Optimal Auditing with Scoring Theory and Application to Insurance Fraud

Georges Dionne
Florence Giuliano
Pierre Picard

November 2005

Cahier n° 2005-037

Optimal Auditing with Scoring Theory and Application to Insurance Fraud

Georges Dionne[1]
Florence Giuliano[2]
Pierre Picard[3]

November 2005

Cahier n° 2005-037

Résumé: Cet article établit une liaison entre la théorie de l'audit optimal et la méthodologie du scoring dans un contexte d'asymétrie d'information. L'application retenue concerne la fraude à l'assurance, mais la même approche peut être appliquée à d'autres activités qui utilisent le scoring. Nous montrons que la stratégie et l'audit optimal consistent à transmettre les demandes d'indemnité à une cellule antifraude lorsque certains indicateurs de fraude sont observés. Les indicateurs de fraude sont classés en fonction d'une probabilité de fraude croissante. Une telle stratégie demeure optimale lorsque la politique d'enquête est budgétairement contrainte. De plus, la politique d'audit agit comme un mécanisme de dissuasion et nous expliquons pourquoi elle nécessite un engagement de l'assureur et comment elle devrait affecter les incitations financières du personnel de la cellule antifraude. Le modèle est calibré avec les données d'un grand assureur européen. Nous calculons une valeur critique d'un indice de suspicion de fraude qui fournit un seuil au delà duquel toutes les demandes d'indemnité doivent être soumises à audit et nous évaluons le gain potentiel qui pourrait être dérivé de la politique d'audit optimal. Nous montrons qu'il est possible d'améliorer ces résultats en séparant différents groupes d'assurés avec des coûts psychologiques de fraude différents. Enfin nos résultats montrent comment l'effet de dissuasion peut être pris en compte et comment il affecte la stratégie optimale d'audit.

Abstract: This article makes a bridge between the theory of optimal auditing and the scoring methodology in an asymmetric information setting. Our application is meant for asurance claims fraud, but it can be applied to many other activities that use the scoring approach. We show that the optimal auditing strategy takes the form of a "Red Flags Strategy" which consists in referring claims to a Special Investigative unit (SIU) when certain fraud indicators are observed. Fraud indicators are classified based on the degree to which they reveal an increasing probability of fraud. This strategy remains optimal even when the investigation policy is budget constrained. Moreover, the auditing policy acts as a deterrence device and we explain why it requires the commitment of the insurer and how it should affect the incentives of SIU staffs. the models is calibrated with data from a large European insurance company. We compute a critical suspicion index for fraud, providing a threshold above which all claims must be audited and we estimate the potential gain that could be derived from the optimal auditing policy. We show that it is possible to improve these results by separating different groups of insureds with different moral costs of fraud. Finally, our results indicate how the deterrence effect of the audit scheme can be taken into account and how it affects the optimal auditing strategy.

Mots clés : Audit, scoring, Fraude à l'assurance, Stratégie d'indicateurs, Indicateurs de fraude, Indice de suspicion, Coût moral de la fraude, Effet de dissasion.

Key Words : Audit, scoring fraud, Red flags strategy, Fraud indicators, Suspicion index, Moral cost of fraud, Deterrence effect.

Classification JEL: D81, G14, G22

---

[1] HEC Montréal.
[2] Université Paris X.
[3] Ecole Polytechnique et HEC School of Management.

Optimal Auditing with Scoring:
Theory and Application to Insurance Fraud

Georges Dionne, Florence Giuliano and Pierre Picard

This version: November 8, 2005

Abstract

This article makes a bridge between the theory of optimal auditing and the scoring methodology in an asymmetric information setting. Our application is meant for insurance claims fraud, but it can be applied to many other activities that use the scoring approach. We show that the optimal auditing strategy takes the form of a "Red Flags Strategy" which consists in referring claims to a Special Investigative Unit (SIU) when certain fraud indicators are observed. Fraud indicators are classified based on the degree to which they reveal an increasing probability of fraud. This strategy remains optimal even when the investigation policy is budget constrained. Moreover, the auditing policy acts as a deterrence device and we explain why it requires the commitment of the insurer and how it should affect the incentives of SIU staffs. The model is calibrated with data from a large European insurance company. We compute a critical suspicion index for fraud, providing a threshold above which all claims must be audited and we estimate the potential gain that could be derived from the optimal auditing policy. We show that it is possible to improve these results by separating different groups of insureds with different moral costs of fraud. Finally, our results indicate how the deterrence effect of the audit scheme can be taken into account and how it affects the optimal auditing strategy.

1. Introduction

Auditing has been a major topic of interest in the economic and financial literature since the path-breaking articles published by Townsend (1979) and Gale and Hellwig (1985). It is indeed widely accepted that the prevalence of auditing arises from the informational asymmetries between principals (bankers, insurers, regulators, tax inspectors…) and agents (borrowers, insureds, regulated firms, tax payers…), asymmetries which lead to implement costly state-verification strategies. The trade-off between monitoring costs and mitigating informational asymmetries between principal and agent is in fact the core of the economic analysis of auditing.

On the empirical side, the importance of auditing in corporations, financial institutions or governmental agencies has given rise to serious analysis of the design of optimal auditing procedures (e.g. Should auditing be internal or external? How should auditors be rewarded? How can collusion between auditors and those audited be avoided? How frequent should auditing be?…) and has motivated firms and governments to search for relevant information on ways to cut auditing costs. Nowadays, the search for optimal auditing procedures is a major concern for a number of players: banks and insurance companies seeking better risk assessment of their customers; prudential regulators of the banking and insurance industries; governments pursuing better compliance by taxpayers; and regulatory agencies in the field of environmental law, food safety or working conditions.

On the theoretical side, many extensions of the basic models have been proposed. In particular, Townsend (1988) and Mookherjee and Png (1989) have shown that random auditing dominates deterministic models. Among many other issues, the effect of collusion between auditees and auditors (e.g. between a tax inspector and a

tax payer or between a manager and an internal auditor) or between auditees and third parties (such as a health care provider in the case of an insurance claim) has received special attention (Kofman and Lawarrée, 1993). The consequences of commitment vs. no-commitment assumptions in an auditing procedure have also been examined with the analytical tools of modern incentives theory (Graetz, Reinganum and Wilde, 1986; Melumad and Mookherjee, 1989).

In such an asymmetric information setting, it is in the interest of principals to use signals on agents' types or actions when deciding whether a costly verification should be performed. Scoring techniques are then most useful since they help to identify suspicious files to be audited as a priority by associating scores — i.e. numerical valuations — to an unobservable default risk, expected profitability or residual value. It is now widely used by banks in credit-risk management[1], by corporations in hiring decisions, by tax authorities in tracking tax compliance, by insurers in detecting claims fraud and even by police officers in airport passengers screening procedures[2].

Although scoring and auditing are complementary ways to reduce the asymmetry of information between principals and agents, almost all theoretical approaches have completely ignored the role od scoring techniques in the design of an optimal auditing strategy[3]. This paper tries to fill this gap by bringing together auditing and scoring within a unified costly state verification approach and by applying the model to insurance claim fraud. In doing so, we shall build a model of optimal auditing which is much more closely related to auditing procedures actually used by insurers (or other principals) than abstract auditing models. Scoring will signal whether an audit should be performed or not, depending on the signals perceived by the insurer: This will be called a red flags strategy. In fact, we shall build a model where the optimal auditing strategy actually takes the form of such a red flags strategy. Though designed to audit insurance claims, it will appear clearly that our approach can be used for many other activities that apply the scoring techniques.

Insurance fraud provides a fascinating case study for the theory of optimal auditing, and particularly for connecting the scoring methodology with the theory of

optimal costly state verification. In recent years, the economic analysis of insurance fraud has developed along two branches. The first branch is mostly theoretical and its foundations may be found in the theory of optimal auditing. It aims at analyzing the strategy adopted by insurers faced with claims or application fraud.[4] This approach focuses mainly on questions such as: What should be the frequency of claim auditing and how do opportunistic policyholders react to the auditing strategy? What are the consequences of potential fraud on the design of insurance contracts, especially with regard to the indemnity schedule? What is the deterrence effect of an auditing policy? What is the role of good faith when insurance applicants may misrepresent their risk? The most usual setting for this literature is a costly state-verification model in which insureds have private information about their losses and insurers can verify claims by incurring an audit cost. Important assumptions are made relative to the ability of insurers to commit to an auditing policy and to the skill of defrauders in manipulating audit costs, i.e. to make the verification of claims more difficult.[5]

The second branch of the literature on insurance fraud is more statistically based: It focuses mainly on the significance of fraud in insurance portfolios, on the practical issue of how insurance fraud can be detected, and on the scope of automated detection mechanisms in lowering the cost of fraudulent claims.[6] The scoring methodology is one of the key ingredients in this statistical approach. Two of the key questions in this literature are: How do insurers actually react to fraud indicators (the so-called red flags) and how can automated early detection of fraud be performed by relying on fraud scores. As shown by Derrig (2002) and Tennyson and Salsas-Forn (2002), when there is suspicion of fraud, claims are usually handled with a two-stage procedure: After careful examinations, the claim is either paid under routine settlement or subjected to more intensive investigation. This investigation may take different forms: referral to a Special Investigative Unit (SIU), request for recorded or sworn statements from the claimant, the policyholder or a witness to the accident, on site investigation, etc. Furthermore, the reaction to red flags may vary depending on the particulars of policyholders. Developing automated methods (via the scoring approach) capable of

using the informational content of red flags as efficiently as possible is currently the subject of intense research by some insurance companies, particularly in the automobile insurance sector.

Our approach reconciles an important result of optimal audit theory with the widespread practice of insurance companies. Indeed, on one hand, theory predicts that an optimal auditing strategy should be random. In other words, claims should be audited with probability less than one. Indeed, filing fraudulent claims is a strictly dominated strategy for opportunistic policyholders if claims are always audited (assuming that audit allows the insurer to detect fraud with certainty). In such a case, slightly decreasing the audit probability would allow insurers to reduce their audit costs without inciting policyholders to fraud. On the other hand, factual evidence suggests that insurers tend to systematically audit certain claims and to directly pay the other ones, depending on the available information on policyholders' and on the perceived red flags. Our model will show that such a red flag strategy is in fact the key ingredient of a random auditing strategy, when individuals cannot control the fraud signals perceived by the insurers. From the insurers' standpoint, the auditing strategy is deterministic (the optimal audit decision is actually a non-random function of red flags) but it is random for the policyholders.

We will also derive some consequences of our results for the implementation of the optimal auditing strategy. Firstly, it will be shown that such a strategy requires a strong commitment from insurance companies. More explicitly, the optimal strategy *does not* amount to perform an audit if and only if the expected benefits of auditing (i.e. the expected value of possibly detected fraudulent claims) exceed the audit cost. Indeed, claims are verified in order to detect fraud but auditing also acts as a deterrence device. The deterrence objective requires auditing some claims although the investigation cost may exceed the expected benefits that the insurer may get from verification. As a consequence, the SIU should not be managed as a profit center of the insurance company. Secondly, we will show that the optimal proportion of successful audits depends on the policyholders' type (more precisely on the

4

particulars that can be observed by the insurer)[7], which affects the incentive scheme that should be offered to the insurance staffs in charge of verifying claims.

In the empirical part of the paper, we will calibrate our model by using data on automobile insurance (theft and collision) from a large European insurance company and we will derive the optimal auditing strategy. As a final outcome, our analysis yields an easily automated procedure which may be viewed as a prototype for the kind of insurance fraud detection mechanisms that are now more and more used by large insurance companies.

Section 2 presents the theoretical model while Section 3 derives the optimal auditing strategy. The application of the model to the portfolio of an insurer begins with Section 4 where the regression analysis is presented. The Section 5 outlines the model calibration and its results. Section 6 concludes. Proofs, data set presentation, and variables description are in the appendices.

2.   The model

We consider a population of policyholders who differ from one another in the moral cost of filing a fraudulent claim. For the sake of notational simplicity, all individuals own the same initial wealth and they all face the possibility of a monetary loss L with probability π with $0 < \pi < 1$. We simply describe the event leading to this loss as an "accident".

All individuals are expected utility maximizers and they display risk aversion with respect to their wealth. Let u be the state dependent utility of an individual drawn from this population. u depends on final wealth W but it also depends on the moral cost incurred in case of insurance fraud: $u = u (W, \omega)$ in case of fraud and $u = u (W, 0)$ otherwise, where ω is a non-negative parameter which measures the moral cost of fraud to the policyholder. We assume $u_1' > 0, u_{11}'' < 0$ and $u_2' < 0$ and that ω is distributed over R_+ among the population of policyholders. In other words, individuals

5

who choose to defraud incur more or less high moral costs. Some of them are purely opportunistic (their moral cost is very low) whereas others have a higher sense of honesty (their moral cost is thus higher). Note that moral cost is private information held by the insured: it cannot be observed by the insurer.

