CIRPÉE
Centre interuniversitaire sur le risque, les politiques économiques et l'emploi

Cahier de recherche/Working Paper **03-29**

Optimal Auditing for Insurance Fraud

Georges Dionne
Florence Giuliano
Pierre Picard

Août/August 2003

Dionne: Risk Management Chair, HEC Montréal, CREF and CIRPÉE, Canada. Corresponding author: Risk Management Chair, HEC Montréal, 3000, Chemin de la Côte-Sainte-Catherine, Montréal (QC) Canada, H3T 2A7. Phone: (514) 340-6596; Fax: (514) 340-5019
georges.dionne@hec.ca
Giuliano: THEMA, Université de Paris X-Nanterre, France
Picard: THEMA, Université de Paris X-Nanterre, CEPREMAP, France, and Hong Kong University of Science and Technology

The paper was presented to the 2003 Risk Theory Seminar in Atlanta, the Insurance Fraud Conference in Newport, 2002, and in seminars in Hong Kong and Paris. Comments by R. Derrig and S. Tennyson were very useful in improving previous versions. Financial support by the Fédération Française des Sociétés d'Assurances (FFSA) and a private insurer is gratefully acknowledged. Claire Boisvert and Sybil Denis improved the different versions of the paper with great competence.

Abstract: This article aims at making a bridge between the theory of optimal auditing and current procedures applied to audit files in different markets where scoring is the instrument used to implement an audit strategy. The literature has not yet developed an optimal audit policy for the scoring methodology. Our application is meant for the audit of insurance fraud but can be applied to many other activities that use the scoring approach. On the theoretical side, we show that the optimal auditing strategy takes the form of a "Red Flags Strategy" which consists in referring claims to a Special Investigation 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 and this strategy remains optimal if the investigation policy is budget constrained. Moreover, the auditing policy acts as a deterrence device. On the empirical side, four significant results are obtained with data from a large European insurance company. First, we are able to compute a critical suspicion index for fraud, providing a threshold above which all claims must be audited. Secondly, we obtain that if the insurer applies this policy, he will save more than €22 million, which represents 43% of the current cost of fraudulent claims. Thirdly, we show that it is possible to improve these results by using information capable of isolating different groups of insureds with different morale 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.

Keywords: Optimal audit, scoring methodology, insurance fraud, red flags strategy, fraud indicators, suspicion index, morale cost of fraud.

JEL Classification: D81, G14, G22

I 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 (Melumad and Mookherjee, 1989; Graetz, Reinganum and Wilde, 1986).

Scoring is an alternative approach to auditing which differs from standard deterministic or random methodologies. Scoring helps in identifying suspicious files to be audited as a priority. It is now widely used by banks in credit-risk management, by corporations in hiring decisions, by tax authorities in tracking tax compliance or by insurers in detecting claims or application fraud. In a sense, scoring is a way out of the deterministic costly state-verification model put forward by Townsend (1979) and Gale and Hellwig (1985), a way out which also differs from the random auditing approach: Scoring says when a file should undergo in-depth investigation and when it should not. In fact, scoring is a way to cut investigation costs by specifying what information should be used.

For example, Moody's Investors Service (2000) has developed the RiskCalc model for private firms. In this model, Moody's uses financial variables to assess client-default risk by computing individual scores. This type of model can also be applied to consumer-credit modeling (Dionne et al., 1996) and to assess the expected profitability of an investment project or the residual value of a bankrupted firm. The scoring methodology associates scores — i.e. numerical valuations — to an unobservable default risk, expected profitability or residual value. However, the current methodology does not lead to the identification of the optimal score that takes into account the various costs and benefits associated with a loan (optimal interest rate, audit cost, default probability, and recovery rate). The same remark applies to other applications of the scoring model, indicating that the literature has not yet developed an optimal audit policy for the scoring methodology.

The main goal of this article is to develop an integrated approach to auditing and scoring, where the scoring methodology can be used to implement an optimal auditing strategy. In doing so, we shall build a model of optimal auditing which is much more closely related to auditing procedures actually used by insurers, bankers or governmental regulators than to abstract costly state-verification models. We shall also show how the scoring methodology can be used in an optimal auditing strategy. Scoring will signal whether an audit should be performed or not: 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 approach.

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.[1] 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 setting 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. The focus is on the deterrence effect of the auditing strategy and on the consequences of insurance fraud on the design of insurance contracts. Important assumptions are made relative to the ability of

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

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.[2]

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.[3] The scoring methodology is one of the key ingredients in this statistical approach.

Two of the key questions 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 individuals. Developing automated methods (in particular the scoring approach) capable of using the informational content of red flags as efficiently as possible is currently the subject of really intense research by some insurance companies, particularly in the automobile insurance sector.

In this paper, we shall link the theory of optimal auditing and the scoring methodology by building a costly verification model that predicts a red flag investigation strategy. We then calibrate our model by using data on automobile insurance from a large European insurance company and derive the optimal auditing strategy. As a final outcome, our analysis yields an easily automated procedure for detecting insurance fraud. Section II presents the theoretical model while Section III derives the optimal auditing strategy. The application of the model to the portfolio of an insurer begins with Section IV where the data set is introduced. Section V presents the regression analysis and Section VI outlines the model calibration and its results. Section VII concludes.

---

[2] 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 (1999) and Picard (1996, 1999).

[3] See Derrig (2002) and Dionne (2000). See also 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 Chiappori and Salanié (2003) for a recent comprehensive survey.

II Model

We consider a population of policyholders who differ from one another in the morale cost of filing a fraudulent claim. For the sake of notational simplicity, all individuals own the same initial wealth W_0 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 morale cost incurred in case of insurance fraud:

$$u = u(W, \omega) \quad \text{in case of fraud}$$

$$u = u(W, 0) \quad \text{otherwise}$$

where ω is a non-negative parameter which measures the morale cost of fraud to the policyholder. We assume $u_1^{'} > 0, u_{11}^{''} < 0$ and $u_2^{'} < 0$ and that ω is distributed over $\Re_+$ among the population of policyholders. In other words, individuals who choose to defraud incur more or less high morale costs. Some of them are purely opportunistic (their morale cost is very low) whereas others have a higher sense of honesty (their morale cost is thus higher). Note that morale 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 P that should be paid to the insurer. Hence, if there is no fraud, we have:

$$W = W_0 - L - P + t \quad \text{in case of an accident}$$

and

$$W = W_0 - P \quad \text{if no accident occurs.}$$

Each individual in the population is characterized by a vector of observable exogenous variables θ, with $\theta \in \Theta \subset \Re^m$. The morale cost of fraud may be statistically linked to some of these variables. Let $H(\omega|\theta)$ be the conditional, cumulated, probability distribution of ω for a type-θ individual, 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.[4] Let Q^f be the probability that a fraudulent claim is detected; this

[4] 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.

probability is the outcome of the insurer's antifraud policy and it depends on the observable variables θ as we shall see in Section III.

