The Impact of Insurance Fraud Detection Systems

Jörg Schiller
Institute of Risk and Insurance
University of Hamburg, Germany
phone: +49 40 428386143, fax: +49 40 428385505
schiller@econ.uni-hamburg.de

October 1, 2003
Abstract

The purpose of this paper is to characterize the impact of fraud detection systems on the auditing procedure and the equilibrium insurance contract, when a policyholder can report a loss that never occurred. Insurers can only detect fraudulent claims through a costly audit (costly state verification). With a fraud detection system insurers can condition their audits on the signal of the system and auditing becomes more effective. This paper presents conditions under which insurance fraud and the resulting welfare losses can be reduced by the implementation of a costly fraud detection system in a competitive insurance market that is supplied by an external third party.

Key words: insurance fraud, auditing, detection systems

JEL-Classification: D 82, G 22
The Impact of Insurance Fraud Detection Systems

1 Introduction

Fraud is a well known phenomenon in insurance markets. Policyholders have the ability to use their informational advantage about the occurrence of an insured loss to report losses that never have happened. Economists usually concentrate on an incentive compatible truth-telling mechanism that induces the agent to reveal his private information honestly. But in many situations that mechanism is not feasible, especially in the insurance fraud context, when the principal cannot commit ex ante to his audit strategy. Townsend [1979] introduced the costly state verification approach, which is used in many environments, like credit markets (e.g. Gale/Hellwig [1985]) and taxation problems (e.g. Mookherjee/Png [1989]). The most important shortcoming of these earlier approaches is that insurance fraud never happens in equilibrium, because it is assumed that the principal can credibly commit ex ante to an audit probability which makes fraud unattractive. The crucial assumption of that commitment approach means that the principal either decides prior to the agent’s action or that he can offer contracts with a specified audit probability. This kind of modeling is not very realistic in most auditing problems, because the principal would always revise his previous decision after the reception of the claims, if this is profitable.

Commitment is a very powerful tool for principals, but the analysis of auditing problems should concentrate on so-called strategic commitment devices. If the principal cannot credibly commit himself ex ante to an audit strategy, policyholders will have an incentive to report some fraudulent claims as Picard [1996] and Boyer [2000] show. The only possible equilibrium is a Perfect Bayesian Nash equilibrium (PBNE) in mixed strategies, where policyholders defraud and insurance companies audit randomly. A very interesting feature of the non-commitment problem is that policyholders are over-compensated for their loss, because the indemnity is the only device for insurers to limit fraud, as Boyer [2000] shows.

Informative noisy information systems are very common in principal-agent models. Firstly, this paper concentrates on the impact of fraud detection systems on the auditing procedure of insurers and on the over-compensation of policyholders. Although some papers, like Belhadjj/Dionne/Tarkhani [2000] and Artis/Ayuso/Guillén [2002], analyzed fraud detection models, to the best of our knowledge only the paper of Dionne/Guilliano/Picard [2003] has taken information systems into account. This is surprising, because these systems are very popular and effective in existing insurance markets with considerable size, like the auto insurance market. Due to the system, the know-how of fraud experts can be duplicated and
suspicious claims can be identified more easily. Furthermore, an insurer can concentrate his audits on more suspicious claims and will obviously not audit all claims with the same probability and can therefore improve the effectiveness of his audits. The findings of Dionne/Guiliano/Picard [2003] concerning the optimal audit strategy corresponds to ours, but they focus on the optimal audit strategy and neither regard the impact of the system on the underlying insurance contract, nor further implementation problems of the system. Our approach is similar to that of Macho-Stadler/Pérez-Castrillo [2002], but they refer to tax audits with commitment and risk-neutral agents. We consider a fraud detection system, which provides noisy information about the true state of the world. The information of the system is observed privately by the insurer and is therefore non-contractible. This information is modeled similar to Holmström [1979], as an additional signal that cannot be intentionally manipulated by the policyholder.1

The insurer’s inability to commit himself credibly causes an inevitable market inefficiency. Some proposed solutions for that non-commitment problem concentrate on the strategic commitment force of external third parties. For example, a common agency that takes wholly or partly charge of the insurer’s audit costs and is financed ex ante by participating companies might help to decrease fraud as Picard [1996] shows. In contrast, Melumad/Mookherjee [1989] suggest that insurers can simply sign an incentive-compatible audit contract with an investigator and can therefore commit themselves credibly to any desired audit probability. Both approaches to overcome the non-commitment problem are not universally applicable. The ex ante collection of auditing expenditures by the common agency is impossible in non-regulated markets, where companies compete in premiums, because the transfer to the agency is sunk at the competition stage. The latter solution of Melumad/Mookherjee rests as well upon very critical assumptions. First of all, the underlying contract is not renegotiation-proof, and the principal has incentives to renegotiate the delegation contract. Secondly, the authors abstract from the moral hazard problem between the principal and the external investigator.

The main goal of this paper is to analyze the following problems. What effects does a detection system have on the auditing game and the underlying insurance contract? How can a detection system with fixed costs be implemented on a competitive insurance market without any market intervention? As our first result shows, an informative system leads in equilibrium - compared to a situation without fraud detection - to a lower fraud and audit probability, which is quite intuitive. As a further consequence, the strategic incentive device over-compensation can be reduced with an informative system, because auditing becomes more effective. The reduction of the fraud and the audit probability as well as the reduction of over-compensation depends only on the quality of the signal. We explore conditions under which a fraud detection system will be applied and give, in addition to Boyer [2000]
and Picard [1996], a new motivation for the use of an external party, like a private supplier or an Insurance Fraud Bureau (IFB). The important role of the third party in our model is to transform the prevailing fixed implementation costs of a fraud detection system, which can lead to a market breakdown, if firms compete in prices a la Bertrand.