All the individuals in the insurer portfolio have taken out the same insurance contract. This contract specifies a level of coverage t in case of an accident and a premium that should be paid to the insurer. If there is no fraud, we have $W = W_0 - L + t$ in case of an accident and $W = W_0$ if no accident occurs where W_0 is the initial wealth net of the insurance premium.

Each individual in the population is characterized by a vector of observable exogenous variables θ, with $\theta \in \Theta \subset \mathsf{R}^m$. The moral cost of fraud may be statistically linked to some of these variables. Let $H(\omega|\theta)$ be the conditional, cumulated distribution of ω over the population of type-θ individuals, with a density $h(\omega|\theta)$.

Our model describes insurance fraud in a very crude way. A defrauder simply files a claim to receive the indemnity payment t although he has not suffered any accident. If a policyholder is detected to have defrauded, he will receive no insurance payment and must in addition pay a fine B to the government.[8] Let Q^f be the probability that a fraudulent claim is detected; this probability is the outcome of the insurer's antifraud policy and it depends on the observable variables θ as we shall see in Section 3.

When an individual has not suffered any loss, his utility is written as $u(W_0, 0)$ if he does not defraud. If he files a fraudulent claim (i.e. if he claims to have suffered an accident although this is not true), his final wealth is: $W = W_0 + t$ if he is not detected and $W = W_0 - B$ if he is detected. An individual with moral cost ω decides to defraud if he expects greater utility from defrauding than staying honest, which is written as:

$$(1 - Q^f)\, u(W_0 + t, \omega) + Q^f u(W_0 - B, \omega) \geq u(W_0, 0).$$

This inequality holds if $\omega \leq \phi(Q^f)$, where function $\phi(p): [0,1] \to \mathbb{R}^+$ is implicitly defined by:

$$(1-p)\, u(W_0 + t, \phi) + pu\,(W_0 - B, \phi) = u(W_0, 0) \quad \text{if } 0 \leq p \leq p_0$$

and $\phi(p) = 0$ if $p_0 \leq p \leq 1$ where p_0 is given by

$$(1 - p_0)\, u\,(W_0 + t, 0) + p_0\, u\,(W_0 - B, 0) = u\,(W_0, 0)$$

with $0 < p_0 < 1$ if $B > 0$ and $p_0 = 1$ if $B = 0$. p_0 is the audit probability that deters an individual with no moral cost ($\omega = 0$) from defrauding. We have $\phi(0) > 0$, $\phi(p_0) = 0$ and $\phi'(p) < 0$ if $0 < p < p_0$. $\phi(Q^f)$ is the critical value of the moral cost under which cheating overrides honesty as a rule of behavior. When $Q^f < p_0$, the higher the probability of being detected, the lower the threshold of the moral cost and thus the lower the frequency of fraud. When $Q^f \geq p_0$, there is no more fraud.

When a policyholder files a claim — be it honest or fraudulent — the insurer privately perceives a k-dimensional signal σ that can take on anyone of ℓ possible configurations. We assume:

$$\sigma \in \{\sigma_1, \sigma_2, \ ..., \ \sigma_\ell\} = \Sigma$$

with $\sigma_i \in \mathbb{N}^k$, $k \geq 1$ for all $i = 1,...,\ell$. Hereafter, k will be interpreted as the number of fraud indicators (or red flags) privately observed by insurers. Fraud indicators are claim-related signals that cannot be controlled by the defrauder and that should make the insurer more suspicious[9]. If indicator j takes N_j possible values[10] – say $0,1,...,N_j-1$ –, we have $\ell = \displaystyle\prod_{j=1}^{k} N_j$. When all indicators are binary (i.e. when $N_j = 2$ for all $j = 1,...,k$), then $\ell = 2^k$ and σ is a vector of dimension k all of whose components are 0 or 1: component j is equal to 1 when indicator j is "on" and it is equal to 0 when it is "off".

Let p_i^f and p_i^n be, respectively, the probability of the signal vector σ taking on configuration σ_i when the claim is fraudulent and when it corresponds to a true accident (non-fraudulent claim), i.e.:

$$p_i^f = P(\sigma = \sigma_i \mid F)$$

$$p_i^n = P(\sigma = \sigma_i \mid N)$$

with $i = 1,...,\ell$, where F and N refer respectively to "fraudulent" and "non-fraudulent" and $P(\bullet)$ denotes probability[11]. Of course, we have:

$$\sum_{i=1}^{\ell} p_i^n = \sum_{i=1}^{\ell} p_i^f = 1.$$

The probability distribution of signals is supposed to be common knowledge to the insurer and to the insureds. For simplicity of notations, we assume $p_i^n > 0$ for all $i = 1,...,\ell$ and w.l.o.g. we rank the possible signals in such a way that[12]

$$\frac{p_1^f}{p_1^n} < \frac{p_2^f}{p_2^n} < ... < \frac{p_\ell^f}{p_\ell^n}$$

This ranking allows us to interpret $i \in \{1,...,\ell\}$ as an index of fraud suspicion. Let $P(F \mid \theta)$ be the proportion of fraudulent claims among the claims filed by type - θ individuals. Section 3 will show how $P(F \mid \theta)$ can be deduced from the insurer's auditing strategy. Bayes law shows that the probability of fraud conditional on signal σ_i and type θ is:

$$P(F \mid \sigma_i, \theta) = \frac{p_i^f \, P(F \mid \theta)}{p_i^f \, P(F \mid \theta) + p_i^n (1 - P(F \mid \theta))} \tag{1}$$

which is increasing with i. In other words, as index i increases so does the probability of fraud.

3. Auditing strategy

The insurer may channel dubious claims to a Special Investigative Unit (SIU) where they will be verified with scrupulous attention. Other claims are settled in a routine way. The SIU referral serves to detect fraudulent claims as well as to deter fraud. We assume for simplicity that an SIU referral always allows the insurer to

determine beyond the shadow of a doubt whether a claim is fraudulent or not. In other words, the SIU performs perfect audits. An SIU claim investigation costs c to the insurer with $c < t$.

The insurer's investigation strategy is characterized by the probability of an SIU referral, this probability being defined as a function of individual-specific variables and claim-related signals. Hence, we define an investigation strategy as a function q : $\Theta \times \Sigma \rightarrow [0,1]$. A claim filed by a type-θ policyholder is transmitted to the SIU with probability $q(\theta, \sigma)$ when signal σ is perceived.

Let $Q^f(\theta)$ – respectively $Q^n(\theta)$ – be the probability of an SIU referral for a fraudulent –non-fraudulent – claim filed by a type-θ individual. $Q^f(\theta)$ and $Q^n(\theta)$ result from the insurer's investigation strategy through:

$$Q^f(\theta) = \sum_{i=1}^{\ell} p_i^f \, q(\theta, \sigma_i) \tag{2}$$

$$Q^n(\theta) = \sum_{i=1}^{\ell} p_i^n \, q(\theta, \sigma_i) \tag{3}$$

In particular, a type-θ defrauder knows that his claim will be subjected to careful scrutiny by the SIU with probability $Q^f(\theta)$. The insurer knows that, given his investigation strategy, truthful claims filed by type-θ individuals are mistakenly channeled to the SIU with probability $Q^n(\theta)$.

An optimal investigation strategy minimizes the total expected cost of fraud over the whole population of insureds. Cost of fraud includes the cost of investigation in the SIU and the cost of residual fraud (fraudulent claims not detected). Let *IC* denote the expected investigation cost. A type-θ individual has an accident with probability π and in such a case his claim will be channeled to the SIU with probability $Q^n(\theta)$. If such an individual has not had an accident, he may decide to file a fraudulent claim, and he will actually do so if his moral cost ω is lower than $\phi(Q^f(\theta))$ which occurs with probability $H(\phi(Q^f(\theta)) | \theta)$. Hence, the expected investigation cost is:

$$IC = c\pi \, E_\theta \, Q^n(\theta) + c(1-\pi) \, E_\theta \, Q^f(\theta) \, H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right) \qquad (4)$$

where E_θ denotes the mathematical expectation operator with respect to the probability distribution of θ over the whole population of insureds.

Let RC be the cost of residual fraud, which corresponds to the cost of undetected fraudulent claims. We have:

$$RC = t(1-\pi) \, E_\theta \left(1 - Q^f(\theta)\right) H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right) \qquad (5)$$

Let $TC = IC + RC$ be the total expected cost of fraud. An optimal investigation strategy minimizes TC with respect to $q(.) : \Theta \times \Sigma \to [0,1]$ under the constraints

$$0 \le q(\theta, \sigma) \le 1 \quad \text{for all } (\theta, \sigma) \text{ in } \Theta \times \Sigma \qquad (6)$$

Such a strategy is characterized in the following proposition.

Proposition 1 : *An optimal investigation strategy is such that*

$$q(\theta, \sigma_i) = 0 \quad \text{if } i < i^*(\theta)$$
$$q(\theta, \sigma_i) \in (0,1] \quad \text{if } i = i^*(\theta)$$
$$q(\theta, \sigma_i) = 1 \quad \text{if } i > i^*(\theta)$$

where $i^*(\theta) \in \{1,...,\ell\}$ is a critical suspicion index that depends on the vector of individual-specific variables.

Proposition 1 says that an optimal investigation strategy consists in subjecting claims to an SIU referral when the suspicion index i exceeds the individual-specific threshold $i^*(\theta)$, or equivalently when the probability of fraud is larger than a type-dependent threshold. Hence the insurer plays a socalled red flags strategy: for some signals σ_i – those for which $i > i^*(\theta)$ – claims are systematically audited at SIU, while there is no audit when $i < i^*(\theta)$[13]. Furthermore the optimal red flags strategy is type-dependent : if σ_i is perceived with $i^*(\theta_0) < i < i^*(\theta_1)$, then the claim should be sent to SIU if it has been filed by a type-θ_0 individual but no special investigation should be performed for a type-θ_1 insured.

Let

$$\tau(Q,\theta) = (1-\pi)\ H\big(\phi(Q)\big|\theta\big)$$

and

$$\eta(Q,\theta) = -\frac{Q\phi'(Q)h\big(\phi(Q)\big|\theta\big)}{H\big(\phi(Q)\big|\theta\big)} > 0.$$

$\tau(Q,\theta)$ is the fraud rate, i.e. the average number of fraudulent claims for a type-θ insured, when the probability of being detected is equal to Q. Note in particular that $\tau(Q, \theta_0) < \tau(Q, \theta_1)$ for all Q, if moving from θ_1 to θ_0, shifts the distribution of ω in the first-order stochastic dominance direction. $\eta(Q, \theta_1)$ is the elasticity of the fraud rate (in absolute value), i.e. the percentage decrease in the fraud rate following a one percent increase in the probability of detection.

Proposition 2 : *Assume that $(1-Q)H\big(\phi(Q)\big|\theta\big)$ is convex in Q. If $\tau(Q, \theta_1) \geq \tau(Q, \theta_0)$ and $\eta(Q, \theta_1) \geq \eta(Q, \theta_0)$, with at least one strong inequality, then $Q^f(\theta_1) > Q^f(\theta_0)$, $Q^n(\theta_1) > Q^n(\theta_0)$ and $i*(\theta_1) \leq i*(\theta_0)$.*

$(1-Q)H\big(\phi(Q)\big|\theta\big)$ is the rate of undetected fraud among the type-θ individuals who have not suffered any accident. It is decreasing from $H\big(\phi(0)\big|\theta\big) > 0$ to 0 when Q goes from 0 to p_0 : in words, the larger the probability for a defrauder to be detected, the lower the residual fraud. The convexity assumption made in Proposition 2 conveys the decrease in the marginal deterrence effect when the fraud detection probability is increasing. Auditing will cut fraud costs all the more efficiently if the insured belongs to a group with a high fraud and/or elasticity rate, hence the statement in Propostion 2. Indeed, the higher the fraud rate, the greater the direct benefits auditing provides by detecting fraudulent claims, and the greater the elasticity of fraud, the greater the indirect deterrence effect. If the rate and elasticity of fraud are

higher for θ_1 than for θ_0, then claims should receive more scrutiny when they are filed by type-θ_1 rather than by type-θ_0 individuals.