When an individual has not suffered any loss, his utility is written as $u(W_0 - P, 0)$ 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:

$$W = W_0 - P + t$$

if he is not detected and

$$W = W_0 - P - B$$

if he his detected. Hence, an individual with morale cost ω decides to defraud if he expects greater utility from defrauding than staying honest, which is written as:

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

This inequality holds if $\omega \leq \phi(p)$, where function $\phi : [0,1] \rightarrow \Re^+$ is implicitly defined by:

$$(1 - Q^f)\, u(W_0 - P + t, \phi) + Q^f u\,(W_0 - P - B, \phi) = u(W_0 - P, 0)$$

with $\phi(0) > 0$, $\phi(1) = 0$ and $\phi'(Q^f) < 0$. $\phi(Q^f)$ is the critical value of the morale cost under which cheating overrides honesty as a rule of behavior. The higher the probability of being detected, the lower the threshold and thus the lower the level of fraud.

When a policyholder files a claim — be it honest or fraudulent — the insurer privately perceives a multidimensional signal σ. We assume:

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

with

$$\sigma_i \in \Re^k, \ k \geq 1 \quad \text{for all } i = 1,...,\ell.$$

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[5]. If indicator j takes N_j possible values[6] – say 0, 1,...,N_j –, we have $\ell = \prod_{j=1}^{k} N_j$. When all indicators are binary (i.e. when $N_j = 2$ for all $j = 1,...,\ell$), then $\ell = 2^k$ and σ is a vector of dimension k all of whose

[5] 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.

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

components may be taken as equal to 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 σ_i when the claim is fraudulent and when it corresponds to a true accident (non-fraudulent claim), i.e.:

$$p_i^f = \text{Prob}\,(\sigma = \sigma_i \mid F)$$
$$p_i^n = \text{Prob}\,(\sigma = \sigma_i \mid N)$$

with $i = 1,...,\ell$, where F and N refer respectively to "fraudulent" and "non-fraudulent". 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 we rank the possible signals in such a way that[7]

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

This ranking allows us to interpret $i \in \{1,...,\ell\}$ as an index of *fraud suspicion*. Indeed, assume that the proportion of fraudulent claims is equal to x, with $0 < x < 1$. Then Bayes law shows that the probability of fraud is:

$$\frac{p_i^f x}{p_i^f x + p_i^n (1-x)} \tag{1}$$

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

III 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

[7] 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.

referral serves to detect fraudulent claims as well as 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 $c < t$. Under this assumption, it would be profitable to channel a claim to an SIU if the insurer were sure that the claim is fraudulent. Unfortunately, non-fraudulent claims may also be channeled to an SIU by mistake! An optimal audit scheme should minimize this possibility.

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 an 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 an SIU with probability $Q^f(\theta)$. The insurer knows that, given his investigation strategy, the probability of mistakenly channeling a truthful claim to an SIU is $Q^n(\theta)$ if the policyholder is of type θ.

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.

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 an 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 morale cost ω is lower than $\phi\left(Q^f(\theta)\right)$ which occurs with probability $H\left(\phi\left(Q^f(\theta)\right) \mid \theta\right)$. Hence, the expected investigation cost is:

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

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:

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

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

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

Such a strategy is characterized in the following proposition.

Proposition 1
An optimal investigation strategy is such that

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

$$q(\theta,\sigma_i) = 1 \quad \text{if } i \geq i^*(\theta)$$

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

Proof
Using equations (2) to (6), pointwise minimization of *TC* with respect to $q(\theta,\sigma)$ gives:

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

where

$$A(Q,\theta) = \left(cQ + t(1-Q)\right) H\left(\phi(Q)\middle|\theta\right)$$

Note that $\phi' < 0$ and $t > c$ give $A_1' < 0$. Consequently, we have:

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

and

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

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

$$\frac{p_{i^*(\theta)-1}^f}{p_{i^*(\theta)-1}^n} < \frac{-c\pi}{(1-\pi)\, A_1^{'}\left(Q^f(\theta),\theta\right)} \leq \frac{p_{i^*(\theta)}^f}{p_{i^*(\theta)}^n} \tag{7}$$

Q.E.D.

Proposition 1 says that an optimal investigation strategy consists in ranking the multidimensional claim – related signals σ_i in such a way that p_i^f / p_i^n is increasing from i = 1 to $i = \ell$: claims should be subjected to an SIU referral when the suspicion index i exceeds an individual-specific threshold $i^*(\theta)$.

Note that:

$$p_i^f = \frac{P(F\,|\,\sigma_i)\,P(\sigma_i)}{P(F)} \tag{8}$$

and

$$p_i^n = \frac{P(N\,|\,\sigma_i)\,P(\sigma_i)}{P(N)} = \frac{(1-P(F\,|\,\sigma_i))\,P(\sigma_i)}{(1-P(F))} \tag{9}$$

which gives

$$\frac{p_i^f}{p_i^n} = \frac{(1-P(F))\,P(F\,|\,\sigma_i)}{P(F)(1-P(F\,|\,\sigma_i))}. \tag{10}$$

Hence, as index i increases so does the conditional probability of fraud. The critical index $i^*(\theta)$ corresponds to a threshold of this probability above which the claim is forwarded to an SIU. This critical probability depends on the individual specific variables.

Given Proposition 1, we may write:

$$Q^f(\theta) = \lambda\,(i^*(\theta))$$

and

$$Q^n(\theta) = \mu(i^*(\theta))$$

where $\lambda(i)$ and $\mu(i)$ are defined by:

$$\lambda(i) = \sum_{j=1}^{\ell} p_j^f$$

$$\mu(i) = \sum_{j=1}^{\ell} p_j^n$$

$\lambda(i)$ and $\mu(i)$ respectively denote the probability of channeling a fraudulent claim and a non-fraudulent claim to an 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.

Hence, $i^*(\theta)$ minimizes

$$c\pi\mu(i) + (1-\pi)H(\phi(\lambda(i))|\theta)(c\lambda(i) + t(1-\lambda(i))) \tag{11}$$

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 an SIU referral, and of

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

which is the expected cost of fraudulent claims. This cost includes the investigation cost of the claim channeled to an 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 an 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 $((\mu(i^*), \lambda(i^*)), i^*=1,...\ell)$ 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)

Let

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

and

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

$\tau(Q,\theta)$ 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 fraud (in absolute value), i.e. the percentage decrease in the fraud rate following a one percent increase in the probability of detection.

Proposition 2 says that a higher fraud rate and/or a larger elasticity of fraud should entail more systematic auditing by an SIU.

