

SUR L'EFFICACITE DES CRITERES DE TARIFICATION
DE L'ASSURANCE CONTRE LES ACCIDENTS
D'AUTOMOBILES

PIERRE DELAPORTE
Paris (France)

Les accidents d'automobiles surviennent lorsqu'il se produit un ensemble de circonstances défavorables plus ou moins fortuites. Ces circonstances sont liées au véhicule, à son conducteur, à l'usage qui en est fait, à la région dans laquelle il circule et à beaucoup d'autres faits souvent difficiles à caractériser.

L'assureur ne peut évidemment connaître qu'une faible partie de ces caractéristiques, c'est pourquoi il considère la survenance de l'accident comme un fait aléatoire. Cependant, les dépouillements statistiques des nombres d'accidents arrivés à des véhicules ayant un même ensemble de caractéristiques montrent que la fréquence moyenne, c'est-à-dire le nombre d'accidents survenus divisé par le nombre de voitures-année, varie avec cet ensemble de caractéristiques communes qui constitue alors une classe du tarif d'assurance. La détermination du montant de la prime à demander pour garantir le véhicule est alors faite par l'assureur en demandant une prime hiérarchisée selon la fréquence moyenne et selon le coût moyen des sinistres de la classe du tarif.

Si l'on ajoutait de nouveaux critères de tarification à ceux déjà utilisés, il est évident qu'on trouverait une diversification plus grande des primes; ceci montre donc qu'à l'intérieur d'une classe de tarif tous les risques des véhicules ne sont pas égaux.

Dans la présente étude, nous rechercherons d'abord quelle part de l'hétérogénéité des risques des voitures est utilisée par les diverses caractéristiques figurant dans le tarif français d'assurance de la responsabilité civile des propriétaires de véhicule. Nous rechercherons alors s'il existe une limite à la précision du tarif.

Pour cette recherche, nous montrerons que la survenance des

sinistres aux véhicules d'une classe de tarif obéit à une loi de probabilité composée, les lois composantes étant l'une la loi de probabilité des risques des voitures à l'intérieur de la classe de tarif et l'autre la loi de probabilité de la survenance au hasard des sinistres.

Nous avons donné les premiers moments de chacune des lois de probabilité composantes et nous avons établi au moyen de dépouillements statistiques faits sur plusieurs classes du tarif que la loi de probabilité de la survenance au hasard des sinistres est la loi de Poisson, ce qui était seulement une hypothèse habituellement admise.

Cette loi de probabilité de la survenance au hasard des sinistres permet de déterminer la limite de précision que peut avoir un tarif où la prime est modelée sur le risque. On montre que cette précision atteint environ 70 % de la variance totale après 5 ans, 81 % après 10 ans et 90 % après 20 ans, alors que le tarif actuel explique seulement 22 %, 25 % et 28 % respectivement.

Analyse de l'hétérogénéité des fréquences d'accidents des véhicules

Un dépouillement statistique¹⁾ a été fait sur l'ensemble des véhicules à 4 roues identifiés, assurés en France en responsabilité civile par la Compagnie d'Assurances l'Urbaine et la Seine: voitures particulières, camionnettes, petits camions en excluant seulement les véhicules des agriculteurs et de transports publics, ainsi que les camions de plus de 3,5 tonnes de poids total. Nous les désignerons indistinctement dans ce qui suit par voitures ou véhicules.

Pour chacun d'eux, on a relevé:

Z la zone de garage habituel (l'ensemble du territoire de la France continentale est divisé en 7 zones correspondant à des fréquences moyennes d'accidents différentes, les plus fortes étant observées dans la zone de Paris);

U l'usage du véhicule et la profession de son propriétaire (promenade, affaires, commerce, artisan . . .);

F la puissance fiscale du moteur du véhicule.

¹⁾ Nous remercions vivement M. J. Mouillard de l'aide qu'il nous a apportée pour l'ensemble de cette étude.

Ces trois premières caractéristiques sont celles figurant dans le tarif d'assurance Auto français;

T le modèle et le constructeur du véhicule (seulement pour les véhicules de grandes séries);

\bar{s}'_{ZUFTi} le nombre d'accidents observé sur le véhicule i de caractéristiques Z, U, F, T pendant l'année 1959.