The remainder of the paper is organized as follows. In section 2 we present the basic model and the optimal insurance contract without a fraud detection system. Section 3 concentrates on the effects of a fraud detection system on the equilibrium of the audit game and the underlying insurance contract. Furthermore, in section 4 we derive conditions under which a system will be implemented and show how the fraud detection should be organized. The conclusions follow in section 5.

2 Equilibrium without fraud detection

2.1 The model and the sequence of play

We assume an insurance market with free entry, where \( I \geq 2 \) insurance companies compete through premium offers. The traded contracts \( C = (\alpha, \beta) \in \Psi \) consist of the premium \( \alpha \in R^+_0 \) and an indemnity \( \beta \in R^+_0 \). Insurers are risk-neutral, they face homogenous and independent risks of the policyholders, and therefore they have constant marginal costs. If insurers charge the same premium, they will share the market equally. In the Bertrand equilibrium premium offers correspond to the expected costs of a policyholder and all companies will charge the same premium. In absence of any fixed costs, what we will suppose until section 4, insurers make zero expected profits \( \Pi(\cdot) = 0 \) in equilibrium. The \( N \) risk-averse policyholders are homogenous. Therefore, they have the same attitude towards insurance fraud and the same initial income \( W_0 \). Additionally, each policyholder possesses a continuous and twice differentiable utility function \( u(W) \) with \( u'(\cdot) > 0 \) and \( u''(\cdot) < 0 \). By their contract choice they maximize the von Neumann-Morgenstern utility function \( U = E[u(W)] \), which is a continuous, twice differentiable and concave function throughout \( \Psi \).

There are only two states of the world, ”no accident” \( (\omega_0) \) and ”accident” \( (\omega_1) \), with \( \{\omega_0, \omega_1\} = \Omega \). The probability of an accident is \( \pi \) and the loss in the state of an accident is \( L > 0 \). It is assumed that the loss is not higher than the policyholder’s initial income \( L < W_0 \), and that insured consumers have private information about the state of world. After the policyholder has observed the move of nature, he decides whether to file a claim \( (m_1) \) or not \( (m_0) \), with \( \{m_0, m_1\} = M \). Consequently, the insurance company has to decide, whether to audit a received claim \( (a_1) \) or not \( (a_0) \), with \( \{a_0, a_1\} = A \). The audit is assumed to be perfect, thus after an audit the insurer can observe the state of world. The audit costs, which the insurer
has to bear under any circumstances and which can not be influenced by policyholders, are \( c > 0 \). Policyholders who are caught defrauding must pay a penalty \( k \) that is sunk and is not collected by the insurance company.\(^4\)

The sequence of play is:

- **Stage 1:** Insurance companies simultaneously offer insurance contracts \( C \);
- **Stage 2:** The nature chooses the state of the world. This move is private information of the policyholder;
- **Stage 3:** The policyholder decides whether or not to file a claim;
- **Stage 4:** The insurance company decides whether or not to audit a claim;
- **Stage 5:** The payoffs are paid and the game ends.

The PBNE of the game consists of sequentially rational strategies given a belief system \( \mu \) that is derived from the strategy profile through Bayes rule whenever possible.

### 2.2 The claiming game

Policyholders can observe the state of the world. In equilibrium it is quite obvious, that every policyholder who suffered a loss will file a claim. The other way around it is trivial, that it will never be a best response for the insurance company to audit, if no claim was made.\(^5\)

**Lemma 1** The only PBNE in mixed strategies has the following properties:

- **The insurance company has sequential consistent ex ante beliefs** \( \mu = \frac{\pi}{\beta} \), that a filed claim is fraudulent;
- **Policyholders will always file a claim, if an insured loss occurred. Otherwise, if they have not suffered a loss, they will file a fraudulent claim with a probability**

\[
\eta^* = \left( \frac{\pi}{1-\pi} \right) \frac{c}{\beta - c}; \tag{1}
\]

- **Insurance companies audit all claims with the same probability**

\[
\nu^* = \frac{u(W_0 - \alpha + \beta) - u(W_0 - \alpha)}{u(W_0 - \alpha + \beta) - u(W_0 - \alpha - k)}. \tag{2}
\]
Proof. See Boyer [2000] ■

In equilibrium all claims are audited with the same probability and the policyholders who suffered a loss will always file a claim and get the indemnity, whether the insurer audits or not. But those, who have not suffered a loss, file a fraudulent claim with probability $\eta^*$. An important point with respect to the following analysis in section 3.3 is that insurers choose their audit probability in order to make policyholders without a loss indifferent between filing a claim and not filing a claim. But the relevant probability for these policyholders is the probability that a fraudulent claim is detected, because their payoffs depend on that probability. Therefore, insurers have to choose an audit strategy in order to reach a critical detection probability $\kappa^c$, which makes policyholders without a loss indifferent between their strategies. In the case without fraud detection this distinction is irrelevant, because $\kappa^c$ corresponds to the optimal auditing probability (2).