Proposition 2

Assume that $A(Q,\theta) = cQ + t(1-Q)\, H(\phi(Q)\,|\,\theta)$ *is convex in* Q. *If*

$$\tau(Q^f(\theta_0),\theta_1) \geq \tau(Q^f(\theta_0),\theta_0) \tag{12}$$

and

$$\eta(Q^f(\theta_0),\theta_1) \geq \eta(Q^f(\theta_0),\theta_0) \tag{13}$$

then

$$i*(\theta_1)\le i*(\theta_0).$$

Proof

Assume that $A(Q,\theta)$ is convex in Q. Let θ_0 and θ_1 in Θ such that (12) and (13) hold. Assume moreover that $i*(\theta_1) > i*(\theta_0)$, which gives:

$$Q^f(\theta_1) > Q^f(\theta_0) \tag{14}$$

Let $i \in \{1,...\ell\}$ such that:

$$i*(\theta_0)\le i< i*(\theta_1).$$

Proposition 1 then gives:

$$q(\theta_0,\sigma_i)=1$$
$$q(\theta_1,\sigma_i)=0.$$

Writing optimality conditions as in the proof of Proposition 1 yields:

$$c\pi p_i^n+(1-\pi)\,A_1^{'}\left(Q^f(\theta_0),\theta_0\right)p_i^f\le 0 \tag{15}$$

and

$$c\pi p_i^n+(1-\pi)\,A_1^{'}\left(Q^f(\theta_1),\theta_1\right)p_i^f\ge 0 \tag{16}$$

Using (14) and the convexity of $Q\to A(Q,\theta)$ gives:

$$A_1^{'}\left(Q^f(\theta_0),\theta_1\right)> A_1^{'}\left(Q^f(\theta_1),\theta_1\right). \tag{17}$$

(16) and (17) give:

$$c\pi p_i^n+(1-\pi)\,A_1^{'}\left(Q^f(\theta_0),\theta_1\right)p_i^f> 0. \tag{18}$$

(13) and (16) then imply:

$$A_1^{'}\left(Q^f(\theta_0),\theta_1\right)> A_1^{'}\left(Q^f(\theta_0),\theta_0\right). \tag{19}$$

We have

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

which may be rewritten as:

$$A_1^{'}(Q,\theta) = -\frac{\tau(Q,\theta)}{1-\pi}\left(t - c + \frac{cQ + t(1-Q)}{Q}\eta(Q,\theta)\right)$$

Using (12) and (13) gives

$$A_1^{'}\left(Q^f(\theta_0),\theta_1\right) < A_1^{'}\left(Q^f(\theta_0),\theta_0\right)$$

which contradicts (19). Hence, we may conclude that $i^*(\theta_1) \leq i^*(\theta_0)$, which completes the proof.
Q.E.D.

We know that $A(Q,\theta)$ is decreasing in Q, thereby reflecting the fact that increasing the audit probability allows the insurer to cut fraud costs, either directly through the detection of fraudulent claims, or indirectly by deterring fraud. Assuming that $A(Q,\theta)$ is convex in Q means that the marginal benefit of auditing is decreasing. The logic at work in Proposition 2 is the following: Auditing will cut fraud costs all the more efficiently if the insured belongs to a group with a high fraud and/or elasticity rate. 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, undoubtedly, claims should not receive less scrutiny when they are filed by type-θ_1 than by type-θ_0 individuals.

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

$$c\pi\, E_\theta Q^n(\theta) + c(1-\pi)\, E_\theta Q^f(\theta)\, H\left(\phi\left(Q^f(\theta)\right) \mid \theta\right) \leq K\,. \tag{20}$$

An optimal investigation strategy then minimizes TC with respect to $q(\cdot) : \Theta \times \Sigma \rightarrow [0,1]$ subject to (6) and (20). Proposition 3 shows that the qualitative characterization of the antifraud policy is not affected by the addition of this upper limit on possible investigation expenditures.

Proposition 3

Propositions 1 and 2 are still valid when the investigation policy is budget constrained.

Proof

Let α be a (non-negative) Kuhn-Tucker multiplier associated with (20) when TC is minimized with respect to $q(\theta,\sigma)$ subject to (6) and (20).
Pointwise minimization gives:

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

where

$$\widetilde{A}(Q,\theta) = \left(c\,(1+\alpha)\,Q + t(1-Q)\right) H\left(\phi(Q) \middle| \theta\right)$$

Proposition 3 can then be proved in the same way as Propositions 1 and 2.
Q.E.D.

Let $P\left(F \middle| \sigma_i, \theta\right)$ be the probability of fraud depending on the perceived signal and on the type of policyholder. $P\left(F \middle| \sigma_i, \theta\right)$ is given by (1) with x = $P\left(F \middle| \theta\right)$; $P\left(F \middle| \theta\right)$ denotes the probability of fraud for type θ individuals, and it is given by:

$$P\left(F \middle| \theta\right) = \frac{(1-\pi)\, H\left(\phi\left(Q^f(\theta)\right) \middle| \theta\right)}{\pi + (1-\pi)\, H\left(\phi\left(Q^f(\theta)\right) \middle| \theta\right)}. \tag{21}$$

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

$$P\left(F \middle| \sigma_i, \theta\right) t - c$$

Proposition 4 shows that the optimal investigation strategy involves transmitting suspicious claims to an SIU in cases where the expected benefit of such a special investigation may be negative.

Proposition 4
When there is no upper limit on SIU expenditures, the optimal investigation strategy is such that:

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

Proof
We have:

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

Hence

$$\frac{-c\pi}{(1-\pi)\,A_1^{'}\left(Q^f(\theta),\theta\right)} < \frac{-c\pi}{(1-\pi)(c-t)\,H\left(\phi\left(Q^f(\theta)\right)\mid\theta\right)} \tag{22}$$

which gives:[8]

$$\frac{p^f_{i*(\theta)}}{p^n_{i*(\theta)}} < \frac{-c\pi}{(1-\pi)(c-t)\,H\left(\phi\left(Q^f(\theta)\right)\mid\theta\right)} \tag{23}$$

Using (21), (23) and

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

gives:

$$P\left(F\mid\sigma_{i*(\theta)},\theta\right)t < c\,.$$

Q.E.D.

Since $P\left(F\mid\sigma_i,\theta\right)$ is increasing in i, Proposition 4 means that there exists $i^{**}(\theta)$ larger than $i^*(\theta)$ such that:

$$P\left(F\mid\sigma_{i**(\theta)},\theta\right)t < c < P\left(F\mid\sigma_{i**(\theta)+1},\theta\right)t\,.$$

Forwarding the claim to an SIU is profitable only if the suspicion index i is larger than $i^{**}(\theta)$. Hence, it is optimal to channel the claim to an SIU when $i^*(\theta) \le i \le i^{**}(\theta)$, although in such a case the expected profit drawn from investigation is negative. This result follows from the fact that the investigation strategy acts as a deterrent: It dissuades some insureds (those with the highest morale costs) from defrauding. Such a strategy involves a stronger investigation policy than the one that would consist in transferring a claim to an SIU when the direct monetary benefits expected from investigation are positive. The indirect deterrence effects of the investigation policy should also be taken into account, which leads to more frequent investigation. We are now ready to test the main propositions of the article.

[8] We here assume that $(p^f_{i*(\theta)} / p^n_{i*(\theta)}) - (p^f_{i*(\theta)-1} / p^n_{i*(\theta)-1})$ is small enough for (23) to be implied by (7) and (22).

IV Data

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 1 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 1 about here)

The second group of files (B) was randomly selected from the population of claims that the insurer did not think contained any type of fraud during the same period of time. We chose to 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 identified by members of the SIU (about 50).

Chart 2 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.