Désignons par :

$\bar{s}'_{ZUFT} = (\sum_i s'_{ZUFTi}) / n_{ZUFT}$ la moyenne arithmétique des nombres d'accidents survenus aux véhicules ayant mêmes Z, U, F, T .

$\bar{s}'_{ZUF} = (\sum_T n_{ZUFT} \bar{s}'_{ZUFT}) / n_{ZUF}$ la moyenne arithmétique des nombres d'accidents survenus aux véhicules ayant mêmes Z, U, F .

$\bar{s}'_{ZU} = (\sum_F n_{ZUF} \bar{s}'_{ZUF}) / n_{ZU}$ la moyenne arithmétique des nombres d'accidents survenus aux véhicules ayant mêmes Z, U .

$\bar{s}'_Z = (\sum_U n_{ZU} \bar{s}'_{ZU}) / n_Z$ la moyenne arithmétique des nombres d'accidents survenus aux véhicules ayant la même zone de garage Z .

$\bar{s}' = (\sum_Z n_Z \bar{s}'_Z) / n = (\sum_Z \sum_U \sum_F \sum_T \sum_i s'_{ZUFTi}) / n$ la moyenne arithmétique des nombres d'accidents survenus à l'ensemble des véhicules.

La somme des carrés des écarts entre le nombre de sinistres survenus à chaque voiture et le nombre moyen peut se décomposer ainsi :

$$\left. \begin{aligned} \sum_Z \sum_U \sum_F \sum_T \sum_i (s'_{ZUFTi} - \bar{s}')^2 &= \sum_Z \sum_U \sum_F \sum_T \sum_i (s'_{ZUFTi} - \bar{s}'_{ZUFT})^2 \\ &+ \sum_Z \sum_U \sum_F \sum_T n_{ZUFT} (\bar{s}'_{ZUFT} - \bar{s}'_{ZUF})^2 + \sum_Z \sum_U \sum_F n_{ZUF} (\bar{s}'_{ZUF} - \bar{s}'_{ZU})^2 \\ &+ \sum_Z \sum_U n_{ZU} (\bar{s}'_{ZU} - \bar{s}'_Z)^2 + \sum_Z n_Z (\bar{s}'_Z - \bar{s}')^2 \end{aligned} \right\} (I)$$

dont les valeurs numériques trouvées pour les 55.562 accidents observés sur 169.950 voitures étudiées sont respectivement :

$$80.296,5 = 72.290,1 + 1.001,3 + 262,2 + 1.570,0 + 5.172,9$$

chacun des termes ainsi obtenus étant proportionnel à la variance due à chacun des caractères relevés.

	<i>Variance</i>	%
Part de variance due		
à la zone de garage habituel	0,0304	6,5
à l'usage et à la profession	0,0092	2,0
à la puissance du moteur	0,0015	0,3
au type et au constructeur	0,0059	1,2
à d'autres causes	0,4254	90,0
Total	0,4724	100,0

Si la zone de garage habituel du véhicule explique 6,5% de la variance, les autres critères ont des importances très faibles et ne représentent au total que 10% de la variance totale. Il reste donc à expliquer 90% de cette variance. Ils correspondent à l'hétérogénéité dans la survenance des accidents pour les voitures d'une même classe du tarif d'assurance, c'est-à-dire pour des voitures ayant toutes la même zone de garage habituel, le même usage, la même puissance fiscale et le même modèle.

Pour une classe quelconque, par exemple: Paris, Affaires, 8 chevaux fiscaux, Peugeot, on a indiqué dans la 2e colonne du tableau 1 les nombres des voitures ayant eu 0, 1, 2, 3 . . . accidents pendant l'année 1959.

Dans la colonne No 3, on a mentionné les nombres probables de voitures qu'on aurait dû observer d'après la loi de probabilité de Poisson, si ces véhicules avaient eu tous un même nombre probable de sinistres, donc un même risque.

