

Université de Montréal

La fraude à l'assurance

Par

Louis Caron

Rapport de recherche présenté au
département de sciences économiques
en vue de l'obtention du grade de
maître ès sciences (M.Sc.)
en sciences économiques

Août 1996

Remerciements

Je tiens à remercier mon directeur de recherche, M. Georges Dionne, pour les données qui ont servies à la présente recherche ainsi que pour son excellent support pédagogique. Les données ont été obtenues d'un projet sur la fraude à l'assurance financé par le Bureau d'Assurance du Canada et la Fondation Conrad Leblanc.

J'aimerais aussi remercier M. Michel Dubois et M. Marc Parisien de la compagnie d'assurance Standard Life pour leur support technique ainsi que le Centre de Recherche sur les Transports pour son support financier.

Index

Résumé	3
Introduction	4
Théorie	6
1) Introduction	6
2) Townsend	9
Introduction	9
Hypothèses	9
Modèle	12
Conclusion	13
3) Mookherjee et Png	14
Introduction	14
Hypothèses	15
Modèle	16
Conclusion	18
3) Crocker et Tennyson	20
Introduction	20
Hypothèses	20
Modèle	22
Conclusion	23
4) Discussion et conclusion	25
Problématique	28
1) Description:	28
2) Illustration et historique:	29
Modèle Statistique	31
Méthodologie	36
1) Les données:	36
2) Traitement des données	37
Section 1) Estimation Unique:	38
1 ^{ère} approche:	38
2 ^{ème} approche:	40
3 ^{ème} approche:	41
4 ^{ème} approche:	43
Section 2) Itérations:	44
Section 3) Résultats	45
Section 4) Les Groupes:	48
Section 5) Evolution:	50
Conclusion	52
Bibliographie	56
Théorique	56
Modèle statistique	57

Résumé

Ce travail a pour but premier d'estimer le pourcentage de fraude à l'assurance automobile au Québec. Il fait suite à l'étude de Dionne et Belhadji¹ et se sert de la même banque de données. On trouve dans le présent ouvrage environ 10% de fraudes soit trois fois plus de fraudes que le niveau plancher calculé dans Dionne-Belhadji. Le travail présente aussi les modèles théoriques sous-jacents au problème de risque moral rencontré, ainsi que le modèle statistique pour l'estimation du phénomène caché qu'est la fraude à l'assurance.

Abstract

This paper's first goal is to estimate fraud in automobile insurance in Quebec. It follows a previous study from Dionne and Belhadji and uses the same data bank. The present study finds a fraud rate of approximately 10% which is three times more than the floor level calculated in Dionne-Belhadji. This Study presents the theoretical models related to the moral hazard problem encountered and the statistical model for the estimation of the hidden phenomenon which is insurance fraud.

¹Dionne G. et B. Belhadji (1996), Evaluation de la fraude à l'assurance automobile au Québec, CRT-96-08

Introduction

A la base, la première étape est de définir la fraude à l'assurance pour le présent travail. La fraude à l'assurance au sens de ce texte sera alors: "Un acte volontaire pour profiter d'un contrat d'assurance. Cet acte suppose la mauvaise foi de l'assuré. Il n'y a donc pas de fraude si l'assuré a agi par erreur ou dans la croyance légitime qu'il ne commettait aucune irrégularité".²

Il est intéressant de noter que cette définition de fraude est conforme à la notion de "faute intentionnelle" du Code Civil. (C.c. 2563, al. 2)

Il peut paraître étrange à première vue de s'attarder sur ce point, mais on peut constater dans la littérature que les auteurs ne définissent pas tous la fraude à l'assurance de la même manière.

Pour illustrer ce point un exemple peut se révéler utile:
Le vol d'automobiles assurées :

.Certains incluent le vol d'une automobile assurée comme étant d'emblée de la fraude à l'assurance.

.Comme le véhicule était assuré, l'agent qui se fait voler est effectivement la compagnie d'assurance.

²Dionne, Gibbens, St-Michel; Conséquences de la fraude à l'assurance; Colloque, Université de Montréal.

.Face au vol d'automobiles, on pourrait objecter que celui-ci fait partie des risques assumés par la compagnie, qu'elle en tient compte dans ses probabilités de sinistres et ainsi maintient les réserves actuarielles adéquates pour cette éventualité.

.Mais voici, la même objection tient aussi pour la fraude à l'assurance au sens décrit plus haut!

.En effet, si, par exemple, 10% des vols de voitures sont effectués avec le consentement de (ou par) l'assuré, la probabilité du sinistre "vol d'auto" que prend l'assureur, par le biais des études d'expériences, se calcule avec 100% des vols (y compris la fraude).

.La prime chargée au client est donc calculée en tenant compte de la possibilité pour le client d'être un fraudeur. La fraude à l'assurance est donc aussi un risque assumé par la compagnie qui maintient des réserves actuarielles adéquates pour y faire face. Le problème considéré n'en est pas un de vol.

.Le vol d'automobiles assurées et la fraude à l'assurance au sens de ce texte ont tous deux pour victimes les assurés par le biais des primes.

.La différence est que la fraude à l'assurance présuppose un manquement à un contrat par une des deux parties liées à ce contrat. Ce manquement dénature le risque assuré et met en danger le principe même de l'assurance.

La prochaine section traite de la littérature sur le problème de la fraude et définit le modèle théorique le plus approprié pour la présente étude. Les sections subséquentes traitent du problème d'estimation.

Théorie

1) Introduction

On doit premièrement s'interroger sur la nature de la fraude qu'on considère dans l'industrie de l'assurance automobile, si l'on veut ensuite en déduire le modèle qui colle le mieux à la réalité.

L'étude de Dionne et Belhadji s'est penché sur la question et ce que les auteurs ont trouvé est présenté au tableau suivant.

Nature de la Fraude	%
Fausses déclarations sur souscription	6
Sinistre intentionnel	18
Faux-sinistre	16
Fausses déclarations sur réclamations	21
Gonflement pour obtenir plus de compensations	32
Gonflements pour récupérer la franchise	7
TOTAL	100

Comme on peut le constater dans le tableau 76% des fraudes (les quatre dernières rangées), proviennent clairement de l'information imparfaite de l'assureur en ce qui a trait au résultat de l'aléa.

La catégorie "Sinistre intentionnel" est cependant plus difficile à cataloguer. Cette catégorie n'entre pas dans les problèmes de risque moral traditionnels, c'est-à-dire où l'agent a des comportements qui influent sur les paramètres de la distribution probabilistique de l'aléa. Dans le cas du sinistre intentionnel l'agent supprime l'aléa. Ce qui nous ramène à un problème de vérification pour identifier la fraude et porte le pourcentage du problème de cette nature à 94%.

La catégorie "Fausses déclarations sur souscription" (6%) est cependant reliée à un problème tout autre, c'est à dire la sélection adverse.

On peut donc considérer le problème de la fraude dans notre échantillon comme un problème sur le contrat entre un principal (l'assureur) et un agent (l'assuré). L'asymétrie d'information est à l'avantage de l'agent et concerne, du moins à 94%, le résultat, ex-post, de l'aléa assuré.

L'assuré, ayant plus d'informations sur l'aléa que l'assureur, peut dissimuler cette information et déclarer un sinistre réellement survenu mais d'une envergure supérieure en terme de perte monétaire. L'assuré peut aussi déclarer un sinistre qui n'a jamais eu lieu. La difficulté pour l'assureur est de vérifier les dires de l'assuré.

Il est à noter que la nature ex-post à l'aléa du risque moral de la présente étude distingue bien la présente littérature de celle du problème de risque moral classique. La nature aléatoire du sinistre distingue la présente littérature des travaux de Spence et Zeckhauser(1971), Shavell (1979) et Harris et Raviv (1978) où cette nature aléatoire n'est pas exogène mais dépend de certains comportements de la part de l'assuré.

L'assureur ne peut vérifier les dires de l'assuré qu'au moyen d'un processus de vérification coûteux (dans les premiers articles présentés). Les hypothèses quant à ce processus vérificationnel varient dans la littérature théorique. Par exemple la vérification peut être systématique (Townsend 1979 sect. 2) ou aléatoire (Mookherjee et Png 1989).

Dans la réalité, les assureurs se servent d'une batterie "d'indicateurs de fraude" pour choisir les dossiers à investiguer. L'indicateur le plus important pour les assureurs est présenté dans le tableau 7a de l'étude Dionne et Belhadji et est reproduit dans le tableau qui suit.

Indicateur le plus important (5% et + d'importance)	
Indicateur	%
Pas de rapport de police	26
Agent n'a jamais vu véhicule	8
Témoignages contradictoires (accident)	6
Beaucoup de réclamations dans le passé	6
Accident impliquant un seul véhicule	5

Sur cette base unique, la vérification est-elle systématique ou aléatoire? Quel est le modèle qui représente le mieux la réalité? Telles sont les questions qui occupent la présente partie du travail.

Aux prochaines sections sont présentés les principaux modèles théoriques, ensuite la conclusion de la présente partie tentera de choisir le modèle à adopter et de justifier cette adoption.

2) Townsend¹

Introduction

Le but de l'article de Townsend était d'expliquer, par le biais d'un problème de risque moral ex-post, la présence de déductibles dans l'industrie de l'assurance. En effet, en information parfaite le contrat d'allocation optimale est un transfert de risque simple entre le principal et l'agent.

Nous allons procéder en adoptant la notation de Townsend. Dans son article il définit deux agents. L'agent 1 peut être vu comme le principal ou l'assureur, et l'agent 2 comme l'agent ou le demandeur d'assurance. Le problème d'information est défini de sorte que s'il n'y a pas de vérification, le résultat de l'aléa est connu uniquement de l'agent 2.

Hypothèses

Le modèle présenté par Townsend est un modèle où la vérification est parfaite. C'est à dire un modèle où toutes les fraudes sont détectées dans les dossiers frauduleux lorsque le processus vérificationnel est appliqué.

Dans le texte, on définira un contrat comme étant *cohérent* si l'agent 2 se soumet volontairement à la vérification. Le contrat présenté en conclusion pour ses caractéristiques devra donc être cohérent et Pareto optimal.

¹Townsend R. (1979), Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory*

Les fonctions d'utilité des deux agents sont définies von Neumann-Morgenstern, strictement croissantes, différentiables et concaves. De plus U_2 est défini tel que: $U_2'(0) = \infty$, et $U_2'(\infty) = 0$. La dotation de l'agent 2 est aléatoire et notée $y_2 \in [\alpha, \beta]$ où $\alpha > 0$ et possède une fonction de densité $f(y_2)$ strictement positive. La dotation pour l'agent 1 n'est pas aléatoire et est noté y_1 .

Soit $c_j(y_2)$ la consommation de l'agent j . Cette consommation est présentée comme étant fonction de y_2 . Nous sommes alors en mesure d'identifier une contrainte de possibilité de la réalisation, c.à d. $c_1(y_2) + c_2(y_2) \leq y_1 + y_2$.

Chaque agent veut maximiser son espérance d'utilité:

$$\int U_j[c_j(y_2)] dF(y_2)$$

L'agent 2 doit être vu ici comme un individu qui est confronté à un perte aléatoire $\beta - y_2$ et qui veut acheter de l'assurance à l'agent 1.

On pose $g(y_2)$ le transfert du bien de consommation de l'agent 2 vers l'agent 1, et $g!$ le choix contractuel (avant l'aléa) de la fonction g .

Avant la réalisation de y_2 les agents décident s'il y aura ou non vérification, dépendemment de y_2 . La région S (avec complément S') délimite l'ensemble des réalisations de y_2 où il y aura vérification. Soit $S!$ et $S!'$, les choix contractuels, avant état de la nature de S et S' respectivement. Donc, clairement, s'il y a vérification (que $y_2 \in S$) alors le montant transféré est celui qui a été décidé avant l'aléa, c'est à dire que $g(y_2) = g!(y_2)$.

S'il n'y a pas de vérification alors l'agent 2 voudra transférer le plus petit montant possible ($\min_{x \in S!} g!(x)$), ou en assurance, déclarer la perte la plus lourde.

Le coût de la vérification peut être soit fixe à $u > 0$, soit dépendant de y_2 avec la fonction $\xi[g!(y_2)]$.

A l'aide de cette notation, un contrat est défini cohérent, au sens décrit plus haut si:

$$(i) \quad S = S!$$

$$(ii) \quad g(y_2) = g!(y_2) \quad \forall y_2 \in [\alpha, \beta]$$

Il est important de noter que sous l'hypothèse du contrat cohérent l'agent 2 n'a aucun intérêt à mentir, qu'il y ait vérification ou non. L'hypothèse de contrat cohérent introduit deux lemmes.

Lemme 2.1: Un contrat $[g!, S!]$ est cohérent si et seulement si, $g!(y_2)$ égale une certaine constante $C!$ sur $S!$ et $g!(y_2) + \xi[g!(y_2)] < C!$ sur $S!$.

Autrement dit, le transfert de l'agent 2 à l'agent 1 ne peut dépendre d'informations connues de l'agent 2 uniquement. De plus, l'agent 2 doit être motivé à la vérification, donc le transfert doit être supérieur au coût de vérification.

Lemme 2.2: \forall contrat $[g!, S!]$ \exists un contrat cohérent $[h!, T!]$ qui donne la même allocation des ressources.

La façon de voir ceci pour le problème d'assurance est que l'agent 2 s'assure contre l'aléa $\beta - y_2$ en achetant le contrat d'assurance $[g!, S!]$ de l'agent 1. $C!$ est la prime, indépendante de la perte $\beta - y_2$ et $I!(y_2) = C! - g!(y_2)$ est le bénéfice d'assurance pour la perte $\beta - y_2$ si une réclamation est faite, en quel cas il y a toujours vérification.