(Chart 2 about here)

In order to obtain a final sample representing the true proportion of fraudulent claims in the company, we used the bootstrapping method. We applied two complementary techniques. The first consisted in replicating the original B' subsample six times, yielding 6,674 observations (6 × 1,129). Then we took a random sample (with replacement) from these 6,674 observations in order to obtain the additional 953 observations needed to produce a fraud rate of 8%, which is supposed to be the fraud rate in the insurer's portfolio. The final sample contained 8,400 files, 673 files (A') containing fraud and 7,727 files (B'') with no fraud. Chart 3 presents the final sample.

(Chart 3 about here)

V Regression Analysis

Regression Model

First, an 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. The insured either has filed a fraudulent claim or has not. The fraud indicators may affect the status of the file. So we can write:

$$\text{Prob}\ (Y = \text{Fraud}) = F(\beta' X)$$

and

$$\text{Prob}\ (Y = \text{No Fraud}) = 1 - F(\beta' X)$$

where X is the vector of explanatory variables (fraud indicators or individual characteristics) and β is the vector of parameters. If we assume that $F(\cdot)$ is the Logistic cumulative distribution function, then we estimate the Logit model. 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).

Regression Results

Table 1 reports the regression results. A detailed description of the variables is presented in Appendix A. 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%. The second column (with θ variables that represent the 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 morale cost component but also a monetary cost component. We have restricted 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 morale cost or that they draw a larger monetary benefit from fraud, hence a higher probability for filing a fraudulent claim.

(Table 1 about here)

Figure 3 presents the Gain Chart corresponding to the model without the θ variables. On the horizontal axis of the figure, the files are ordered by decreasing fraud probability (i.e. decreasing fraud score). The vertical axis of the figure indicates the percentage of captured responses (% of the files with fraud) according to three different methods. The first 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 an 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 8^{th} 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 8^{th} 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)

VI Model Calibration

Data

Let $\hat{\pi}(\theta)$ be the probability that a claim will be filed by a type-θ individual 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 morale costs), t does not depend on θ. For our purpose $\hat{\pi}(\theta)$ = 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 proportion of claims with fraud $z(\theta)$ is 8%.

Since $\hat{\pi}(\theta)$ contains fraudulent claims, the true loss (theft and accident) probability π is given by:

$$\pi = \hat{\pi}(\theta)\left(1 - z(\theta)\right) = 0.2024$$

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 morale costs in the population of policyholders as well as on the relationship between the audit probability, the morale 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 θ. Such information is obviously unobservable. This is why we will content ourselves with an approximation of $\tau(Q,\theta)$.

We will assume:

$$\tau(Q,\theta) = \hat{\tau}(\theta)\,(1-Q)^{\gamma(\theta)} \tag{24}$$

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.

Using (24) allows us to rewrite (11) as:

$$c\pi\mu(i) + \hat{\tau}(\theta)\,(1-\lambda(i))^{\gamma(\theta)}\,(t - \lambda(i)(t-c)) \tag{25}$$

where $\lambda(i) = \sum_{j=1}^{i} p_j^f$ and $\mu(i) = \sum_{j=1}^{i} p_j^n$.

In what follows we do not use information on the types of documentation available on claimants: In other words, 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 (25) with respect to *i*. For that purpose, we must first compute the values of $\lambda(i)$ and $\mu(i)$. From (8) and (9) we have:

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

$$p_i^n = \frac{(1-P(F/\sigma_i))\;P(\sigma_i)}{[1-P(F)]} = P(\sigma_i/N).$$

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 fraud indicators and, consequently, 8,192 values for σ_i. Since our data set is limited to 9,171 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 $(q_j = 1)$ or absence $(q_j = 0)$ of the indicator j in a given file. So we can write:

$$\sigma_{ij} = 1 \text{ if } q_j = 1$$
$$\sigma_{ij} = 0 \text{ if } q_j = 0.$$

Let

$$\alpha_j^f = \text{Prob } (q_j = 1/F)$$

and

$$\alpha_j^n = \text{Prob } (q_j = 1/N)$$

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. This conditional independence assumption allows us to write:

$$p_i^f = P(\sigma_i / F) = \prod_{j/\sigma_{ij}=1} \alpha_j^f \prod_{j/\sigma_{ij}=0} \left(1-\alpha_j^f\right) \tag{26}$$

$$p_i^n = P(\sigma_i / N) = \prod_{j/\sigma_{ij}=1} \alpha_j^n \prod_{j/\sigma_{ij}=0} \left(1-\alpha_j^n\right) \tag{27}$$

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 8th 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 Table 2. The corresponding values are 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 will not be optimal. The optimal critical suspicion index, denoted i^*, trades off the benefits and the costs of auditing. Another example would be to choose $i^* = 10$ as a critical suspicion index, which means that all files with a ratio p_i^f / p_i^n higher than 0.17 will be audited. This would mean that 95% of the fraudulent claims would be audited and that 45% of the non-fraudulent claims would also be audited (see $\lambda(10)$ and $\mu(10)$ in column 7). This would also be a very costly strategy. Let us now consider in detail the different auditing costs.

Column 6 presents the expected investigation cost of a non-fraudulent claim 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:

$$€280 \times 0.2024 = €56.67$$

since $\pi = 0.2024$ is the accident probability, c = €280 is the audit cost and $\mu(i) = 1$. For line 10, this cost is reduced to €25.66 because $\mu(i)$ is now equal to 0.45291. 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 for $\eta = 0$ and $\tau = 0.018$. In line 1, this expected cost is very low because it is reduced to €280 × τ. Moreover, here $\gamma = 0$ which means that there is no incentive or deterrent effect associated with a variation in $\lambda(i)$. Column 9 computes the expected total cost of fraud per policyholder which is the sum of columns 7 and 8. The optimal $i^*(\theta)$ will be 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 $i^* = 1$, all claims are audited and the probability of fraud is equal to the average fraud rate in the sample because there is no incentive effect here ($\gamma = 0$). However, if $i^* = 10$, the auditing strategy will be more focused on claims with suspected fraud ($\lambda(i) = 0.95$ and $\mu(i) = 0.45$) and the probability of fraud in audited claims is then equal to 0.1544.

The optimal solution is at line 238 = $i^*(\theta)$. We then have $\lambda(i^*) = 0.6805$, which means that 68% of the fraudulent claims will be audited. So the optimal expected cost of a fraudulent claim in the total insurer portfolio (Column 8) is €10.57. We also have $\mu(i^*) = 0.04$, which means that only 4% of the non-fraudulent claims will be audited. The corresponding optimal expected cost of a non-fraudulent claim in the insurer portfolio is equal to €2.25. So the optimal expected total cost of fraud reaches its minimal value at €12.83. Note that the corresponding cost at line 1 (audit all claims) is €61.60 and that at line 8192 it is €22.56.

The optimal strategy entails auditing 9.10% of the files (column 10) and the optimal audit cost per claim is €25.50. Finally,

$$P(F/i > i^*) = \frac{z(\theta)\,\lambda(i^*)}{z(\theta)\,\lambda(i^*) + (1 - z(\theta))\,\mu(i^*)} = 59.8\%$$

which means that 59.8% of audited claims prove to be fraudulent, which can also be illustrated in Figure 3 at the 9.10% value.

