



# Le droit, la statistique et l'assurance

*Jean-Hubert SMITH LACROIX\**

*André BÉLANGER\*\**

## Résumé

Peu de limites sont imposées à la tarification en assurance automobile car la statistique et la logique de marché dans laquelle elle s'inscrit sont perçues comme des mécanismes autonomes assurant une certaine équité. La statistique module jusqu'à nos droits et libertés fondamentaux et nous nous contentons souvent de justifications basées sur de simples corrélations. Il semble toutefois que cela s'explique par une certaine incompréhension de la statistique. En établissant une démonstration de la présence de subjectivité en statistique, nous présentons une situation où un critère de tarification est acceptable pour l'actuaire s'il ne crée pas de discrimination aux yeux du juriste, tandis que le juriste tolère une discrimination lorsque suggérée par l'actuaire. Bien que nous ne prétendions pas que de graves injustices soient actuellement causées par le fait que

## Abstract

Few limits are imposed on the pricing of automobile insurance since statistics and the free market paradigm in which it operates are perceived to be independent mechanisms that insure some level of fairness. Statistics even affects our basic rights and freedoms and we often are satisfied with justifications based on mere correlations. However, this tendency seems heavily predicated on a lack of understanding of statistics. By showing that statistics implies some degree of subjectivity, we paint the picture of an environment in which a rating variable is deemed acceptable by the actuary where it does not create illegal discrimination in the eyes of the jurist, but where the jurist tolerates discrimination when suggested by the actuary. While we do not claim that grave injustices are presently caused by the situation wherein statistics no longer

\* Chargé de cours à l'Université Laval, B. Sc. (mathématiques); M. Sc. (statistique); M. Sc. (finance); doctorant en sciences de l'administration (finance et assurance), Faculté des sciences de l'administration, Université Laval.

\*\* Professeur à la Faculté de droit, Université Laval. Les auteurs remercient les évaluateurs de la version préliminaire de ce texte pour leurs commentaires et suggestions.



la statistique ne garantit plus un résultat *juste parce qu'objectif*, nous formulons ici un questionnement essentiel aux plans juridique et collectif afin de ne pas laisser la technique répondre à une question de justice sociale.

ensures (objective fairness based) results, we introduce an essential debate on both the collective and legal fronts which is necessary in order to prevent a letting technique settle matters of social justice.

---

## Plan de l'article

<b>Introduction</b> .....	103
<b>I. Le système actuel et la recherche d'une tarification adéquate</b> ....	106
<b>II. L'attrait pour la statistique</b> .....	107
<b>III. La subjectivité dans la statistique</b> .....	110
<b>IV. Choix des méthodes</b> .....	119
<b>V. Le <i>data mining</i> abusif</b> .....	122
<b>VI. Des choix lourds de conséquences</b> .....	126
<b>Conclusion</b> .....	129





*Aucune loi en matière de droits de la personne ne pourrait jamais atteindre ses objectifs si la discrimination pouvait être justifiée par l'affirmation intéressée qu'une pratique donnée « existe depuis toujours », ce qui constitue l'élément central de l'argument de l'intimée qui se fonde sur l'inexistence d'autres systèmes de classification. Si tel était le cas, ce serait récompenser l'inaction et la discrimination traditionnelle au détriment du progrès et de la reconnaissance de normes sociales plus exigeantes.*

L'Honorable Claire L'Heureux-Dubé, dissidente  
dans l'affaire *Zurich Compagnie d'Assurances c. Ontario*  
(*Commission des droits de la personne*), [1992] 2 R.C.S. 321.

En droit des assurances, le recours justificatif à la statistique est courant<sup>1</sup>. À titre de science qualifiée d'« exacte », celle-ci permet de soutenir certaines positions législatives sous couvert d'impartialité. Il semble toutefois que ce soit davantage la relative incompréhension du juriste face à ce domaine, qu'une réelle objectivité du processus, qui explique cet état de fait. En ce sens, il nous paraît pertinent de nous pencher ici sur les liens entre la pratique commerciale de l'assurance et les sciences tant statistique que juridique en matière de droit des assurances. Nous souhaitons ainsi démontrer qu'une adhésion aveugle au discours scientifique exotique au domaine légal prend le pas sur ce qui ne devrait constituer qu'une argumentation complémentaire, entraînant de ce fait un déficit de légitimité. Afin d'illustrer notre propos, nous nous intéresserons au rôle du législateur dans le cadre de la tarification de l'assurance de dommages reliés à l'utilisation d'une automobile dans un contexte où il semble que, comme c'est le cas dans plusieurs juridictions, « [the] legislators and the courts appear rather uninterested in the ways in which the insurance industry renders the future calculable »<sup>2</sup>. Un tel examen implique une réflexion sur l'obligation d'assurance et les techniques de la statistique qui sont importantes en matière de tarification.

Nous verrons que le principal obstacle – sans être le seul – à une réflexion critique sur la question est probablement notre trop grande foi dans la statistique. Il est d'ailleurs important de se rappeler de la mise en garde de

<sup>1</sup> La statistique module même nos droits et libertés fondamentaux, comme en témoigne l'article 20.1 de la *Charte des droits et libertés de la personne*, L.R.Q., c. C-12.

<sup>2</sup> Pat O'MALLEY, *Risk, Uncertainty and Government*, London, The Glass House Press, 2004, p. 117.



Fereyabend selon laquelle « toutes les méthodologies, même les plus évidentes, ont leurs limites »<sup>3</sup>, que « le savoir d'aujourd'hui peut devenir le conte de fées de demain, et que le mythe le plus risible peut éventuellement devenir un élément très solide de la science »<sup>4</sup>. Ainsi, il semble peu avisé de faire reposer une trop grande part de la réflexion relative à la réglementation de la tarification sur les épaules des actuaires<sup>5</sup>. À cet égard, il importe de souligner que, « [le] processus d'établissement des prix en assurance automobile est un processus complexe et le jugement y occupe une place importante »<sup>6</sup>. Notre intention ici n'est pas de tomber dans un excès de méfiance, mais d'indiquer certaines failles du système et d'inciter un questionnement essentiel aux plans juridique et collectif. Notre interrogation questionne davantage la politique juridique que la structure du droit positif, qu'elle débordé largement. C'est d'ailleurs l'intérêt du regard interdisciplinaire que nous voulons porter ici. Par conséquent, le lecteur doit comprendre que l'approche critique adoptée ne réfère que très partiellement au droit comparé, méthodologie qui n'est aucunement utilisée aux fins de notre exposé, mais qui pourrait évidemment être tout à fait pertinente dans le cadre d'une étude complémentaire éventuelle. L'objectif du présent texte consiste plutôt à mettre l'accent sur un champ de connaissances spécifique de l'un des coauteurs, soit la statistique, et à l'intégrer au sein d'une réflexion de nature juridique. Somme toute, il s'agit de permettre à la critique de la statistique de se faire entendre en matière de droit de l'assurance.

Dans les prochaines pages, nous exposerons donc le point de vue selon lequel il est aujourd'hui souhaitable que le législateur québécois justifie différemment et avec plus de rigueur la réglementation de la tarification en assurance automobile. Dans ce contexte, l'un des éléments à prendre en considération est évidemment la *Convention d'indemnisation directe*<sup>7</sup>. Celle-

<sup>3</sup> Paul FERAYABEND, *Contre la méthode: esquisse d'une théorie anarchiste de la connaissance*, Paris, Éditions du Seuil, 1979, p. 30.

<sup>4</sup> *Id.*, p. 53.

<sup>5</sup> Voir à ce sujet: Leah WORTHAM, « Insurance classification: too important to be left to the actuaries », (1985) 19 *Mich. J.L. Reform* 349; ainsi que Leah WORTHAM, « The Economics of Insurance Classification: The Sound of One Invisible Hand Clapping », (1986) 47 *Ohio St L.J.* 835.

<sup>6</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, *Rapport annuel sur la tarification en assurance automobile*, Québec, 1978, p. 11.

<sup>7</sup> *Convention d'indemnisation directe pour le règlement des sinistres automobiles*, R.Q., c. A-25, r.2.3 (ci-après « *Convention d'indemnisation directe* »).

ci réduit les frais en permettant que la plus petite partie possible des coûts de la réparation du préjudice soit absorbée par des intermédiaires (avocats, tribunaux, etc.) et que le traitement des dossiers soit aussi simple et rapide que possible. Pour réaliser sa mise en place, nous avons collectivement choisi, puisque cela est nécessaire au bon fonctionnement de la convention, de rendre obligatoire ce qu'on appelle communément le Chapitre A du contrat d'assurance automobile portant sur la responsabilité civile<sup>8</sup>. Nous considérons ici que cette obligation d'assurance est légitime et ne doit pas être remise en question.