TABLEAU 1

Nombre d'accidents 1	Nombre de voitures		
	Observé 2	calculé par	
		Loi de Poisson 3	Méthode proposée 4
0	764	687,3	759,8
1	347	450,6	360,6
2	146	147,7	135,3
3	45	32,3	46,3
4	18	5,3	15,1
5	2	0,7	4,8
6	2	0,1	2,1
Total	1.324	1.324	1.324

On voit immédiatement que la représentation donnée par la loi de Poisson est mauvaise. Ceci provient de ce que toute classe pourrait être décomposée en plusieurs sous-classes, si l'on ajoutait un nouveau critère de distinction entre les voitures à condition qu'il soit légèrement lié au risque. Ceci montre qu'à l'intérieur d'une classe toutes les voitures n'ont pas une même fréquence probable et que la distribution des nombres de véhicules, selon le nombre de sinistres qu'ils ont eus, est formée par la superposition de deux lois de probabilité : la loi de probabilité de la survenance au hasard des sinistres pour une voiture donnée et la loi de probabilité des risques des voitures, c'est-à-dire la loi de distribution des fréquences probables des diverses voitures appartenant à une même classe.

Désignons par :

$F(s)$ la loi de probabilité des risques des voitures d'une classe de tarif,

$G(s'/s)$ la loi de probabilité de la survenance au hasard des sinistres, c'est-à-dire du nombre de sinistres s' survenus à une voiture dont le risque, ici mesuré par l'espérance mathématique du nombre de sinistres, est s .

On a souvent admis que la loi de probabilité $G(s' | s)$ est une loi de Poisson. Il est possible de le démontrer de la manière suivante :

Nous avons montré précédemment ¹⁾ que lorsqu'un caractère X est mesuré avec une erreur de mesure ϵ , il est possible d'obtenir la suite infinie des estimations absolument correctes des moments de l'erreur ϵ , et par conséquent la loi de probabilité de ϵ à condition de disposer d'au moins 3 mesures du caractère sur chaque objet.

Dans le cas présent :

le caractère X à mesurer est le risque s_i de chaque voiture i

la mesure de s_i est le nombre de sinistres s'_i observés sur la voiture i

l'erreur de mesure ϵ'_i est la différence entre le nombre de sinistres observés s'_i et le nombre probable s_i , c'est-à-dire $\epsilon'_i = s'_i - s_i$.

Dans le cas de 3 mesures s'_i, s''_i, s'''_i sur chaque voiture, on obtient

¹⁾ P. J. DELAPORTE. Etude statistique des erreurs de mesure. Bulletin de l'Institut International de Statistique, t. 38, Tokyo 1960.

les estimations absolument correctes des moments centrés de l'erreur de mesure par (P. Delaporte loc. cit. formules 43 et 67 à 69)

$$\epsilon\mu_2'' = \frac{1}{2n} \sum_{i=1}^n \left[(s_i' - \bar{s}_i)^2 + (s_i'' - \bar{s}_i)^2 + (s_i''' - \bar{s}_i)^2 \right]$$

où $\bar{s}_i = (s_i' + s_i'' + s_i''') / 3$

n est le nombre de voitures étudiées

$$\left. \begin{aligned} \epsilon\mu_3'' &= \frac{3}{2n} \sum_{i=1}^n \left[(s_i' - \bar{s}_i)^3 + (s_i'' - \bar{s}_i)^3 + (s_i''' - \bar{s}_i)^3 \right] \\ \epsilon\mu_4'' &= \frac{3}{2n} \sum_{i=1}^n \left[(s_i' - \bar{s}_i)^4 + (s_i'' - \bar{s}_i)^4 + (s_i''' - \bar{s}_i)^4 \right] - 3\epsilon\mu_2''^2 \\ \epsilon\mu_5'' &= \frac{27}{16n} \sum_{i=1}^n \left[(s_i' - \bar{s}_i)^5 + (s_i'' - \bar{s}_i)^5 + (s_i''' - \bar{s}_i)^5 \right] - \frac{5}{4}\epsilon\mu_2'' \cdot \epsilon\mu_3'' \\ \epsilon\mu_6'' &= \frac{81}{22n} \sum_{i=1}^n \left[(s_i' - \bar{s}_i)^6 + (s_i'' - \bar{s}_i)^6 + (s_i''' - \bar{s}_i)^6 \right] \\ &\quad - \frac{105}{11}\epsilon\mu_2'' \cdot \epsilon\mu_4'' + \frac{50}{11}\epsilon\mu_3''^2 - \frac{60}{11}\epsilon\mu_2''^3 \end{aligned} \right\} (2)$$

Les premiers cumulants s'obtiennent par :

$$\left. \begin{aligned} k_2 &= \epsilon\mu_2'' \\ k_3 &= \epsilon\mu_3'' \\ k_4 &= \epsilon\mu_4'' - 3\epsilon\mu_2''^2 \\ k_5 &= \epsilon\mu_5'' - 10\epsilon\mu_2'' \epsilon\mu_3'' \end{aligned} \right\} (3)$$

Pour rechercher quelle est la loi de probabilité de ϵ on a obtenu 3 mesures du risque de chaque voiture en notant, pour des voitures assurées pendant les 12 mois de l'année 1958, les nombres de sinistres :

s_i' des mois de Janvier, Avril, Juillet et Octobre

s_i'' des mois de Février, Mai, Août et Novembre

s_i''' des mois de Mars, Juin, Septembre et Décembre.

