

Université de Montréal

**Étude des indicateurs pouvant prédire la présence
De fraude : une application à l'assurance
Automobile du Québec**

par

Mathieu Maurice

Département de sciences économiques

Faculté des arts et sciences

Novembre 1997

Remerciements

Je tiens à remercier monsieur Georges Dionne pour ses nombreux conseils, sa grande disponibilité tout au long de la rédaction de ce rapport et tiens à le féliciter pour ses grandes qualités didactiques.

Je voudrais également remercier le FCAR, le CRSH et le Centre de Recherche sur les Transports pour son soutien financier.

28 11 1998

TABLE DES MATIÈRES

INTRODUCTION	1
Chapitre 1 : Étude des caractéristiques concernant la fraude	3
1.1. Données	3
1.2. Le modèle Probit	4
1.3. Couvertures	5
1.4. Indicateurs de fraude	11
Chapitre 2 : Étude du modèle Probit bivarié	25
2.1. Introduction	25
2.2. Probit bivarié avec observation complète	25
2.3. Probit bivarié avec observation partielle	26
2.4. Probit bivarié avec observation conditionnelle	27
2.5. Estimation des indicateurs de fraude selon le modèle Probit bivarié avec observation conditionnelle	28
2.6. Analyse des résultats	40
2.7. Comparaison entre la méthode du probit bivarié et celle du rapport Belhadji- Dionne pour résoudre le biais de sélection	42
2.8. Analyse des différences entre la méthode avec processus d'élimination de Belhadji-Dionne et celle du probit bivarié	45
CONCLUSION	46
BIBLIOGRAPHIE.....	48
ANNEXE A	51

Introduction

Depuis quelques années, la fraude à l'assurance est devenue une préoccupation importante. Plusieurs explications des causes de la fraude ont été suggérées mais aucune étude n'a vraiment identifié sa principale cause. Par contre, on a souvent évoqué le fait que les cas de fraude les plus fréquents fassent intervenir des petits montants coûteux à vérifier et à prouver, ce qui réduit les bénéfices nets des enquêtes à court terme.

Certaines études tentent de définir quelle est la véritable proportion de dossiers frauduleux dans l'ensemble des réclamations. Plusieurs recherches américaines disent en effet qu'une réclamation sur dix contiendrait de la fraude à l'assurance. Une étude réalisée au Centre de recherche sur les transports (C.R.T.) réalisée par messieurs Georges Dionne et El-Bachir Belhadji (1996) évalue quant à elle le montant total de la fraude à l'assurance automobile au Québec pour la période allant d'avril 94 à mars 95. Celle-ci se chiffrait entre 28 et 61 millions de dollars. On comprend donc aisément que la fraude à l'assurance automobile représente un problème sérieux pour l'ensemble des compagnies d'assurances. Nous utiliserons les données de cette enquête dans notre recherche.

Le but de ce rapport est de déterminer quels sont les différents facteurs qui nous permettraient d'établir la possible présence de fraude dans les dossiers de réclamation. On peut en effet penser qu'il existe certaines caractéristiques présentes dans les dossiers qui augmentent la probabilité qu'une fraude soit perpétrée. En fait, l'enquête Dionne et Belhadji a demandé aux répondants d'inclure pour chaque dossier les indicateurs de fraude présents. Tirés d'une liste de cinquante indicateurs choisis, nous voulons vérifier si la présence de quelques-uns de ceux-ci amène une augmentation quant à la probabilité qu'une fraude soit commise. Ainsi, une compagnie d'assurances pourrait se servir de l'absence ou non d'indicateurs entraînant une hausse de la probabilité de fraude. Elle pourrait alors décider si une enquête plus poussée est requise lorsqu'un dossier est jugé comme pouvant contenir une part de fraude.

Dans un premier temps, nous allons étudier les différents indicateurs de fraude à l'aide d'un modèle probit. Nous trouverons ainsi les indicateurs qui augmentent de façon significative la probabilité qu'un dossier contienne de la fraude. Cependant, cette méthode introduit un biais de sélection quant à l'estimation des coefficients des indicateurs de fraude. En effet, nous ignorons si l'absence d'indicateurs de fraude dans un dossier est due à la non-réponse ou encore si celui-ci ne possède réellement aucun indicateur.

Pour résoudre ce problème, nous utiliserons dans un deuxième temps le modèle probit bivarié avec observation conditionnelle. Grâce à ce modèle, nous obtiendrons des estimations sans biais de sélection pour les coefficients de nos indicateurs.

Finalement, nous comparerons nos résultats avec ceux de l'étude Belhadji-Dionne (1997). Dans cette étude, les auteurs ont résolu le problème du biais de sélection en effectuant une étude exhaustive des réponses au sondage données par chaque répondant. Ainsi, un enquêteur devait au moins inscrire un indicateur parmi tous les dossiers qu'il avait remplis pour que ceux-ci soient retenus dans la banque de données. Si l'enquêteur n'avait jamais inscrit d'indicateurs de fraude, Belhadji-Dionne considéraient tous les dossiers remplis par celui-ci comme possédant de la non-réponse et les retiraient de l'étude. Ce processus d'élimination a conduit au retrait de 441 observations. La comparaison des résultats entre les deux études permettra de voir si les deux techniques donnent des résultats semblables ou si on remarque bel et bien des différences entre l'une et l'autre.

Chapitre 1

Étude des caractéristiques concernant la fraude

1.1 Données

Depuis quelques années, la fraude à l'assurance est devenue une préoccupation importante. Plusieurs études tentent d'évaluer le montant total de la fraude, tandis que d'autres essaient de trouver quelle fraction des réclamations peut être considérée comme frauduleuse. Le but de ce travail est plutôt de définir quelles sont les principales caractéristiques récurrentes lors des cas de fraude. Ainsi, les résultats trouvés pourraient donner des indications aux enquêteurs quant aux caractéristiques importantes dont il faudrait tenir compte lors des études de dossiers. Une démarche a été réalisée auprès des 20 compagnies d'assurance automobile ayant les plus grosses parts de marché au Québec afin de réaliser une enquête. 18 de ces compagnies ont accepté de participer au sondage, ces compagnies représentant 70 pour cent du marché de l'assurance automobile de la province de Québec. Nous pouvons donc dire que les données recueillies représentent bien le marché étudié.

Aux différentes compagnies d'assurances retenues, des questionnaires ont été acheminés au printemps 1995. Le nombre de questionnaires alloués à une compagnie était proportionnel à sa part de marché. Pour remplir les questionnaires, les compagnies devaient choisir des dossiers aléatoirement parmi tous leurs dossiers fermés durant la période du 1er avril 1994 au 31 mars 1995. Seules les réclamations pour «bris de vitre seulement» devaient être laissées de côté puisque celles-ci ne représentent pas vraiment un attrait pour les études de ce genre. Plusieurs renseignements spécifiques ont été demandés relativement aux réclamations des clients. Parmi les plus importantes: des questions objectives sur les types d'accidents, sur les montants payés et demandés des réclamations et sur le fait qu'il y ait eu fraude établie ou non, ainsi que des questions subjectives comme

le degré de suspicion de fraude lorsque celle-ci n'était pas établie ou encore le montant de la fraude suspecte ou établie compris dans le montant qui a été payé par la compagnie. Finalement, des renseignements sur la présence d'indicateurs de fraude dans les dossiers ont également été demandés. Qu'il y ait eu fraude ou non, les répondants devaient indiquer par ordre d'importance les indicateurs qui étaient présents dans le dossier. La liste des cinquante indicateurs de fraude utilisés lors de cette enquête se retrouve à l'annexe A.

1.2 Le modèle Probit

Afin d'estimer la probabilité qu'une fraude soit commise, nous avons utilisé le modèle Probit. Soit y_i^* une variable continue non observable qui représente la propension qu'il y ait fraude pour le dossier i . Cette variable est reliée aux variables explicatives x_i de la façon suivante:

$$y_i^* = \alpha + \beta x_i + \varepsilon_i$$

où α et β sont des coefficients et ε le terme d'erreur aléatoire.

Il y a fraude lorsque y_i^* dépasse un certain seuil. Dans notre modèle, ce seuil est zéro. Posons la variable indicatrice y_i pouvant prendre les valeurs 0 ou 1 selon qu'il y a eu fraude ($y_i = 1$) ou non ($y_i = 0$). Autrement dit:

$$y_i = \begin{cases} 1 & \text{si } \alpha + \beta x_i + \varepsilon_i > 0 \\ 0 & \text{sinon} \end{cases}$$

La vraisemblance d'un échantillon de n observations est donnée par :

$$L = \prod_{i=1}^n [F(-\alpha - \beta x_i)]^{y_i} [1 - F(-\alpha - \beta x_i)]^{(1-y_i)}$$

où F représente la fonction de répartition de la variable ε_i .

Si ε_i suit une loi de distribution normale, la maximisation de L par rapport aux coefficients α et β nous donne les expressions du maximum de vraisemblance du modèle Probit qui permettent d'estimer les paramètres.

