

**Développement d'un système expert
de détection automatique de la
fraude à l'assurance automobile**

**par El Bachir Belhadji
et Georges Dionne**

Cahier de recherche 97-04
Juin 1997

ISSN : 1206-3290

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.

Cette étude a été réalisée avec la collaboration de M. Éric Le Saux qui a été responsable de la partie programmation du logiciel développé.

Développement d'un système expert de détection automatique de la fraude à l'assurance automobile

**El Bachir Belhadji
et Georges Dionne**

El-Bachir Belhadji est assistant de recherche à la Chaire de gestion des risques de l'École des Hautes Études Commerciales.

Georges Dionne est titulaire de la Chaire de gestion des risques et professeur de finance à l'École des Hautes Études Commerciales.

Développement d'un système expert de détection automatique de la fraude à l'assurance automobile

par El Bachir Belhadji et Georges Dionne

Résumé

Le but de cette étude était 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 18 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 en annexe.

Summary

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 18 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 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 any one 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 given in the appendix.

Introduction

Le but de cette recherche était 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 cinq 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. La section IV discute sommairement des coûts des enquêtes et la section V présente des avenues de recherche qui pourraient intéresser les assureurs.

Une disquette est disponible auprès des auteurs. Elle contient le logiciel FRAUDE1.0 et peut être utilisée sur des micro-ordinateurs de type «pc» équipés de Windows 95/NT et d'un navigateur Web.

I. 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.

I.1 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.

I.2 Déroulement de l'enquête

Les dossiers ont été choisis au hasard parmi tous les dossiers fermés durant la période du 1er 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 .

Tableau 1
Nombre de questionnaires reçus des différentes compagnies

No Cie	Nombre de questionnaires complétés	% échantillon
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
Total	2 509	100

I.3 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.

I.4 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 correspondants 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. Pour le moment, nous avons retenu cette méthode, mais d'autres avenues de correction sont à l'étude. 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

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

II. 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 automobile¹ 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.

II.1 Le 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.

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.

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.

II.2 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 (3ème 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 $Z_{\alpha/2}$ 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 14 dossiers et moins. C'est ainsi qu'il ne faudrait pas interpréter les indicateurs possédant un astérisque (5^e 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.

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.

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*
43	Prim	1	1	1*
44	Motel	1	0	0*
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.

II.3 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 :

$$\begin{aligned} y &= 1 && \text{si } y_i^* > 0 \\ y &= 0 && \text{autrement.} \end{aligned}$$

Dans notre cas particulier :

$$\begin{aligned} y &= 1 && \text{si le dossier a été jugé frauduleux (suspçon ou fraude établie)} \\ y &= 0 && \text{autrement.} \end{aligned}$$

Il s'ensuit que :

$$\begin{aligned} \text{prob}(y_i = 1) &= \text{prob}(u_i > -b' x_i) \\ &= 1 - F(-b' x_i) \end{aligned}$$

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

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

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

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

Les résultats de la régression Probit sont reportés au tableau 4 de la page suivante. 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
	constante	2,53842	23,87243
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.

III. 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.

III.1 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 %².

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

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.

de l'échantillon à sélectionner, on se base sur le taux de fraude de l'industrie³. 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 :

α : 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 %.

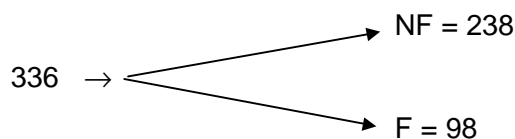
β : est le taux (en %) de précision des cas non frauduleux selon le modèle. Dans le cas où le seuil de probabilité est de 10 %, ce taux est égal à 98,09 % .

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 :

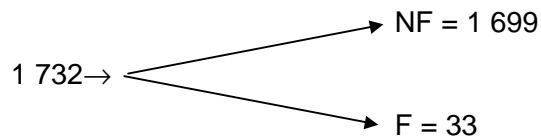


Taux de l'échantillon à réexaminer = X = 16,21 %.

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

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.

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 :



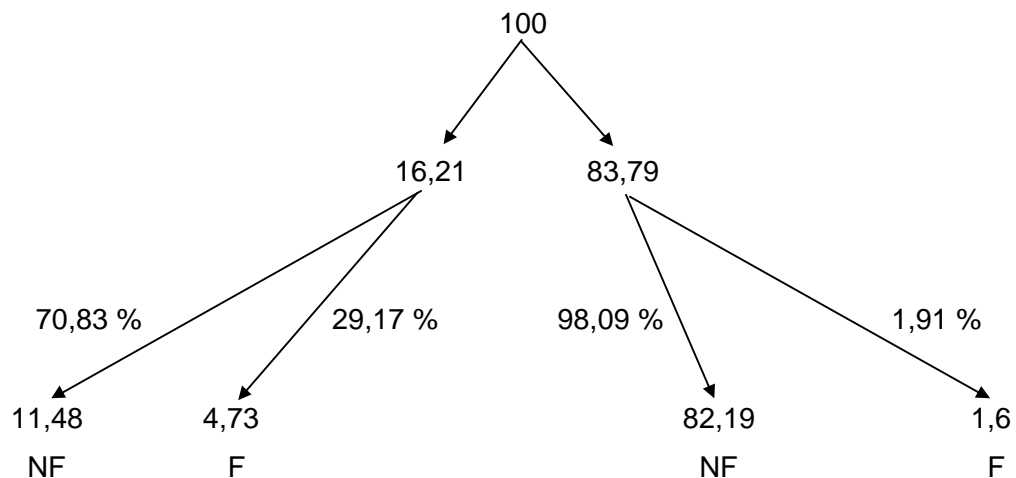
Taux de précision des «non fraude» (NF) = $1\,699/1\,732 = 98,09\%$.