Donc si $y_2 \in S'$ alors $I(y_2) = 0$. De plus, si $y_2 \in S$ alors la cohérence requiert $g(y_2) + \xi[g(y_2)] < C$ pour que $I(y_2) - \xi[g(y_2)] > 0$.

Modèle

Le problème revient à maximiser l'utilité sous trois contraintes. La première étant que l'agent 1 se trouve au moins aussi bien qu'en autarcie. La seconde, que le contrat soit cohérent, et la dernière est une contrainte de non-négativité de la consommation de l'agent 1.

Le résultat important à ce problème est que la zone de vérification se trouve dans un intervalle $[\alpha, \gamma]$ où $\gamma \leq \beta$.

Suivant la notation de Townsend on a le paiement d'assurance $I(y_2) = C - g(y_2)$ et on sait, d'une part que $I = 0$ en S' et d'autre part que $I > 0$ en S .

On pose le coût de vérification $\xi[g(y_2)] = \psi[I(y_2)]$ dépendant donc du paiement d'assurance, avec $\psi(0) = 0$, $\psi'(0) < 1$, et on pose de plus la fonction $\psi(\cdot)$ continue différentiable et convexe.

Le problème de maximisation devient alors:

$$\max_{I, C} \int_{\alpha}^{\beta} U_2 \{ Y_2 + I(Y_2) - C - \psi[I(Y_2)] \} dF(Y_2)$$

$$\begin{aligned} \text{S.A. (1)} \quad & \int U_1 [Y_1 - I(Y_2) + C] dF(Y_2) \geq K \\ (2) \quad & I(Y_2) \geq 0 \\ (3) \quad & Y_1 - I(Y_2) + C \geq 0 \end{aligned}$$

La résolution de ce système nous amène aux conclusions du modèle à la section suivante.

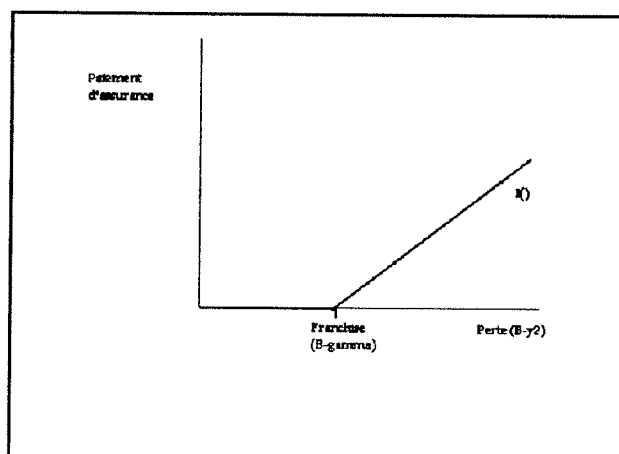
Conclusion

Proposition 3.1: Toute solution I^* , C^* au problème précédent a la propriété que $S^* = \{Y_2 : Y_2 < \gamma\}$ pour un certain paramètre γ .

A l'aide d'un exemple Townsend montre que la région de vérification peut dépendre du coût de vérification de façon non-triviale. C'est à dire que S^* n'est pas un ensemble vide, ni qu'il couvre tout le domaine.

Le résultat qui nous occupe est que ce contrat n'est optimal au sens de Pareto que si l'assuré effectue une réclamation seulement dans les cas où il subit une perte $(\beta - Y_2)$ plus grande que $(\beta - \gamma)$. Donc on peut voir $(\beta - \gamma)$ comme le déductible du contrat.

Ce contrat d'assurance est représenté à la figure ci-contre. La perte, en abscisse, doit atteindre un certain seuil, c'est-à-dire le déductible $(\beta - \gamma)$, pour que le bénéfice d'assurance, en ordonnée, soit différent de zéro. Ensuite, le bénéfice d'assurance suit la perte avec une pente de quarante-cinq degrés.



3) Mookherjee et Png²

Introduction

L'objectif de cet article est d'étudier le problème soulevé par Townsend, dans l'article précédent au sujet de l'hypothèse de vérification. Celui-ci a noté que l'optimalité des contrats qu'il trouvait, dépendait de façon cruciale de l'hypothèse de vérification déterministique des contrats. Aussi, laissait-il ouverte la question du caractère général des contrats optimaux lorsque les vérifications aléatoires sont permises.

L'objectif de cet article est donc de réouvrir cette question.

L'article de Mookherjee et Png introduit aussi le problème de risque moral classique, en supplément au problème de risque moral ex-ante considéré. En d'autres termes, dans cet article, l'agent, en plus d'être le seul à connaître le résultat de l'aléa, s'il n'y a aucune vérification, influe aussi sur les paramètres de la distribution probabilistique de l'aléa par ses comportements.

Nous sommes, dans ce cadre, à la recherche de la forme du contrat optimal.

²Mookherjee D and I. Png (1989), Optimal Auditing Insurance and redistribution, *Quarterly Journal of Economics*

Hypothèses

Soient Y_1, Y_2, \dots, Y_n les niveaux possibles de revenus, avec $Y_{i-1} < Y_i \forall i$. L'agent doit choisir un comportement dans l'ensemble A des comportements possibles. Le comportement choisi détermine la distribution probabilistique des niveaux de revenus: $\lambda_i(a)$ donne la probabilité du revenu Y_i résultante du comportement a , où $\lambda_i(a) \in [0, 1]$ et $\sum_{i=1}^n \lambda_i(a) = 1$.

Le principal est supposé neutre au risque. L'agent est riscophobe avec une fonction d'utilité W définie von Neumann-Morgenstern.

On pose de plus que W est additivement séparable en consommation et en action: $W = U(C) - G(a)$. U est strictement croissante et concave. De plus, la consommation de l'agent est bornée inférieurement à zéro.

Cette dernière hypothèse borne la consommation inférieurement pour empêcher un résultat trivial de quasi-"first-best" d'être réalisable. En effet un tel résultat peut être obtenu en imposant de larges pénalités si l'agent ne déclare pas la vérité ou s'il choisit un comportement indésirable.

De façon similaire, on doit empêcher que le principal n'encourage la révélation de la vérité en vérifiant avec probabilité très faible, mais en accordant de très grandes récompenses aux individus vérifiés qui ont déclaré la vérité. Ainsi on pose $\lim_{c \rightarrow \infty} [U(C)/C] = 0$.

On note le coût de la vérification E_i , et celui-ci peut dépendre de Y_i .

Soit un agent avec un revenu réalisé Y_i . Si cet agent révèle Y_i , il doit alors payer un transfert T_i au principal. Le principal, quant à lui, vérifie les dires de l'agent avec probabilité p_i . S'il y a vérification, Y_i va être découvert, avec certitude (hypothèse de vérification parfaite), comme étant le revenu réel de l'agent. Celui-ci se verra dans l'obligation de payer la pénalité F_{ii} .

Les auteurs ne définissent pas un contrat cohérent, comme l'a fait Townsend, mais spécifient plutôt qu'à l'équilibre l'agent révélera toujours la vérité. Ils stipulent donc qu'on peut favoriser cette révélation, sans affecter l'utilité de l'agent, en imposant la pénalité pour une fausse déclaration (F_{ii}) comme étant la plus grande possible. Dans un tel cas, la consommation de l'agent est portée à zéro. C'est à dire $T_i - F_{ii} = 0$. Pour simplifier, la notation F_{ii} devient simplement F_i .

Modèle

Le problème du principal est alors de trouver le contrat $(\{p_i\}, \{T_i\}, \{F_i\})$ avec un choix de comportement tel qu'il résout la maximisation suivante:

$$\sum_{i=1}^n \lambda_i(a) [p_i U(Y_i - T_i - F_i) + (1 - p_i) U(Y_i - T_i)] - G(a)$$

S.A.

$$(1) \quad p_i U(Y_i - T_i - F_i) + (1 - p_i) U(Y_i - T_i) \geq (1 - p_h) U(Y_i - T_h) \quad \forall i, h$$

$$(2) \quad \sum \lambda_i(a) V_i - G(a) \geq \sum \lambda_i(a') V_i - G(a') \quad \forall a' \in A$$

$$(3) \quad \sum \lambda_i(a) [T_i - p_i (E_i - F_i)] \geq R$$

$$(4) \quad Y_i - T_i - F_i \geq 0 ; \quad Y_i - T_i \geq 0 ; \quad p_i \in [0, 1] \quad \forall i.$$

La contrainte (1) représente la contrainte de révélation, obligeant l'agent à dire la vérité. La contrainte (2) représente la contrainte d'effort, celle-ci impose le comportement (a), celui qui a été choisi par le principal. La contrainte (3) requiert que l'espérance de profit du principal rencontre un certain niveau R. Les dernières contraintes (4) stipulent que la consommation ne peut devenir négative et que les probabilités doivent être comprises en $[0,1]$.

Les auteurs présentent, et montrent, ensuite, les deux lemmes suivants:

Lemme 1: Tout contrat ne respectant pas la contrainte de profit, n'est pas optimal.

Lemme 2: Sans perte de généralité, le principal ne doit pas pénaliser l'agent si la vérification révèle que celui-ci a dit la vérité: que $F_i \leq 0 \quad \forall i$.

Conclusion

La résolution du système décrit plus haut, amène les auteurs aux deux propositions suivantes:

- Proposition 1:
- (a) Le plan optimal existe
 - (b) Pour un tel plan, toutes les vérifications doivent être faites de façon aléatoire. Autrement dit: $p_i < 1 \quad \forall i$.
 - (c) Pour un plan optimal, si la vérification démontre que l'agent n'a pas menti, alors le principal doit le récompenser. En d'autres termes $F_i < 0, \quad \forall p_i \in [0,1]$.

- Proposition 2: Pour tout plan optimal:
- (a) La révélation correspondant au plus grand des transferts possible ne sera pas vérifiée mais toutes les autres révélations le seront avec une probabilité positive.
 - (b) Les révélations correspondant aux plus grands transferts seront vérifiées avec une probabilité égale ou inférieure aux autres.

Un fait à noter est que ce modèle prédit que si le contrat est optimal, il y aura toujours vérification aléatoire. Or la thèse de vérification aléatoire suppose que pour un certain pourcentage de réclamation, l'assureur paie sans vérifier quoi que ce soit. En réalité, l'assureur vérifie plutôt de façon systématique tous les dossiers pour, au moins, trouver les indicateurs de fraude.

Cet étude introduit le problème de risque moral classique au modèle. Ceci généralise le modèle, car nous n'avons pas, bien entendu, que le problème de risque moral ex-post à considérer pour établir le contrat optimal. D'autres problèmes tel que le risque moral classique, sont à considérer.

Il est toutefois à noter que cette généralisation du modèle ne s'est effectué qu'au prix d'une perte de généralité en ce qui concerne la fonction d'utilité. Celle-ci est, dans la présente étude, additivement séparable en consommation et en action.

3) Crocker et Tennyson³

Introduction

Le but de l'article est de pallier à ce qui, de l'avis des auteurs, manque à la littérature empirique. C'est-à-dire la considération des mécanismes contractuels par lesquels les assureurs peuvent gérer le problème de la fraude.

Hypothèses

Les assurés possèdent de l'information privée sur leurs pertes. De plus, ils peuvent falsifier ces pertes moyennant un certain coût.

Il est important de noter que pour ce modèle, c'est le processus de falsification, et non le processus vérificationnel, qui engendre des coûts.

On définit par conséquent un contrat efficace comme un contrat sélectionné par rapport à la pondération de deux choses. La première est l'effet bénéfique de distribution du revenu entre l'état de la nature "accident" et "non-accident" qu'on reconnaît à l'assurance. La seconde, que le paiement d'assurance contingent à la taille de la perte n'induit pas l'assuré à déclarer une perte plus lourde qu'elle ne l'est en vérité, afin de réclamer plus.

³Crocker K.J., Tennyson S. (1996), Contracting with Costly State Falsification: Theory and Empirical Results from Automobile Insurance, *Pennsylvania State University*

Les agents sont riscophobes, avec fonction d'utilité $U(W_i)$ définie von Neumann-Morgenstern. W_i est la richesse de l'agent dans l'état de la nature i , tel que $i=N$ s'il n'y a pas de perte. Si l'agent subit une perte alors $i=H$ dans le cas d'une perte élevée et $i=L$ dans le cas d'une faible perte.

On définit γ la probabilité de perte et λ est la probabilité conditionnelle de faible perte étant donné que l'agent subit une perte. L'espérance d'utilité est alors notée:

$$V(W_N, W_L, W_H) = (1-\gamma)U(W_N) + \gamma\lambda U(W_L) + \gamma(1-\lambda)U(W_H)$$

La perte est notée θ_H (θ_L) dans le cas où celle-ci est grande (petite), avec $\theta_H > \theta_L$.

Un contrat d'assurance $C = (p, I_H, I_L)$ comprend une prime p payée dans le cas où il n'y a pas de perte, et le paiement I_i de la compagnie d'assurance à l'assuré en cas de perte ($i \in \{H, L\}$). La compagnie d'assurance ne fait aucune espérance de profit. C'est à dire que:

$$(1-\gamma)p - \gamma\lambda I_L - \gamma(1-\lambda)I_H = 0$$

On pose que l'assureur peut observer si l'état de la nature "pas de perte" est survenu, mais l'assuré peut lui masquer l'étendu de la perte, si perte il y a.

