

## DÉVELOPPEMENT D'UN SYSTÈME EXPERT DE DÉTECTION AUTOMATIQUE DE LA FRAUDE À L'ASSURANCE AUTOMOBILE

El Bachir Belhadji et Georges Dionne

Volume 67, numéro 2, 1999

URI : <https://id.erudit.org/iderudit/1105264ar>

DOI : <https://doi.org/10.7202/1105264ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0004-6027 (imprimé)

2817-3465 (numérique)

[Découvrir la revue](#)

Citer cet article

Belhadji, E. & Dionne, G. (1999). DÉVELOPPEMENT D'UN SYSTÈME EXPERT DE DÉTECTION AUTOMATIQUE DE LA FRAUDE À L'ASSURANCE AUTOMOBILE. *Assurances*, 67(2), 251-274. <https://doi.org/10.7202/1105264ar>

Résumé de l'article

Le but de cette étude est de développer un outil d'aide à la décision permettant aux enquêteurs des compagnies d'assurance d'être mieux équipés pour combattre la fraude à l'assurance. Cet outil est basé sur l'utilisation de façon systématique des indicateurs de fraude. Dans une première étape, nous proposons une procédure afin d'isoler les indicateurs les plus significatifs pour prédire la probabilité qu'un dossier soit frauduleux. Nous avons appliqué la procédure aux données recueillies de l'enquête Dionne-Belhadji (1996). Le modèle nous a permis de constater que 19 des 50 indicateurs utilisés étaient significatifs pour prédire la probabilité de fraude. Nous avons également discuté de la précision et de la capacité de détection du modèle. Cette discussion avait comme point de référence les taux de détection obtenus des enquêteurs ayant participé à l'enquête. Or, comme démontré dans Caron-Dionne (1997), il est possible que ces taux représentent une sous-estimation de la fraude.

Dans une seconde étape, nous avons préparé un logiciel qui permet d'utiliser les résultats du modèle statistique afin de calculer les probabilités de fraude des dossiers, et de décider de l'opportunité d'effectuer ou non une enquête approfondie. Ce logiciel contient l'équation mathématique et les valeurs des paramètres calculés par le modèle Probit. Comme indiqué dans le rapport, ces paramètres reflètent les données des entreprises qui ont participé à l'enquête et non nécessairement celles d'une compagnie en particulier. Il n'est pas évident que les mêmes indicateurs soient significatifs, ni même que les valeurs des coefficients soient les mêmes pour un assureur en particulier. Il est recommandé de refaire une enquête systématique à partir des dossiers de l'assureur qui veut utiliser le logiciel.

Une fois le logiciel adapté à un assureur, celui-ci peut être utilisé facilement par les enquêteurs; il s'agira d'entrer les indicateurs présents dans les dossiers. Ce logiciel calculera la probabilité de fraude du dossier et aidera l'enquêteur à décider de la pertinence d'effectuer une enquête approfondie.

Une disquette contenant le logiciel proposé est disponible auprès des auteurs. Elle peut être utilisée sur un «pc» doté du système d'exploitation Windows95/NT et d'un navigateur Web. Une procédure d'installation est également disponible.

# DÉVELOPPEMENT D'UN SYSTÈME EXPERT DE DÉTECTION AUTOMATIQUE DE LA FRAUDE À L'ASSURANCE AUTOMOBILE

El Bachir Belhadji et Georges Dionne

## RÉSUMÉ

Le but de cette étude est de développer un outil d'aide à la décision permettant aux enquêteurs des compagnies d'assurance d'être mieux équipés pour combattre la fraude à l'assurance. Cet outil est basé sur l'utilisation de façon systématique des indicateurs de fraude. Dans une première étape, nous proposons une procédure afin d'isoler les indicateurs les plus significatifs pour prédire la probabilité qu'un dossier soit frauduleux. Nous avons appliqué la procédure aux données recueillies de l'enquête Dionne-Belhadji (1996). Le modèle nous a permis de constater que 19 des 50 indicateurs utilisés étaient significatifs pour prédire la probabilité de fraude. Nous avons également discuté de la précision et de la capacité de détection du modèle. Cette discussion avait comme point de référence les taux de détection obtenus des enquêteurs ayant participé à l'enquête. Or, comme démontré dans Caron-Dionne (1997), il est possible que ces taux représentent une sous-estimation de la fraude.

Dans une seconde étape, nous avons préparé un logiciel qui permet d'utiliser les résultats du modèle statistique afin de calculer les probabilités de fraude des dossiers, et de décider de l'opportunité d'effectuer ou non une enquête approfondie. Ce logiciel contient l'équation mathématique et les valeurs des paramètres calculés par le modèle Probit. Comme indiqué dans le rapport, ces paramètres reflètent les données des entreprises qui ont participé à l'enquête et non nécessairement celles d'une compagnie en particulier. Il n'est pas évident que les mêmes indicateurs soient significatifs, ni même que les valeurs des coefficients soient les mêmes pour un assureur en particulier. Il est recommandé de refaire une enquête systématique à partir des dossiers de l'assureur qui veut utiliser le logiciel.

---

### Les auteurs :

El-Bachir Belhadji est conseiller à Conseils et vérification Canada.

Georges Dionne est titulaire de la Chaire de gestion des risques, professeur de finance à l'École des Hautes Études Commerciales et chercheur associé au Centre de recherche sur les transports, Université de Montréal.

Cette recherche a été rendue possible grâce au financement accordé par le Bureau d'Assurance du Canada (BAC-Québec). Nous remercions M. Raymond Medza pour sa collaboration à différentes étapes de l'étude, ainsi que MM. Gilles Chevrier de Axa Assurances et Gilles Beaulieu de Axa Boréal pour leur aide dans la recherche de certaines informations utiles. Un arbitre nous a aidés à améliorer la première version.

Une fois le logiciel adapté à un assureur, celui-ci peut être utilisé facilement par les enquêteurs; il s'agira d'entrer les indicateurs présents dans les dossiers. Ce logiciel calculera la probabilité de fraude du dossier et aidera l'enquêteur à décider de la pertinence d'effectuer une enquête approfondie.

Une disquette contenant le logiciel proposé est disponible auprès des auteurs. Elle peut être utilisée sur un «pc» doté du système d'exploitation Windows95/NT et d'un navigateur Web. Une procédure d'installation est également disponible.