2.3 The equilibrium insurance contract

The insurer has to design a contract as a combination of coverage and premium which maximizes the policyholder’s expected utility. Because of the non-existence of any fixed costs, the unique symmetric equilibrium premium offer of all $I$ insurers corresponds to the constant marginal costs of a policy. Since the premium equals the marginal costs of a policy, insurers make an expected profit of zero. Until section 4, we will abstain from taking a closer look at the competition between insurers and therefore impose a zero expected profits constraint. The designed contract, in particular the indemnity $\beta$, alters the payoffs of the policyholder and influences the equilibrium randomization of the insurer and the policyholder. In the last section the equilibrium strategies for a given contract were specified. If insurance is sold at a fair premium, risk-averse consumers maximize their utility by a full-insurance contract, but in the considered auditing context over-compensation is optimal, as the following Lemma shows.

Lemma 2 In competitive insurance markets without fraud detection systems the unique utility maximizing contract $C^* = (\pi^\beta^*, \beta^*)$ entails over-insurance with $\beta^* > L$.

Proof. See Appendix. ■
At first glance it seems quite surprising that the equilibrium contract entails over-compensation. There are two questions:

- Why should the insurer supply contracts that give policyholders more incentives to commit insurance fraud?
- Why do risk-averse consumers maximize their utility by taking an income risk?

Firstly, over-compensating the loss does not only provide additional incentives for fraud, but it also makes auditing more attractive for insurers. The difference between the audit costs and the possible savings from detected fraudulent claims increases ceteris paribus with an increasing coverage. Therefore, the equilibrium fraud probability diminishes with an increase of coverage $\beta$ as (3) shows.

$$\frac{\partial \eta^*}{\partial \beta} = -\left( \frac{\pi}{1 - \pi} \right) \frac{c}{|\beta - c|^2} < 0$$

(3)

Over-compensation is optimal for policyholders, because the slope of the zero-profit premium function is always smaller than that of the fair premium $\alpha = \pi \beta$. Hence, for $\beta < L$ an increase in coverage raises the expected payoffs and reduces the income risk of a policyholder. Due to the risk-neutrality of policyholders at $\beta = L$, the increase of the expected payoffs leads to a positive marginal utility. Therefore, the optimal indemnity must be greater than the loss. For $\beta > L$, there is a trade-off between increasing the expected payoffs and risk taking, and there exists a unique optimal contract with $\beta^* > L$.

The effects of over-compensation are similar to that of the unlimited liability discussion concerning principal-agent problems. The standard principal-agent problem has an approximate first-best solution without limited liability, if the penalty is set infinitely high, because that drastic penalty deters the agent from cheating. In our context the solution works the other way around. If the indemnity tends to infinity, insurance companies cannot allow that policyholders defraud, and therefore have to increase their audit probability which causes a decrease of the fraud probability. The indemnity cannot be increased infinitely, because the risk-averse policyholder has to bear an income risk for all indemnities $\beta > L$. The binding constraint is the utility maximization of the policyholder.
3 Equilibrium with a fraud detection system

3.1 Modifications

Let us assume that insurance companies are now able to implement a fraud detection system with a given technology that assigns an exogenous signal $s$ to each reported loss. For the moment, we desist from the costs of the fraud detection system and specify its properties. But later on, in section 4, we examine the consequences that the costs of the system might have.

The fraud detection system generates one of two possible signals $\{s_0, s_1\} \in S$, when a policyholder files a claim. The following analysis concentrates on the case $m_1 \in M$. The whole set of claims consists of two different types of reports:

- truthful claims with the probability $p(\omega_1) = \pi$ and
- fraudulent claims with the probability $p(\omega_0) = (1 - \pi)\eta$.

In addition to a filed claim, the insurance company receives a private signal from the fraud detection system. The properties of the fraud detection system are common knowledge. The conditional probabilities of the signal $s \in S$ given the state of world $\omega \in \Omega$ are:

\[
\begin{array}{l}
\text{[Insert Table 1 here]} \\

\text{The probabilities imply that the fraction } \delta \text{ (respectively } (1 - \delta)) \text{ of all honest claims get the signal } s_1 (s_0) \text{ and that the fraction } \phi \text{ (resp. } (1 - \phi)) \text{ of all fraudulent claims get the signal } s_1 (s_0). \text{ The resulting probabilities of the signals are}
\end{array}
\]

\[
p(s_1) = \delta \pi + \phi(1 - \pi)\eta
\text{ (4)}
\]

and

\[
p(s_0) = (1 - \delta)\pi + (1 - \phi)(1 - \pi)\eta
\text{ (5)}
\]

respectively.

We assume, that insurance companies will update their fraud beliefs $\mu$ according to Bayes rule. The posterior fraud beliefs $\mu(s)$ of insurers after observing the signal $s$ are

\[
\mu(s_1) = \frac{\phi(1 - \pi)\eta}{\delta \pi + \phi(1 - \pi)\eta}
\text{ (6)}
\]

and

\[
\mu(s_0) = \frac{(1 - \phi)(1 - \pi)\eta}{(1 - \delta)\pi + (1 - \phi)(1 - \pi)\eta}
\text{ (7)}
\]
3.2 Sequence of play

The earlier stated sequence of play must be modified with respect to the fraud detection system. After the implementation of a system, insurers are able to audit contingent on the signal. Consequently, after the observation of the signal they have to decide whether to audit a claim or to pay the indemnity immediately without any audit.

