice Covid-19 qui vaut 0. En revanche, pour la deuxième ligne, nous avons la durée du chômage est égale à la durée de la première ligne plus la durée du sinistre après janvier 2020 et l'indicatrice Covid-19 vaut 1. La figure ci-dessous illustre bien les cas décrits précédemment.

ID_INDIVIDU	ID_PRET	DATE_PREM_ECH	DATE_FIN_SIN	Duree_PE	censure	covid
24598947	35197795129	10/09/2019	30/12/2019	6	1	0_avant
24598947	35197795129	10/09/2019	10/08/2020	12	1	1_après
26209261	34197502072	03/03/2013	03/03/2013	6	0	0_avant
27206076	34199678755	20/04/2019	20/01/2020	10	1	1_après

Figure 25: Illustration du reformatage de la nouvelle base de données

En résumé, la nouvelle base de données compte 1.919 lignes.

IV.2.3.3 ***Résultats***

Le modèle de Cox mesurant l'effet du Covid-19 est globalement significatif. Les trois statistiques de significativité globale ainsi que la statistique de Student du coefficient de la variable ont des p-valeurs inférieures à 5% (voir tableau 18 et 19).

Modalités	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Après Covid-19	-0,86	0,42	0,20	-4,33	0,00

Tableau 18: Tests de significativité du coefficient de la variable du modèle de Cox de l'effet du Covid-19

Test	z	df	p
Likelihood ratio test	24,24	1	0,00
Wald test	198,78	1	0,00
Score test	20,23	1	0,00

Tableau 19: Tests de significativité globale du modèle de Cox de l'effet du Covid-19

D'après ce modèle, les taux de sortie du chômage après la crise sanitaire valent 0,42 fois les taux de sortie après le début de la crise. Autrement dit, on observe un allongement de la durée du chômage pendant la crise, ce qui est logique vu le ralentissement de l'activité économique.

En revanche, le manque de recul sur le phénomène empêche la possibilité de bien challenger la valeur du coefficient estimé bien que son signe semble cohérent. Le seul moyen de vérifier la vraisemblance du coefficient estimé par ce modèle est de faire un backtesting en 2020 en utilisant les taux de sortie réajusté de la table de maintien retenue.

Par ailleurs, l'hypothèse de risques proportionnels est bien vérifiée avec ce modèle de Cox étendu.

Variables	chisq	df	p
Covid-19	0,083	1	0,82
GLOBAL	0,083	1	0,82

Tableau 20: Tests de l'hypothèse de risques proportionnels du modèle de Cox de l'effet du Covid-19

Pour vérifier la significativité de la différence entre les taux de sortie avant et après crise sanitaire, nous avons estimé les fonctions de survie (loi de maintien en perte d'emploi) avant et après le Covid-19. Force est de constater qu'elles sont distinctes et que la fonction de survie

après Covid-19 est toujours au-dessus de celle d'avant Covid-19. Cela traduit le fait qu'après 2020, on met plus de temps à de trouver un emploi qu'avant. Il faut toutefois faire attention puisqu'il y a moins de données disponibles pour la période après Covid-19, par conséquent, ces estimations moins robustes et moins fiables.