The conditional probability of fraud $P\big(F\big|\sigma_i,\theta\big)$ is given by (1) with

$$P\big(F\big|\theta\big) = \frac{(1-\pi)H\big(\phi\big(Q^f(\theta)\big)\big|\theta\big)}{\pi + (1-\pi)H\big(\phi\big(Q^f(\theta)\big)\big|\theta\big)}. \tag{7}$$

When signal σ_i is perceived, the expected benefit of an SIU investigation is:

$$P\big(F\big|\sigma_i,\theta\big)t - c.$$

Proposition 3 shows that the optimal investigation strategy involves transmitting suspicious claims to the SIU in cases where the expected benefit of such a special investigation may be negative and consequently it highlights the importance of the insurer's commitment.

Proposition 3 : *The optimal investigation strategy is such that:*

$$P\Big(F\Big|\sigma_{i*(\theta)},\theta\Big)\,t < c \ \ \text{for all } \theta \text{ in } \Theta.$$

Since $P\big(F\big|\sigma_i,\theta\big)$ is increasing in i, Proposition 3 means that there exists $i**(\theta)$ larger than $i*(\theta)$ such that:

$$P\Big(F\Big|\sigma_{i**(\theta)},\theta\Big)t < c < P\Big(F\Big|\sigma_{i**(\theta)+1},\theta\Big)t.$$

Forwarding the claim to the SIU is profitable only if the suspicion index i is larger than $i**(\theta)$. Hence, it is optimal to channel the claim to the SIU when $i*(\theta) \le i \le i**(\theta)$, although in such a case the expected profit drawn from investigation is negative. Indeed the investigation strategy acts as a deterrent: It dissuades some insureds (those with the highest moral costs) from defrauding. Such a strategy involves a stronger investigation policy than the one that would consist in transferring a claim to the SIU when the direct monetary benefits expected from investigation are positive. A consequence of this result is that the SIU should not be

organized as a profit center of the insurance company, for otherwise its objective would be in conflict with the implementation of the optimal auditing strategy.

In practice, when ℓ is large, p_i^n and p_i^f are very small and we can write $Q^f(\theta)$ = $\lambda(i^*(\theta))$ and $Q^n(\theta) = \mu(i^*(\theta))$, where[14]

$$\lambda(i) = \sum_{j=i}^{\ell} p_j^f \tag{8}$$

$$\mu(i) = \sum_{j=i}^{\ell} p_j^n. \tag{9}$$

$\lambda(i)$ and $\mu(i)$ respectively denote the probability of channeling a fraudulent claim and a non-fraudulent claim to the SIU when the critical index of suspicion is i. $\lambda(i)$ and $\mu(i)$ are decreasing functions: In other words, the higher the index-of-suspicion threshold, the lower the probability of subjecting a claim (be it fraudulent or not) to special investigation by an SIU. Using Proposition 1 allows us to reduce the insurer's optimization problem to the choice of the type-dependent suspicion threshold: $i^*(\theta)$ minimizes the expected cost of fraud

$$c\pi\mu(i) + (1-\pi)H\big(\phi(\lambda(i))|\theta\big)\big(c\lambda(i) + t(1-\lambda(i))\big) \tag{10}$$

with respect to $i \in \{1,...,\ell\}$. For a type-θ individual, the expected cost attributable to fraud is the sum of $C^n(i) \equiv c\pi\mu(i)$ which is the expected investigation cost of non-fraudulent claims that are incorrectly referred to the SIU (i.e. the cost of type-2 errors) and of

$$C^f(\theta,i) = (1-\pi)H\big(\phi(\lambda(i))|\theta\big)\big(c\lambda(i) + t(1-\lambda(i))\big)$$

which is the expected cost of fraudulent claims. This cost includes the investigation cost of the claim channeled to the SIU and the cost of paying out unwarranted insurance indemnities. $\lambda(i)$ and $\mu(i)$ are decreasing functions, which implies that $C^n(i)$ and $C^f(\theta,i)$ are respectively decreasing and increasing with respect to i. The optimal investigation strategy trades off excessive auditing of non-fraudulent claims against

inadequate deterrence and detection of fraudulent claims. The optimal critical suspicion index $i^*(\theta)$ minimizes $C^n(i) + C^f(\theta, i)$ as represented in Figure 1.

(Figure 1 about here)

The optimal auditing policy is also illustrated in Figure 2. When i^* goes from ℓ to 1, $\mu(i^*)$ and $\lambda(i^*)$ are both increasing: $\mu(i^*)$ is the probability of transmitting a non-fraudulent claim to the SIU and may thus be considered as a false alarm rate. $\lambda(i^*)$ is a true alarm rate since it corresponds to the probability of transmitting a fraudulent claim to an SIU. In the literature on classification techniques, the locus $\left\{ \left(\mu(i^*), \lambda(i^*) \right), i^* = 1, ...\ell \right\}$ is known as the Receiver Operating Characteristic (ROC) curve; see Viaene, Derrig, Baesens and Dedene (2002). It allows us to visualize the performance of the signals in terms of fraud detection. Using the monotonicity of p_i^f / p_i^n with respect to i shows that the ROC curve is concave. The optimal auditing procedure minimizes the expected cost of fraud with respect to (μ, λ), under the constraint that (μ, λ) is on the ROC curve. Figure 2 shows the dependence of the optimal solution on the agent's type.

(Figure 2 about here)

In practice (and particularly for the calibration of real data), we may assume that the activity of the SIU is budget-constrained: Antifraud expenditures should be less than some (exogenously given) upper limit K, which gives the following additional constraint:

$$c\pi\, E_\theta Q^n(\theta) + c(1-\pi)\, E_\theta Q^f(\theta)\, H\!\left(\phi\!\left(Q^f(\theta)\right)\middle|\theta\right) \le K. \qquad (11)$$

An optimal investigation strategy then minimizes TC with respect to $q(\cdot): \Theta \times \Sigma \to [0,1]$ subject to (6) and (11). Proposition 4 shows that the qualitative

characterization of the antifraud policy is not affected by the addition of this upper limit on possible investigation expenditures.

Proposition 4 : *Propositions 1 and 2 are still valid when the investigation policy is budget constrained.*

Constraint (11) may also be interpreted as the consequence of a fixed number of investigators at SIU, each of them being able to audit at most a certain number of claims. K/c then corresponds to the maximum number of audits per policyholder.

Under the optimal decision rule, a claim with signal σ_i is audited when $i \geq i^*(\theta)$. For type-θ individuals, the hit rate (i.e. the proportion of successful audits for each investigator at SIU) is then:

$$\frac{P(F|\theta) Q^f(\theta)}{P(F|\theta) Q^f(\theta) + \left[1 - P(F|\theta)\right] Q^n(\theta)}.$$

We may check that this hit rate varies with θ. In other words, the optimal investigation policy does *not* equalize the probability of success (i.e. of catching defrauders) across individuals' types.

This can be illustrated through an example which is a simplified version of our model. Assume that there is a proportion $q(\theta)$ of potential defrauders among type-θ individuals and a proportion $1 - q(\theta)$ of honest persons. Honest policyholders never defraud because their moral cost of fraud is very large. Potential defrauders incur a moral cost of fraud ω, which is distributed according to the (type independent) cumulated distribution function $G(\omega)$, with density $g(\omega) = G'(\omega)$.[15] We then have $H(\omega|\theta) = q(\theta) G(\omega)$ and $h(\omega|\theta) = q(\theta) g(\omega)$.

Let $\lambda_\theta \equiv \lambda(i^*(\theta))$ and $\mu_\theta \equiv \mu(i^*(\theta))$. We may write the equation of the ROC curve as $\mu = f(\lambda)$, with $f(0) = 0$, $f(1) = 1$, $f' > 0$, $f'' > 0$.

15

When (11) is binding, the optimal audit strategy minimizes the expected cost of residual fraud subject to the constraint on the maximum number of audits. This is written as:

$$\text{Minimize } E_\theta \left\{ tq(\theta) G(\phi(\lambda_\theta))(1 - \lambda_\theta) \right\}$$

with respect to λ_θ, $\theta \in \Theta$, subject to:

$$E_\theta \left\{ \pi f(\lambda_\theta) + (1 - \pi) q(\theta) G(\phi(\lambda_\theta)) \lambda_\theta \right\} \leq \frac{K}{c} \qquad (12)$$

and $\mu_\theta = f(\lambda_\theta)$. The first-order optimality condition for this problem is written as:

$$tp(\theta) \left[g(\phi(\lambda_\theta)) \phi'(\lambda_\theta)(1 - \lambda_\theta) - G(\phi(\lambda_\theta)) \right]$$

$$+ \xi \left[\pi f'(\lambda_\theta) + (1 - \pi) q(\theta) g(\phi(\lambda_\theta)) \phi'(\lambda_\theta) \lambda_\theta + G(\phi(\lambda_\theta)) \right] = 0 \qquad (13)$$

where ξ is a Lagrange multiplier associated to constraint (12).

For type-θ policyholder, the hit rate is $\dfrac{1}{(1 + X_\theta)}$ with:

$$X_\theta = \frac{\left[1 - P(F|\theta) \right] Q^n(\theta)}{P(F|\theta) Q^f(\theta)} = \frac{\pi f(\lambda_\theta)}{(1 - \pi) q(\theta) G(\phi(\lambda_\theta)) \lambda_\theta} \qquad (14)$$

After some simple calculations, (13) and (14) give:

$$X_\theta = \frac{1}{\xi(1 - \pi) \gamma_\theta} \left[(1 - \xi(1 - \pi))(1 - \eta_\theta) + \frac{\eta_\theta}{\lambda_\theta} \right]$$

where

$$\eta_\theta \equiv \eta(\lambda_\theta, \theta)$$

and

$$\gamma_\theta = \frac{f'(\lambda_\theta) \lambda_\theta}{f(\lambda_\theta)}.$$

η_θ is the fraud elasticity for type-θ individuals and λ_θ is the elasticity of the ROC curve, i.e. it is the increase in the number of false alarms, following a one-percent increase in the number of true alarms.

16

Assume $f(\lambda) = \lambda^{\alpha}$ with $\alpha > 1$, which gives $\gamma_{\theta} = \alpha$ for all θ. Assume also $\eta_{\theta} = \eta$ for all θ. In other words, the fraud rate and the ROC curve have constant elasticity. We then have:

$$X_{\theta} = \frac{1}{\xi(1-\pi)\alpha} \left[\left(1 - \xi(1-\pi)\right)(1-\eta) + \frac{\eta}{X_{\theta}} \right] \qquad (15)$$

for all θ.

For all θ_0, θ_1 in Θ such that $q(\theta_1) > q(\theta_0)$, we have $\lambda_{\theta_1} > \lambda_{\theta_0}$ from Proposition 2, and thus $X_{\theta_1} < X_{\theta_0}$ from (15). The hit rate is then larger for type θ_1 than for type θ_0.

The fact that the optimal hit rate is type-dependent affects the incentive mechanism that should be used by the insurance company to stimulate the activity of its staffs at SIU. In particular, paying a constant bonus each time SIU staffs catch a defrauder whatever his type is *not* an optimal incentive mechanism. Indeed, such an incentive scheme would lead investigators to concentrate on the claims with the highest hit rates and to neglect the other claims. At equilibrium, the hit rate would be the same for all audited claims and it would be lower for non-audited claims, which would not be optimal. On the contrary, if SIU staffs receive a bonus $b_{\theta} = k(1 + X_{\theta})$ for any type-θ hit, with $k > 0$, and if they are risk neutral with respect to their global earnings, then their own financial interest will not be in conflict with the optimal audit policy of the insurance company. In other words, the lower the probability of success, the larger must be the bonus to the auditor when a defrauder is caught.

4. Regression Analysis

Data are drawn from the automobile claims (theft and collision) of a large European insurer. Some of these claims had been spotted as fraudulent by the investigators of the company, while other ones could be reasonably considered as

truthful. The econometric analysis allowed us to identify relevant fraud indicators that are correlated with the frequency of fraudulent claiming, i.e. signals or individual characteristics. For that purpose, we used the standard Logit model for binary choice.

In our modeling, an explanatory variable takes the value +1 when the indicator is present and the value –1 otherwise. There is no clear evidence that the Logit model is more appropriate than the Probit model for our purpose. Our choice was explained only by mathematical convenience (See Green, 1997, for a longer discussion).