On observe tout d'abord qu'il existe une remarquable égalité entre les cumulants K_2 et K_3 et que malgré la croissance rapide des moments avec leur ordre, les cumulants K_4 et K_5 sont du même ordre de grandeur que K_2 et K_3 . Ceci suggère que tous les cumulants de la loi de probabilité de l'erreur ϵ seraient égaux, les écarts aléatoires, dûs aux petites tailles des échantillons, croissant naturellement avec l'ordre du cumulant. Cette propriété d'égalité des cumulants est caractéristique de la loi de probabilité de Poisson.

TABLEAU 2

Zone de garage	Province Normal	Paris	Paris	Paris
Usage	Promenade	Promenade	Affaires	Affaires
Puissance	7	7	7	4
Constructeur	Simca	Simca	Simca	Renault
Nbre de voitures	3,276	1.216	1.373	1.234
K_2	0,0414	0,1162	0,2088	0,2242
K_3	0,0400	0,1184	0,1979	0,2358
K_4	0,0446	0,1832	0,3682	0,4817
K_5	0,0216	0,0960	0,1104	0,3205
Estimation du paramètre moyen de la loi de Poisson				
par μ_2^*	0,0414	0,1162	0,2088	0,2242
μ_3^*	0,0400	0,1184	0,1979	0,2358
μ_4^*	0,0439	0,1533	0,2738	0,3218
μ_5^*	0,0295	0,1108	0,1842	0,2457
μ_n^*	0,0440	0,1608	0,2998	

Dans la deuxième partie du tableau ci-dessus, on a indiqué les estimations du paramètre de la loi de Poisson tiré des divers moments. On constate que les écarts observés sont d'un ordre de grandeur acceptable par rapport à leurs écarts types, ils peuvent donc être attribués au hasard. Par conséquent, la *survenance au hasard des accidents de chaque voiture se fait selon la loi de probabilité de Poisson*.

La loi de probabilité des risques des voitures à l'intérieur d'une classe du tarif peut alors être étudiée ¹⁾ au moyen des moments de $F(s)$ exprimés en fonction des moments des nombres observés

¹⁾ P. Delaporte. Un problème de tarification de l'assurance accidents d'automobiles examiné par la Statistique Mathématique. Comptes rendus du XVIe Congrès International d'Actuaires, Bruxelles 1960, t. 2, p. 121-135.

s' de sinistres, sachant que la loi de probabilité de ε est une loi de Poisson.

$$s'm_1 = s'm_1$$

$$s'\mu_2 = s'\mu_2 - s'm_1$$

$$s'\mu_3 = s'\mu_3 - 3 s'\mu_2 + 2 s'm_1$$

$$s'\mu_4 = s'\mu_4 - 6 s'\mu_3 + (11 - 6 s'm_1) \cdot s'\mu_2 + 3 s'm_1 (s'm_1 - 2)$$

Dans le premier exemple d'étude des nombres de voitures ayant eu 0, 1, 2 . . . sinistres pendant l'année 1959, les nombres probables ainsi trouvés sont indiqués dans la colonne 4 du tableau No 1. Ils sont statistiquement en excellent accord avec les nombres observés.