III.2 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 :



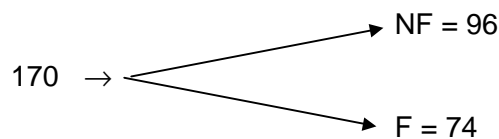
Taux de fraude total = $4,73\% + 1,6\% = 6,33\%$ ⁴

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

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

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% :

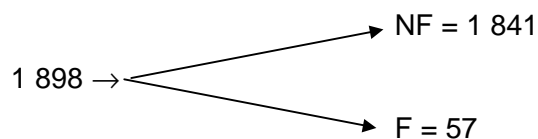


Taux de l'échantillon à réexaminer = $X = 8,22\%$.

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.

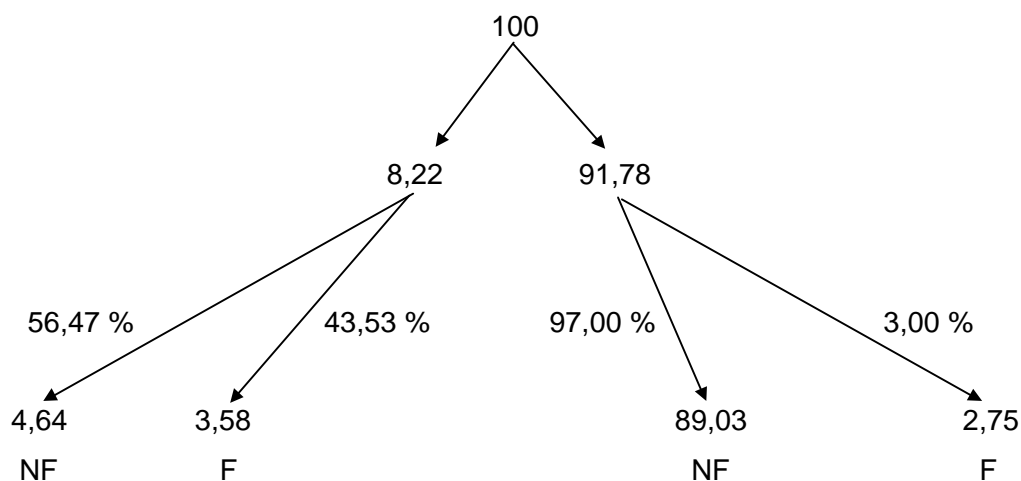
Parmi ceux-ci, 1 841 cas sont classés non frauduleux par les enquêteurs :



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 de la page suivante.

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 %

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

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			
P > 25,5 %	6,85	49,30	53,40
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 dossiers⁶ (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).

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).

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 dossiers⁷ (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).

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.

III.3 Cas où les prévisions de notre modèle sont correctes

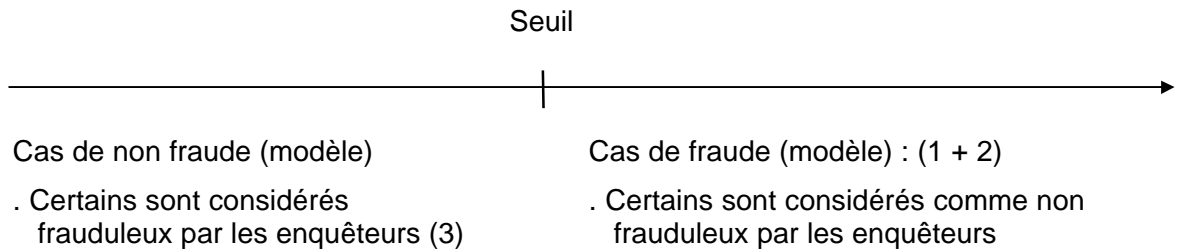
Selon cette approche, nous nous basons sur notre modèle de prévision de la fraude pour comptabiliser celle-ci. À un seuil donné, on considère tous les cas dont la probabilité dépasse ce seuil comme étant frauduleux. Les prévisions des enquêteurs interviennent cependant dans la proportion des cas en deçà 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 correctes⁸. 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

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.

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 :



IV. Quelques notions sur les coûts et les enquêtes approfondies

Lors de sa décision d'enquête poussée, la compagnie d'assurance fait face au choix suivant : si le coût de règlement sans enquête approfondie est inférieur à l'espérance de coût avec enquête, elle ne fera pas d'enquête; sinon, elle procédera à l'enquête⁹.