Étant donné la portée du présent texte, nous n'entrerons pas dans le débat sur la nécessité de posséder une automobile. Nous nous contenterons de dire que, *à tout le moins pour certaines personnes*, renoncer à la possession d'une voiture serait très pénalisant, par exemple pour des gens habitant des régions où le transport en commun n'est pas accessible, dont le travail dépend de la possibilité de se déplacer en automobile ou qui résident à des endroits éloignés. Pour ces personnes, la couverture définie au chapitre A est donc un produit obligatoire et elles représentent par le fait même une clientèle forcée pour les assureurs. Nous considérons ainsi que la tarification du chapitre A doit être examinée d'une façon différente de la tarification du reste de l'assurance automobile. Par conséquent, notre argumentaire se développe dans le contexte de la tarification de la protection prévue au chapitre A qui, en plus d'être obligatoire, ne peut être obtenue autrement que par l'entremise d'intérêts privés. C'est particulièrement ici que la justification des normes, qui se devrait d'être claire, souffre d'un important déficit. Nous tenterons d'en faire la démonstration en abordant la délicate question de déterminer ce qui caractérise une tarification adéquate (I), mise en perspective avec l'important attrait pour la statistique en matière de normativité d'assurance (II). De plus, plusieurs éléments troublants pour le juriste doivent être mis à jour. Ainsi, la présence inévitable de subjectivité dans l'application de la statistique (III), par exemple lors du choix des méthodes (IV) ou encore pour éviter le Data Mining abusif (V), entraîne des conséquences qui, bien qu'importantes, sont souvent ignorées en droit (VI).

<sup>8</sup> *Loi sur l'assurance automobile*, L.R.Q., c. A-25, art. 84 (ci-après « chapitre A »).

## I. Le système actuel et la recherche d'une tarification adéquate

L'administration québécoise, dans son examen du fonctionnement du système de tarification en vigueur, fait appel au concept de tarification dite *adéquate* en ces termes :

« Une tarification adéquate doit avant tout permettre de couvrir tous les coûts liés aux opérations d'assurance et de générer des bénéfices raisonnables en tenant compte des risques associés aux activités d'assurance. »<sup>9</sup>

Il est intéressant de comparer ce passage du *Rapport annuel sur la tarification automobile* à la position adoptée dans le Rapport Gauvin selon laquelle :

« [...] un système de tarification doit être conçu de façon à établir des taux qui ne soient pas « excessifs, inadéquats ou arbitrairement discriminatoires ». Les deux premières contraintes exigent [...] la prévision par l'assureur du coût total des sinistres et des frais d'administration en rapport avec l'ensemble de ses affaires. Il est évident a priori que des taux excessifs tendraient à lui faire perdre des clients tandis que des taux inadéquats le placeraient dans une mauvaise situation financière. L'imposition d'un taux unique, commun à l'ensemble des assurés pourrait fort bien remplir ces deux premiers critères, mais ce système produirait cependant des résultats très discriminatoires à l'égard de l'assuré individuel, ce qui viole évidemment la troisième contrainte. »<sup>10</sup>

Selon le Rapport Gauvin, « [de] façon générale, l'équité individuelle entre les assurés constitue l'objectif ultime de la tarification actuelle et c'est par rapport à cet objectif qu'elle doit être évaluée »<sup>11</sup>. Ainsi, là où l'on faisait autrefois une référence explicite à l'équité entre assurés, semble s'être opéré un glissement vers la seule notion de l'équité entre d'un côté l'ensemble des assurés et de l'autre l'ensemble des assureurs<sup>12</sup>. Pourtant, on

<sup>9</sup> GOUVERNEMENT DU QUÉBEC, AUTORITÉ DES MARCHÉS FINANCIERS, *Rapport annuel sur la tarification en assurance automobile*, Québec, 2007, p. 31 (ci-après « RTA 2007 »).

<sup>10</sup> GOUVERNEMENT DU QUÉBEC, *Rapport du comité d'étude sur l'assurance automobile*, Québec, Éditeur officiel du Québec, 1974, p. 162 (ci-après « Rapport Gauvin »).

<sup>11</sup> *Id.*, p. 243.

<sup>12</sup> Précisons que le concept d'équité entre assurés n'a pas complètement disparu de la documentation pertinente, l'administration continuant par exemple à étudier les différences entre les assurés de sexes différents dans les rapports annuels. Toutefois, il semble à la lecture des documents complets que l'importance relative de l'équité entre



peut aisément faire la distinction entre un *niveau* adéquat et une *répartition* adéquate des primes. On peut aussi se demander si, considérés séparément, ces deux critères sont satisfaits. Or, il nous semble légitime de croire que le premier n'implique pas nécessairement le deuxième. En effet, un rendement raisonnable de l'industrie n'implique pas une tarification adéquate. Or, si l'organisme responsable de la supervision du marché de l'assurance s'attarde peu à l'examen de la justesse de la répartition des coûts, le législateur ne semble pas lui rappeler son devoir de le faire. Ceci s'explique peut-être par la conviction que des mécanismes autonomes sont suffisants pour soutenir le droit. Deux principaux mécanismes sont ici visés : l'utilisation de la statistique et la logique de marché dans laquelle elle s'inscrit<sup>13</sup>.

## II. L'attrait pour la statistique

La pratique de l'assurance est aujourd'hui caractérisée par une utilisation intensive de la statistique. Nul ne peut nier que cette dernière a permis de grandes avancées et a introduit un peu de rationalité dans un exercice qui, à une certaine époque, pouvait dans une large mesure être apparenté à un simple jeu de hasard<sup>14</sup>. Toutefois, on doit se questionner sur la profondeur du changement de perspective qui s'est opéré avec le temps. L'actuaire et la statistique ont obtenu le statut convoité de *sciences*, et ce dernier leur procure une aura d'inafaillibilité et d'incontestabilité. Il en découle une

---

les assurés dans l'analyse du caractère adéquat de la tarification ait diminuée avec le temps.

<sup>13</sup> En ce sens, voir François EWALD, « Génétique et assurance », (1999) 3 *R.G.D.A.* 539, 545 : « L'opération d'assurance est liée à la notion de justice actuarielle, d'équité, de fairness. Celle-ci s'entend en plusieurs sens : d'abord, vis-à-vis de la mutualité des assurés, chacun ne représente pas le même risque. [ ... ] L'idée d'équité actuarielle consiste à faire en sorte que la contribution de chacun à la mutualité corresponde à son risque. C'est-à-dire à faire en sorte que chacun paie pour ce qu'il doit. Ni plus, ni moins. [ ... ] Enfin, l'idée de justice actuarielle est en même temps la conséquence d'un marché concurrentiel : dans un tel contexte, une société qui ne tarifierait pas le risque à son juste prix se verrait vite concurrencée par une autre qui pourrait offrir à ses assurés des tarifs plus adéquats pour une même garantie. »

<sup>14</sup> Voir : Didier DACUNHA-CASTELLE, *Chemins de l'aléatoire, Le hasard et le risque dans la société moderne*, Paris, Flammarion, 1996 ; Alain PLESSIS et André STRAUS, « Le développement des assurances en Europe jusqu'au début de l'industrialisation », (2000) 68-1 *Assurances* 65.

confiance dont l'importance peut être remise en question. Des auteurs nous rappellent que :

« [an] influential claim is that, as a matter of historical fact, modern science was born in the early seventeenth century when the strategy of taking the facts of observation seriously as the basis for science was first seriously adopted. It is held by those who embrace and exploit this story about the birth of science that prior to the seventeenth century the observable facts were not taken seriously as the foundation for knowledge. Rather so the familiar story goes, knowledge was based largely on authority, especially the authority of the philosopher Aristotle and the authority of the Bible. »<sup>15</sup>

Dans le même ordre d'idées, Ewald souligne que :

« La sélection d'un risque n'est pas discriminatoire, dès lors que son critère est un déterminant du risque. C'est ce que constate le droit y compris en matière de santé. *Le principe de l'équité actuarielle* est : à risque égal, traitement égal. »<sup>16</sup>

Il indique également que :

« Les accidents sont prévisibles, assurables, calculables. [...] L'accident moderne a une objectivité propre qui le fait échapper à l'analyse traditionnelle des comportements en termes de prudence ou de vigilance individuelles. »<sup>17</sup>

Il est toutefois intéressant de réaliser que tous ne sont pas d'accord avec ces positions. Un an après la parution de ce dernier commentaire d'Ewald, on lisait dans une étude se penchant sur les conséquences d'un phénomène important appelé le *redlining*<sup>18</sup> que « underwriting is not entirely rational – that is predicated solely on objective risk assessment »<sup>19</sup>. Il faudrait, en

<sup>15</sup> Alan F. CHALMERS, *What is this thing called Science?*, 3<sup>e</sup> éd., Indianapolis, Hackett Publishing Company, 1999, p. 2.

<sup>16</sup> François EWALD, « Les valeurs de l'assurance », dans François EWALD et Jean-Hervé LORENZI (dir.), *Encyclopédie de l'assurance*, Paris, Economica, 1998, p. 399, à la page 419 (nos italiques).