*Mots clés* : Fraude à l'assurance, indicateurs, Probit, logiciel

#### ABSTRACT

*The goal of this study is to develop a tool to aid insurance company adjusters in their decision making and to ensure that they are better equipped to fight fraud. This tool is based on the systematic use of fraud indicators. We first propose a procedure to isolate those indicators which are most significant in predicting the probability that a claim may be fraudulent. We applied the procedure to data collected in the Dionne-Belhadji study (1996). The model allowed us to observe that 19 of the 50 indicators used were significant in predicting the probability of fraud. Our study also discusses the model's accuracy and detection capability. The detection rates obtained by the adjusters who participated in the study constitute the reference point of this discussion. As shown in Caron-Dionne (1997), there is the possibility that these rates underestimate the level of fraud.*

*Our second step was to develop a software allowing us to use the results of the statistical model to estimate the probability of fraud in files and to decide whether or not an in-depth investigation should be conducted. This software contains the mathematical equation and the parameters calculated by the Probit model. As indicated in the report, these parameters reflect the data from all the firms having participated in the study and not from any one company in particular. It is not obvious that the same indicators would be significant or even that the coefficients would be the same for an insurer in particular. Any insurer wishing to use the software is advised to carry out a systematic study of the company's own files.*

*Once adapted to the insurer's use, the software can easily be used by claims adjusters. It would then be a matter of entering the indicators present in the files. The software will calculate the probability of fraud in a file and help the adjuster to decide whether an in-depth investigation is warranted.*

*A floppy disk containing the software proposed is available from the authors. It can be used on a PC with a Windows 95/NT system and a Web navigator. Procedures for installing the software are also available from the authors.*

*Keywords* : Insurance Fraud, Indicators, Probit, Software

## ■ INTRODUCTION

Le but de cette recherche est de développer un outil d'aide à la décision qui permettra aux professionnels enquêteurs des compagnies d'assurance de détecter la fraude de façon automatique (son application dans ce rapport est limitée au secteur automobile).

Le travail consiste en la mise en place d'un système informatique simple qui permet à l'enquêteur de générer la probabilité de fraude de chaque dossier étudié. En effet, après avoir déterminé les indicateurs de fraude les plus significatifs, ainsi que leurs poids respectifs, nous avons formulé une fonction qui lie ces indicateurs à une probabilité de fraude.

Afin de rendre le système facile d'utilisation, nous avons développé une interface qui fait le lien entre la fonction décrite ci-dessus et les indicateurs de fraude que l'enquêteur doit entrer sur le système.

Nous proposons le développement d'un logiciel convivial qui aura une interface graphique semblable aux formulaires traditionnellement utilisés lors de sondages. L'utilisateur fournira à l'écran les informations requises pour les différents indicateurs. Le score de fraude (probabilité) apparaîtra en bas de la «feuille questionnaire». Sur la base de cette probabilité, les superviseurs peuvent décider (ou non) de mener une enquête approfondie pour certains dossiers. Une partie complémentaire a été développée afin de connaître la rentabilité d'une poursuite d'enquête pour chaque cas particulier. Cette deuxième partie reste cependant une partie exploratoire, du fait de l'absence de certaines données.

Il est important de souligner que le modèle ne peut pas être appliqué directement par une entreprise car les paramètres utilisés proviennent de calculs effectués avec les données de l'industrie (voir Dionne et Belhadji, 1996). L'utilisateur potentiel devra préalablement refaire une enquête semblable à partir de la totalité de ses dossiers sur une période donnée (ou d'un échantillon aléatoire) et ensuite, recalculer les paramètres correspondants.

Ce rapport est divisé en quatre sections. Dans la première section, nous décrivons la procédure d'échantillonnage des dossiers. Dans la section II, nous revoyons la littérature sur les indicateurs de fraude et nous motivons notre choix des indicateurs retenus. Puis nous calculons les paramètres qui nous permettent de dégager les indicateurs significatifs pour prédire la probabilité qu'un dossier contienne de la fraude. La section III utilise les paramètres (qui sont statistiquement significatifs) pour calculer les probabilités de fraude

des différents dossiers. Cette section explique aussi les procédures à suivre pour se constituer un échantillon de dossiers à réviser, en fonction du budget alloué à la fraude par la compagnie. Finalement, cette section commente l'importance de la fraude si on se fie exclusivement à notre modèle. Une courte conclusion termine l'article.

## ■ PROCÉDURE D'ÉCHANTILLONNAGE

Dans cette section, nous étudierons, de façon successive, le degré de représentativité de notre échantillon, la manière dont l'enquête s'est déroulée, les notions de fraude suspecte et établie, ainsi que les méthodes qui nous ont permis de sélectionner l'échantillon actuel.

### Représentativité de l'échantillon

Après avoir étudié les parts de marché des assureurs en collaboration avec le Bureau d'assurance du Canada, nous avons retenu vingt compagnies d'assurances parmi les plus importantes dans le secteur «assurance automobile» au Québec. Selon les données du «Rapport annuel sur les assurances» (1994) de l'Inspecteur général des institutions financières, les vingt compagnies retenues détenaient 78,5 % du marché de l'assurance automobile au Québec. Dix-huit (18) des vingt (20) entreprises retenues ont accepté de participer à l'enquête, soit un taux de participation de quatre-vingt-dix pour cent (90 %). En tenant compte des parts relatives de marché de ces dix-huit entreprises, nous avons donc obtenu la participation d'entreprises représentant soixante-dix pour cent (70 %) du marché, ce qui est très satisfaisant. Nous pouvons donc conclure que notre enquête est représentative du marché étudié.

Ces entreprises ont accepté d'analyser en détail des dossiers de réclamations fermés afin de nous transmettre différentes informations utiles à la compréhension du phénomène de la fraude. Leur collaboration nous a également permis d'évaluer, d'une part, l'importance de la fraude et d'étudier, d'autre part, les principaux indicateurs de celle-ci en assurance automobile.

### Déroulement de l'enquête

Les dossiers ont été choisis au hasard parmi tous les dossiers fermés durant la période du 1<sup>er</sup> avril 1994 au 31 mars 1995. Le nombre de questionnaires alloués à une compagnie était proportionnel à sa part de marché. Nous avons suggéré aux entreprises une méthode d'échantillonnage aléatoire (décrite dans Dionne et Belhadji, 1996).

L'enquête a eu lieu au printemps 1995. Le choix des enquêteurs pour répondre aux questionnaires a été fait par les assureurs. Nous leur avons demandé de désigner les enquêteurs qui avaient traité les dossiers échantillonnés. Des procédures rigoureuses ont été mises en place pour conserver l'anonymat des enquêteurs et des compagnies.

Les dix-huit compagnies nous ont retourné 2 509 questionnaires complétés, ce qui correspond à un taux de réponse de 98 % pour les 18 compagnies participantes. Le tableau 1 donne la répartition des dossiers reçus en fonction des compagnies qui ont répondu.

<b>TABLEAU I NOMBRE DE QUESTIONNAIRES REÇUS DES DIFFÉRENTES COMPAGNIES</b>		
<b>No Cie</b>	<b>Nombre de questionnaires complétés</b>	<b>% échantillon</b>
1	100	4
2	89	4
3	166	7
4	108	4
5	274	11
6	48	2
7	370	15
9	45	2
10	39	2
11	60	2
12	275	11
13	346	14
15	69	3
16	39	2
17	100	4
18	40	2
19	161	6
20	180	7
<b>Total</b>	<b>2 509</b>	<b>100</b>

## **Fraude établie et fraude suspecte dans les données de l'échantillon**

Le tableau 2 donne le nombre de cas de fraude établie et suspecte. Ces nombres reflètent l'opinion des enquêteurs qui ont répondu aux questionnaires. Par fraude établie, nous entendons qu'il y a eu effectivement fraude pour un dossier particulier, qu'il y ait eu ou non poursuite. Une fraude est suspecte si l'enquêteur avait des soupçons de fraude pendant le traitement du dossier mais le dossier n'a pas fait l'objet d'une enquête plus approfondie afin d'établir la fraude.