Maximum de précision d'un tarif d'assurance automobile

Nous venons de montrer que pour une voiture donnée i , donc pour un risque donné i , les sinistres surviennent au hasard selon une loi de probabilité de Poisson de paramètre s_i . C'est dire que:

la probabilité d'observer ${}_1s'_i$ sinistres en 1 an sera :

$$G ({}_1s'_i | s_i) = e^{-s_i} \frac{s_i^{s'_i}}{s'_i!}$$

l'espérance mathématique du nombre de sinistres du véhicule i sera : pour 1 année quelconque h

$$E [{}_1s'_i] = \sum_{s'_i=0}^{\infty} {}_1s'_i \cdot G ({}_1s'_i | s_i) = s_i$$

pour t années

$$E_h [{}_h s'_i] = \sum_{h=1}^t E [{}_1 s'_i] = t \cdot s_i$$

le moment d'ordre 2 du nombre des sinistres du véhicule i sera : pour 1 an

$$s'_i m_2 = E [{}_1 s'^2_i] = \sum_{s'_i=0}^{+\infty} {}_1 s'^2_i \cdot G ({}_1 s'_i | s_i) = s_i + s_i^2$$

pour t années

$$\begin{aligned}
 {}_t s'_i m_2 &= E \left[\left(\sum_{h=1}^t {}_h s'_i \right)^2 \right] = E \left[\sum_{h=1}^t {}_h s_i'^2 + 2 \sum_h \sum_j {}_h s'_i \cdot {}_j s'_i \right] \\
 &= t \cdot {}_t s'_i m_2 + 2 \frac{t(t-1)}{2} \left\{ E [{}_1 s'_i] \right\}^2 \\
 &= t (s_i + s_i^2) + t(t-1) s_i^2 = t s_i + t^2 s_i^2
 \end{aligned}$$

Si l'on désigne par $F(s)$ la loi de probabilité des risques des véhicules dans l'ensemble d'une classe du tarif ou dans l'ensemble du parc assuré, on aura pour t années d'observation :

$$\begin{aligned}
 {}_t s'_i m_1 &= E_i [{}_t s'_i] = \int_{s_0}^{+\infty} t \cdot s_i dF(s_i) = t \cdot {}_t s m_1 \\
 {}_t s'_i m_2 &= E_i [{}_t s_i'^2] = \int_{s_0}^{+\infty} (t \cdot s_i + t^2 s_i^2) dF(s_i) = t \cdot {}_t s m_1 + t^2 s m_2
 \end{aligned}$$

où s_0 est la limite inférieure du risque.

La variance des nombres de sinistres par voiture en t années sera :

$$\begin{aligned}
 {}_t s' \mu_2 &= {}_t s' m_2 - {}_t s' m_1^2 = t \cdot {}_t s m_1 + t^2 \cdot {}_t s m_2 - t^2 \cdot {}_t s m_1^2 \\
 &= t {}_t s m_1 + t^2 {}_t s \mu_2 = t \cdot \bar{s}' + t^2 \cdot {}_t s \mu_2 \tag{5}
 \end{aligned}$$

Ainsi, la variance totale des nombres d'accidents en t années est-elle formée :

d'une partie proportionnelle à la fréquence des sinistres et au temps ;

d'une autre partie proportionnelle à la variance des risques et au carré du temps.

La variance totale trouvée au début de cette étude pour un an se décompose ainsi :

$${}_1 s' \mu_2 = 0,4724 = 0,3269 + {}_1 s \mu_2 \text{ d'où } {}_1 s \mu_2 = 0,1455$$

car la fréquence moyenne était $\bar{s} = \frac{55\ 562}{169\ 950} = 0,3269$

Pour une durée de t années, la variance totale devient :

$${}_t s' \mu_2 = 0,3269 t + 0,1455 t^2 \tag{6}$$

Le dernier terme est lui-même formé par une variance expliquée par le tarif actuel (zone de garage, usage et profession, puissance du moteur), soit $0,0411 t^2$ et une variance complémentaire $(0,1455$

— $0,0411) t^2 = 0,1044 t^2$ correspondant au gain de précision de tarification qui pourrait être obtenu en utilisant une prime modelée sur le risque au lieu de la tarification actuelle à la prime moyenne.