Soit une perte réelle θ , l'assuré peut mentir en déclarant une perte y , à un coût $c(\theta, y) = \xi(\theta-y)^2$, où ξ est un paramètre exogène. Alors, l'utilité d'un individu subissant la perte θ et déclarant la perte y pour recevoir la compensation I sera:

$$U(W - \theta + I - \xi(\theta-y)^2)$$

Pour un coût de falsification suffisamment élevé mentir peut se révéler moins intéressant que révéler la vérité. Dans un tel cas, l'équilibre "first best" (C^*) est atteignable. Pour un coût de falsification moins élevé, l'individu ayant subi θ_L peut préférer déclarer θ_H pour recevoir la compensation I_H . C^* est atteignable ssi $\theta_L < \theta_H - 1/\xi$, sinon le contrat sera optimal de second rang uniquement.

Modèle

Cette étude utilise l'approche de *mécanisme direct* développée par Myerson (1979). Selon cette approche, un assuré qui se présente comme étant de type θ_i reçoit l'allocation contingente $A^i = \{I_i, y_i\}$. Pour être réalisable, une allocation doit satisfaire les conditions d'auto-sélection:

$$U(W - \theta_i + I_i - \xi(\theta_i - y_i)^2) \geq U(W - \theta_i + I_j - \xi(\theta_i - y_j)^2) \quad ; \forall i, j \in \{H, L\}$$

En d'autres termes, l'utilité à révéler la vérité est plus grande que celle de mentir.

Le contrat efficace est donc une solution au problème de bien-être qui maximise l'espérance de l'utilité des assurés:

$$\text{Max}_{p, I, y} V(W_N, W_L, W_H)$$

S.A.

$$(1) \quad (1-\gamma)p - \gamma\lambda I_L - \gamma(1-\lambda)I_H = 0$$

$$(2) \quad U(W - \theta_i + I_i - \xi(\theta_i - y_i)^2) \geq U(W - \theta_i + I_j - \xi(\theta_i - y_j)^2)$$

où $W_N = W - p$ et $W_i = W - \theta_i + I_i - \xi(\theta_i - y_i)^2 \quad \forall i \in \{H, L\}$

Conclusion

La solution du problème décrit plus haut nous amène au théorème suivant:

Théorème: Si $\theta_L > \theta_H - 1/\xi$, alors un contrat d'assurance est optimal de second rang, et satisfait aux conditions suivantes:

- (1) Pour les individus de type θ_L : $y_L = \theta_L$ et $W_L > W-p$.
- (2) Pour les individus de type θ_H : $y_H > \theta_H$ et $W_H > W-p$.
- (3) $U(W_L) = U(W - \theta_L + I_H - \xi(\theta_L - y_H)^2)$

Les conclusions de cette étude sont ensuite énoncées sous forme d'hypothèses, le but de l'article étant de tester empiriquement celles-ci.

Hypothèse 1: Dans un environnement avec falsification coûteuse de l'état de la nature, les agents qui subissent de faibles pertes seront sur-assurés et ceux qui subissent de lourdes pertes seront sous-assurés.

Hypothèse 2: Dans un environnement avec falsification coûteuse de l'état de la nature, le paiement d'assurance pour de faibles (grandes) pertes devrait être plus (moins) élevé que le montant réclamé.

Hypothèse 3: Pour des accidents où la falsification est plus (moins) coûteuse, la fonction de paiement d'assurance devrait être plus (moins) abrupte.

Cette étude se distingue des deux premières par son hypothèse de falsification coûteuse des états de la nature au lieu de l'hypothèse de vérification coûteuse.

Les données de cette étude portent sur des réclamations relatives à des lésions corporelles encourues lors d'accidents automobiles. Dans cette catégorie, la littérature contient des approximations de coûts de falsification. En effet, dans leurs études de réclamations sur la compensation des travailleurs, Dionne et St-Michel (1991) classifient les lésions par degré de difficulté à établir un diagnostic précis. En assurance de biens, par contre, rien ne nous indique qu'une révélation fautive des dommages encourus engendre des coûts pour l'assuré.

Finalement une conclusion étonnante du modèle est qu'en présence du problème considéré, si le contrat est optimal alors la compagnie d'assurance donnera plus que le montant réclamé aux individus qui auront réclamé de petites sommes. Nous aborderont ceci à la prochaine section.

4) Discussion et conclusion

Les trois articles considérés ont en commun qu'avec les contrats optimaux trouvés, les assurés ne fraudent jamais. Or dans la réalité, on observe des cas de fraude. Quel est le problème?

Une hypothèse implicite, qui ressort des deux premiers articles, est celle de plein engagement des deux parties. Or une des deux parties aura intérêt à renégocier. C'est-à-dire que les deux parties s'entendent sur un contrat et s'engagent pleinement à le respecter, mais qu'une des deux parties n'aura aucun intérêt à le respecter. Cette hypothèse est dangereuse pour le contrat qui nous occupe car on peut lui appliquer un raisonnement similaire à celui de Dionne et Doherty (1990) concernant, dans ce cas, le problème de sélection adverse:

.L'agent et le principal s'entendent sur un contrat et les deux parties s'engagent à le respecter.

.Le principal sait que l'agent ne ment jamais car, selon le contrat, ce n'est jamais dans son intérêt.

.Quand le principal sélectionne, de façon aléatoire (Mookherjee et Png) ou systématiquement (Townsend) des dossier à vérifier, il lui est coûteux d'effectuer la vérification.

.De plus, dans Mookherjee et Png, le principal doit récompenser les agents qui ont, (de façon certaine), révélé la vérité sur leurs pertes.

.Donc, dans les deux modèles, le principal sait que le processus vérificationnel est coûteux et, de surcroît, inutile, car les agents révèlent la vérité.

.Le principal choisira donc de ne pas vérifier.

.L'agent anticipe ce raisonnement de l'assureur, alors, sachant qu'il n'y aura pas de vérification il choisira de frauder.

Le même problème sur l'hypothèse plein engagement tient aussi pour l'étude de Crocker et Tennyson, et ce, de la façon suivante. En effet, les deux parties s'accordent sur le contrat avant sa signature. Les deux parties s'engagent pleinement à respecter ce contrat. Lors de la réclamation, l'agent révèle la vérité car il n'est pas dans son intérêt de mentir. Le principal n'a par conséquent aucune motivation à payer plus que le montant de la réclamation, si celle-ci est petite. Il ne paie donc pas plus en cas de petites réclamations.

Townsend a mentionné dans son article, que de dessiner le patron du contrat optimal avec comme toile de fond un problème de risque moral ex-post, amenait à avoir, pour ce contrat optimal, un déductible positif. Le déductible positif est un résultat bien connu de la littérature en assurance, mais, dans cette même littérature, il s'obtient en incluant un taux de chargement positif au contrat. Donc, le taux de chargement positif et le problème de risque moral, pris séparément, entraînent tous deux des déductibles. On sait que les deux sont présents dans la réalité. Qu'arrive-t-il quand les deux sont présents en même temps? Comment, alors, interagissent-ils? Sans modèle adéquat, nous ne pouvons que spéculer. Nous y reviendrons à la conclusion du présent travail.

Comme nous l'avons mentionné plus haut, Mookherjee et Png ont introduit le problème de risque moral classique au modèle. Ceci généralise le modèle, car le risque moral classique est lui aussi à considérer pour bâtir le contrat optimal. Ce faisant, les auteurs introduisent une perte de généralité en ce qui concerne la fonction d'utilité.

Les trois articles, enfin, supposent que la population est homogène. Cette homogénéité de la population veut dire que, par exemple, si il y a une punition pour la fraude, celle-ci a le même effet dissuasif, au même degré, pour tous les membres de la population. C'est une hypothèse qui, à notre avis, fait perdre de la généralité aux modèles considérés.

En résumé, la construction du contrat optimal devrait considérer:

- .Le risque moral ex-post
- .Le taux de chargement
- .Le risque moral ex-ante
- .L'hétérogénéité de la population
- .La sélection adverse

Tout ces facteurs sont dans la réalité en interaction.

Nous allons, dans les sections subséquentes, nous intéresser à mesurer l'étendue du problème de la fraude à l'assurance. Dans les deux premiers articles, l'hypothèse de vérification stipulait que celle-ci était parfaite. Qu'en est-il en réalité? Nous savons qu'il existe de la fraude, car une partie est détectée, mais jusqu'où connaissons-nous l'étendue du problème? Est-ce que la vérification est, comme dans la littérature, parfaite, ou est-elle pitoyablement inefficace à détecter les fraudeurs? Observons-nous tous les cas de fraudes, ou la pointe d'un iceberg?

Telles sont les questions qui stimulent la présente étude. D'autant plus que le problème de la fraude à l'assurance, par sa définition même, ne doit pas être pris à la légère. En effet il est sournois puisque en falsifiant le résultat de l'aléa, l'assuré détruit le principe même de l'assurance.

Problématique

1) Description:

Le problème majeur, pour le niveau de la fraude, en est un d'estimation et donc de statistique. Ce problème d'estimation peut se rencontrer dans la littérature pour tous les phénomènes cachés (marché noir, fraude fiscale, braconnage, ...).

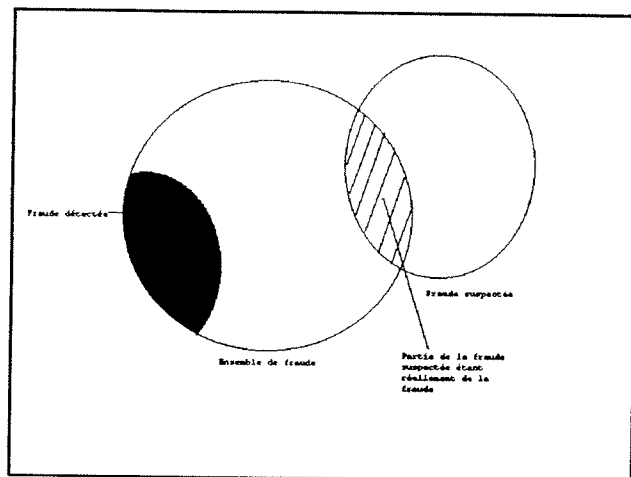
Pour estimer (et non évaluer), un phénomène on a besoin d'une loi probabilistique (exacte ou asymptotique) et de trouver un estimateur convergent et efficace de la quantité, proportion ou relation qu'on cherche.

Pour ce qui est des phénomènes cachés, le même problème revient sans cesse: les données. On ne peut pas procéder par sondage car, par définition, les gens mentent sur ce phénomène. Donc tout ce qu'on peut obtenir dans les données est le nombre de détections du phénomène.

Les estimateurs qui correspondent à ce problème sont connus dans la littérature sous le nom de "count data estimators", et le problème majeur de ces estimés est leur robustesse.

2) Illustration et historique:

La figure ci-contre illustre le problème. On suppose que toutes les couvertures sont présentes. L'ensemble du milieu est l'ensemble des fraudes dont on recherche le cardinal (à estimer). À gauche, en foncé nous avons les fraudes détectées, (l'information qu'on possède pour estimer le cardinal de tout l'ensemble!).



On peut voir ici qu'on doit se servir d'une partie des données pour estimer le cardinal d'un ensemble plus grand.

Ce problème a été étudié dans la littérature d'abord par Student, qui a considéré la loi de Poisson. La loi de Poisson possède la caractéristique d'absence de mémoire et est particulièrement pratique pour représenter des phénomènes rares.

Ensuite Fisher a considéré la binomiale négative ayant écarté la binomiale comme étant arbitraire. Il ne spécifie pas dans son article ce qu'il entend par là.

La binomiale a quant à elle, été considérée par Binet. On peut lui trouver deux avantages majeurs.

.Premièrement l'estimateur de n est borné inférieurement par l'occurrence maximale de l'échantillon. C'est à dire, par exemple, qu'on ne peut avoir "100" provenant d'une $\text{Bin}(70,p)$, donc l'observation maximale est une borne inférieure pour le nombre d'essais de Bernouilli implicites à la binomiale.

.Deuxièmement, il y a une raison intuitive à prendre la binomiale pour le problème considéré et nous y reviendrons dans la prochaine section.

Le problème de robustesse quant à lui, a été étudié par Olkin, Petkau et Zidek (1981). Ces chercheurs ont développé une méthode pour augmenter la robustesse du " n " estimé. C'est cette méthode que nous décrivons à la section "Modèle statistique" et que nous avons appliqué pour obtenir les résultats présentés à la fin du présent travail.

Dans notre cas, nous possédons aussi de l'information supplémentaire, c'est-à-dire les couvertures où la fraude est suspectée. Les éléments de cet ensemble peuvent être vus comme "plus susceptibles" de faire partie de l'ensemble des fraudes que les autres couvertures. La partie de cet ensemble qu'on inclut à l'ensemble des fraudes observées est représenté par la partie hachurée du graphique.

Cet ensemble doit être créé à l'aide d'une hypothèse, nous y reviendrons dans la section "Traitement des données".

Modèle Statistique

Si on suppose que le nombre de fraudes détectées suit une binomiale (n,p) où p = la probabilité de détecter une fraude étant donné qu'il y a fraude, notre problème en est un d'estimation conjointe de n et p pour la binomiale.

Pour estimer les paramètres n et p de la binomiale nous procédons par la méthode des moments. Nous avons deux paramètres à estimer, il nous faut donc deux moments. Bien entendu, pour calculer deux moments, l'espérance et la variance, nous devons avoir plus d'un groupe. Nous reviendrons plus tard sur le nombre K de groupes à prendre.