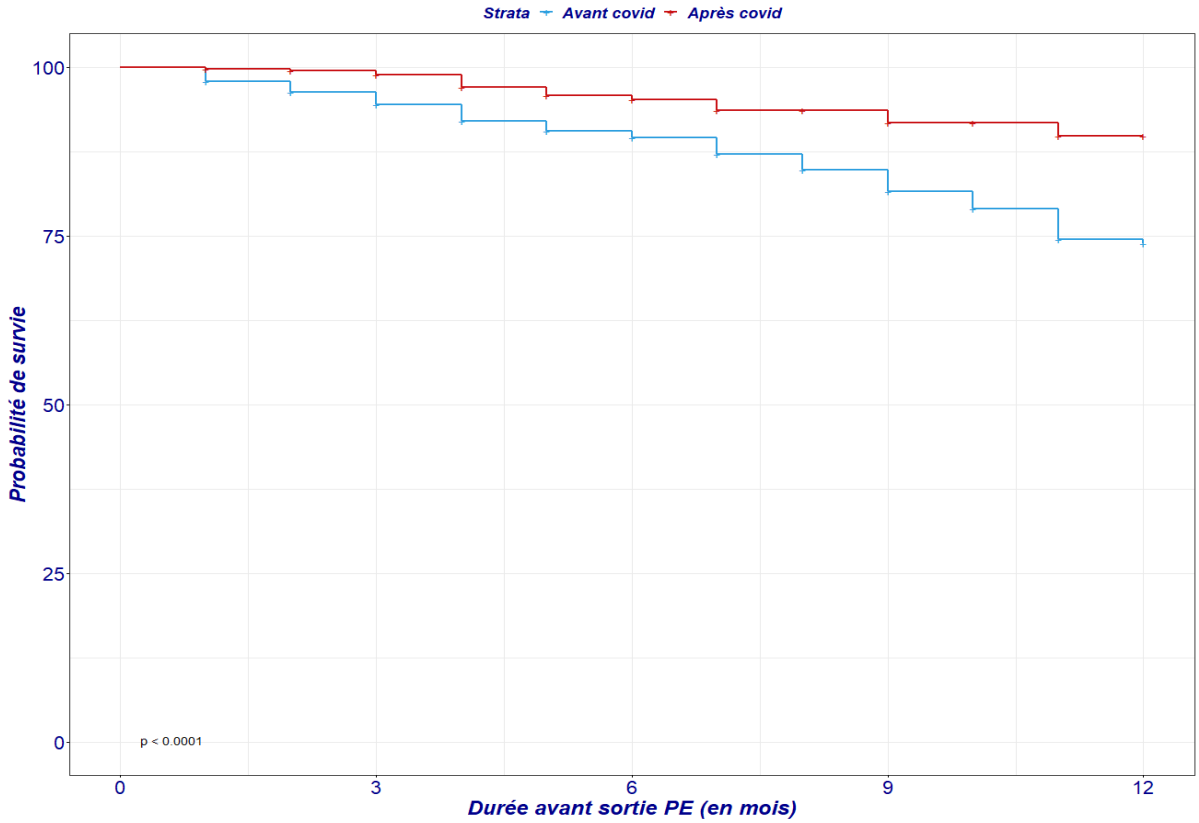


Figure 26: Comparaison des fonctions de survie avant et après début du Covid-19

Comme évoqué plus haut, nous allons tester les estimations du modèle de Cox à travers un backtesting en 2020. En effet, pour cela nous utiliserons les taux de sortie de la table retenue multipliés par le coefficient d'ajustement (exp ($\hat{\beta}$)) pour prédire le nombre de sortie en 2020 que l'on comparera avec les sorties effectivement observées.

La mise en œuvre de cette approche conduit à un nombre de sorties du chômage prédit (26) assez proche du nombre observés (23) (cf. tableau 21). Toutefois, ces résultats manquent de fiabilité dans la mesure où les données de 2020 ont servi à calculer le coefficient $\hat{\beta}$. C'est comme si on testait le pouvoir prédictif d'un modèle sur les données qui ont permis à son calibrage. Les résultats fournis par ces tests ont tendance à être trop optimiste. Pour avoir plus de fiabilité, il faudra réaliser le backtesting sur une autre période, 2022 par exemple.