Nous connaissons cependant un peu plus d'information sur la probabilité qu'il y ait eu fraude, et ce grâce à la variable de suspicion donnée par l'enquête. Nous pourrions donc ranger dans la classe fraude les dossiers ayant un degré de suspicion supérieur à 20 pour cent par exemple, ou 40 pour cent, ou 50 pour cent et ainsi de suite. Ainsi, nous pourrions vérifier si nos résultats possèdent une certaine robustesse quant à la classification des dossiers suspectés frauduleux. Ceci entraînera cependant plusieurs régressions, ce qui affecte le niveau des tests employés. En effet, si on exécute par exemple 80 régressions avec un niveau critique de cinq pour cent, on doit s'attendre à avoir en moyenne quatre résultats significatifs résultant uniquement du hasard. Il sera donc préférable de considérer les tests significatifs à plusieurs reprises lorsqu'il y aura plusieurs régressions.

1.3 Couvertures

La première question qui nous vient à l'esprit est de savoir si la probabilité qu'il y ait fraude varie selon les différentes couvertures. Il existe cinq genres de couverture lorsque l'on parle d'assurance automobile. La première, le chapitre A, est la responsabilité civile et couvre tous dommages aux autres véhicules et aux infrastructures causés par la personne fautive. Il est obligatoire au Québec de posséder cette couverture. Viennent ensuite les chapitres B1 (assurances tous risques), B2 (assurance collision ou versement), B3 (assurance sans collision) et B4 (risques spécifiés). On peut également se procurer des avenants tels que remise valeur à neuf, incapacité totale ou mutilation. Comme le nombre de couvertures activées de ces avenants est peu élevé, nous les regrouperons dans une sixième catégorie.

Tableau 1.3.1 : Fraude par couverture

Chapitres	Non fraude (N)	Suspecte (S)	Établie (E)	nombre de couvertures	% S+E	% E
A	1285	40	3	1328	3.2 %	0.23 %
B1	130	13	2	145	10.3 %	1.38 %
B2	710	28	7	745	4.7 %	0.94 %
B3	416	37	6	459	9.4 %	1.31 %
B4	24	2	1	27	11.11 %	3.70 %
Avenants	66	3	0	69	4.3 %	0 %
Total	2631	123	19	2773	5.1 %	0.69 %

Lorsque l'on étudie attentivement les pourcentages du tableau précédent, on remarque que les chapitres B1, B2 et B4 sont ceux ayant les plus hauts taux de fraude. Effectuons un Probit sur l'équation:

$$\text{fraude} = \alpha + \beta_1 \text{catB1} + \beta_2 \text{catB2} + \beta_3 \text{catB3} + \beta_4 \text{catB4} + \beta_5 \text{avenants} + \varepsilon$$

où les variables **catB1**, **catB2**, **catB3**, **catB4** et **avenants** sont dichotomiques et prennent la valeur 1 si la couverture activée est respectivement B1, B2, B3, B4 ou un avenant et la valeur 0 autrement. ε représente le terme d'erreur. Finalement, la variable **fraude** prend 1 comme valeur si la couverture étudiée a été considérée comme contenant une part de fraude, 0 sinon. On constate que les variables du modèle sont toutes mutuellement exclusives puisque nos observations sont constituées de l'ensemble des couvertures ayant été activées. Il faut donc retirer une des variables, dans notre cas celle représentant le chapitre A, qui nous servira de référence. À noter que dans le tableau de la page suivante, le nombre entre parenthèses représente la valeur-p du coefficient. Ainsi, en supposant comme frauduleux les cas suspects à plus de 50%, nous avons :

$$P\{ (\beta_1=0.5951) = 0 \} = 0.0008$$

Tableau 1.3.2 : Coefficients de régression pour les différents chapitres

groupe de référence : Chapitre A

variables	(E+S) avec S ≥ 0 %	(E+S) avec S ≥ 20 %	(E+S) avec S ≥ 40 %	(E+S) avec S ≥ 50 %	(E+S) avec S ≥ 60 %	(E+S) avec S ≥ 80%	fraude établie (E)
constante	-1.8681 (0.0001)	-1.9851 (0.0001)	-2.0786 (0.0001)	-2.0786 (0.0001)	-2.2322 (0.0001)	-2.3065 (0.0001)	-2.8395 (0.0001)
catB1	0.6060 (0.0001)	0.7270 (0.0001)	0.7772 (0.0001)	0.5951 (0.0008)	0.6946 (0.0002)	0.7098 (0.0003)	0.6364 (0.0544)
catB2	0.1932 (0.0640)	0.2258 (0.0451)	0.2477 (0.0389)	0.2105 (0.0841)	0.1528 (0.2853)	0.1301 (0.4016)	0.4899 (0.0341)
catB3	0.5366 (0.0001)	0.6167 (0.0001)	0.6629 (0.0001)	0.6325 (0.0001)	0.7209 (0.0001)	0.6412 (0.0001)	0.6155 (0.0111)
catB4	0.6475 (0.0474)	0.5430 (0.1392)	0.6325 (0.0860)	0.6325 (0.0860)	0.7861 (0.0343)	0.8604 (0.0211)	1.0534 (0.0300)
avenants	0.1564 (0.5693)	0.2775 (0.3159)	0.3670 (0.1874)	0.3670 (0.1874)	0.5205 (0.0651)	0.4106 (0.2016)	-4.1618 (0.9998)
# fraude=1	139	119	105	97	75	59	19
Log-likelihood	-534.5885	-470.8247	-425.203	-403.768	-324.522	-270.384	-107.729

Comme une variable subjective nous donne une certaine information quant au degré de doute lors d'une fraude suspecte, nous avons effectué plusieurs régressions en changeant toujours le critère de sélection lors de ces cas. Ainsi, une première régression a été réalisée en considérant comme frauduleux les seuls dossiers où la fraude a été établie. Ensuite, les dossiers étant suspectés à plus de 20, 40, 50, 60 et 80 pour cent ont été tour à tour mis au nombre des frauduleux. Enfin, tous les dossiers suspects ont rejoint ceux de fraude établie. Ainsi, nous pourrions vérifier s'il y a une certaine robustesse des résultats quant à la spécification de la variable dépendante.

On constate premièrement que la probabilité d'avoir une fraude dans les chapitres B1 et B3 est plus élevée que dans le chapitre A, et ce quel que soit le critère de sélection des dossiers suspectés. Tout porte à croire que le chapitre B4 a également une plus forte probabilité de fraude que le chapitre A puisque qu'un seul résultat n'est pas significatif, soit celui où tous les dossiers suspects sont incorporés avec les dossiers frauduleux. Pour le chapitre B2, il est difficile de dire si la probabilité de fraude est plus grande que pour le chapitre A puisqu'à l'instant où seulement les dossiers suspects à plus de 50 pour cent sont considérés comme frauduleux, aucune différence ne peut être détectée entre les deux chapitres sauf pour les dossiers avec fraude établie. Finalement, on ne peut dire qu'il y a une différence sur la probabilité d'avoir une fraude entre le chapitre A et la catégorie des avenants.

En analysant d'un peu plus près ces derniers résultats, on constate que les chapitres contenant les sinistres sans collision sont ceux qui ressortent le plus lors des dernières régressions. Cela concorde avec l'intuition qu'un gonflement de facture dans le but de toucher une plus grande réclamation est plus facile à perpétrer lorsqu'il n'y a pas de collision avec un autre véhicule ou qu'il n'y a pas de témoin. Nous allons nous intéresser aux couvertures B de façon agrégée. D'un côté, les sinistres avec collision (chapitre B2 et assurances tous risques (B1) avec collision) et les accidents sans collision (chapitre B3, chapitre B4 et assurances tous risques (B1) sans collision).

Dans la première étape (tableau 3), le groupe de référence est demeuré le chapitre A. Les résultats s'apparentent beaucoup à ceux du tableau 2. Pour la deuxième partie (tableau 4), on a changé le groupe de référence, passant du chapitre A à celui des accidents avec collision. On remarque que les dossiers provenant d'un accident sans collision ont une probabilité plus élevée de contenir de la fraude que ceux ayant une collision. Globalement, ce sont donc les sinistres n'impliquant pas de collision qui sont les plus sujets aux fraudes à l'assurance automobile.