Ce problème peut sembler trivial à première vue, mais il est en réalité complexe car les estimateurs connus de n sont très instables (non-robustes).

$$\begin{array}{llll} \text{Soit } X \sim \text{Bin}(n,p) & \text{alors} & E(X) & = np \\ & & \text{Var}(X) & = np(1-p) \end{array}$$

Donc: $\text{Var}(X)/E(X) = (1-p)$ qui peut être près de 1.

$$\begin{array}{ll} \text{Si p.ex. } \hat{\text{Var}}(X)/\hat{E}(X) \geq 1 & \text{maximisé à } n \rightarrow \infty \\ & \text{et } p \rightarrow 0, \end{array}$$

donc clairement, de petits chocs dans les données vont donner lieu à de grandes différences dans les estimateurs, d'où l'instabilité.

***Note:** L'instabilité ne provient pas de l'utilisation de la binomiale, le même problème se rencontre avec la Poisson ou la binomiale négative par exemple.

La première partie du travail doit être de rendre l'estimateur plus robuste. C'est ce qu'on fait Olkin, Petkau et Zidek en 1981, dans leur article "A Comparison of n Estimators for the binomial distribution". Ils ont proposé une procédure pour stabiliser le n estimé.

Pour commencer nous allons réécrire le problème en adoptant leur notation:

Soit S_1, S_2, \dots, S_K des observations de détection du phénomène étudié.

Pose $P = \text{Pr}[\text{détection}/\text{fraude}]$

$n = \text{Nb. de fraudes dans l'échantillon}$

Ici, on peut voir combien la binomiale est intuitive pour ce genre de problème car si tout est aléatoire, on devrait avoir sensiblement le même nombre de fraudes par groupe. Si on pose ce nombre égal à 20 et que la probabilité $p = 0.5$ (étant donné que la réclamation est frauduleuse, les évaluateurs ont une chance sur deux de détecter cette fraude) alors intuitivement on devrait s'attendre à avoir une moyenne de $0.5 \times 20 = 10$ fraudes détectées par groupe. Ceci sous-entend que:

$$S_i \sim \text{Bin}(n, p)$$

$$\rightarrow E[S_i] = nP$$

$$\rightarrow V[S_i] = nP(1-P)$$

Posons u, σ^2 l'espérance et la variance échantillonnale.

$$\rightarrow u = \sum s_i / K$$

$$\rightarrow \sigma^2 = \sum (S_i - u)^2 / K$$

Donc, par la méthode des moments:

$$P = u / n$$

$$n = u^2 / (u - \sigma^2)$$

Ici, on peut noter qu'on écrit le n estimé comme $n(u, \sigma^2)$ une fonction de l'espérance et de la variance échantillonnale.

De plus on peut mq.: n, p : convergents pour $k \rightarrow \infty$.⁴

L'exemple suivant servira à démontrer à quel point ces estimateurs peuvent être non robustes:

Soit des observations (16, 18, 22, 25, 27) provenant d'une binomiale (n, p) ; par la méthode des moments (avec l'estimateur trouvé plus haut) on trouve: $n = 102$.

On suppose maintenant qu'on s'était trompé de 1 sur la dernière observation, donc on a le même échantillon sauf pour la dernière observation qui passe de 27 à 28, et on recalcule notre estimation à: $n = 195$!

Avec une légère variation dans l'échantillon notre estimateur a presque doublé. Ce problème de stabilité est connu dans la littérature comme un problème de robustesse.

Graphiquement le problème est que la fonction de vraisemblance est peu accidenté autour de notre estimateur.

⁴Olkin I. and A.J. Petkau and J.V. Zidek (1981), A Comparison of n Estimators for the Binomial Distribution, *Journal of the American Statistical Association*

Corrections: (Olkin, Petkau, Zidek)

Posons $n(u) = u^2 / (u - \sigma^2)$

$\rightarrow n'(u) = [2u(u - \sigma^2) - u^2] / (u - \sigma^2)^2$

Test de stabilité:

si $n'(u) \geq -1 \quad \rightarrow \quad$ MME stable

sinon: On doit corriger pour la robustesse.

(MME = Method of Moment Estimator)

Procédure:

Posons $\epsilon > 0$ ajouté à chaque S_i

$\rightarrow u + u + \epsilon = x$

$\rightarrow \sigma^2$ reste pareil

Conditions sur ϵ : (endogénéisation)

1) $x > \sigma^2$

2) $n'(x) \geq -1$

3) $n(x) \geq S_{\max} + \epsilon$

Ces conditions peuvent être vues comme un processus de "graduation". C'est-à-dire qu'on impose des contraintes rationnelles (ou triviales) à nos estimés. Ceci revient à entrer notre idée à priori pour limiter nos estimés.

Par exemple, la première contrainte revient à limiter les probabilités à l'intervalle $]0,1[$ (ce qui est assez raisonnable) et la seconde vient directement du test de robustesse. La dernière est une condition implicite de la binomiale, à savoir qu'on ne peut avoir une réalisation de la binomiale plus grande que n , (ce dont nous avons discuté à la section précédente en commentant le travail de Binet), où S_{\max} = occurrence maximale.

1) On prend la 2^{ème} contrainte comme "serrée" à $\epsilon = \epsilon_0$

$$\rightarrow \epsilon_0 = (1 + 1/\sqrt{2})\sigma^2 - u$$

Si ceci respecte la 3^{ème} contrainte

$$\text{c.a.d si: } S_{\max} \leq n(u + \epsilon_0) - \epsilon_0$$

$$\rightarrow z/\sigma \leq (1 + \sqrt{2})$$

$$\text{où } z = (S_{\max} - u)/\sigma$$

2) Si ceci ne respecte pas la troisième contrainte

On prend $\epsilon = \epsilon_0$:

$$\text{tq } \epsilon_0 = [S_{\max}(u - \sigma^2) - u^2]/[u + \sigma^2 - S_{\max}]$$

ie: La troisième contrainte devient serrée.

Donc la version stabilisée du MME:

$$\text{MME:S} = \text{Max}\{\sigma^2\psi^2/(\psi-1), S_{\max}\}$$

où:

$$\begin{aligned} \psi &= u/\sigma^2 && \text{si } u/\sigma^2 \geq 1 + 1/\sqrt{2} \\ &= \max\{z/\sigma^2, 1 + \sqrt{2}\} && \text{si } u/\sigma^2 < 1 + 1/\sqrt{2} \end{aligned}$$

et:

$$z = (S_{\max} - u)/\sigma.$$

Si on revient sur l'exercice et si on applique cette fois-ci la méthode décrite ci-haut on obtient $n = 70$ pour le premier échantillon et $n = 80$ pour le second. Ce n'est pas parfait mais c'est beaucoup plus stable.

Cette façon d'écrire le processus de stabilisation est particulièrement bien adaptée pour la programmation. C'est d'ailleurs de cette façon qu'il a été appliqué. Cette procédure a été programmée en APLC, le programme est inclut en annexe et les résultats sont décrits dans la section "Traitement des données".

Méthodologie

1) Les données:

Les données proviennent de 18 compagnies d'assurance où l'on a réouvert un échantillon aléatoire de 2509 dossiers fermés, pour 2773 couvertures.

La fraude est donnée par couverture et est classée comme suit par les évaluateurs:

1) Détection	(19)
2) Suspicion	(123)
3) Ni (1), ni (2)	(2631)

Les suspicions de fraudes sont données par les évaluateurs des compagnies participantes avec leur degré de suspicion (de 1 à 10). Un degré de suspicion peut être vu comme une probabilité, tel qu'estimée par l'évaluateur du dossier, que la réclamation soit frauduleuse.

Les 18 compagnies qui ont participé représentent 70% du marché. On peut donc parler d'une étude sur l'industrie en général.

2) Traitement des données

Nous allons d'abord donner les principales approches utilisées et les résultats obtenus avec chacune d'elle pour ensuite mieux voir la problématique.

La première étape du problème considéré vise à recréer des échantillons d'une même loi binomiale (n,p) .

Le "p" dans notre modèle représente la probabilité de détecter une fraude, étant donné qu'on ait entre les mains une couverture frauduleuse. Il ne nous cause pas de problème autre que le fait qu'il est supposé constant pour chaque couverture et que cette hypothèse est assez forte. Cependant, si les couvertures sont ordonnées de façon stochastique on estime réellement l'espérance mathématique de "p" et ceci sans biais.

Le "n" quant à lui, représente le nombre de fraudes dans chacun des groupes. Il est donc par hypothèse constant pour chaque groupe. Ce qui suppose qu'on a le même nombre de fraudes dans chaque groupe. Comme les données ne sont pas déjà séparées en groupe avec cette hypothèse, comme dans l'étude de l'université de Floride⁵, nous devons donc randomiser avant de séparer en groupe.

⁵Automobile Insurance Fraud Study (1991), Florida Insurance Research Center, University of Florida, Florida Insurance News Service.

Comme nous devons estimer deux paramètres, notamment "n" et "p", nous devons nous servir de deux moments (au moins), ce qui implique au moins deux groupes. Plus il y aura de groupes plus les mesures de moments seront précises, car elles convergent quand: $K \rightarrow \infty$.

Mais plus les groupes sont nombreux, moins l'hypothèse de la binomiale, qu'il y ait le même nombre de fraudes "n" dans chaque groupe, est acceptable. A la limite pour cette hypothèse nous devrions avoir que $K = 1$.

Nous pouvons ainsi voir la relation d'arbitrage (trade-off), entre l'hypothèse de la binomiale et la qualité des estimateurs, pour ce qui est du nombre de groupes.

Dans les premières tentatives d'estimation nous avons pris $K = 6$ de façon arbitraire. C'est ce qui est décrit plus bas dans la section 1 "Estimation Unique". Ensuite nous décrivons à la section suivante comment nous nous y sommes pris pour trouver l'incidence du nombre de groupes et pour stabiliser les estimateurs.

Section 1) Estimation Unique:

Ici, le but est de trouver une méthode d'estimation cohérente avec les hypothèses de départ et de vérifier la robustesse obtenue en répétant l'expérience.

1^{ère} approche:

Dans la première approche utilisée, nous avons d'abord procédé à la randomisation des 2509 dossiers à l'aide d'une fonction de processus stochastique en APL.

Nous avons ensuite séparé les dossiers en six groupes $K=6$ pour un total de 418 dossiers par groupe. Il est à noter que le dossier ainsi exclu de l'étude l'est de façon aléatoire. L'important est d'avoir sensiblement le même nombre de fraudes par groupe.

L'étape suivante consistait à compiler les fraudes détectées (**D**), suspectées (**S**) et non-fraudes (**N**) puis de comparer les résultats avec ceux obtenus par Dionne et Belhadji⁶ sur la même base de données.

Ensuite il s'agissait de calculer u , σ^2 ainsi que n et p par la méthode des moments avec les formules décrites plus haut. Les résultats sont présentés au tableau 1.

Tableau 1

Fraudes par groupe						
Groupe	Fraude			Non-Fraude	D	D + S
	Suspectée	Détectée				
1	20	2	437	2	22	
2	27	1	443	1	28	
3	21	0	449	0	21	
4	19	10	425	10	29	
5	23	4	429	4	27	
6	13	2	448	2	15	
Total :	123	19	2631	19	142	
E(x)	20.500	3.167	438.500	3.167	23.667	
V(x)	21.500	12.967	99.100	12.967	28.667	
(Programme: ESTIMATE)						
	Estim. de (1-p) =	4.094737	1.211268			
	Estim. de p =	-3.09474	-0.21127			
	Estim. de n =	-1.02324	-112.022			
	Estim. de n / N =	-0.00221	-0.24247			

⁶Dionne G. et B. Belhadji (1996), Evaluation de la fraude à l'assurance automobile au Québec, CRT-96-08.

Au tableau 1 nous pouvons voir le problème de cette approche. La variance est énorme par rapport à l'espérance ce qui nous donne des probabilités négatives. Cela nous dit que nos hypothèses de départ ne sont pas respectées. En effet, trouver une probabilité négative ne résiste pas à l'analyse très longtemps.

Le problème majeur est que nous avons randomisé sur les dossiers. Pour avoir une véritable binomiale en procédant de cette façon il faudrait que les probabilités de détection (ou de suspicion) de fraude dans un même dossier soient indépendantes. Les **D** (ou **S**) cependant sont donnés par couverture, et non par dossier, ce qui fait "exploser" la variance.

2^{ème} approche:

A l'inverse cette fois-ci nous avons d'abord procédé à la compilation des **D**, **S** et **N** puis nous avons placé leurs valeurs dans une matrice de 2773 couvertures.

Nous avons ensuite effectué une randomisation sur les 2773 couvertures, pour encore une fois séparer les données en six groupes $K=6$ mais cette fois-ci avec 462 couvertures par groupe.

Enfin, par la méthode des moments nous avons calculé u , σ^2 , n et p . Les résultats sont présentés cette fois-ci au tableau 2.

Le pourcentage de fraude estimé = $100xn/462 = 2.17\%$ en ne tenant compte uniquement que des fraudes détectées et de $100xn/462 = 20.9\%$ dans l'hypothèse extrême que toute fraude suspectée soit une fraude détectée.