The sequence of play is as follows:

- **Stage 1**: Insurance companies simultaneously offer insurance contracts $C$;
- **Stage 2**: The nature chooses the state of the world. This move is private information of the policyholder;
- **Stage 3**: The policyholder decides whether or not to file a claim;
- **Stage 4**: When the policyholder files a claim, the insurance company receives a private signal $s \in S$. Otherwise the payoffs are paid and the game ends;
- **Stage 5**: The insurance company decides whether or not to audit a claim;
- **Stage 6**: The payoffs are paid and the game ends.

3.3 The claiming game

The signal of a fraud detection system should carry some relevant information about the probability that a particular claim is fraudulent. We assume without loss of generality that the system is valuable or informative in the sense of Holmström [1979] and $\phi > \delta$ holds. Since the fraction of fraudulent claims in the set of claims with the signal $s_1$ is higher than in set of claims with the signal $s_0$, it is obvious that fraud beliefs of insurers are higher after the observation of the signal $s_1$ than after the observation of $s_0$. In order to simplify the analysis, we assume that $\phi$ is greater than the crucial detection probability $\kappa^c$. Consequently, insurers exclusively audit claims with the signal $s_1$, as the following Lemma states.

**Lemma 3** **Insurers will exclusively audit claims with the signal $s_1$, if $\phi \geq \kappa^c$ holds.**

**Proof.** See Appendix. ■

If the fraction of fraudulent claims in the set of claims with the signal $s_1$ is higher than the critical detection probability, an insurer will always be able to make policyholders without a loss indifferent between their strategies
by exclusively auditing claims with the signal $s_1$. Policyholders have to take the equilibrium audit strategy of the insurance company, and particularly, the technology of the detection system $(\delta, \phi)$ into their account. The latter is very important, because it has a major influence on the exposure of fraudulent claims. As mentioned earlier, the signal has an impact on the insurer’s incentive to audit. However, after the implementation of the fraud detection system the structure of the game is unaffected.

**Proposition 1** If $\phi \geq \kappa^e$ holds, the unique PBNE for a given fraud detection system with $\phi > \delta$ has the following properties:

- The insurance company has an ex ante sequential consistent fraud belief $\mu = \frac{\delta c}{\phi \beta - (\phi - \delta) c}$.
- The equilibrium fraud probability is
  \[
  \eta^* = \left( \frac{\pi}{1 - \pi} \right) \left( \frac{\delta c}{\phi \beta - (\phi - \delta) c} \right):
  \]
  \[
  \tag{8}
  \]
- The insurer audits only claims with the signal $s_1$ with the probability
  \[
  \nu^*(s_1) = \frac{1}{\phi} \frac{u(W_0 - \alpha + \beta) - u(W_0 - \alpha)}{u(W_0 - \alpha + \beta) - u(W_0 - \alpha - k)} = \frac{\kappa^e}{\phi},
  \]
  which leads to an overall audit probability
  \[
  \nu^* = \frac{\delta \beta}{\phi \beta - (\phi - \delta) c} \kappa^e.
  \]
  \[
  \tag{10}
  \]

**Proof.** See Appendix. □

Not surprisingly, a perfect fraud detection system $(\delta = 0)$ leads to a fraud probability $\eta^* = 0$, because there is no trade-off for insurers, when they audit claims with the signal $s_1$, since all claims with that signal are fraudulent. Auditing all claims with the signal $s_1$ leads to a higher detection than the critical detection probability. Therefore, it is a dominant strategy for policyholders to only file honest claims. Thus, in equilibrium the insurance company does not receive the signal $s_1$ and no audit is necessary. In any other case the system is not perfect, and in equilibrium insurers audit and policyholders defraud with a positive probability.

An increase (decrease) of $\phi$ leads to a decreasing (increasing) fraud, conditional audit and overall audit probability. An increase (decrease) of $\delta$ raises (diminishes) the fraud and overall audit probability, since the fraction of honest claims in the set of claims with the signal $s_1$ increases (decreases). At first glance, it is surprising that the conditional probability $\nu^*(s_1)$ is
unchanged by a variation of $\delta$. But policyholders without a loss choose their fraud probability in order to make insurers after the observation of the signal $s_1$ indifferent between their strategies. Hence, they compensate the lower or higher incentives of insurers to audit claims with the signal $s_1$ by a variation of the fraud probability. The overall audit probability rises (declines) ceteris paribus with an increasing (decreasing) fraction $\delta$, because insurers get the signal $s_1$ more (less) often and audit all claims with that signal $s_1$ with the same conditional probability as before. This argumentation explains why the fraud probability is only zero, when $\delta = 0$ holds. This case is the only situation where policyholders without a loss are not able to keep insurers indifferent between their strategies.

3.4 The equilibrium insurance contract

The signal is non-contractible, because it is private information of the insurer. For that reason, we can derive the optimal insurance contract $\hat{C}$ in the same way as in section 2.3 and maximize the expected utility of the policyholder under the zero profit constraint of the insurance company.

