

L'impact de la sinistralité passée sur la sinistralité future (2) : une modélisation des classes de risques

Olga A. Vasechko et Michel Grun-Rehomme

Volume 79, numéro 3-4, 2011–2012

URI : <https://id.erudit.org/iderudit/1091878ar>

DOI : <https://doi.org/10.7202/1091878ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Faculté des sciences de l'administration, Université Laval

ISSN

1705-7299 (imprimé)

2371-4913 (numérique)

[Découvrir la revue](#)

Citer cet article

Vasechko, O. A. & Grun-Rehomme, M. (2011). L'impact de la sinistralité passée sur la sinistralité future (2) : une modélisation des classes de risques. *Assurances et gestion des risques / Insurance and Risk Management*, 79(3-4), 279–311. <https://doi.org/10.7202/1091878ar>

Résumé de l'article

Dans ce marché concurrentiel de l'assurance automobile, l'assureur se doit d'inciter les bas risques (les bons conducteurs) à rester dans son portefeuille et d'attirer, dans un même temps, à lui les bas risques des autres compagnies. Encore faut-il bien estimer ces risques dans un contexte particulier d'information incomplète.

Dans cet article, on construit différents indicateurs de sinistralité antérieure, obtenus par le croisement du coefficient réduction majoration (CRM) et de l'ancienneté de permis, qui ont empiriquement un meilleur effet prédictif que le CRM sur la sinistralité future, ainsi l'assureur dispose de plusieurs stratégies. En l'absence de données longitudinales, ces indicateurs sont utilisés également pour revisiter la problématique de l'asymétrie d'information.

On constate, à l'aide de modèles Logit bivariés, utilisés pour éviter les biais liés à l'endogénéité du choix de contrat et obtenir des résultats robustes quant à l'estimation des probabilités de sinistre, que l'effet marginal (+ 9 % en moyenne) de cet indicateur de sinistralité passée sur la sinistralité actuelle est plus important que celui des autres variables. Les assurés qui choisissent un contrat RC ont en moyenne des probabilités plus élevées de sinistralité.

**L'impact de la sinistralité passée
sur la sinistralité future (2) :
une modélisation des classes de risques
par Olga A. Vasechko et Michel Grun-Rehomme**

RÉSUMÉ

Dans ce marché concurrentiel de l'assurance automobile, l'assureur se doit d'inciter les bas risques (les bons conducteurs) à rester dans son portefeuille et d'attirer, dans un même temps, à lui les bas risques des autres compagnies. Encore faut-il bien estimer ces risques dans un contexte particulier d'information incomplète.

Dans cet article, on construit différents indicateurs de sinistralité antérieure, obtenus par le croisement du coefficient réduction majoration (CRM) et de l'ancienneté de permis, qui ont empiriquement un meilleur effet prédictif que le CRM sur la sinistralité future, ainsi l'assureur dispose de plusieurs stratégies. En l'absence de données longitudinales, ces indicateurs sont utilisés également pour revisiter la problématique de l'asymétrie d'information.

On constate, à l'aide de modèles Logit bivariés, utilisés pour éviter les biais liés à l'endogénéité du choix de contrat et obtenir des résultats robustes quant à l'estimation des probabilités de sinistre, que l'effet marginal (+ 9 % en moyenne) de cet indicateur de sinistralité passée sur la sinistralité actuelle est plus important que celui des autres variables. Les assurés qui choisissent un contrat RC ont en moyenne des probabilités plus élevées de sinistralité.

Mots clés : Assurance automobile, sinistralité, modèles Logit bivariés.

Les auteurs :

Les auteurs sont attachés au ERMES (CNRS), Université Paris II / Sorbonne Universités
12 place du Panthéon, 75005 Paris, France.

Courriel : o.vasechko_stat@yahoo.fr; grun@u-paris2.fr

ABSTRACT

Facing a competitive insurance market, the insurers have to adopt the best marketing strategy to attract the lowest risky agents, (i.e. the good drivers with the lowest probability of accidents), and to retain them within their portfolios of policyholders. Still insurers need to estimate the low risks in the context of asymmetric information.

In this paper, we construct an indicator of past claims, which combines the Bonus-Malus coefficient (BM) and the seniority of the driver. This indicator is a better estimate of the past claims than the BM coefficient alone. Since longitudinal data are not available, this indicator is also used to study the issue of information asymmetry.

Using a bivariate logit model to avoid bias and provide robust results, we find that the marginaffect (+9% on average) of this indicator of past claims, on the actual claims, is more significant than other exogenous variables. The policyholders with a third-party insurance contract have, on average, a larger probability to incur claims.

Keywords: Automobile insurance, insurance claims, bivariate logistic regression.

I. INTRODUCTION

Le marché des assurances est un contexte privilégié d'information incomplète entre l'agent (l'assuré) et le principal (l'assureur). L'assureur ne connaît pas tout de l'assuré, et en particulier les risques financiers qu'il lui fait courir à travers sa sinistralité. En situation d'anti-sélection, l'assuré dispose d'un avantage informationnel par rapport à l'assureur avant la conclusion du contrat d'assurance. Cette hypothèse d'anti-sélection stipule que les hauts risques ont tendance à choisir de plus fortes couvertures que les bas risques (Rothschild et Stiglitz, 1976, Wilson, 1977, Spence, 1978). Mais de nombreux auteurs ont montré empiriquement que cette hypothèse d'anti-sélection est très secondaire en assurance automobile (Chiappori et Salanié, 2000, Dionne *et al.*, 2001 *a, b*, Grun-Réhomme et Joly, 2003, Cohen, 2005, Grun-Rehomme et Benlagha, 2007, Saito, 2008, Young *et al.*, 2009).

L'assureur essaye donc d'appréhender, d'estimer les risques supplémentaires liés à cette asymétrie d'information à l'aide des renseignements demandés au moment de la souscription du contrat et du suivi du comportement de l'assuré au niveau de sa sinistralité.

Pour Daniel Bernoulli (1738)¹ : «Le risque est l'espérance mathématique d'une fonction de probabilité d'évènements». Il considère ainsi la valeur moyenne des conséquences d'évènements

affectés de leur probabilité. Dans son acceptation actuelle (conforme à l'approche de Bernoulli), la notion de probabilité est attachée au risque. Le risque naît de l'aléa, dès que l'on ne peut prévoir avec certitude un futur.

Dans notre démarche pour tenter de répondre à la question posée dans le titre, nous partons de deux hypothèses généralement admises sur le comportement des assurés : l'existence de différents degrés d'aversion au risque et l'existence de différents niveaux de risques.

Devant cette situation d'aversion au risque, l'assureur propose différents contrats d'assurance pour capter plusieurs segments du marché et plus particulièrement les bas risques. Dans cette mutuelle d'assurance, dont les données seront utilisées ultérieurement, quatre formules de garantie sont proposées :

- 1- Responsabilité Civile (RC, assurance minimale obligatoire).
Sont inclus dans cette formule des garanties : Défense-recours
– Attentats – Catastrophes Naturelles – Corporel du conducteur
– Assistance.
- 2- Dommages au véhicule (DV1) : RC + Garantie Dommage au véhicule toutes causes, avec une franchise importante.
- 3- DV2 : RC + Garantie Dommage au véhicule toutes causes, avec une franchise moyenne.
- 4- DV3 : RC + Garantie Dommage au véhicule toutes causes, avec une franchise faible.

Les contrats DV ou « tous risques » garantissent l'assuré pour ses propres dommages en cas d'accident où celui-ci est responsable avec ou sans adversaire (perte de contrôle) mais aussi dans les événements « vandalisme », accident de « parking » (tiers non identifié) ou bien encore selon les compagnies pour l'évènement « grêle ». Le « tous risques » intervient également en « avance sur recours » : l'assureur indemnise l'assuré dans une situation où la responsabilité de celui-ci n'est pas engagée sans attendre l'argent de l'autre compagnie (exemple des accidents corporels ou des accidents avec un étranger qui peuvent être clos seulement après plusieurs années).

Avec un contrat d'assurance contenant une franchise (DV1-DV3), aucune indemnité n'est versée lorsque la perte ne dépasse pas le montant de la franchise. Dans le cas contraire, l'indemnité est égale à la perte dont on déduit la franchise.

Le caractère obligatoire de l'assurance, lié au risque de responsabilité civile de l'assuré, a pour avantage de limiter les effets d'anti-

sélection susceptibles de conduire à l'inassurabilité du risque, au cas où seuls les conducteurs à hauts risques s'assureraient.

La problématique de cet article est la suivante : Quelles sont les informations dont dispose l'assureur qui peuvent lui permettre d'avoir une bonne estimation des mauvais risques (mauvais conducteurs) et des bons risques (bons conducteurs) dans son portefeuille ? L'hypothèse d'anti-sélection est-elle invalidée au profit d'un indicateur de sinistralité passée (ou antérieure), « toutes choses égales par ailleurs » ?

Nous avons abordé ces questions dans un précédent article (Vasechko, Grun-Rehomme, 2010) sous un angle différent, basé sur la fréquence des sinistres et non, comme ici, sur le Coefficient Réduction-Majoration et des régressions logistiques bivariées. On supposait, en fonction de la sinistralité passée, que chaque classe de risques est constituée de deux catégories de conducteurs : les assurés à bas risques et ceux à hauts risques. À l'aide d'une loi binomiale négative et d'une approche bayésienne, on a montré que la probabilité d'être un conducteur à bas risques est plus importante en l'absence de sinistres (ou avec un seul sinistre) et qu'à l'inverse la probabilité d'être un assuré à hauts risques augmente fortement dès que l'assuré a 2 ou 3 sinistres au cours de l'année de référence. Dans presque tous les cas, la sinistralité passée est un bon indicateur de la sinistralité future.