Soit les coûts suivants :

C_1 : coût de règlement sans enquête approfondie

C_2 : coût de règlement après enquête approfondie

C_c : coût de mener une enquête approfondie¹⁰

$C_1 = R_1$ (valeur de la réclamation nette de la franchise)

$C_2 = C_c + p \cdot R_2 + (1-p) \cdot R_1$

où :

R_2 : montant de la réclamation versée si l'enquête réussit

p : probabilité de réussite de l'enquête.

⁹ Cette déduction n'est valide que dans un raisonnement statique à court terme. Une entreprise qui base ses décisions sur des actions de plus long terme peut vouloir poursuivre l'enquête même si, en moyenne, ses coûts avec enquête sont supérieurs à ses coûts sans enquête. En agissant ainsi, cette entreprise va acquérir une réputation de compagnie active dans la lutte à la fraude et les fraudeurs vont se déplacer vers d'autres compagnies plus tolérantes (à la fraude). Ceci va résulter en des paiements de réclamations et des coûts d'enquête totaux (actualisés) plus faibles.

¹⁰ Ce montant peut être évalué approximativement par le coût moyen de mener une enquête approfondie. Il est cependant plus juste de considérer chaque coût comme étant spécifique à un dossier en particulier. Il existe des dossiers plus difficiles à traiter que d'autres.

Notons que la probabilité de réussite de l'enquête va généralement dépendre de plusieurs facteurs tels la probabilité de fraude (que nous calculons et notée P), l'expérience et la formation des enquêteurs (notées respectivement EX et F) et l'organisation de l'entreprise (notée USE). Dans l'organisation de l'entreprise, on peut trouver par exemple l'existence ou non d'Unités Spéciales d'Enquête. Cette probabilité de réussite peut donc être représentée par la fonction suivante :

$$p = f(P, EX, F, USE\dots).$$

Mentionnons toutefois que la probabilité de fraude est déterminante dans la fonction ci-dessus : un dossier qui possède une très faible probabilité de fraude (P) ne devrait pas générer une probabilité de succès élevée (p) et ce, quelles que soient les valeurs des autres facteurs (EX, F, USE\dots).

Enfin, la décision de l'entreprise va revêtir la forme suivante :

Si $C_1 \geq C_2 \Rightarrow$ On mène l'enquête.

Si $C_1 < C_2 \Rightarrow$ On ne mène pas d'enquête.

En remplaçant les valeurs de C_1 et C_2 dans la première inégalité ci-dessus, on trouve que l'entreprise va mener l'enquête dans le cas où :

$$p(R_1 - R_2) - C_c \geq 0.$$

En d'autres termes, on mène l'enquête si l'espérance du différentiel de réclamation couvre au moins le coût certain de l'enquête.

V. Avenues de recherche

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 (voir figure 1).

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 (voir figure 2). En effet, ce sont là deux étapes complémentaires qu'il faudrait mettre au point de façon simultanée : la figure 3 en résume la démarche.

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.

Figure 1
Détermination de la probabilité de fraude
(prise de décision au niveau des enquêteurs régisseurs)

