

“ L’INFOTECH ”

Tarification : Faut-il céder aux sirènes du « Big Data » ?



L’édito

Comment la réglementation force la dérégulation de la tarification ?

Comment la réglementation force la dérégulation de la tarification ? La tendance réglementaire actuelle viendrait-elle perturber un certain conformisme de place en matière de tarification des risques de masse ? La directive européenne gender en est un exemple flagrant, puisqu’en interdisant l’usage de la variable sexe comme critère discriminant du prix des contrats, elle contraint les acteurs de l’assurance à explorer des pistes nouvelles pour leur tarification. On peut d’ailleurs imaginer dans le futur que l’utilisation d’autres variables socio-démographiques (âge, CSP...) sera proscrit.

Le projet de loi Hamon (ou aménagement de la Loi Châtel) agite également les esprits avec la crainte d’un fort impact probable sur les taux de résiliation des polices, et induit une réflexion en profondeur sur les structures tarifaires.

La réglementation joue le rôle d’amplificateur de concurrence à un moment où l’environnement est fortement réceptif : la crise incite les assurés à être plus attentifs au message prix de leurs assureurs alors qu’il est dans le même temps très aisé de comparer ce prix sur internet. Autant de raisons d’inciter les assureurs à mettre en place un positionnement tarifaire dynamique et réactif.

Les enjeux de ce positionnement seront à la fois de pouvoir réagir très rapidement face aux mouvements de la concurrence, d’éviter l’anti-sélection (éliminer les cases mal tarifées systématiquement exploitées par les comparateurs) et d’optimiser la performance des portefeuilles (ménager l’équilibre entre part de marché et rentabilité).

Les données, méthodologies et outils informatiques permettant d’établir en temps réel des tarifs efficacement différenciés existent bien sûr (avec recours au Big Data ou pas) mais ne sont pas encore suffisamment et systématiquement exploités en France, alors que nos voisins anglo-saxons en font un usage massif. D’aucuns diront que cet usage se pratique chez eux jusqu’à l’excès, mais il est néanmoins intéressant de réfléchir dès aujourd’hui au choix des futures armes dont il faut se doter dans l’arsenal de la sophistication tarifaire.

Améliorer l’appréciation et la segmentation du risque pour permettre de mieux le mutualiser, mais aussi de mieux réagir en période de guerre tarifaire - sans surexploitation des assurés fidèles - est désormais l’équation à résoudre par les assureurs. Il y a fort à penser que de nombreux travaux informatiques et actuariels seront à mener. Stéphane Chappellier nous en donne un aperçu passionnant dans cette Infotech.

Pierre ARNAL – Directeur Général

« Big-data » : LE concept du moment

Tout le secteur assurance bruisse aujourd’hui autour du « Big Data », concept à la fois très large et relativement flou, mais pourtant souvent présenté comme une sorte de Saint Graal apportant une solution miraculeuse à toutes les problématiques de tarification.

Dans ce contexte, il nous semble important de commencer par bien identifier les contours du phénomène « Big Data » et en quoi ce dernier pourrait constituer un vecteur efficace de l’innovation tarifaire.

Afin d’éviter que ce concept « fasse pschitt », il faut repartir des fondamentaux : la tarification en assurances cherche, à partir d’algorithmes basés sur des données, à identifier l’espérance de sinistralité d’un risque. Si des données et/ou des algorithmes permettaient de mieux expliquer la sinistralité, il est indiscutable que cela constituerait un différentiel concurrentiel incontestable pour celui qui saurait les exploiter correctement. D’où l’engouement actuel pour un phénomène qui promet d’apporter cette manne de façon quasi magique. Qu’en est-il réellement ?