Tableau 2

Fraudes par groupe					
Groupe	Fraude			D	D + S
	Suspectée	Détectée	Non-Fraude		
1	27	3	431	3	30
2	17	5	440	5	22
3	22	4	436	4	26
4	17	4	441	4	21
5	24	1	437	1	25
6	16	2	444	2	18
Total :	123	19	2629	19	142
E(x)	20.500	3.167	438.167	3.167	23.667
V(x)	20.300	2.167	20.567	2.167	17.867

(Programme: ESTIMATE2)

Estim. de (1-p) =	0.684211	0.75493
Estim. de p =	0.315789	0.24507
Estim. de n =	10.02778	96.57088
Estim. de n / N =	0.021705	0.209028

Ici, le problème de l'approche est qu'on n'entre pas l'information supplémentaire que nous donne la fraude autrement que par le biais d'une hypothèse extrême.

3^{ème} approche:

Pour pouvoir entrer l'information supplémentaire on doit aussi faire d'autres hypothèses, p.ex. compter les fraudes suspectées (S) comme des fraudes détectées (D) partielles.

Une façon de faire ceci est de procéder par degré de suspicion. Chaque fraude suspectée l'était à un certain degré par l'évaluateur qui a étudié le dossier et ceci est inclut dans la banque de données.

Nous avons alors compté comme .3 fraude détectée une fraude suspectée avec un degré de suspicion de 3 et ceci pour 1 à 10. Les résultats ainsi obtenus sont exposés au tableau 3.

Tableau 3

Fraudes par groupe					
Groupe	Fraude			D	D + S
	Suspectée	Détectée	Non-Fraude		
1	21	3	437	3	24
2	7.3	3	444	3	10.3
3	27.9	6	437	6	33.9
4	12.4	2	441	2	14.4
5	10.2	3	437	3	13.2
6	14.9	2	434	2	16.9
Total :	93.7	19	2630	19	112.7
E(x)	15.617	3.167	438.333	3.167	18.783
V(x)	57.846	2.167	12.667	2.167	76.326

Estim. de (1-p) =	0.684211	4.063478
Estim. de p =	0.315789	-3.06348
Estim. de n =	10.02778	-6.13137
Estim. de n / N =	0.021705	-0.01327

Avec les résultats présentés ci-haut nous pouvons voir le problème de cette approche. Premièrement, il est difficile de soutenir que ces données proviennent d'une binomiale. La binomiale ne donne que des résultats discrets, alors 7.3 comme occurrence, par exemple, est difficile à justifier. De plus, des comptes de .1 à .9 sont loins du résultat binaire d'une épreuve de Bernouilli, ce qui a pour effet de faire "exploser" une fois de plus la variance.

Il est cependant intéressant de noter que le résultat de l'estimateur de n avec seulement la fraude détectée est resté constant à 10.03 ce qui démontre une certaine robustesse.

4^{ème} approche:

Enfin nous avons procédé à l'application de la méthode de Olkin, Petkau et Zidek pour stabiliser nos estimateurs.

Avec l'entrée de l'information supplémentaire nous n'avons pas obtenu une robustesse énorme (les résultats varient de 18 à 42% de fraude estimée) mais nous ne trouvons jamais de probabilités négatives ou de n "très grands".

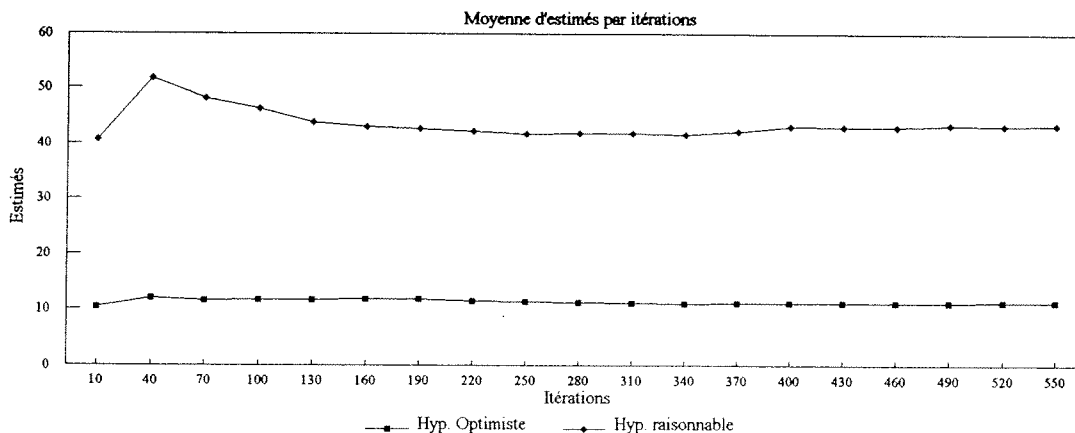
Section 2) Itérations:

Le but ici était de répéter l'expérience un certain nombre de fois et d'ensuite prendre la moyenne des estimateurs du nombre de fraudes pour ainsi acquérir une robustesse supplémentaire.

La question ici est de trouver à quel nombre d'itérations les estimateurs commencent à être stables et comment cette stabilité évolue.

La procédure utilisée pour répondre à ces questions est fort simple. Il s'agit de prendre la moyenne sur des quantités de plus en plus grandes d'estimateurs. C'est ce que nous avons fait comme on peut le constater sur le graphique qui suit. L'estimateur du nombre de fraudes totales, en ne considérant que la fraude détectée, est stable même pour de petits échantillons.

L'autre estimateur considère comme "fraude" toute fraude détectée et toutes les suspicions de fraude avec un degré de suspicion plus grand que cinq. C'est notre hypothèse raisonnable (ni optimiste, ni pessimiste). Les résultats ne sont pas très robustes pris individuellement mais le sont assez sur une moyenne prise sur un échantillon assez grand.



Il serait ici intéressant de noter qu'on doit utiliser l'approche d'Olkin, Petkau et Zidek obligatoirement avant d'enclencher le processus itératif et de prendre la moyenne. L'estimateur de la méthode des moments est en effet sans biais. C'est à dire que

$$E[u^2 / (u - \sigma^2)] = n$$

mais que si $u \rightarrow \sigma^2$ alors $n \rightarrow \infty$. Ce qui donne un très grand poids aux écarts positifs par rapport à la moyenne. La procédure des trois chercheurs s'avère indispensable pour écarter ce biais.

Section 3) Résultats

Nous avons effectué mille itérations pour chaque hypothèse représentée dans le tableau suivant:

Six Groupes (N=462)					
1000 Itérations					
	Occurrence		Estimation		
	(n)	(n/N%)	(n)	(n/N%)	(p)*
D	19	0.6852	10.9449	2.3690	0.2893
D + (S > 9)	38	1.3704	22.1362	4.7914	0.2861
D + (S > 8)	48	1.7310	27.7267	6.0015	0.2885
D + (S > 7)	62	2.2358	35.449	7.6729	0.2915
D + (S > 6)	71	2.5604	41.9	9.0693	0.2824
D + (S > 5)	78	2.8128	43.7889	9.4781	0.2969
D + (S > 4)	100	3.6062	56.4136	12.2107	0.2954
D + (S > 3)	108	3.8947	59.4966	12.8781	0.3025
D + (S > 0)	127	4.5799	69.5356	15.0510	0.3044
D + S	142	5.1208	76.3894	16.5345	0.3098

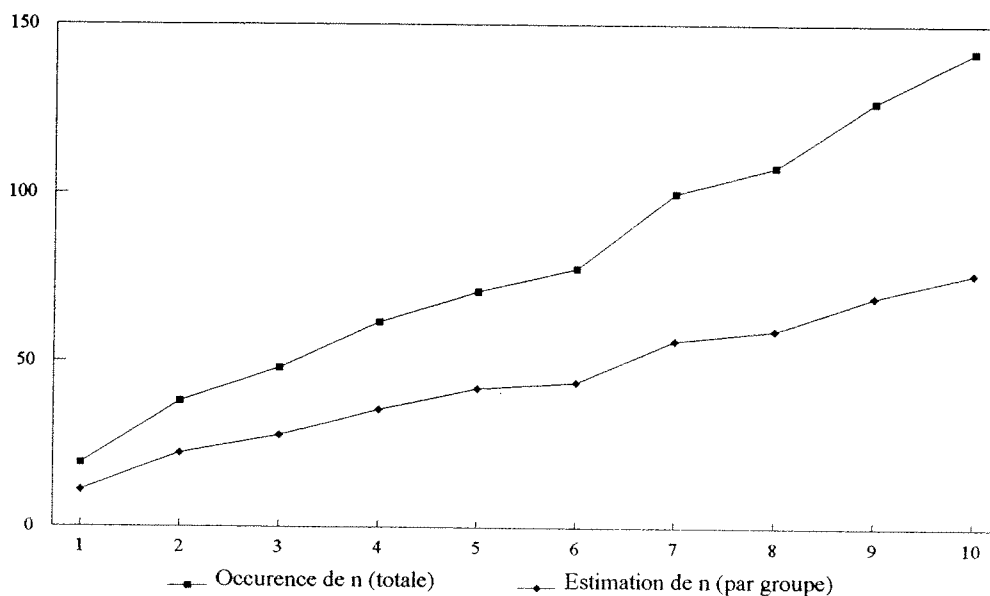
* Où p = Pr [Détecter / Fraude]

Les hypothèses vont de la plus optimiste "D" ,où seulement la fraude détectée est considérée, à la plus pessimiste "D+S" où toute couverture suspecte est prise en compte.

Les résultats nous donnent l'estimation de p (la probabilité de détecter une fraude si fraude il y a). Ceci peut être vu comme un indice d'efficacité des dispositifs de détection de la fraude en place dans l'industrie. On peut noter que cette probabilité passe de 0.289 à .31 ce qui est cohérent avec l'idée qui veut que plus on inclut des couvertures suspectes à l'analyse plus on détecte effectivement de fraudes.

Le pourcentage réel de fraude estimé oscille aux alentours de 9.5% sous l'hypothèse mitoyenne que les couvertures suspectes avec un degré de suspicion de 6 et plus " $D + (S > 5)$ " sont considérées comme frauduleuses. Les bornes inférieures et supérieures du pourcentage de fraude réel peuvent être obtenues en considérant les deux hypothèses extrêmes. Ceci nous donne un intervalle assez large de 2.4 à 16.5%.

Le graphique qui suit représente bien l'évolution des estimateurs de n (par groupe) par rapport aux observations de fraude (occurrence de n). Cette évolution prend une allure concave ce qui indique bien qu'on "détecte" de plus en plus toutes les fraudes dans le groupe. Avec, à la limite, qu'on ne pourrait avoir moins de fraudes estimées que détectées (de par l'hypothèse implicite dans le choix de la loi binomiale).



Le nombre de fraudes détectées dans l'étude (19) sur le nombre total de couvertures étudiées (2773) nous donne un pourcentage de fraudes détectées de 0.685%. L'estimation de n par la méthode que nous avons employée (10.945) en ne tenant compte que de ces même fraudes détectées, sur le nombre total de couvertures par groupe (462) donne un pourcentage de fraudes estimé (2.369%). Donc le *facteur de multiplication* de l'estimation par rapport à l'occurrence est de 3.5. Sous cette hypothèse, ce même facteur évolue inversement avec le "degré de pessimisme de l'hypothèse" jusqu'à 3.2. Ceci encore une fois nous confirme qu'on "détecte" de plus en plus de fraudes en introduisant les fraudes suspectes.

Pour l'hypothèse mitoyenne ce facteur est de 3.3684. C'est donc dire que le nombre réel de fraude serait 3.4 fois plus élevé que ce qu'on détecte généralement dans l'industrie. Ces résultats sont présentés dans le tableau suivant.

Facteur de Multiplication de n	
S	3.4563
S + (D > 9)	3.4952
S + (D > 8)	3.4658
S + (D > 7)	3.4305
S + (D > 6)	3.5408
S + (D > 5)	3.3684
S + (D > 4)	3.3848
S + (D > 3)	3.3054
S + (D > 0)	3.2851
S + D	3.2277

Section 4) Les Groupes:

Jusqu'à présent, nous avons divisé les observations en $K=6$ groupes de 462 couvertures. Ce nombre de groupes étant arbitraire, l'étape suivante consistait à prendre des nombres de groupes différents afin de voir si les résultats restaient sensiblement les mêmes.

On sait que plus le nombre de groupes est élevé plus l'estimateur de la variance est fiable. Mais plus le nombre de groupe est élevé, plus il est difficile de soutenir qu'il y a le même nombre de fraudes dans chaque groupe, c'est à dire de maintenir que chaque groupe contient une binomiale (n,p) avec le même n .

Nous avons effectué la même expérience de mille itérations qu'à la section précédente avec, cette fois-ci, $K = 5, 7, 8, 9, 10, 11, 12$ et 18 groupes.

On peut commencer par regarder les résultats que nous avons obtenus avec $K = 7$ groupes, donc avec 396 couvertures par groupe. Ils sont présentés dans le tableau suivant:

Sept Groupes (N=396) 1000 Itérations					
	Occurrence		Estimation		
	(n)	(n/N%)	(n)	(n/N%)	(p)*
D	19	0.6852	9.4477	2.3858	0.2873
D+(S>9)	38	1.3704	19.5406	4.9345	0.2778
D+(S>8)	48	1.7310	24.0245	6.0668	0.2854
D+(S>7)	62	2.2358	29.909	7.5528	0.2961
D+(S>6)	71	2.5604	36.7226	9.2734	0.2762
D+(S>5)	78	2.8128	38.8868	9.8199	0.2865
D+(S>4)	100	3.6062	48.0278	12.1282	0.2974
D+(S>3)	108	3.8947	52.0648	13.1477	0.2963
D+(S>0)	127	4.5799	61.8068	15.6078	0.2935
D+S	142	5.1208	70.1754	17.7211	0.2891

* Où $p = \text{Pr}[\text{Détecter} / \text{Fraude}]$

On peut noter dans ce tableau que l'estimateur de n diminue sensiblement pour toutes les hypothèses.