## **Échantillon retenu pour l'analyse des indicateurs**

Une des questions à laquelle il fallait répondre (question 17) consistait en un dénombrement de certaines caractéristiques de dossiers que nous avons alors appelés «indicateurs de fraude». Une liste de 50 indicateurs était proposée aux enquêteurs, qui devaient sélectionner ceux correspondant à chaque dossier en inscrivant leurs numéros respectifs. Lors de la saisie de données, il a été remarqué que certains enquêteurs ne rentraient jamais d'indicateurs, et ce, quel que soit le dossier qu'ils étudiaient. Il y avait deux raisons qui pouvaient expliquer ces actions : soit qu'aucun des 50 indicateurs ne s'appliquait à aucun des dossiers de ces enquêteurs, ou bien que ces derniers ne voyaient pas la pertinence de la question et ne mettaient jamais d'indicateurs. Notons que la liste des 50 indicateurs a peut-être découragé certains enquêteurs, qui ont alors préféré s'abstenir de répondre à la question 17. Afin de résoudre ce problème, nous avons décidé de limiter notre échantillon à l'ensemble des dossiers qui ont été complétés par des enquêteurs ayant inscrit au moins un indicateur parmi tous les dossiers qu'ils ont remplis. Les dossiers d'un enquêteur particulier étaient tous retenus si au moins un des dossiers comportait au moins un indicateur. Cette procédure nous a permis d'écartier un possible biais des réponses associé au fait que des enquêteurs n'ont pas répondu à la question. Cette façon de procéder a le mérite d'enlever les dossiers de ceux qui n'ont pas cru bon de répondre à la question, mais elle est radicale dans le sens qu'elle peut avoir éliminé des dossiers qui ne contenaient pas d'indicateurs, alors que l'enquêteur était disposé à répondre à la question. Dans cet article, nous avons retenu cette méthode, mais d'autres avenues de correction ont été proposées (Belhadji, Dionne, Tharkani, 1998). Enfin, ce processus d'élimination a résulté en un échantillon plus petit que l'échantillon d'origine : on est passé de 2 509 cas à 2 068 cas, dont la répartition est présentée au tableau 2. Nous avons donc perdu 441 observations.

**TABLEAU 2**  
**CLASSEMENT DES DOSSIERS**

Classe- ment	N	S	E	$\frac{S+E}{N+S+E}$	$\frac{E}{N+S+E}$	N+S+E
	Non fraude	Fraude suspectée	Fraude établie			
Total	1 937	113	18	6,33 %	0,87 %	2 068

## ■ IMPORTANCE DES INDICATEURS DE FRAUDE

Le but de l'enquête administrée au printemps 1995 était double : d'une part, elle consistait en une évaluation de l'importance de la fraude à l'assurance automobile<sup>1</sup> et d'autre part, elle visait l'élaboration d'un système expert de détection automatique de la fraude. Cette deuxième partie de l'étude s'appuie sur la détermination d'un ensemble d'indicateurs qui sont significatifs dans la détection et la suspicion de la fraude. C'est dans cette section que nous allons motiver le choix des indicateurs retenus, ainsi que les critères qui nous ont obligés à en écarter certains.

### □ Choix des indicateurs

Comme indiqué plus haut, nous avons présenté aux enquêteurs une liste de 50 indicateurs de fraude. Nous leur avons demandé d'inclure un ou plusieurs indicateurs dans tout dossier présentant une ou plusieurs de ces caractéristiques.

Les 50 indicateurs sont énumérés en annexe et classés par catégories. Les indicateurs inscrits en gras sont ceux qui sont significatifs pour expliquer la probabilité que le dossier soit frauduleux.

Dans la littérature sur les indicateurs, il existe très peu d'évidence empirique qui pourrait nous démontrer scientifiquement la pertinence de certains indicateurs de fraude par rapport à d'autres. C'est aussi la raison pour laquelle nous nous sommes intéressés à ce sujet.

La liste que nous avons mise sur pied (voir annexe) et qui a été incluse dans le questionnaire à l'intention des enquêteurs-régisseurs, était basée sur la littérature existante au moment où nous rédigeons notre questionnaire.

Notre liste a été dressée à partir de deux types de littérature : d'une part, une littérature appliquée où on a essayé de montrer les indicateurs pertinents et, d'autre part, une littérature qui se base sur



le vécu des professionnels de l'assurance. Parmi la littérature appliquée, notons le travail du «*Florida Institute Research Center, 1991*» de l'Université de Floride. Cette étude a porté sur la fraude à l'assurance automobile en Floride. Ses auteurs ont calculé des probabilités conditionnelles de fraude pour chacun des 90 indicateurs proposés dans leur liste. Certains d'entre eux s'appliquaient à très peu de dossiers (moins de 10) et nous avons dû les éliminer. Nous avons retenu tous les indicateurs pertinents aux couvertures (chapitres) offertes au Canada et dont la probabilité conditionnelle de fraude était supérieure à 10 %. Notons cependant que cette étude américaine n'a à aucun moment étudié tous les indicateurs en même temps : elle a seulement calculé des probabilités de fraude étant donné qu'un indicateur déterminé est présent dans le dossier (un seul indicateur à la fois).

L'autre étude, Weisberg et Derrig, 1993, nous a servi à compléter notre liste en choisissant les indicateurs qui étaient significatifs. Cette étude, qui emploie une régression multiple, possède le défaut du choix d'un échantillon non aléatoire de dossiers.

Afin que notre liste contienne une composante professionnelle, nous nous sommes basés sur des études descriptives menées pour le compte du Bureau d'Assurance du Canada, du SACA et d'une note interne sur les indicateurs du *Property and Casualty Claims Services*, mars 1993. L'idée de la sélection était de reprendre les indicateurs communs à plusieurs études différentes.

Enfin, afin de rajouter une composante proche du marché québécois de l'assurance automobile à notre étude, nous avons discuté avec des responsables d'entreprises qui nous ont révélé leurs indicateurs les plus utilisés.

C'est donc à partir de ces études et discussions que la liste des 50 indicateurs qui se trouve en annexe a été dressée.

#### **Critères de limitation du nombre des indicateurs à inclure dans la régression**

Étant donné que le nombre d'indicateurs de notre liste était très élevé, nous avons voulu trouver un moyen de réduire celui-ci afin de pouvoir inclure dans nos régressions un nombre raisonnable d'indicateurs. Le moyen qui nous semblait le plus efficace et qui, selon nous, éliminerait les indicateurs non pertinents, serait de calculer les probabilités conditionnelles de fraude pour chacun des indicateurs. Le tableau 3 indique ces probabilités conditionnelles. La colonne 1 reporte le numéro d'indicateur tel qu'il est reproduit

en annexe. À chaque indicateur, nous avons attribué un nom : celui-ci est inscrit dans la deuxième colonne. La troisième colonne reprend, quant à elle, le nombre total de dossiers où l'indicateur (numéro «1» par exemple) est inscrit. La quatrième colonne est similaire à la colonne qui la précède, sauf que le nombre inscrit est celui des dossiers étiquetés E (fraude établie) ou S (fraude suspecte). Enfin, la dernière colonne retrace la probabilité que le dossier soit frauduleux étant donné qu'un indicateur précis (exemple : *police*) est présent : ce chiffre est obtenu en divisant le nombre de la quatrième colonne par celui de la troisième.