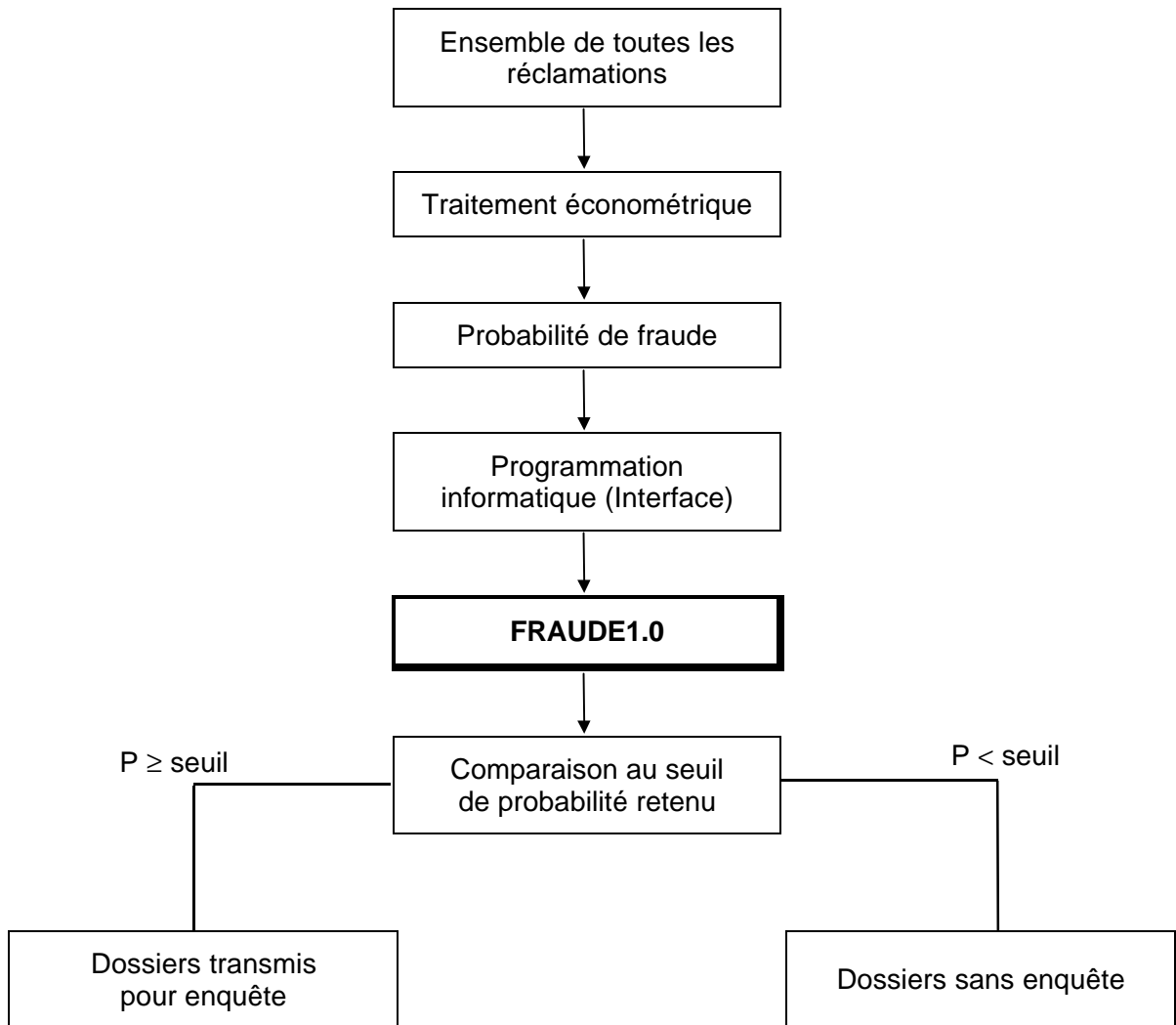


Figure 2
Détermination de la probabilité de réussite d'une enquête
(prise de décision au niveau des USE)