Par exemple, pour l'hypothèse optimiste "D" il passe de 10.9 à 9.4. Mais ceci est presque exactement compensé par la diminution de couvertures au total dans chaque groupe (de 462 à 396). Ce qui fait que l'estimation de la fraude en terme de pourcentage ne varie presque pas (de 2.369 à 2.386%).

L'estimation de la fraude en pourcentage reste similaire mais augmente un peu sur toutes les hypothèses. Par exemple, pour l'hypothèse mitoyenne, l'estimateur passe de 9.478 à 9.82.

Evidemment le facteur de multiplication, par définition, se trouve lui à augmenter un peu avec cette méthode. Par exemple il passe de 3.37 à 3.49 sous l'hypothèse mitoyenne.

Si on refait l'expérience avec 8 groupes cette fois-ci, on retrouve encore une fois sensiblement les mêmes estimateurs. Pour l'hypothèse optimiste, par exemple l'estimateur du nombre de fraudes par groupe diminue jusqu'à 8.6078. Quand on prend en compte le nombre de couvertures dans chaque groupe, la situation se corrige pour nous donner une estimation du pourcentage de fraudes de 2.4878.

Les résultats importants, sont évidemment les estimateurs de pourcentage de fraude. Ces résultats sont analysé à la prochaine section.

Section 5) Evolution:

Les estimations de pourcentage de fraudes ont tendance à évoluer de façon proportionnelle au nombre de groupes. Cependant cette évolution n'est pas très grande.

Pour l'hypothèse mitoyenne, par exemple, les pourcentages varient de 9.21, pour 5 groupes, à 10.76 pour 18 groupes. Le choix du nombre de groupes est arbitraire. Toutefois on peut dire que la variation des résultats attribuable à ce choix est tolérable.

L'évolution des estimateurs par rapport au nombre de groupes choisi est représentée dans le tableau qui suit.

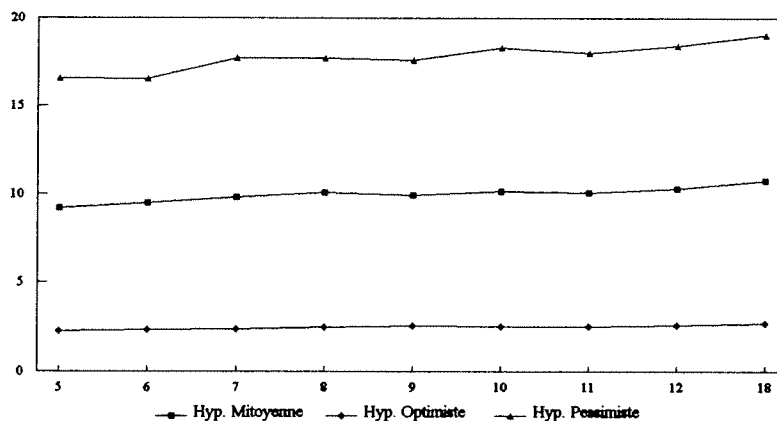
Hypothèses	Estimation du Pourcentage (n/N%)								
	5 Groupes (N = 554)	6 Groupes (N = 462)	7 Groupes (N = 396)	8 Groupes (N = 346)	9 Groupes (N = 308)	10 Groupes (N = 277)	11 Groupes (N = 252)	12 Groupes (N = 231)	18 Groupes (N = 154)
D	2.2860	2.3690	2.3858	2.4878	2.5819	2.5293	2.5501	2.6135	2.7077
D+(S > 9)	4.5631	4.7914	4.9345	4.8232	5.1174	4.9690	4.9756	5.1630	5.2368
D+(S > 8)	5.7705	6.0015	6.0668	6.0588	6.4005	6.2617	6.2785	6.4957	6.6488
D+(S > 7)	7.2835	7.6729	7.5528	7.8712	8.0999	8.0238	8.0044	8.3267	8.5361
D+(S > 6)	8.3811	9.0693	9.2734	9.1111	9.1558	9.1961	9.1621	9.5338	9.7969
D+(S > 5)	9.2065	9.4781	9.8199	10.0528	9.9097	10.1279	10.0604	10.3020	10.7623
D+(S > 4)	11.7235	12.2107	12.1282	12.9225	12.4940	13.1121	12.8471	13.1145	13.6389
D+(S > 3)	12.3135	12.8781	13.1477	13.7772	13.4125	14.0680	13.9661	14.1109	14.7088
D+(S > 0)	14.8829	15.0510	15.6078	15.9181	15.6921	16.5478	16.3490	16.5917	17.2218
D+S	16.5630	16.5345	17.7211	17.7121	17.6176	18.2694	18.0135	18.3934	19.0118

Comme on peut le constater dans ce tableau, moins l'hypothèse est optimiste, plus le choix du nombre de groupes a de l'influence sur notre estimation du pourcentage de couvertures frauduleuses. Donc l'hypothèse où ce choix a le plus d'influence est l'hypothèse pessimiste "D + S". Sous cette hypothèse le pourcentage de fraude varie de 16.6%, pour 5 groupes, à 19% pour 18 groupes.

Ceci nous donne un intervalle de pourcentage de fraude de $17.8 \pm 1.2 \%$, pour l'hypothèse pessimiste. Pour l'hypothèse mitoyenne cet intervalle devient $10 \pm 0.8\%$, alors qu'il passe à $2.5 \pm 0.2\%$ sous l'hypothèse optimiste.

Il serait important de noter que cet intervalle ne doit pas être interprété comme un intervalle de confiance au sens propre. Cet intervalle n'a pas de justification statistique, mais doit être compris comme un intervalle où les résultats fluctuent dépendamment de la méthodologie suivie.

Le graphique ci-dessous représente l'évolution des estimés par rapport au nombre de groupes choisi dans la méthode.



Il serait intéressant ici de noter qu'on ne peut pas déterminer le nombre de groupes optimal en considérant, par exemple, la variance des estimations. En effet, plus il y a de groupes plus les résultats sont stables. Une façon de voir cette stabilité est de considérer ce qui arrive à la limite, c'est à dire avec 2773 groupes de une couverture. A ce stade les deux moments dont on se sert pour notre estimateur sont stables. Ceci implique que le même n est estimé et que par conséquent il n'existe pas de variance entre nos différentes expériences.

Conclusion

La présente étude estime le taux de fraude dans l'industrie à $10 \pm 0.8 \%$ sous l'hypothèse la plus vraisemblable et selon la méthode employée. Ce résultat s'avère stable. Les taux de fraude estimés dans le présent travail sont plus du triple de ceux calculés par Dionne et Belhadji, qui étaient présentés comme des estimations plancher ou, en d'autres termes, comme les observations de fraude.

Un corollaire intéressant de la présente étude est l'estimation de p autour de $1/3$. Encore une fois, p est la probabilité conditionnelle, qu'un évaluateur découvre la fraude dans un dossier frauduleux.

On se souvient dans les modèles théoriques présentés au début de l'hypothèse de vérification parfaite. Pour que la présente étude corrobore cette hypothèse, le p aurait dû être estimé près de un, au lieu de un tiers. Donc la vérification est, d'après nos données, imparfaite. Le p peut être vu comme un indice d'efficacité très bas du processus vérificationnel. Ce qui nous amène logiquement à la question suivante: " Pourquoi cet indice est-il si bas? "

Les candidats de réponse à cette question sont légions. Cet index peut refléter l'incompétence des évaluateurs à détecter les dossiers frauduleux. Il peut aussi nous amener à douter très sérieusement de la pertinence des indicateurs de fraude utilisés par les compagnies. Il peut aussi refléter la pauvre qualité et/ou quantité des processus amenant à la détection des fraudes. Il peut enfin refléter un certain laxisme de la part des compagnies d'assurance face au problème de la fraude.

Le choix de l'une ou l'autre de ces réponses ne peut être effectué sans un modèle théorique adéquat pour appuyer nos dires. Or, aucun des modèles considérés n'inclut la possibilité que le processus vérificationnel soit imparfait.

Une réponse intéressante, cependant, serait la dernière; le laxisme des compagnies d'assurance. En effet, cette dernière peut couvrir les autres explications. L'incompétence des évaluateurs à détecter les dossiers frauduleux, peut s'expliquer, par exemple, par leurs ressources inadéquates ou par un manque de motivation de leur employeur. Ce qui nous ramène au laxisme des compagnies.

L'étude de Dionne et Belhaddji révèle que la plupart des cas de fraudes (93%) ne sont pas poursuivis. Les causes de non-poursuite sont énoncées dans le tableau suivant.

Cause de non-poursuite	
Définition	%
Preuves insuffisantes	59
Manque de ressources	3
Montant de la fraude trop petit	11
Loi inadéquate	1
Peur d'une mauvaise publicité	1
Charge de travail trop grande	2
Pression du courtier	2
Pression du service à la clientèle	2
Pas de raison particulière	5
Autre raison	16

On notera dans ce tableau que 59% des dossiers n'étaient pas poursuivis car les preuves n'étaient pas suffisantes. Pourquoi alors, ne pas avoir poussé l'investigation plus à fond?

Dans ce cas nous pouvons interpréter le refus de procédures d'investigation supplémentaires comme un laxisme de la part des compagnies ou par les coûts élevés des procédures de détectations. Mais nous sommes portés à pencher un peu plus du côté du laxisme de la compagnie comme réponse à la question de l'indice d'efficacité trouvé si bas.

Cela enfin, nous amène à la dernière question: " Pourquoi y aurait-il du laxisme dans l'industrie au sujet de la fraude à l'assurance? "

Si les montants de réclamations sont trop petits pour justifier le coût d'investiguer plus à fond, alors des déductibles plus élevés seraient peut être de mise dans l'industrie. Ceci est en accord avec le modèle de Townsend. Des déductibles plus élevés pourraient augmenter le niveau des réclamations jusqu'au point où l'investigation pourrait être motivée.

Le plus bas déductible (250\$) est resté au même niveau depuis les vingt dernières années. L'industrie automobile fonctionne, pour sa part, en dollars réels. Ceci veut dire que les biens assurés, les véhicules, coûtent maintenant environ quatre fois plus, en terme nominal, qu'il y a vingt ans. Le déductible de 250\$, ne représente plus le même pourcentage du bien assuré. Cette dernière remarque provient d'une enrichissante discussion avec M. Dionne.

Dans le tableau précédent les trois premières causes de non-poursuite peuvent suivre le même raisonnement. Ceci porte à 73% les causes où le problème de déductible trop petit peut être présenté en explication.

Si, d'un autre côté, des investigations plus poussées ou une poursuite, sont vues comme étant de la mauvaise publicité par une compagnie d'assurance, alors le temps est peut être venu d'établir une agence d'information et d'investigation de la fraude à travers l'industrie.

Le problème de la fraude n'est pas seulement un problème de vol des assurés honnêtes, mais est plus pernicieux puisqu'il met en danger le principe même de l'assurance.

Bibliographie

Théorique

- Crocker K.J., Morgan J.** (1995), The Optimality of Costly State Falsification: Sharecropping and Insurance Fraud, *Pennsylvania State University*
- Crocker K.J., Tennyson S.** (1996), Contracting with Costly State Falsification: Theory and Empirical Results from Automobile Insurance, *Pennsylvania State University*
- Dionne G. et B. Belhadji** (1996), Evaluation de la fraude à l'assurance automobile au Québec, CRT-96-08
- Dionne G.** (1994), Economie de l'assurance, *Nanterre*
- Dionne G. and N. Doherty** (1991), Adverse selection in insurance markets: a selective survey, *Contributions to Insurance Economics*, Kluwer Academic Press.
- Dionne G., P. St-Michel and C. Vanasse** (1992), Moral Hazard, Optimal Auditing and Worker's Compensation, *Ninth Annual Seminar "Economic Issues in W.C."*
- Florida Insurance Research Center** (1991), Automobile Insurance Fraud Study, University of Florida, *Florida Insurance News Service*.
- Laffont J.J.** , *Economie de l'incertain et de l'information*, Chap. 10
- Mookherjee D and I. Png** (1989), Optimal Auditing Insurance and redistribution, *Quarterly Journal of Economics*
- Rothschild M. and J. Stiglitz** (1976), Equilibrium in competitive insurance markets: an essay on the economics of imperfect information, *Quarterly Journal of Economics*, vol. 80.
- Townsend R.** (1979), Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory*
- Townsend R.** (1988), Information Constrained Insurance: The Revelation Principle Extended, *Journal of Monetary Economics*
- Winter R.A.** (1991), Moral Hazard in Insurance Contracts, *Contributions to Insurance Economics*, Kluwer Academic Press.

