

a) Problématique

L'un des grands objectifs de la science actuarielle est d'estimer les coûts futurs de produits d'assurance. La tâche de déterminer la valeur de ces coûts pour un contrat en particulier, appelée la tarification en assurance, est depuis longtemps un objet de recherche dans la communauté actuarielle. À partir de techniques plutôt intuitives, la science actuarielle s'est de plus en plus développée et offre maintenant, et depuis plusieurs décennies, un appareil théorique rigoureux contenant plusieurs outils et techniques pouvant être utilisés, mais nécessitant – évidemment - de nombreuses améliorations. Bien que de nombreux changements technologiques et sociaux aient déjà amené dans le passé des transformations de la science actuarielle, les fondements sont toujours plus ou moins restés les mêmes. Avec l'avènement des systèmes de géolocalisation sur les voitures, ou d'ordinateurs de bord compilant et enregistrant diverses données télémétriques potentiellement pertinentes pour analyser la qualité de conduite, nous estimons que le milieu de l'assurance vit une sorte de petite révolution ayant la possibilité de modifier fortement les fondements théoriques et pratiques de la tarification en actuariat.

b) Objectifs de recherche

La disponibilité de nouvelles données télématiques en assurance automobile fait en sorte qu'il est maintenant possible d'avoir un nouveau regard sur les activités d'assurance. La recherche scientifique sur ce sujet en est à ses débuts (Ayuso et al. 2016, Verbelen et al. 2016, Boucher et al. 2017) et les résultats sont prometteurs. Sous l'hypothèse que les nouvelles informations récoltées par les assureurs arriveront à mieux expliquer et comprendre le risque que les variables de segmentation usuelles (âge, sexe et état civil du conducteur, entre autres), nous développerons de nouvelles approches semi-paramétriques flexibles pour tarifier les produits d'assurance, ce qui permet d'estimer les probabilités d'accident et le coût de ces accidents.

c) Méthodologie

Pour développer une approche semi-paramétrique flexible permettant de mieux comprendre l'impact de la distance parcourue sur la probabilité d'avoir un accident d'automobile, la théorie des modèles additifs généralisés pour paramètres de location-échelle (GAMLSS), généralisant les GAM sera ainsi une voie à étudier dans un premier temps. La flexibilité de ce modèle fait en sorte que des dépendances complexes entre variables aléatoires pourraient être considérées. En effet, les diverses couvertures ou chapitres d'assurance, les différents véhicules d'un même contrat d'assurance et les contrats d'assurance d'un même assuré observés à travers le temps créent une structure de dépendance hiérarchique complexe qu'il convient d'étudier pour mieux comprendre le risque.

Par la suite, toujours dans le cadre d'une estimation non-paramétrique de l'utilisation simultanée de la distance parcourue et de la durée contractuelle, il pourrait être pertinent d'utiliser d'autres techniques d'apprentissage statistique adaptées à l'actuariat comme on peut les consulter par exemple chez Gao et al. (2018). À la lumière des résultats obtenus, il sera intéressant de voir de quelle manière une distribution de comptage paramétrique, plus concise, pourrait être déduite. On rappelle que plusieurs distributions de comptage peuvent être construites à partir de certaines hypothèses de base. En ayant une exposition au risque nouvellement définie, nous estimons qu'il sera intéressant de faire une revue exhaustive de plusieurs modèles de tarification en assurance pour voir l'impact d'une telle redéfinition. Une comparaison de l'ajustement et du pouvoir

prédictif sont à prévoir, et des nouveaux outils de comparaison de modèles typiquement adaptés à la science actuarielle seront développés. Similairement, nous serons bien positionnés pour étudier convenablement le pouvoir prédictif des caractéristiques usuelles de segmentation, comme le sexe ou l'âge du conducteur, dont l'utilisation en tarification est de plus en plus contestée, voire même interdite dans certaines juridictions. Nous vérifierons de quelle manière l'ajout d'une exposition au risque nouvellement définie pourra toucher ces régresseurs et de quelle manière il serait le plus approprié d'enlever les effets de ces variables de segmentation interdites. Finalement, nous estimons qu'il sera pertinent de voir comment généraliser les approches Boucher et Couture-Piché (2016) dans le nouveau paradigme d'une autre définition de l'exposition au risque.

d) Contribution du projet à l'avancement des connaissances

- Le projet de recherche proposé développera de nouveaux modèles statistiques adaptés aux caractéristiques des données télématiques dans lesquels une structure de dépendance hiérarchique sera incluse. En généralisant les modèles GAM, GAMLSS ou les techniques d'apprentissage statistique pour tenir compte de la dépendance hiérarchique présente dans nos données d'assurance (réelles ou simulées), les approches développées auront le potentiel d'être utilisés dans d'autres contextes.
- L'approche actuarielle qui sera développée devrait provoquer des changements importants dans les méthodes d'analyse du risque et au niveau de la tarification en assurance automobile.
- Nous verrons de quelle manière la télématique éliminera le besoin d'utiliser l'âge, le sexe, ou l'état civil du conducteur, variables de segmentation socialement discutables et maintenant proscrites d'utilisation dans plusieurs pays. En effet, plusieurs pays incluant le Canada ont pris conscience du caractère délicat, voire discriminatoire de l'utilisation du sexe, de l'état matrimonial, de l'état familial et de l'âge d'un assuré en tarification (on peut référer à l'arrêt *Bates c. Zurich Insurance* pour le Canada). Des solutions alternatives doivent donc être trouvées, et l'avènement des systèmes de géolocalisation sur les voitures est devenu la meilleure voie à explorer.

e) Bibliographie

Ayuso, M., Guillen, M. et Pérez-Marin, A.M. (2016). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C: Emerging Technologies*, **68**, 160-167.

Boucher, J.-P. & Couture-Piché, G. (2016), Modeling the Number of Insured Households in an Insurance Portfolio Using Queuing Theory, *ASTIN Bulletin*, 46(2), 401-430.

Boucher, J.-P., Côté, S. et Guillen, M. (2017). Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks* **5**(54).

Gao, G., Meng, S. et Wüthrich, M.V. (2018). Claims frequency modeling using telematics car driving data. Technical Report.

Verbelen, R., Antonio, K. et Claeskens, G. (2016). Unraveling the predictive power of telematics data in car insurance pricing. *SSRN ID: 2872112*.