Dans ce texte, on utilise des régressions logistiques bivariées

La section suivante présente le contexte du marché de l'assurance automobile en France. Les données sont présentées dans la troisième section. Une variable de sinistralité passée est construite dans la quatrième section et les modélisations de la sinistralité actuelle proposées en section 5. Enfin, une conclusion, une bibliographie et des annexes terminent cet article.

2. LE MARCHÉ DE L'ASSURANCE AUTOMOBILE EN FRANCE

Depuis ces cinq dernières années, la concurrence s'est accentuée dans le secteur de l'assurance automobile, avec la venue des banques, de l'assurance directe et l'ouverture européenne. De plus, il faut noter que le développement des offres sur internet permet aux consommateurs de comparer rapidement les avantages des contrats de différents assureurs. Les assurés changent maintenant plus fré-

quement d'assureurs. Ces derniers devraient être en mesure de préserver leurs positions en adaptant leurs stratégies commerciales. La baisse de la sinistralité observée ces dernières années avait permis une diminution des primes d'assurance. Mais si effectivement le nombre d'accidents déclarés est en recul, le coût des sinistres corporels et des réparations ne cesse d'augmenter. Tout laisse à penser que les assureurs vont subir une dégradation de leur ratio du fait d'une part de ces augmentations et d'autre part de tarifs plus attrayants. Aujourd'hui, les marges des assureurs ne leur permettent pas de pouvoir baisser significativement leurs tarifs. Des hausses sont même prévues (variables selon les compagnies de 3 à 5 %) pour cette année 2011. Afin de préserver la pérennité de l'entreprise dans ce marché concurrentiel, il est nécessaire pour les assureurs de bien cerner les coûts de sinistralité.

Cette pérennité de l'entreprise repose donc sur une maîtrise des flux financiers et en particulier elle cherche à diminuer les risques financiers auxquels elle doit faire face dans ses contrats, en cas de sinistralité responsable de la part de ses assurés.

Les stratégies commerciales peuvent s'orienter dans quatre directions (non exclusives) :

- Inciter les bas risques (assurés dont la probabilité de sinistralité est faible, les bons conducteurs) à rester dans le portefeuille : fidélisation de la clientèle,
- Attirer les bas risques des autres compagnies,
- Inciter les hauts risques à faire des efforts,
- Surtaxer (voire exclure) les hauts risques, qui répugnent à l'effort.

2.1 Le coefficient réduction majoration ou bonus-malus

Le coefficient réduction majoration, noté CRM, ou Bonus-Malus, est un mécanisme qui permet d'ajuster la prime d'assurance automobile payée par l'assuré, en fonction de sa conduite observée au cours des périodes antérieures. Il permet donc à la compagnie d'assurance d'intégrer le niveau de risque de l'assuré dans la tarification comme élément de crédibilité de l'historique individuel. Il devient alors un outil de tarification *a posteriori* à l'intérieur de chaque classe de tarification, basée sur d'autres facteurs explicatifs du risque liés au conducteur, au véhicule et à la zone géographique (Dionne et Vanasse, 1992, Henriet et Rochet, 1991). Ce coefficient est conservé par les assureurs et disponible pour l'ensemble des acteurs du marché. Il ne fait intervenir que la fréquence des sinistres responsables survenus, et non le coût de ces sinistres.

La clause française du bonus-malus a été adoptée en 1984 et c'est le seul système européen multiplicatif. La Commission Européenne de Bruxelles avait formulé des injonctions à la France de ne plus utiliser de façon réglementaire le système bonus-malus. Les assureurs français (mutuelles et compagnies traditionnelles), pour une fois solidaires, avaient souhaité continuer d'appliquer le système bonus-malus tel qu'il est. L'argumentation reposant sur le désir des assureurs de garder ce coefficient comme résumé du comportement de l'assuré lorsque celui-ci change d'assureur. Ils ont donc fait pression auprès du gouvernement et ont obtenu gain de cause : La France a soutenu la position des assureurs et obtenu de conserver ce système obligatoire après différents procès avec la Commission Européenne.

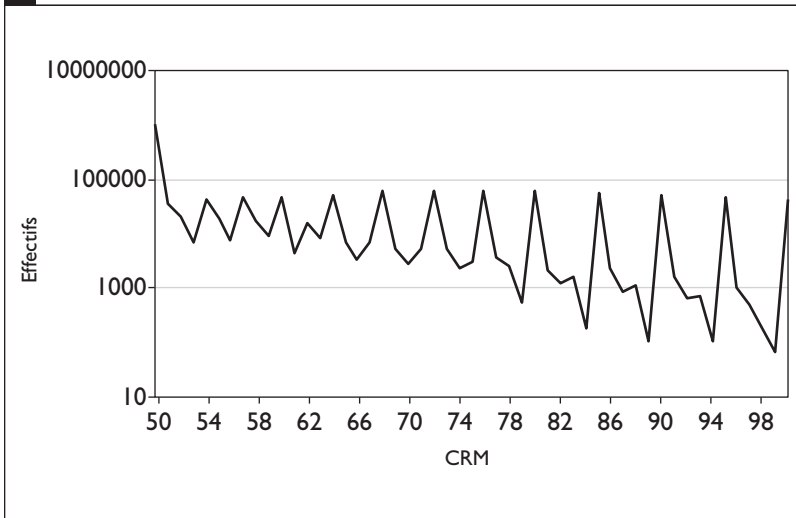
Rappelons que la première année d'assurance, le coefficient réduction majoration est égal à 1.00 (ou 100), et qu'il est réévalué chaque année. Si l'assuré n'a eu aucun sinistre responsable, le coefficient est multiplié par 0.95 avec une règle pour les arrondis, en utilisant à chaque étape, la partie entière et il est multiplié par 1.25 pour chaque sinistre 100 % responsable et par 1.125 pour les sinistres à torts partagés. En aucun cas, le coefficient ne pourra dépasser 3.50 (ou 350) ou passer en dessous de 0.50 (ou 50, dispositions législatives). Sans sinistre, le bonus passe de 100 à 50 après 13 ans.

**TABLEAU I
QUELQUES EXEMPLES DE LA VALEUR
DU BONUS-MALUS, SUR 5 ANS, EN FONCTION
DE LA SINISTRALITÉ PASSÉE**

Situation	1	2	3	4	5
Aucun sinistre la première année	1.00	0.95	0.90	0.85	0.80
Aucun sinistre sur 10 ans	0.76	0.72	0.68	0.64	0.60
Aucun sinistre sur 15 ans	0.57	0.54	0.51	0.50	0.50
1 sinistre 100% responsable la 1ère année	1.00	1.25	1.18	1.00	0.95
1 sinistre 100% responsable la 3ème année	1.00	0.95	0.90	1.12	1.06
1 sinistre la 1ère année à torts partagés	1.00	1.12	1.06	1.00	0.95
2 sinistres 100% responsables aux dates 2 et 4	1.00	0.95	1.19	1.03	1.29

Dans un portefeuille, on trouve plus de 50 % des assurés ayant un CRM égal à 50. Bien évidemment certaines valeurs du CRM sont plus fréquentes que d'autres (cf. figure 1 ci-dessous).

FIGURE I
DISTRIBUTION DES EFFECTIFS (ÉCHELLE LOGARITHMIQUE) SELON LE BONUS-MALUS (ENTRE 50 ET 100)

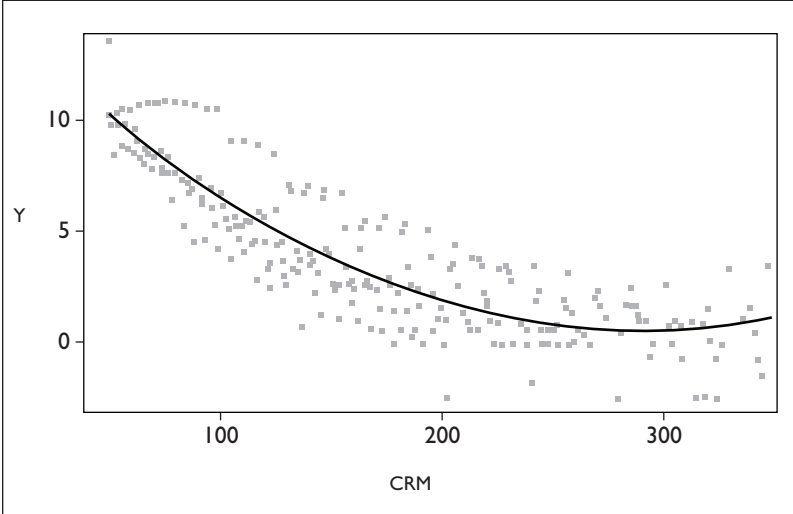


Comme le montre la figure précédente, la répartition des assurés (en réalité, le nombre de véhicules années assurés) selon le CRM présente une périodicité qui varie avec le CRM. Selon la règle des 5% de baisse relative des CRM en l'absence de sinistres, une « période » de 5 concerne l'évolution du CRM au début (0.95, 0.90, 0.85, 0.80), puis une « période » de 4 (0.76, 0.72, 0.68, 0.64), puis de 3 (0.60, 0.57, 0.54, 0.51). Les valeurs modales correspondent aux trajets les plus courants : retour au bonus maximum des conducteurs qui viennent d'être responsables d'un sinistre (d'où une valeur moyenne à 0.62), et trajet vers le bonus maximum pour les jeunes conducteurs. Pour des malus supérieurs à 2.00, l'effectif est faible.