Durée	Exposition	Sorties observées (1)	Sorties prédites par taux bruts	Sorties prédites taux lissés ajustés (2)	Borne inférieure	Borne supérieure	Ratio ((1)-(2))/(1)
1	237	1	2	2	0,462	4,351	1,000
2	226	1	3	3	0,760	5,006	2,000
3	219	1	3	3	0,546	4,540	2,000
4	210	4	2	2	0,302	3,973	-0,500
5	198	3	2	2	0,231	3,788	-0,333
6	180	1	2	2	0,265	3,860	1,000
7	144	4	2	2	0,185	3,646	-0,500
8	125	0	2	2	0,185	3,625	-
9	115	4	2	2	0,256	3,797	-0,500
10	104	0	2	2	0,344	3,986	-
11	94	4	3	2	0,395	4,087	-0,500
12	83	0	2	2	0,231	3,688	-
Total		**23**	**27**	**26**	4,162	48,346	0,130

Tableau 21: Comparaison des sorties observées et prédites par la table retenue après ajustement en 2020

Vu que l'effet du Covid-19 n'est pas considéré comme fiable pour le moment, le coefficient d'ajustement n'est pas utilisé dans le calcul des provisions lors de l'utilisation de la table. De ce fait, une provision additionnelle est ajoutée au best estimate des provisions donné par l'utilisation de la table. Tel est l'objet de la partie suivante.

Partie V. Utilisation de la table de loi de maintien dans le cadre du calcul des provisions

Dans cette partie, il sera question de montrer, dans un premier temps, comment est utilisée la table de loi de maintien en prenant en compte l'effet du Covid-19 pour le calcul des provisions, puis de mettre en exergue l'impact de son utilisation pour le calcul des provisions dans un second temps.

V.1 Application de la table dans le calcul des provisions

Les provisions sont calculées tête-par-tête à l'aide d'un programme sous le logiciel SAS. Nous allons voir d'abord le principe général de calcul des provisions puis nous l'illustrerons à l'aide d'un exemple.

Nous disposons de la base sinistre d'une part et de la table de maintien en perte d'emploi d'autre part. La base sinistre contient tous les assurés au chômage au moment de l'inventaire avec les informations comme l'âge, le temps passé au chômage (appelé ancienneté en perte d'emploi), CRD (Capital Restant Dû), le montant de l'échéance mensuelle, le nombre maximum d'échéances restant, etc. Il convient de souligner que le nombre maximum d'échéances restant est calculé en se basant sur l'ancienneté de l'assuré en perte d'emploi et sur les termes du contrat d'Expresso comme l'atteinte d'âge limite, l'atteinte du nombre limite de prestations auquel peut bénéficier l'assuré, la nullité du capital restant dû, etc.

La table de maintien est jointe à la base sinistre ce qui permet de calculer le taux de maintien en perte d'emploi pour chaque sinistré. Ce dernier est la somme de deux taux de maintien définis comme suit :

$$Taux\ de\ maitntien\ 1 = \sum_{i=\lfloor a \rfloor+1}^{K} (\lfloor a \rfloor + 1 - a) \times \frac{S(i)}{S(\lfloor a \rfloor)}$$

$$Taux\ de\ maitntien\ 2 = \sum_{i=\lfloor a \rfloor+2}^{K} (a - \lfloor a \rfloor) \times \frac{S(i)}{S(\lfloor a \rfloor + 1)}$$

K : nombre maximum d'échéances restant

a : ancienneté en perte d'emploi

$\lfloor a \rfloor$: partie entière de l'ancienneté en perte d'emploi

$S(i)$: fonction de survie en i

Ces deux taux permettent de prendre en compte le fait que l'ancienneté en perte d'emploi d'un assuré ne soit pas forcément un nombre entier en mois lors de l'inventaire. Par exemple, lorsqu'un assuré a 10 mois d'ancienneté en perte d'emploi au moment de l'inventaire, on voit que seule la première formule est utilisée (car le taux maintien 2 vaut zéro) et le taux de maintien en perte d'emploi devient directement la somme des probabilités de maintien ($\frac{S(11)}{S(\lfloor 10 \rfloor)} + \frac{S(12)}{S(\lfloor 10 \rfloor)}$). Cependant, quand l'ancienneté n'est pas entière et vaut, par exemple 9,6 le taux de maintien 1 devient $(9 + 1 - 9{,}6) \times \left[\frac{S(\lfloor 10 \rfloor)}{S(\lfloor 9 \rfloor)} + \frac{S(\lfloor 11 \rfloor)}{S(\lfloor 9 \rfloor)} + \frac{S(\lfloor 12 \rfloor)}{S(\lfloor 9 \rfloor)}\right]$ et le taux de maintien 2 est $(9{,}6 - 9) \times \left[\frac{S(\lfloor 11 \rfloor)}{S(\lfloor 10 \rfloor)} + \frac{S(\lfloor 12 \rfloor)}{S(\lfloor 10 \rfloor)}\right]$.