Tableau 1.3.3 : Régression par couvertures agrégées

Groupe de référence : Chapitre A

variables	(E+S) avec S ≥ 0 %	(E+S) avec S ≥ 20 %	(E+S) avec S ≥ 40 %	(E+S) avec S ≥ 50 %	(E+S) avec S ≥ 60 %	(E+S) avec S ≥ 80 %	fraude établie (E)
constante	-1.8681 (0.0001)	-1.9891 (0.0001)	-2.0786 (0.0001)	-2.0786 (0.0001)	-2.2322 (0.0001)	-2.3065 (0.0001)	-2.8395 (0.0001)
collision	0.2373 (0.0164)	0.2878 (0.0067)	0.3114 (0.0059)	0.2206 (0.0593)	0.1997 (0.1382)	0.1991 (0.1677)	0.5325 (0.0167)
sans collision	0.5744 (0.0001)	0.6497 (0.0001)	0.7029 (0.0001)	0.6777 (0.0001)	0.7639 (0.0001)	0.6969 (0.0001)	0.6261 (0.0078)
avenants	0.1564 (0.5693)	0.2775 (0.3159)	0.3670 (0.1874)	0.3670 (0.1874)	0.5205 (0.0651)	0.4106 (0.2016)	-4.1618 (0.9998)
Log- likelihood	-535.6232	-472.7892	-427.0097	-402.2977	-324.1357	-271.3182	-108.3341

Valeurs-p entre parenthèses

Tableau 1.3.4 : Régression par couvertures agrégées

Groupe de référence : Accidents avec collision

variables	(E+S) avec $S \geq 0\%$	(E+S) avec $S \geq 20\%$	(E+S) avec $S \geq 40\%$	(E+S) avec $S \geq 50\%$	(E+S) avec $S \geq 60\%$	(E+S) avec $S \geq 80\%$	fraude établie (E)
constante	-1.6308 (0.0001)	-1.7013 (0.0001)	-1.7672 (0.0001)	-1.8581 (0.0001)	-2.0325 (0.0001)	-2.1073 (0.0001)	-2.3070 (0.0001)
Chapitre A	-0.2373 (0.0164)	-0.2878 (0.0067)	-0.3114 (0.0059)	-0.2205 (0.0059)	-0.1997 (0.1382)	-0.1991 (0.1677)	-0.5325 (0.0167)
sans collision	0.3372 (0.0012)	0.3618 (0.0008)	0.3915 (0.0004)	0.4571 (0.0001)	0.5642 (0.0001)	0.4977 (0.0003)	0.0937 (0.6269)
avenants	-0.0808 (0.7694)	-0.0104 (0.9701)	-0.0555 (0.8415)	0.1464 (0.6002)	0.3208 (0.2577)	0.2114 (0.5121)	-4.6943 (0.9998)
Log- likelihood	-535.6232	-472.7892	-427.0097	-402.2977	-324.1357	-271.3182	-108.3341

Valeurs-p entre parenthèses

1.4 Indicateurs de fraude

Lors de l'enquête, il avait été demandé aux répondants d'inscrire par ordre d'importance les indicateurs de fraude présents dans le dossier, qu'ils soient frauduleux ou non. Une liste de 50 indicateurs leur avait été auparavant remise. Cette liste se trouve à l'annexe A.

Des 2773 observations, seules 1345 possèdent au moins un indicateur de fraude, soit une proportion de 48,5 pour cent. La première difficulté est de définir si l'absence d'indicateur est due à un taux élevé de non-réponse ou encore à l'absence réelle de ceux-ci puisque les dossiers avec fraude ne représentent que 0.69% (fraude établie) ou 5,1% (fraude suspecte et établie) de l'échantillon. En effet, il n'est pas facile de répondre à la question sur les indicateurs et il est possible que dans les cas sans fraude, les personnes interrogées n'aient pas pris la peine d'y répondre. Une façon de contourner ce problème aurait été de vérifier si une autre des questions jugées difficiles à été complétée. Malheureusement, celles-ci se retrouvent dans les cas de dossiers suspects possédant presque tous au moins un indicateur de fraude. Il est donc préférable, à partir de maintenant, d'éliminer des régressions les dossiers ne contenant pas d'indicateur. Ceci peut entraîner cependant un biais. Effectivement, les 1428 observations retirées ne représentent pas toutes nécessairement des cas de non-réponse. Ainsi, étant à très forte majorité des cas sans fraude (1423 observations sur 1428, soit une proportion de 99.65%) les observations laissées pour compte auront pour effet de biaiser en baisse la valeur réelle du coefficient. Les résultats trouvés seront donc conservateurs, c'est-à-dire que les indicateurs influençant la probabilité de fraude d'une façon significative trouvés dans ce modèle devraient également hausser la probabilité qu'une fraude soit commise dans un modèle sans biais de sélection. Cependant, certains indicateurs pourraient être considérés comme n'affectant pas la probabilité de fraude dans le modèle avec biais de sélection alors qu'en réalité, ils augmentent eux aussi la probabilité qu'une fraude soit commise.

Vérifions d'abord si le nombre d'indicateurs présent dans le dossier augmente la probabilité d'avoir une fraude. On peut le faire en testant l'équation :

$$\text{fraude} = \alpha + \beta \text{ nombre} + \varepsilon$$

où **nombre** est une variable discrète prenant comme valeur le nombre d'indicateurs présents dans le dossier et ε est un terme d'erreur. En effectuant les régressions pour trois des critères de sélection des dossiers suspects retenus auparavant, on s'aperçoit qu'un nombre élevé d'indicateurs présents dans le dossier augmente de beaucoup la probabilité que celui-ci renferme de la fraude.

Tableau 1.4.1 : Nombre d'indicateurs

variables	(E+S) avec S ≥ 0 %	(E+S) avec S ≥ 50 %	fraude établie (E)
constante	-2.3571 (0.0001)	-2.5123 (0.0001)	-2.9990 (0.0001)
nombre	0.4477 (0.0001)	0.4504 (0.0001)	0.3105 (0.0001)
Log-likelihood	-384.3195	-308.5790	-89.1369

Valeurs-p entre parenthèses

Vérifions maintenant quels sont les indicateurs modifiant de façon significative la probabilité d'avoir une fraude. On peut le faire en employant une variable dichotomique quantifiant la présence ou non de l'indicateur dans la réclamation. Les tableaux 1.4.2 à 1.4.6 qui suivent présentent ces régression en considérant respectivement les cas suspectés comme frauduleux à plus de 0%, 30%, 50%, 80% et les cas de fraude établie comme variable dépendante.

Tableau 1.4.2 : Indicateurs de fraude
Fraude suspecte et établie (E+S)

Variables	coefficients	t	valeurs-p
Constante	-2.4037	-16.155	0.0000
Indicateur1	0.3718	2.249	0.0245
Indicateur2	0.4906	1.938	0.0526
Indicateur3	1.1997	5.349	0.0000
Indicateur4	0.6130	1.869	0.0617
Indicateur7	0.1715	0.599	0.5492
Indicateur10	-0.3210	-1.196	0.2318
Indicateur11	0.7984	2.903	0.0037
Indicateur13	0.1991	0.791	0.4288
Indicateur14	0.6365	1.536	0.1244
Indicateur15	0.5202	2.022	0.0432
Indicateur16	0.7934	2.869	0.0041
Indicateur18	1.6517	5.239	0.0000
Indicateur19	0.3799	1.690	0.0911
Indicateur21	0.6565	1.753	0.0796
Indicateur22	1.7300	4.877	0.0000
Indicateur24	1.2657	2.781	0.0054
Indicateur26	0.9580	2.639	0.0083
Indicateur28	1.5941	2.091	0.0365
Indicateur30	0.6623	3.224	0.0013
Indicateur32	0.6732	3.404	0.0007
Indicateur36	0.8524	1.344	0.1788
Indicateur39	0.8671	2.916	0.0035
Indicateur40	0.9248	2.233	0.0256
Indicateur45	0.4919	1.482	0.1382
Indicateur47	0.9232	3.565	0.0004

Tableau 1.4.2 : Indicateurs de fraude(suite)
Fraude suspecte et établie (E+S)

Variables	coefficients	t	valeurs-p
Indicateur48	0.1895	0.854	0.3931
Nombre	-0.0357	-0.396	0.6918
Catb1	0.8352	3.727	0.0002
Catb2	0.0041	0.025	0.9798
Catb3	0.4220	2.413	0.0158
Catb4	0.1328	0.255	0.7984
Catav	-0.0119	-0.031	0.9754
Noncol	0.0015	0.138	0.8899
Log-likelihood	-280.58		
N	1345		

Tableau 1.4.3 : Indicateurs de fraude
Fraude suspecte à 30% et plus (E+S>2)

Variables	coefficients	t	valeurs-p
Constante	-2.5848	-15.739	0.0000
Indicateur1	0.3681	2.127	0.0334
Indicateur2	0.4788	1.744	0.0761
Indicateur3	1.2285	5.350	0.0000
Indicateur4	0.6015	1.821	0.0685
Indicateur7	0.2053	0.708	0.4790
Indicateur10	-0.2779	-1.020	0.3079
Indicateur11	0.9126	3.276	0.0011
Indicateur13	0.2398	0.925	0.3550
Indicateur14	0.3787	0.897	0.3699
Indicateur15	0.5327	2.001	0.0453
Indicateur16	0.7833	2.691	0.0071
Indicateur18	1.7075	5.351	0.0000
Indicateur19	0.4682	2.053	0.0401
Indicateur21	0.7952	2.107	0.0346
Indicateur22	1.6470	4.672	0.0000
Indicateur24	1.3707	2.992	0.0028
Indicateur26	0.5468	1.205	0.2282
Indicateur28	1.6289	2.113	0.0346
Indicateur30	0.7052	3.275	0.0011
Indicateur32	0.6189	3.031	0.0024
Indicateur36	0.7962	1.243	0.2140
Indicateur39	0.6891	2.149	0.0316
Indicateur40	0.9652	2.277	0.0228
Indicateur45	0.4445	1.304	0.1923
Indicateur47	0.8879	3.307	0.0009