Données : Etat des lieux

Les limites à l’accès aux données

Traditionnellement, la tarification s’est attachée à expliquer la sinistralité en se basant sur les données décrivant le risque (l’auto, l’habitation), le détenteur du risque (l’assuré) et la sinistralité historique. Les modèles de tarification ont d’abord privilégié l’utilisation des données déjà disponibles en interne dans les compagnies ; données qui ont été par la suite enrichies progressivement. Mais cet accroissement a des limites, notamment :

- **Les limites des systèmes d’informations**, à différents niveaux :
 - (a) **conservation des données** : pour mesurer l’élasticité au prix de la demande (propension de l’assuré à résilier sa police en fonction du pourcentage d’augmentation / réduction

tarifaire), il est indispensable de disposer dans le système des devis non-transformés et des contrats non-renouvelés avec les circonstances ayant prévalu à ces situations. Ensemble d'informations qui n'ont trop souvent pas été conservées, ou pas avec le niveau de profondeur suffisant.

- (b) **outils de cotation à disposition du réseau** : ces outils ne disposent souvent pas d'une performance suffisante pour traduire parfaitement l'ensemble de l'algorithme tarifaire que les actuaires souhaiteraient mettre en place. Lorsqu'ils le permettent, c'est souvent au prix de développements informatiques longs et très coûteux, qui aboutissent à une solution pas forcément très réactive.
- **Les limites au recueil d'information** : tout en restant dans le cadre réglementaire lié à la collecte d'informations personnelles, il existe une limite au nombre de données susceptibles d'être recueillies auprès de l'assuré. Ceci est d'autant plus vrai avec le développement de la souscription via Internet. Les comparateurs aboutissent à une limitation et une uniformisation des données collectées.

Ces limites peuvent certes être partiellement levées en investissant dans la capture et la conservation des données, ainsi que dans l'amélioration des outils de

cotation à destination du réseau, mais le recours à des données externes semble désormais inévitable.