Notons que des intervalles de confiance peuvent être calculés à partir des données du tableau ci-dessous. Pour tous les estimés du tableau, des écarts types peuvent être calculés comme suit :

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

où  $\hat{p}$  est la valeur de l'estimé et « $n$ » est la fréquence de cet indicateur en particulier (3<sup>e</sup> colonne). L'intervalle de confiance sera donc :

$$\left[ \hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}; \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right]$$

où  $Z_{\frac{\alpha}{2}}$  est le  $1 - \alpha/2$  percentile d'une loi normale.

L'écart type de l'estimé de l'indicateur numéro 3, par exemple, se calcule comme suit :

$$\sqrt{\frac{(0,316)(0,684)}{57}} = 0,062$$

L'intervalle de confiance à 95 % de l'estimé de cet indicateur serait donc de :

$$\{0,316 - 1,96(0,062); 0,316 + 1,96(0,062)\}$$

ou bien : {19,5 % ; 43,7 %}

Ce tableau, qui n'était qu'une ligne directrice dans le choix des indicateurs, nous a permis d'éliminer tous les indicateurs qui étaient présents dans 10 dossiers et moins. C'est ainsi qu'il ne faudrait pas interpréter les indicateurs possédant un astérisque (5<sup>e</sup> colonne) comme étant non significatifs car ils ont été écartés pour des raisons d'insuffisance de données. Toute étude ultérieure qui porterait sur un grand nombre de dossiers (beaucoup plus que 2 500 dossiers) devrait les inclure afin de les tester.

**TABLEAU 3**  
**PROBABILITÉ CONDITIONNELLE DE FRAUDE (E OU S)**  
**ÉTANT DONNÉ QUE L'INDICATEUR «X» EST PRÉSENT**

Numéro	Indicateur	Nombre total	Nombre (E + S)	Probabilité conditionnelle %
1	Police	363	34	9,40
2	Mineur	74	6	8,10
3	Incoh	57	18	31,60
4	Vole	32	12	37,50
5	Recent	80	7	8,75
6	Facture	30	3	10,00
7	Interet	71	13	18,31
8	Commer	24	3	12,50
9	Etendu	41	5	12,20
10	Agent	148	9	6,08
11	Diffic	43	16	37,21
12	Occup	7	3	42,86*
13	Recoit	71	12	16,90
14	Petit	15	8	53,33
15	Rapide	56	15	26,79
16	Jargon	48	13	27,08
17	Deplace	19	7	36,84
18	Enthou	17	8	47,06
19	Proche	87	13	14,94
20	Loi68	3	0	0,00*
21	Taxi	17	7	41,18
22	Preuve	19	12	63,16
23	Coupabl	15	5	33,33
24	Histoir	14	7	50,00
25	Tierce	22	2	9,09
26	Docum	21	5	23,81
27	Memega	21	2	9,52
28	Represe	4	1	25,00*
29	Repare	20	3	15,00
30	Temoig	97	19	19,59
31	Nie	13	4	30,77
32	Seul	142	19	13,38
33	Nonide	69	7	10,14
34	Garanti	9	3	33,33*
35	Loue	3	2	66,66*
36	Contact	8	3	37,50*
37	Signe	34	4	11,76
38	Rembour	6	3	50,00*
39	Comptan	33	8	24,24
40	Chomage	18	7	38,88
41	Retrouv	6	3	50,00*
42	Surassu	8	0	0,00*
43	Prim	1	1	1,00*
44	Motel	1	0	0,00*
45	Agressi	26	9	34,62
46	Refuse	15	1	6,66
47	Nerveu	42	17	40,48
48	Beauco	129	12	9,30
49	Titre	5	2	40,00*
50	Precede	7	2	28,57*

\* Il faut être prudent en interprétant ces résultats car ils contiennent 10 cas et moins.

Notons que ce tableau n'est pas suffisant pour indiquer la possibilité de fraude car les différents indicateurs sont pris isolément. Une régression qui prendrait en compte tous les indicateurs présents dans un dossier serait préférable. C'est ce que nous ferons dans les régressions *Probit* qui vont suivre. Elles permettront de déterminer les indicateurs les plus significatifs : ceux-ci sont reproduits en gras dans l'annexe.

## □ **Modèle de régression et résultats**

Le modèle probit que nous utilisons suppose une variable de réponse  $y_i^*$ , définie par la relation suivante :

$$y_i^* = b' x_i + u_i$$

où  $y_i^*$  est non observable. Le vecteur  $x_i$ , représente les indicateurs présents dans le dossier, alors que le vecteur  $b'$  renferme les valeurs de leurs paramètres.

Nous observons cependant la variable binaire  $y$  qui est définie par :

$$y = 1 \quad \text{si } y_i^* > 0$$

$$y = 0 \quad \text{autrement.}$$

Dans notre cas particulier :

$y = 1$  si le dossier a été jugé frauduleux (soupçon ou fraude établie)

$$y = 0 \quad \text{autrement.}$$

Il s'ensuit que :

$$\text{prob}(y_i = 1) = \text{prob}(u_i > -b' x_i)$$

$$= 1 - F(-b' x_i)$$

où  $F$  est la fonction de distribution cumulative de  $u$ . La fonction de vraisemblance sera donc :

$$L = \prod_{y_i=0} F(-b' x_i) \prod_{y_i=1} [1 - F(-b' x_i)]$$

Dans le modèle Probit,  $u_i$  suit une loi normale  $N(0, \sigma^2)$ . Dans ce cas :

$$F(-b' x_i) = \int_{-\infty}^{-\frac{b' x_i}{\sigma}} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{t^2}{2}\right) dt$$

Les résultats de la régression Probit sont reportés au tableau 4. Dans ce tableau, seuls les indicateurs significatifs sont présentés. Ce sont ces indicateurs qui seront utilisés pour calculer les probabilités de fraude des dossiers. La première et la seconde colonnes donnent respectivement les numéros et les noms des indicateurs retenus. La troisième colonne présente les coefficients estimés, alors que la dernière indique le niveau de signification de ces coefficients. Il est à noter que les indicateurs suivis du chiffre 1 sont ceux qui ont été jugés comme étant les plus importants par l'enquêteur. Par contre, les indicateurs suivis du chiffre 2 représentent ceux dont le classement par ordre d'importance se situe entre 2 et 12.