<sup>17</sup> François EWALD, *L'État providence*, Paris, Grasset, 1986, p. 17.

<sup>18</sup> Au sujet du *redlining*, voir : Michel GENDRON et Geneviève CHOQUETTE, « L'accès à l'assurance habitation dans les quartiers centraux, *redlining* et conditions socio-économiques », (1999) 69 *Assurances* 571 ; Michel GENDRON et Étienne MARCEAU, « L'accès à l'assurance habitation dans les quartiers centraux de quatre villes québécoises » (1999) 67 *Assurances* 479 ; Jay D. SCHULTZ, « An Analysis of Agent Location and Homeowners Insurance Availability », (1995) 14 *Journal of Insurance Regulation* 65.

<sup>19</sup> Gregory D. SQUIRES et William VÉLEZ, « Insurance Redlining and the Transformation of Urban Metropolis », (1987) 23 *Urban Affairs Quarterly* 63, 74. Par ailleurs, nous souli-

effet, se rappeler que la modélisation statistique ne représente qu'une partie de l'exercice de l'assurance, une activité commerciale dans notre contexte. Or, même si la scientificité de la statistique la rendait intrinsèquement *juste*, cette discipline n'est que le point de départ de l'assurance et ne garantirait aucunement l'infailibilité du commerce dans son ensemble. On pourrait encore voir des assureurs accorder des rabais à certains groupes, non pas pour des raisons scientifiques, mais par exemple afin de s'accaparer une plus grande part de marché à l'intérieur d'une certaine niche ou pour affecter leur image de manière stratégique<sup>20</sup>. Il s'agit d'ailleurs d'une position qui, malgré des développements majeurs dans les techniques de la modélisation statistique, est toujours présente dans des écrits plus récents comme en fait foi cet extrait :

« While insurance is regarded as the quitessentially risk-based institution, even it is strongly marked by educated guesswork, traditionnal routines and hunches »<sup>21</sup>.

Malgré cela, il appert que la science a transformé le sinistre, qui était autrefois au moins partiellement le résultat du hasard, en un phénomène quasi déterministe. Le résultat est parfois surprenant ; François Ewald va par exemple jusqu'à affirmer qu'« il n'y a pas d'événements qui soient par nature des accidents »<sup>22</sup>. Pourtant, il n'en a pas toujours été ainsi. Les membres du Comité d'étude sur l'assurance automobile écrivaient en 1974 :

« Ces chiffres illustrent bien que le facteur humain, malgré son importance, ne saurait expliquer tous les accidents [...]. C'est une erreur de vouloir toujours attribuer à l'humain la responsabilité d'un accident ; comme semble l'être le fait de ne rechercher qu'une cause unique à un accident : différents facteurs interviennent, bien qu'à des degrés divers. Un accident peut bien avoir pour cause immédiate l'erreur humaine ; mais l'environnement eut-il été mieux conçu que cette erreur aurait peut-être été moins susceptible de se produire

---

gnons au passage l'utilisation du terme « underwriting », car il y a une distinction importante, mais souvent négligée dans les débats sur la réglementation de l'assurance, entre tarification et souscription.

<sup>20</sup> Par exemple, on peut se demander si les « rabais verts » proposés par certaines compagnies aux conducteurs de voitures hybrides sont à l'heure actuelle motivés par des considérations statistiques.

<sup>21</sup> Pat O'MALLEY, préc., note 2, p. 7.

<sup>22</sup> François EWALD, préc., note 17, p. 22.

et moins désastreuse dans ses conséquences. Tous les facteurs sont bien souvent tellement interdépendants que *l'accident peut être considéré comme parfaitement fortuit.*»<sup>23</sup>

De plus, la lecture des propos cités ci-haut laisse croire que *l'erreur* en tarification est impossible étant donné la certitude qui est conférée à cet exercice par la statistique. L'impact du caractère scientifique de la statistique devient tel qu'on substitue ses techniques au jugement humain. Tel que l'écrit Ewald :

« Dans la logique des jugements de responsabilité, le juge part de la réalité de l'accident, du dommage, pour en inférer l'existence de sa cause dans une faute de conduite. Il suppose que sans faute il n'y aurait pas eu d'accident. L'assureur fonde ses calculs sur la probabilité *objective* d'un accident indépendamment de toute volonté : peu importe qu'il relève de la faute de l'un ou de l'autre, qu'on eut pu l'éviter, le fait est que quelle que soit la bonne ou mauvaise volonté des hommes, quoi qu'ils aient pu vouloir faire ou ne pas faire, l'accident se produit avec telle ou telle régularité. »<sup>24</sup>

Et il ajoute que « [l'assurance] est un mode d'administration de la justice concurrent du droit. [...] [L]'assurance pense la justice dans la catégorie de répartition dont il s'agit de fixer équitablement la règle »<sup>25</sup>. Pourtant, un examen plus approfondi révèle qu'il est trompeur de prétendre que l'application elle-même de la statistique, en plus de l'utilisation des résultats qu'elle produit, se fait de manière complètement *objective*.

### III. La subjectivité dans la statistique

Les juristes doivent réaliser que le domaine de la statistique n'est pas imperméable à la dogmatique et qu'il existe des courants de pensée qui ont une influence majeure sur son évolution. Par exemple, le développement de l'approche Bayésienne nous rappelle que les statistiques classiques, tel qu'elles sont souvent appliquées dans la pratique, ne sont pas infaillibles et qu'elles ne sont jamais totalement objectives. Howson et Urbach tiennent le discours suivant :

<sup>23</sup> GOUVERNEMENT DU QUÉBEC, préc., note 10, p. 40 (nos italiques).

<sup>24</sup> François EWALD, préc., note 17, p. 176 (nos italiques).

<sup>25</sup> *Id.*, p. 180.

« Indeed, we shall argue that the ideal of total objectivity is unattainable and that classical methods, which pose as guardians of that ideal, actually violate it at every turn; virtually none of those methods can be applied without a generous helping of personal judgment and arbitrary assumption. »<sup>26</sup>

De tels propos nous rappellent que, contrairement à la prétention habituelle, il y a nécessairement une part de subjectivité dans l'application des méthodes de la statistique. En effet, il est inévitable que le jugement du statisticien intervienne dans l'ajustement de modèles ou dans tout traitement statistique en général, et en tarification d'assurance en particulier. Par exemple, le choix de la forme d'un modèle, notamment par la simple question d'opter pour des effets multiplicatifs ou additifs, repose nécessairement sur un certain nombre de suppositions. Les passages qui suivent, tirés d'un document intitulé *A practitioner's Guide to Generalized Linear Models* rédigé par la firme de consultants Watson Wyatt, illustrent bien ce fait. On y lit que « [T]he statistical framework of GLMs allows *explicit assumptions* to be made about the nature of the insurance data and its relationship with predictive variables »<sup>27</sup>.

Et les auteurs ajoutent que « [T]he basic ingredients for a linear model thus consist of two elements: a. *a set of assumptions* about the relationship between  $Y$  and the predictor variables and b. [...] »<sup>28</sup>.

Qui plus est, ces suppositions peuvent affecter directement la valeur des paramètres obtenus comme on le voit ici :

« The vector of responses  $Y_i$ , the design matrix  $X_{ij}$ , the prior weights  $\omega_i$ , and the offset term  $\xi_i$  are based on data *in a manner determined by the practitioner.* »<sup>29</sup>

En fait, on va jusqu'à reconnaître explicitement que le statisticien devra nécessairement appliquer à l'occasion son propre jugement dans le processus, comme en témoigne ce dernier passage :

« To distinguish whether a factor effect is systematic or random (and therefore unlikely to be repeated in the future) there are a number of criteria which can

<sup>26</sup> Colin HOWSON et Peter URBACH, *Scientific reasoning: the Bayesian approach*, Chicago, Open Court, 2006, p. 9.

<sup>27</sup> Duncan ANDERSON et al., *A practitioner's Guide to Generalized Linear Models – Third edition*, Watson Wyatt Worldwide, 2007, p. 3 (nos italiques).

<sup>28</sup> *Id.*, p. 9 (nos italiques).

<sup>29</sup> *Id.*, p. 21 (nos italiques).

be considered, including parameter estimate standard errors, deviance tests, consistency with time, *common sense*.»<sup>30</sup>

Ainsi, même s'il existe des méthodes pour orienter le statisticien dans les choix qu'il a à faire, le processus n'est jamais complètement indépendant de son jugement personnel. Bien entendu, celui-ci intervient loin en amont du produit final, mais les résultats, et éventuellement la tarification qui en résulte dans le contexte de l'assurance, n'en demeurent pas moins ultimement influencés.