Table 1 reports the regression results. A detailed description of the data and the variables are presented in the appendix. The first column (without θ variables) in Table 1 is limited to variables identifying fraud indicators (the so-called red flags). The notation q_j^d refers to variables j that were directly available in the data warehouse of the insurer. The notation q_j^p corresponds to variables j that required some searching in the paper files. All these variables are significant in explaining (positively) the probability that a file may contain either suspected or established fraud at a level of at least 95% (with one exception at 90%). The second column (with θ variables that represent characteristics of policyholders) yields similar results but takes into account two additional variables affecting the probability that a file is fraudulent. In connection with the theoretical part of the paper, these variables are used to approximate the individual private cost of fraud which includes a pure moral cost component but also a monetary cost component. For the sake of brevity, we here restrict attention to two significant variables: q_7^p and q_{16}^p respectively indicate owners of vehicles whose value does not match the policyholder's income and which are not covered by damage insurance. Implicitly, it is suggested either that such people have a lower moral cost or that they draw a larger monetary benefit from fraud, hence a higher probability for filing a fraudulent claim.

(Table 1 about here)

18

5. Model Calibration

Methodology

Our sample didn't include a realistic proportion of fraudulent claims. Having the true proportion of fraudulent claims is not important for regression analysis but it is essential for the calibration of the optimal auditing model. We used bootstrapping techniques in order to obtain a final sample representing the estimated proportion of fraudulent claims in the company's portfolio (8%). Details are in Appendix A2.

Figure 3 presents the Gain Chart corresponding to the econometric model without the θ variables and with 8,400 files, each of them corresponding to a claim which may be fraudulent or non-fraudulent. The chart gives the percentage of detected fraud as a function of the audit probability, according to three different methods. The first one corresponds to a random sampling of the files and is illustrated by the 45º degree line: n% of the fraudulent claims will be captured if n% of the files are randomly sampled. The upper line corresponds to the performance of a "perfect expert" who would capture 100% of the fraudulent claims without any mistake. Such an expert would need to channel 8% of the files to the SIU in order to capture all the fraudulent claims. The line in the middle corresponds to the econometric model without the θ variables, where the fraud probabilities are estimated by the model. This method allows the insurer to transmit a small number of suspicious files to the SIU and to detect a significant number of fraudulent claims. For instance, about 55% of the fraudulent claims are captured by the model, if we use the 8th percentile (i.e. 8% of files with the highest fraud probability) as a reference percentile. This may be considered a very good outcome, given that we used only thirteen variables. The score can be improved easily by adding variables in the θ vector. Of course, it is not necessarily optimal to stop at the 8th percentile. The decision must trade off the

benefits and the costs of investigating the files. We now tackle the innovative part of the empirical analysis related to the calibration of the theoretical model.

(Figure 3 about here)

Let $\hat{\pi}(\theta)$ be the probability that a type-θ individual files a claim during a one-year time period when there is no auditing (which is supposed to correspond to the status quo situation in the insurance company) and let t be the average cost of a claim for the insurer (average amount paid above the deductible). Since in our model all the heterogeneity between insureds is related to the attitude toward fraud (i.e. to their moral costs), t does not depend on θ. For the time being, we do not distinguish the groups of insureds: a type-θ individual is thus a representative policyholder of the company. Data from the company give $\hat{\pi}(\theta) = 22\%$ and $t = €1,284$. The audit cost c of a claim is equal to €280 (including investigation costs, lawyers fees, SIU overheads, …) and we take the insurer's opinion for granted that the current proportion of claims with fraud is $z(\theta) = P(F \mid \theta) = 8\%$.

Since $\hat{\pi}(\theta)$ contains fraudulent claims, the true loss (theft and accident) probability π is given by $\pi = \hat{\pi}(\theta)\,(1 - z(\theta)) = 0.2024$. From the above data $\tau(0,\ \theta) = \hat{\tau}(\theta)$ can be approximated by $\hat{\tau}(\theta) = \hat{\pi}(\theta)\,z(\theta) = 0.0176$, which amounts to assuming that the observed current anti-fraud policy of the company does not entail any deterrence effect.

Estimating the elasticity of fraud with respect to the audit probability can only be a matter of approximation: Indeed, the elasticity $\eta(Q,\theta)$ depends on the distribution of moral costs in the population of policyholders as well as on the relationship between the audit probability, the moral cost, and the decision to file a fraudulent claim. In short, the elasticity of fraud with respect to Q depends on $H(\phi(Q)|\theta)$ and

20

θ. Such information is obviously unobservable. This is why we will content ourselves with an approximation of $\tau(Q,\theta)$. We will assume:

$$\tau(Q,\theta) = \hat{\tau}(\theta)(1-Q)^{\gamma(\theta)} \qquad (16)$$

where $\gamma(\theta)$ is a parameter used to define $\eta = \gamma(\theta) \, Q \, / \, (1-Q)$, the elasticity of the fraud rate with respect to Q. The expected cost of fraud can then be rewritten as:

$$c\pi\mu(i) + \hat{\tau}(\theta)(1-\lambda(i))^{\gamma(\theta)}(t - \lambda(i)(t-c)) \qquad (17)$$

where $\lambda(i)$ and $\mu(i)$ are given by (8) and (9).

As previously mentioned, in this part of our analysis θ corresponds to an average policyholder in the portfolio of the company. Of course, the analysis can be replicated for different values of θ, as we shall see in the last part of the article.

The optimal threshold $i*$ is obtained by minimizing (17) with respect to i. For that purpose, we must first compute the values of $\lambda(i)$ and $\mu(i)$ and thus p_i^f and p_i^n for all i. They are given by:

$$p_i^f = \frac{P(F|\sigma_i)P(\sigma_i)}{P(F)}$$

and

$$p_i^n = \frac{P(N|\sigma_i)P(\sigma_i)}{P(N)} = \frac{(1-P(F|\sigma_i))P(\sigma_i)}{(1-P(F))}.$$

The conditional probability $P(F|\sigma_i)$ could be computed directly from the econometric model. Unfortunately, $P(\sigma_i)$ is much more difficult to obtain directly: Indeed the econometric analysis yielded 13 significant binary fraud indicators and, consequently, 8,192 values for σ_i. Since our data set is limited to 8,400 observations or files, many potential values for σ_i should be nil. Using the econometric analysis to estimate p_i^f and p_i^n would then come to a deadlock. An indirect procedure can help us to escape from this difficulty. The procedure below is known in the literature as the

simple Bayes classifier method (Viaene et al., 2002) which is equivalent to the Bayes optimal classifier only when all predictors are independent in a given class. It has been shown that this simple Bayes classifier often outperforms more powerful classifiers (Duda et al., 2001).

From the regression analysis, we know that 13 fraud indicators are significant. q_j, $j = 1 \ldots k$, designates the presence $\left(q_j = 1\right)$ or absence $\left(q_j = 0\right)$ of indicator j in a given file. So we can write $\sigma_{ij} = 1$ if $q_j = 1$ and $\sigma_{ij} = 0$ if $q_j = 0$. Let $\alpha_j^f = P\left(q_j = 1 | F\right)$ and $\alpha_j^n = P\left(q_j = 1 | N\right)$ for $j = 1 \ldots k$, where $\alpha_j^f > \alpha_j^n$ by definition of fraud indicators. Let us assume that the q_j are independent conditional on the fact that the file is F or N^{16}. This conditional independence assumption allows us to write:

$$p_i^f = P\left(\sigma_i | F\right) = \underset{j / \sigma_{ij} = 1}{\Pi} \alpha_j^f \underset{j / \sigma_{ij} = 0}{\Pi} \left(1 - \alpha_j^f\right) \tag{18}$$

$$p_i^n = P\left(\sigma_i | N\right) = \underset{j / \sigma_{ij} = 1}{\Pi} \alpha_j^n \underset{j / \sigma_{ij} = 0}{\Pi} \left(1 - \alpha_j^n\right) \tag{19}$$

which allows us to compute $\lambda(i)$ and $\mu(i)$ for all i. We are now in a position to get a full calibration of our model.

Results

The calibration results are summarized in Table 2. Column 1 presents the identification numbers of the observed σ_i. They can simply be denoted by index i. One of them will also be the $i*$. Table 2 has $2^{13} = 8{,}192$ lines because the regression analysis identified 13 significant binary indicators. So we obtain 8,192 values for σ_i in Column 2 resulting from different combinations of N and Y where N indicates that an indicator is not present and Y indicates that an indicator is present for that line. For example, the first line in Column 2 indicates that no significant fraud indicator is present. Line 2 indicates that only the 9th fraud indicator is present. According to

Proposition 1, the optimal investigation strategy consists in ranking the observations i by using the values p_i^f / p_i^n in an increasing manner, which is done in Column 3.

(Table 2 about here)

Column 4 yields the value of $\lambda(i)$, the probability of channeling a fraudulent claim to the SIU when the critical index of suspicion is i. In line 1, $\lambda(i) = \mu(i) = 1$ and all claims are audited, be they fraudulent or not. Of course, this strategy would be very costly and as we shall see, it is not optimal. The optimal critical suspicion index, denoted i^*, trades off the benefits and the costs of auditing. If i^* were equal to 10, then all files with a ratio p_i^f / p_i^n higher than 0.20 would be audited: 94% of the fraudulent claims would be audited and 44% of the non-fraudulent claims would also be audited (see $\lambda(10)$ and $\mu(10)$ in columns 4 and 5).

Column 6 presents the expected investigation cost of a non-fraudulent claim $-\,C^n(i)\,-$ for different values of $\mu(i)$, the probability that a non-fraudulent claim will be channeled to the SIU. So for line one, we have $C^n(1) = €280 \times 0.2024 = €56.67$, since $\pi = 0.2024$ is the accident probability, $c = €280$ is the audit cost and $\mu(i) = 1$. For $i = 10$, this cost is reduced to €24.73 because $\mu(10) = 0.4364$. Column 7 yields the average cost of fraudulent claims for different values of $\lambda(i)$ and column 8 computes the expected cost of a fraudulent claim $-\,C^f(\theta,i)\,-$ for $\eta = 0$ and $\hat{\tau}(\theta) = 0.0176$. In line 1, this expected cost is very low because it is reduced to $€280 \times \hat{\tau}(\theta)$. Moreover, here $\gamma(\theta) = 0$ which means that there is no deterrence effect associated with a variation in $\lambda(i)$. Column 9 computes $C^n(i) + C^f(\theta,i)$. The optimal $i^*(\theta)$ is obtained by minimizing this expected cost. Finally, Columns 10, 11, and 12 respectively give information on the audit probability, on the expected audit cost and on the probability of fraud for audited claims. Again, if i^* were equal to 1, then all claims would be audited and the probability of fraud among audited claims would be equal to the

23

average fraud rate in the sample. If i^* were equal to 10, then the auditing strategy would be more focused on claims with suspected fraud ($\lambda(10) = 0.94$ and $\mu(10) = 0.44$) and the probability of fraud in audited claims would be equal to 0.1575.

The optimal solution is at line 194 = $i^*(\theta)$ (the corresponding values for the full random sampling are at the bottom of the table). We then have $\lambda(i^*) = 0.6672$, which means that 67% of the fraudulent claims are audited. So the optimal expected cost of a fraudulent claim in the total insurer portfolio (Column 8) is €10.81. We also have $\mu(i^*) = 0.04$, which means that only 4% of the non-fraudulent claims are audited. The corresponding optimal expected cost of a non-fraudulent claim in the insurer portfolio is equal to €2.40. So the optimal expected total cost of fraud reaches its minimal value at €13.20. The optimal strategy entails auditing 9.23% of the files (column 10) and the optimal audit cost per claim is €25.85. Finally, the hit rate is

$$P(F|i > i^*) = \frac{z(\theta)\lambda(i^*)}{z(\theta)\lambda(i^*) + (1 - z(\theta))\mu(i^*)} = 57.8\%$$

which means that 57.8% of audited claims prove to be fraudulent, which can also be illustrated in Figure 3 at the 9.23% value.

Tables 3 and 4 present sensitivity analyses with respect to the parameters $\gamma(\theta)$ and $z(\theta)$. In Table 2, we have indeed assumed that $z(\theta)$ is the fraud rate in the whole portfolio of the insurer and we have neglected the deterrence effect of the auditing policy, i.e. we have assumed $\gamma(\theta)=0$. Table 3 gives four different values of $z(\theta)$ computed by using the observable characteristics which have been highlighted in Table 1. In other words, q_7^p and q_{16}^p approximate different values of θ.

(Table 3 about here)

$z(\theta) = 5.83\%$ corresponds to the case were $q_7^p = q_{16}^p = 0$ which yields the lowest fraud rate in the portfolio. The other value of interest for the sensibility analysis is that

obtained when $q_{16}^p = 1$ and $q_7^p = 0$. The corresponding value for $z(\theta)$ is then 10.01%. The two other cases are not considered because their respective frequency is too low. Table 4 clearly shows that fraud auditing intensifies ($i^*(\theta)$ decreases and $\lambda(i^*(\theta))$ increases) as the fraud rate $z(\theta)$ increases, which is an illustration of Proportion 2.