La figure suivante représente l'ensemble du portefeuille. L'axe des abscisses correspond au coefficient réduction majoration et l'axe des ordonnées au logarithme du nombre de véhicules années assurés. Une modélisation log-quadratique permet d'expliquer les $\frac{3}{4}$ de la dispersion du portefeuille.

Bien sûr, le bonus-malus n'est pas indépendant de l'ancienneté de permis et de l'âge du conducteur, et ceci est particulièrement vrai pour les jeunes. Il est globalement une fonction décroissante de l'ancienneté de permis et de l'âge du conducteur.

FIGURE 2
DISTRIBUTION DE L'ENSEMBLE DU PORTEFEUILLE
(ÉCHELLE LOGARITHMIQUE) EN FONCTION
DU BONUS-MALUS



2.2 Le rôle du bonus-malus dans un contexte de concurrence

La compétition entre les assureurs se joue moins sur les prix que sur le Bonus-Malus. En effet, on constate actuellement différentes stratégies (assez proches) qui vont en ce sens :

- Super bonus : réduction supplémentaire liée au nombre d'années où l'assuré a un bonus égal à 50. Par exemple, après 3 ans avec un CRM à 50, application d'un bonus spécial.
- Absence de pénalités pour les assurés qui ont un CRM à 50 et qui ont un seul accident responsable sur une période de 3 ans.
- Le « Bonus à vie », qui est une réduction commerciale, est obtenu quand le CRM a été à 50 de façon continue, pour le seul véhicule et le seul conducteur principal (titulaire d'un permis de conduire d'au moins 16 ans) visés par le contrat au cours des trois dernières années.
- Le « bonus longue durée », qui ne fait varier la cotisation qu'après le deuxième sinistre responsable pour les bons conducteurs, alors que la loi ne prévoit cette disposition que pour un seul accident provoqué.

Par contre, l'assureur a la possibilité de résilier un contrat à son échéance en raison du nombre de sinistres (responsables et non responsables, y compris le bris de glace) et de la gravité de ceux-ci. Ces critères sont variables selon l'assureur. Ce dernier se doit toutefois de respecter un préavis de deux mois (article L 113.12 du code des assurances) et doit annoncer la résiliation par lettre recommandée. Notons que l'assureur n'est pas obligé d'indiquer le motif de la résiliation. Ajoutons que l'assureur ne peut résilier un contrat immédiatement après un sinistre (sans attendre l'échéance du contrat) que dans les deux cas suivants (article L 211.1.2 du code des assurances) :

- le sinistre a été causé par un conducteur en état d'imprégnation alcoolique,
- le sinistre a été causé par une infraction au code de la route qui a entraîné une suspension du permis de conduire d'au moins un mois ou une annulation.

3. LES FACTEURS DE RISQUE : APPROCHE EMPIRIQUE

L'assureur doit répartir la charge de sinistralité de façon équitable entre tous les assurés, en même temps qu'il mutualise les risques entre les assurés qui présentent des caractéristiques semblables personnelles et de véhicule. L'assureur procède donc à une recherche minutieuse de tous les facteurs disponibles et susceptibles d'expliquer le risque.

Dans ce marché concurrentiel, à partir du moment où l'un des assureurs commence à segmenter le marché, et donc à attirer les « bons » risques, les autres assureurs vont aussi devoir segmenter ce marché; la segmentation devient inévitable.

Quels sont les facteurs qui peuvent avoir un impact sur la sinistralité ?

Ils sont très nombreux et pas forcément observés ou observables par l'assureur.

Toutefois l'assureur dispose d'une panoplie d'informations liées au conducteur et au véhicule.

Dans notre démarche empirique, nous disposons d'un échantillon aléatoire, issu du portefeuille d'un assureur, de 43 245 observations (taille obtenue après la mise en place de différents contrôles et l'élimination des données erronées ou comportant des valeurs man-

quantes). Les données concernent l'année 2004 (année de référence) et sont constituées de trois groupes de variables (caractéristiques du conducteur, du véhicule et la sinistralité) qui sont décrits ci-dessous.

3.1 Les caractéristiques du conducteur

Sexe : il s'agit du sexe du conducteur principal déclaré.

Type du conducteur : il exprime la qualification du conducteur principal déclaré au regard du véhicule (le conducteur principal déclaré est ou non le sociétaire).

Age du conducteur : exprimé en années.

Numéro de département : numéro de département du domicile du conducteur.

Ancienneté de permis : exprimée en années.

Coefficient Bonus Malus ou *CRM*.

Période de couverture : période, en mois, au cours de laquelle l'assuré est couvert par la police qu'il a souscrite, le plus souvent cette période est d'une année (cette variable peut éventuellement servir de pondération).

Garantie choisie (cf. section 1).

3.2 Les caractéristiques du véhicule

Ancienneté de véhicule : elle exprime le millésime de l'année du modèle du véhicule. *Puissance réelle du véhicule* : elle exprime la puissance du moteur en chevaux Din (Deutsch Industrie Normen). Cette mesure donne une vision plus réaliste de la puissance effective au niveau des roues (1 ch. Din = 0,735 Watt).

3.3 Les caractéristiques de la sinistralité actuelle

Pour l'année de référence (2004), on dispose aussi des informations suivantes :

Nombre de sinistre : déclaré à l'assureur pour l'année de référence.

Responsabilité du conducteur : variable binaire qui indique si la responsabilité de l'assuré est engagée, en cas de sinistre.

Montant de dépenses cumulées : lorsque ce montant figure pour un sinistre, il représente les dépenses cumulées depuis l'enregistrement de l'événement ou depuis sa dernière remise en cours. Il est réinitialisé lors d'une remise en cours.

Montant de l'évaluation du sinistre : ce montant correspond à une évaluation *a priori* du montant du sinistre. En effet, pour les accidents corporels, l'assureur peut connaître le montant total de ses dépenses seulement après plusieurs mois, voire plusieurs années, d'où cette nécessité de faire une prévision de ses dépenses.

A toutes ces variables observées par l'assureur, supposées explicatives de la sinistralité, il faut ajouter des variables omises et de l'aléatoire. On peut penser que des facteurs inobservés ou inobservables par l'assureur comme le nombre de kilomètres parcourus, les infractions au code de la route, le comportement au volant, l'état de la route, les conditions météorologiques, peuvent avoir un impact sur la sinistralité.

A propos du nombre de kilomètres parcourus au cours de l'exercice, certains assureurs ont proposé des contrats (« pay as you drive ») dont la prime payée dépend aussi du nombre de kilomètres effectués. Ce système pose deux difficultés : d'une part, l'engagement *a priori* du nombre de kilomètres que l'on va parcourir et d'autre part, le dépassement de ce seuil fixé.

4. UNE MESURE DE LA SINISTRALITÉ

Du fait de l'inversion du cycle de production en assurance, l'assureur se doit d'anticiper le montant des dommages qu'il devra rembourser. La charge financière occasionnée par une police du portefeuille est inconnue au début de la période d'assurance, alors que la prime est déjà encaissée. La prime proposée aux assurés est basée sur des études économétriques et actuarielles des données archivées les années précédentes.

4.1 La sinistralité antérieure : première approche

La sinistralité passée constitue un élément important dans la mesure où elle peut apporter de l'information sur le niveau de risque de l'assuré. Nous essayons donc de mesurer le niveau de risque de chaque conducteur en fonction de l'observation de sa sinistralité passée. Pour cette étude, il nous faudrait des données longitudinales, afin d'évaluer les changements de comportement liés à la sinistralité, toutes choses égales par ailleurs (Abbring *et al.*, 2003a, Abbring *et al.*, 2003b), données que nous ne détenons pas. La prise en compte de données historiques peut permettre effectivement de préciser le sens de causalité entre le choix de contrat et la sinistralité.

Dans un contexte concurrentiel, il est très difficile d'obtenir des données individuelles (certes anonymes) qui concernent à la fois les caractéristiques des cases tarifaires et la sinistralité. En effet, il est toujours possible de trouver, via le web, le tarif proposé par une assurance pour des caractéristiques données. Le rapprochement de ces éléments permettrait d'estimer la rentabilité de chaque case tarifaire et donc de pouvoir faire du dumping et d'attaquer cette entreprise.

Nous disposons uniquement de cette base de données en coupe transversale, qui ne concerne qu'une seule année (2004). Comment alors observer (ou estimer) la sinistralité passée en absence de données longitudinales ?

Nous pouvons aborder ce problème à partir du CRM de chaque assuré, qui fait intervenir le nombre d'accidents antérieurs responsables (et non les montants de ces accidents). Mais le CRM, seul, ne permet pas d'avoir une bonne représentation de la sinistralité passée pour les souscripteurs d'un contrat depuis moins de 13 ans. En effet, un jeune assuré depuis 9 ans sans accident aura un CRM égal à 63, alors qu'un assuré qui avait un CRM à 50 et qui vient d'avoir un sinistre responsable aura un CRM à 62,5. On ne peut pas dire que le premier conducteur a une sinistralité plus importante que le deuxième. Le CRM peut permettre, en tenant compte de l'ancienneté du permis de conduire, d'observer, non pas la fréquence des sinistres passés, mais leur occurrence, c'est-à-dire si l'assuré a déclaré au moins un accident responsable dans le passé ou non.