Corollary 1 If there is an informative fraud detection system with $\phi > \delta$, the equilibrium contract $\hat{C} \left( \frac{\pi \beta c + \delta c}{\beta - c}, \hat{\beta} \right)$ has the following properties:

- As long as $\delta > 0$, the system is imperfect, and the optimal contract entails over-insurance with $\hat{\beta} > L$;
- If and only if $\delta = 0$, the system is perfect, and the optimal contract entails a fair premium and full insurance with $\hat{\beta} = L$;
- The premium and the indemnity increase (decrease) ceteris paribus in $\delta$ ($\phi$). Consequently, the optimal contract with an informative detection system entails less over-insurance and a lower premium than the optimal contract without fraud detection.

Proof. See Appendix. ■

The motivation of over-compensation as given in section 2.3 still holds, but the extent of over-compensation for an informative fraud detection system can be reduced, due to the decreased incentives of the policyholder to commit fraud. For that reason, over-compensation and fraud detection are two strategic devices that help to reduce insurance fraud. Our findings are in the spirit of Demougin/Fluet [2001]. Among other things, they analyze the impact of different information gathering costs (like monitoring or
auditing) on the optimal auditing-incentive-mix in a hidden action principal-agent relationship, where both parties are risk-neutral and the agent faces a limited liability constraint. In our model the agent is risk-averse, and the exogenous penalty is an implicit limited liability constraint. Therefore, the insurer as the principal has to use other monetary incentives to reduce the costs of fraud. Another difference compared to our approach is that over-compensation is used as an indirect incentive for the principal to audit and not as an incentive for the agent to choose a higher effort level. The results are very similar, because, if an informative signal is available, the principal will use more auditing and less monetary incentives to reduce the information rent of the agent. The meaning of ”more auditing” is only valid in a wider sense here, because of the strategic auditing game. The better information of the insurer however leads to a reduction of over-compensation.

4 The organizational design of fraud detection

The effectiveness of the fraud combat is very sensitive to its organizational design. There are many approaches addressing whether or how a third party can help to mitigate the market inefficiency caused by the non-commitment problem of the principal. In Picard’s view, the non-commitment problem can (partly) be solved by a common agency created for all insurance companies, that does not audit, but (partly) takes charge of the audit expenditures and is financed by the whole market through partition fees. However, such a solution is only valid for regulated insurance markets. A different approach employed by Melumad/Mookherjee [1989] suggests, for the case of a tax audit, that the authority can delegate the auditing job to an independent investigation agency. But such a delegation causes additional agency costs, because the investigator gets an incentive compatible audit contract. This solution is not really convincing for two reasons that the authors concede. Firstly, it is not renegotiation-proof, because the insurer has an incentive to privately renegotiate the contract with the audit agency after the audit contract is publicly signed. Secondly, Melumad/Mookherjee abstract from the moral hazard problem between the insurer and the agency. In light of this criticism, their results cannot really explain the existence of an IFB, as the non-commitment problem also arises in the proposed solutions. A further weighty shortcoming of both solutions is that they disregard from the resulting costs and, more importantly, from the impact of the costs on the market equilibrium.

During this section, we present how a fraud detection system can and will be implemented in insurance markets when it mainly causes fixed costs. Our approach will fit to the two oppositional legal frameworks for markets in the United States (US) and the European Union (EU). In the United States many federal states have established State Insurance Fraud Bureaus that
support the fight against insurance fraud or perform investigations. Such a market intervention can be welfare improving, when insurers face different auditing costs, as Boyer [2000] shows. Within the EU, it is nearly impossible to enforce a market intervention and/or merely a private cooperation of insurers. Every company performs its own audits, and there is no way to commonly develop or run a detection system. But as we can observe, there are some ways to implement such a system. For example, in Germany a commercial reinsurer developed a system and sells it to European insurance companies even when they are not their customers.

In contrast to Boyer [2000], we do not consider different audit costs of insurers, because we have assumed that they are completely homogenous. A detection system which can be implemented assigns a signal to all claims prior to an audit decision. Insurers have to accept or reject a participation contract offer of a supplier to pay a fixed charge for using the system and/or a variable charge for each signal that the system refers. Within this paper, we characterize situations where the system can and will be implemented in competitive insurance markets, but our solution is also valid in regulated markets.

In addition to the information about the probability of a fraudulent claim, the fraud detection system will usually generate relevant information for the audit itself. Therefore, we could assume that the audit costs after the implementation of the systems would be lower than before. But the system will cause some costs of data input, because the relevant facts of the claim must be available for the system. Finally, we suppose that the two effects lead to the same audit costs $c$. More relevant than the per unit audit costs are the fixed costs of developing and implementing the system. For the remainder of the section we will concentrate on the arising problem of the fixed costs of fraud detection.

In the presence of fixed costs, the existence of a competitive market equilibrium is questionable, because the fixed costs are sunk at the premium competition stage. In the case where only one insurer can invent and implement the system at some fixed costs $D$, he will undercut all other companies and will force them to withdraw from the market, because of his reduced marginal costs and charges $\alpha_M = \alpha - \varepsilon$. This market configuration is only feasible in the sense of Baumol/Panzar/Willig [1982], if the market has a considerable size and no capacity constraint is present. In equilibrium the monopoly insurer will make an expected profit:

$$\Pi = N [\hat{\alpha} - \alpha_M] - D \geq 0,$$

for some $N$.