Tables 3 and 4 present sensibility analyses with respect to the parameters $\gamma(\theta)$ and $z(\theta)$. Up to now, we have indeed assumed that $z(\theta)$ is the fraud rate in the insurer portfolio and we have neglected the deterrence effect of the auditing policy. In fact, as shown in the theoretical part of the paper, $z(\theta)$ may differ between policyholders whose observable characteristics are correlated with their morale cost, a variable which is not observable. Table 4 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)

$z(\theta)$ = 5.88% 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 9.93%. The two other cases are not considered because their respective frequency is too low.

The results of Table 4 clearly show that fraud auditing intensifies (i^* decreases and $\lambda(i^*)$ increases) as the fraud rate $z(\theta)$ increases. In other words, the benefits of fighting fraud increase as $z(\theta)$ increases.

(Table 4 about here)

Deterrence effect and profitability of optimal auditing

The 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. In Table 2, the value of $\gamma(\theta)$ was fixed at zero, which means that the setting for the optimal audit policy took no account of the incentive effect of fraud deterrence. However, we have seen that the higher the probability Q^f of being detected, the lower the $\phi(Q^f)$ threshold and consequently the lower the level of fraud. When $\gamma(\theta)$ is positive, auditing expects such a deterrence effect on fraud. Proposition 2 has characterized the relationship between, on one side, the intensity of this deterrence effect and the fraud rate, and, on the other side, the optimal auditing strategy. It states that when the function $A(Q,\theta)$ is convex in Q, auditing is increasingly successful at reducing fraud costs as the fraud rate rises (higher direct benefit of auditing) or as the elasticity of fraud grows (higher deterrence or incentive effect). These results are illustrated in Table 4. We observe, for the three different values of the fraud rate $z(\theta)$, that i^* decreases (audit increases) when the elasticity of the fraud rate increases (in absolute value) with respect to Q^f. This is the deterrence effect: When the insurer chooses a tougher audit policy (i.e. a lower i^*), he decreases the threshold $\phi(Q^f)$ or the critical value of the morale cost under which cheating dominates honesty because the audit probability of fraudulent claims increases.

Table 5 presents the monetary gains of auditing with data from the insurer. As already mentioned, the claims rate (over the whole portfolio) of that 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 9.1% of the files will be audited at a cost of €280. Secondly, we also know that 68% of the fraudulent claims will be audited and will not receive any insurance coverage. However, 32% of the fraudulent claims would not be audited. The total claim cost net of audit costs will then be equal to €620 M, a saving of €22 M, which represents 43% 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 introduces the deterrence effect when $\eta(\theta)=0.23$ 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 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:

$$P(F \mid \sigma_i, \theta)\, t - c$$

where $P(F \mid \sigma_i, \theta)$ is given by (1), with

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

For illustrative purpose, consider the case $\hat{\tau}(\theta)=1.76\%$. If $\gamma=0$, there is no deterrence effect which gives:

$$P(F \mid \theta) = z(\theta) = 8\%.$$

Using $i^*(\theta)=238$ yields:

$$P\left(F \mid \sigma_{i^*(\theta)}, \theta\right) = 0.218 = \frac{c}{t}.$$

If $\gamma=0.10$, the elasticity of fraud with respect to the probability of being detected is $\eta(\theta)=0.23$. We then have $i^*(\theta)=224$ and $P\left(F \mid \sigma_{i^*(\theta)}, \theta\right) = 0.193 < \frac{c}{t}$.

In that case, the fraud rate is:

$$P(F \mid \theta) = 7.3\%$$

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 €46.86 M (i.e. 7.3% × €642 M) and the total claim cost net of audit costs is reduced to €618 M.

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 modeling. A complete modeling has been developed for the detection of insurance fraud, but the same methodology could easily be adapted to other hidden information problems, particularly those connected with banking 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 "red flags strategy" which consists in referring claims to an 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.

On the empirical side, four significant results were tested with data from a large European insurance company. First, we were able to compute a critical suspicion index for fraud, providing a threshold above which all claims must be audited. Secondly we showed that if the insurer implements this policy, 68% of the fraud claims are audited while only 4% of the no-fraud claims are audited. We showed that if the insurer applies this policy, he will save more than €22 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 morale cost of fraud and that auditing does not involve any deterrence effect.

Thirdly, we showed that it is possible to improve these results by using information on observable variables which are correlated with the intensity of fraudulent claiming. Such variables were identified by including in the regression analysis different indicators capable of isolating different groups of insureds with different morale costs of fraud. The sensitivity results show that more auditing should be applied to those with higher fraud rates or higher thresholds for the dominance of cheating over honesty. Our numerical results show that the optimal expected audit probability goes from 6.8% to 12.2% when the fraud rate goes from 5.9% (for a low fraud type) to 9.9% (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.

Chiappori, P.A. and Salanié, B. (2003), "Testing Contract Theory: A Survey of Some Recent Work," in *Advances in Economics and Econometrics, Theory and Applications*, Eight World Congress of the Econometric Society, edited by M. Dewatripont, L.P. Hansen and S. J. Turnovsky, vol. 1, 115-149.

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.

Crocker, K.J. and Tennyson, S. (1999), "Costly State Falsification or Verification? Theory and Evidence from Bodily Injury Liability Claims," in *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, G. Dionne and C. Laberge-Nadeau (Eds), Kluwer Academic Publishers, Boston.

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.

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.

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

Picard, P. (2000), "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.

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.

Appendix A
Detailed Description of Variables in Regression Analysis

q_{19}^{p}	Production of questionable or falsified documents (photocopies or duplicates of bills)
qexp	Fraud alert by expert
qext	Fraud alert by external organizations
q_{30}^{p}	Description of circumstances surrounding the accident either lack clarity or seem contrived
q_{21}^{p}	Variations in or additions to the policy-holders initial claims
q_{35}^{p}	Too long a lag between date vehicle was purchased and date of guarantee
q_{36}^{d}	Date of subscription to guarantee and/or date of its modification too close to date of accident
q_{7}^{p}	Vehicle whose value does not match income of policy-holder
q_{20}^{p}	Refusal or reluctance to provide original documents (registration of the car, mechanical check-list, maintenance record, sticker…) and/or comply with the insurer's requests
q_{16}^{p}	Victim with no damage insurance and/or one who would be wronged if found at fault
q_{22}^{p}	Harassment from policy-holder to obtain quick settlement of a claim
q_{32}^{d}	Abnormally high frequency of accidents (more than three accidents a year)
q_{12}^{d}	Retroactive effect of the contract or the guarantee
q_{34}^{p}	Person making the claim not the same as policy-holder
q_{18}^{d}	Delay in filing accident claim