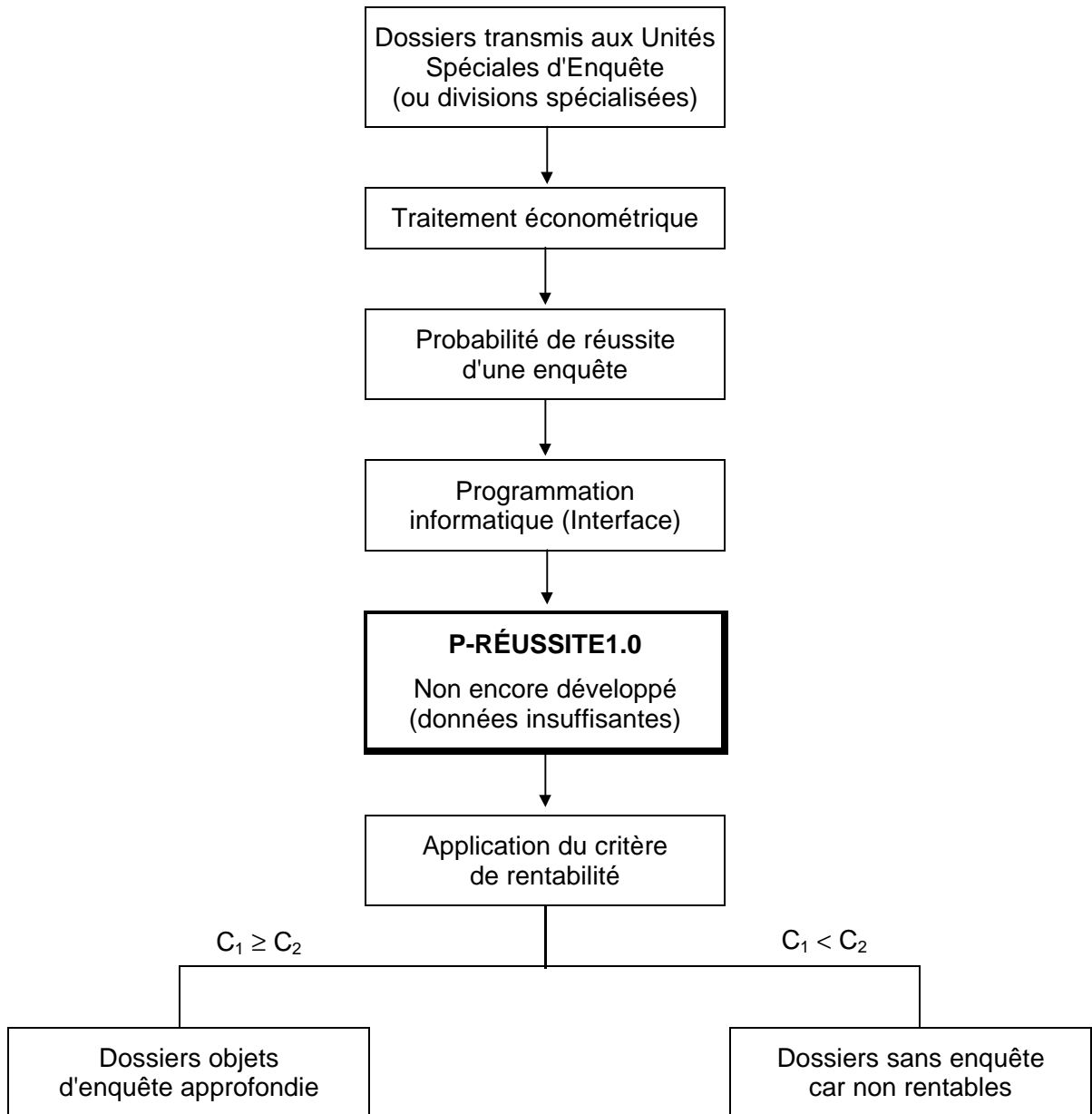
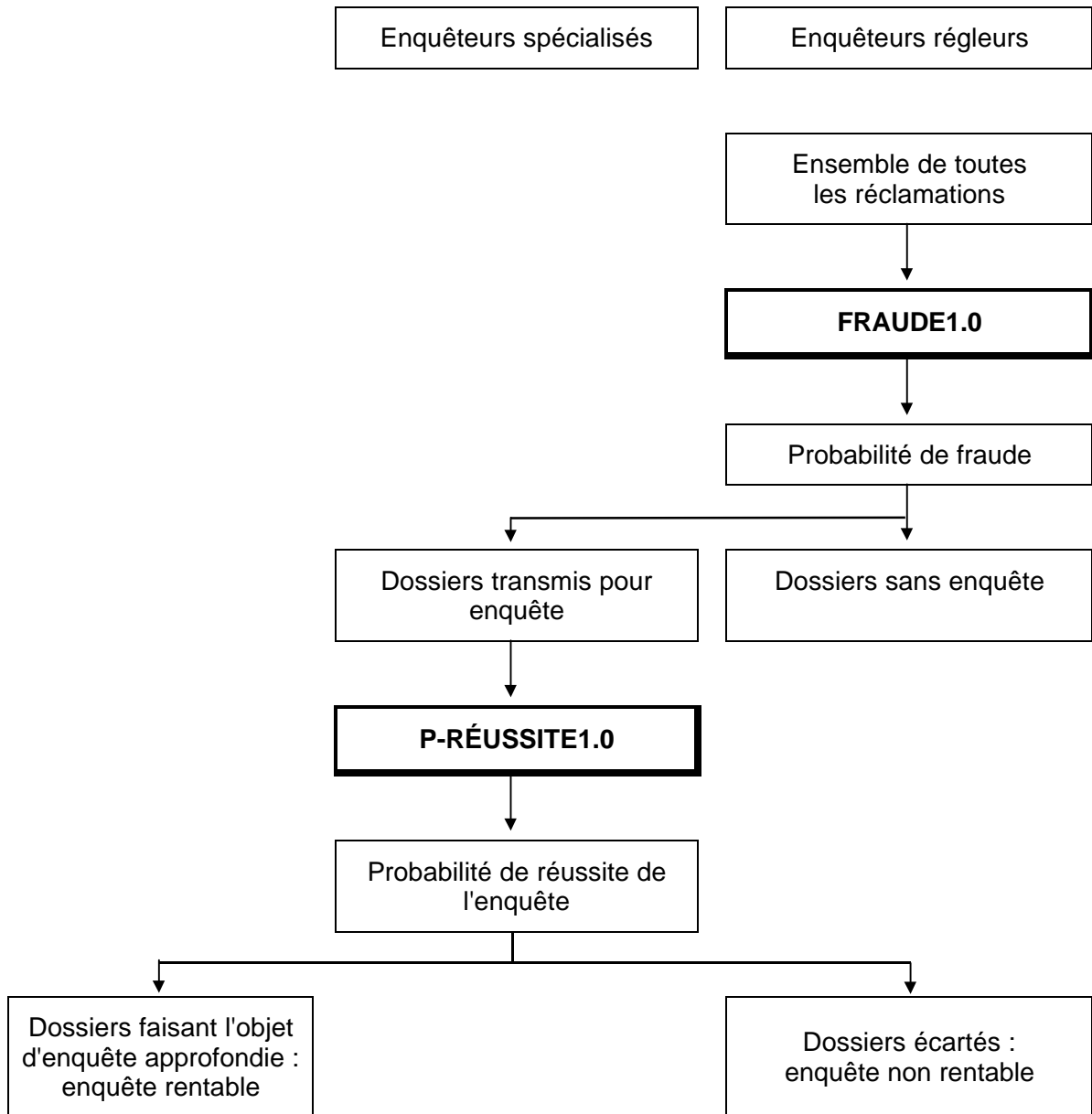


Figure 3
Probabilité de fraude et probabilité de réussite d'une enquête
(prise de décision à deux niveaux différents)



Bibliographie

Automobile Insurance Fraud Study (1991), Florida Insurance Research Center, University of Florida.