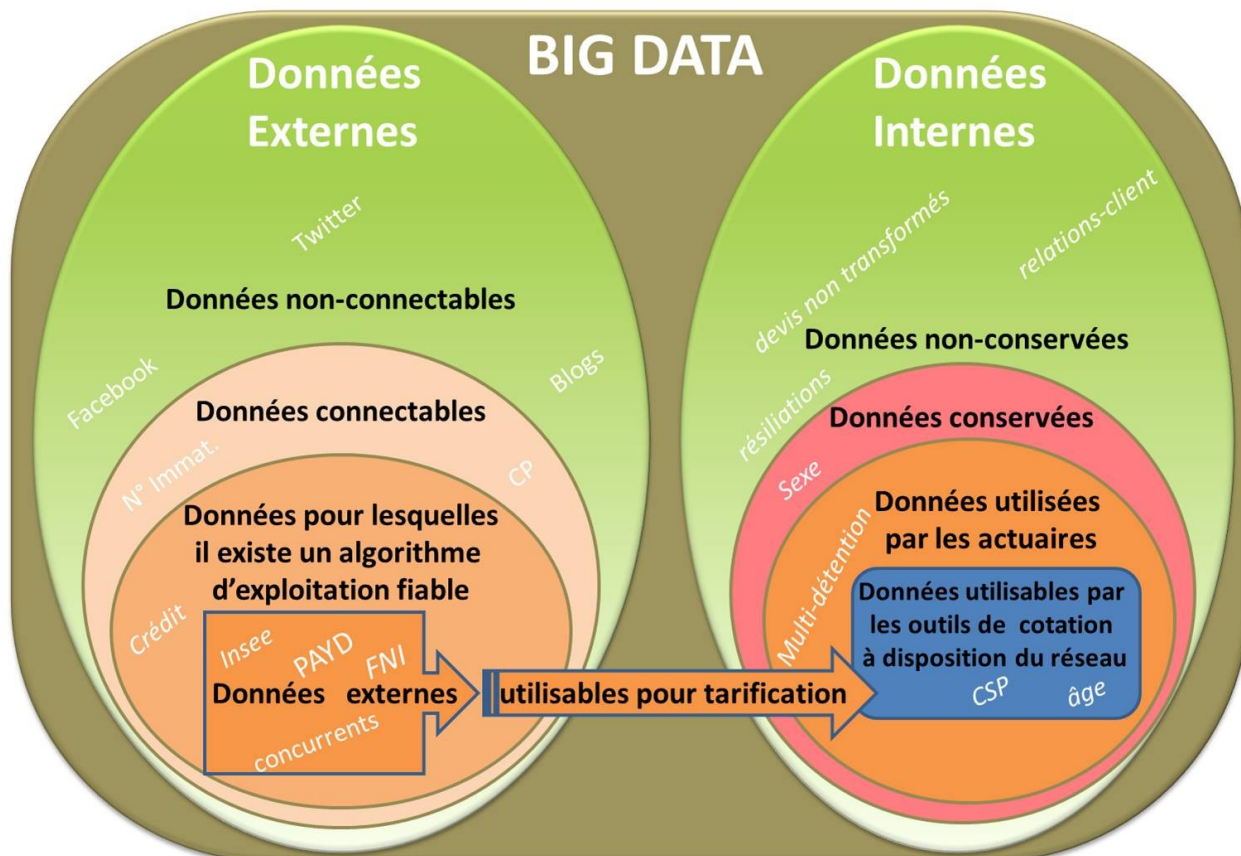
L'apport des données externes

Par exemple, le recours à des données externes est souvent nécessaire pour fiabiliser et enrichir l'information sur le risque assuré. En automobile en particulier, la caractérisation du véhicule peut ne pas être exacte si :

- 1) la description du risque faite par l'assuré est erronée. Il arrive fréquemment que l'assuré se trompe sur le modèle de véhicule et l'année de commercialisation (inconsciemment ou sciemment si une légère imprécision permet d'obtenir un tarif inférieur).
- 2) la base de données de l'assureur est insuffisamment précise et/ou à jour.

Or cette combinaison est malheureusement fréquente, et entraîne *de facto* une qualification défectueuse du risque, ceci impactant le positionnement tarifaire de l'assureur. Le recours à des données externes peut apporter une solution à cette problématique d'identification du véhicule. Le Fichier National des Immatriculations (FNI), récemment commercialisé, permet à partir du numéro de plaque minéralogique de connaître exactement le véhicule et une série de données précises décrivant ses caractéristiques.

Il existe d'autres données externes permettant



Il existe de nombreuses données non exploitées par la tarification sans avoir besoin de recourir à des données posant de réelles problématiques de fiabilité lors de leur exploitation.

d'accroître la connaissance de l'assuré ou la propension à

la sinistralité. Ainsi les bases INSEE recensent le niveau d'instruction et de revenus lié au code postal ou encore l'éloignement des services essentiels par rapport à ce code postal (et donc la propension à utiliser son véhicule pour accéder à ces services).

Données : Apport du Big Data

Données connectables ou pas

Pour qu'une donnée externe soit exploitable, il faut d'abord qu'elle soit connectable donc qu'il existe une variable permettant de mettre en relation cette donnée avec le portefeuille (n° d'immatriculation, code-postal).

Dès lors que les sources de données externes disposent de cette caractéristique, il est possible de les mettre en relation avec le portefeuille et donc d'identifier leur pouvoir explicatif sur la sinistralité.

Parmi les informations disponibles certaines ont un lien plus ou moins proche ou direct en rapport avec le sujet. Parmi les plus lointaines : les traces internet laissées par le détenteur de risques (Facebook, Twitter, Blogs...), aujourd'hui l'intérêt de ces données n'a pas encore été démontré (mais y en a-t-il un ?). Parmi les plus proches, celles qui sont directement attachées aux risques (port OBD des autos, et pourquoi pas domotique dans le futur pour la MRH).