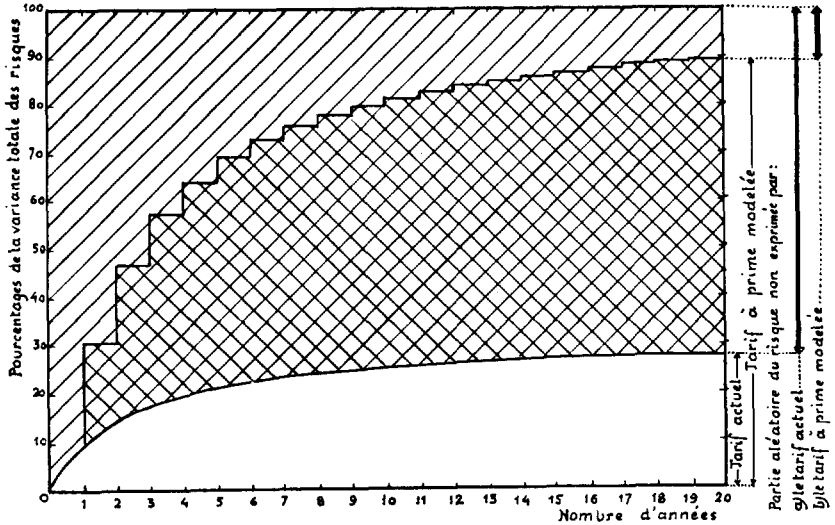
Si l'on divise les deux membres de l'équation (5) par ${}_s\mu_2$ et qu'on les multiplie par 100, on obtient les pourcentages de la variance totale des risques:

expliquée par la zone de garage habituel	$\frac{3,04 t}{0,3269 + 0,1455 t}$
expliquée par l'usage et la profession	$\frac{0,92 t}{0,3269 + 0,1455 t}$
expliquée par la puissance du moteur	$\frac{0,15 t}{0,3269 + 0,1455 t}$
total expliqué par le tarif actuel	$\frac{4,11 t}{0,3269 + 0,1455 t}$
complément explicable par un tarif où la prime est modelée sur le risque	$\frac{10,44 t}{0,3269 + 0,1455 t}$
reliquat dû à la survenance au hasard des sinistres et donc non explicable par un tarif d'assurance	$\frac{32,69}{0,3269 + 0,1455 t}$

Ce dernier terme tend vers zéro lorsque la durée d'assurance augmente. Tous les autres termes, au contraire, ont la forme analytique d'hyperboles qui tendent asymptotiquement vers des limites finies telles que la somme des variances expliquées asymptotiques est égale à 100 pour cent.

Les fonctions ci-dessus indiquent les courbes limites de l'efficacité des diverses méthodes de tarification des risques. En fait, les primes peuvent difficilement être modifiées en cours d'année, c'est pourquoi le tarif à prime modelée sur la base d'une révision annuelle de la prime d'après les antécédents du risque se présente sous la forme indiquée dans le diagramme ci-après:

Ce diagramme indique, dans sa partie inférieure limitée par la courbe inférieure, le pourcentage de la variance totale des risques analysé par le tarif actuel, tout le reste des risques étant mis en mutualité, sans distinction entre les bons risques et les mauvais risques. Après une durée d'assurance très longue, l'efficacité de cette tarification n'est que 28,3 % de l'hétérogénéité des risques.



La zone comprise entre l'axe des nombres d'années et la courbe supérieure montre le pourcentage de la variance totale des risques qui pourrait être analysé par un tarif où la prime du tarif actuel de première année serait ensuite modelée sur le risque pour tenir compte des particularités individuelles de chaque risque. Elle tend vers 100 % pour une durée très longue d'assurance.

La zone située au-dessus de la courbe supérieure indique la part de la variance du risque qui ne peut être analysée par un tarif d'assurance, car elle constitue la *partie aléatoire irréductible du risque qui doit nécessairement être mise en mutualité*. Cette partie aléatoire tend vers zéro quand la durée d'assurance est très longue.

La zone comprise entre les deux courbes (zone doublement hachurée) est très large, elle correspond à la part des risques que les compagnies d'assurance appliquant un tarif commun en France, mettent en mutualité, alors qu'un tarif mieux adapté pourrait permettre de faire payer à chaque assuré une prime proportionnée à son risque. C'est l'étendue de cette zone qui permet à des compagnies n'appliquant pas le tarif commun de sélectionner les risques qu'elles assurent et de leur demander des primes inférieures à celles du tarif commun, ce qui produit une antisélection dans le portefeuille des compagnies du tarif commun.

Des motifs commerciaux feront souvent choisir des tarifs éloignés du tarif à prime modelée sur le risque qui correspond à un maximum d'efficacité de discrimination entre les risques.

Les qualités techniques d'un tarif d'assurance automobile devraient toujours être appréciées sous la forme de l'hétérogénéité des risques décomposée en trois parties: celle analysée par le tarif, celle volontairement laissée en mutualité et l'hétérogénéité résiduelle obligatoirement laissée en mutualité.