Chart 1
Original Data Set

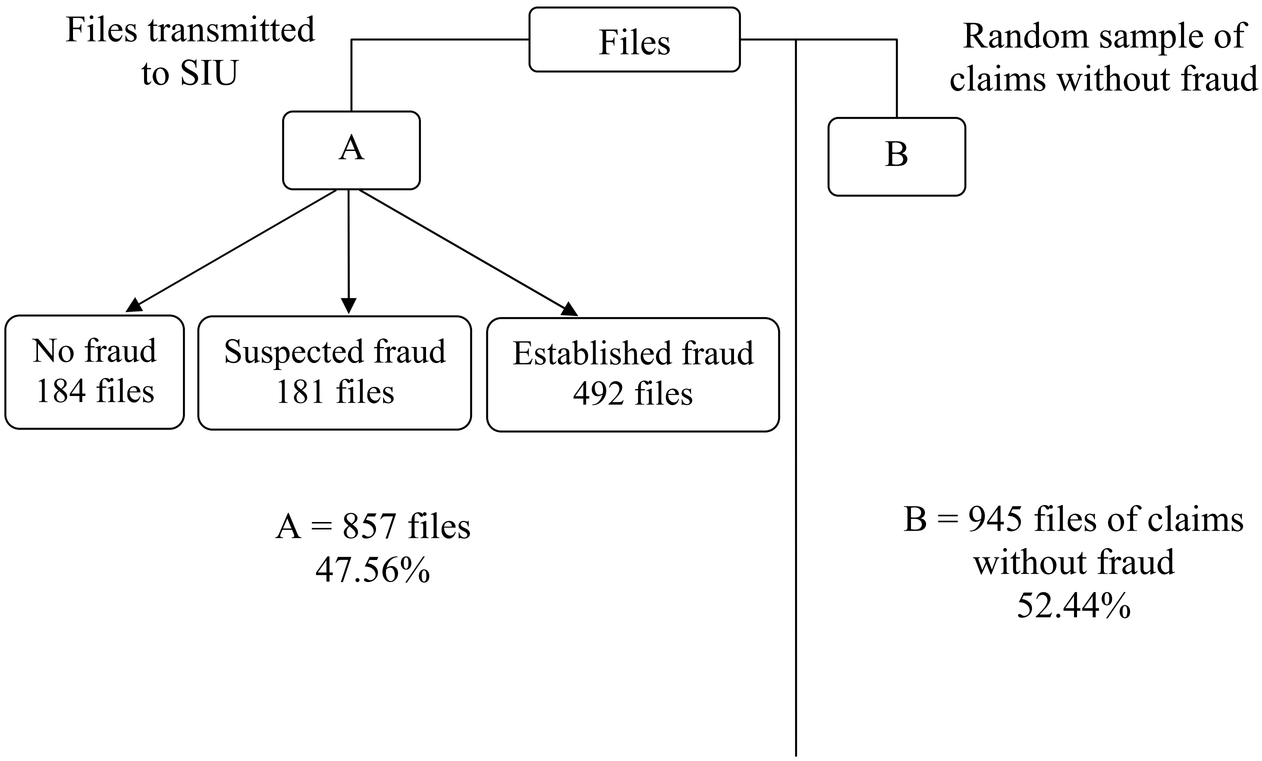

Chart 2
Original Data Set (continued)