(Table 4 about here)

Parameter $\gamma(\theta)$ measures the incentive effect of the optimal audit policy: the higher $\gamma(\theta)$, the higher the elasticity of fraud with respect to the detection probability and hence the more powerful the deterrence effect of auditing. In Table 4, we observe, for the three different values of the fraud rate $z(\theta)$, that $i^*(\theta)$ decreases (audit increases) when $\gamma(\theta)$ increases, which also illustrates Proportion 2.

Table 5 presents the monetary gains of auditing. As already mentioned, the claims rate (over the whole portfolio) of the insurer is 22%, which represents about 500,000 claims for the corresponding time period. Without the optimal audit policy, the fraud rate is 8%. So €51 million are paid for fraudulent claims and the total claim cost is €642 million. Let us now consider the optimal auditing policy of Table 2. First, we know that 9.23% of the files will be audited at a cost of €280. Secondly, we also know that 67% of the fraudulent claims will be audited and will not receive any insurance coverage. However, 33% of the fraudulent claims would not be audited. The total claim cost net of audit costs will then be equal to €621 M, a saving of €21 M, which represents 41% of the current cost of fraudulent claims. Finally, we show that auditing all claims is not efficient, as suspected. Indeed, auditing all claims would generate a total claim and audit cost of €731 M, the total claim cost is reduced to €591 M but the total audit cost is equal to €140 M.

(Table 5 about here)

The last line of Table 5 introduces the deterrence effect when $\eta(\theta) = 0.21$ which corresponds to $\gamma(\theta) = 0.10$. As shown in Proposition 4, because of this deterrence effect, under the optimal auditing policy, the expected benefit of an SIU investigation under the optimal auditing may be negative. More precisely, this expected benefit is negative when the suspicion index is in the neighborhood of $i*(\theta)$, while it is positive for suspicion indexes larger than $i**(\theta)$ with $i**(\theta) > i*(\theta)$.

This can be illustrated as follows with our data. The expected benefit of investigation is $P(F|\sigma_i,\theta) t - c$ where $P(F|\sigma_i,\theta)$ is given by (1), with

$$x = P(F|\theta) = \frac{\hat{\tau}(\theta)[1 - \lambda(i*(\theta))]^{\gamma(\theta)}}{\pi + \hat{\tau}(\theta)[1 - \lambda(i*(\theta))]^{\gamma(\theta)}}$$

For illustrative purpose, consider the case $\hat{\tau}(\theta) = 1.76\%$. If $\gamma = 0$, there is no deterrence effect which gives $P(F|\theta) = z(\theta) = 8\%$. Using $i*(\theta) = 194$ and equation (10) yields:

$$P(F|\sigma_{i*(\theta)},\theta) = 0.218 = \frac{c}{t}.$$

If $\gamma = 0.10$, the elasticity of fraud with respect to the probability of being detected is $\eta(\theta) = 0.21$ (line C in Table 4). We then have $i*(\theta) = 183$ and $P(F|\sigma_{i*(\theta)},\theta) = 0.189 < \frac{c}{t}$. In that case, the fraud rate is $P(F|\theta) = 7.2\%$ which illustrates the deterrence effect of the auditing policy, since the fraud rate would be 8% if no audit were performed or if an audit were performed without a deterrent effect. Hence €51 M is replaced by €46.22 M (i.e. 7.2% × €642 M) in the last line of Table 5 and the total claim cost net of audit costs is reduced to €619 M yielding a supplementary benefit to the audit policy.

26

VII Conclusion

This article aimed at making a bridge between the theory of optimal auditing and the actual claims auditing procedures used by insurers. More generally, we have developed an integrated approach to auditing and scoring which is much more closely related to the actual auditing procedures used by insurers, bankers, tax inspectors or governmental regulatory agencies than the abstract costly state-verification modelling. A complete modelling has been developed for the detection of insurance fraud, but the same methodology could be adapted to other hidden information problems, particularly those connected with banking, tax compliance or with the regulation of productive or financial activities by governmental or international agencies.

On the theoretical side, we have shown that the optimal auditing strategy takes the form of a type dependent red flags strategy which consists in referring claims to the SIU when some fraud indicators are observed. The classification of fraud indicators corresponds to an increasing order in the probability of fraud and such a strategy remains optimal if the investigation policy is budget constrained. Furthermore, the auditing policy acts as a deterrence device and in some cases, the (unconstrained) optimal investigation strategy leads to an SIU referral even if the direct expected gain of such a decision is negative. A strong commitment of the firm is thus necessary for such a policy to be fully implemented. Finally the optimal hit rate depends on the policyholder's type, which affects the optimal incentive mechanism of SIU staffs.

On the empirical side, several significant results were tested with data from a large European insurance company. We have computed a critical suspicion index for fraud, providing a threshold above which all claims must be audited. If the insurer implements this policy, 67% of the fraudulent claims are audited while only 4% of the non fraudulent claims are audited. We have shown that if the insurer applies this policy, he will save more than €21 M (net of audit costs), while he was paying €51 M

for fraudulent claims. These results were obtained under the conservative scenario that all policyholders share the same moral cost of fraud and that auditing does not involve any deterrence effect.

We have shown that these results are improved by using information on observable variables which are correlated with the intensity of fraudulent claiming. The optimal expected audit probability goes from 6.7% to 12.3% when the fraud rate goes from 5.8% (for a low fraud type) to 10.0% (for a high fraud type) which suggests that strongly differentiated audit rates are actually optimal. Finally, our results show how the deterrence effect of the audit scheme can be taken into account and how it affects the optimal auditing strategy.

References

Artis, M., Ayuso, M., and Guillén, M. (2002), "Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims," *Journal of Risk and Insurance* 69, 325-340.

Belhadji, E.B., Dionne, G., and Tarkhani, F. (2000), "A Model for the Detection of Insurance Fraud," *Geneva Papers on Risk and Insurance Issues and Practice* 25, 517-538.

Boyer, M.M. (2004), "Overcompensation as a Partial Solution to Commitment and Renegotiation Problems: The Case of *Ex Post* Moral Hazard," *The Journal of Risk and Insurance* 71, 559-582.

Crocker, K.J. and Morgan, R.J. (1997), "Is Honesty the Best Policy? Curtailing Insurance Fraud Through Optimal Incentive Contracts," *Journal of Political Economy* 106, 355-375.

Crocker, K.J. and Tennyson, S. (2002), "Insurance Fraud and Optimal Claims Settlement Strategies," *Journal of Law and Economics* 45, 469-507.

Derrig, R.A. (2002), "Insurance Fraud," *The Journal of Risk and Insurance* 69, 271-287.

Derrig, R.A. and Weisberg, H.I. (2003), "Auto Bodily Injury Claim Settlement in Massachusetts," Document, Automobile Insurers Bureau of Massachusetts, 36 p.

Dionne, G. (2000), "The Empirical Measure of Information Problems with Emphasis on Insurance Fraud," in *Handbook of Insurance*, G. Dionne (Ed.), Kluwer Academic Publishers, Boston, 395-419.

Dionne, G., Artis, M., and Guillén, M. (1996), "Count Data Models For A Credit Scoring System," *Journal of Empirical Finance* 3, 303-325.

Dionne, G. and Gagné, R. (2001), "Deductible Contracts Against Fraudulent Claims: Evidence from Automobile Insurance," *Review of Economics and Statistics* 83, 290-301.

Dionne, G. and Gagné, R. (2002), "Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance," *Journal of Risk and Uncertainty* 24, 213-230.

Duda, R.O., Hart, P.E., and Stork, E.G. (2001), "Pattern Classification," Wiley, New York.

Gale, D. and Hellwig M. (1985), "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies* 52, 647-663.

Graetz, M.J., Reinganum, J.F. and Wilde, L.L. (1986), "The Tax Compliance Game: Toward and Interactive Theory of Law Enforcement," *Journal of Law, Economics and Organization* 2, 1-32.

Green, W.H. (1997), *Econometric Analysis*, Prentice Hall, New Jersey, 1075 pages.

Knowles, J., Persico N. and Todd P. (2001), "Racial Bias in Motor Vehicle Searches: Theory and Evidence", Journal of Political Economy, 109, N°1, 203-229.

Kofman, F. and Lawarrée, J. (1993), "Collusion in Hierarchical Agency," *Econometrica* 61, 629-656.

Melumad, M.D. and Mookherjee, D. (1989), "Delegation as Commitment: The Case of Income Tax Audits," *The RAND Journal of Economics* 20, 139-163.

Moody's Investor Service 2000, RiskCalc For Private Companies:Moody's Default Model, New York.

Mookerjee, D. and Png, I.P.L. (1989), "Optimal Auditing Insurance and Distribution" *Quarterly Journal of Economics* 104, 399-415.

Persico, N. (2002), "Racial Profiling, Fairness, and Effectiveness of Policing", The American Economic Review, Vol. 92, 5, 1472-1497.

Picard, P. (1996), "Auditing Claims in Insurance Market with Fraud: the Credibility Issue," *Journal of Public Economics* 63, 27-56.

Picard, P. (1999), "On the Design of Optimal Insurance Contracts Under Manipulation of Audit Costs," *International Economic Review* 41, 1049-1071.

Picard, P. (2000), "Economic Analysis of Insurance Fraud," in *Handbook of Insurance*, G. Dionne (Ed.), Kluwer Academic Publishers, Boston, 315-362.

Tennyson, S. and Salmsa-Forn, P. (2002), "Claims Auditing and Automobile Insurance: Fraud, Detection and Deterrence Objectives," *The Journal of Risk and Insurance* 69, 289-308.

Townsend, R. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* XXI, 265-293.

Townsend, R. (1988), "Information Constrained Insurance: The Revelation Principle Extended," *Journal of Monetary Economics*, 21, 411-450.

United States General Accounting Office (2004), *Aviation Security: Computer-Assisted Passenger Prescreening System Faces Significant Implementations Challenges*, GAO-04-385.

Viaene, S., Derrig, R.A., Baesens, B. and Dedene, G. (2002), "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection," *Journal of Risk and Insurance* 69, 373-422.

Appendices

Appendix A1: Proofs

Proof of Proposition 1

If $Q^f(\theta) \geq p_0$, there is no residual fraud (hence RC = 0) and reducing $q(\theta,\sigma_i)$ for any i decreases IC, hence a decrease in TC. Thus the optimal investigation strategy involves $Q^f(\theta) < p_0$, with $\phi'\left(Q^f(\theta)\right) < 0$. Using equations (2) to (6), pointwise minimization of TC with respect to $q(\theta,\sigma)$ gives:

$$c\pi p_i^n + (1-\pi)\; A_1\left(Q^f(\theta),\theta\right) p_i^f \begin{cases} \leq 0 & if \;\; q(\theta,\sigma_i) = 1 \\ = 0 & if \;\; 0 < q(\theta,\sigma_i) < 1 \\ \geq 0 & if \;\; q(\theta,\sigma_i) = 0 \end{cases}$$

where $A(Q, \theta) = \left(cQ + t(1-Q)\right) H\left(\phi(Q)|\theta\right)$ and A_1 denotes the partial derivative of A with respect to Q. Note that $\phi' < 0$ and $t > c$ give $A_1 < 0$. Consequently, we have:

$$q(\theta,\sigma_i) = 1 \;\; if \;\; \frac{p_i^f}{p_i^n} > -\frac{c\pi}{(1-\pi)A_1\left(Q^f(\theta),\theta\right)}$$

and

$$q(\theta,\sigma_i) = 0 \;\; if \;\; \frac{p_i^f}{p_i^n} < -\frac{c\pi}{(1-\pi)A_1\left(Q^f(\theta),\theta\right)}$$

which proves the proposition, with $i^*(\theta)$ given by:

$$\frac{p_{i^*(\theta)-1}^f}{p_{i^*(\theta)-1}^n} < \frac{-c\pi}{(1-\pi)A_1\left(Q^f(\theta),\theta\right)} \leq \frac{p_{i^*(\theta)}^f}{p_{i^*(\theta)}^n} \tag{20}$$

Q.E.D.