**TABLEAU 4**  
**RÉSULTATS DE L'ESTIMATION DU MODÈLE PROBIT**

No	Indicateurs	Coefficients	t
	<b>constante</b>	<b>2,53842</b>	<b>23,87243</b>
2	MINEUR	0,65360	2,88309
*	SANCOL	0,44192	3,14714
4	VOLE	1,26659	4,60894
11	DIFFIC2	0,83390	2,60732
14	PETIT	1,13099	2,78572
16	JARGON2	0,90678	3,19590
18	ENTHOU	1,58303	4,75648
19	PROCHE	0,63276	3,23234
21	TAXI	0,72584	1,89837
22	PREUVE	1,64896	4,79668
26	DOCUM	1,24422	3,68028
30	TEMOIG2	1,03962	3,69397
32	SEUL	0,69536	4,19627
39	COMPTAN	0,81779	2,83649
45	AGRESSI	0,93629	3,10616
47	NERVEU	1,04607	4,49000
**	SANOPOL	0,57129	3,95309
3	INCOH1	1,80978	7,86276
11	DIFFIC1	1,89913	2,34950
16	JARGON1	1,18581	3,03581
30	TEMOIG1	0,95626	4,24095

\* Sinistre sans collision.

\*\* Sinistre avec collision et dont la réclamation n'a pas été accompagnée par un rapport de police, alors que celui-ci était nécessaire.

## ■ MODÈLE OÙ SEULES LES PRÉVISIONS DES ENQUÊTEURS SONT PERTINENTES

Dans un premier temps, nous comparerons les prédictions de notre modèle aux résultats de l'enquête, ensuite nous proposerons des exemples de prise de décision quant aux poursuites éventuelles d'enquêtes.

### □ Comparaison des résultats de régression à ceux de l'enquête

On compare les prédictions de notre modèle aux décisions des enquêteurs quant à leurs suspicions, lesquelles suspicions sont considérées comme étant complètes et qu'il n'y a pas d'autre fraude (ou suspicions) dans les échantillons considérés.

Le modèle que nous avons utilisé a généré des probabilités de fraude pour chaque dossier. Celles-ci varient entre 0,69 % (aucun indicateur) et 99,30 %. Pour l'instant, nous n'avons pas utilisé d'intervalle de confiance.

En choisissant le seuil de probabilité à partir duquel on décide d'examiner les dossiers de plus près, on détermine implicitement le taux de détection de la fraude, ainsi que le taux de précision des cas supposés frauduleux par le modèle.

Nous savons que parmi les 2 068 dossiers de réclamations examinés, les enquêteurs ont trouvé 131 cas soupçonnés ou établis frauduleux. Ceci nous donne un taux de fraude de 6,33 %<sup>2</sup>.

Prenons par exemple le seuil de probabilité de 10 %. À un niveau de 10 %, le modèle génère 336 cas de fraude. Parmi ces 336 cas, les enquêteurs ont soupçonné (et/ou établi) 98 cas de fraude. Le taux de classification correct de fraude est donc égal à 29,17 % (ou 98/336). Par ailleurs, les cas classés «non fraude» par le modèle sont au nombre de 1 732 (ou 2068-336). Parmi ceux-ci, 1 699 cas sont classés «non frauduleux» par les enquêteurs, un taux de classification correct de 98,09 % (ou 1699/1732). Afin de connaître la proportion de l'échantillon à sélectionner, on se base sur le taux de fraude de l'industrie<sup>3</sup>. La compagnie va sélectionner un pourcentage «X» de dossiers à revoir, qui va être calculé en résolvant l'équation suivante :

$$\alpha \cdot X \% + (1 - \beta) (1 - X \%) = 6,33 \%$$

où :

$\alpha$  : est le taux (en %) de précision des cas de fraude obtenu du modèle. Dans le cas où le seuil est égal à 10 %, ce taux est égal à 29,17 %.

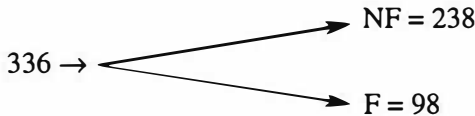
$\beta$  : est le taux (en %) de précision des cas non frauduleux selon le modèle. Dans le cas où le seuil de probabilité est de 10 %, ce taux est égal à 98,09 % .

Notons que l'équation ci-dessus comporte deux parties : la première partie, notée  $\alpha \cdot X$  %, exprime le pourcentage de fraude (tel que prédit par le modèle) parmi les cas sélectionnés (cas où la probabilité de fraude est supérieure au seuil). La deuxième partie, notée  $(1 - \beta) (1 - X)$  %, exprime, quant à elle, la proportion de fraude détectée par les enquêteurs, mais qui est en bas du seuil choisi (cas non sélectionnés pour révision).

En résolvant par exemple l'équation ci-dessus on obtiendra, pour un seuil de 10 %, un taux d'échantillon à réexaminer,  $X$ , égal à 16,21 %.

Récapitulons :

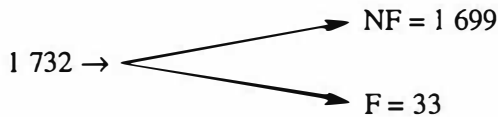
Pour un niveau seuil de probabilité de fraude de 10 %, on devrait sélectionner 336 cas (selon le modèle de régression); parmi ceux-ci, les enquêteurs ont trouvé que 98 cas étaient frauduleux et 238 non frauduleux, c'est-à-dire :



Taux de l'échantillon à réexaminer =  $X = 16,21$  %.

Taux de précision des cas de fraudes (F) :  $98/336 = 29,17$  %.

Le nombre de dossiers en bas de ce seuil de 10 % est égal (selon le modèle) à 1 732. Parmi ceux-ci, 1 699 cas sont classés non frauduleux par les enquêteurs :



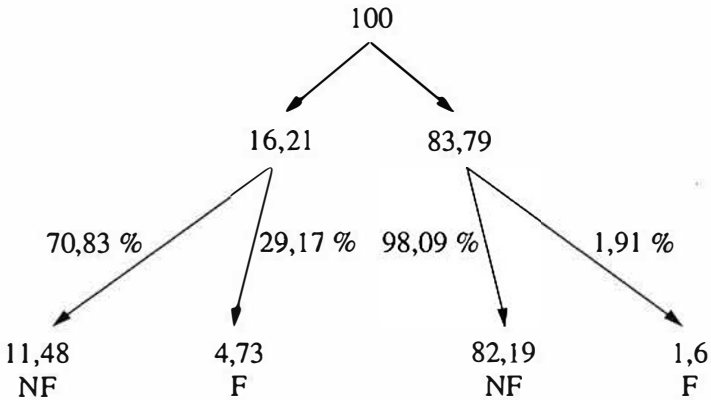
Taux de précision des «non fraude» (NF) =  $1\ 699/1\ 732 = 98,09$  %.

#### **Décision d'enquête**

Nous proposons maintenant un modèle de décision d'enquête.

À un seuil de probabilité  $P=10$  %, l'entreprise devrait sélectionner 16,21 % de ses réclamations afin de les réexaminer, si elle considère qu'elle possède un taux de fraude équivalent à celui de l'industrie et égal à 6,33 %.

Ainsi sur 100 réclamations, on aura les résultats suivants :

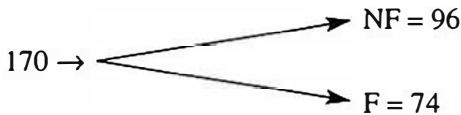


Taux de fraude total =  $4,73 \% + 1,6 \% = 6,33 \% ^4$

Taux de précision des cas de fraude (F) =  $29,17 \%$   
(voir le calcul ci-dessus)

Taux de détection =  $4,73\% / 6,33 \% = 74,72 \%$ .