Données comportementales

Ces données ont un avantage incomparable par rapport aux autres. Elles ne s'attachent plus à apporter des paramètres qui permettraient d'identifier des populations ayant en moyenne un comportement les exposant aux risques, mais à apporter les paramètres permettant d'identifier directement et individuellement ces comportements à risque. Ainsi avec un dispositif équivalent à une clé USB enfichée dans le port OBD (installé désormais en standard dans les voitures) il est possible, grâce à l'algorithme adéquat, d'identifier les comportements à risques : coups de volant intempestifs, freinages brusques, conduites nocturnes sur voies rurales...

Algorithmes : la limite à l'utilisation du Big Data

L'existence d'algorithmes permettant l'exploitation de ces données impose néanmoins une limite à l'utilisation des données externes dans les modèles de tarification. En effet, une fois ces données connectées au portefeuille, il est nécessaire de les faire « parler ». Par exemple, il faut un algorithme extrêmement puissant pour permettre d'identifier ce qui constitue un « coup de volant intempestif » dans un fichier composé de milliers de lignes prises à une seconde d'intervalle comprenant coordonnées GPS, vitesse... Ce type d'algorithme existe aujourd'hui et permet d'établir un score sur le comportement de conduite particulièrement prédictif.

Par contre certains cherchent encore l'algorithme qui permettra d'obtenir une interprétation plus fine en examinant les photos postées par le détenteur de risque sur Facebook ! Mais, il est fortement probable qu'il n'y ait aucune utilité à exploiter les zettaoctets d'information disponible. Cette information extraite sera-t-elle plus pertinente que celle obtenue par les données internes non-utilisées en raison de la limite des outils de cotation ou de leur non-conservation, ou encore celle obtenue par les données externes connectables et non encore exploitées ?

Lorsque les assureurs auront exploité l'univers des données externes facilement accessibles, il est probable qu'ils marquent un temps d'arrêt avant d'engager la coûteuse exploration des données moins facilement exploitables. On peut penser qu'ils s'orienteront en priorité vers la recherche de données permettant de comparer leur vision du prix du risque à celle de leurs concurrents. Ils essayeront aussi d'analyser comment l'assuré prend une décision face aux différentes propositions que le marché lui présente.

Tarification : Evaluation du risque ou de la prime ?

Elasticité et concurrence

La tarification ne s'arrête plus à l'identification du prix du risque mais cherche à optimiser le positionnement de ce prix par rapport à la concurrence, d'une part, et à la propension du client à choisir une offre plutôt que celle d'un concurrent, d'autre part.

Prenons l'exemple d'un assuré en portefeuille qui devrait bénéficier d'une réduction tarifaire en raison d'une amélioration de son risque. Si l'assureur connaissait à la fois : le prix proposé par ses concurrents et le fait que l'assuré soit réceptif ou non au message prix, il serait alors en position de mieux prendre la décision de l'octroi ou pas de cette réduction tarifaire.

La possibilité d'accéder à des données internes ou externes pouvant permettre (a) d'identifier la propension de l'assuré à souscrire en fonction de différents pourcentages d'augmentation tarifaire, ou (b) de connaître le positionnement des concurrents, serait bien sûr un atout considérable pour la compagnie.

Cette analyse est déjà possible et pratiquée dans les pays anglo-saxons mais timidement mise en œuvre par quelques acteurs en France.

- **Elasticité au prix de la demande :**

Il est possible de construire des modèles permettant d'expliquer et de prédire la réaction d'un assuré en fonction du pourcentage d'augmentation qui lui est demandé. Ces modèles requièrent de conserver dans les bases de données un historique des variations tarifaires pratiquées y compris (et surtout) pour les assurés qui ont résilié. Ces modèles sont très gourmands en données car les réactions de résiliation

sont souvent liées à des faits passés (d'où l'importance de la conservation d'une profondeur d'historique).