Ainsi, puisque le taux de prise en charge des échéances de l'assuré par l'assureur est de 100%, le montant de la provision par individu est donné par :

$$Provision = Taux\ de\ maintien \times montant\ échéance\ mensuelle$$

Avec $Taux\ de\ maintien = Taux\ de\ maintien\ 1 + Taux\ de\ maintien\ 2$.

Numéro assuré	Age	CRD	Ancienneté en PE au moment de l'inventaire	Nombre maximum d'échéances restant	Montant échéance	Taux de maintien 1	Taux de maintient 2	Taux de maintien	Provision
1	45	10000	0,4	11,6	290	6,035	3,718	9,753	2828,269
2	21	4000	5	7	154	6,093	0,000	6,720	1034,902
3	**65**	15000	4,5	0	305	0,000	0,000	0,000	0,000
4	30	3126	**12**	0	174	0,000	0,000	0,000	0,000
5	44	7651	2,7	9,3	210	2,555	5,411	8,965	1882,566
6	32	**0**	9,1	0	0	0,000	0,000	0,000	0,000
7	51	3561	**12**	0	317	0,000	0,000	0,000	0,000

Tableau 22: Exemple d'application de calcul des provisions par sinistré

Dans le tableau ci-dessus, nous illustrons le calcul du taux de maintien avec des données fictives. Il convient de bien noter l'impact de la prise en compte des termes du contrat (âge limite pour bénéficier de prestations, CRD nul et nombre maximum de prestations atteints) sur le nombre maximum d'échéances restant et par conséquent sur le montant de provision.

Par ailleurs, à cause de la crise sanitaire, un montant est ajouté à la provision best estimate par mesure de précaution. En effet, un S/P (Sinistre sur Prime) additionnel, intégrant la hausse du nombre de pertes d'emploi et l'allongement de la durée de perte d'emploi, multiplié par les primes donne le montant additionnel. Par exemple, en fixant un S/P à 3%, un montant des

primes à 100.000 euros et en reprenant l'exemple d'application précédent, on obtient un montant additionnel de 3.000 euros et une provision totale de 8.746,738 euros.

V.2 Impact de l'application de la table sur les provisions du mois de mars 2021

L'étude de l'impact de la table de loi de maintien sur les provisions est réalisée en comparant le niveau des provisions du mois de mars 2021 avant et après application de la table de loi.

Avant application de la loi de maintien, les provisions étaient calculées comme suit :

$$\boldsymbol{provision = montant\ de\ l'échéance \times (12 - ancienneté\ en\ perte\ d'emploi)}$$

Cette méthode a l'avantage d'être très prudente mais conduit à un sur-provisionnement et ne permet pas d'avoir le *best estimate*.

Ainsi, pour le mois de mars 2021, les provisions baissent de 12,8% en passant de la méthode de calcul de provisions n'utilisant pas de tables à celle utilisant la table retenue.

Conclusion

Au terme de ce mémoire, nous avons présenté les étapes de construction d'une table de maintien en perte d'emploi pour un produit d'assurance emprunteur.

Nous avons posé le cadre de l'étude en revenant sur certaines notions importantes telles que l'assurance emprunteur, les contrats groupes, la garantie perte d'emploi et ses caractéristiques et en présentant le produit Expresso.