32

Proof of Proposition 2

Assume that $(1-Q)H\big(\phi(Q)|\theta\big)$ is convex in Q. Let θ_0 and θ_1 in Θ such that $\tau(Q, \theta_1) \geq \tau(Q, \theta_0)$ and $\eta(Q, \theta_1) \geq \eta(Q, \theta_0)$ with at least one strong inequality. Assume moreover that $i*(\theta_1) > i*(\theta_0)$, which gives:

$$Q^f(\theta_1) < Q^f(\theta_0). \tag{21}$$

Let $i \in \{1,...\ell\}$ such that $i*(\theta_0) \leq i < i*(\theta_1)$. Writing optimality conditions as in the proof of Proposition 1 yields:

$$c\pi p_i^n + (1-\pi)A_1\big(Q^f(\theta_0),\theta_0\big)p_i^f \leq 0 \tag{22}$$

and

$$c\pi p_i^n + (1-\pi)A_1\big(Q^f(\theta_1),\theta_1\big)p_i^f \geq 0 \tag{23}$$

Using (21) and the convexity of $Q \to A(Q,\theta)$ gives:

$$A_1\big(Q^f(\theta_0),\theta_1\big) > A_1\big(Q^f(\theta_1),\theta_1\big). \tag{24}$$

(23) and (24) give:

$$c\pi p_i^n + (1-\pi)A_1\big(Q^f(\theta_0),\theta_1\big)p_i^f > 0. \tag{25}$$

(22) and (25) then imply:

$$A_1\big(Q^f(\theta_0),\theta_1\big) > A_1\big(Q^f(\theta_0),\theta_0\big). \tag{26}$$

We have

$$A_1(Q,\theta) = (c-t)H\big(\phi(Q)|\theta\big) + \phi'(Q)h\big(\phi(Q)|\theta\big)\big(cQ + t(1-Q)\big)$$

which may be rewritten as:

$$A_1(Q,\theta) = -\frac{\tau(Q,\theta)}{1-\pi}\left(t - c + \frac{cQ + t(1-Q)}{Q}\eta(Q,\theta)\right).$$

The assumptions on $\tau(Q, \theta)$ and $\eta(Q, \theta)$ give

$$A_1\big(Q^f(\theta_0),\theta_1\big) < A_1\big(Q^f(\theta_0),\theta_0\big)$$

which contradicts (26). Hence, we may conclude that $i*(\theta_1) \leq i*(\theta_0)$, which completes the proof. Q.E.D.

Proof of Proposition 3

We have:

$$A_1\left(Q^f(\theta),\theta\right) = \left(cq + t(1-Q)\right) \ h\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right)\phi'\left(Q^f(\theta)\right)$$
$$+ (c-t) \ H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right) < (c-t) \ H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right).$$

Hence

$$\frac{-c\pi}{(1-\pi) \ A_1\left(Q^f(\theta),\theta\right)} < \frac{-c\pi}{(1-\pi)(c-t)H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right)} \tag{27}$$

which gives:[17]

$$\frac{p^f_{i*(\theta)}}{p^n_{i*(\theta)}} < \frac{-c\pi}{(1-\pi)(c-t) \ H\left(\phi\left(Q^f(\theta)\right)\middle|\theta\right)} \tag{28}$$

Using (7), (28) and

$$\frac{p^f_{i*(\theta)}}{p^n_{i*(\theta)}} = \frac{\left(1 - P(F|\theta)\right) P\left(F\middle|\sigma_{i*(\theta)},\theta\right)}{P(F|\theta)\left(1 - P\left(F\middle|\sigma_{i*(\theta)},\theta\right)\right)}$$

gives $P\left(F\middle|\sigma_{i*(\theta)},\theta\right)t < c$. Q.E.D.

Proof of Proposition 4

Let α be a (non-negative) Kuhn-Tucker multiplier associated with (11) when TC is minimized with respect to $q(\theta,\sigma)$ subject to (6) and (11). Pointwise minimization gives:

$$c\pi(1+\alpha)p_i^n + (1-\pi)\tilde{A}_1\left(Q^f(\theta),\theta\right)p_i^f \begin{cases} \leq 0 & if \ q(\theta,\sigma_i) = 1 \\ = 0 & if \ 0 < q(\theta,\sigma_i) < 1 \\ \geq 0 & if \ q(\theta,\sigma_i) = 0 \end{cases}$$

where $\tilde{A}(Q,\theta) = \big(c(1+\alpha)Q + t(1-Q)\big)H\big(\phi(Q)|\theta\big)$ and $\tilde{A}_1$ is the partial derivative of $\tilde{A}$ with respect to Q. Proposition 4 can then be proved in the same way as Propositions 1 and 2. Q.E.D.

Appendix A2: Data

Estimation

The data come from a large insurer in Europe. We draw a sample from the automobile insurance claims files containing information on automobile thefts and collisions. Chart A1 presents the parameters of the original data set. The first group of files (A) comes from the company's SIU. This is the population of claims referred to this unit over a given period by claims handlers suspecting fraud. Of the 857 files referred to the SIU, 184 contained no fraud and 673 were classified as cases of either established or suspected fraud. As in Belhadji et al. (2000), we considered all these files as fraudulent because they all contained enough evidence of fraud to serve in designing a model for forwarding suspicious files to the SIU. Out of these 673 files, 181 were classified as suspicious because there was not enough evidence or proof to convince the SIU that these claims should not be paid.

(Chart A1 about here)

The second group of files (B) was selected from the population of claims that the insurer did not think contained any type of fraud during the same period of time. We first chose to randomly select only about 1,000 files in the reference group, because the cost of compiling information on fraud indicators is very high. In fact, to find significant indicators for fraud detection, our assistants had to read each file in groups (A and B) to search for the potential indicators (about 50) identified by members of the SIU. Over the 1,000 selected claims in group B, 945 were classified as without fraud. The 55 remaining files contained some indication of fraud.

Chart A2 describes the breakdown of files chosen for the analysis: The 184 files without any fraud in A were transferred to B, yielding two groups of files (A' with fraud and B' without fraud) and showing that 37% of the files contained established or suspected fraud. The 55 files with some suspicion of fraud in B were not included in A'.

36

(Chart A2 about here)

Calibration

To obtain the estimated fraud rate of the insurer (8% according to the views of the company managers), we first replicated the non fraud sub-sample six times, yielding 6,774 observations (6 × 1,129). Then we took a random sample (with replacement) from these 6,774 observations in order to obtain the additional 953 observations needed to produce a fraud rate of 8%. This percentage reflects the views of the company managers about the importance of claims fraud in their portfolio of automobile insurance contracts. The final sample contains 8,400 files, 673 files with fraud and 7,727 files without fraud. For a matter of robustness, we did also select randomly 6,598 non-fraud files from the original population of 1,129 to obtain the 7,727 non-fraud files. The results are very similar and are reported at the bottom of Table 2.

Appendix A3: Detailed Description of Variables in Regression Analysis

q_{19}^p	Production of questionable or falsified documents means that the policy holder is submitting fraudulent invoices. The anti-fraud agent is alerted when the policyholder submits photocopies or duplicates of bills.
$qexp$	Fraud alert by expert (expert warns that damages to vehicle do not correspond to those claimed by policyholder).
$qext$	Fraud alert by ARGOS or ALFA professionals (ARGOS alerts by mail, telephone, or via ARVA, a telematic management link with experts).
q_{30}^p	Description of circumstances surrounding the accident either lack clarity or seem contrived (accident report is too perfect, policyholder's descriptions of accident are too detailed or too vague).
q_{21}^p	Variations in or additions to the policyholder's initial claims (additional reports…).
q_{35}^p	Too long a lag between date of purchase and date of guarantee: policyholder takes out insurance a week after purchasing the vehicle whereas, logically, this should be done the same day.
q_{36}^d	Date of guarantee subscription and/or date of its modification too close to date of accident (< 1 month).
q_7^p	Vehicle whose value does not match income of policyholder (ex: an unemployed person or a welfare recipient with a Porsche).
q_{20}^p	Refusal or reluctance to provide original documents.
q_{16}^p	Victim with no damage insurance and/or one who would suffer harm if found at fault (for example, a victim who is insured only for damages caused by a third party asks a friend to stage an accident in order to receive payment).
q_{22}^p	Harassment from policyholder to obtain quick settlement of a claim (abnormally frequent letters or calls from the policyholder).
q_{32}^d	Abnormally high frequency of accidents (more than 3 accidents a year).
q_{12}^d	Retroactive effect of the contract or guarantee (when the date of guarantee date of accident date of accident report).

q_{34}^p	Claimant not the same as policyholder (this can be detected in the claim or police report).
q_{18}^d	Delay in filing accident claim (refer to periods stated in contract: more than 5 business days for all types of accidents and more than 2 days after the accident occurs and/ or the theft is noticed.)

Appendix A4: Correlation tables

Table A – Correlations conditional on F

	q_{12}^d	q_{18}^d	q_{19}^p	q_{21}^p	q_{30}^p	q_{32}^d	q_{36}^d	q_{35}^p	q_{34}^p	q_{22}^p	q_{20}^p	qexp	qext
q_{12}^d	1	-0.02	-0.05	0.02	-0.04	0.02	0.05	-0.02	-0.01	-0.03	-0.00	0.01	-0.00
q_{18}^d		1	0.04	0.02	-0.01	0.01	-0.02	0.06	0.09	0.01	-0.04	-0.11	0.04
q_{19}^p			1	0.16	0.12	-0.02	0.02	-0.01	-0.02	0.06	0.01	0.07	0.01
q_{21}^p				1	0.20	0.05	-0.05	-0.01	0.09	0.03	0.12	0.11	0.07
q_{30}^p					1	0.03	0.04	-0.00	-0.01	0.03	0.12	0.08	0.02
q_{32}^d						1	-0.06	-0.05	0.12	0.04	0.04	0.00	0.02
q_{36}^d							1	0.19	-0.06	0.08	0.10	-0.11	0.02
q_{35}^p								1	-0.07	0.10	0.10	-0.04	-0.00
q_{34}^p									1	0.05	-0.03	-0.06	-0.05
q_{22}^p										1	0.14	-0.02	-0.02
q_{20}^p											1	0.05	0.05
qexp												1	-0.15
qext													1

Table B – Correlations conditional on N

	q_{12}^d	q_{18}^d	q_{19}^p	q_{21}^p	q_{30}^p	q_{32}^d	q_{36}^d	q_{35}^p	q_{34}^p	q_{22}^p	q_{20}^p	qexp	qext
q_{12}^d	1	-0.04	0.06	-0.00	-0.02	-0.02	0.06	0.00	-0.03	0.01	0.00	0.07	-0.02
q_{18}^d		1	0.04	-0.02	0.06	-0.06	0.08	0.04	0.08	-0.03	0.03	0.02	0.03
q_{19}^p			1	0.13	0.07	-0.01	0.04	0.02	0.01	0.10	0.05	0.00	-0.01
q_{21}^p				1	0.03	0.00	-0.03	0.03	0.03	0.08	0.03	0.05	-0.01
q_{30}^p					1	-0.02	0.02	-0.04	0.02	0.01	0.05	0.01	0.01
q_{32}^d						1	-0.10	-0.07	0.02	-0.02	0.01	0.04	0.01
q_{36}^d							1	0.13	0.02	0.12	0.09	0.00	-0.03
q_{35}^p								1	0.04	-0.02	0.01	-0.02	-0.01
q_{34}^p									1	-0.03	0.04	0.03	0.01
q_{22}^p										1	0.02	0.14	-0.01
q_{20}^p											1	0.07	-0.01
qexp												1	-0.01
qext													1