Caron, Louis et Georges Dionne (1997), «Insurance Fraud Estimation : More Evidence from Quebec Automobile Insurance Industry», *Assurances*, n° 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*, n° 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*, n° 65, July, p. 175-184.

Promoters of Creative Claims Solutions (1993), «Fraud Indicators», Nationwide Insurance. Property and Casualty Claims Services, mars.

ANNEXES

Annexe 1

Manuel d'utilisation du logiciel FRAUDE1.0

par

El Bachir Belhadji

Georges Dionne

Eric Le Saux

A1-1 Mise en garde

Ce logiciel a été développé avec des données de l'industrie, et donc ne concerne pas une compagnie en particulier mais plutôt l'ensemble des entreprises ayant participé à l'enquête. Si une compagnie d'assurance est intéressée à avoir le même produit pour ses besoins privés, il serait souhaitable qu'elle développe le même type de produit en prenant en considération ses spécificités particulières. Dans ce cas, cette compagnie utiliserait ses propres données car il n'est pas évident que les paramètres calculés pour l'industrie puissent s'appliquer directement à une entreprise.

A1-2 But du logiciel

Le but de ce logiciel est de fournir une aide à la décision en matière d'enquête de dossiers supposés frauduleux. Ce produit permet à la compagnie qui l'utilise d'accélérer le processus de décision en matière d'enquête et donc, de gagner du temps et des ressources. L'utilisation de ce produit nous permet de générer une probabilité de fraude pour chaque dossier. Sur la base de cette probabilité et de certaines informations propres à l'entreprise, ce logiciel nous dira s'il est rentable ou non de poursuivre l'enquête.

A1-3 Procédure d'installation

1. Systeme d'exploitation requis

Le formulaire d'assistance à la détection de la fraude a été écrit en Java, un nouveau langage informatique dont les programmes peuvent s'exécuter sur la plupart des systèmes d'exploitation existants (Windows 95 et NT, Unix, etc.). Pour utiliser ce programme, il est nécessaire d'utiliser un *Web Browser*, un de ces outils qui permettent de naviguer sur l'Internet. Il n'est cependant pas nécessaire d'être connecté à l'Internet pour que le programme fonctionne correctement.

Plusieurs *Web Browsers* sont actuellement disponibles via l'Internet, certains gratuitement. Nous vous recommandons de vous procurer **Netscape Navigator 3.01** ou **HotJava 1.0**.

Malheureusement, *Microsoft Internet Explorer 3* a encore quelques «bugs» au niveau de l'exécution des programmes Java, ce qui vous empêche de vous en servir.

2. Démarrage du programme

Le logiciel vous est fourni sur une disquette en format PC. Vous pouvez l'exécuter directement à partir de la disquette en demandant à votre *Browser* d'accéder à la référence suivante :

file:/a:/fraude.html

Il vous est également possible de copier tous les fichiers de la disquette sur votre disque dur. Si vous les installez dans le répertoire «C:\assurance», vous pourrez alors y accéder en utilisant la référence «file:/c:/assurance/fraude.html».

Si votre corporation possède déjà son **Intranet**, c'est-à-dire son propre site *Web* à usage interne à la compagnie, vous pourrez demander à vos analystes d'y installer le logiciel. Cela vous évitera d'avoir à l'installer sur tous les PC de vos enquêteurs. Ils pourront démarrer le logiciel en utilisant une adresse semblable à :
«http://www.bac.ca/enquete/fraude.html».

Il vous est également possible de rendre le logiciel accessible au monde entier, via l'Internet, en le plaçant sur un *Web* conventionnel.

3. Où se procurer un *Browser*?

Vous pourrez vous procurer une version démo complètement fonctionnelle de Netscape via FTP anonyme à l'adresse suivante «ftp://ftp6.netscape.com/navigator».

HotJava est disponible via FTP anonyme à «ftp://java.sun.com/pub/hotjava».

A1-4 Comment remplir le formulaire

Interface

L'interface est divisée en trois parties principales :

Interface identification :

A **Dossier**, veuillez entrer le numéro de la police du dossier de réclamation.

A **Matricule de l'enq**, veuillez inscrire votre numéro d'employé.

A **Lieu du sinistre**, veuillez entrer la municipalité où a eu lieu le sinistre.

A **Date du sinistre**, entrez, dans l'ordre, le jour, le mois et l'année du sinistre.

A **Garanties achetées**, inscrire toutes les couvertures que l'assuré a achetées.

Interface indicateurs :

Cette partie sert principalement à introduire tous les indicateurs qui sont présents dans le dossier que vous étudiez. Vous n'avez qu'à « cliquer » avec la souris sur un indicateur en particulier quand vous savez que cet indicateur est présent. Après avoir sélectionné tous les indicateurs pertinents, vous pouvez passer à la section suivante. N'oubliez pas d'indiquer le numéro de l'indicateur le plus important selon vous.

Remarquez que vous ne pourrez pas passer à l'étape suivante si vous n'avez pas complété celle-ci. Veuillez noter qu'au fur et à mesure que vous entrez les données, la probabilité de fraude s'affiche en bas de l'écran. L'ordre d'entrée des indicateurs n'est pas important.