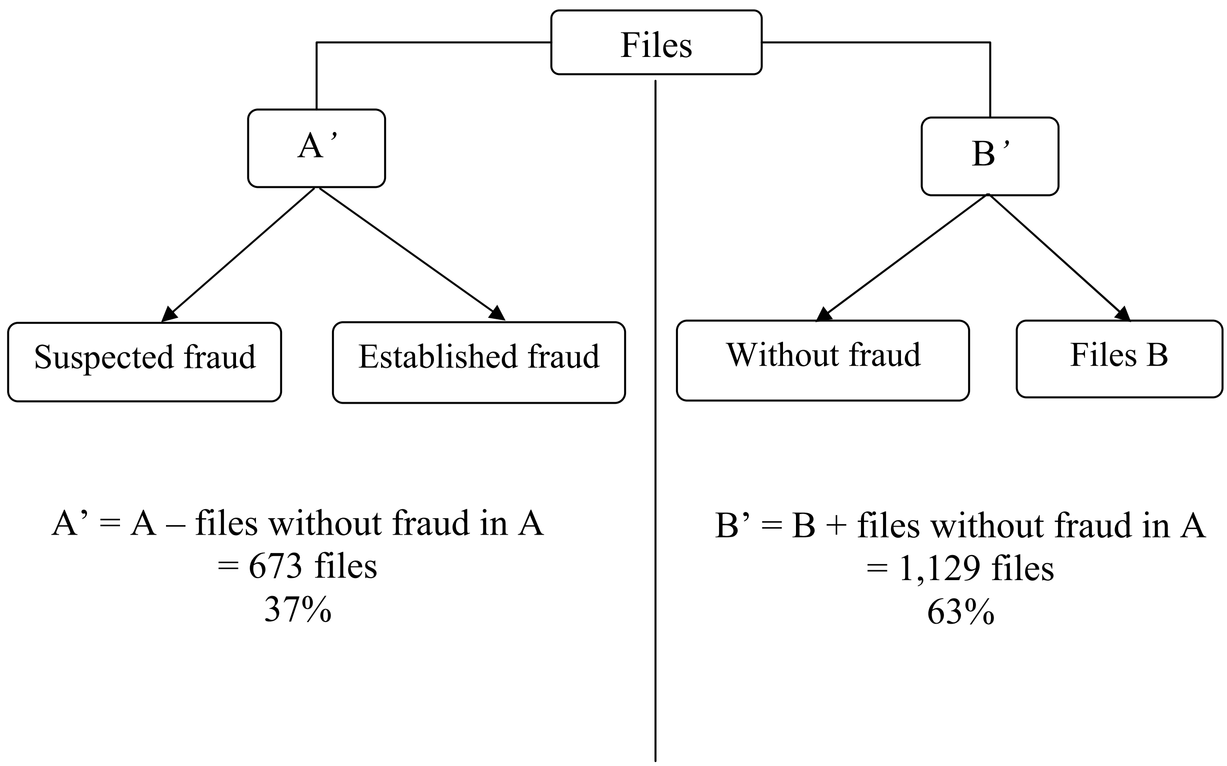

Chart 3
Data Set with Bootstrapping

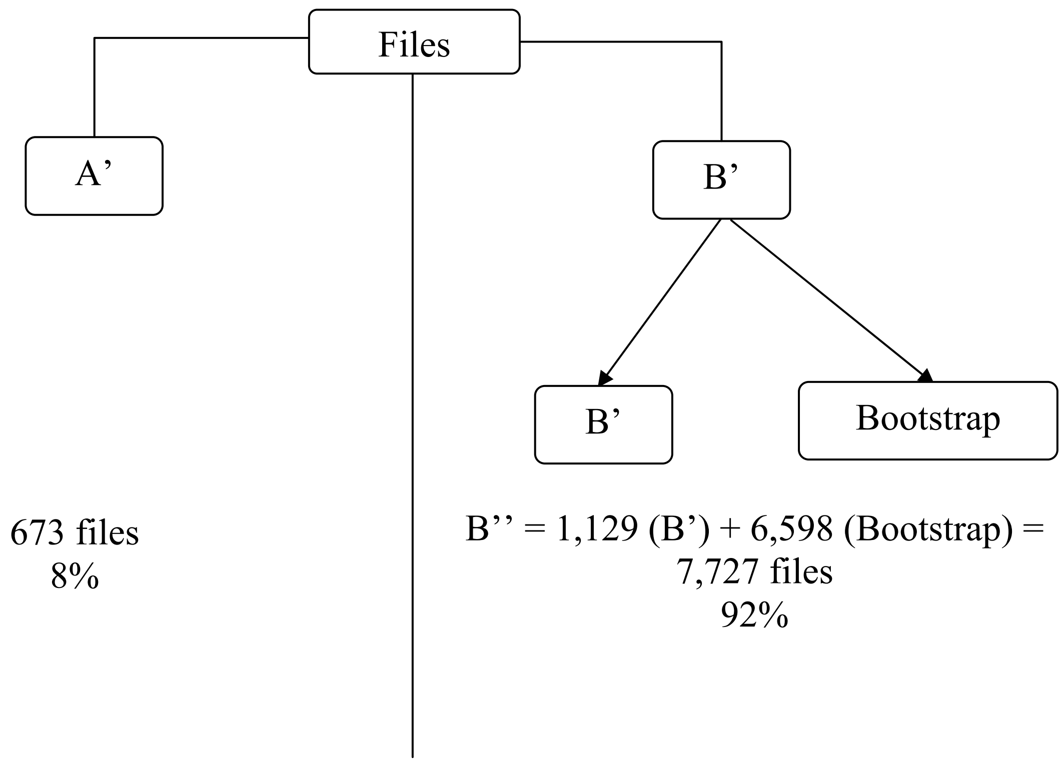

Table 1
Regression Results

		Without θ variables			**With θ variables**		
Symbol	Variable	Parameter	Std Error	P	Parameter	Std Error	P
	Intercept	2.7558	0.2120	<0.0001	3.1465	0.2443	<0.0001
q_{12}^{d}	Retroactive effect of the contract	0.1706	0.0829	0.0396	0.1837	0.0839	0.0285
q_{18}^{d}	Delay in filing accident claim	0.1371	0.0519	0.0083	0.1088	0.0526	0.0385
q_{19}^{p}	Falsified documents or duplicated bills	1.0813	0.0509	<0.0001	1.0771	0.0591	<0.0001
q_{21}^{p}	Variation in the initial claim	0.4677	0.0820	<0.0001	0.4562	0.0825	<0.0001
q_{30}^{p}	Lack of clarity in the accident description	0.4878	0.0604	<0.0001	0.4911	0.0607	<0.0001
q_{32}^{d}	Abnormal high frequency of accident	0.2343	0.0645	0.0003	0.2312	0.0647	0.0004
q_{36}^{d}	Date of subscription to guarantee too close	0.4227	0.0565	<0.0001	0.3870	0.0575	<0.0001
q_{35}^{p}	Too long lag between car purchase and guarantee	0.7014	0.1020	<0.0001	0.6085	0.1064	<0.0001
q_{34}^{p}	Claim maker different to policyholder	0.1701	0.0665	0.0105	0.1520	0.0669	0.0230
q_{22}^{p}	Harassment to obtain quick settlement	0.3257	0.1016	0.0013	0.2861	0.1030	0.0055
q_{20}^{p}	Reluctance to provide original documents	0.4177	0.0963	<0.0001	0.4117	0.0971	<0.0001
$qexp$	Fraud alert by expert	1.3929	0.0617	<0.0001	1.3835	0.0617	<0.0001
$qext$	Fraud alert by external organizations	1.2787	0.0990	<0.0001	1.2567	0.1013	<0.0001
q_{7}^{p}	Too high value of vehicle				0.1584	0.0529	0.0028
q_{16}^{p}	No damage insurance				0.5953	0.1436	<0.0001
Log Likelihood		-1,407.04			-1,394.04		
Number of observations		8,400			8,400		

Table 2
Calibration Results

1	**2**	**3**	**4**	**5**	**6**	**7**
i or $i*$	σ_i	$p_i^f \,/\, p_i^n$	$\lambda(i)$	$\mu(i)$	$(1) = c\,\pi\,\mu(i)$	$t - \lambda(i)(t-c)$
1	NNNNNNNNNNNNN	0.06	1.00000	1.00000	56.67200	80.00000
2	NNNNNNNYNNNNN	0.09	0.98548	0.77033	43.65614	294.57808
3	NNNNNNNNNYNNN	0.09	0.98201	0.73297	41.53888	298.06196
4	NNNNNNNNYNNNN	0.10	0.97777	0.68780	38.97900	302.31892
5	NNNNNNNNNNYNN	0.11	0.96277	0.53558	30.35239	317.37892
6	NNNNNNNYNYNNN	0.14	0.96075	0.51784	29.34703	319.40700
7	NNNNNNNYYNNNN	0.15	0.95974	0.51049	28.93049	320.42104
8	NNNNNNNNYYNNN	0.15	0.95615	0.48573	27.52729	324.02540
9	NNNNNNNYNNYNN	0.17	0.95177	0.45579	25.83053	328.42292
10	NNNNNNNNNYYNN	0.17	0.95129	0.45291	25.66732	328.90484
…						
…						
238	**NYNYNNNYNYYNN**	**3.20**	**0.68050**	**0.03976**	**2.25328**	**600.77800**
…						
…						
8192	YYYYYYYYYYYYY	7567.03	0.00214	0.00000	0.00000	1281.85144

Table 2
Calibration Results (continued)

8	**9**	**10**	**11**	**12**
$(2)=\hat{\tau}(\theta)(t-\lambda(i)(t-c))(1-\lambda(i))^0$	$(1)+(2)$	$(3)=z(\theta)\lambda(i)+(1-z(\theta))\mu(i)$	$c\times(3)$	$P(F/i>i^*)$
4.92800	61.60000	1.00000	280.00000	0.08000
5.18457	48.84072	0.78754	220.51176	0.10011
5.24589	46.78477	0.75289	210.81010	0.10435
5.32081	44.29981	0.71100	199.07933	0.11002
5.58587	35.93826	0.56976	159.53146	0.13518
5.62156	34.96859	0.55327	154.91638	0.13892
5.63941	34.56990	0.54643	153.00040	0.14051
5.70285	33.23014	0.52336	146.54181	0.14615
5.78024	31.61077	0.49547	138.73115	0.15368
5.78873	31.45604	0.49278	137.97851	0.15444
10.57369	**12.82697**	**0.09102**	**25.48538**	**0.59812**
22.56059	22.56059	0.00017	0.04794	1.00000

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

q_7^p / q_{16}^p	Yes	No
Yes	$z(\theta)$ = 51.50% (46 policyholders)	$z(\theta)$ = 26.37% (77 policyholders)
No	$z(\theta)$ = 9.93% (3506 policyholders)	$z(\theta)$ = 5.88% (4771 policyholders)

Table 4
Sensibility of Optimal Solutions With Respect to γ(θ) and z(θ)