Nous construisons une variable binaire S_p égale à 1 si l'assuré est considéré comme un bon conducteur et égale à 2 sinon (mauvais conducteur), de la façon suivante :

Notons n_i le nombre d'années d'ancienneté de permis de l'assuré i et CRM_i son coefficient bonus malus.

Plus précisément, la variable S_p est définie ainsi :

$S_p = 1$, si l'on est dans l'un des cas suivants :

- $n_i \geq 13$ et $CRM_i = 50$

- $n_i < 13$ et $CRM_i = E^{n_i}(95)$ (on ne peut pas avoir $CRM_i < E^{n_i}(95)$).

Où $E^{n_i}(95)$ désigne la partie entière de la valeur obtenue à la $n_i^{\text{ème}}$ itération des puissances de 0,95, sachant qu'à chaque étape, on considère également la partie entière, notée E . La relation de récurrence s'écrit :

$$E^{n_i}(95) = E[E^{n_i-1}(95) \times (0,95)] \quad (1)$$

$S_p = 2$, si l'on est dans l'un des cas suivants :

- $n_i \geq 13$ et $CRM_i > 50$
- $n_i < 13$ et $CRM_i = E^{n_i}$ (95)

La construction de cet indicateur de sinistralité passée, proche du CRM, mais qui fait aussi intervenir l'ancienneté de permis, permet d'élargir la classe des bons conducteurs à des conducteurs ayant moins de 13 ans d'ancienneté de permis et n'ayant pas occasionné d'accidents responsables.

Dans notre échantillon aléatoire de 43 245 observations du portefeuille, environ 56 % des assurés ont atteint le bonus maximal², comme le présente le tableau 2 en croisant ces deux critères. Nous n'avons pas comptabilisé, dans les bonus à 0.50, les jeunes conducteurs qui ont bénéficié du CRM des parents, ou les conjoints (sinon, on obtient un pourcentage équivalent à celui de la FFSA).

TABEAU 2
RÉPARTITION (EN POURCENTAGE) DE
L'ÉCHANTILLON EN FONCTION DU CRM ET DE
L'ANCIENNETÉ DU PERMIS DE CONDUIRE

	<i>CRM</i> > 0.50	<i>CRM</i> = 0.50	Total
$n_i < 13$	21,63	0	21,63
$n_i \geq 13$	22,17	56,20	78,37
Total	43,80	56,20	100

Avec cette variable binaire S_p , 65,60 % des assurés de notre échantillon peuvent être considérés comme de bons conducteurs ($S_p = 1$). Cette variable a l'avantage d'être simple dans sa construction et de suivre strictement le système de calcul du CRM, mais elle présente l'inconvénient d'avoir un pourcentage important de mauvais conducteurs (dans cette définition, aucune chance n'est laissée au conducteur qui est responsable d'un accident).

Pour pallier à ce défaut, on propose une deuxième approche.

4.2 La sinistralité antérieure : deuxième approche

Cette deuxième approche a l'avantage d'être plus souple et de permettre à l'assureur de choisir entre plusieurs stratégies. Un indicateur de sinistralité antérieure est construit à partir des écarts au centre de la distribution du CRM dans un groupe donné (ancienneté de permis fixée).

Quelques notations préalables :

j = le nombre d'années d'ancienneté du permis de conduire, où j varie de 1 à 25 (nous ne prenons pas en compte les assurés qui viennent d'avoir leur permis l'année de référence ou qui prennent pour la première fois une assurance l'année de référence). De plus on regroupe ensemble tous les assurés dont l'ancienneté de permis est supérieure ou égale à 25, en effet cela n'a pas de sens de faire une distinction entre ceux, par exemple qui ont 30 ans ou 33 ans de permis.

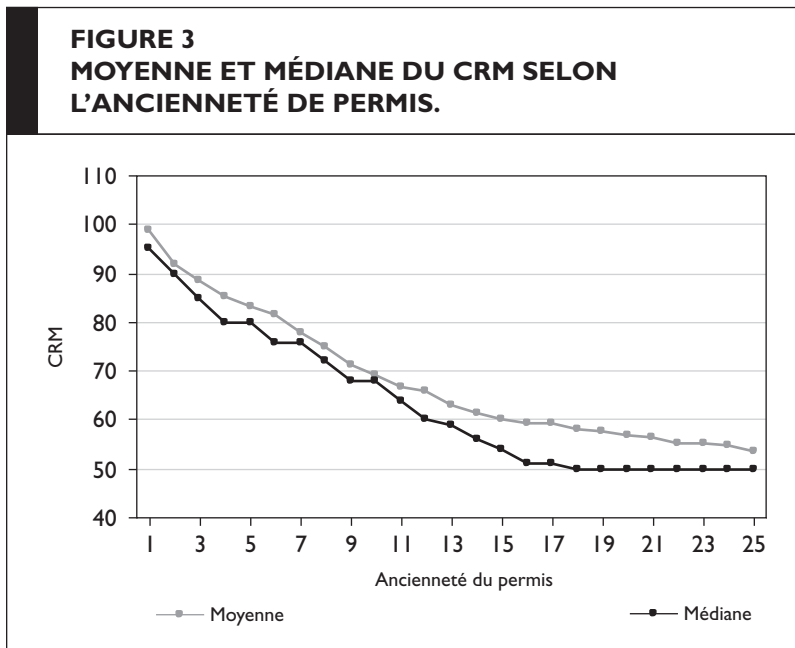
i = l'indice pour un assuré.

$X_{i,j}$ = le Coefficient Réduction Majoration de l'assuré i ayant j années d'ancienneté de permis.

X_j = la moyenne des $X_{i,j}$ pour tous les assurés du groupe j (ancienneté de permis).

$Med(X_j)$ = la médiane des $X_{i,j}$ pour tous les assurés du groupe j .

La figure suivante représente ces deux indicateurs de tendance centrale (les valeurs numériques de ces indicateurs ainsi que les effectifs des groupes sont donnés dans l'annexe 1).



$M_{i,j} = \bar{X}_j - X_{i,j}$ mesure les écarts de CRM d'un assuré i par rapport à la moyenne des assurés de son groupe.

$Med_{i,j} = Med(X_j) - X_{i,j}$ mesure les écarts de CRM d'un assuré i par rapport à la médiane des assurés de son groupe.

Les variables $M_{i,j}$ et $Med_{i,j}$ sont discrètes.

Dans une approche binaire (comparable à l'approche précédente), les bons risques correspondent aux $M_{i,j} \geq 0$ (ou respectivement $Med_{i,j} \geq 0$).

Dans nos données, avec la moyenne (M), on trouve 75,08 % de bas risques et avec les écarts à la médiane (Med), 69,64 %.

4.3 L'historique et le futur de la sinistralité : Quelles stratégies ?

La question initiale « prévisionnelle » peut être remplacée par celle-ci :

La sinistralité passée est-elle un bon indicateur de la sinistralité actuelle, dans la mesure où l'on observe un portefeuille de grande taille et que l'on suppose globalement peu d'évolution des comportements des assurés sur ce point (même si le renforcement des mesures de sécurité routière a permis de réduire le nombre d'accidents) ?

L'avenir étant inconnu, une approche probabiliste s'impose.

Schématiquement, les réponses pourraient être globalement, toutes choses égales par ailleurs :

- Oui, les hauts risques d'hier restent des hauts risques,
- Non, les hauts risques deviennent des bas risques avec l'expérience, l'âge et les changements sociaux,
- Non, tout le monde aura des accidents responsables dans sa vie de conducteur, on se trouve en face d'un processus « uniforme ».

A partir de ces différents éléments, trois stratégies peuvent être proposées à l'assureur :

1. Utiliser l'un de ces trois indicateurs (binaires) pour obtenir une répartition du portefeuille entre bas risques et hauts risques. Ces stratégies ont l'avantage d'être simples et elles sont plus ou moins « sévères » dans la définition des bas risques (bons conducteurs).

2. Envisager une stratégie intermédiaire entre la moyenne et médiane. Effectivement, on peut concevoir une combinaison convexe de ces deux statistiques.

Soit

$$MM = \alpha M + (1 - \alpha) Med \quad (2)$$

avec α entre 0 et 1.

Etant donné que nous sommes dans une situation d'échantillonnage (et il en est de même du portefeuille complet qui n'est pas le même d'une année à l'autre), un choix judicieux de α doit correspondre à la meilleure estimation. Par conséquent une valeur de α qui minimise la variance de MM : $\alpha = \text{Arg min } \text{Var}(MM)$. Comme il s'agit d'une forme quadratique, la solution est évidente :

$$\alpha = \frac{\text{Var}(Med) - \text{Cov}(M, Med)}{\text{Var}(M) + \text{Var}(Med) - 2\text{Cov}(M, Med)} \quad (3)$$

Dès le début du XIX^{ème} siècle, Pierre Simon de Laplace (1818) avait étudié la distribution jointe de la Moyenne et de la Médiane. Ce qui, en l'absence du terme de covariance, revient à définir la « Médienne » ou « Meadian » (Josselin, 2000, Josselin et Ladiray, 2001).

La méthode du bootstrap introduite à la fin des années 70 par Bradley Efron (Efron, 1979) permet d'estimer la précision des estimateurs de paramètres, comme la variance, sans connaître la vraie valeur de cette précision et même une expression de cette précision. Cette méthode procède par ré-échantillonnage (Efron et Tibshirani, 1993, Ardilly, 1994).