Figure 1
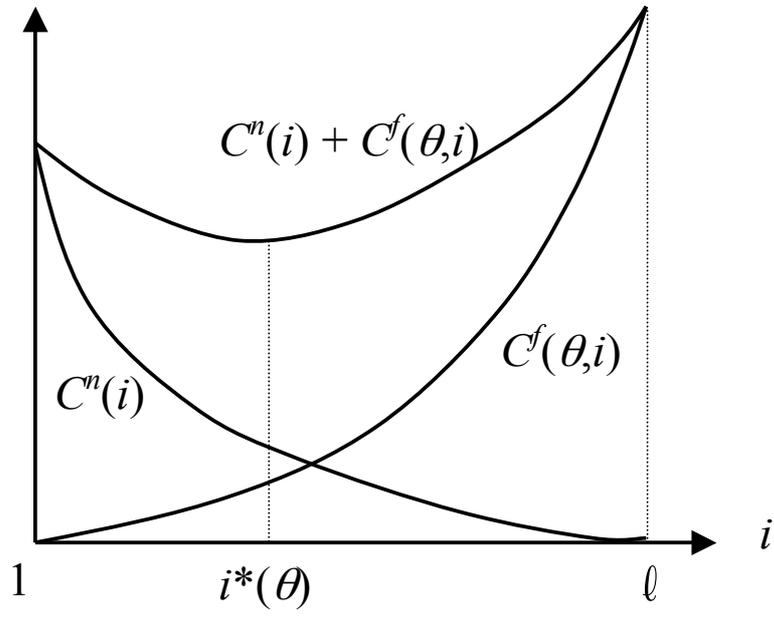
Figure 2
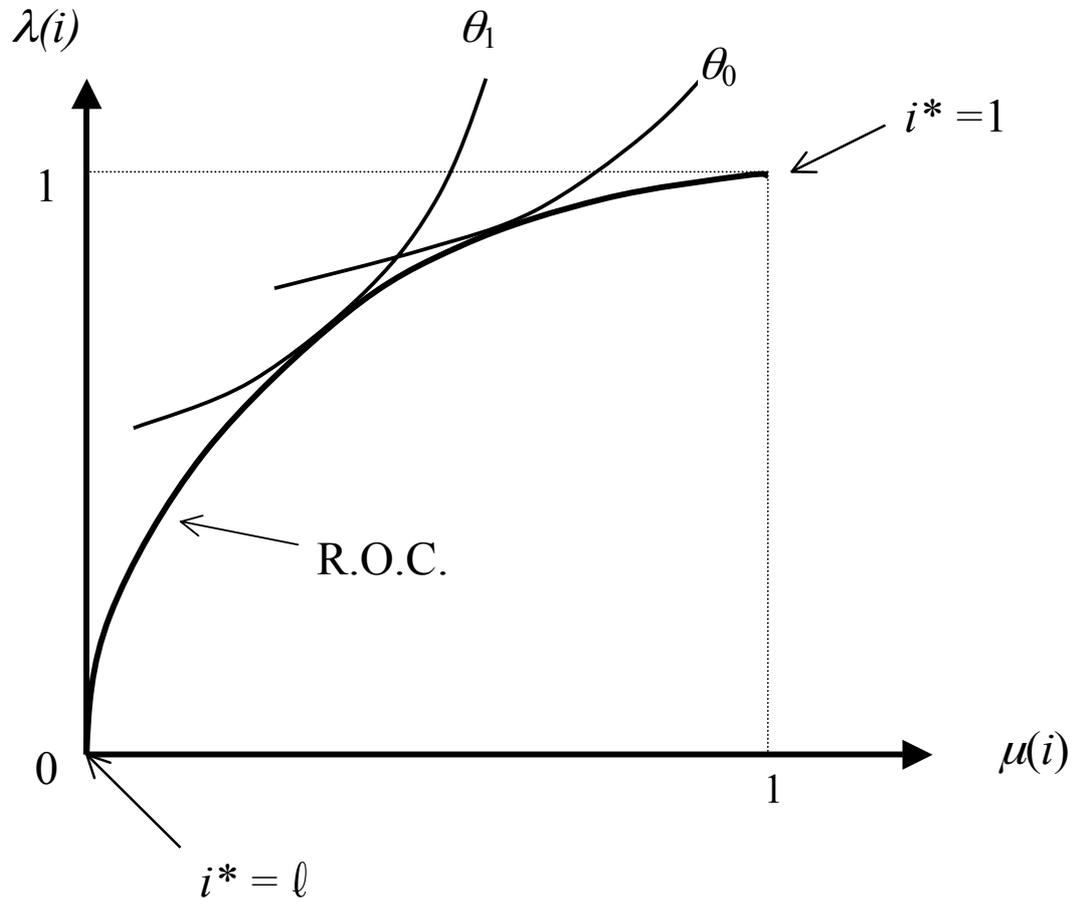
Figure 3
Gain Chart

43
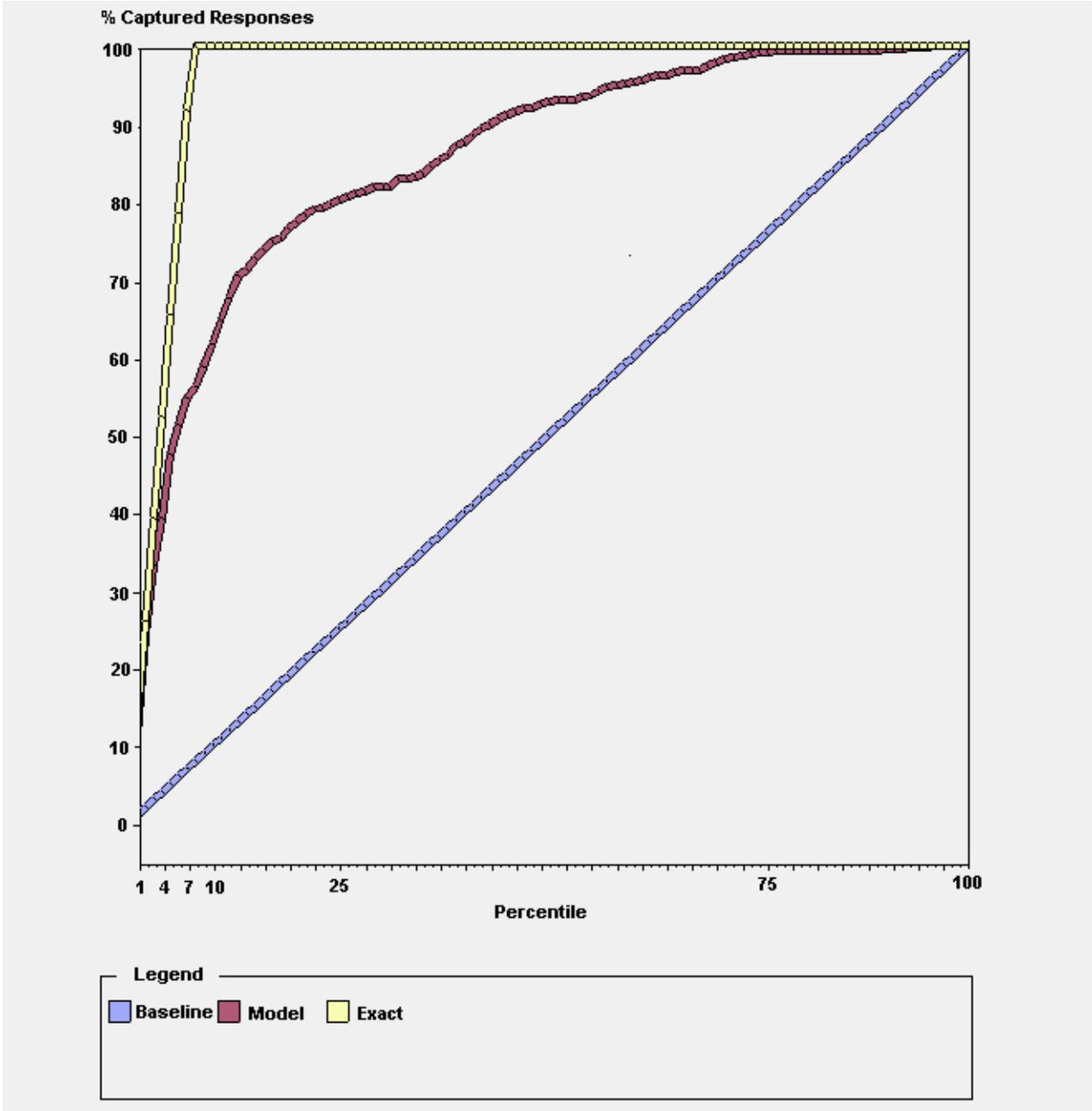
Chart A1
Original Data Set

Files transmitted to SIU

Files

Sample of claims without fraud

A

B

No fraud
184 files

Suspected fraud
181 files

Established fraud
492 files

A = 857 files
47.56%

B = 945 files of claims
without fraud
52.44%

```
                              ┌─────────┐
                              │  Files  │
                              └─────────┘
            ┌────────────────────┘  │  └────────────────────┐
        ┌───────┐                   │                   ┌───────┐
        │  A '  │                   │                   │  B '  │
        └───────┘                   │                   └───────┘
      ┌────┘  └────┐                │                 ┌────┘  └────┐
```

| Suspected fraud | Established fraud | | Without fraud | Files B |

A' = A – files without fraud in A
= 673 files
37%

B' = B + files without fraud in A
= 1,129 files
63%
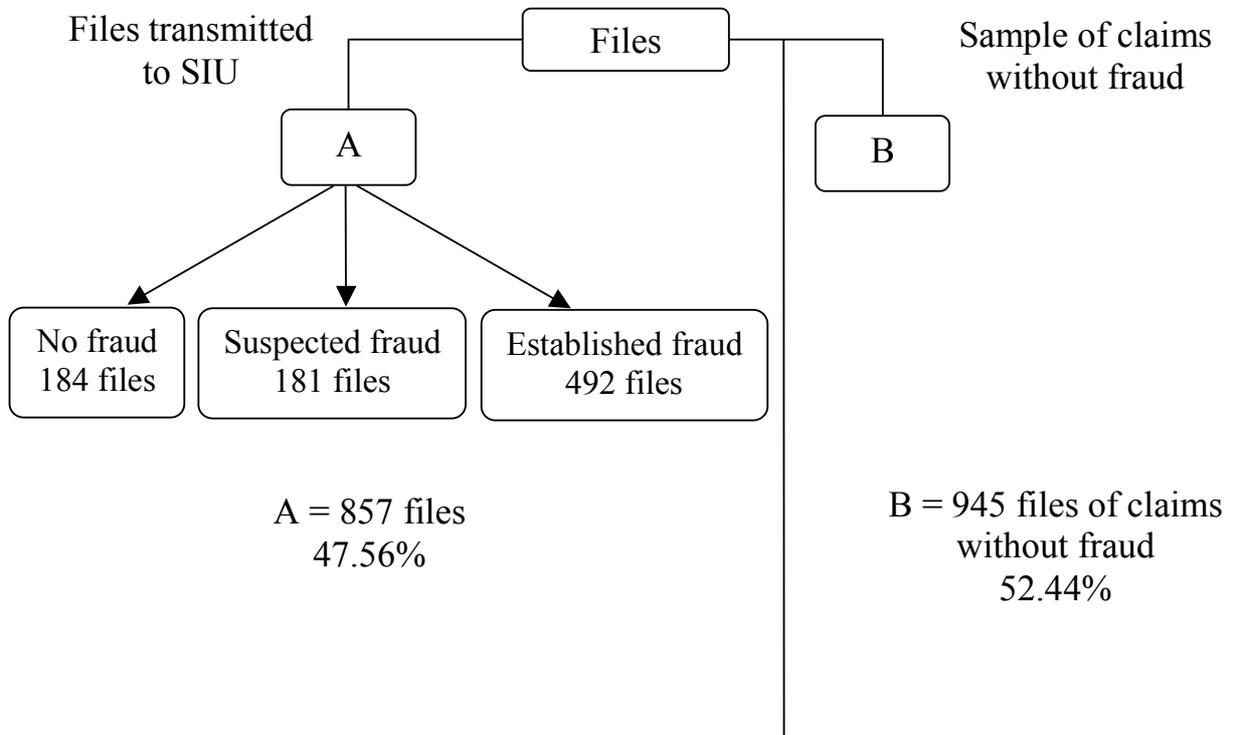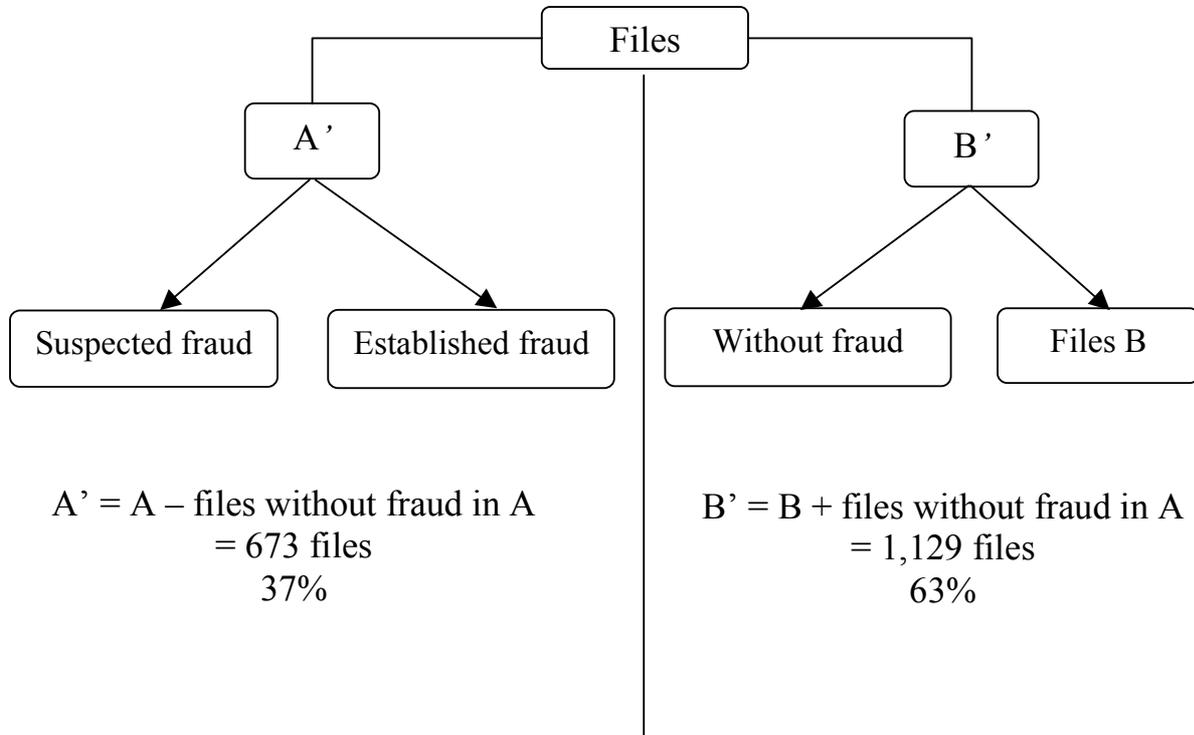
Table 1
Regression Results