Dans cette optique, rappelons ici que le simple choix du seuil de significativité<sup>31</sup> approprié peut susciter de profonds débats<sup>32</sup> et que la procédure selon laquelle on procède à l'inspection visuelle de divers graphiques pour évaluer le respect des hypothèses d'un modèle est un exemple évident de la place laissée au jugement humain dans l'application de la statistique<sup>33</sup>. Contrairement à la perception généralement véhiculée, les méthodes de la statistique appliquée ne sont pas purement objectives. En ce sens, Howson et Urbach soutiennent que « [classical] methods necessarily introduce arbitrary elements that are at variance not only with scientific practice and intuition, but also with the objectivist ideals that motivated them »<sup>34</sup>. Dans ce contexte, une confiance aveugle dans la statistique appliquée favorise le glissement qui entraîne une exonération de toute responsabilité de celui qui l'utilise. C'est aussi elle qui nous rend susceptibles de tolérer des pratiques autrement critiquables simplement parce qu'elles seraient suggérées par les chiffres. Cela est particulièrement important dans le domaine de l'assurance. En effet, là où nous avons tendance à ne voir aujourd'hui qu'un

<sup>30</sup> *Id.*, p. 45 (nos italiques).

<sup>31</sup> Le seuil de significativité, habituellement représenté par le paramètre  $\alpha$ , est la force probante jugée nécessaire par le statisticien pour arriver à rejeter certaines hypothèses. Par exemple, dans un test classique, on considèrera qu'il est justifié d'abandonner une hypothèse de départ pour considérer une hypothèse alternative seulement si la probabilité d'observer des résultats aussi extrêmes que ceux contenus dans nos données, calculée en fonction d'un modèle choisi et de l'hypothèse de départ, était inférieure à ce seuil.

<sup>32</sup> Voir : Robert S. PINDYCK et Daniel L. RUBINFELD, *Econometric Models and Economic Forecasts*, 4<sup>e</sup> éd., Boston, Irwin McGraw Hill, 1999, p. 67.

<sup>33</sup> À cet égard, on peut consulter : Ronald CHRISTENSEN, *Plane Answers to Complex Questions: The Theory of Linear Models*, 3<sup>e</sup> éd., New York, Springer, 2002, p. 342-359, qui présente une multitude de graphiques couramment analysés et les descriptions des patrons qu'on souhaite ou ne souhaite pas y déceler.

<sup>34</sup> C. HOWSON & P. URBACH, préc., note 26, p. 182.

processus mécanique de recherche de corrélation, on y voyait autrefois un exercice bien différent. Par exemple, le RTA 1980 indiquait :

« Les assurés doivent payer une prime qui reflète le degré de risque qu'ils représentent. Pour y arriver, il faut grouper, classer ensemble les assurés exposés à un degré similaire de risque et leur fixer la même prime. Les caractéristiques propres à un système de classification qui permettent le traitement équitable des assurés sont multiples : les principales ont trait à l'homogénéité des classes, la simplicité, la facilité d'application, *la causalité, le contrôle par l'assuré*, l'incitation à la réduction du risque, *l'acceptation sociale*, et, enfin, la justification actuarielle. »<sup>35</sup>

Ainsi, tandis que la justification actuarielle venait derrière le désir de moduler la contribution demandée en fonction des *choix* de l'assuré et de facteurs pouvant *causer* davantage de pertes, nous avons maintenant tendance à ne voir qu'un exercice mathématique, un exercice technique. Il est d'ailleurs important de souligner que la statistique repose sur des relations de corrélation et non des relations de causalité<sup>36</sup>. La différence n'est pas sans conséquence. Cette causalité, qui n'est pas *démontrable* simplement par l'entremise de la statistique, était pourtant recherchée explicitement en tarification il n'y a pas si longtemps. Par exemple, le RTA 1980 mentionnait :

« [Un] système de classification est valable s'il permet de prévoir, de façon fiable, les coûts des sinistres dans chaque classe. Il est souhaitable que les éléments, les critères propres à chaque classe aient *une relation de cause à effet* avec le potentiel de risque des assurés qui en font partie. »<sup>37</sup>

Depuis, cette préoccupation s'estompe. Ainsi, il est par exemple aujourd'hui secondaire dans le débat sur l'utilisation du dossier de crédit en assurance de se demander si une mauvaise cote *cause* davantage d'accidents.

<sup>35</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, *Rapport annuel sur la tarification en assurance automobile*, Québec, 1980, p. 37 (nos italiques) (ci-après « RTA 1980 »).

<sup>36</sup> Voir : R.S. PINDYCK et D. L. RUBINFELD, préc., note 32, qui mentionnent à la page 74 : « The least-squares method is appropriate only if the causal structure of the model can be determined before the data are examined. If a model [of a certain type] is specified, one may interpret a significant *t* statistic [...] as evidence tending to *validate* the model. » Le modèle de régression viendrait donc appuyer la croyance en une relation de causalité qu'on aurait découverte par d'autres méthodes.

<sup>37</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, préc., note 35, p. 38 (nos italiques).

À cet égard, on trouve un exemple frappant de cette nouvelle philosophie dans les écrits récents de Brockett et Golden. Ces auteurs commencent leur plaidoyer de la façon suivante :

« However, in spite of [an individual's credit history] obvious success as an underwriting tool, and the clear actuarial substantiation of a strong association between credit score and insured losses over multiple methods and multiple studies, the use of credit scoring is under attack because there is not an understanding of *why* there is an association. »<sup>38</sup>

Les mêmes auteurs ajoutent :

« One reason for this controversy is that it is not clear why there is a correlation between credit score and insured losses (unlike the use of driving history to predict losses, building construction type to predict homeowners insurance losses, etc.). [...] A problem confronting the insurance industry is exhibiting to the legislatures (and the public) that there is an underlying rationale for the observed relationship between credit scores and insured losses; that it is not a statistical fluke. »<sup>39</sup>

Or, ce discours s'avère plutôt étonnant. Outre le fait que le besoin de se faire comprendre du public ne se présente qu'en toile de fond, on remarque dans le choix des mots qu'il n'est pas question de chercher la causalité. Il est tenu pour acquis que l'existence d'un lien, quel qu'il soit, est suffisante. Ici, on explique que le fait d'avoir un mauvais dossier de crédit et celui de représenter un plus grand risque pour l'assureur découleraient tous les deux d'une cause commune. Cette cause n'est ni plus ni moins qu'un assemblage de caractéristiques physiques, voire même génétiques, des individus. C'est ainsi que les auteurs écrivent dans le passage suivant :

« Credit scoring "works" because it provides a numerical proxy for the biopsychobehavioral makeup of the individual which affects insurance losses. It yields additional information about one's responsibility and stability, stress level, and distractibility, all of which influence the amount of insured losses paid by the insurance company. »<sup>40</sup>

<sup>38</sup> Patrick L. BROCKETT et Linda L. GOLDEN, « Biological and Psychobehavioral Correlates of Credit Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works », (2007) 74 *The Journal of Risk and Insurance* 23.

<sup>39</sup> *Id.*, 25.

<sup>40</sup> *Id.*, 26.



Quoique cette approche nous semble plutôt dangereuse en raison des possibilités qu'elle crée, nous sommes prêts à y voir une position qu'on peut défendre. Là où le bât blesse cependant, c'est dans le standard qui est retenu pour justifier toute la démarche. En effet, les auteurs mentionnent en conclusion de leur texte :

« Actuaries, recognizing the impossibility of proving “causality” for all useful and predictive variables, have addressed the fact that not all useful predictor variables can be clearly visible as “causal.” The result is a standard that helps determine when a variable is permissible for use and when it is not (Actuarial Standard 12). This standard prescribes that it is not necessary to understand the “black box” mechanism by which a variable predicts losses, *if it can be exhibited statistically to be related to the loss and does not discriminate in an unlawful way*. This has been publicly acceptable for decades, but the acceptability of Actuarial Standard 12 as a standard is questioned with the new underwriting variable “credit score”. »<sup>41</sup>

Il serait donc acceptable du point de vue de l'actuaire d'utiliser une variable si elle ne crée pas de discrimination inacceptable du point de vue du juriste pendant que le juriste met en place des exceptions qui admettent explicitement la discrimination si elle est acceptable du point de vue de l'actuaire ! Face à un tel raisonnement circulaire, il est permis de se demander qui est réellement responsable de l'acceptation de l'utilisation du dossier de crédit en pratique. Cela est d'autant plus troublant dans un contexte où il semble que la population ne soit pas universellement confortable avec ce choix.

Dans la même veine, l'exemple de l'utilisation du sexe du preneur est intéressant. Il serait difficile d'argumenter que le fait d'être une femme, à lui seul, change le risque<sup>42</sup>. On prétendra plutôt que cette condition est cor-

<sup>41</sup> *Id.*, 36 (nos italiques).