Tableau 1.4.3 : Indicateurs de fraude (suite)
Fraude suspecte à 30% et plus (E+S>2)

Variables	coefficients	t	valeurs-p
Indicateur48	0.2151	0.934	0.3505
Nombre	-0.0473	-0.514	0.6076
Catb1	1.0289	4.450	0.0000
Catb2	0.1432	0.818	0.4132
Catb3	0.6436	3.491	0.0005
Catb4	0.3783	0.729	0.4658
Catav	0.1648	0.420	0.6743
Noncol	0.0014	0.113	0.9104
Log-likelihood	-255.46		
N	1345		

Tableau 1.4.4 : Indicateurs de fraude
Fraude suspecte à 50% et plus (E+S>4)

Variables	coefficients	t	valeurs-p
Constante	-2.7379	-15.471	0.0000
Indicateur1	0.5102	2.718	0.0066
Indicateur2	0.6359	2.283	0.0224
Indicateur3	1.3234	5.523	0.0000
Indicateur4	0.5653	1.548	0.1217
Indicateur7	0.4015	1.340	0.1802
Indicateur10	-0.2813	-0.930	0.3523
Indicateur11	1.0691	3.713	0.0002
Indicateur13	0.3857	1.447	0.1480
Indicateur14	0.4597	1.066	0.2865
Indicateur15	0.6522	2.385	0.0171
Indicateur16	0.9154	3.040	0.0024
Indicateur18	1.5868	4.750	0.0000
Indicateur19	0.6044	2.485	0.0129
Indicateur21	0.8613	2.246	0.0247
Indicateur22	1.8843	5.265	0.0000
Indicateur24	1.3701	2.792	0.0052
Indicateur26	0.7535	1.615	0.1062
Indicateur28	1.8733	2.391	0.0168
Indicateur30	0.7801	3.374	0.0007
Indicateur32	0.5847	2.535	0.0113
Indicateur36	0.9419	1.439	0.1503
Indicateur39	0.4074	1.077	0.2814
Indicateur40	1.0838	2.521	0.0117
Indicateur45	0.6073	1.750	0.0802
Indicateur47	0.9253	3.343	0.0008

Tableau 1.4.4 : Indicateurs de fraude (suite)
Fraude suspecte à 50% et plus (E+S>4)

Variables	coefficients	t	valeurs-p
Indicateur48	0.3861	1.591	0.1116
Nombre	-0.0849	-0.864	0.3877
Catb1	0.6742	2.331	0.0197
Catb2	0.0771	0.413	0.6796
Catb3	0.5341	1.753	0.0796
Catb4	0.3991	0.687	0.4918
Catav	0.2606	0.635	0.5254
Noncol	0.0713	0.254	0.7998
Log-likelihood	-220.60		
N	1345		

Tableau 1.4.5 : Indicateurs de fraude
Fraude suspecte à 80% et plus (E+S>7)

Variables	coefficients	t	valeurs-p
Constante	-3.2456	-12.705	0.0000
Indicateur1	0.8944	3.550	0.0004
Indicateur2	0.1410	0.292	0.7700
Indicateur3	1.5248	5.105	0.0000
Indicateur4	0.1642	0.310	0.7568
Indicateur7	0.6069	1.620	0.1052
Indicateur10	0.2187	0.663	0.5071
Indicateur11	1.5121	4.422	0.0000
Indicateur13	0.2308	0.646	0.5182
Indicateur14	1.0428	2.299	0.0215
Indicateur15	0.7629	2.302	0.0213
Indicateur16	1.3378	3.734	0.0002
Indicateur18	1.6610	4.115	0.0000
Indicateur19	0.9903	3.367	0.0008
Indicateur21	1.0421	2.377	0.0175
Indicateur22	2.0067	4.952	0.0000
Indicateur24	1.7607	3.034	0.0024
Indicateur26	0.1280	0.165	0.8686
Indicateur28	2.5320	3.156	0.0016
Indicateur30	0.9018	3.015	0.0026
Indicateur32	0.5836	1.928	0.0539
Indicateur36	1.1183	1.441	0.1497
Indicateur39	0.1346	0.234	0.8153
Indicateur40	0.7916	1.419	0.1560
Indicateur45	1.1402	2.887	0.0039
Indicateur47	0.6180	1.735	0.0827

Tableau 1.4.5 : Indicateurs de fraude (suite)
Fraude suspecte à 80% et plus (E+S>7)

Variables	coefficients	t	valeurs-p
Indicateur48	0.1848	0.493	0.6220
Nombre	-0.1983	-1.584	0.1132
Catb1	0.9675	3.104	0.0019
Catb2	0.0129	0.051	0.9595
Catb3	0.7159	2.833	0.0046
Catb4	0.9921	1.792	0.0732
Catav	0.4826	0.969	0.3324
Noncol	0.0010	0.066	0.9474
Log-likelihood	-135.78		
N	1345		

Tableau 1.4.6 : Indicateurs de fraude
Fraude établie (E)

Variables	coefficients	t	valeurs-p
Constante	-4.3980	-6.607	0.0000
Indicateur1	-0.0535	-0.088	0.9295
Indicateur2	-4.6092	-0.057	0.9544
Indicateur3	2.0499	3.782	0.0002
Indicateur4	0.1695	0.139	0.8895
Indicateur7	-2.9042	-1.314	0.1888
Indicateur10	-4.1204	-0.076	0.9397
Indicateur11	2.6582	4.156	0.0000
Indicateur13	0.0499	0.076	0.9398
Indicateur14	0.9426	0.969	0.3324
Indicateur15	-0.0464	-0.052	0.9584
Indicateur16	1.1410	1.471	0.1412
Indicateur18	2.6193	3.891	0.0001
Indicateur19	0.8621	1.402	0.1610
Indicateur21	0.4554	0.229	0.8189
Indicateur22	1.5506	1.850	0.0643
Indicateur24	2.0851	2.134	0.0329
Indicateur26	-4.1057	-0.030	0.9761
Indicateur28	-1.5619	-0.004	0.9966
Indicateur30	1.2322	2.109	0.0350
Indicateur32	0.2761	0.406	0.6849
Indicateur36	1.9383	1.684	0.0923
Indicateur39	-2.8009	-0.026	0.9790
Indicateur40	1.2518	1.065	0.2868
Indicateur45	-7.9175	-0.060	0.9523
Indicateur47	1.3498	2.109	0.0350

Tableau 1.4.6 : Indicateurs de fraude (suite)

Fraude établie (E)

Variables	coefficients	t	valeurs-p
Indicateur48	0.7124	1.073	0.2831
Nombre	-0.1956	-0.743	0.4575
Catb1	1.3056	1.747	0.0806
Catb2	0.7961	1.609	0.1075
Catb3	1.0731	1.250	0.2112
Catb4	1.2266	0.856	0.3918
Catav	-2.6557	-0.023	0.9816
Noncol	0.0798	0.102	0.9187
Log-likelihood	-37.171		
N	1345		

1.5 Résultats

À partir des derniers tableaux, on s'aperçoit que certains indicateurs varient moins que d'autres selon la définition de la variable dépendante. On note en effet que les indicateurs 3, 11, 18, 24, 30 et 47 augmentent la probabilité de fraude, et ce que l'on considère les seuls dossiers ayant été établis comme contenant de la fraude ou encore en incorporant graduellement les dossiers suspects. Ces indicateurs sont :

- 3 : Existence de tout dommage non relié à la perte ou incohérent avec les faits de l'accident.
- 11 : L'assuré éprouve des difficultés financières personnelles ou reliées à ses affaires.
- 18 : L'assuré est trop enthousiaste ou trop franc pour prendre le blâme de l'accident.
- 24 : L'historique de la propriété du véhicule ne peut être établie.
- 30 : Témoignages contradictoires concernant les circonstances de la perte.
- 47 : Lors de l'enquête, l'assuré est nerveux et semble confus.

Considérons maintenant les indicateurs qui augmentent la probabilité de fraude pour les dossiers de fraude établie et suspects à 50 pour cent et plus. En plus des six nommés précédemment s'ajoutent les indicateurs 1, 2, 15, 16, 19, 21, 22, 28, 32 et 40 qui sont :

- 1 : Pas de rapport de police quand celui-ci était nécessaire.
- 2 : Une collision mineure a entraîné des coûts de réparation excessif.
- 15 : L'assuré est très insistant pour un règlement rapide.
- 16 : L'assuré est extraordinairement familier avec le jargon des assurances.
- 19 : L'accident 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.
- 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.
- 28 : Le représentant des réclamations se voit refusé l'examen du véhicule.
- 32 : Accident impliquant un seul véhicule.
- 40 : L'assuré est au chômage ; travaille dans une industrie en dépression ; vit dans une région défavorisée.