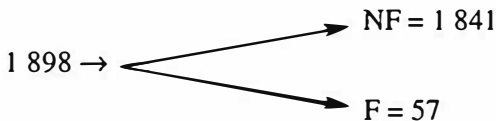
De la même manière et pour un seuil de probabilité  $P = 20\%$ , le modèle détecte 170 cas de fraude. De ceux-ci, 74 ont été classés «frauduleux» par les enquêteurs. Le taux d'exactitude du modèle est donc de  $43,53 \%$  (ou  $74/170$ ). Par ailleurs, en bas de  $20 \%$ , le modèle génère 1 898 cas de «non fraude»; les enquêteurs ont déterminé que sur ces 1 898 cas, il y avait 1 841 cas non frauduleux. Le taux de précision des «non fraude» est donc de  $97 \%$  (ou  $1\ 841/1\ 898$ ). Si on récapitule, on aura, à un seuil de probabilité de fraude de  $20 \%$  :



Taux de l'échantillon à réexaminer =  $X = 8,22 \%$ .

Taux de précision des cas de fraudes (F) :  $74/170 = 43,53 \%$ .

Le nombre de dossiers en bas de ce seuil de  $20 \%$  est égal (selon le modèle) à 1 898. Parmi ceux-ci, 1 841 cas sont classés non frauduleux par les enquêteurs :

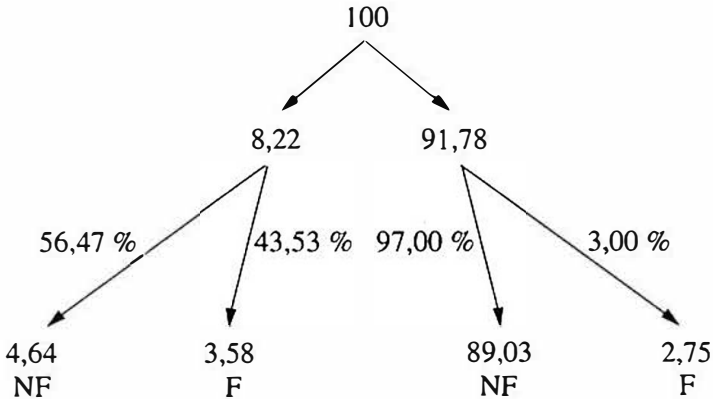


Taux de précision des «non fraude» (NF) =  $1\ 841/1\ 898 = 97 \%$ .



À un seuil de probabilité  $P = 20 \%$ , l'entreprise devrait sélectionner  $8,22 \%$  de ses réclamations afin de les réexaminer, si elle considère qu'elle possède un taux de fraude équivalent à celui de l'industrie et égal à  $6,33 \%$ .

Alors sur 100 réclamations, on aura les résultats suivants :



Taux de fraude total =  $3,58 \% + 2,75 \% = 6,33 \%^5$

Taux de précision des cas de fraude (F) =  $43,53 \%$

(voir le calcul ci-dessus)

Taux de détection =  $3,58 \% / 6,33 \% = 56,56 \%$ .

De la même manière, on calcule que pour un seuil de probabilité de  $P = 90 \%$ , on trouve les résultats suivants :

Taux d'échantillonnage =  $X = 0,38 \%$

Taux de précision =  $75 \%$

Taux de détection de la fraude =  $4,58 \%$ .

Si on répète ces mêmes calculs pour différents niveaux de seuil (de  $P$ ), on obtient le tableau récapitulatif suivant.

**TABLEAU 5**  
**TAUX DE PRÉCISION ET DE DÉTECTION EN FONCTION**  
**DU NIVEAU SEUIL DE LA PROBABILITÉ DE FRAUDE**

Niveau seuil de probabilité de fraude	Pourcentage d'échantillon (%)	Taux de précision en %	Taux de détection en %
P > 10 %	16,21	29,17	74,72
P > 15 %	10,66	35,75	60,35
P > 20 %	8,22	43,53	56,56
P > 25 %	6,90	48,95	53,40
P > 30 %	5,22	55,56	45,81
P > 35 %	4,59	58,95	42,81
P > 50 %	2,26	61,70	21,96
P > 65 %	1,20	64,00	12,16
P > 85 %	0,52	72,73	6,00
P > 90 %	0,38	75,00	4,58
Moyenne de l'échantillonnage <b>P &gt; 25,5 %</b>	<b>6,85</b>	<b>49,30</b>	<b>53,40</b>
Colonne S	Colonne E	Colonne R	Colonne D

Le tableau 5 nous dicte les conclusions suivantes.

Si on ne s'intéresse qu'à la détection, on va échantillonner un grand nombre de dossiers. Noter qu'un échantillon où  $P > 10\%$  contient plus de dossiers qu'un échantillon où  $P > 20\%$ . Cette méthodologie possède l'avantage d'aller chercher un maximum de cas suspects. Elle renferme cependant deux inconvénients : le premier est qu'un large échantillonnage de révision de dossiers coûte cher à l'entreprise. Une entreprise moyenne qui décide de réexaminer tout dossier dont la probabilité de fraude est supérieure à  $10\%$  va en fait devoir engager des fonds pour revoir 2468 dossiers<sup>6</sup> (sinistres automobiles excluant les bris de vitre). Le deuxième inconvénient résulte du fait qu'on réexamine un large éventail de dossiers; ce processus va engendrer certaines «injustices» envers les assurés qui ne sont pas des fraudeurs, mais qui se voient l'objet d'enquête de façon poussée par leur compagnie d'assurance. À titre d'exemple, si on se base sur le seuil de  $10\%$ , on se rend compte qu'on n'est précis qu'à  $29,17\%$  (voir tableau 5) et que, donc, plus des deux tiers de l'échantillon retenu ne représente pas des fraudeurs (selon les réponses des enquêteurs de notre étude).

Si, par contre, on ne s'intéresse qu'au taux de précision, on va échantillonner un très petit nombre de dossiers afin de les réétudier. L'avantage d'appliquer une telle méthode réside dans les coûts totaux de révision relativement bas, car l'échantillon est assez petit. En effet, une entreprise moyenne qui décide de réexaminer tout dossier dont la probabilité de fraude est supérieure à 90 % va en fait devoir engager des fonds pour revoir seulement 58 dossiers<sup>7</sup> (sinistres automobiles excluant les bris de vitre). L'autre avantage se trouve au niveau de la précision : alors que, dans le cas où le seuil était de 10 %, la précision se trouvait égale à 29,17 %, celle-ci équivaut à 75 % si on déplace notre seuil à 90 % : dans ce cas, on détecte les fraudeurs trois fois sur quatre. L'inconvénient d'une telle décision de seuil (à 90 %) réside dans le fait qu'on ne détecte réellement que 4,58 % de la fraude; la majorité des fraudeurs (c'est-à-dire 95,42 % des cas) va échapper au contrôle de la compagnie.