<sup>42</sup> Autrement que par le biais d'un point de vue *impressionniste* pour reprendre le terme de l'Honorable L'Heureux-Dubé qui écrivait dans sa dissidence de l'affaire *Zurich Compagnie d'Assurances c. Ontario (Commission des droits de la personne)*, 1992 CanLII 67 (C.S.C.), p. 66 : « L'industrie des assurances ne sait tout simplement pas si l'âge, le sexe et l'état matrimonial ont un lien de causalité avec le risque élevé et elle tente de combler cette lacune en se fiant aux mythes et aux stéréotypes ».

rélée avec des comportements qui, à leur tour, influencent le risque. En ce sens, le RTA 2007 contenait le passage suivant :

« Certaines personnes œuvrant dans l'industrie avancent que la fréquence des réclamations plus élevée chez les conducteurs féminins s'explique en partie par le fait que les femmes achètent plus de couvertures d'assurance pour leur véhicule que les hommes. Or, si une personne n'est pas assurée pour la garantie « collision » par exemple, cette dernière n'est pas assurée pour ses propres dommages. Donc, si elle est responsable d'un accident, elle ne pourra pas réclamer à son assureur les dommages subis à son véhicule. Par conséquent, [...] il est normal que la fréquence de réclamations totale soit plus élevée que celle des hommes. »<sup>43</sup>

Et le rapport indiquait également que :

« Par contre, selon certaines sources œuvrant dans le secteur de l'assurance automobile au Québec, l'écart entre les hommes et les femmes tendrait à diminuer en ce qui a trait aux primes qui leur sont demandées. Ces mêmes sources constatent que les femmes sont beaucoup plus nombreuses sur les routes du Québec qu'il y a 30 ans et qu'elles ont à jongler avec des emplois du temps serrés. Ainsi, pour certains assureurs, il n'est pas exclu que le sexe ne soit plus un facteur aussi prédictif que par le passé. *À ce moment-là, il était légitime d'accorder un rabais additionnel aux conducteurs féminins, mais maintenant pour ces assureurs, ce rabais ne semble plus être aussi justifié.* »<sup>44</sup>

On soulevait ainsi la possibilité de mettre fin à une pratique *a priori* discriminatoire, mais exceptionnellement légalisée par une modification à la *Charte des droits et libertés de la personne*<sup>45</sup>. Toutefois, il est

<sup>43</sup> GOUVERNEMENT DU QUÉBEC, AUTORITÉ DES MARCHÉS FINANCIERS, préc., note 9, p. 40.

<sup>44</sup> *Id.*, p. 44 (nos italiques).

<sup>45</sup> L.R.Q., c. C-12., art. 20.1 qui énonce : « Dans un contrat d'assurance ou de rente, un régime d'avantages sociaux, de retraite, de rentes ou d'assurance ou un régime universel de rentes ou d'assurance, une distinction, exclusion ou préférence fondée sur l'âge, le sexe ou l'état civil est réputée non discriminatoire lorsque son utilisation est légitime et que le motif qui la fonde constitue un facteur de détermination de risque, basé sur des données actuarielles ». Soulignons que le cœur du débat dans l'affaire *Zurich Compagnie d'Assurances c. Ontario (Commission des droits de la personne)*, [1992] 2 R.C.S. 321 qui portait sur le caractère raisonnable et de bonne foi de la discrimination opérée par les assureurs est écarté par l'article 20.1 qui, en raison de son libellé, laisse le champ libre aux actuaires. Cependant, la considération de la « bonne foi » des parties, exigée par ailleurs en Ontario en vertu de l'article 22 du *Code des droits de la personne*, L.R.O. 1990, Ch. H.19 est loin d'être une panacée et l'intrusion de cette notion morale dans le droit des assurances peut faire l'objet de critiques. En ce sens, voir : André BÉLANGER,

permis de douter de la fin prochaine de cette pratique<sup>46</sup> dans la mesure où le RTA 1980 énonçait déjà :

« Si l'on considère que les rôles traditionnels des deux sexes et leurs habitudes de conduite se transforment et tendent à se rapprocher, *on ne peut que conclure à la désuétude de plus en plus évidente de la classification selon le sexe.* »<sup>47</sup>

Il s'agit là d'un débat qui a créé des remous dans plusieurs juridictions. Par exemple, Miller s'insurge suite au fait qu'un tribunal de Pennsylvanie ait jugé discriminatoire l'utilisation du sexe en tarification de l'assurance automobile. Il souligne que :

« The Pennsylvania Supreme Court has thus declared the use of gender classifications to set insurance rates illegal under the Pennsylvania Equal Rights Amendment. This prohibition will not reduce or eliminate discrimination in insurance pricing; instead it will engender increased discrimination, since cost differences between the sexes cannot be considered by actuaries in pricing insurance coverage. Gender, at least in Pennsylvania, is no longer a variable to be allowed. This policy treats unequals as equals in the name of equality treatment. »<sup>48</sup>

Toutefois, cette position et son affirmation selon laquelle « there will now be a redistribution of income or wealth from women to men »<sup>49</sup> sont ultimement entièrement appuyées sur le passage suivant :

---

« Interrogations critiques sur l'*uberrima fides* dans le contrat d'assurance », (2010) 77 *Assurances et gestion des risques* 203.

<sup>46</sup> À titre d'exemple des débats sur la question, voir : Jill GAULDING, « Race, Sex and Genetic Discrimination in Insurance: What's Fair? », (1995) 80 *Cornell L. Rev.* 1646; Kenneth S. ABRAHAM, « Efficiency and fairness in insurer risk classification », (1985) 71 *Va. L. Rev.* 403. Il va sans dire que la réflexion en la matière est intimement liée au recours accru à la segmentation dans la tarification et à l'effacement progressif d'une mutualité assurantielle basée sur des principes de solidarité; voir respectivement : Catherine PARIS, *Les dérives de la segmentation en assurance*, Bruxelles, Les Éditions Larcier, 2005; André BÉLANGER et Joelle MANEKENG TAWALLI, « Le spectre de la mutualité dans le contrat d'assurance », (2009) 39(2) *R.G.D.* 8.

<sup>47</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, préc., note 35, p. 42 (nos italiques).

<sup>48</sup> Richard A. MILLER, « Discrimination by Gender in Automobile Insurance: A Note on Hartford Accident and Indemnity Co. v. Insurance Commissioner », (1985) 23 *Duq. L. Rev.* 621, 630.

<sup>49</sup> *Id.*, 627.

« Statistical evidence clearly indicates that young men have more and bigger accidents than young women; costs of providing insurance to young men are thus higher than those costs for young women, and the price of providing such insurance should reflect those differences in costs. »<sup>50</sup>

Or, rien de plus dans le texte de Miller ne vient soutenir cette position selon laquelle le coût, pour l'assureur, de la couverture du risque que représentent un homme et une femme est réellement différent. C'est plutôt la panacée statistique, à laquelle nul ne saurait s'opposer, qui a été invoquée une fois de plus.

Cela est d'autant plus cocasse que des « experts » dans d'autres États en arrivent à des conclusions différentes. Par exemple, on lisait en 1984 dans un ouvrage à caractère hautement statistique produit suite à l'implantation d'une interdiction de différenciation de tarifs basée sur le sexe dans l'État du Michigan que :

« The NAII average premium data and the Michigan residual market survey information show that elimination of gender as a rating factor for automobile insurance is not nearly as radical an idea as it might appear to some on its surface. [...] Also, most people find it reasonable and *fair* to base rates on a driver's actual performance instead of on gender. »<sup>51</sup>

Par ailleurs, et de façon à retourner plus directement à l'idée principale de nos propos, l'exemple de l'utilisation du sexe met en lumière un autre problème majeur qui se présente, si ce n'est dans la tarification elle-même, mais dans la justification de celle-ci. Certains passages cités expriment l'idée que la différence perçue entre le risque représenté par les hommes et celui des femmes peut, du moins en partie, être expliquée par des différences dans leurs comportements moyens (choix de la couverture, fréquence de conduite, etc.). Tarifier en fonction du sexe pourrait donc être injuste pour les individus ayant des comportements différents de ceux qui sont habituels dans leur groupe respectif si on ne corrigeait<sup>52</sup> pas notre évaluation du risque à l'aide d'autres variables. Par exemple, il faudrait ajuster la présentation de la sinistralité « homme/femme » pour l'étendue de la couverture et la fréquence de conduite avant de conclure à un effet véritable

<sup>50</sup> *Id.*, 625 et 626.

<sup>51</sup> Frances K. WALLACE, « Unisex Automobile Rating: The Michigan Experience », (1984) 3 *Journal of Insurance Regulation* 127,139 (nos italiques).

<sup>52</sup> Le terme *contrôler* est aussi employé dans ce contexte.

du sexe. C'est d'ailleurs là un problème qui a déjà été présenté dans la littérature juridique<sup>53</sup>.

