

Le processus de tarification en assurance non-vie : panorama des pratiques du marché

Gabriel Fauchet, IA



Le marché de l'assurance dommages fait face à de nombreux défis qui pourraient s'accroître dans les années à venir, notamment : dérive des charges sinistres, flambée du coût des climatiques, nouveaux entrants, risques émergents, etc.

Une des façons d'y répondre est d'aller vers davantage de sophistication tarifaire afin d'améliorer sa connaissance du risque, et d'augmenter sa réactivité en automatisant le plus possible l'ensemble du processus de tarification.

Cette sophistication tarifaire se traduit différemment selon le profil de l'assureur. Dans le panorama que nous présentons dans cet article, nous avons donc cherché à analyser les pratiques actuelles au sein des différentes étapes du processus de tarification.

Nous remercions les différents participants à cette étude qui nous ont permis de constituer un échantillon représentatif du marché de l'assurance non-vie.

Tarif technique et tarif commercial : deux cycles tarifaires qui coexistent

Un tarif assurantiel peut généralement se décomposer en deux grands éléments distincts : d'une part la prime pure, qui est l'évaluation financière la plus juste possible du risque couvert (l'espérance des pertes), et d'autre part la prime finale payée par l'assuré ou le sociétaire. Ainsi, on distingue nettement deux grands cycles de revue des modèles tarifaires parmi les répondants :

- La prime pure et les zoniers, revus au mieux tous les ans ;
- La prime commerciale, les modèles de demande et de prix de marché, mis à jour plus fréquemment : deux tiers des répondants revoient leur segmentation commerciale au moins annuellement, jusqu'à trimestriellement pour un quart d'entre eux.

Cet asynchronisme peut s'expliquer. A moins d'une rupture de tendance majeure impactant structurellement la fréquence ou le coût moyen d'un sinistre, on peut légitimement penser qu'une revue annuelle de la prime pure est suffisante ; pour autant, l'assureur doit pouvoir réagir aux évolutions du marché, à la pression concurrentielle et à de nouveaux objectifs de rentabilité : la grille tarifaire peut ainsi évoluer plus fréquemment.

Lors de la revue des modèles tarifaires, les équipes en charge du projet suivent généralement les grandes étapes décrites ci-dessous.

Tout d'abord, les données d'intérêt pour l'étude sont récupérées, souvent à partir de différentes sources, et agrégées dans une même base ; une phase de mise en forme est alors nécessaire : traitement des valeurs manquantes, discrétisation de variables continues, ajout de valeurs planchers, regroupements de classe, etc. Une part importante du processus de tarification est allouée à cette préparation des données : en moyenne 25% du temps du processus de tarification, soit autant que pour la création du modèle. Cet investissement est justifié car il est primordial de constituer des bases de données fiables pour interpréter correctement les

sorties des modèles (en informatique, « Garbage In, Garbage Out »).

Une fois la base d'entraînement créée, la modélisation en tant que telle est réalisée (25% du temps également) ; l'analyse exploratoire préalable des données, la connaissance du phénomène à expliquer et l'expertise du modélisateur permettent de sélectionner le modèle statistique le plus adapté.

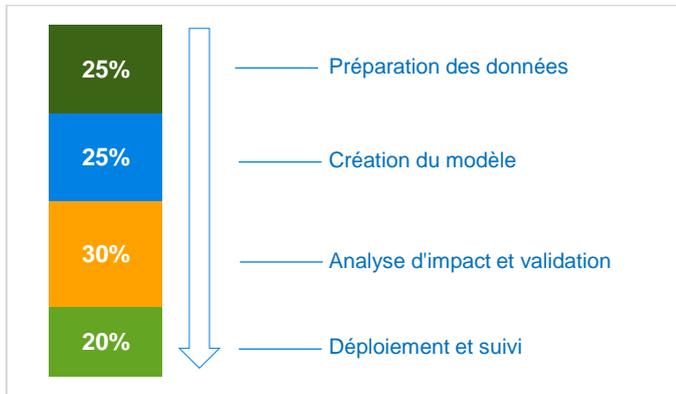


FIGURE 1 : TEMPS MOYEN ALLOUÉ PAR ÉTAPE DU PROCESSUS DE TARIFICATION

Environ 30% du temps est ensuite alloué à analyser les prédictions du modèle, et à décider de son déploiement en production : des comparaisons entre l'ancien et le nouveau tarif sont effectuées, et les impacts attendus sont estimés (volume d'affaires nouvelles, primes acquises, taux de résiliation, etc.). Le déploiement d'un nouveau modèle est ensuite généralement validé par au moins deux niveaux hiérarchiques, d'abord au sein des équipes techniques puis lors d'un comité qui regroupe les différents acteurs responsables du produit.