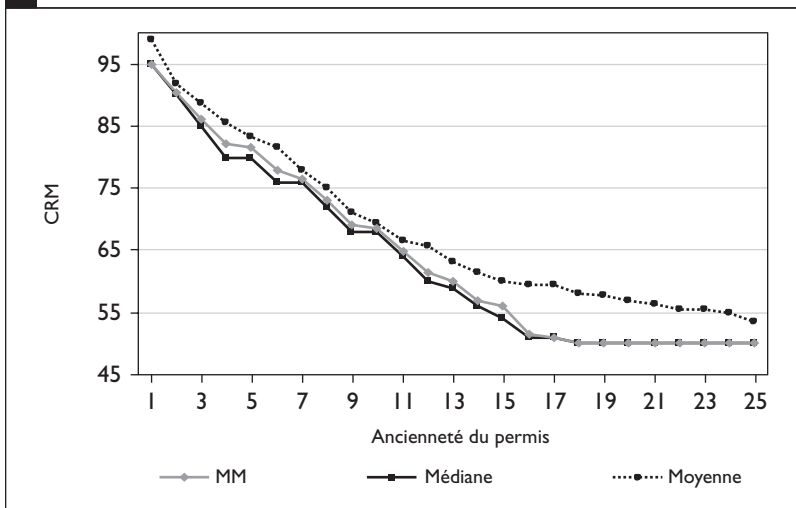
Les variances et la covariance (Beran et Srivastava, 1985) ont donc été estimées par la méthode du bootstrap qui consiste à tirer, avec remise, K échantillons de même taille que l'échantillon considéré. On a choisi $K=1000$. Les temps de mise en œuvre et de calcul sont assez longs.

Pour les différents groupes j (d'ancienneté de permis), la variance de la médiane est plus faible (voire nulle) que celle de la moyenne, par contre la variance de la moyenne n'est jamais nulle. 71 % des assurés ayant plus de 13 ans de permis ont un CRM égal à 50 (médiane).

3. Construire une variable polytomique ordonnée à partir de ces indicateurs pour obtenir une répartition plus fine des risques. Dans cette voie, on définit une variable R de la façon suivante :

$$R_i = \begin{cases} 1 & \text{si } S_p = 1 \\ 2 & \text{si } S_p = 2 \text{ et } Med_i \geq 0 \\ 3 & \text{si } S_p = 2 \text{ et } Med_i < 0 \\ 4 & \text{si } S_p = 2 \text{ et } M_i < 0 \end{cases} \quad (4)$$

FIGURE 4
MOYENNE, MÉDIANE ET MEADIAN DU CRM SELON
L'ANCIENNETÉ DE PERMIS



On peut vérifier l'« inclusion » des trois variables : Si $S_p(i) = 1$ (65,60 %) alors $Med_i \geq 0$ (69,64 %) et si $Med_i \geq 0$ alors $M_i \geq 0$ (75,08 %).

On pourrait aussi proposer à l'assureur une stratégie intermédiaire en construisant une variable R à trois modalités en utilisant MM à la place de la moyenne et de la médiane.

TABLEAU 3
RÉPARTITION DE L'ÉCHANTILLON SELON
LA VARIABLE DE RISQUE R

R	1 (bas risques)	2 (risques faibles)	3 (risques moyens)	4 (hauts risques)
Effectif (pourcentage)	28 370 (65,60)	1 744 (4,03)	2 355 (5,45)	10 776 (24,92)

Cette section 4 avait comme objet de proposer plusieurs approches de la sinistralité passée qui tiennent compte à la fois du CRM et de l'ancienneté de permis, afin de ne pas pénaliser systématiquement les conducteurs ayant moins de 13 ans de permis qui constitueront aussi le portefeuille de demain.

La question posée maintenant est double :

- La sinistralité passée (selon les différentes approches) est-elle un facteur explicatif de la sinistralité actuelle (année de référence) ?
- Si oui, quel est son impact par rapport aux autres caractéristiques du couple (assuré, véhicule) ?

Et par la même occasion, on revisite la question de l'anti-sélection.

5. MODÉLISATION

De façon générale, une analyse des comportements reste souvent incomplète quand on se limite à l'observation de tableaux croisés qui répartissent la sinistralité selon un ou plusieurs facteurs. L'assureur utilise des approches analytiques (fichier d'apprentissage des années précédentes contenant les caractéristiques des conducteurs et des véhicules assurés, ainsi que des informations sur la sinistralité) et des approches statistiques et économétriques (analyses multidimensionnelles, modèles Logit, modèles de comptage, équations simultanées,...).

Un modèle est une représentation simplifiée d'une réalité et s'appuie sur une ou plusieurs équations déterministes ou probabilistes. L'intérêt pragmatique d'un modèle réside dans sa capacité à apporter une réponse satisfaisante à une problématique posée (Vasyechko et *al.*, 2008). Pour cette partie de modélisation, après de nouveaux contrôles sur les données, la taille de l'échantillon est ramenée à 42 853 observations. Tous les résultats numériques peuvent être obtenus auprès des auteurs.

Notations

Pour chaque assuré i ,

C correspond à la variable binaire du choix de contrat :

$$C_i = \begin{cases} 0 & \text{si l'assuré } i \text{ a un contrat } RC \\ 1 & \text{si l'assuré } i \text{ a un contrat } DV \end{cases}$$

Y désigne la variable binaire de sinistralité actuelle :

$$Y_i = \begin{cases} 0 & \text{si l'assuré } i \text{ n'a déclaré aucun accident responsable} \\ 1 & \text{si l'assuré } i \text{ a déclaré au moins un accident responsable} \end{cases}$$

$X = (X_1, X_2)$ le vecteur des caractéristiques du couple (assuré, véhicule), et ε la composante résiduelle (terme d'erreur).

On trouve que 15 % des assurés de notre échantillon ont eu un accident responsable (totalement ou en partie) durant l'année de référence.

5.1 Modèles Logit

L'économétrie des variables qualitatives s'est considérablement développée depuis le début des années 80 et les assureurs ont utilisé des modèles Logit ou Probit pour expliquer la sinistralité (occurrence d'accidents). Le modèle Logit (utilisé pour la première fois par Mc Fadden, 1976) donne une plus grande importance aux valeurs extrêmes, mais il a l'avantage d'être plus facile à interpréter et à communiquer.

Modèle Logit avec la variable exogène S_p

Pour expliquer la variable binaire de sinistralité, nous utilisons classiquement un modèle de régression logistique.

Soit $p = \Pr(Y = 1 / X_1, X_2, C, S_p)$, le modèle s'écrit :

$$\text{Log}\left(\frac{p}{1-p}\right) = \alpha + \beta_1' X_1 + \beta_2' X_2 + \beta_3' C + \beta_4' S_p + \varepsilon \quad (5)$$

où α est un paramètre constant et β_1 le vecteur de pente à estimer. La modalité de référence pour la variable endogène est l'absence de sinistre.

« Toutes choses égales par ailleurs », la probabilité d'avoir au moins un accident responsable durant l'année de référence augmente avec l'ancienneté du véhicule, en zone urbaine et la sinistralité passée (19 % vs 13 %). Elle diminue avec la puissance du véhicule et la garantie DV (11 % vs 20 %) (cf. annexe 2). Les autres variables ne sont pas significatives au seuil de 5 %. On obtient 64 % de paires concordantes. On ne valide pas empiriquement l'hypothèse d'anti-sélection, puisque la probabilité de sinistralité est plus importante pour les assurés ayant choisi une garantie minimale.

Modèle Logit pour la variable exogène polytomique ordonnée R

L'équation du modèle est la même que précédemment (5), avec $p = \Pr(Y = 1 / X_1, X_2, C, R)$.

Bien évidemment les résultats sont semblables à ceux de la régression précédente et la variable R de sinistralité passée explique la probabilité de sinistralité actuelle (dans le même sens, cf. annexe 3). La variable sexe est maintenant significative (au seuil de 5 % et non de 1 %) : les hommes ont plus de comportements à risque. On obtient 66 % de paires concordantes.

Les variables explicatives R et S_p sont comparables. Qu'apporte cette variable R par rapport à la variable binaire S_p ?

Les trois autres catégories d'assurés ont une probabilité de sinistralité plus importante que les bas risques, conformément aux résultats précédents, mais ce qui est plus surprenant, c'est l'importance des risques « faibles » et « moyens ». La classe des risques « faibles » qui ne représente que 4 % des données, est caractérisée par de « jeunes conducteurs (moyenne de 10 ans de permis avec un écart type de 2, contre plus de 16 ans de moyenne pour les autres classes) et qui ont à 70 % des contrats DV (vs 50 % pour la catégorie des bas risques). Une sensibilisation préventive à la sinistralité pourrait être mise en place de la part de l'assureur pour cette catégorie d'assurés. Un phénomène d'aléa moral peut se produire dans cette classe. La possession de données longitudinales serait nécessaire pour faire la distinction entre les effets d'anti-sélection et d'aléa moral (Dionne et al., 2006).

Les caractéristiques de la classe des risques moyens sont proches de la catégorie des hauts risques.

Remarques

Dans un modèle Logit, les équations de vraisemblance ne sont pas linéaires pour les paramètres qui doivent être estimés en utilisant des algorithmes. La convergence et la normalité asymptotique ne sont analysées que sous certaines conditions sur les variables explicatives. En particulier, les variables explicatives sont supposées exogènes et les variables omises doivent être presque orthogonales aux résidus. Rien ne permet *a priori* de valider ces hypothèses. Par exemple, le choix de garantie ne peut être supposé *a priori* indépendant de la gamme du véhicule (Grun-Rehomme, Joly, 2003). Il peut aussi dépendre de X_2 (par exemple, on contracte plus souvent une assurance « tous risques » pour un véhicule neuf qu'un véhicule ancien). Avec la franchise, la garantie « tous risques » ne sert pas à grand-chose pour une voiture ancienne ayant un kilométrage élevé. Utiliser un tel modèle peut introduire des biais dans les estimations des paramètres. Le choix de garantie pourrait aussi refléter l'influence de caractéristiques individuelles inobservées.