Ayant examiné ces cas extrêmes ( $P = 10\%$  et  $P = 90\%$ ), on peut voir qu'il y a une relation d'arbitrage entre la détection et la précision (voir tableau) : plus le seuil de la probabilité de fraude est grand, plus la précision est élevée et plus la détection est faible. Une entreprise très conservatrice qui ne veut pas trop s'impliquer dans la lutte à la fraude va probablement opter pour des seuils assez élevés. Par ailleurs, une entreprise agressive qui veut s'impliquer dans la lutte à la fraude et qui veut réduire ses coûts va opter pour des seuils assez bas. Notons enfin qu'un seuil autour de 25 % (seuil moyen de l'industrie) va nous permettre de détecter à peu près la moitié de la fraude et d'être précis en même temps à environ 50 %.

Ce tableau peut aussi servir de guide aux entreprises désireuses de connaître leurs taux de précision et de détection de la fraude. Si une compagnie connaît sa contrainte budgétaire et décide d'allouer un montant  $Z$  aux enquêtes approfondies, alors elle connaîtra le nombre de dossiers à sélectionner afin de conduire ces enquêtes. À ce taux d'échantillonnage correspondra un taux de détection et de précision de l'enquête (si elle considère que son taux de fraude est le même que celui de l'industrie).

#### **Cas où les prévisions de notre modèle sont correctes**

Selon cette approche, nous nous basons sur notre modèle de prévision de la fraude pour comptabiliser celle-ci. À un seuil donné, on considère tous les cas dont la probabilité dépasse ce seuil comme étant frauduleux. Les prévisions des enquêteurs interviennent cependant dans la proportion des cas en deçà de la probabilité-seuil choisie. Ce raisonnement peut avoir pour motivation le fait que

notre modèle (par le biais de ses indicateurs) n'arrive pas à cerner tous les aspects de la fraude; nous pensons ici aux aspects essentiellement subjectifs lors de l'enquête routinière.

Il est possible également que les enquêteurs n'observent pas tous les éléments qui permettent de détecter la fraude. Par conséquent, il se peut que le nombre total de fraude se trouve augmenté par rapport au cas où seules les prévisions des enquêteurs étaient correctes<sup>8</sup>. Toute la portion des fraudes détectées par le modèle mais non détectées par les enquêteurs est rajoutée. Rappelons que dans le cas précédent, la fraude s'élevait à 6,33 % (équivalant à un niveau seuil de  $P = 25,5 \%$ ). Dans ce cas, où tous les dossiers dont la probabilité de fraude est supérieure à ce seuil sont considérés frauduleux, la fraude s'élèverait à 9,82 %; le différentiel de taux (i.e 3,49 %) pourrait être dû à de la fraude cachée qui n'a pas pu être détectée par les enquêteurs. Nous avons en fait trois cas de figure qui se présentent à nous :

1. Une partie de la fraude est détectée simultanément par notre modèle et par les enquêteurs. Ce sont les dossiers dont la probabilité est supérieure au seuil et qui ont été détectés par les enquêteurs.
2. Une autre partie est détectée par le modèle mais non détectée par les enquêteurs. Ce sont les dossiers dont la probabilité est supérieure au seuil mais qui n'ont pas été détectés par les enquêteurs. C'est ce que nous avons appelé de la fraude cachée.
3. Enfin, une troisième partie est constituée de dossiers détectés par les enquêteurs mais non détectés par le modèle. Nous avons évoqué ci-dessus les aspects subjectifs pour motiver ces dossiers frauduleux. Ces dossiers ont une probabilité de fraude inférieure au seuil. Ces trois cas (1, 2 et 3) sont reproduits à la figure suivante :

Seuil	
Cas de non fraude (modèle)	Cas de fraude (modèle) : (1 + 2)
<ul style="list-style-type: none"> <li>• Certains sont considérés frauduleux par les enquêteurs (3)</li> </ul>	<ul style="list-style-type: none"> <li>• Certains sont considérés comme non frauduleux par les enquêteurs</li> </ul>

## ■ CONCLUSION

Le produit qui a été développé est basé sur des dossiers de réclamations qui ont été tirés de façon aléatoire de la population des dossiers des assureurs participants. Il traite donc de tous les cas possibles de fraude d'une réclamation quelconque. Le résultat généré consiste en une probabilité de fraude d'un dossier ordinaire.

En plus de s'intéresser à une probabilité de fraude d'un dossier ordinaire, les dirigeants des compagnies d'assurance pourraient s'intéresser à la probabilité de réussite de l'enquête approfondie. En effet, ce sont là deux étapes complémentaires qu'il faudrait mettre au point de façon simultanée.

Dans un premier temps, on cherche à savoir si c'est nécessaire d'enquêter sur un dossier ordinaire : c'est le but du travail actuel où, à l'aide d'un modèle, nous avons généré une probabilité de fraude. Notre échantillon dans ce travail était constitué de 0,64 % de tous les dossiers de réclamations automobiles (excluant les bris de vitres) de l'industrie.

Dans un deuxième temps, il est nécessaire de se consacrer à l'étude de cas dits «sérieux» et qui ont fait l'objet d'enquêtes particulières. Dans ces cas, il faudrait échantillonner auprès des cas qui ont fait l'objet d'enquête poussée. L'idéal serait d'échantillonner une partie des dossiers qui sont du ressort des Unités Spéciales d'Enquête (ou de toute autre division chargée de mener des enquêtes spécialisées). En plus des indicateurs propres à la première étape de l'enquête, on pourrait rajouter des indicateurs spécifiques à la deuxième étape (enquête approfondie). Cette étude prendrait en considération des indicateurs propres, la formation des enquêteurs spécialisés, ainsi que leur expérience. Le but de cette étape serait de générer une probabilité de réussite de l'enquête approfondie. À la suite de celle-ci, on pourrait calculer la rentabilité de l'enquête en termes monétaires. C'est, selon nous, cette deuxième étape qu'il faut compléter afin que le processus d'enquête automatisée soit achevé.

## ■ ANNEXE : LES INDICATEURS DE FRAUDE

### □ Indicateurs

Les indicateurs ci-dessous sont reproduits exactement tels qu'ils ont été présentés aux enquêteurs. Notons toutefois qu'une liste détaillée d'explication de ces indicateurs leur a été fournie en même temps que cette liste. Ceux en caractères foncés ont été désignés significatifs pour expliquer la probabilité de fraude d'un dossier.

#### ■ Accident / dommage

- 1) Pas de rapport de police quand celui-ci est nécessaire (ou bien, rapport demandé au poste de police, alors que l'accident a eu lieu dans des endroits où la police a l'habitude d'agir rapidement).
- 2) Une collision mineure a entraîné des coûts de réparation excessifs.
- 3) Existence de tout dommage non relié à la perte ou incohérent avec les faits de l'accident.
- 4) Le véhicule est rapporté volé et retrouvé peu de temps après avec de lourds dommages.

#### ■ Véhicule

- 5) Véhicule de modèle récent et cher.
- 6) Les factures de maintenance du véhicule ne peuvent être fournies par l'assuré.
- 7) Le véhicule n'a pas d'intérêt pour le voleur (i.e. vieille voiture ordinaire).
- 8) Véhicule volé dans un centre commercial.