Ajoutons finalement à cette liste les indicateurs augmentant la probabilité de fraude lorsque l'on considère comme étant frauduleux les dossiers de fraude établie et tous les dossiers suspects. On devra être plus prudent avec ces indicateurs puisque l'on ajoute aux dossiers frauduleux des cas suspects à moins de 50 pour cent, donc moins solides que pour les deux étapes précédentes. Il est cependant intéressant de vérifier quand même ces résultats puisque les dossiers suspectés le sont par les représentants des compagnies et il est fort possible qu'une certaine part de la fraude passe inaperçue à leurs yeux, augmentant ainsi la proportion réelle de cas de fraude à l'assurance. Deux seuls indicateurs, 26 et 39, s'ajoutent aux précédents. Ce sont :

- 26 : La documentation de l'estimation et de la réparation n'est pas disponible.
- 39 : L'achat du véhicule s'est fait au comptant.

Notons finalement que la variable représentant le nombre d'indicateurs présents dans le dossier n'influence plus la probabilité de fraude lorsque mise avec les indicateurs. La variable dichotomique Noncol tenant compte de la présence ou non d'une collision lors du sinistre n'influence également pas la probabilité de fraude présente. Au prochain chapitre, nous emploierons une méthode d'estimation différente dans le but de corriger le biais de sélection potentiel introduit lors de la mise de côté des observations ne contenant pas d'indicateurs de fraude.

Chapitre 2

Étude du modèle probit bivarié

2.1 Introduction

Dans ce chapitre, nous étudierons les mécanismes du modèle probit bivarié. Celui-ci est utilisé lorsque l'on doit estimer conjointement deux équations tirées d'un modèle probit. Le modèle probit bivarié peut être séparé en trois grandes catégories. Il y a tout d'abord le cas où les deux équations du modèle sont observées complètement, vient ensuite le cas où les deux équations sont observées partiellement et enfin lorsqu'une des équations est observée complètement et l'autre partiellement. Les sections qui suivent décrivent en détail les techniques d'estimation reliées à ces trois catégories.

2.2 Probit bivarié avec observation complète

Ce modèle a été pour la première fois utilisé par Zellner et Lee [1965]. Ils ont considéré les équations :

$$\begin{aligned}y_1^* &= X_1' \beta_1 + \varepsilon_1 \\ y_2^* &= X_2' \beta_2 + \varepsilon_2\end{aligned}$$

où y_j^* n'est pas observée, mais est reliée à la variable dépendante y_j de la façon suivante :

$$y_j = \begin{cases} 1 & \text{si } y_j^* > 0 \\ 0 & \text{sinon} \end{cases} \quad j = 1, 2$$

les erreurs $(\varepsilon_1, \varepsilon_2)'$ sont considérées comme suivant une loi binormale avec un coefficient de corrélation égal à ρ :

$$(\varepsilon_1, \varepsilon_2)' \sim \text{BN} \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right]$$

Comme les deux équations sont observées complètement, il est toujours possible de les estimer séparément. Mais lorsque $\rho \neq 0$, Zellner et Lee ont démontré qu'il est plus efficace d'estimer simultanément par la méthode de maximum de vraisemblance les deux équations. La fonction de maximum de vraisemblance est :

$$\begin{aligned} \ln L(\beta_1, \beta_2, \rho) = & \sum_{i=1}^n \{ y_{i1} y_{i2} \ln F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho) \\ & + y_{i1} (1 - y_{i2}) \ln [\Phi(X_{i1}'\beta_1) - F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho)] \\ & + (1 - y_{i1}) y_{i2} \ln [\Phi(X_{i2}'\beta_2) - F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho)] \\ & + (1 - y_{i1}) (1 - y_{i2}) \cdot \\ & \ln [1 - \Phi(X_{i1}'\beta_1) \Phi(X_{i2}'\beta_2) - F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho)] \end{aligned}$$

où $F(\bullet, \bullet, \rho)$ est la fonction de distribution d'une binormale et

$\Phi(\bullet)$ est la fonction de distribution de la loi normale.

2.3 Probit bivarié avec observation partielle

Ce cas a été traité pour la première fois par Poirier[1980]. Pour bien comprendre le fonctionnement de ce modèle, considérons un comité de deux personnes votant anonymement sous une règle d'unanimité. On observe seulement si une motion est acceptée (les deux membres ont voté « oui ») ou rejetée (au moins un des membres a voté « non »).

À la place de y_{i1} et y_{i2} , on observe que la variable $z_i = y_{i1} \cdot y_{i2}$ $i = 1, 2, 3, \dots, n$
Ainsi :

$$z_i \begin{cases} 1 \text{ si } y_{i1} = y_{i2} = 1 \\ 0 \text{ sinon} \end{cases}$$

Des quatre combinaisons possibles [(oui, oui) , (non, oui) , (oui, non) , (non, non)], les trois dernières ne peuvent être dissociées puisque l'on observe seulement que la motion n'est pas acceptée. Poirier a démontré qu'il faut alors estimer simultanément les

deux équations par la méthode du maximum de vraisemblance. La fonction de maximum de vraisemblance est alors :

$$\ln L(\beta_1, \beta_2, \rho) = \sum_{i=1}^n \{ z_i \ln F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho) + (1 - z_i) \ln [1 - F(X_{i1}'\beta_1; X_{i2}'\beta_2; \rho)] \}$$

2.4 Probit bivarié avec observation conditionnelle

Ce modèle se trouve à mi-chemin entre celui de Zellner et Lee et celui de Poirier. En fait, on observe toujours la variable $z_i = y_{i1} \cdot y_{i2}$ mais en plus, on connaît la valeur de y_{i1} . Ainsi, on n'observe y_{i2} que dans le cas où $y_{i1} = 1$. En effet, si $y_{i1} = 0$, alors $z_i = 0$ et on n'a plus aucune idée de la valeur que prend y_{i2} . Le modèle a été utilisé pour la première fois par Farber[1982] pour étudier la demande d'adhésion d'un individu à un syndicat.

Soit :

$y_{i1} = 1$ si l'individu i veut faire partie du syndicat, 0 sinon

$y_{i2} = 1$ si le syndicat accepte l'individu i , 0 sinon

Il est clair que l'on observe y_{i2} , la décision du syndicat, seulement si l'individu i a fait une demande pour y entrer. Comme la première équation est observée complètement, on peut l'estimer sans tenir compte de la deuxième. Cependant, cela sera inefficace sauf si $\rho = 0$. En plus, il y aura biais de sélection si l'on sépare les deux équations et que $\rho \neq 0$. On doit donc estimer les deux équations simultanément par la méthode du maximum de vraisemblance, où la fonction de maximum de vraisemblance est :

$$\ln L(\beta_1, \beta_2, \rho) = \sum_{i=1}^n \{ y_{i1} y_{i2} \ln F(X_{i1}'\beta_1, X_{i2}'\beta_2, \rho) + y_{i1} (1 - y_{i2}) \ln [F(X_{i1}'\beta_1, -X_{i2}'\beta_2, -\rho)] + (1 - y_{i1}) \ln \Phi(-X_{i1}'\beta_1) \}$$

2.5 Estimation des indicateurs de fraude selon le modèle du probit bivarié avec observation conditionnelle.

Notre problème consiste à retirer la possibilité qu'un biais de sélection se trouve lorsque nous étudions les indicateurs de fraude présents dans un dossier. Pour ce faire, nous allons utiliser le modèle de Farber du probit bivarié avec observation conditionnelle. Nous pouvons en effet estimer le modèle posé au chapitre un où seules les observations possédant un ou plusieurs indicateurs de fraude sont intégrées au modèle, conditionnellement à une nouvelle équation tentant d'expliquer la probabilité qu'au moins un indicateur soit présent dans l'analyse du dossier. Les variables qui ont été choisies pour évaluer la probabilité qu'un indicateur de fraude soit présent au dossier sont les indicatrices des différentes compagnies (englobant ainsi les enquêteurs ayant participé au sondage) et les indicatrices des couvertures activées (contrôlant pour la différence des proportions des cas de fraude entre les différentes couvertures). Les deux équations de notre modèle sont donc :

$$y_{i1} = \alpha_1 + \sum_j \beta_{j,1} \text{cie}_{ij} + \beta_{18,1} \text{catb1} + \beta_{19,1} \text{catb2} + \beta_{20,1} \text{catb3} + \beta_{21,1} \text{catb4} + \beta_{22,1} \text{catav} \\ + \varepsilon_1$$

$$y_{i2} = \alpha_2 + \sum_k \beta_{k,2} \text{ind}_{ik} + \beta_{27,2} \text{catb1} + \beta_{28,2} \text{catb2} + \beta_{29,2} \text{catb3} + \beta_{30,2} \text{catb4} + \beta_{31,2} \text{catav} \\ + \beta_{32,2} \text{noncol} + \varepsilon_2$$

où $y_{i1} = 1$ si au moins un indicateur est présent dans le dossier i , 0 sinon

$y_{i2} = 1$ si le dossier i contient de la fraude, 0 sinon

α_p est le terme constant de l'équation p ; $p = 1, 2$

ε_p est le terme d'erreur associé à l'équation p

$\beta_{q,1}$ sont les coefficients de l'équation 1 à estimer ; $q = 1, 2, \dots, 22$