5.2 Modèles Logit bivariés

Un modèle bivarié récursif qui estime simultanément la sinistralité et le choix de garantie permet de tester l'endogénéité du choix de garantie sur la sinistralité (des facteurs peuvent affectés à la fois le choix de garantie et la sinistralité) et donne des résultats robustes quant à l'estimation des probabilités de sinistre. Cette modélisation est utilisée pour déterminer ce qui des caractéristiques de l'assuré ou du véhicule et du choix de garantie est le plus déterminant dans la probabilité de sinistre ou de non sinistre.

Le modèle s'écrit :

$$\begin{aligned} C^* &= k_1 + \beta_1' X_1 + \beta_2' X_2 + \beta_3' S_p + \varepsilon \\ Y^* &= k_2 + \gamma_1' X_1 + \gamma_2' X_2 + \gamma_3' C + \gamma_4' S_p + e \end{aligned} \quad (6)$$

où Y^* et C^* sont deux variables latentes continues associées respectivement à Y et C , pour lesquelles on observe $Y = I(Y^* > 0)$ et $C = I(C^* > 0)$, où I désigne la fonction caractéristique. Il en est de même pour un tel modèle avec la variable R à la place de S_p .

Dans ces modèles, les erreurs sont supposées corrélées :

$$\begin{pmatrix} \varepsilon \\ e \end{pmatrix} \rightarrow N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right] \quad (7)$$

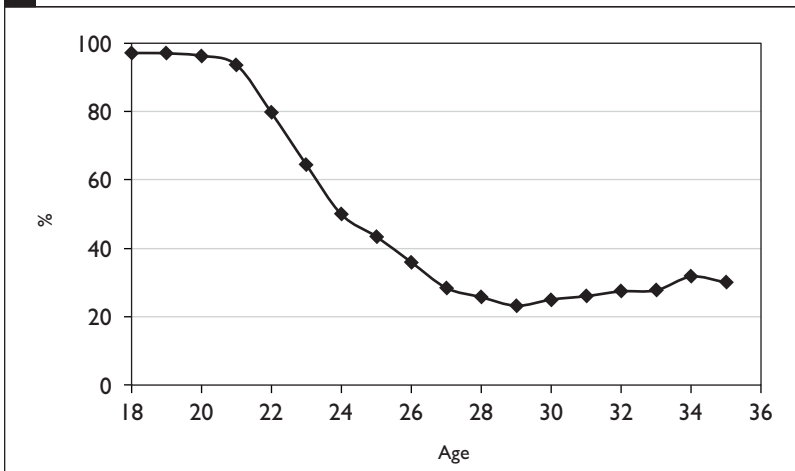
où σ_1 (resp. σ_2) est l'écart-type du terme d'erreur ε (resp. e) et ρ le coefficient de corrélation entre les résidus des deux équations. Il mesure la corrélation entre les variables omises (inobservables ou inobservées) dans les deux équations (6) qui influent sur les variables endogènes. Dans ce problème, les paramètres peuvent être déterminés à une constante multiplicative près. On impose alors aux termes d'erreur d'avoir une variance unitaire ($\sigma_1 = \sigma_2 = 1$). Les paramètres du modèle sont estimés à l'aide d'un algorithme de Newton.

Le signe des coefficients du modèle indique si la variable associée influe à la hausse ou à la baisse sur la probabilité de l'événement considéré (accident). Rappelons que la valeur numérique d'un coefficient n'a pas d'intérêt. Il est possible de calculer l'effet marginal associé à chaque variable explicative. Si on désigne par X le vecteur des variables explicatives x^j et par β le vecteur des paramètres associés, l'effet marginal associé à la $j^{\text{ème}}$ variable explicative x^j est défini

$$\text{par : } \frac{\partial p_i}{\partial x_i^j} = \frac{e^{-x_i \beta}}{(1 + e^{-x_i \beta})^2} \beta_j, \text{ où } p_i = \text{Prob}(Y_i = 1).$$

A vrai dire, les vecteurs X_1 et X_2 ne sont pas exactement les mêmes dans les deux équations. Pour des raisons d'identification, l'équation de la sinistralité doit comporter au moins une variable différente de celle de l'équation du choix de contrat (Maddala, 1983, p. 122). Cette variable est dite instrumentale, son choix n'est pas aléatoire mais doit être justifié. Une variable instrumentale est supposée agir sur le choix de contrat, mais pas directement sur la sinistralité. Elle assure ainsi l'identification non paramétrique des modèles à équations simultanées. La variable instrumentale retenue est le type de conducteur (l'assuré est ou n'est pas le conducteur principal). En effet, le type de conducteur affecte directement le choix de garantie. D'une part, si l'assuré est le conducteur principal, la garantie choisie ne dépendra que de ses caractéristiques propres. A l'inverse, si l'assuré n'est pas le conducteur principal, il choisira un contrat qui tient compte, simultanément de ses caractéristiques et des caractéristiques liées à ce conducteur principal (par exemple, l'un des enfants de l'assuré, ou le conjoint). Le pourcentage de conducteurs principaux qui ne sont pas les assurés varie avec l'âge (cf. figure 1). Il est très important pour les moins de 22 ans (on peut penser qu'il s'agit des enfants des assurés), puis il diminue avec l'âge du conducteur principal pour se stabiliser autour de 30% au-delà de 35 ans. Le conducteur principal devenant, le plus souvent, le conjoint de l'assuré. Dans cette mutuelle d'assurance, il est avantageux pour l'assuré de procéder ainsi.

FIGURE 5
RÉPARTITION, EN POURCENTAGES,
DES CONDUCTEURS PRINCIPAUX QUI
NE SONT PAS LES ASSURÉS SELON L'ÂGE



Cette variable instrumentale, type de conducteur, affecte la variable explicative endogène (choix de garantie), dont tout l'effet sur la variable à expliquer (la sinistralité) transite par son effet sur la variable explicative endogène. Cette variable possède aussi la propriété de ne pas être corrélée avec les résidus des deux équations.

Dans les deux modèles bivariés (avec S_p ou avec R), les facteurs explicatifs du choix de contrat et de la sinistralité actuelle sont les mêmes. Ce qui est cohérent, puisque les deux modèles ne changent que sur la définition de la sinistralité passée, ou plus exactement sur deux modes d'estimation du même facteur endogène de sinistralité passée (en l'absence de données longitudinales).

Avec la variable S_p

On peut de cette modélisation bivariée en déduire les principaux éléments suivants («toutes choses égales par ailleurs» avec la variable (cf. annexe 4):

- Pour l'équation du choix de contrat, toutes les variables sont significatives. La probabilité de choisir un contrat DV est plus importante pour les femmes (5,5 %), et elle augmente avec l'ancienneté du véhicule et sa puissance. La probabilité de prendre un contrat RC est plus importante pour les jeunes conducteurs, lorsque l'assuré est le conducteur principal et pour les assurés des zones rurales. On a considéré deux zones géographiques : rurales et urbaines (critère de densité). Cette probabilité de contrat RC est plus importante pour les mauvais conducteurs.
- Pour l'équation de la sinistralité, le choix de contrat n'est pas une variable significative. La probabilité d'avoir au moins un accident responsable augmente avec l'âge du conducteur, la puissance du véhicule et la sinistralité antérieure (+7,2 %), elle diminue pour les véhicules anciens en zone rurale et pour les femmes. La sinistralité antérieure se présente comme un bon indicateur explicatif de la sinistralité future.

Du fait que cet indicateur de sinistralité passée a un bon pouvoir prédictif de la sinistralité actuelle, l'assureur pourrait l'utiliser pour accorder une réduction de prime à ses assurés afin de les fidéliser. Ils constitueront son portefeuille dans le futur. Les mauvais conducteurs restent de mauvais conducteurs. Ce résultat va à l'encontre de l'idée qu'avec l'âge, l'expérience de la conduite, le changement de situation familiale et professionnelle, les conducteurs deviennent plus responsables, plus respectueux des codes de la société et des autres.

Dans cette modélisation, le coefficient ρ de corrélation positif entre les termes d'erreur n'est pas significatif au seuil de 5 % (mais à 10 %). Deux hypothèses peuvent être émises :

- L'estimation séparée des deux équations n'est pas utile, on ne trouve pas de biais d'endogénéité et l'information détenue par l'assureur est suffisante pour expliquer la sinistralité.
- Les classes des variables binaires (Choix de contrat et sinistralité) sont trop hétérogènes en leur sein ou trop semblables entre elles.

Avec la variable R

Les conclusions sur les variables explicatives, dans les deux équations, sont les mêmes. Il faut simplement noter que dans l'équation de la sinistralité, l'âge du conducteur n'est plus significatif au seuil de 5 % (mais seulement 10 %). Les effets marginaux sont du même ordre pour les deux modèles, mais la variable *R* (« plus fine ») permet d'obtenir des effets marginaux significativement différents pour les différentes classes de risque pour le choix de contrat et de façon moins différenciée sur la sinistralité actuelle (cf. annexe 6). L'effet marginal de la variable de sinistralité passée (environ 9 %) sur la sinistralité actuelle est plus important que celui des autres variables exogènes retenues dans le modèle.