Modèle statistique

Binet, F.E. (1954), The Fitting of the Positive Binomial Distribution When Both Parameters are Estimated From the Sample, *Annals of Eugenics*

Fisher R.A. (1941, 1942), The Negative Binomial Distribution, *Annals of Eugenics*

Hoerl A.E. and R.W. Kennard (1970), Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*

Olkin I. and A.J. Petkau and J.V. Zidek (1981), A Comparison of n Estimators for the Binomial Distribution, *Journal of the American Statistical Association*

Schumacher P. and J.V. Zidek (1993), Using Prior Information in Designing Intervention Detection Experiments, *The Annals of Statistics*

Simpson D.G. (1987), Minimum Hellinger Distance Estimation for the Analysis of Count Data, *Journal of the American Statistical Association*

Winkelmann R. (1994), Count data models, *Lecture Notes in Economics and Mathematical Systems*, 410, Springer-Verlag, Berlin

Olkin

```

V Olkin;I;LINE;STOP;FRAUD;STOPI;CIE;FR;FR123;TOT;IND1;NB;RAN;SUSP;E1;V1P1;N1
;E2;P2;V2;N2;PSI1;Z1;SIG1;PSI2;Z2;SIG2;NEUF;HUIT;SEPT;SIX;CINQ;QUATRE;TROI
S;ZERO;TOUT;g;e
[1]
[2] A Programme pour sélectionner et traiter les données de fraude
[3] A Modifications pour les bénéfices multiples
[4] A Division en 'g' groupes de 'e' elements
[5] A Suspicion traite comme D si > que 5,6,7... et N sinon
[6] J 1
[7] g 2773 ◇ e 1
[8] L1:
[9]
[10]
[11] STOP (DATAAX) [1] ◇ FRAUD 10 ◇ RES (0 g) ρ(10) ◇ SUSP 10
[12] I 1
[13]
[14] L2:
[15]
[16] LINE DATAAX [I;]
[17] CIE ϕLINE [24+12] ◇ →(CIE=99)/L3
[18] FR ϕLINE [146+19] ◇ SUS ϕLINE [330+29]
[19] IND1 (0≠FR)
[20] FR123 (IND1/FR)
[21] FRAUD FRAUD,FR123 ◇ SUSP SUSP, (IND1/SUS)
[22]
[23] L3:→(STOP≥I I+1)/L2
[24]
[25] 'Le nb de fraude est : '
[26] +/(FRAUD=3)
[27] 'Le nb de suspicions est : '
[28] +/(FRAUD=2)
[29] 'Le nb de non-fraude est : '
[30] +/(FRAUD=1)
[31]
[32] NB ρFRAUD ◇ RAN NB?NB
[33]
[34] A 'Le nb total de dossiers est : '
[35] A NB
[36]
[37] FRAUD FRAUD [RAN] ◇ SUSP SUSP [RAN]
[38]
[39] FRAUD (g e) ρFRAUD ◇ SUSP (g e) ρSUSP
[40]
[41] RES RES; (+/(FRAUD=1))
      ┌──────────┐
      │            │
[42] RES RES; (+/( (FRAUD=2) × (SUSP>9) ))
      ┌──────────┐
      │            │
      └──────────┘

```

Olkin

[43] RES RES; (+/ (FRAUD=3))

[44] RES RES; (+/ ((FRAUD=2) × (SUSP>8)))

[45] RES RES; (+/ ((FRAUD=2) × (SUSP>7)))

[46] RES RES; (+/ ((FRAUD=2) × (SUSP>6)))

[47] RES RES; (+/ ((FRAUD=2) × (SUSP>5)))

[48] RES RES; (+/ ((FRAUD=2) × (SUSP>4)))

[49] RES RES; (+/ ((FRAUD=2) × (SUSP>3)))

[50] RES RES; (+/ ((FRAUD=2) × (SUSP>0)))

Olkin

[51]

$$RES \text{ RES; } (+ / ((FRAUD=2) \times (SUSP \geq 0)))$$

[52]
 [53] A 'TADAM'
 [54] RES
 [55]
 [56] +/+ /RES
 [57]
 [58] 'Itération :'
 [59] J

[62] A 'A) Espérance ET variance du nb. de fraude détecté :'
 [63]

$$E1 \left(\frac{+/RES[3;]}{g} \right) \diamond V1 \left(\frac{+ / ((RES[3;] - E1) * 2)}{g} \right) \diamond E1, V1 \diamond SIG1 (V1) * (0.5)$$

[54] A 'B) Estimation de p ET de n'
 [65] P1 1 - V1 ÷ E1 \diamond \rightarrow (P1=0) / \square LC+1 \diamond N1 E1 ÷ P1 \diamond P1, N1

[66] A 'C) Estimation du pourcentage de fraude'
 [67] A (N1 ÷ 554) * 100
 [68] '*****'
 [69]

[70] NEUF RES [3;] + RES [2;]
 [71] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'
 [72]

$$E2 \left(\frac{+/NEUF}{g} \right) \diamond V2 \left(\frac{+ / ((NEUF - E2) * 2)}{g} \right) \diamond E2, V2 \diamond SIG2 (V2) * (0.5)$$

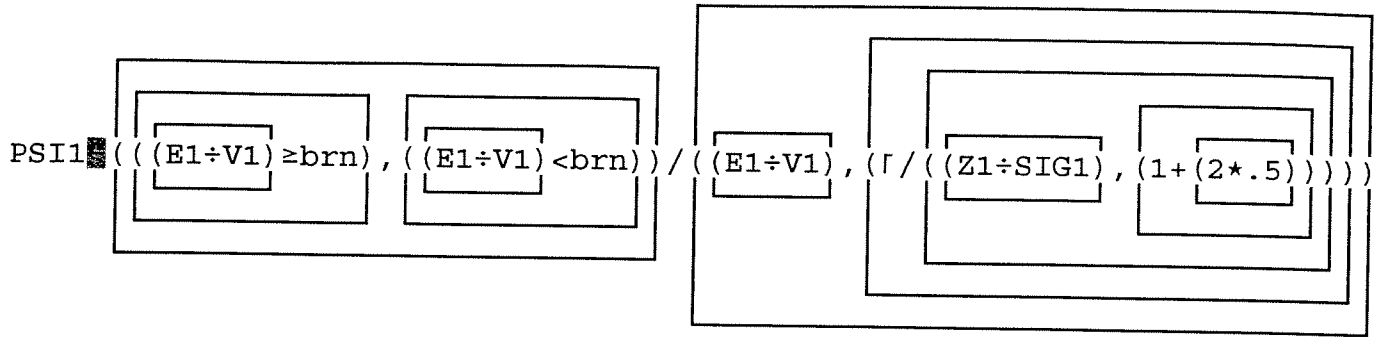
[73] A 'E) Estimation de p ET de n'
 [74] P2 1 - V2 ÷ E2
 [75] A 'C) Estimation du pourcentage de fraude'
 [76] A (N2 ÷ 554) * 100
 [77]
 [78] A Debut de la procedure d'Olkin, Petkau, Zidek
 [79]

[80] Z1 $\left(\left(\frac{+/RES[3;]}{g} \right) - E1 \right) \div SIG1$

[91]

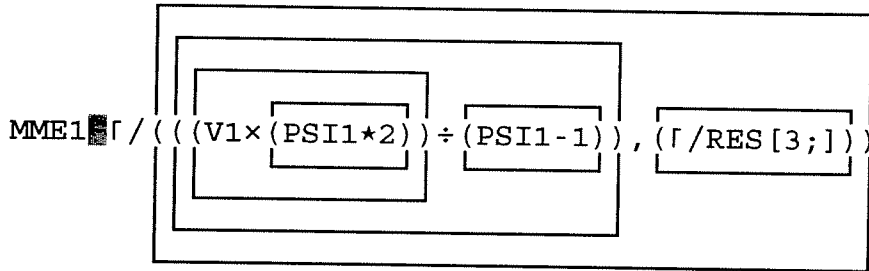
Olkin

[32]



[83]

[84]



[85]

[86]

[87] 'MME:S de la procedure d''Olkin, Petkau, Zidek (Sans Susp.)'

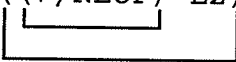
[88] MME1

'89] A 'Donc le n estime est:'

90] A MME1 ÷ 554

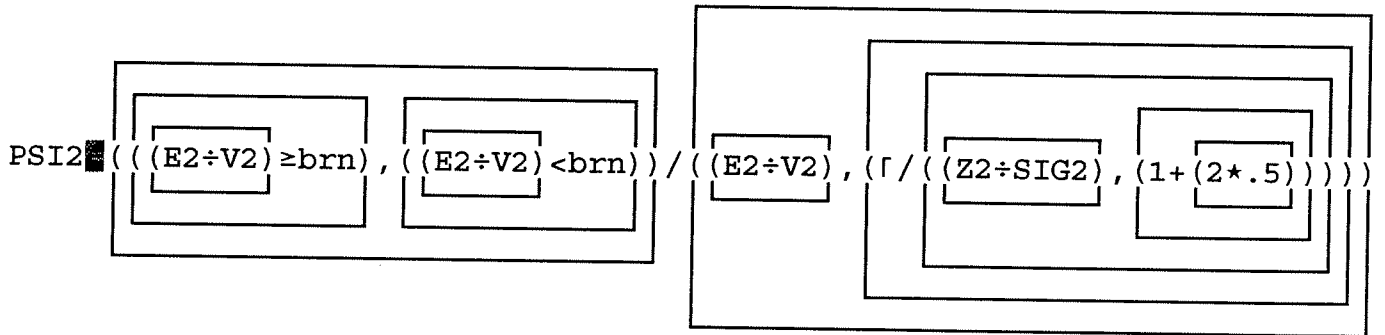
[91]

[92] Z2 $\left(\left(\Gamma / \text{NEUF} \right) - E2 \right) \div \text{SIG2}$



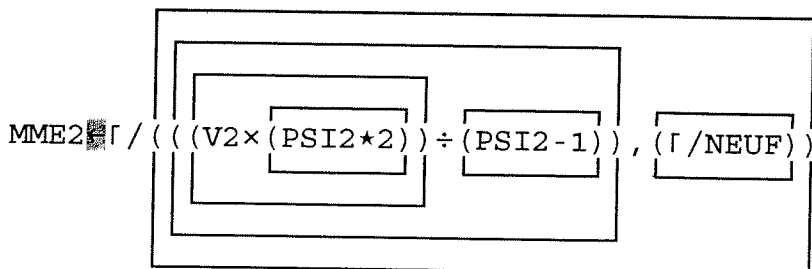
[93]

[94]



[95]

[96]



[97]

[98]

Olkin

[9] 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[100] MME2

[101]

[102] A A, MME1

[103] B B, MME2

[104]

[105] A*****

[106]

[107] HUIT RES [3;]+RES [4;]

[108] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[109]

$$E2 \left(\left(\frac{+}{\text{HUIT}} \right) \div g \right) \diamond V2 \left(\left(\frac{+}{\left(\text{HUIT} - E2 \right) * 2} \right) \right) \div g \diamond E2, V2 \diamond \text{SIG2} (V2) * (0.5)$$

[110] A 'E) Estimation de p ET de n'

[111] P2 1 - V2 ÷ E2

[112] A 'C) Estimation du pourcentage de fraude'

[113] A (N2 ÷ 554) × 100

[114]

[115] Z2 $\left(\left(\frac{\Gamma}{\text{HUIT}} \right) - E2 \right) \div \text{SIG2}$

[116]

$$\text{PSI2} \left(\left(\left(\frac{E2}{V2} \right) \geq \text{brn} \right), \left(\left(\frac{E2}{V2} \right) < \text{brn} \right) \right) / \left(\left(\frac{E2}{V2} \right), \left(\frac{\Gamma}{\left(\frac{Z2}{\text{SIG2}} \right)}, \left(1 + (2 * .5) \right) \right) \right)$$

[117]

$$\text{MME2} \left(\frac{\Gamma}{\left(\left(\frac{V2 * (\text{PSI2} * 2)}{\text{PSI2} - 1} \right) \right)}, \left(\frac{\Gamma}{\text{HUIT}} \right) \right)$$

[118]

[119]

[120] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[121] MME2

[122]

[123] C C, MME2

[124]

[125] A*****

[126]

[127] SEPT RES [3;]+RES [5;]

Olkin

[128] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'
[129]

$$E2 \left(\frac{+/SEPT}{g} \right) \diamond V2 \left(\frac{+ / ((SEPT-E2) * 2)}{g} \right) \diamond E2, V2 \diamond SIG2 (V2) * (0.5)$$

[130] A 'E) Estimation de p ET de n'

[131] P2 1-V2÷E2

[132] A 'C) Estimation du pourcentage de fraude'

[133] A (N2÷554)×100

[134]

[135] Z2 ((Γ/SEPT) - E2) ÷ SIG2

[136]

$$PSI2 \left(\left(\frac{E2}{V2} \geq brn \right), \left(\frac{E2}{V2} < brn \right) \right) / \left(\frac{E2}{V2}, \left(\frac{\Gamma}{Z2 \div SIG2}, (1 + (2 * .5)) \right) \right)$$

[137]

$$MME2 \left(\frac{\Gamma}{\left(\frac{V2 * (PSI2 * 2)}{PSI2 - 1} \right)}, \left(\frac{\Gamma}{SEPT} \right) \right)$$

[138]

[139]

[140] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[141] MME2

[142]

[143] D D, MME2

[144] A*****

[145]

[146] SIX RES[3;]+RES[6;]

[147] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[148]

$$E2 \left(\frac{+/SIX}{g} \right) \diamond V2 \left(\frac{+ / ((SIX-E2) * 2)}{g} \right) \diamond E2, V2 \diamond SIG2 (V2) * (0.5)$$

[149] A 'E) Estimation de p ET de n'

[150] P2 1-V2÷E2

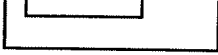
Olkin

[151] A 'C) Estimation du pourcentage de fraude'