La création et la fiabilisation de la base de données, étape primordiale de l'étude, a été présentée. Sur la base de données initiale de sinistres, nous avons effectué des traitements pour consolider en une seule ligne chaque sinistre et avons créé des variables supplémentaires comme l'indicatrice de censure ainsi que d'autres covariables. Ces travaux ont fourni une base de données avec 1.379 lignes portant sur les sorties de perte d'emploi réalisées entre 01/01/2012 et 31/12/2018. Pour mieux cerner le profil de risque du portefeuille, nous avons réalisé des statistiques descriptives.

Nous avons appliqué au portefeuille d'Expresso la méthode de Kaplan-Meier pour l'estimation des taux bruts et la méthode de Whittaker-Henderson pour le lissage des taux bruts. Pour le choix des paramètres de lissage et par conséquent de la courbe de taux lissés finale, nous avons retenu la courbe suffisamment lisse qui prédisait les sorties de perte d'emploi les plus proches de celles observées sur la période du 01/01/2019 au 31/12/2019.

En outre, nous avons également étudié l'effet des covariables à travers le test du LogRank et les modèles de Cox (simple et étendu). Selon le test du LogRank, le sexe, le montant initial du prêt et le montant des échéances du prêt n'affectent pas les probabilités de sortie du chômage mais l'âge et l'ancienneté lors de la survenance du sinistre l'influencent. En revanche, le modèle de Cox révèle que toutes ces covariables, à l'exception du sexe, influencent significativement les chances de sortie de la perte d'emploi. Toutefois ce résultat doit être pris avec précaution puisque l'hypothèse fondamentale de risques proportionnels dans le modèle de Cox simple n'est pas vérifiée pour la variable ancienneté. Le passage à un modèle de Cox étendu visant à estimer deux coefficients pour chaque modalité de la variable ancienneté regroupée en classe n'a pas réussi à résoudre le problème. Par ailleurs, afin de prendre en compte l'effet de la crise sanitaire survenue depuis 2020 sur la durée de la perte d'emploi, nous avons calibré un autre modèle de Cox étendu. Selon ce dernier, les taux de sortie de la perte d'emploi ont diminué de 0,42 fois

depuis le début de la crise mais la fiabilité de ce coefficient reste à valider avec un backtesting à réaliser en 2022.

Enfin, nous avons étudié l'impact de l'utilisation de la table de maintien sur les provisions en mars 2021 après avoir montré comment la table est utilisée en pratique pour calculer les provisions de la garantie perte d'emploi. L'utilisation de celle-ci a fait baisser les provisions de 12,8%.

Il est toutefois très important de garder à l'esprit les hypothèses faites lors des traitements des données (identification des sinistres et conditions de censures). Par ailleurs, l'insuffisance des observations dans la base de données utilisée peut causer des instabilités de la table dans le temps et par conséquent la table doit toujours faire l'objet de suivi surtout lorsqu'on observe des changements sur la composition du portefeuille.

Bibliographie

ALLISON, P. D. (2010). *Survival Analysis Using SAS : A Practical Guide, Second Edition.* SAS Institute.

CLEVELAND, W. S. (1974). Robust locally weighted regression and smoothing scatterplot. *Journal of the American Statistical Association 74(368)*, 829-836.

COLLETAZ, G. (2020, Septembre 29). Econométrie des durrées de survie Note de Cours Master 2 ESA voies professionnelle et recherche. *Econométrie des durrées de survie*.

GRAMBSCH, P., & THERNEAU, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika 81(3)*, 515-526.

HARELL, F. E. (2006). *Regression Modeling Strategies : With Applications to LinearModels, Logistic Regression, and Survival Analysis.* New York: Springer-Verlag.

MARCHAL, T. (2011, 10 17). Construction de tables d'entrée et de maintien en chômage sur des contrats collectifs emprunteurs rattachés à des crédits à la consommation. Lyon: Mémoire ISFA.

MOORE, D. F. (2016). *Applied Survival Analysis Using R.* Switzerland: Springer International Publishing.

THOMAS, L., & REYES, E. M. (2014). *Journal of Statistical Software*, Volume 61, Code Snippet 1.