Symbol	Variable	Without θ variables			With θ variables		
		Parameter	Std Error	P	Parameter	Std Error	P
	Intercept	5.8157	0.4246	<0.0001	6.4780	0.4749	<0.0001
q^d_{12}	Retroactive effect of the contract	0.2103	0.1150	0.0674	0.2329	0.1165	0.0455
q^d_{18}	Delay in filing accident claim	0.1808	0.0689	0.0087	0.1567	0.0697	0.0247
q^p_{19}	Falsified documents or duplicated bills	1.2227	0.0917	<0.0001	1.2287	0.0928	<0.0001
q^p_{21}	Variation in the initial claim	0.5681	0.1262	<0.0001	0.5218	0.1261	<0.0001
q^p_{30}	Lack of clarity in the accident description	0.3590	0.0849	<0.0001	0.3546	0.0859	<0.0001
q^d_{32}	Abnormal high frequency of accident	0.2352	0.0879	0.0075	0.2404	0.0888	0.0068
q^d_{36}	Date of subscription to guarantee too close	0.4520	0.0748	<0.0001	0.4264	0.0760	<0.0001
q^p_{35}	Too long lag between car purchase and guarantee	0.6714	0.1484	<0.0001	0.6340	0.1525	<0.0001
q^p_{34}	Claim maker different to policyholder	0.2044	0.0910	0.0246	0.1999	0.0920	0.0297
q^p_{22}	Harassment to obtain quick settlement	0.6197	0.1629	0.0001	0.5933	0.1677	0.0004
q^p_{20}	Reluctance to provide original documents	0.6099	0.1506	<0.0001	0.5980	0.1503	<0.0001
$qexp$	Fraud alert by expert	1.6365	0.1341	<0.0001	1.6299	0.1349	<0.0001
$qext$	Fraud alert by external organizations	1.4072	0.2612	<0.0001	1.4387	0.2632	<0.0001
q^p_7	Too high value of vehicle				0.7742	0.2055	0.0002
q^p_{16}	No damage insurance				0.2302	0.0696	0.0009
Log Likelihood				−694.72			−682.55
Number of observations				1,802			1,802

46

Table 2
Calibration Results

1	2	3	4	5	6	7
i or $i*$	σ_i	p_i^f / p_i^n	$\lambda(i)$	$\mu(i)$	$C^n(i)$	$t - \lambda(i)(t-c)$
1	NNNNNNNNNNNN	0.0779	1.0000	1.0000	56.672	280.00
2	NNNNNNNNNNNN	0.1146	0.9816	0.7633	43.259	298.51
3	NNNNNYNNNNNN	0.1155	0.9772	0.7248	41.076	302.94
4	NYNNNNNNNNNN	0.1219	0.9718	0.6782	38.437	308.34
5	YNNNNNNNNNNN	0.1362	0.9527	0.5220	29.580	327.46
6	NNNNNYNYNNNN	0.1699	0.9502	0.5032	28.516	333.03
7	NYNNNNNYNNNN	0.1793	0.9489	0.4956	28.087	331.32
8	NYNNNYNNNNNN	0.1808	0.9443	0.4702	26.645	335.90
9	YNNNNNYNYNNN	0.2003	0.9388	0.4394	24.903	341.48
10	YNNNNYNNNNNN	0.2019	0.9382	0.4364	24.729	342.10
...						
...						
194	**NNNNNYYNNNYNNN**	**3.1979**	**0.6672**	**0.0423**	**2.400**	**614.13**
...						
...						
8192	YYYYYYYYYYYY	63291108.94	0.0000	0.0000	0.0000	1284.00

180	**NNNNNYNNNYYNNN**	**3.1530**	**0.6723**	**0.0431**	**2.4439**	**608.96**

Table 2
Calibration Results (continued)

8	9	10	11	12
$C^f(\theta,i)$	$C^n(i)+C^f(\theta,i)$	$z(\theta)\lambda(i)+(1-z(\theta))\mu(i)$	$c\times 10$	$P(F/i>i^*)$
4.928	61.600	1.0000	280.00	0.08000
5.254	48.513	0.7808	218.62	0.10057
5.332	46.407	0.7450	208.60	0.10493
5.427	43.864	0.7017	196.48	0.11079
5.763	35.344	0.5564	155.80	0.13698
5.808	34.325	0.5389	150.90	0.14104
5.831	33.918	0.5319	148.92	0.14273
5.912	32.557	0.5081	142.27	0.14868
6.010	30.913	0.4794	134.22	0.15667
6.021	30.750	0.4765	133.42	0.15751
10.809	**13.208**	**0.0923**	**25.85**	**0.57810**
22.598	22.598	0.0000	0.00	1.00000
10.718	**13.167**	**0.094**	**26.17**	**0.5755**

Table 3
z(θ) Values from Regression Results in Table 1

$q_7^p \, / \, q_{16}^p$	Yes	No
Yes	*z(θ)* = 41.67% (48 policyholders)	*z(θ)* = 30.38% (79 policyholders)
No	*z(θ)* = 10.01% (3518 policyholders)	*z(θ)* = 5.83% (4755 policyholders)

Table 4

Sensitivity of Optimal Solutions With Respect to $\gamma(\theta)$ and $z(\theta)$

Case	Parameters $z(\theta)$	$\eta(\theta)^{(1)}$	$\gamma(\theta)$	π	$\pi(\theta)$	Optimal Treshold i^*	Expected Cost of Fraud $C^n(i^*(\theta)) + C^f(\theta, i^*(\theta))$	Expected Audit Probability $z(\theta)\lambda(i^*) + (1 - z(\theta)) \times \mu(i^*)$	Audit Probabiliy of Defrauders $\lambda(i^*)$
A	0.0800	0.00	0.00	0.2024	0.0176	194	13.21	0.0923	0.6672
B	0.0800	0.10	0.05	0.2024	0.0176	187	12.63	0.0937	0.6709
C	0.0800	0.21	0.10	0.2024	0.0176	183	12.08	0.0971	0.6798
D	0.0800	0.75	0.35	0.2024	0.0176	178	9.74	0.0979	0.6818
E	0.0583	0.00	0.00	0.2024	0.0128	250	10.15	0.0644	0.6210
F	0.0583	0.08	0.05	0.2024	0.0128	246	9.75	0.0646	0.6216
G	0.0583	0.17	0.10	0.2024	0.0128	240	9.37	0.0652	0.6238
H	0.0583	0.59	0.35	0.2024	0.0128	223	7.71	0.0669	0.6295
I	0.1001	0.00	0.00	0.2024	0.0220	157	15.84	0.1164	0.6952
J	0.1001	0.11	0.05	0.2024	0.0220	155	15.10	0.1170	0.6965
K	0.1001	0.23	0.10	0.2024	0.0220	153	14.39	0.1174	0.6974
L	0.1001	0.85	0.35	0.2024	0.0220	149	11.45	0.1232	0.7095

[1] $\eta(\theta) = \gamma(\theta)\lambda(i^*(\theta)) / 1 - \lambda(i^*(\theta))$

Table 5
Monetary Values of the Results
for the Insurer Portfolio

Without optimal audit the total claim cost =

$$500,000 \times € 1,284 = € 642 \text{ M} < \begin{matrix} 92\% = € 591 \text{ M} \\ \\ 8\% = € 51 \text{ M} \end{matrix}$$

With optimal audit the expected total claim **and** audit cost (with $\eta = 0$) =

$$9.23\% \times 500,000 \times € 280 + € 591 \text{ M} + 33\% \times €51 \text{ M} = € 621 \text{ M}$$

With audit of all files we obtain:

$$500,000 \times €280 + €591 \text{ M} = €731 \text{ M}$$

When $\eta(\theta) = 0.21$, the expected total claim **and** audit cost becomes:

$$9.71\% \times 500,000 \times €280 + €591 \text{ M} + 32\% \times €46.22 \text{ M} = €619.38 \text{ M}$$

Notes

[1] See for instance Dionne et al. (1996) on consumer credit scoring and Moody's Investors Service (2000) on corporate default risk scoring.

[2] See United States General Accounting Office (2004).

[3] The papers by Knowles et al. (2001) and Persico (2002) are noteworthy exceptions to this separation between the literature strands on audit and scoring. They are about auditing strategies (more specifically automobile control by police officers) in a setting where auditors can condition the audit probability on the agents' observable particulars and they focus on the detection of possible bias in the auditors' behavior (e.g. racial bias in vehicle searches). The present paper focuses on the role of signals perceived by the principal when agents can defraud and on the design and the practical implementation of an optimal auditing strategy.

[4] See Picard (2000) for an overview.

[5] Crocker and Morgan (1997) have developed a *costly state falsification* approach to insurance fraud which has conceptual similarities with the models of costly state verification with audit cost manipulation. Other references on this issue are Crocker and Tennyson (2002), Picard (1996, 1999), and Boyer (2004).

[6] See Dionne and Gagné (2001, 2002), Derrig and Weisberg (2003), Artis et al. (2002) and Crocker and Tennyson (2002) for different econometric applications, and volume 69, no 3, of the *Journal of Risk and Insurance*, September 2002, for a state-of-the-art presentation of claims fraud detection methods. There is also a literature on the measurement of information problems in economic activity that interprets insurance fraud as an ex-post moral hazard problem. See Derrig (2002) and Dionne (2000) for comprehensive surveys.

[7] A similar result is obtained by Persico (2002) in a different context. He analyses the behavior of police officers who choose whom to investigate when citizens of two groups may engage in crime. The equilibrium search is obtained when both groups have the same fractions of criminals, while maximal search effectiveness usually entails different hit rates between the two groups.

[8] The indemnity B does not play any crucial role in the model (apart from affecting the equilibrium intensity of fraud) and $B = 0$ is a possible case.

[9] Hence we assume that the red flags cannot be manipulated by the defrauders. Note that such signals used in auditing strategy are usually kept as confidential by insurers. Characterizing an optimal auditing strategy under costly signal manipulation would be an important extension of the present analysis.

[10] We then have $\sigma_i = (\sigma_{i1}, \sigma_{i2},..., \sigma_{ik})$ for all $i = 1, ..., \ell$ with $\sigma_{ij} \in \{0, 1, ..., N_j-1\}$ for all $j = 1, ..., k$.

[11] For notational simplicity, we assume that p_i^f and p_i^n do not depend on θ. Our analysis can be straightforwardly extended to the case where signals are correlated with individuals' types.

[12] Of course if $p_i^n = 0$ and $p_i^f > 0$ then the optimal investigation strategy involves channeling the claim to SIU (see the definition and the role of SIU hereafter) when $\sigma = \sigma_i$. Indeed the claim is definitely fraudulent in such a case.

[13] The only case where a random investigation may be optimal is for $i = i^*(\theta)$.

[14] Hence, we here disregard the fact that we may have $q(\theta, \sigma_i) < 1$ when $i = i^*(\theta)$.

[15] In this specification of the model, ω may be interpreted as reflecting the difficulty to defraud rather than a psychological cost.

[16] Tables A and B in Appendix A4 show that the conditional correlations are very small in our data set.

[17] We here assume that $(p_{i^*(\theta)}^f / p_{i^*(\theta)}^n) - (p_{i^*(\theta)-1}^f / p_{i^*(\theta)-1}^n)$ is small enough for (28) to be implied by (20) and (27).