$\beta_{r,2}$ sont les coefficients de l'équation 2 à estimer ; $r = 1, 2, \dots, 32$

$\text{cie}_{ij} = 1$ si le dossier i vient de la compagnie d'assurances j , 0 sinon ; $j = 1, 2, \dots, 17$

$\text{ind}_{ik} = 1$ si l'indicateur k est présent dans le dossier i , 0 sinon ; $k = 1, 2, \dots, 26$

Notre modèle est bien conforme à celui de Farber puisque y_{i2} et X_{i2} ne sont observées que si $y_{i1} = 1$. De plus, on ne peut supposer à priori que $\rho = 0$ puisqu'une absence d'indicateur dans un dossier a de bonnes chances de résulter en une absence de fraude dans celui-ci. Nous allons donc estimer les équations conjointement par la méthode du maximum de vraisemblance et ce pour les dossiers de fraude établie auquel nous ajouterons, comme au chapitre précédent, les dossiers suspects à 0%, 30%, 50% et 80% et plus. On note immédiatement que seule la deuxième équation est touchée par le changement de définition des cas de fraude. Également, contrairement à ce que nous avons fait au chapitre un, nous n'estimerons pas les équations pour les cas de fraude établie seulement. En effet, la complexité du modèle jumelée au peu d'observations contenant une fraude établie entraînent des problèmes de convergences. Finalement, comme les indicatrices des compagnies sont mutuellement exclusives, nous avons enlevé la compagnie possédant le plus de dossier au sondage. Celle-ci, la compagnie no.7, servira donc de référence quant aux résultats concernant les autres compagnies d'assurances. Puisque deux compagnies avaient déjà refusé de participer au sondage, 17 coefficients devront donc être estimés pour les compagnies restantes.

Il faut cependant noter que la spécification de notre modèle, bien que suivant l'idée générale de celui de Farber, entraîne un problème quant au caractère endogène des variables ind_{ik} de notre deuxième équation. Il existe en effet un lien entre la présence d'indicateurs de fraude, calculée selon les variables ind_{ik} de l'équation 2, et la variable endogène y_{i1} de l'équation 1 qui mesure si le dossier i possède ou non au moins un indicateur de fraude. Ainsi, le modèle de Probit bivarié avec observation conditionnelle de Farber employé ici se révèle plus simple que le véritable modèle auquel il faudrait recourir afin de solutionner ce problème. Il faudra donc prendre les prochains résultats avec une certaine réserve.

Les tableaux des pages suivantes représentent les résultats obtenus lorsque nous avons considéré les cas de fraude établie auxquels nous avons ajouté respectivement tous les dossiers suspects, les dossiers suspects à 30% et plus, ceux à 50% et plus et enfin ceux à 80% et plus. Comme les coefficients obtenus pour l'équation un ne varient pas entre les

différentes spécifications des cas de fraude, les résultats ne seront inscrits que dans le premier tableau.

Finalement, tous les résultats indiqués dans les tableaux des pages suivantes ont été obtenus grâce à la procédure Probit bivarié du logiciel Limdep.

Tableau 2.5.1 : Probit bivarié
Fraude suspecte et établie (E+S)

variables	coefficients	t	valeurs-p
Équation1			
constante	-0.38693	-5.402	0.0000
cie1	0.56179	3.769	0.0002
cie2	0.63196	4.366	0.0000
cie3	-0.14946	-1.267	0.2053
cie4	0.96051	6.915	0.0000
cie5	-0.22933	-2.288	0.0221
cie6	0.38253	2.046	0.0407
cie9	0.42875	2.102	0.0356
cie10	-1.2145	-3.879	0.0001
cie11	0.52780	2.939	0.0033
cie12	-0.08159	-0.866	0.3863
cie13	0.24981	2.262	0.0086
cie15	0.64212	3.971	0.0001
cie16	2.0469	4.971	0.0000
cie17	0.02273	0.165	0.8692
cie18	0.88599	4.029	0.0001
cie19	0.71220	6.268	0.0000
cie20	-0.53200	-4.145	0.0000
catb1	0.45024	3.760	0.0002
catb2	0.31270	5.176	0.0000
catb3	0.42432	5.690	0.0000
catb4	0.46527	1.877	0.0605
catav	0.53878	3.620	0.0003

Tableau 2.5.1 : Probit bivarié (suite)
Fraude suspecte et établie (E+S)

variables	coefficients	t	valeurs-p
Équation2			
constante	-2.4355	-8.780	0.0000
indicateur1	0.37292	2.037	0.0416
indicateur2	0.49201	1.975	0.0482
indicateur3	1.1983	5.581	0.0000
indicateur4	0.60971	1.670	0.0949
indicateur7	0.17725	0.511	0.6093
indicateur10	-0.3141	-0.912	0.3618
indicateur11	0.79597	3.489	0.0005
indicateur13	0.19839	0.729	0.4658
indicateur14	0.63762	1.212	0.2256
indicateur15	0.51969	1.871	0.0613
indicateur16	0.79223	2.624	0.0087
indicateur18	1.6490	5.203	0.0000
indicateur19	0.37736	1.617	0.1059
indicateur21	0.65248	1.809	0.0704
indicateur22	1.7275	4.908	0.0000
indicateur24	1.2695	2.633	0.0085
indicateur26	0.96141	2.587	0.0097
indicateur28	1.5816	1.791	0.0734
indicateur30	0.66397	3.079	0.0021
indicateur32	0.67155	3.110	0.0019
indicateur36	0.85535	0.915	0.3604
indicateur39	0.86509	2.761	0.0058
indicateur40	0.92643	1.666	0.0957
indicateur45	0.48938	1.552	0.1206
indicateur47	0.92050	3.753	0.0002

Tableau 2.5.1 : Probit bivarié (suite)
Fraude suspecte et établie (E+S)

variables	coefficients	t	valeurs-p
Equation2 (suite)			
indicateur48	0.19044	0.749	0.4541
nombre	-0.05528	-0.405	0.6857
catb1	0.83889	3.219	0.0013
catb2	0.00861	0.047	0.9628
catb3	0.43039	1.418	0.1562
catb4	0.13976	0.196	0.8450
catav	-0.0045	-0.010	0.9921
noncol	0.00149	0.006	0.9955
ρ	0.03744	0.143	0.8866
log-likelihood	-2018.0		

Tableau 2.5.2 : Probit bivarié
Fraude établie et suspecte à 30% et plus (E+S>2)

variables	coefficients	t	valeurs-p
constante	-2.6054	-8.451	0.0000
indicateur1	0.36583	1.890	0.0587
indicateur2	0.47723	1.771	0.0766
indicateur3	1.2247	5.495	0.0000
indicateur4	0.59930	1.627	0.1036
indicateur7	0.20579	0.575	0.5651
indicateur10	-0.2761	-0.801	0.4232
indicateur11	0.90809	3.829	0.0001
indicateur13	0.23528	0.842	0.3997
indicateur14	0.37551	0.738	0.4606
indicateur15	0.52979	1.739	0.0821
indicateur16	0.78002	2.288	0.0221
indicateur18	1.7027	5.237	0.0000
indicateur19	0.46739	1.956	0.0504
indicateur21	0.78538	2.147	0.0318
indicateur22	1.6449	4.717	0.0000
indicateur24	1.3667	2.762	0.0057
indicateur26	0.54430	1.005	0.3150
indicateur28	1.6147	1.794	0.0728
indicateur30	0.70615	3.088	0.0020
indicateur32	0.61700	2.785	0.0054
indicateur36	0.77047	0.802	0.4224
indicateur39	0.68468	2.001	0.0454
indicateur40	0.96270	1.721	0.0853
indicateur45	0.43865	1.390	0.1645
indicateur47	0.88264	3.363	0.0008
indicateur48	0.21449	0.823	0.4103

Tableau 2.5.2 : Probit bivarié (suite)
Fraude établie et suspecte à 30% et plus (E+S>2)

variables	coefficients	t	valeurs-p
nombre	-0.04557	-0.507	0.6121
catb1	1.0263	3.720	0.0002
catb2	0.14605	0.725	0.4685
catb3	0.64301	1.831	0.0671
catb4	0.37738	0.501	0.6165
catav	0.16222	0.339	0.7343
noncol	0.00719	0.022	0.9822
ρ	0.02757	0.103	0.9182
log-likelihood	-1993.0		