Enfin, le temps restant (20%) est consacré à la mise en œuvre du modèle en production et au suivi de sa performance. La vérification de l'absence d'écarts entre l'environnement de test et de production est une étape nécessaire et chronophage, le plus souvent car il faut traduire le nouveau tarif dans le langage du moteur de règles tarifaires. Une fois en production, la performance du modèle doit être suivie pour attester de sa qualité. Une des façons d'y parvenir est de procéder par A/B test (technique employée par un tiers des répondants) qui permet de distinguer les effets du nouveau modèle des autres effets (action marketing, pression concurrentielle, évolution de l'environnement économique, etc.).

Des logiciels commerciaux qui résistent face aux langages open sources

Le processus de tarification n'est pour le moment pas centralisé sur une même plateforme, et les répondants utilisent en moyenne quatre types d'outils ou de logiciels différents pour mener à bien leurs études tarifaires. On peut distinguer deux tendances principales dans l'utilisation de ces différents outils.

D'une part, SAS®, Excel® et R / Python sont des logiciels et langages de programmation polyvalents qui sont employés à toutes les étapes du processus de modélisation.

- SAS® reste le logiciel communément utilisé pour créer la base de données d'étude, et plus de quatre répondants sur cinq l'utilisent.
- Excel® est largement utilisé pour les analyses pre et post modélisation ; sa simplicité d'utilisation et la facilité à créer des visualisations graphiques sont ses principaux atouts en comparaison d'autres outils (notamment R / Python).
- Les langages de programmation R et Python sont utilisés par l'ensemble des répondants en complément des deux autres outils ; la profusion de bibliothèques disponibles pour les différentes étapes du processus de tarification, la communauté d'utilisateurs répondant bénévolement à la plupart des questions (par exemple sur stackoverflow.com) et le nombre croissant d'utilisateurs au sein des compagnies d'assurance expliquent leur diffusion progressive.

D'autre part, la modélisation qui est une étape cruciale encore dominée par les logiciels de tarification commerciaux, la quasi-intégrité des acteurs interrogés en utilisant un.

- Les logiciels commerciaux continuent de dominer le marché, les deux tiers des répondants en utilisant un lors de la modélisation (prime pure, zonier) ; on peut noter la percée très rapide d'Akur8® sur ce marché des outils de tarification : le logiciel est déjà utilisé par un quart des répondants, après seulement 3 années de commercialisation en France.
- R / Python sont utilisés conjointement à ces logiciels commerciaux par une large majorité des répondants. De nombreuses bibliothèques de modélisations statistiques existent en effet dans ces deux langages (par exemple sklearn pour Python, ou caret pour R) et permettent de créer la plupart des modèles statistiques les plus populaires, mais sans pour autant fournir une interface utilisateur intuitive. Par ailleurs, certains algorithmes de Machine Learning sont disponibles sur R / Python mais ne sont pas proposés par ces logiciels commerciaux alors qu'ils peuvent être utiles pour certaines parties du processus de tarification (par exemple pour détecter des valeurs aberrantes).

Le GLM, toujours au cœur de la modélisation en non-vie

Les assureurs cherchent à estimer le plus précisément possible les risques qu'ils couvrent. Pour y parvenir, ils ont à leur disposition différents algorithmes et modèles statistiques ; depuis quelques années, les modèles ensemblistes (Random Forest, Gradient Boosting) et, dans une moindre mesure, les réseaux de neurones, ont été testés pour estimer la prime pure,

notamment par de nombreux étudiants en actuariat dans le cadre de la rédaction de leur mémoire. Généralement, des performances prédictives supérieures aux différents types de modèles linéaires sont observées.