TOMAS, J., & PLANCHET, F. (2016). Note de travail (II1291-14 v1.4); Critère de validation : Aspects méthodologiques. *Critère de validation : Aspects méthodologiques*. France: Institut des actuaires.

Yves Jauneau, Joëlle Vidalenc (Insee). (2021). Une photographie du marché du travail en 2020. *INSEE PREMIERE*.

Note de synthèse

Dans un contexte réglementaire en pleine mutation, avec la directive solvabilité II et la norme IFRS 17 qui entrera en vigueur à partir du 1er janvier 2023, les assureurs sont plus que jamais contraints à calculer le best estimate de leurs provisions. En effet, ces nouvelles normes encouragent les assureurs à évaluer les risques en se basant sur des hypothèses stochastiques.

Dans des garanties comme le décès, l'incapacité ou l'invalidité présentes dans les contrats d'assurance emprunteur, cela se traduit par l'utilisation de tables réglementaires ou tables d'expérience du portefeuille. Contrairement aux autres garanties, la garantie perte d'emploi ne dispose pas de tables réglementaires (ni en entrée, ni en maintien de perte d'emploi). Ainsi, le pôle Valeur de l'équipe Actuariat Prévoyance de la Société Générale Assurance, est obligé de calculer des provisions extrêmement prudentes pour les garanties perte d'emploi par manque de table de maintien. Ainsi, afin de répondre aux exigences réglementaires avec le calcul du best estimate et d'éviter de sur-provisionner le risque de perte d'emploi, il a été décidé de mettre en place un ensemble d'études visant à construire des tables d'expérience de maintien sur la perte d'emploi.

Ce mémoire a pour but de retracer les différentes étapes de construction d'une table de maintien en perte d'emploi sur le portefeuille du produit Expresso, produit d'assurance emprunteur de la Société Générale Assurance, ainsi que l'impact de son utilisation en provisionnement.

Le travail a été divisé en cinq parties. Dans la première, il a été question de poser le cadre général de l'étude en revenant sur certaines notions importantes telles que l'assurance emprunteur, les contrats groupes, la garantie perte d'emploi et ses caractéristiques. Nous avons également présenté le produit Expresso.

La deuxième partie s'est focalisée sur la création et la fiabilisation de la base de données ainsi que la réalisation de statistiques descriptives sur le portefeuille. Sur la base de données initiale de sinistres, nous avons effectué des traitements pour consolider en une seule ligne chaque sinistre et nous avons créé des variables supplémentaires comme l'indicatrice de censure ainsi que d'autres covariables. Ces travaux ont fourni une base de données finale avec 1.379 lignes portant sur les sorties de perte d'emploi réalisées entre 01/01/2012 et 31/12/2018.

Après avoir présenté théoriquement la méthode de Kaplan-Meier pour l'estimation des taux bruts et la méthode de Whittaker-Henderson pour le lissage des taux bruts, dans la troisième partie, nous les avons appliquées au portefeuille d'Expresso dans la quatrième partie. Pour le choix des paramètres de lissage et par conséquent de la courbe de taux lissés finale, nous avons retenu la courbe suffisamment lisse qui prédisait les sorties de perte d'emploi les plus proches de celles observées sur la période du 01/01/2019 au 31/12/2019. Ci-dessous est présenté le tableau résumant les probabilités de survie (fonction de survie) et les taux de sortie de perte d'emploi de la table retenue.

Durée avant sortie de PE (en mois)	Probabilité de survie de PE (en %)	Taux de sortie de PE (en %)
0	100,00	2,29
1	97,71	2,35
2	95,42	2,43
3	93,10	2,51
4	90,76	2,32
5	88,66	2,17
6	86,73	2,73
7	84,36	3,59
8	81,33	4,43
9	77,72	5,14
10	73,73	5,44
11	69,72	4,47
12	66,60	

En outre, nous avons également étudié l'effet des covariables à travers le test de LogRank et les modèles de Cox (simple et étendu). Selon le test de LogRank, le sexe, le montant initial du prêt et le montant des échéances du prêt n'affectent pas les probabilités de sortie du chômage mais l'âge de l'assuré et l'ancienneté du prêt lors de la survenance du sinistre l'influencent. En revanche, le modèle de Cox révèle que toutes les covariables, à l'exception du sexe, influencent significativement les chances de sortir de la perte d'emploi. Le tableau ci-dessous récapitule les coefficients estimés via le modèle de Cox.