Interface décision :

Cette partie permet au superviseur de décider de poursuivre ou non l'enquête de façon plus approfondie.

À **Montant estimé de remboursement s'il n'y a pas d'enquête**, veuillez reporter le montant du sinistre (estimé à cette date) en ayant au préalable déduit la franchise si celle-ci s'applique. Si vous ne possédez pas ce montant pour le moment, laissez l'espace libre. Vous reviendrez plus tard pour le compléter.

Montant estimé de remboursement si la preuve de fraude est établie suite à une enquête. Pour chaque dossier, on peut savoir (approximativement) quel sera le montant qu'on versera à l'assuré si l'enquête réussit. Dans certains cas, ce montant sera nul, alors que dans d'autres, la compagnie va verser un certain montant, même si elle considère que son enquête a réussi. D'après nos données sur la fraude à l'assurance automobile, ce montant était égal en moyenne à 1 671\$ (en 1995).

Coût estimé pour faire une enquête approfondie. Chaque dossier possède un niveau de difficulté en terme d'enquête qui lui est propre. En général, plus le niveau de difficulté est élevé, plus coûteuse sera l'enquête. Ce montant peut être déterminé plus ou moins adéquatement par un enquêteur professionnel qui possède de l'expérience. Dans le milieu des assurances, on évoque cependant un coût se situant entre 20 % et 30 % de la valeur du sinistre. En d'autres termes, pour une réclamation moyenne¹¹ de 2 514 \$, le coût de l'enquête approfondie se situerait entre 500 \$ et 750 \$.

Ce coût reste, bien sûr, une moyenne et ne peut s'appliquer qu'aux cas dits «moyens» en termes de difficulté d'enquête et de valeur de sinistre.

Probabilité subjective que l'enquête réussisse. L'estimation inscrite en bas des indicateurs représente la probabilité de fraude du dossier analysé. En général, plus cette probabilité est élevée, plus vos chances de réussir l'enquête le sera aussi. Votre «chance» de réussir l'enquête va dépendre de la probabilité (ci-dessus) que le dossier soit frauduleux, de votre expérience, de votre formation, ainsi que de l'existence ou non d'Unités Spéciales d'Enquête dans votre compagnie. C'est la

¹¹ Moyenne selon les données de l'enquête Dionne-Belhadji (1996).

connaissance de tous ces paramètres qui va vous permettre de donner votre probabilité subjective que l'enquête réussisse. Dans les cas où vous vous sentez incapable de donner votre probabilité subjective, veuillez inscrire la probabilité qui est reportée ci-dessus.

Résultat. Après avoir introduit toutes les informations pertinentes, le programme effectue un calcul de rentabilité et vous conseille de poursuivre ou non l'enquête, sur la base du seul critère de rentabilité.

Remarque : Étant donné que le programme se base uniquement sur des critères de rentabilité (de court terme), il se peut que dans certains cas (notamment les cas où le montant de réclamation est trop petit) il vous dicte de ne pas poursuivre l'enquête et cela même si la probabilité de fraude (et/ou de réussite de l'enquête) est élevée. Dans ces cas, le superviseur peut se baser sur la valeur de la probabilité de fraude pour décider de poursuivre ou non l'enquête. Pour plus d'information sur la valeur de cette probabilité, se reporter à la section intitulée *seuil de probabilité* dans les pages qui suivent.

A1-5 Exemple d'un cas particulier

À titre d'exemple sur la façon de compléter l'interface, nous allons reprendre un exemple d'un dossier particulier.

Interface description:

Dossier : le numéro de la police est le : *a1397*

Matricule de l'enq : le numéro d'employé de l'enquêteur est le : *0074*.

Lieu du sinistre : le sinistre a eu lieu à *Ste Thérèse*.

Date du sinistre : le sinistre est survenu le *04 octobre 1997*.

Montant estimé du sinistre net de franchise : Dans ce cas, le montant estimé du sinistre est de : *5 527 \$*. La franchise applicable dans le cas de ce sinistre est de *250 \$*. Le montant net de franchise à reporter est donc de *5 277 \$*.

Garanties achetées: l'assuré a acheté les couvertures suivantes: *A; B2; B3*.

Interface indicateurs :

L'enquêteur va « cliquer » avec la souris sur les indicateurs suivants car il observe leur présence :

- L'assuré est trop enthousiaste à prendre le blâme de l'accident.
- 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 couverture.

Pour l'enquêteur, l'indicateur le plus important est le deuxième inscrit ci-dessus. Après avoir entré ces indicateurs, la probabilité de fraude s'affichant en bas de l'écran indique 40 %.

Interface décision :

Cette partie permet au superviseur de décider de poursuivre ou non l'enquête de façon plus approfondie.

Montant estimé de remboursement s'il n'y a pas d'enquête. Dans ce cas, le montant estimé du sinistre est de 5 527 \$. La franchise applicable dans le cas de ce sinistre est de 250 \$. Le montant net de franchise est donc de 5 277 \$.