Pourtant, ces modèles linéaires, et plus particulièrement les Modèles Linéaires Généralisés (GLM) et les Modèles Additifs Généralisés (GAM) sont toujours privilégiés : l'ensemble des répondants emploie un de ces modèles dans ses tarifs en production. Cette prépondérance s'explique notamment par l'interprétabilité (le fonctionnement de l'algorithme doit être compréhensible par un expert, dans un format interprétable) et l'explicabilité (un néophyte doit comprendre explicitement les éléments responsables de la prise de décision de l'algorithme) de ces modèles, deux qualités essentielles pour communiquer aisément des changements tarifaires à différents interlocuteurs (comité de direction, réseau de distribution, régulateur, assuré ou sociétaire). Corolaire à cette prépondérance des modèles linéaires, la plupart des répondants construisent le modèle final en suivant la procédure stepwise (ajout de variables une à une maximisant un KPI et arrêt lorsque le KPI ne s'améliore plus), hormis les utilisateurs d'Akur8® qui automatise le processus de sélection des variables de manière globale.

Pour autant, un tarif d'assurance ne repose pas uniquement sur la bonne estimation de la prime pure ; la sophistication d'un tarif repose souvent sur l'addition d'éléments complémentaires, pour certains issus de la prédiction d'autres modèles ne nécessitant pas systématiquement l'utilisation de modèles linéaires (par exemple la modélisation du prix de marché). Ainsi, quasiment la moitié des répondants fait appel à d'autres types de modèles, permettant notamment de capturer des effets non linéaires et des interactions plus riches entre variables.

Des méthodes d'explicabilité de ces modèles complexes se démocratisent peu à peu : les méthodes d'analyse globale (PDP, ALE, etc.) et d'analyse locale (valeurs de Shapley, LIME, etc.) viennent aider l'actuaire à rendre plus transparent l'impact de chaque variable sur la prédiction finale ; l'utilisation de ces modèles est donc amenée à s'étendre¹.

Cet intérêt pour d'autres types de modélisation est légitime, notamment pour répondre aux limites des GLM : on peut citer sa structure rigide, sa difficulté à gérer de nombreuses variables explicatives et interactions (en partie lié au fait que les variables ne doivent pas être trop corrélées), et le fait de devoir faire des modifications en amont sur les variables, modifications parfois jugées subjectives (discrétisation des variables continues).

Données internes et externes : une richesse encore sous-exploitée

"Data is the new oil" énonçait Clive Humby au début du millénaire, et c'est d'autant plus juste pour les assureurs. Mais

¹ Milliman: Machine Learning et transparence des modèles en assurance

comme pour le pétrole, ces données n'ont de valeur qu'une fois transformées, raffinées.

90% des répondants sont ainsi conscients qu'ils disposent de gisements de données non exploitées ; ces données ne sont pas mises en valeur par manque de ressources et de temps : en effet, un travail préalable de documentation et de fiabilisation de ces bases est nécessaire avant de pouvoir envisager de les exploiter (déterminer les clefs d'identification, etc.). Par ailleurs, certaines données d'intérêt peuvent être dans des formats moins communs pour les équipes actuarielles : la quasi-totalité des répondants manipule soit des données structurées dans des datamarts soit des fichiers tabulaires (xlsx, csv, etc.) ; il y a pourtant également de l'information à valoriser parmi les documents et images acquis à la souscription ou lors d'un sinistre (factures, justificatifs, etc.).