Une différence notable entre les deux modèles provient du coefficient ρ , qui est maintenant significatif au seuil de 5 % (seuil habituellement retenu en économie). On peut donc en conclure que les facteurs omis (variables inobservées ou inobservables) qui expliquent la sinistralité et le choix de garantie sont corrélés. Dans ce cas, l'estimation autonome de la seconde équation peut comporter un biais d'endogénéité et les coefficients des variables explicatives de cette équation de la sinistralité actuelle sont sous estimés. Et comme ρ est positif, les caractéristiques individuelles des conducteurs qui expliquent le choix de garantie, expliquent aussi positivement la probabilité de sinistralité.

De nombreux auteurs (Chiappori, 2000, Chiappori et Salanié, 2000, Chiappori et *al.*, 2004) ont souligné, qu'en se basant sur des données en coupe instantanée décrivant une seule période, une corrélation positive entre la couverture d'assurance et les sinistres déclarés conditionnellement aux variables observables permet seulement d'avoir une idée globale de la présence d'asymétrie d'information. Cette corrélation n'est pas simplement expliquée par l'anti-sélection, mais aussi par l'aléa moral. Avec des données statiques, il n'est pas facile d'identifier le sens de la causalité entre le risque et la couverture. Des données de panel décrivant plusieurs années, plusieurs périodes d'assurance permettent d'identifier ces deux formes d'asymétrie d'information (Dionne, 1998, Chiappori, 2000, Chiappori et *al.*, 2004). En utilisant des données françaises, Dionne et *al.* (2006)

valident empiriquement l'hypothèse d'aléa moral et distinguent cette asymétrie d'information de l'anti-sélection en utilisant un test de causalité de Granger (1969).

6. CONCLUSION

L'utilisation d'un modèle bivarié a permis de confirmer que l'estimation séparée des équations de choix de contrat et de sinistralité comporte des biais d'endogénéité (Chiappori et Salanié, 2000, Cohen, 2005, Grun-Rehomme et Benlagha, 2007, Vasyechko, 2008). De plus, on invalide empiriquement l'hypothèse d'anti-sélection conformément aux études récentes antérieures (Chiappori et Salanié, 2000, Dionne et al., 2001 *a, b*, Grun-Réhomme et Joly, 2003, Cohen, 2005, Saito, 2008). Le choix de garantie traduit plutôt l'aversion au risque, l'aversion à la perte de l'investissement (valeur du véhicule) que l'assuré a réalisé. Il cherche à protéger cet investissement. A ce niveau, les risques relatifs de chaque individu, que ce soit l'information « officielle » du CRM ou le risque estimé dans ce texte, entrent peu en jeu. L'aversion au risque de l'individu correspond non pas à l'espérance mathématique de la perte, mais au niveau de perte maximal. On retrouve ainsi les résultats de Dionne et al. (2001), à savoir que les bas risques ont plus d'aversion pour le risque que les hauts risques.

La corrélation positive et significative entre les termes d'erreur signifie premièrement que le choix de garantie est bien une variable endogène et deuxièmement que toutes choses observables égales par ailleurs, les assurés qui choisissent un contrat RC ont en moyenne des probabilités plus élevées de sinistralité. Cette corrélation entre les équations laisse à penser que l'asymétrie d'information existante entre l'assuré et l'assureur ne se traduit pas en assurance automobile par l'anti-sélection et l'aléa moral (même si nous ne disposons pas dans cette étude de données longitudinales, les fraudes sont financièrement marginales par rapport aux sinistres graves corporels), mais par des facteurs inobservés ou inobservables. Ces derniers peuvent expliquer le choix de garantie et la sinistralité. Plusieurs éléments méritent de retenir notre attention et des études prospectives pourraient être réalisées en ce sens.

Si le montant de la prime était le même pour un contrat DV que pour un contrat RC, *a priori* tout le monde opterait pour un contrat DV, mais tel n'est pas le cas. Il existe un seuil à partir duquel on ne peut plus souscrire un contrat « tous risques ». En effet, on peut sous-

crir un niveau de couverture élevé si on est assez riche pour payer la prime correspondante et en même temps craindre la perte financière liée à l'absence de ce niveau de couverture. Il faudrait mettre en rapport le niveau de cotisation et le niveau de revenu. On peut également penser à faire intervenir le patrimoine de l'assuré, à la place de son revenu, mais cette information est plus difficile à obtenir et elle doit être corrélée avec le revenu. Une étude sur l'élasticité de la prime pourrait être réalisée, en utilisant la richesse de l'assuré comme variable de contrôle.

Autre élément que l'on pourrait prendre en considération est l'effet de prescription du réseau : au moment de la souscription, le choix est influencé par le gestionnaire du contrat; les assurés sont d'ailleurs très demandeurs de conseils à ce moment là et dans ce cas, l'employé de l'assurance préfère souvent proposer un contrat « tous risques » afin d'éviter, en cas de sinistre, le mécontentement de l'assuré.

Effectivement il existe bien une information d'asymétrie entre l'assureur et l'assuré (l'assureur ne sait pas tout de l'assuré, comme le psychanalyste ne sait pas tout du vécu de son patient; la comparaison s'arrête là). Il nous semble donc que ces hypothèses d'anti-sélection et d'aléa moral sont secondaires en assurance automobile par rapport aux facteurs socio-économiques et psychologiques mentionnés précédemment.

L'accident responsable est-il une manifestation de l'Inconscient ?

D'autre part l'approche de la sinistralité doit être ouverte et multiforme, et en ce sens, il n'y a pas une méthode pour un problème. Un modèle a ses limites et ne peut donner qu'une image imparfaite de la réalité.

Références

- Abbring J., Chiappori P-A., Heckman J., Pinquet J. (2003a) *Adverse selection and moral hazard in insurance: can dynamic data help to distinguish?* Journal of the European Economic Association, 1, 512-521.
- Abbring J., Chiappori P-A., Heckman J., Pinquet J. (2003b) *Moral hazard and dynamic insurance data*, Journal of the European Economic Association, 1(4), 767-820.
- Ardilly P. (1994), *Les techniques de sondage*, Ed. Technip.
- Beran R., Srivastava M.S. (1985), Bootstrap tests and confidence regions for functions of a covariance matrix, *The Annals of Statistics*, 13(1), 95-115.
- Bernoulli D. (1738), *Specimen Theoriae Novae de Mensura Sortis*, *Comentarii Academiae Scientiarum Imperialis Petropolitanae*, 5175-192, translated by L. Sommer in *Econometrica*, 1954, 22, 23-36.

- Chiappori P-A., Salanié B. (2000), *Testing for Asymmetric Information in Insurance Markets*, Journal of Political Economy, 108, 1, 56-78.
- Cohen A. (2005), *Asymmetric information and learning: evidence from the automobile insurance market*, The Review of Economics and Statistics, 87, 197-207.
- Dionne G., Vanasse C. (1992), *Automobile Insurance Ratemaking in the Presence of Asymmetrical Information*, Journal of Applied Econometrics, 7, 149-165.
- Dionne G., Doherty N., Fombaron N. (2001 a), *Adverse Selection in Insurance Markets*, Handbook of Insurance, Kluwer Academic Publishers, Boston, 185-243.
- Dionne G., Gouriéroux C., Vanasse C. (2001 b), *Testing for Evidence of Adverse Selection in the Automobile Insurance Market*, Journal of Political Economy, 109, 2, 444-453.
- Dionne G., Michaud P. C., Dahchour M. (2006), *Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France*, American Risk and Insurance Association (ARIA), Meeting 2006, Washington.
- Efron B. (1979), *Bootstrap methods: another look at the Jackknife*, The Annals of Statistics, 7, 1-26.
- Efron B., Tibshirani R. J. (1993), *An introduction to the Bootstrap*, New York, Chapman and Hall, 436.
- Granger C. (1969), *Investigating Causal Relations by Econometrics Models and Cross-spectral Methods*, Econometrica, 37, 3, 424-438.
- Grun-Rehomme M., Joly V. (2003), *Risque individuel et choix de contrat : Le cas de l'assurance automobile*, revue Assurances et gestion des risques, HEC Montréal, Canada, 71(1), 145-162.
- Grun-Rehomme M., Benlagha N. (2007), *Choix de contrat et sinistralité chez les jeunes conducteurs*, revue Assurances et gestion des risques, HEC Montréal, Canada, 74(4), 505-532.
- Henriet D., Rochet J.C. (1991), *Microéconomie de l'assurance*, Economica.
- Josselin D. (2000), *Méthodologies d'analyse exploratoire des données géographiques : tout centrer sur les distributions statistiques et spatiales et les liens dynamiques*, Actes du colloque PRISM-INRETS sur le Data Mining Spatial, 24-25 février 2000.
- Josselin D., Ladiray D. (2001), *Combining L1 and L2 Norms for a more Robust Spatial Analysis : the Median Attitude*, European Colloquium on Theoretical and Quantitative Geography, Saint-Valery en Caux, Cybergeo, September 2001.
- Laplace P. S. de (1818), *Deuxième supplément à la Théorie Analytique des Probabilités*, Courcier, Paris.
- Mc Fadden D. (1976), *A comment on discriminant analysis versus Logit analysis*, Annals of Econ. and Soc. Meas., 5, 511-524.
- Maddala G. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.
- Rothschild M., Stiglitz J. (1976), *Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*, Quarterly Journal of Economics, 90, 629-649.