Montant de réclamation si l'enquête réussit. Étant donné l'expérience et la formation de l'enquêteur, ainsi que la non-existence des USE au sein de la compagnie (voir ci-dessus), l'enquêteur estime à 2 400 \$ le montant que la compagnie paierait si l'enquête réussissait.

Coût estimé pour faire une enquête approfondie. Ne connaissant pas le coût exact d'une enquête approfondie, l'enquêteur se base sur les données de l'industrie pour l'évaluer à 600 \$. Il a supposé que le dossier était de difficulté moyenne et que son coût d'enquête était semblable à celui de l'industrie (entre 500\$ et 750\$).

Probabilité subjective que l'enquête réussisse. La probabilité inscrite en bas des indicateurs représente la probabilité « objective » de fraude du dossier analysé. Si

l'enquêteur n'a pas beaucoup de formation ni d'expérience et si la compagnie ne dispose pas d'Unités Spéciales d'Enquête, alors on peut inscrire 40 % comme «probabilité subjective» pour que l'enquête réussisse.

On peut par contre supposer, par exemple, que notre enquêteur dispose d'un diplôme d'A.I.A.C., qu'il a sept ans d'expérience en assurances et que sa compagnie ne dispose pas d'USE. Dans ce cas, il pourrait estimer que la probabilité que l'enquête réussisse est de 45 % (étant entendu que la probabilité de fraude est de 40 %).

Résultat : Le résultat que le programme génère est que *OUI*, c'est rentable d'effectuer une enquête approfondie.

A1-6 Notions sur la probabilité seuil

Il se peut que la dernière partie de l'interface ne génère que des résultats où l'enquête ne serait pas profitable : c'est le cas par exemple où les montants de réclamations impliqués sont très petits. Pour une entreprise très impliquée dans la lutte à la fraude, elle peut décider de poursuivre l'enquête même si le montant est assez petit; elle basera alors sa décision que sur le niveau de probabilité de fraude. Si cette probabilité de fraude dépasse un certain seuil, alors elle mènera une enquête approfondie; sinon, elle paiera la réclamation et fermera le dossier. Afin de connaître le niveau seuil de cette probabilité, la compagnie pourra se baser sur le tableau de la page suivante.

Comme nous disposons de données moyennes de l'industrie, nous présenterons ici un cas typique d'une compagnie «moyenne».

Supposons un assureur qui ferme 15 220 dossiers de réclamations annuellement. Il dispose d'un budget anti-fraude de 630 108 \$ pour effectuer des enquêtes approfondies. Chaque enquête approfondie coûte en moyenne 600 \$. Avec ses 630 108 \$, il peut enquêter sur 1 050 dossiers de façon approfondie. Ceci représente 6,9 % du total des réclamations qu'il traite chaque année.

En se basant sur le tableau ci-dessous (concernant la moyenne de l'industrie), la compagnie en question saura que son niveau seuil de probabilité de fraude sera de 25 % : chaque dossier dont la probabilité est supérieure à 25 % sera l'objet d'une enquête de façon approfondie. Si les données de cette compagnie sont comparables à la moyenne de l'industrie, alors la dite compagnie aura épuisé ses ressources dédiées à la fraude à la fin de l'année (ni plus ni moins de ressources ne seront consacrées à la fraude)¹².

Tableau 1
L'échantillonnage, la détection et la précision en fonction
les seuils de 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 P > 25,5 %	6,85	49,30	53,40
Colonne S	Colonne E	Colonne R	Colonne D

Ainsi donc, en connaissant la valeur du budget alloué à la fraude, il est possible de connaître le seuil de probabilité de fraude.

Relations entre la précision et la détection

¹² Nous conseillons cependant l'adoption d'un tableau propre à chaque compagnie, même si une entreprise croit être représentative de la moyenne. Dans les cas où la compagnie est trop importante ou trop petite, le tableau précité n'est pas adapté.

Supposons toujours que notre entreprise possède le taux moyen de fraude de l'industrie qui est de 6,33 %. Alors, sur les 15 220 dossiers fermés annuellement, il existe 963 dossiers qui sont supposés frauduleux.

À un niveau seuil de probabilité de 25 %, le taux de détection est de 53,40 %. Sur les 963 dossiers frauduleux, on en détecte 514.

Lorsqu'on fait une enquête, il n'est pas sûr que tous les cas objet d'enquêtes représentent des fraudeurs réels. Selon le tableau ci dessus, plus le seuil de la probabilité est élevé, plus faible sera la marge d'erreur. À un niveau seuil de probabilité de fraude de 25 %, nous avons déduit qu'on détectait 514 cas de fraude. Parmi ces 514 cas, il y en a 48,95 % qui sont des *vrais* fraudeurs. En bout de ligne, pour un seuil de 25 %, on détecte 252 cas de fraude réelle.

Pour plus d'informations sur l'interprétation des taux de détection et de précision, se reporter à la section III du présent document.

Aperçu de l'écran de navigation dans le cahier de recherche seulement.

Annexe 2

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.