#### ■ Agent / assureur

- 9) Peu de temps avant la perte, l'assuré a vérifié auprès de son agent l'étendue de sa couverture d'assurance.
- 10) L'agent d'assurance n'a jamais vu le véhicule assuré.

#### ■ Financiers

- 11) L'assuré éprouve des difficultés financières personnelles ou reliées à ses affaires.
- 12) L'occupation de l'assuré ne justifie pas la valeur élevée de son véhicule (et ses contenus).

### ■ **Règlement**

- 13) L'assuré (le réclamant) est trop enthousiaste à l'idée de recevoir un paiement monétaire au lieu de faire effectuer des réparations sur son véhicule.
- 14) L'assuré est prêt à accepter un montant de règlement relativement petit plutôt que d'avancer tous les documents reliés à la perte.

### ■ **Réclamant / assuré**

- 15) L'assuré est très insistant pour un règlement rapide.
- 16) L'assuré est extraordinairement familier avec le jargon des assurances ou des réparations de véhicules.
- 17) L'assuré propose de se déplacer au bureau des réclamations pour paiement.
- 18) L'assuré (ou le réclamant) est trop enthousiaste ou trop franc pour prendre le blâme de l'accident.
- 19) L'accident (ou la perte) a eu lieu peu de temps après l'enregistrement et l'achat d'assurance pour le véhicule, ou dans le mois qui précède la fin de la police (ou de la couverture).
- 20) S'appuyant sur la loi 68, l'assuré refuse de donner son consentement (pour une enquête plus approfondie).
- 21) Nombreux reçus de taxi, ou bien factures de location de véhicule provenant d'un magasin de débosselage.
- 22) Factures ou preuves semblent fausses ou fabriquées.
- 23) Antécédents de l'assuré : il a déjà été reconnu coupable de fraude, ou a déjà commis des méfaits qui laissent supposer que c'est un fraudeur potentiel.

### ■ **Titre et propriété**

- 24) L'historique de la propriété du véhicule ne peut être établie.

### **AUTRES INDICATEURS**

#### ■ **Accident / dommage**

- 25) La réclamation (et/ou la perte) est rapportée par une tierce personne.
- 26) La documentation de l'estimation et de la réparation n'est pas disponible.
- 27) Tous les véhicules endommagés sont acheminés vers le même garagiste.

- 28) Le représentant des réclamations (évaluateur) se voit refusé l'examen du véhicule.
- 29) Le véhicule a été réparé avant son examen par le représentant des réclamations (évaluateur).
- 30) Témoignages contradictoires concernant les circonstances de la perte.
- 31) L'assuré nie les versions des témoins de l'accident.
- 32) Accident impliquant un seul véhicule.
- 33) Accident impliquant une tierce personne non identifiée.

#### ■ Véhicule

- 34) Véhicule ayant connu des problèmes mécaniques récurrents; la garantie du manufacturier a expiré.
- 35) Véhicule loué avec un kilométrage élevé.
- 36) Verrouillage du contact intact après recouvrement du véhicule.
- 37) Pas de signes d'entrée par effraction (serrure de la porte intacte...).

#### ■ Financiers

- 38) Remboursements du prêt d'achat du véhicule en retard.
- 39) L'achat du véhicule s'est fait au comptant.
- 40) L'assuré est au chômage; travaille dans une industrie en dépression; vit dans une région défavorisée.

#### ■ Réclamant / assuré

- 41) Véhicule retrouvé par l'assuré.
- 42) Plusieurs couvertures; plusieurs polices; perte surassurée.
- 43) Prime payée en personne et au comptant.
- 44) Problèmes avec l'adresse : C.P.; motel; fausse adresse; assuré absent; vit chez des amis...etc.
- 45) Réclamant est très agressif (menace de faire appel à un avocat, au gouvernement, etc.)
- 46) Réclamant refuse de répondre aux questions qui ont trait au sinistre.
- 47) Lors de l'enquête, l'assuré est nerveux et semble confus.
- 48) Beaucoup de réclamations dans le passé.

#### ■ Titre et propriété

- 49) Titre récemment transféré d'une autre province (ou d'un autre état).
- 50) Le titre de la propriété est encore au nom du propriétaire précédent.



## □ Bibliographie

- AUTOMOBILE INSURANCE FRAUD STUDY (1991), Florida Insurance Research Center, University of Florida.
- BELHADJI, El Bachir, Georges DIONNE, et Faouzi THARKANI (1998), «A Model for the Detection of Insurance Fraud», Mimeo, HEC-Montréal.
- CARON, Louis et Georges DIONNE (1997), «Insurance Fraud Estimation: More Evidence from Quebec Automobile Insurance Industry», *Assurances*, 64, janvier, p. 567-579.
- DERRIG, R.A. et H.I. WEISBERG (1993), «Fraud and Automobile Insurance», Insurance Bureau of Massachusetts.
- DIONNE, Georges et El-Bachir BELHADJI (1996), «Évaluation de la fraude à l'assurance automobile au Québec», *Assurances*, 64, octobre, p. 365-395.
- INSPECTEUR GÉNÉRAL DES INSTITUTIONS FINANCIÈRES (1994), «Rapport annuel sur les assurances».
- JOHNSTON, Daniel J. (1997), «Combating Fraud: Handcuffing Fraud Impact Benefits», *Assurances*, 65, July, p. 175-184.
- PROMOTERS OF CREATIVE CLAIMS SOLUTIONS (1993), «Fraud Indicators», Nationwide Insurance. Property and Casualty Claims Services, mars.

## □ Notes

1. Cette partie de l'étude est achevée et une copie du rapport est disponible auprès du Bureau d'assurance du Canada (Dionne et Belhadji, 1996). Ses grandes lignes sont également publiées dans le numéro d'octobre 1996 de la revue *Assurances*.

2. Ici il est supposé que les enquêteurs détectent tous les cas de fraude dans les dossiers. Cette hypothèse a été remise en cause par Caron et Dionne (1997) qui montrent que les enquêteurs observent seulement le tiers de la fraude.

3. Ce pourcentage peut être remplacé par un autre si la compagnie considère que son taux de fraude est différent de celui de la moyenne de l'industrie.

4. Noter que ce chiffre peut ne pas être exactement égal à 6,33 % à cause des arrondissements des pourcentages.

5. Noter que ce chiffre peut ne pas être exactement égal à 6,33 % à cause des arrondissements des pourcentages.

6. Ce chiffre est équivalent à 16,21 % de 15 224 dossiers de réclamations que traite annuellement une compagnie moyenne dans le secteur de l'assurance automobile. Ces chiffres proviennent de l'étude de Dionne et Belhadji (1996).

7. Ce chiffre est équivalent à 0,38 % de 15 224 dossiers de réclamations que traite annuellement une entreprise moyenne dans le secteur de l'assurance automobile.

8. Voir Caron et Dionne (1997) pour plus de détails. Le meilleur estimateur de leur modèle génère un taux de fraude de 10 % ou 113,5 millions de dollars dans l'industrie.