Mais pourquoi ne faudrait-il pas aussi procéder à un ajustement selon l'âge, le type de véhicule, la profession et *tutti quanti*? À l'ultime, ne pourrions-nous jamais être certains qu'il n'existe pas une combinaison de facteurs qui effacerait complètement l'effet du sexe et serait, contrairement à celui-ci, sous le contrôle de l'assuré? Ces questions mettent en évidence une problématique liée aux interactions et à la *dangerosité* de présenter des comparaisons basées sur une seule variable qui est fondamentale et dont l'industrie est bien consciente. En ce sens, il a été écrit que :

« One way analyses can be distorted by correlations between rating factors. For example, young drivers may in general drive older cars. A one-way analysis of age of car may show high claims experience for older cars, however this may result mainly from the fact that such cars are in general driven more by high risk younger drivers. »<sup>54</sup>

Qui plus est, peu importe le nombre de facteurs étudiés, il sera toujours possible que subsistent des variables ignorées qui s'avèreraient pertinentes. Tout comme il est possible qu'une de celles-ci rende caduque celle qui apparaissait importante antérieurement.

#### IV. Choix des méthodes

Une autre illustration de la présence de subjectivité dans la statistique se situe dans le choix des méthodes. En effet, il existe parfois plusieurs options différentes parmi lesquelles il faut choisir et dont aucune ne peut être objectivement déclarée la seule raisonnable. Un exemple se présente dans le contexte de la détermination des valeurs des paramètres d'un modèle de régression simple. La plus connue des techniques utilisées est sans doute celle de la minimisation des carrés moyens ou *ordinary least squares* (OLS). Elle est fréquemment enseignée dans les cours d'introduction à la statistique et se retrouve également dans les manuels avancés<sup>55</sup>. Nous la présentons ici

<sup>53</sup> Voir par exemple : Lea BRILMAYER, Douglas LAYCOCK & Teresa A. SULLIVAN, « The efficient use of group averages as nondiscrimination : a rejoinder to professor Benston », (1983) 50 *U. of Chicago L. Rev.* 222.

<sup>54</sup> D. ANDERSON et al., préc., note 27, p. 4.

<sup>55</sup> Voir par exemple : Ronald CHRISTENSEN, préc., note 33, lire notamment la section 2.2.

dans le contexte du modèle de régression linéaire simple, un modèle bien moins complexe que ce qui peut maintenant être utilisé en assurance.

Ce modèle sert à expliquer ou prédire la valeur d'une variable selon une relation linéaire avec une autre par un modèle de la forme suivante :

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

où les  $Y_i$  sont les observations de la variable à prédire, les  $X_i$  sont les observations de la variable prédictive, les  $\varepsilon_i$  sont des termes d'erreur et  $\alpha$  et  $\beta$  sont les *paramètres*. On peut l'utiliser par exemple pour expliquer le revenu annuel en fonction des années d'études, le poids en fonction de la grandeur, le coût d'une course de taxi en fonction de la distance parcourue, etc. L'objectif est alors de déterminer, à partir d'observations, les valeurs optimales pour les paramètres  $\alpha$  et  $\beta$ . La technique habituelle pour ce faire, celle des OLS, revient à calculer l'erreur de prédiction pour chaque donnée,  $(\alpha + \beta X_i) - Y_i$ , à élever ensuite la valeur de ces erreurs au carré, en calculer la somme et minimiser la fonction de  $\alpha$  et  $\beta$  ainsi obtenue.

Le lecteur aguerri demandera peut-être pourquoi on a pris le carré des erreurs avant d'en faire la somme. La raison la plus simple est que l'on souhaite accorder un *coût* positif à toutes les erreurs, peu importe leur signe, afin d'éviter que les surestimations<sup>56</sup> ne compensent pour les sous-estimations<sup>57</sup>. Le lecteur encore plus aguerri pourra maintenant demander pourquoi on choisit de prendre l'erreur élevée au carré plutôt que simplement sa valeur absolue qui, elle aussi, accorderait un coût positif à toutes les erreurs. La réponse habituelle sera soit que la technique choisie facilite les calculs nécessaires à l'optimisation, qu'elle correspond à celle qui est suggérée par d'autres raisonnements statistiques<sup>58</sup>, ou encore qu'elle permet d'accorder une plus grande importance aux erreurs de grande envergure<sup>59</sup>. Le lecteur persévérant demandera alors peut-être pourquoi, en

<sup>56</sup> Quand la prédiction est supérieure à la valeur observée, donc  $(\alpha + \beta X_i) > Y_i$  et l'erreur est positive.

<sup>57</sup> Quand la prédiction est inférieure à la valeur observée, donc  $(\alpha + \beta X_i) < Y_i$  et l'erreur est négative.

<sup>58</sup> Par exemple la technique de la maximisation de la vraisemblance qui est d'une complexité supérieure à ce qu'il est pertinent d'examiner ici.

<sup>59</sup> En effet, en minimisant les erreurs au carré, on préfère deux erreurs de 2 unités à la combinaison d'une erreur de 1 unité et d'une erreur de 3 unités puisque  $1^2 + 3^2 > 2^2 + 2^2$ .

vertu de la troisième justification que l'on vient de donner, on ne choisirait pas de minimiser les erreurs élevées à la puissance huit<sup>60</sup>.

La présentation peut sembler obscure – voire farfelue, c'est selon! –, mais l'idée qui doit en ressortir ne l'est aucunement. Elle est reliée au fait que lorsque nous avons parlé, ci-haut, de *valeurs optimales des paramètres*, nous n'avons pas identifié la mesure d'optimalité qui nous intéressait. Or, les résultats dépendent de cette mesure d'optimalité et il n'y a pas de choix *objectivement supérieur* pour celle-ci. Même si la présentation technique faite ici était difficilement accessible au juriste, le passage suivant de Howson et Urbach devrait éclairer notre propos :

« It is said almost universally by classical statisticians that if the regression of  $y$  on  $x$  is linear, then least squares provides the best estimates of the regression parameters. The term “best” in this context intimates some epistemic significance and suggests that the least squares method is for good reason preferable to the infinitely many other conceivable methods of estimation. Many statistics textbooks adopt the least squares method uncritically [...]. »<sup>61</sup>

Et le problème devient encore plus évident à la lumière de cet extrait qui traite de l'option de minimiser la somme des valeurs absolues des erreurs :

« [This] too is often regarded as unsatisfactory. For example, Wonnacott and Wonnacott rejected the method unequivocally, and signalled their emphatic disapproval of the method by giving it the acronym MAD [minimum absolute deviations]. But this highlights a weakness of arguments that are based on intuition, namely, that people often disagree over the intuitions themselves. Thus, unlike the Wonnacotts, Brook and Arnold found the MAD method perfectly sane; indeed they regarded it as “a sensible approach which works well”,

<sup>60</sup> Si un tel lecteur se présentait, il pourrait être intéressé par les considérations suivantes. En choisissant une puissance paire de plus en plus élevée, on obtient des valeurs des paramètres  $\alpha$  et  $\beta$  qui produisent sur l'ensemble des observations des erreurs moyennes de plus en plus élevées, mais des erreurs maximales de plus en plus petites. On pourrait donc croire qu'un groupe d'individus voulant répartir le coût des sinistres de façon plus *solidaire* devrait faire le choix d'utiliser des puissances de plus en plus grandes. Toutefois, une telle approche entraîne d'autres problèmes. Son principal défaut est qu'on obtient alors un modèle de plus en plus fortement influencé par les données éloignées de la moyenne. La présence de données aberrantes devient donc plus problématique et le modèle peut ultimement n'être déterminé que par un petit sous-ensemble des observations.

<sup>61</sup> C. HOWSON et & P. URBACH, préc., note 26, p. 209.

their sole reservation being the practical one that “the actual mathematics is difficult when the distribution of estimates are sought”.<sup>62</sup>

On voit donc qu’il n’y a pas toujours consensus dans les choix des méthodes<sup>63</sup>, et ce, même pour des cas simples. Tous ne sont pourtant pas enclins à laisser transparaître ces débats dans leurs écrits. La firme d’actuaire citée plus tôt mentionne uniquement ce qui suit au sujet du modèle de régression linéaire simple :

« The parameters  $\beta_1, \beta_2, \beta_3$ , which explain best the observed data are then selected. For the classical linear model this is done by minimizing the sum of squared errors (SSE): [...].<sup>64</sup> »

Si des interrogations subsistent quant aux choix des méthodes dans des contextes comme celui de la régression linéaire simple, il est facile de s’imaginer que l’utilisation de modèles plus complexes faite aujourd’hui ouvre la porte à bien des interrogations. En fait, on peut aussi dire que la progression de la statistique appliquée, notamment grâce au développement rapide de l’informatique, est la source même de certaines difficultés. À ce titre, le point suivant présente l’un de ces nouveaux problèmes auxquels nous sommes aujourd’hui confrontés.