Modalités	coef	exp(coef)	se(coef)	z	P-valeur
Sexe : Masculin	-0,01	0,99	0,11	-0,12	0,90
Age survenance : Plus de 40 ans	-0,38	0,69	0,11	-3,48	**0,00**
Ancienneté : Moins de 18 mois	-0,37	0,69	0,13	-2,79	**0,01**
Ancienneté : Plus de 30 mois	0,13	1,14	0,13	0,97	0,33
Montant initial : Entre 9000 et 14000	0,44	1,55	0,18	2,41	**0,02**
Montant initial : Moins de 9000	0,51	1,66	0,23	2,26	**0,02**
Montant initial : Plus de 22000	-0,27	0,76	0,21	-1,28	0,20
Montant échéance : Entre 248 et 386	0,23	1,26	0,18	1,28	0,20
Montant échéance : Moins de 158	-0,46	0,63	0,19	-2,44	**0,01**
Montant échéance : Plus de 386	0,58	1,79	0,25	2,31	**0,02**

Toutefois ce résultat doit être considéré avec précaution puisque l'hypothèse fondamentale de risques proportionnels dans le modèle de Cox simple n'est pas vérifiée pour la variable ancienneté. Le passage à un modèle de Cox étendu visant à estimer deux coefficients pour chaque modalité de l'ancienneté regroupée en classe n'a pas réussi à résoudre le problème.

Par ailleurs, afin de prendre en compte l'effet de la crise sanitaire, survenue depuis 2020, sur la durée de la perte d'emploi, nous avons calibré un autre modèle de Cox étendu. Selon ce dernier, les taux de sortie de perte d'emploi après la crise ont diminué et valent 0,42 fois les taux de sortie avant le début de la crise. Toutefois, la fiabilité de ce coefficient reste à valider avec un backtesting à réaliser en 2022.

Enfin, dans la cinquième partie, nous avons montré d'abord comment est utilisée la table de maintien en prenant en compte l'effet du Covid-19 pour le calcul des provisions. Puis, nous avons mis en exergue l'impact de son utilisation sur les provisions en mars 2021. L'utilisation de la table de maintien a fait baisser les provisions de 12,8%.

Il est toutefois très important de garder à l'esprit les hypothèses faites lors du traitement des données (identification des sinistres et conditions de censures). Par ailleurs, l'insuffisance des observations dans la base de données utilisée peut causer des instabilités de la table dans le temps et par conséquent la table doit toujours faire l'objet de suivi surtout lorsqu'on observe des changements sur la composition du portefeuille.

Executive summary

In a rapidly changing regulatory context, with the Solvency II Directive and IFRS 17 which will come into force from January 1, 2023, insurers are more than ever forced to calculate the best estimate of their reserves. Indeed, these new standards encourage insurers to assess risks based on stochastic assumptions.

In guarantees such as death, incapacity, and invalidity present in credit insurance contracts, this results in the use of regulatory tables or portfolio experience tables. Unlike other guarantees, the job loss guarantee does not have a regulatory table (both entry and maintenance of loss of employment). Thus, the Value division of the Actuarial and Provident team of Société Générale Assurance is obliged to calculate extremely cautious reserves for job loss guarantees due to the lack of a maintenance table. Thus, in order to meet regulatory requirements with the calculation of the best estimate and to avoid over-reserving the risk of job loss, it was decided to set up a set of studies aimed at building experience tables maintenance on job loss.