Tableau 2.5.3 : Probit bivarié
Fraude établie et suspecte à 50% et plus (E+S>4)

variables	coefficients	t	valeurs-p
constante	-2.6501	-6.690	0.0000
indicateur1	0.50956	2.367	0.0179
indicateur2	0.63056	2.288	0.0221
indicateur3	1.3241	5.666	0.0000
indicateur4	0.57178	1.223	0.2214
indicateur7	0.38483	1.001	0.3167
indicateur10	-0.2951	-0.734	0.4632
indicateur11	1.0737	4.429	0.0000
indicateur13	0.38856	1.319	0.1871
indicateur14	0.45522	0.898	0.3690
indicateur15	0.65387	2.078	0.0377
indicateur16	0.91949	2.576	0.0100
indicateur18	1.5916	4.764	0.0000
indicateur19	0.60949	2.329	0.0199
indicateur21	0.87159	2.364	0.0181
indicateur22	1.8826	5.373	0.0000
indicateur24	1.3591	2.402	0.0163
indicateur26	0.74724	1.284	0.1990
indicateur28	1.0075	2.034	0.0419
indicateur30	0.77243	3.200	0.0014
indicateur32	0.58838	2.312	0.0208
indicateur36	0.94782	0.952	0.3412
indicateur39	0.41824	0.900	0.3679
indicateur40	1.0786	1.875	0.0608
indicateur45	0.61569	1.921	0.0547
indicateur47	0.93096	3.487	0.0005
indicateur48	0.38339	1.356	0.1752

Tableau 2.5.3 : Probit bivarié (suite)
Fraude établie et suspecte à 50% et plus (E+S>4)

variables	coefficients	t	valeurs-p
nombre	-0.0868	-0.874	0.3823
catb1	0.66090	1.758	0.0788
catb2	0.06280	0.286	0.7750
catb3	0.50834	1.363	0.1728
catb4	0.37596	0.484	0.6281
catav	0.24066	0.502	0.6160
noncol	0.07569	0.220	0.8255
ρ	-0.0921	-0.304	0.7611
log-likelihood	-1958.0		

Tableau 2.5.4 : Probit bivarié
Fraude établie et suspecte à 80% et plus (E+S>7)

variables	coefficients	t	valeurs-p
constante	-2.9636	-4.248	0.0000
indicateur1	0.87068	2.221	0.0264
indicateur2	0.14664	0.305	0.7607
indicateur3	1.5074	3.900	0.0001
indicateur4	0.16922	0.240	0.8107
indicateur7	0.55441	0.883	0.3774
indicateur10	0.16551	0.368	0.7126
indicateur11	1.4953	3.564	0.0004
indicateur13	0.23158	0.498	0.6187
indicateur14	1.0157	1.717	0.0860
indicateur15	0.75102	1.584	0.1132
indicateur16	1.3237	2.619	0.0088
indicateur18	1.6537	2.873	0.0041
indicateur19	0.98682	2.450	0.0143
indicateur21	1.0602	2.450	0.0143
indicateur22	1.9611	3.979	0.0001
indicateur24	1.7204	1.768	0.0770
indicateur26	0.13385	0.146	0.8839
indicateur28	2.5595	2.519	0.0119
indicateur30	0.85807	2.151	0.0315
indicateur32	0.58962	1.463	0.1434
indicateur36	1.0925	0.777	0.4373
indicateur39	0.16859	0.212	0.8319
indicateur40	0.76146	0.892	0.3724
indicateur45	1.1436	2.819	0.0048
indicateur47	0.60792	1.363	0.1728
indicateur48	0.18662	0.296	0.7671

Tableau 2.5.4 : Probit bivarié (suite)
Fraude établie et suspecte à 80% et plus (E+S>7)

variables	coefficients	t	valeurs-p
nombre	-0.1973	-1.201	0.2296
catb1	0.90883	1.961	0.0499
catb2	-0.0293	-0.084	0.9327
catb3	0.63788	1.356	0.1753
catb4	0.92276	1.125	0.2605
catav	0.41956	0.586	0.5577
noncol	0.00383	0.010	0.9921
ρ	-0.2449	-0.637	0.5239
log-likelihood	-1873.0		

2.6 Analyse des résultats

On note tout d'abord, pour ce qui est de la première équation, qu'il existe une grande variabilité dans la façon dont les compagnies ont répondu au sondage. Comparativement à la compagnie no.7 qui est celle de référence, les compagnies 1, 2, 4, 9, 11, 13, 15, 16, 18 et 19 augmentent significativement la probabilité qu'au moins un indicateur soit inscrit au dossier. Par contre, toujours comparé à la compagnie ayant la plus grande part de marché au Québec, les compagnies 5, 10 et 20 diminuent cette probabilité. Ces différences ne peuvent s'expliquer par la part de marché détenue par les compagnies puisque les grandes comme les plus petites sont uniformément distribuées entre les deux catégories. On remarque finalement que les chapitres B1, B2, B3 et le groupe comprenant les avenants ont plus de chances de contenir un indicateur de fraude que les sinistres provenant du chapitre A.

Le plus important des résultats provenant de l'analyse du probit bivarié est cependant le fait que le terme de covariance entre les erreurs des deux équations (ρ) n'est pas différent de zéro. Ainsi, il n'existe pas vraiment de relation entre le fait d'avoir ou non un indicateur au dossier et la probabilité qu'une fraude soit commise, annulant donc la présence possible d'un biais de sélection dû à la non-réponse. On constate d'ailleurs que les estimés obtenus par la méthode du probit bivarié pour la deuxième équation sont similaires à ceux trouvés au chapitre un lorsque l'on ne se préoccupait pas du biais de sélection, et à ceux trouvés par Belhadji et Dionne et leur méthode de classification des dossiers selon les enquêteurs. Les résultats obtenus au chapitre un sont donc valides et les indicateurs trouvés augmentent bel et bien la probabilité qu'une fraude soit commise. On doit cependant prendre ces résultats avec quelques réserves puisque les variables exogènes ind_{ik} de l'équation 2 possèdent un certain lien avec la variable endogène y_{i1} de l'équation 1.

2.7 Comparaison entre la méthode du probit bivarié et celle du rapport Belhadji-Dionne pour résoudre le biais de sélection.

Une première méthode d'estimation tenant compte de la présence possible d'un biais de sélection dans l'équation étudiée a été mise en place par Belhadji et Dionne (1997). Ils ont tenté de diminuer le biais de sélection en faisant une analyse rétrospective des données récoltées lors du précédent sondage. La méthode consiste à vérifier pour chaque enquêteur s'il a ou non inclus des indicateurs de fraude dans les dossiers analysés. Si un enquêteur a inscrit un indicateur de fraude dans au moins un des dossiers ne contenant pas de fraude qu'il a remplis, Belhadji et Dionne considéraient alors que tous les dossiers provenant de cet enquêteur avaient été remplis adéquatement et les dossiers ne contenant pas d'indicateurs de fraude étaient classés comme ne possédant pas d'indicateurs de fraude. À l'opposé, si un enquêteur n'avait jamais inscrit d'indicateur de fraude dans ses dossiers sans fraude, Belhadji et Dionne considéraient alors tous ces dossiers comme des cas de non-réponse quant à la question concernant les indicateurs de fraude. Cette procédure, si elle a le mérite d'écarter un possible biais des réponses associé au fait que des enquêteurs n'ont pas répondu à la question sur les indicateurs de fraude, est cependant radicale puisqu'elle peut avoir éliminé des dossiers qui ne contenaient pas d'indicateurs, alors que l'enquêteur était disposé à répondre à la question.

Ce processus d'élimination a conduit à la perte de 441 observations. Un modèle probit a été ajusté aux observations restantes. La variable dépendante est la présence établie ou suspecte de fraude dans le dossier alors que des variables dichotomiques ont été insérées pour mesurer la présence ou non des différents indicateurs de fraude. Il est à noter que Belhadji et Dionne ont quelque fois différencié le rang de l'indicateur de fraude. Ainsi, certains indicateurs ont été séparés sachant qu'ils étaient jugés les plus importants selon l'enquêteur (indicateur suivi du chiffre 1) ou non (indicateur suivi du chiffre 2). Les résultats de cette régression sont présentés au tableau 2.7.1 de la page suivante. Seuls les indicateurs ayant des résultats significatifs ont été insérés dans la régression.

Pour pouvoir comparer adéquatement la méthode Belhadji-Dionne avec la notre, nous avons répliqué parfaitement la spécification de leur modèle (même définition de la variable dépendante mesurant s'il y a eu fraude ou non et mêmes variables explicatives) tout en évaluant les coefficients des indicateurs à la lumière du résultat trouvé à la section précédente, soit que le biais de sélection ne joue aucun rôle dans l'estimation des coefficients des indicateurs de fraude. Ces résultats sont quant à eux présentés au tableau 2.7.2.