For the remainder of the section, we will concentrate on the case, where an external supplier invents and implements an informative fraud detection system with a given technology. To simplify matters, we set $D = 1$ without
loss of generality. The supplier offers every insurer a participation contract 
\( P = (F, f) \) that entails a fixed fee \( F \) for using the system and a variable 
fee \( f \) for every signal \( s \) that the system refers. The sequence of play is as 
follows:

- **Stage 1:** The supplier offers the contract \( P \) to all insurers;
- **Stage 2:** The insurers decide simultaneously whether or not to accept 
the offered contract;
- **Stage 3:** Insurance companies simultaneously offer insurance contracts 
\( C \);
- **Stage 4:** The nature chooses the state of the world. This move is 
private information of the policyholder;
- **Stage 5:** The policyholder decides whether or not to file a claim;
- **Stage 6:** If the policyholder files a claim, the insurance companies that 
participated at stage 2, receive a private signal \( s \in S \);
- **Stage 7:** The insurance companies decide whether or not to audit a 
claim;
- **Stage 8:** The payoffs are paid and the game ends.

We only analyze stages 1-3, because they determine whether the system 
can be implemented or not. There are only two possible types \( t_j, j = 0, d \) 
of insurance companies that can offer insurance contracts, where 0 denotes 
companies without and \( d \) companies with a fraud detection system. Next, 
we exclude dominated participation contracts, that can never be a part of 
an equilibrium of the whole game of stages 1-3.

**Lemma 4** The optimal participation offer entails no fix participation fee \( F \).

**Proof.** See Appendix □

Due to the Bertrand competition, there is no way that insurers can engage 
in activities that cause any fixed costs, because at the competition stage 
such fixed costs are sunk. Consequently, the insurer can only offer insurance 
coverage at marginal costs and makes expected losses in the amount of the 
fixed costs. The only possibility to implement the system is to transform 
fixed into variable costs. Hence, the external supplier should implement the 
system centrally and charge insurers a fee \( f \) for each received signal. A 
market equilibrium with fraud detection can only be reached, if:
i. The market configuration with fraud detection is feasible and sustainable, and consequently the marginal costs with fraud detection are lower than without.

\[
\pi \frac{\phi \beta - [\phi - \delta] c}{\phi [\beta - c]} \leq \pi \frac{\beta}{\beta - c} \quad \text{(C1)}
\]

ii. The external supplier makes non-negative profits.

\[
f \geq \frac{1}{N \pi} \frac{\phi \beta - [\phi - \delta]}{\phi [\beta - c]}
\quad \text{(C2)}
\]

If condition (C1) holds, there will be no market configuration where an insurance company can make expected profits by not participating in the fraud detection system and undercutting the other companies. On the other hand, as can be seen from (C2), if the market size (number of policyholders) does not exceed a critical value, no fraud detection will be introduced.

**Proposition 2** A sustainable market equilibrium with fraud detection by all insurance companies is feasible, if and only if the insurance market has a considerable size.

**Proof.** In Bertrand equilibrium the premium offer corresponds to the marginal costs of the insurer with the lowest marginal costs. Due to Corollary 1 for some small \( f \) (C1) will hold. Therefore, if (C1) holds, no insurer, whether he participates or not, has an incentive to undercut the premium \( \alpha^d = \pi \frac{\phi \beta - [\phi - \delta]}{\phi [\beta - c]} \leq [\beta + f] \), because he would make expected losses with such an offer. If (C1) holds, it is a dominant strategy for every insurance company to participate. Since \( f \) corresponds to the relative development costs, the market size is crucial, because \( f \) vanishes, if \( N \) tends to infinity. Ceteris paribus \( f \) declines with a rising market size, and therefore fraud detection is becoming more advantageous.

Insurers will only accept participation contracts with a variable and no fixed fee. If the insurance premium with fraud detection and the variable fee is not higher than the premium without fraud detection, all insurers will accept the participation contract. The supplier will only offer such a system, if he makes expected non-negative profits with the system. Although we disregarded from the strategic problem of the external supplier and only considered a break even situation, Proposition 2 can explain the fact that fraud detection systems are only used in insurance market with a high number of policyholders.
5 Conclusions

One possible strategic commitment device to reduce fraud in the non-commitment context is over-compensation of the agent, because it commits the insurer to audit more often. In this paper, we have characterized the impact of a fraud detection system on the equilibrium of the audit game and the insurance contract. We show that informative fraud detection systems are another possible strategic commitment device to reduce the fraud and audit probability in the economy, because insurers can condition their audits on the signal and do not audit all claims with the same probability. Due to the properties of Nash equilibria in mixed strategies, the detection system affects the randomization of the policyholders and insurers. Since the fraud probability and the over-compensation of the policyholder diminish with respect to the signal quality, the non-commitment problem of insurers remains as long as the signal is imperfect.

The amount and structure of the costs of a detection system is very crucial in our context. Fraud detection systems base on computer programs that cause high implementation costs. Operating a system will only be remunerative, if the benefits from the declining fraud and audit probability exceed the relative costs of the system related to an individual consumer. For that reason, the number of policyholders on the market is critical for the effectiveness of fraud detection. Thus, we can confirm the fact that fraud detection is only performed in insurance markets with a considerable size, such as the auto insurance and the workers compensation insurance market. Moreover, we deduce, that due to the fixed costs of the system and the Bertrand competition on the considered insurance market, the contract between insurers and the external supplier of the system can only consist of a variable transfer from the insurer to the supplier.