This thesis aims to show the different stages of construction of a job loss maintenance table on the portfolio of Expresso, Société Générale Assurance's credit insurance product, as well as the impact of its use in reserving.

The thesis was divided into five parts. The first part sets the general framework of the study by defining certain important concepts such as creditor insurance, group contracts, job loss guarantee and its characteristics. We also presented the Expresso product.

The second part focused on the creation and the improvement of the reliability of the database. In addition, descriptive statistics on the portfolio were also made. Based on the initial claims database, we performed treatments to consolidate each claim into a single row and created additional variables such as the censorship dummy variable and other covariates. This work provided a final database with 1379 rows about job loss exits occurred between 01/01/2012 and 12/31/2018.

After presenting the theoretical Kaplan-Meier methods for estimating crude rates and the Whittaker-Henderson method for smoothing crude rates, in part three, we applied them to the Expresso portfolio in part four. For the choice of the smoothing parameters and therefore of the final smoothed rate curve, we retained the sufficiently smooth curve which performed the best predictions of exits from job losses compared to those actually observed over the period from

01/01/2019 to 12/31/2019. Below is the table summarizing the survival probabilities (survival function) and the job loss exit rates from the selected table.

Duration before exiting unemployment (in month)	Probability surviving in unemployment (in %)	Rate of exiting unemployment (in %)
0	100.00	2.29
1	97.71	2.35
2	95.42	2.43
3	93.10	2.51
4	90.76	2.32
5	88.66	2.17
6	86.73	2.73
7	84.36	3.59
8	81.33	4.43
9	77.72	5.14
10	73.73	5.44
11	69.72	4.47
12	66.60	

In addition, we also investigated the effect of covariates through the LogRank test and Cox models (simple and extended). According to the LogRank test, sex, the initial loan amount, and the amount of monthly financial commitment of the loan do not affect the probabilities of exiting unemployment but the insured's age and the seniority of the loan when the claim occurs impact it. In contrast, Cox's model reveals that all covariates, except sex, significantly influence the odds of exiting from job loss. The table below summarizes the coefficients estimated using the simple Cox model.

Factors	coef	exp(coef)	se(coef)	z	P-value
sex : Male	-0.01	0.99	0.11	-0.12	0.90
Age: Plus de 40 ans	-0.38	0.69	0.11	-3.48	**0.00**
Seniority : Less than 18 mois	-0.37	0.69	0.13	-2.79	**0.01**
Seniority: More than 30 mois	0.13	1.14	0.13	0.97	0.33
Initial loan amount: between 9000 and 14000	0.44	1.55	0.18	2.41	**0.02**
Initial loan amount: Less than 9000	0.51	1.66	0.23	2.26	**0.02**
Initial loan amount: More than 22000	-0.27	0.76	0.21	-1.28	0.20
Commitment: Between 248 and 386	0.23	1.26	0.18	1.28	0.20
Commitment: Less than 158	-0.46	0.63	0.19	-2.44	**0.01**
Commitment: More than 386	0.58	1.79	0.25	2.31	**0.02**

However, this result should be considered with caution since the fundamental assumption of proportional hazards in the simple Cox model is not verified for the seniority variable. Switching to an extended Cox model aiming to estimate two coefficients for each modality of seniority grouped into class failed to resolve the problem.

Moreover, in order to take into account the effect of the health crisis that has arisen since 2020 on the duration of unemployment, we have calibrated another extended Cox model. According to the latter, exit rates from job loss have decreased by 0.42 times since the start of the crisis, but the coefficient remains unreliable until a backtesting in 2022 is run.

Finally, in the last part, we first showed how the maintenance table is used by considering the effect of Covid-19 for the calculation of reserves. Then, we highlighted the impact of the use of sound on reserves in March 2021. The use of the table lowered by 12.8% the reserves for March 2021.

However, it is very important to remember the assumptions made during data processing (identification of claims and censorship conditions). In addition, the low size of the used database can cause instabilities in time of the table and therefore the table should always be followed up especially when the composition of the portfolio changes.