En reprenant l'exemple précédent, si l'assuré n'est que très peu influencé dans son choix par la variation tarifaire (« inélastique au prix »), il ne sera pas nécessaire pour l'assureur de réduire sa prime. Il pourrait même envisager de l'augmenter et conserver alors une réduction de prime pour un assuré qui serait plus influencé par la variation tarifaire (« élastique au prix »).

- **Positionnement par rapport à la concurrence :**

L'extraction de devis sur internet communément appelé le « webscraping » est aujourd'hui communément pratiquée par de nombreux acteurs du marché. Cette pratique a avantageusement remplacé celle du « client mystère ». Elle permet surtout le décodage de la formule tarifaire des concurrents, et donc de connaître le positionnement de sa propre offre sur chaque case tarifaire.

En reprenant le même exemple que précédemment, si pour l'assuré observé, les tarifs des concurrents sont supérieurs à celui proposé, il n'est alors pas nécessaire de proposer une réduction tarifaire qui pourra alors être réservée aux assurés auxquels les concurrents proposent des prix inférieurs.

On conçoit bien que la combinaison des deux modèles permet d'optimiser l'octroi de rabais.

Valeur client et optimisation tarifaire

Néanmoins, pour que cette approche aboutisse à un modèle qui ne surexploite pas les profils des assurés les plus inélastiques (la surexploitation conduirait inévitablement à leur résiliation !), il est nécessaire de mettre en place un processus tarifaire permettant de piloter les équilibres en toute connaissance des enjeux.

Un processus tarifaire commence par la détermination de la prime pure, c'est à dire le prix du risque à l'exclusion d'autres éléments. Cette prime pure doit utiliser toutes les variables permettant de segmenter le plus finement possible le risque. Segmenter insuffisamment dans un

marché ouvert aux comparateurs, conduit inmanquablement à récupérer uniquement les risques sous-tarifés (les sur-tarifés dans une case tarifaire partant chez les concurrents).

Sur cette base il faut ensuite positionner astucieusement chargement et marge pour optimiser le positionnement tarifaire vis-à-vis des concurrents en prenant en compte l'élasticité au prix des assurés.

Cette optimisation suppose un algorithme propre à chaque assureur selon ses objectifs de part de marché et de rentabilité. Cet algorithme de « valeur client » doit également prendre en compte un horizon temporaire pluriannuel en vue de veiller à conserver les bons risques. Ces derniers doivent en effet assurer une marge suffisante pour financer la remise à niveau progressive des risques mal tarifés en évitant qu'il ne partent avant d'avoir pu être rentabilisés.

Cette « valeur client » devra être suivie de près. En effet, une rotation accélérée des portefeuilles (liée à des conditions favorisées de résiliation), pourrait nécessiter un amortissement du coût d'acquisition d'un prospect sur une plus courte durée.

Faut-il céder aux sirènes du « Big Data » ?

En conclusion, il apparaît inévitable que le processus tarifaire intègre un niveau de sophistication beaucoup plus élevé dans les prochaines années. Cette sophistication passera par l'utilisation d'un plus grand nombre de données, utilisées pour la segmentation, le positionnement des mesures tarifaires par rapport aux attentes de l'assuré et à la concurrence. Où trouver ces données ? D'abord en interne mais également en externe. Il existe déjà des pistes encore trop peu utilisées de données « connectables » disposant d'algorithmes permettant de les exploiter avant d'entamer une recherche d'autres données. Il est donc vrai que l'assurance va devoir s'attaquer au domaine du « Big Data », mais le sous-ensemble qui doit être exploité est déjà un terrain connu et exploité par d'autres marchés, il n'est pas (encore) nécessaire d'explorer l'inconnu.



Stéphane CHAPPELLIER
Associé
Responsable du Pôle IARD

ACTUARIS INTERNATIONAL annonce en avant-première le prochain lancement (Septembre 2013) de son logiciel de Tarification : ACTU-PRICING™

Un logiciel novateur de dernière génération dédié à assister les actuaires pour répondre aux nouveaux enjeux techniques en tarification des risques de particuliers.