The analysis neglects two aspects. First of all, we do not consider any negotiations. As Macho-Stadler/Pérez-Castrillo [2001] show for the commitment audit case, a fraud detection system which generates an informative signal, gives the insurer incentives for renegotiations with the policyholder. After observing the signal, insurers have an incentive to settle with policyholders in order to save audit expenditures. The settlement is ex post efficient, but dilutes the deterrence effect of the fraud detection system. Future research should concentrate on the renegotiation incentives and their effects in a non-commitment environment. Another shortcoming of our approach is the binary signal of the system. Realistically, the model should be enlarged and should consider more than two signals and their impact on the auditing game. Such an extension will presumably not cause any structural changes of the equilibrium properties, as the results of Dionne/Guilliano/Picard [2003] indicate, because insurers would start auditing claims with the highest fraud belief and go on with claims with the second highest belief and so on until they reach the critical detection probability.
Appendix

Proof of Lemma 2. The optimal insurance contract maximizes the expected utility of policyholders

\[ U = \pi u(W_0 - \alpha - L + \beta) + (1 - \pi) (1 - \eta) u(W_0 - \alpha) \]
\[ + (1 - \pi) \eta [(1 - \nu) u(W_0 - \alpha + \beta) + \nu u(W_0 - \alpha - k)] \]

under the zero-profit constraint

\[ \Pi = \alpha - \pi \beta - (1 - \pi)(1 - \nu) \eta \beta - \nu \beta [\pi + (1 - \pi) \eta] c = 0. \]

The insurance premium can be expressed as a function of the indemnity by rearranging the zero-profit condition of insurance companies.

\[ \alpha^*(\beta) = \pi \beta + (1 - \pi)(1 - \nu^*) \eta^* \beta + \nu^* [\pi + (1 - \pi) \eta^*] c \]

With (1) it simplifies to

\[ \alpha^*(\beta) = \pi \beta \frac{\beta}{\beta - c} \]

The first and second order derivatives of (15) with respect to \( \beta \) are

\[ \frac{\partial \alpha}{\partial \beta} = \pi \frac{\beta - 2c}{[\beta - c]^2} \]

and

\[ \frac{\partial^2 \alpha}{\partial \beta^2} = \pi \frac{2c^2}{[\beta - c]^3}. \]

Since (15) is a hyperbolic function, we will concentrate on the economic interesting interval \( \beta \in (c, \infty) \). Given \( \frac{\beta}{\beta - c} > 1 \), the zero-profit premium function \( \alpha^*(\beta) \) is always greater than a fair premium \( \alpha = \pi \beta \). From (16), we can easily deduce that its slope is always smaller than \( \pi \). Finally, (15) has its only minimum in the relevant interval at \( \beta = 2c \). The second order derivative (17) is positive for \( \beta \in (c, \infty) \), thus (15) is convex in the relevant interval.

Because of (15) and (1), the maximization problem (12) simplifies to

\[ \max_{\beta} U = \pi u \left( W_0 - \pi \beta \frac{\beta}{\beta - c} - L + \beta \right) + (1 - \pi) u \left( W_0 - \pi \beta \frac{\beta}{\beta - c} \right) \]

The concave shape of the utility function \( u \) with \( u' > 0 \) and \( u'' < 0 \) implies a strictly increasing and concave expected utility function \( U \) throughout the contract space \( \Psi \). Differentiating (18) with respect to \( \beta \) yields

\[ \frac{dU}{d\beta} = \pi u' \left( W_0 - \pi \beta \frac{\beta}{\beta - c} - L + \beta \right) \left[ 1 - \pi \frac{\beta - 2c}{[\beta - c]^2} \right] \]
\[ - (1 - \pi) u' \left( W_0 - \pi \beta \frac{\beta}{\beta - c} \right) \left[ \frac{\beta - 2c}{[\beta - c]^2} \right]. \]
Over-insurance will be optimal, if the slope of the indifference curve at \( \beta = L \) is positive and
\[
\frac{dU}{d\beta} \bigg|_{\beta=L} = \pi u' \left( W_0 - \pi L \frac{L}{L-c} \right) - u' \left( W_0 - \pi L \frac{L}{L-c} \right) \left[ \pi \frac{L}{L-c} \right] > 0 \tag{20}
\]
holds.

Rearranging (20) leads to
\[
1 > \frac{L [L - 2c]}{[L - c]^2}. \tag{21}
\]

Clearly, condition (21) is always valid, if \( c > 0 \). Thus, the optimal contract \( C^* \) entails over-insurance, because at \( \beta = L \) policyholders are marginally risk neutral and only regard the expected change of their payoffs. Since the marginal costs of an increase of coverage \( \frac{\partial \alpha}{\partial \beta} \) are smaller than the marginal benefits \( \pi \), a further increase of the indemnity is profitable at \( \beta = L \).

There exists a contract \( C^* \) with an indemnity \( \beta^* > L \) that solves the first order condition
\[
\frac{u' \left( W_0 - \pi \beta^* \frac{L}{L-c} - L + \beta^* \right)}{\pi u' \left( W_0 - \pi \beta^* \frac{L}{L-c} - L + \beta^* \right) + (1 - \pi) u' \left( W_0 - \pi \beta^* \frac{L}{L-c} \right)} = \beta^* \frac{[\beta^* - 2c]}{[\beta^* - c]^2}. \tag{22}
\]

There exists an unique optimal contract \( C^* \), because \( U \) is concave while the zero-profit premium is convex in the relevant interval.