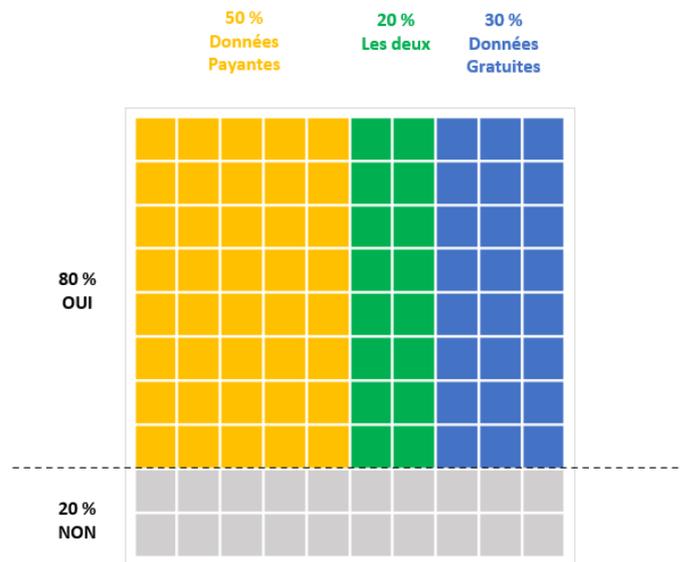


FIGURE 2 : UTILISATION DES DONNEES EXTERNES DANS LES MODELES TARIFAIRES

Les assureurs sont également à la recherche de données externes pour enrichir la connaissance des risques qu'ils couvrent. Plus de trois quarts des répondants utilisent ces sources de données dans leurs modèles tarifaires, pour la plupart provenant d'organismes publics. En effet, ces données sont de plus en plus nombreuses et accessibles en France : en témoigne l'évolution du site data.gouv.fr, plateforme ouverte des données publiques françaises, et la volonté gouvernementale de stimuler l'open data des données publiques². Pour les assureurs, l'enjeu est avant tout de ne pas être distancés par leurs pairs dans l'intégration de ces données très riches (par exemple pour construire des zoniers climatiques). Pour autant, bien que ces données soient en libre accès, leur utilisation en production demande souvent un travail conséquent :

² Milliman: Open Data et ses enjeux en assurance

- Un travail de recherche, pour définir quelles bases, parmi les dizaines de milliers disponibles, ont potentiellement de la valeur pour l'assureur : présence d'une documentation, granularité et historique suffisants, fréquence de mises à jour, proportion de valeurs manquantes, etc. ;
- Un travail d'analyse, de valorisation et d'intégration de ces données, souvent volumineuses et dans des formats plus difficiles à exploiter (données cartographiques Geojson, shapefile, etc.).

Les cas d'usage sont en tout cas nombreux, de l'utilisation d'une donnée externe structurée pour enrichir un zonier à la mise en œuvre de techniques de deep learning pour détecter des fissures sur des façades de bâtiment ou d'habitations à risques d'après la vue aérienne.

La modernisation du processus de tarification, portée par une nécessaire évolution des infrastructures IT

La sophistication tarifaire n'est pas envisageable sans un investissement en temps et ressources en R&D, en formation et développement des équipes techniques, mais aussi dans une infrastructure IT plus moderne.

C'est d'ailleurs ce dernier point qui est le plus souvent mentionné par les répondants parmi les trois sources de frustration les plus importantes du processus de tarification : l'innovation tarifaire est souvent bridée par des contraintes IT. Ainsi, pour presque deux tiers des répondants, ces contraintes limitent l'inclusion de nouvelles variables dans leur tarif. Autre contrainte répandue, la nécessité pour trois quarts des répondants de traduire le nouveau modèle tarifaire dans le langage informatique spécifique du moteur de règles : environ 10% du temps est consacré à ce travail de traduction et de validation des écarts entre environnements de développement et de production. Parmi les solutions possibles, on peut

envisager une intégration plus rapide des tables tarifaires en homogénéisant langage de production et de développement, ou en facilitant les transferts de ces tables via des API dédiées : en réduisant au maximum les transferts de données entre différents environnements, on réduit d'autant plus les risques opérationnels et les procédures de validation.

La technologie du cloud se démocratise et commence à être adoptée par une majorité de répondants, en particulier afin de réaliser les études tarifaires ; en effet, les ressources extensibles du cloud permettent de réaliser des gains de temps opérationnels et d'utiliser des algorithmes plus gourmands en mémoire. Certaines entreprises technologiques appuient d'ailleurs leur proposition de valeur sur cette technologie, comme certaines plateformes de data science ou certains logiciels de tarification. La réalisation des études tarifaires sur un environnement cloud permet par ailleurs de ne plus effectuer certaines tâches de modélisation sur son ordinateur en local, une pratique courante mais qui peut soulever des problématiques de RGPD lorsque l'on manipule des données d'assurés ou de sociétaires. Enfin, le développement d'un environnement de travail unique permet une meilleure automatisation des tâches.

Remerciements

Nous remercions vivement l'ensemble des sociétés qui ont participé à notre étude de marché et ainsi contribué à l'établissement de ce panorama des pratiques tarifaires en assurance non-vie, qui montre notamment les évolutions du marché en termes de sophistication méthodologique, de sources de données, d'outil et d'infrastructure IT.



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.



milliman.com

© 2022 Milliman, Inc. All Rights Reserved. The materials in this document represent the opinion of the authors and are not representative of the views of Milliman, Inc. Milliman does not certify the information, nor does it guarantee the accuracy and completeness of such information. Use of such information is voluntary and should not be relied upon unless an independent review of its accuracy and completeness has been performed. Materials may not be reproduced without the express consent of Milliman.

CONTACTS

Fabrice Taillieu, IA
fabrice.taillieu@milliman.com
Gabriel Fauchet, IA
gabriel.fauchet@milliman.com