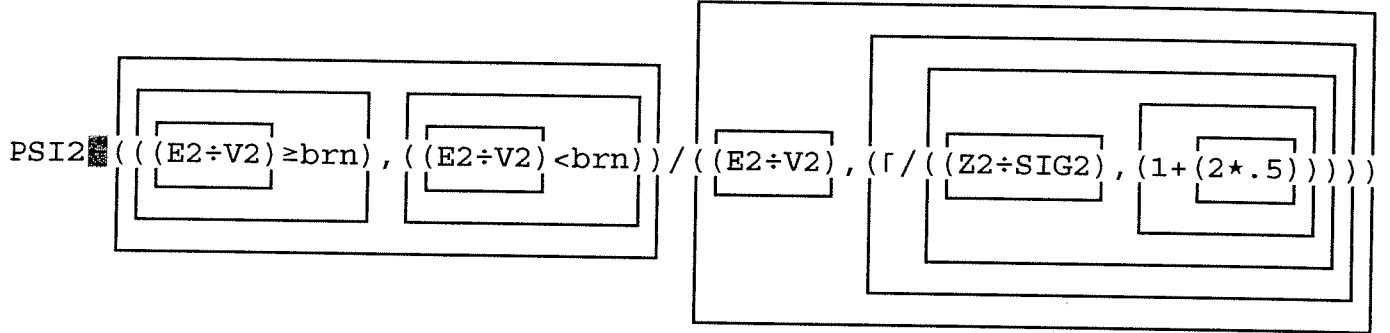
[152] A (N2÷554)×100

[153]

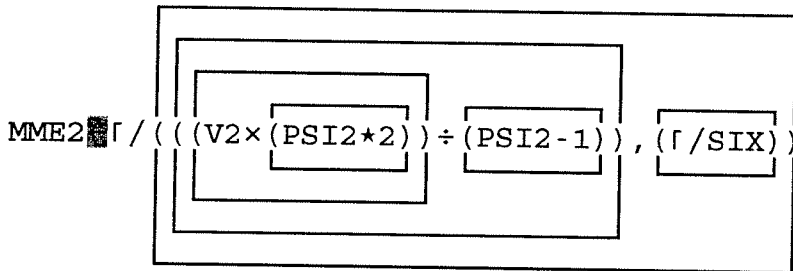
[154] Z2 (((r/SIX) - E2) ÷ SIG2



[155]



[156]



[157]

[158]

[159] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[160] MME2

[161]

[162] E E, MME2

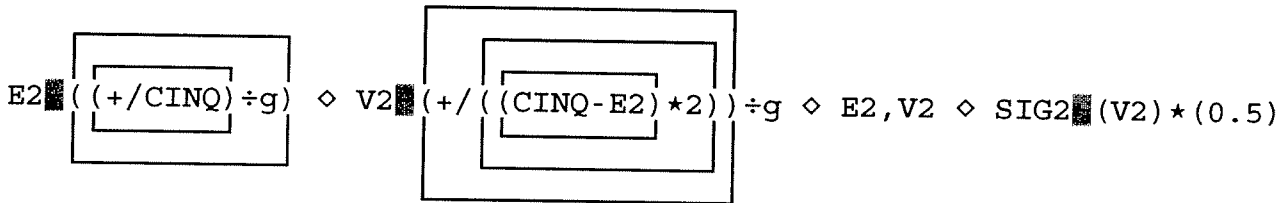
[163] A*****

[164]

[165] CINQ RES [3;] + RES [7;]

[166] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[167]



[168] A 'E) Estimation de p ET de n'

[169] P2 1 - V2 ÷ E2

[170] A 'C) Estimation du pourcentage de fraude'

[171] A (N2 ÷ 554) × 100

[172]

[173] Z2 (((r / CINQ) - E2) ÷ SIG2



Olkin

[174]

$$PSI2 = \left(\left(\frac{E2}{V2} \geq brn \right), \left(\frac{E2}{V2} < brn \right) \right) / \left(\frac{E2}{V2}, \left(\frac{\Gamma}{Z2 \div SIG2}, (1 + (2 * .5)) \right) \right)$$

[175]

$$MME2 = \frac{\Gamma}{\left(\left(V2 * (PSI2 * 2) \right) \div (PSI2 - 1) \right), \left(\frac{\Gamma}{CINQ} \right)}$$

[176]

[177]

[178] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[179] MME2

[180]

[181] F F, MME2

[182] A*****

[183]

[184] QUATRE RES [3;] + RES [8;]

[185] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[186]

$$E2 = \left(\frac{+}{QUATRE} \right) \div g \diamond V2 = \left(\frac{+}{(QUATRE - E2) * 2} \right) \div g \diamond E2, V2 \diamond SIG2 = (V2) * (0.5)$$

[187] A 'E) Estimation de p ET de n'

[188] P2 = 1 - V2 ÷ E2

[189] A 'C) Estimation du pourcentage de fraude'

[190] A (N2 ÷ 554) × 100

[191]

[192] Z2 = ((Γ / QUATRE) - E2) ÷ SIG2

Olkin

[193]

$$PSI2 \left(\left((E2 \div V2) \geq brn \right), \left((E2 \div V2) < brn \right) \right) / \left((E2 \div V2), \left(\Gamma / \left((Z2 \div SIG2), (1 + (2 * .5)) \right) \right) \right)$$

[194]

$$MME2 \left(\Gamma / \left((V2 * (PSI2 * 2)) \div (PSI2 - 1) \right), \left(\Gamma / QUATRE \right) \right)$$

[195]

[196]

[197] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[198] MME2

[199]

[200] G G, MME2

[201] A*****

[202]

[203] TROIS RES[3;]+RES[9;]

[204] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[205]

$$E2 \left((+/TROIS) \div g \right) \diamond V2 \left((+/((TROIS - E2) * 2)) \div g \right) \diamond E2, V2 \diamond SIG2 \left((V2) * (0.5) \right)$$

[206] A 'E) Estimation de p ET de n'

[207] P2 1 - V2 ÷ E2

[208] A 'C) Estimation du pourcentage de fraude'

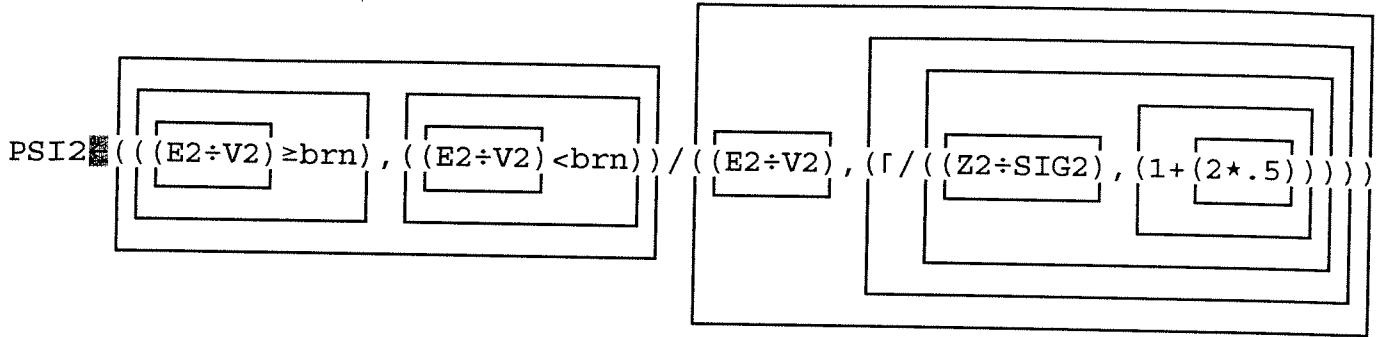
[209] A (N2 ÷ 554) * 100

[210]

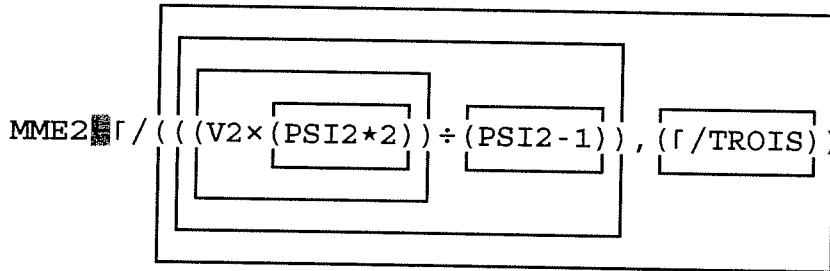
[211] Z2 ((Γ/TROIS) - E2) ÷ SIG2

Olkin

[12]



[213]



[214]

[215]

[216] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[217] MME2

[218]

[219] H H, MME2

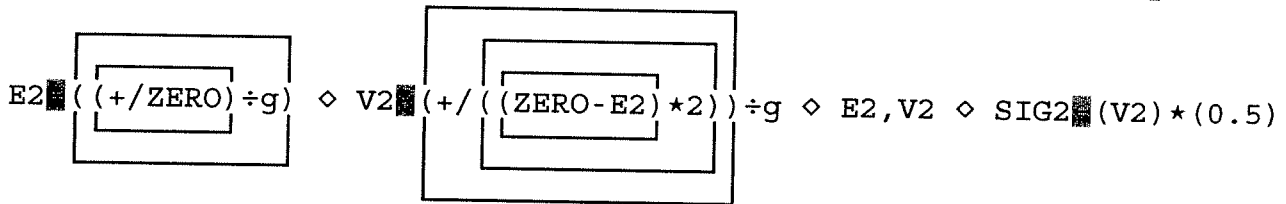
[220] A*****

[221]

[222] ZERO RES[3;]+RES[10;]

[223] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[224]



[225] A 'E) Estimation de p ET de n'

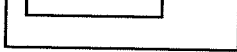
[226] P2 1-V2÷E2

[227] A 'C) Estimation du pourcentage de fraude'

[228] A (N2÷554)×100

[229]

[230] Z2 ((Γ/ZERO) - E2) ÷ SIG2



Olkin

(31]

$$PSI2 = \left(\left((E2 \div V2) \geq brn \right), \left((E2 \div V2) < brn \right) \right) / \left((E2 \div V2), \left(\Gamma / \left((Z2 \div SIG2), (1 + (2 * .5)) \right) \right) \right)$$

[232]

$$MME2 = \Gamma / \left(\left((V2 * (PSI2 * 2)) \div (PSI2 - 1) \right), \left(\Gamma / ZERO \right) \right)$$

[233]

[234]

[235] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[236] MME2

[237]

[238] K K, MME2

[239] A*****

[240]

[241] TOUT RES[3;]+RES[11;]

[242] A 'D) Espérance ET variance du nb. de fraude détecté ET suspecté :'

[243]

$$E2 = \left((+ / TOUT) \div g \right) \diamond V2 = \left((+ / ((TOUT - E2) * 2)) \div g \right) \diamond E2, V2 \diamond SIG2 = (V2) * (0.5)$$

[244] A 'E) Estimation de p ET de n'

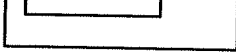
[245] P2 = 1 - V2 ÷ E2

[246] A 'C) Estimation du pourcentage de fraude'

[247] A (N2 ÷ 554) × 100

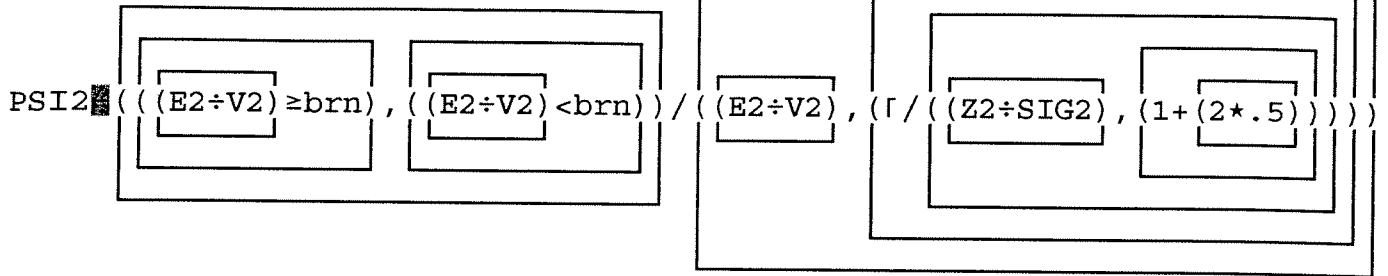
[248]

[249] Z2 = ((Γ / TOUT) - E2) ÷ SIG2

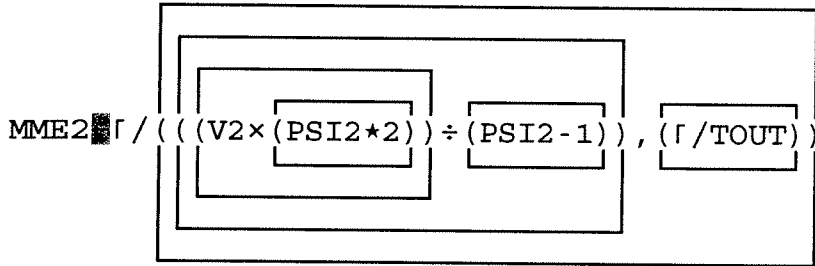


Olkin

[50]



[251]



[252]

[253]

[254] A 'MME:S de la procedure d''Olkin, Petkau, Zidek '

[255] MME2

[256]

[257] L L, MME2

[258] A*****

[259]

[260] L4: -> (1000 ≥ J J+1) / L1

[261]

[262]

[263] 'Estimes'

[264] (ρA), (ρB)

[265]

[266] +/A ÷ ρA

[267]

[268] +/B ÷ ρB

[269] +/C ÷ ρC

[270] +/D ÷ ρD

[271] +/E ÷ ρE

[272] +/F ÷ ρF

[273] +/G ÷ ρG

[274] +/H ÷ ρH

[275] +/K ÷ ρK

[276] +/L ÷ ρL

[277]

[278]