- Saito K. (2008), *Testing for Asymmetric Information in the Automobile Insurance Market under rate Regulation*, The Journal of Risk and Insurance, 73(2), 335-356.
- Spence M.A. (1978), *Product Differentiation and Performance in Insurance Markets*, Journal of Political Economy, 10, 427-447.
- Vasechko O., Grun-Rehomme M., Benlagha N. (2008), *Sélection adverse, sinistralité et choix de contrat chez les jeunes conducteurs : La pertinence des modèles*, Risques, Les cahiers de l'assurance, 75, 133-141.
- Vasechko O., Grun-Rehomme M. (2010), *L'impact de la sinistralité passée sur la sinistralité future : approche empirique en assurance automobile*, revue Assurances et gestion des risques, HEC Montréal, Canada, vol. 78(1-2), 71-91, 2010.
- Wilson C.A. (1977), *A Model of Insurance Markets with Incomplete Information*, Journal of Economic Theory, 16, 167-207.
- Young G., Valdez E., Kohn R. (2009), *Multivariate probit models for conditional claim-types*, Insurance: Mathematics and Economics, 44(2), 214-228.

**ANNEXE I
PROPRIÉTÉS STATISTIQUES DU CRM SELON
L'ANCIENNETÉ DU PERMIS**

Ancienneté de permis (années)	Effectif	Moyenne	Médiane	Ecart type
1	641	99.65	95	17.71
2	626	91.88	90	14.00
3	726	88.68	85	14.43
4	697	85.50	80	14.54
5	669	83.17	80	14.98
6	762	81.60	76	17.22
7	790	77.88	76	15.67
8	754	75.19	72	16.70
9	847	71.14	68	16.11
10	842	69.41	68	15.42
11	990	66.64	64	15.30
12	1012	65.81	60	15.9
13	998	63.07	59	15.32
14	990	61.46	56	15.02
15	909	60.04	54	15.04
16	945	59.43	51	14.49
17	934	59.36	51	14.05
18	958	57.97	50	14.03
19	1032	57.64	50	13.37
20	1017	56.92	50	13.85
21	1076	56.38	50	13.65
22	1024	55.38	50	11.41
23	1154	55.36	50	12.41
24	1287	54.89	50	12.02
25	21565	53.44	50	11.10
Total	43245	60.04	50	16.47

**ANNEXE 2
RÉSULTATS DE LA RÉGRESSION LOGISTIQUE SIMPLE
(SINISTRALITÉ ACTUELLE) AVEC LA VARIABLE
BINAIRE DE SINISTRALITÉ PASSÉE S_p**

Estimation des paramètres (significatifs) du modèle (méthode stepwise)

Variabes	Coefficient	Std Error	Wald	Odds Ratio
Constante	1.3441	0.0614	478.90 **	
Garantie DV	-0.5653	0.0290	379.35 **	0.601
Puissance du véhicule	-0.00211	0.00052	16.48 **	0.998
Ancienneté du véhicule	0.0659	0.00335	388.13 **	1.068
Zone urbaine	0.1658	0.0355	21.86 **	1.180
Sinistralité passée	0.5619	0.0284	391.57 **	1.754

(*) Significatif au seuil de 5%, (**) au seuil de 1%.

**ANNEXE 3
RÉSULTATS DE LA RÉGRESSION LOGISTIQUE SIMPLE
(SINISTRALITÉ ACTUELLE) AVEC LA VARIABLE POLY-
TOMIQUE ORDONNÉE DE SINISTRALITÉ PASSÉE R**

Variabes	Coefficient	Std Error	Wald	Odds Ratio
Constante	1.1618	0.0623	351.80 **	
Garantie DV	-0.5729	0.0292	384.31 **	0.597
Homme	0.0593	0.0283	4.38 *	1.122
Puissance du véhicule	-0.00246	0.00053	21.05 **	0.998
Ancienneté du véhicule	0.0657	0.00337	380.52 **	1.075
Zone urbaine	0.1652	0.00356	21.57 **	1.265
Risques faibles	0.7467	0.0305	600.20 **	2.110
Risques moyens	0.8269	0.0810	104.20 **	2.286
Hauts risques	0.7479	0.0675	122.88 **	2.113

(*) Significatif au seuil de 5%, (**) au seuil de 1%.

**ANNEXE 4
RÉSULTATS DU MODÈLE PROBIT BIVARIÉ
(AVEC LA VARIABLE S_p)**

Estimation des paramètres du modèle

Variabes	Coefficient	Std Error	t Value
Equation du choix de contrat			
Constante	-0.549102	0.034811	-15.77 ***
Age du conducteur	0.026964	0.000503	53.56 ***
Femme	-0.172675	0.013546	-12.75 ***
Le conducteur est l'assuré	0.029086	0.013776	2.11 **
Puissance du véhicule	-0.003605	0.000263	-13.71 ***
Ancienneté du véhicule	-0.089924	0.001519	-59.18 ***
Zone urbaine	0.095850	0.016187	5.92 ***
Sinistralité passée	0.276163	0.014404	19.17 ***
Equation de la sinistralité actuelle			
Constante	-0.756312	0.061053	-12.39 ***
Garantie DV	0.063640	0.144680	0.44 n.s.
Age du conducteur	0.003136	0.001433	2.19 **
Femme	-0.045058	0.017700	-2.55 **
Puissance du véhicule	0.000956	0.000351	2.73 ***
Ancienneté du véhicule	-0.044287	0.004474	-9.90 ***
Zone rurale	-0.079901	0.019675	-4.06 ***
Sinistralité passée	0.291171	0.023119	12.59 ***
Rho	0.144834	0.086156	1.68 *

(*) Significatif au seuil de 10%, (**) au seuil de 5%, (***) au seuil de 1%.

n.s. : non significatif

**ANNEXE 5
RÉSULTATS DU MODÈLE PROBIT BIVARIÉ
(AVEC LA VARIABLE R)**

Estimation des paramètres du modèle

Variabes	Coefficient	Std Error	t Value
Equation du choix de contrat			
Constante	-0.605654	0.035957	-16.84 ***
Age du conducteur	0.027161	0.000509	53.35 ***
Femme	-0.172497	0.013554	-12.73 ***
Le conducteur est l'assuré	0.029090	0.013779	2.15 **
Puissance du véhicule	-0.003636	0.000263	-13.80 ***
Ancienneté du véhicule	-0.090064	0.001528	-58.94 ***
Zone urbaine	0.095677	0.016194	5.91 ***
Risques faibles	0.326708	0.015982	20.44 ***
Risques moyens	0.142218	0.036085	3.94 ***
Hauts risques	0.212341	0.030593	6.94 ***
Equation de la sinistralité actuelle			
Constante	-0.624577	0.058560	-10.67 ***
Garantie DV	0.035372	0.140414	0.25 n.s.
Age du conducteur	0.002738	0.001456	1.88 *
Femme	-0.048149	0.017569	-2.74 ***
Puissance du véhicule	0.000988	0.000350	2.82 ***
Ancienneté du véhicule	-0.046323	0.004325	-10.48 ***
Zone rurale	-0.078234	0.019967	-3.92 ***
Risques faibles	0.380585	0.026116	14.57 ***
Risques moyens	0.437961	0.044165	9.92 ***
Hauts risques	0.391306	0.038736	10.10 ***
Rho	0.168169	0.083324	2.02 **

(*) Significatif au seuil de 10%, (**) au seuil de 5%, (***) au seuil de 1%.

n.s. : non significatif

ANNEXE 6 EFFETS MARGINAUX DES VARIABLES EXOGÈNES DANS LES DEUX MODÈLES PROBIT BIVARIÉS

Variables	Equation du choix de contrat		Equation de la sinistralité	
	(avec S_p)	(avec R)	(avec S_p)	(avec R)
Age du conducteur	0.00904	0.00910	0.00072	0.00062
Femme	-0.05790	-0.05780	-0.01026	-0.01091
Le conducteur est l'assuré	0.00220	0.00211	(*)	(*)
Puissance du véhicule	-0.00120	-0.00122	0.00021	0.00022
Ancienneté du véhicule	-0.03017	-0.03017	-0.01008	-0.01027
Zone urbaine	0.03216	0.03206	-0.01826	-0.01773
Sinistralité passée (S_p)	0.09265	-	0.06629	-
Risques faibles	-	0.10947	-	0.08628
Risques moyens	-	0.04762	-	0.09929
Hauts risques	-	0.07115	-	0.08868

(*) Cette variable est un instrument dans les modèles bivariés

n.s. : non significatif

Notes

1. Exposé d'une théorie nouvelle sur l'évaluation du risque, traduit du latin en anglais par Louise Sommer, American University, Washington D.C., de « Specimen Theoriae Novae de Mensura Sortis », Commentaires de l'Académie Impériale des Sciences de Petersbourg, tome V 1738, pages 175-192. Ce texte a été ensuite traduit en français dans la Revue de Statistique Appliquée, 19(3), 5-18, 1971.

On trouve aussi un texte en français, plus proche de la version originale, traduit directement du latin par R. Charreton et présenté dans les Cahiers du Séminaire d'Histoire des Mathématiques, tome 6(1985), pp. 61-77.

On peut ajouter que dans ce texte, Bernoulli, pour répondre au « paradoxe de St Petersburg » introduit une fonction d'utilité et considère le cas d'un joueur ayant une aversion au risque à travers la concavité de sa fonction d'utilité et l'espérance mathématique d'une variable aléatoire de gain dans ce jeu.

2. Selon les chiffres publiés en l'année 2008 par la Fédération Française des Sociétés d'Assurance (FFSA), 64 % des assurés automobiles français ont atteint le bonus maximal, soit un coefficient de réduction – majoration égal à 50.