## V. Le *data mining* abusif

Le dernier aspect que nous présentons pour illustrer les dangers d’une confiance aveugle dans les statistiques est le phénomène du *data mining* abusif. Celui-ci découle de la recherche des relations statistiquement significatives dans les données faite en négligeant de se préoccuper de la pertinence de telles relations. Le *data mining* abusif est rendu possible par le fait que les développements de l’informatique ont, dans une large mesure, réduit le coût de l’assemblage, du stockage et du traitement des données. Tandis qu’on peut maintenant voir un assureur offrir une prime différenciée en fonction du signe astrologique du preneur<sup>65</sup>, aucun dirigeant de compagnie d’assurance n’aurait investi il y a 30 ans les sommes requises pour explorer cette possibilité. À cette époque, on devait, avant d’investi-

<sup>62</sup> *Id.*, p. 210, références omises.

<sup>63</sup> Ou, à tout le moins, dans leur justification.

<sup>64</sup> D. ANDERSON et al., préc., note 27, p. 7.

<sup>65</sup> Rémi MOREAU, « Faits d’actualité », (2002) 70-3 *Assurances* 491, 503.



guer la relation entre une variable et le phénomène à modéliser, avoir une raison légitime de supposer l'existence d'une telle relation. Ce simple fait limitait les abus. De nos jours, n'importe qui peut, à l'aide d'un ordinateur personnel et d'une base de données suffisamment étoffée, découvrir une relation entre deux variables que l'on n'aurait jamais imaginée auparavant<sup>66</sup>. Il peut arriver que cette nouvelle capacité entraîne des découvertes intéressantes, mais la question de déterminer la frontière entre l'examen d'une relation pertinente et non pertinente, le point où le *data mining* devient abusif, fait une fois de plus appel à notre subjectivité.

Devant cette réalité, un recul critique face aux choix des variables à conserver s'impose. En effet, le statisticien (ou l'actuaire) ne devrait pas ajouter une variable aveuglément à son modèle simplement afin de favoriser l'amélioration apparente de son ajustement, mais bien faire preuve de parcimonie dans ses choix. En ce sens, il a été écrit que: «Adding any factor will improve the fit on the data in question – what matters is whether the improvement is significant given the extra parameterization»<sup>67</sup>. L'abondance de données et la grande capacité de calcul créent donc de nouveaux problèmes, tel l'établissement des variables les plus *justifiées*. Or, comme le passage qui suit en témoigne, les techniques disponibles ne suffisent pas en général pour déterminer le meilleur ensemble de variables:

«It is not generally possible to determine from a single GLM which set of factors are significant since the inclusion or exclusion of one factor will change the observed effects and therefore possibly the significance of other correlated factors in the model.»<sup>68</sup>

Dans la pratique, on proposera par exemple des techniques itératives pour déterminer l'ensemble des variables à retenir. Toutefois, même lorsque

<sup>66</sup> À titre d'exemple, voir: Peter COY, «He who mines data may strike fool's gold», *Business Week* du 16 juin 1997, p. 40: «Finance is rife with wrong-headed data mining. David J. Leinweber, managing director of First Quadrant Corp. in Pasadena, Calif., which manages \$20 billion in assets, likes to illustrate the problem with "Stupid Data-Miner Tricks." For example, he sifted through a United Nations CD-ROM and discovered that historically, the single best predictor of the Standard & Poor's 500-stock index was butter production in Bangladesh.»

<sup>67</sup> D. ANDERSON et al., préc., note 27, p. 38.

<sup>68</sup> *Id.*, p. 51 (nos italiques).

l'on emploie ces techniques, on suggèrera de forcer l'inclusion de certaines d'entre elles qui sont jugées plus importantes :

« To determine the theoretically optimal set of factors, therefore, it is generally necessary to consider an iterated series of models. [...] If a very large number of factors are to be considered it can be impractical to start the factor iteration process with all possible factors in the model. In such cases it is possible to select a model with certain factors which are *known* to be important, and then to test all other excluded factors [...]. »<sup>69</sup>

Toute la question consiste évidemment à déterminer quelles sont ces variables que l'on *sait* importantes. On se demande alors si, dans le cas d'un produit obligatoire distribué par un intérêt privé, la collectivité ne devrait pas pouvoir influencer explicitement ces choix ou si l'on doit donner carte blanche aux assureurs<sup>70</sup>. Bref, il ne faudrait pas hésiter à s'interroger, ne serait-ce qu'au plan normatif, sur l'étendue du rôle que doit jouer le statisticien. C'est d'ailleurs peut-être pour susciter cette réflexion que les rapporteurs écrivaient il y a 15 ans :

« Rappelons que l'argumentation des assureurs pour justifier les pratiques actuelles repose essentiellement sur les indications produites par les statistiques. L'utilisation de statistiques a pour but de présenter une justification objective des pratiques utilisées par les assureurs. Encore faut-il, cependant, que l'utilisation qui en est faite soit appropriée. Les statistiques permettent d'établir des corrélations entre certains facteurs et la probabilité d'avoir un accident. Cependant, des corrélations statistiques entre des facteurs ne signifient pas qu'il y a une relation de cause à effet entre eux. »<sup>71</sup>

<sup>69</sup> *Id.* (nos italiques).

<sup>70</sup> En ce sens, l'Honorable McLachlin écrivait dans sa dissidence de l'affaire *Zurich Compagnie d'Assurances c. Ontario (Commission des droits de la personne)*, préc., note 45, p. 86 : « Cette tendance des autorités réglementaires d'une industrie de voir le monde par les yeux de cette industrie, quelquefois appelée « vision de l'industrie », est compréhensible et peut à certains égards être bénéfique. Toutefois il s'agit d'un bon motif pour que l'assemblée législative confère la détermination des violations des droits de la personne non pas aux autorités réglementaires de l'industrie, mais à un organisme distinct. »

<sup>71</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, *Rapport annuel sur la tarification en assurance automobile*, Québec, 1995, p. 23 (ci-après « RTA 1995 »).

Pour exprimer l'idée selon laquelle la statistique ne doit pas être utilisée seule pour guider nos choix, Howson et Urbach écrivent :

« Classical estimation theory and significance tests, in their various forms, are still immensely influential; they are advocated in hundreds of books that are recommended texts in thousands of institutions of higher education, and required reading for hundreds of thousands of students. [...] Yet, as we have shown, classical “estimates” are not estimates in any normal or scientific sense, and, like judgments of “significance” and “non-significance”, they carry no inductive meaning at all. Therefore, they cannot be used to arbitrate between rival theories or to determine practical policy.»<sup>72</sup>

Quoi qu'il en soit, même en l'absence de tout ce qui a été dit pour mettre en relief le fait que la simple utilisation de la statistique ne garantit pas un résultat *juste parce qu'objectif*, il y aurait des voix pour défendre la position que celle-ci doit être encadrée et dirigée en certaines circonstances. Par exemple, dans le contexte d'une étude sur le *redlining*, Squires écrivait ce qui suit :

« Whether these underwriting patterns reflect nonrational behavior (that is, racial discrimination with no foundation in terms of objective risk assessment) on the part of a few significant actors in the insurance marketplace or are simply the logical outcome of rational actors responding to objective market signals, or are a combination of both dynamics, some intervention of or alternative to these patterns appears justifiable, if not essential, given the substantial social costs associated with these practices.»<sup>73</sup>

Le problème revient ainsi à déterminer comment et qui doit encadrer le travail du statisticien ? Or, si des experts s'expriment en ces termes, il va sans dire que le public en général risque fort d'être tout simplement désemparé quand vient le temps d'évaluer la pertinence des choix qui sont faits. Au strict plan juridique, cette situation ajoute au déséquilibre contractuel dans la relation entre l'assureur et preneur ; déséquilibre causé notamment par l'accès aux données dont disposent les assureurs, mais non les assurés. Tout cela n'est pas sans conséquence étant donné, faut-il le rappeler, que le chapitre A est obligatoire au Québec.

<sup>72</sup> C. HOWSON & P. URBACH, préc., note 26, p.181 et 182.

<sup>73</sup> G.D. SQUIRES, & W. VÉLEZ, préc., note 19, 75.

## VI. Des choix lourds de conséquences

S'il est permis de douter que la statistique garantisse une tarification *juste* pour un produit obligatoire, dans la mesure où le législateur s'abstient d'intervenir, il est justifié de se demander si la situation ne risque pas de produire des iniquités importantes. Nous ne prétendons pas être au fait de l'existence d'injustices graves ayant actuellement cours au Québec. Nous estimons toutefois à propos d'encourager la réflexion sur la pertinence du *statu quo*. Une telle réflexion semble en effet nécessaire afin de ne pas simplement confier une question de justice sociale, du moins pour l'assurance obligatoire, uniquement aux actuaires. Ce serait alors faire le choix d'une « solution technocratique [constituée par] la vieille idée qu'une bonne politique est une politique indexée sur la science »<sup>74</sup>. Il semble pourtant que ce soit la voie dans laquelle nous soyons engagés. Un certain malaise à l'idée d'intervenir dans le contrôle de la tarification ressort d'ailleurs du RTA 1978 :

« [le] législateur a donc confié au surintendant un rôle de surveillance des pratiques de tarification en assurance automobile. *Ce rôle n'est pas, il est important de le préciser, un rôle de contrôle* de la tarification. Ainsi, le surintendant n'a pas le pouvoir d'approuver ou de désapprouver les tarifs, pas plus qu'*il ne peut intervenir directement* dans le processus de tarification en imposant aux assureurs des normes. »<sup>75</sup>

Néanmoins, le rapport principal qui avait engendré la réforme quatre ans plus tôt faisait la recommandation suivante :

« Que le Surintendant des assurances ait pouvoir d'exiger d'un assureur qu'il fournisse une justification concernant les différents aspects de sa tarification, et qu'il puisse même suspendre le permis d'une compagnie en certaines circonstances. »<sup>76</sup>

Or, on peut se demander, dans les contextes juridique, assurantiel et économique en vigueur, s'il est aujourd'hui concevable d'utiliser ce pouvoir et quel barème serait pertinent pour apprécier la justification apportée.