Case	Parameters					Optimal Treshold	Expected Cost of Fraud	Expected Audit Probability	Audit Probabiliy of Defrauders
	$z(\theta)$	$\eta(\theta)^{(1)}$	$\gamma(\theta)$	π	$\tau(\theta)$	i^*	$\hat{\tau}(\theta)\,(1-\lambda(i^*))^{\gamma(\theta)} \times (t-\lambda(i^*)\,(t-c))$	$z(\theta)\,\lambda(i^*) + (1-z(\theta)) \times \mu(i^*)$	$\lambda(i^*)$
A	0.0800	0.00	0.00	0.2024	0.0176	238	12.83	0.0910	0.6805
B	0.0800	0.11	0.05	0.2024	0.0176	232	12.24	0.0952	0.6916
C	0.0800	0.23	0.10	0.2024	0.0176	224	11.68	0.0985	0.7004
D	0.0800	0.82	0.35	0.2024	0.0176	217	9.31	0.0995	0.7030
E	0.0588	0.00	0.00	0.2024	0.0126	304	9.74	0.0676	0.6485
F	0.0588	0.03	0.05	0.2024	0.0126	298	9.34	0.0680	0.6500
G	0.0588	0.19	0.10	0.2024	0.0126	293	8.96	0.0683	0.6509
H	0.0588	0.66	0.35	0.2024	0.0126	279	7.31	0.0693	0.6543
I	0.0993	0.00	0.00	0.2024	0.0242	183	16.62	0.1217	0.7240
J	0.0993	0.14	0.05	0.2024	0.0242	179	15.77	0.1221	0.7248
K	0.0993	0.26	0.10	0.2024	0.0242	179	14.98	0.1221	0.7248
L	0.0993	0.93	0.35	0.2024	0.0242	171	11.72	0.1234	0.7274

(1) $\eta(\theta) = -(\lambda(i^*) / 1-(\lambda(i^*))\ \gamma(\theta)$ in absolute value.

Table 5
Monetary Values of the Results
for the Insurer Portfolio

Without optimal audit the total claim cost =

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

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

$$9.10\% \times 500{,}000 \times € \ 280 + € \ 591 \text{ M} + 32\% \times €51 \text{ M} = €620.06 \text{ M}$$

With audit of all files we obtain:

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

When $\eta(\theta) = 0.23$, the optimal total claim cost become:

$$9.85\% \times 500{,}000 \times €280 + €591 \text{ M} + 30\% \times €46.86 \text{ M} = €618.43 \text{ M}$$

Figure 1

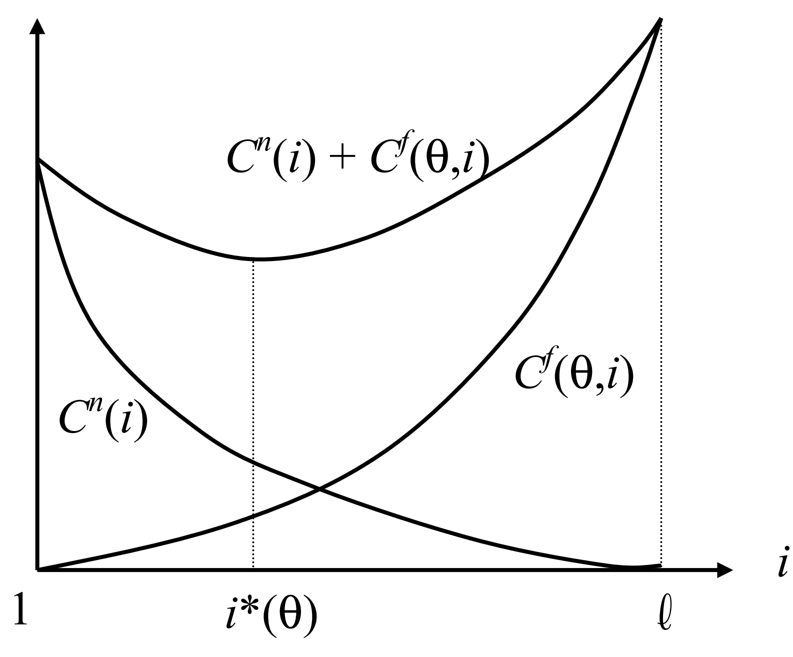

Figure 2

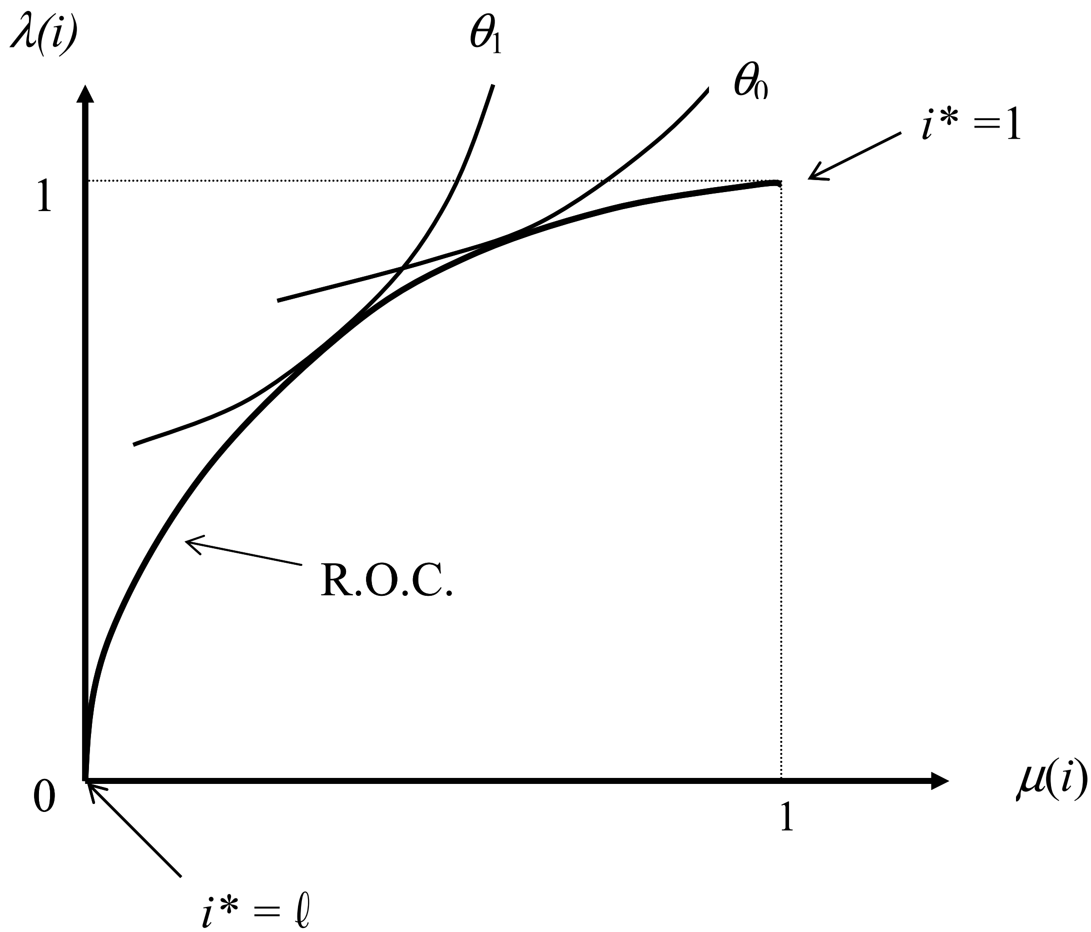

Figure 3
Gain Chart