**Tableau 2.7.1 : Résultats du modèle Belhadji-Dionne
avec le processus d'élimination**

variables	coefficients	t
constante	2.53842	23.87243
Indicateur1	0.57129	3.95309
Indicateur2	0.65360	2.88309
Indicateur3(1)	1.80978	7.86276
Indicateur4	1.26659	4.60894
Indicateur11(1)	0.89913	2.34950
Indicateur11(2)	0.83390	2.60732
Indicateur14	1.13099	2.78572
Indicateur16(1)	1.18581	3.03581
Indicateur16(2)	0.90678	3.19590
Indicateur18	1.58303	4.75648
Indicateur19	0.63276	3.23234
Indicateur21	0.72584	1.89837
Indicateur22	1.64896	4.79668
Indicateur26	1.24422	3.68028
Indicateur30(1)	0.95626	4.24095
Indicateur30(2)	1.03962	3.69397
Indicateur32	0.69536	4.19627
Indicateur39	0.81779	2.83649
Indicateur45	0.93629	3.10616
Indicateur47	1.04607	4.49000
Noncol	0.44192	3.14714

**Tableau 2.7.2 : Résultats du modèle Belhadji-Dionne
sans le processus d'élimination**

variables	coefficients	t
constante	-2.2844	-17.881
Indicateur1	0.32582	2.393
Indicateur2	0.46727	1.985
Indicateur3(1)	1.4878	6.100
Indicateur4	1.68952	2.432
Indicateur11(1)	0.79473	2.489
Indicateur11(2)	0.89490	2.975
Indicateur14	0.74445	1.993
Indicateur16(1)	0.35352	0.713
Indicateur16(2)	0.90707	3.259
Indicateur18	1.6913	5.838
Indicateur19	0.46961	2.463
Indicateur21	0.52925	1.565
Indicateur22	1.6480	4.999
Indicateur26	0.78453	2.290
Indicateur30(1)	0.47419	2.001
Indicateur30(2)	0.94150	3.706
Indicateur32	0.63432	3.822
Indicateur39	0.81310	2.926
Indicateur45	0.65685	2.270
Indicateur47	1.0389	4.831
Noncol	0.37431	2.618

2.8 Analyse des différences entre la méthode avec processus d'élimination de Belhadji-Dionne et celle du probit bivarié

On constate qu'il y a de légères différences entre les deux méthodes. Tout d'abord, deux indicateurs n'influencent plus la probabilité qu'une fraude soit commise. En effet, le fait que l'assuré est extraordinairement familier avec le jargon des assurances ou des réparations de véhicules (indicateur 16) alors que cet indicateur est considéré comme le plus important selon l'enquêteur et également les nombreux reçus de taxis ou de factures de location de véhicule (indicateur 21) ne sont plus significativement différents de zéro. On peut alors présager que la méthode de Belhadji-Dionne avec processus d'élimination a elle-même introduit un léger biais de sélection dans les estimations des coefficients des indicateurs de fraude. On constate également que les valeurs des coefficients trouvés pour les indicateurs toujours différents de zéro sont habituellement plus bas dans le modèle du probit bivarié. Ainsi, la méthode d'élimination de Belhadji-Dionne serait trop restrictive et aurait retiré de la banque de données des dossiers ne possédant pas d'indicateurs de fraude en les considérant à tort comme possédant de la non-réponse.

Conclusion

La fraude à l'assurance étant un problème de plus en plus préoccupant pour les compagnies d'assurances, il est intéressant de chercher les caractéristiques qui ressortent le plus souvent des dossiers frauduleux. Les compagnies pourront alors, selon la présence ou l'absence de ces caractéristiques, décider de pousser ou non l'enquête d'un dossier suspecté de contenir de la fraude.

Nous avons montré, en nous servant de données recueillies lors d'un sondage auprès des vingt compagnies d'assurances ayant les plus grosses parts de marché au Québec, que certains indicateurs de fraude entraînaient bel et bien par leur présence une augmentation de la probabilité qu'un dossier soit frauduleux. En étant conservateur et en supposant que seuls les dossiers où la fraude a été établie ou encore les dossiers avec un doute très élevé de contenir de la fraude (80% et plus) sont frauduleux, nous trouvons quatorze indicateurs augmentant la probabilité de fraude. Ces indicateurs sont : l'absence de rapport de police lorsque celui-ci était nécessaire, l'existence de tout dommage non relié à la perte ou incohérent avec les faits de l'accident, l'assuré qui éprouve des difficultés financières personnelles ou reliées à ses affaires, celui qui est prêt à accepter un montant de règlement relativement petit plutôt que d'avancer tous les documents reliés à la perte, l'assuré qui est très insistant pour un règlement de la perte, celui qui est extraordinairement familier avec le jargon des assurances ou des réparations de véhicule, l'assuré qui est trop enthousiaste ou trop franc à prendre le blâme de l'accident, l'accident qui a eu lieu peu de temps après l'enregistrement et l'achat d'assurance pour le véhicule ou encore dans le mois qui précède la fin de la police, l'existence de nombreux reçus de taxi ou de voitures de location, des factures qui semblent fausses ou fabriquées, lorsque l'historique du véhicule ne peut être établie, lorsque le représentant des évaluations se voit refusé l'examen du véhicule, des témoignages contradictoires concernant les circonstances de la perte sont évoqués et enfin, lorsque le réclamant est très agressif et lance diverses menaces de recours (avocat, gouvernement, ...).

En étant moins restreint et en insérant les dossiers suspects à 50% et plus dans nos dossiers frauduleux, s'ajoutent quatre nouveaux indicateurs de fraude. Ce sont : une collision mineure a entraîné des coûts de réparation excessifs, lorsqu'un accident n'implique qu'un seul véhicule, si l'assuré est au chômage ou vit dans une région défavorisée et finalement lorsque l'assuré est nerveux ou semble confus lors de l'enquête.

Dans une dernière étape, si l'on considère que la fraude à l'assurance est difficilement perceptible par les enquêteurs et que celle-ci est en réalité plus fréquente que lors du sondage, nous pouvons être tenté de considéré tous les dossiers suspects comme contenant de la fraude. Ce faisant, deux derniers indicateurs s'ajoutent aux listes précédantes, soient lorsque la documentation de l'estimation et de la réparation n'est pas disponible et lorsque l'achat du véhicule s'est fait au comptant.

Nous avons également démontré que le fait de ne pas savoir la cause de l'absence d'indicateur de fraude dans un dossier (qui peut être due à la non réponse de l'employé de la compagnie d'assurances ou à l'absence réelle d'indicateur dans la réclamation) n'entraînait pas de biais de sélection de nos estimés. Nous devons être cependant très prudent sur ce résultat puisqu'il existe un lien entre la variable endogène de notre première équation du modèle probit bivarié et les variables exogènes représentant les indicateurs de fraude dans notre seconde équation.

Pour conclure, nous avons trouvé des différences entre les coefficients trouvés lors de cette recherche et ceux de l'étude Belhadji-Dionne (1997). Les auteurs, afin d'enrayer le biais de sélection potentiel sur la question des indicateurs de fraude, ont mis en place un processus d'élimination des observations étant susceptible de contenir de la non-réponse. Comme la méthode du probit bivarié avec observation conditionnelle n'a pas décelé de biais de sélection, on peut penser que les différences trouvés entre les deux méthodes sont causés par un possible biais de sélection introduit par le processus d'élimination instauré par Belhadji-Dionne. On doit cependant être très prudent quant à cette affirmation puisqu'aucune technique rigoureuse n'a été employée afin de vérifier que les observations restantes de Belhadji-Dionne avaient effectivement introduit un biais.

Bibliographie

- Belhadji E., Dionne G., Développement d'un système-expert de détection automatique de la fraude à l'assurance automobile, Cahier 97-04, Chaire de gestion des risques, HEC-Montréal..
- Boyes W.J, Hoffman D.L, Low S.A., An econometric Analysis of the Bank Credit Scoring Problem, *Journal of Econometrics*, 40 (1989), p. 3-14.
- Bruneau C., Sélection endogène, Notes de cours, Chapitre 6.
- Dionne G., Artis M., Guillén M., Count Data Models for a Credit Scoring System, *Journal of Empirical Finances*, 1996
- Dionne G., Belhadji E., Évaluation de la fraude à l'assurance automobile du Québec, *Assurances*, Octobre 1996.
- Farber H.S., Worker Preference for Union Representation, *Research in Labor Economics*, Supplement 2 (1983), p. 171-205.
- Greene W.H., *Limdep User's Guide*, Econometric Software, 1992.
- Heckman J.J., Sample selection bias as a specification error, *Econometrica*, 47 (janvier 1979), p. 153-162.
- Meng C., Schmidt P., On the cost of partial observability in the bivariate Probit Model, *International Economic Review*, 26 (février 1985), p. 71-85.
- Poirier D.J., Partial Observability in Bivariate Probit Models, *Journal of Econometrics*, 12 (février 1980), p. 210-217.

Zellner A., Lee J.H., Joint Estimation of Relationships Involving Discrete Random Variables, *Econometrica*, 33 (mai 1965), p. 382-394.

ANNEXE A

Indicateurs de fraude

Principaux indicateurs

Accident / Dommage

- 1) Pas de rapport de police quand celui-ci était nécessaire (ou 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 entretien du véhicules ne peuvent être fournies par l'assuré.
- 7) Le véhicule n'a pas d'intérêt pour le voleur (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'assurances n'a jamais vu le véhicule assuré.

Financiers

- 11) L'assuré éprouve des difficultés financières personnelles ou liées 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 exagérément 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.

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) Antécédants 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 rapporté 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 l'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 signe d'entrée par effraction.

Financiers

- 38) Remboursement 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 sur assurée.
- 43) Prime payée en personne et au comptant.
- 44) Problème avec l'adresse: C.P; motel; fausse adresse; assuré absent; vit chez des amis...
- 45) Réclamant est très agressif.
- 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.