<sup>74</sup> F. EWALD, préc., note 17, p. 423.

<sup>75</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, préc., note 6, p. 14 (nos italiques).

<sup>76</sup> GOUVERNEMENT DU QUÉBEC, préc., note 10, p. 47.

Finalement, on peut aussi constater qu'il n'y a pas que le rôle du législateur qui se soit dégradé. L'assuré lui-même est progressivement expulsé du processus. Par exemple, tandis que le Fichier central des sinistres automobiles (FCSA) devait à l'origine servir à vérifier l'*exactitude* des informations concernant les sinistres passés fournies par l'assuré, il semble désormais constituer pour certains un substitut au dialogue entre assureur et assuré. Dans l'affaire *Fillion c. La Capitale*, le juge résume ainsi les témoignages :

« [59] La preuve en demande et en défense de la part de la compagnie d'assurance La Capitale est nettement contradictoire, particulièrement quant à savoir si l'agente d'assurance, Mme Rodrigue, a posé ou non des questions à M. Fillion, concernant les réclamations antérieures, au cours des cinq dernières années.

[60] Comme on l'a vu, M. Fillion affirme que Mme Rodrigue ne lui a jamais demandé d'informations à ce sujet, sinon il aurait donné les renseignements pertinents. *Selon lui, la compagnie d'assurance n'a qu'à consulter le fichier central.* Son épouse, qui était présente lors d'une conversation téléphonique de son mari avec Mme Rodrigue, déclare qu'elle n'a pas entendu quoique ce soit à ce sujet.»<sup>77</sup>

Du point de vue de l'assuré, il n'est tout simplement plus de son ressort de donner l'information sur son historique de sinistres<sup>78</sup>, puisque celle-ci est trouvée par l'assureur en s'adressant à une source externe, le FCSA. La tendance de plus en plus forte à consulter la cote de crédit des assurés<sup>79</sup> a le même effet puisque le preneur est généralement peu au fait de son dossier de crédit, voire même de la façon dont il est constitué. Tout cela contribue à diminuer le rôle du cocontractant qu'est le preneur lors de la proposition d'assurance. Au lieu de fournir de l'information, celui-ci se limite à autoriser une enquête à son sujet. La discussion entre assureur et assuré est donc remplacée par un simple branchement à des bases de

<sup>77</sup> *Fillion c. La Capitale*, [2001] R.R.A. 1108 (C.Q.) (nos italiques).

<sup>78</sup> Notons au passage qu'un glissement s'est fait de la notion d'historique de sinistres à historique de réclamations. Pour certains, l'assureur ne s'intéresse plus aux *sinistres* qui ont été subis, mais bien aux *réclamations* qui ont été faites (et donc, en un certain sens, aux sinistres qu'il a lui-même subis).

<sup>79</sup> Le passage suivant témoigne de cette tendance : « The personal lines insurance industry in the US has been using credit-based insurance scoring for over a decade. A 2001 Conning & Company survey reported that 92 % of the respondents of a survey of the 100 largest personal automobile insurance writers in the US use some form of credit scoring », D. ANDERSON et al., préc., note 27, p. 82.

données où sont fichés les assurés. Il est désormais raisonnable d'imaginer le jour où les seules questions posées par l'assureur à l'assuré au moment de la négociation du contrat seront les options souhaitées, le numéro de permis de conduire, la date de naissance et l'autorisation à consulter telle ou telle banque de données. Qui plus est, ces bases de données sont aujourd'hui habituellement gérées par des intérêts privés<sup>80</sup> et paraissent, dans une large mesure, hermétiques pour les assurés, – bien qu'elles proposent des mécanismes permettant aux particuliers d'en vérifier l'exactitude<sup>81</sup>.

Aussi, relativement au fait que les automobilistes québécois ont le droit de consulter le FCSCA afin de vérifier l'exactitude des renseignements qu'il contient à leur sujet (et qui influencent la prime qui leur est demandée), on lisait dans le RTA 2008 le passage suivant :

« Dans la mesure où l'inscription au FCSCA est obligatoire à la suite d'un accident et que ce fichier contient des renseignements personnels sur les conducteurs, des mécanismes ont été mis en place afin d'en assurer la protection. En effet, conformément aux dispositions de la Loi sur l'accès aux documents des organismes publics et à la protection des renseignements personnels (L.R.Q., c. A-2.1), un titulaire de permis de conduire peut demander une confirmation et une rectification, le cas échéant, des données inscrites à son dossier au FCSCA. »<sup>82</sup>

Toutefois, le rapport ne présente que le nombre total de consultations du fichier, qui se situe autour de 1 600 000, mais ne dit pas combien de ces consultations sont faites par des particuliers. Le RTA 1995 présentait cette information. À l'époque, des 837 140 consultations, seulement 7 019 avaient été faites par des titulaires de permis de conduire<sup>83</sup>. On se demande alors s'il est surprenant que le preneur ne se sente plus partie entière à la relation contractuelle d'assurances. En effet, quelles sont les conséquences de ce fait dans le cadre de *l'achat* d'une protection obligatoire ?

<sup>80</sup> Même le FCSCA, qui est la propriété de l'Autorité des marchés financiers, est géré par une entreprise privée. C'est l'article 178 de la *Loi sur l'assurance automobile* (L.R.Q., c. A-25) qui permet cette situation.

<sup>81</sup> On peut toutefois souligner qu'il semble plus compliqué d'obtenir sa propre cote de crédit sans frais qu'en payant les frais associés à des services supplémentaires.

<sup>82</sup> GOUVERNEMENT DU QUÉBEC, AUTORITÉ DES MARCHÉS FINANCIERS, *Rapport annuel sur la tarification en assurance automobile*, Québec, 2008, p. 8 et 9 (ci-après « RTA 2008 »).

<sup>83</sup> GOUVERNEMENT DU QUÉBEC, AUTORITÉ DES MARCHÉS FINANCIERS, préc., note 71, p. 13.



\*  
\* \*

Il ne faut pas s'étonner de la dominance du discours de la *justice actuarielle*. Cette position est souvent plus facile à défendre que la position contraire qui fait appel d'un côté au doute et à la remise en question et de l'autre, à la notion apparemment dépassée de solidarité mutualiste. Chose certaine, le débat visant à établir l'ampleur du rôle que devrait prendre le législateur dans un domaine comme celui de l'assurance automobile ne peut être *gagné* objectivement. Par conséquent, la situation nécessite de s'interroger sur divers choix de société. Le législateur exprimait il y a près de 30 ans le souhait que le public, sans nécessairement se les approprier complètement, soit à tout le moins confortable avec les mécanismes en place :

« Les distinctions entre les assurés dans un système de classification doivent être acceptées par la société. Ces distinctions doivent être justes et traiter les assurés d'une façon uniforme. »<sup>84</sup>

Plus récemment, Ewald écrivait :

« [le] concept de solidarité (au sens de l'assurance) est un concept plus fort, plus exigeant que celui de mutualité. Mutualité est un concept technique quand solidarité est un concept moral. Toute mutualité n'est pas solidaire (au sens de l'assurance). »<sup>85</sup>

Certes, il est possible que nous souhaitions collectivement un système d'assurance obligatoire qui fait abstraction de cette solidarité et dont les règles – notamment juridiques – ne s'en remettent qu'à une doctrine de la technique actuarielle. Il serait toutefois regrettable que ce choix soit fait par inconscience ou par insouciance, voire que les règles juridiques y fassent référence par manque de rigueur<sup>86</sup>. Les apports de la statistique au processus assurantiel sont indéniables. Il faut simplement se rappeler à titre de juristes et de citoyens qu'elle constitue un excellent serviteur, mais un très mauvais maître.

<sup>84</sup> GOUVERNEMENT DU QUÉBEC, INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES, préc., note 35, p. 39.

<sup>85</sup> F. EWALD, préc., note 16, à la page 415.

<sup>86</sup> A. BÉLANGER et J. MANEKENG TAWALLI, préc., note 46.