Proof of Lemma 3. The critical detection probability \( \kappa^c \) makes policyholders without a loss indifferent between filing and not filing a fraudulent claim. Since no payoffs change compared to the case without fraud detection, the indifference condition of policyholders without a loss is given by
\[
u(W_0 - \alpha) = \kappa^c u(W_0 - \alpha - k) + (1 - \kappa^c) u(W_0 - \alpha + \beta) \tag{23}
\]
with
\[
\kappa^c = \frac{u(W_0 - \alpha + \beta) - u(W_0 - \alpha)}{u(W_0 - \alpha + \beta) - u(W_0 - \alpha - k)}. \tag{24}
\]

After the implementation of the system, insurers can condition their audits on the signal. The expected detection probability \( \bar{\kappa} \) of a fraudulent claim is a linear combination of the conditional audit probabilities \( \nu(s) \) weighted by the fractions of fraudulent claims with the signal \( s \in S \).
\[
\bar{\kappa} = \phi \nu(s_1) + (1 - \phi) \nu(s_0) \tag{25}
\]
In equilibrium the expected detection probability \( \bar{\kappa} \) must correspond to the critical detection probability of policyholders without a loss \( \kappa^c \). Using (24) and (25) yields
\[
\phi \nu^*(s_1) + (1 - \phi) \nu^*(s_0) = \kappa^c
\] (26)

From \( \phi > \delta \), (6) and (7) follow that the insurer’s fraud beliefs are higher for claims with the signal \( s_1 \) than for those with the signal \( s_0 \). Consequently, insurers begin to audit claims with the signal \( s_1 \). In equilibrium the conditional audit probability \( \nu^*(s_0) \) can only be zero, if insurers reach the critical detection probability by auditing claims with the signal \( s_1 \) exclusively. The fraction \( \phi \) must therefore satisfy
\[
\phi \nu(s_1) \geq \kappa^c
\] (27)

Using \( \nu(s_1) \in [0, 1] \) leads to the condition \( \phi \geq \kappa^c \), which completes the proof.

**Proof of Proposition 1.** The equilibrium fraud probability \( \eta^* \) must solve the indifference condition of insurers, which is given by
\[
-\beta = \mu(s_1)[-c] + (1 - \mu(s_1))[-\beta - c].
\] (28)

Substituting \( \mu(s_1) \) in (28) by (6) and rearranging the term yields
\[
\eta^* = \left( \frac{\pi}{1-\pi} \right) \left( \frac{\delta c}{\phi [\beta - c]} \right).
\] (29)

Given (29) the resulting ex ante fraud belief \( \mu \) of insurers is
\[
\mu = \frac{(1-\pi)\eta^*}{\pi + (1-\pi)\eta^*} = \frac{\delta c}{\phi \beta - [\phi - \delta] c}.
\] (30)

After the observation of the signal \( s_1 \) the posterior fraud belief \( \mu(s_1) \) is
\[
\mu(s_1) = \frac{\phi(1-\pi)\eta^*}{\delta \pi + \phi(1-\pi)\eta^*} = \frac{c}{\beta}.
\] (31)

The equilibrium audit probability \( \nu^*(s_1) \) satisfies the following indifference condition of policyholders
\[
u(W_0 - \alpha) = \phi \nu^*(s_1) u(W_0 - \alpha - \kappa)
+ (1 - \phi \nu^*(s_1)) U(W_0 - \alpha + \beta).
\] (32)

After some manipulations one obtains
\[
\nu^*(s_1) = \frac{1}{\phi} \frac{u(W_0 - \alpha + \beta) - u(W_0 - \alpha)}{u(W_0 - \alpha + \beta) - u(W_0 - \alpha - k)} = \frac{\kappa^c}{\phi}.
\] (33)
The overall audit probability is given by proportion of audited claims to all received claims, which is
\[ \nu^* = \frac{\delta \pi + \phi (1 - \pi) \eta^* v^*(s_1)}{\pi + (1 - \pi) \eta^*}. \] (34)

By using (29) and (33) the overall probability simplifies to
\[ \nu^* = \frac{\delta \beta}{\phi \beta - c} \frac{u(W_0 - \alpha + \beta) - u(W_0 - \alpha)}{u(W_0 - \alpha + \beta) - u(W_0 - \alpha - k)}. \] (35)

Since all elements of the PBNE have been found, the prove is done. ■

**Proof of Corollary 1.** The procedure of this proof corresponds to that of Lemma 2. The zero-profit premium with fraud detection is given by
\[ \alpha^*(\beta) = \pi \beta + (1 - \phi \nu^*(s_1)) (1 - \pi) \eta^* \beta + \nu^*(s_1) [\pi + (1 - \pi) \eta^*] c. \] (36)

Using (29) and rearranging (36) leads to
\[ \alpha^*(\beta, \delta, \phi) = \pi \beta - c + \frac{\delta c}{\beta - c}. \] (37)

If \( \phi > \delta \), the zero-profit premium with fraud detection (37) is strictly lower than that without fraud detection (15) as (38) shows.
\[ \pi \beta \frac{\beta}{\beta - c} > \pi \beta - c + \frac{\delta c}{\beta - c}. \] (38)

Given that \( \alpha^*(\beta, \delta, \phi) \) is twice differentiable with respect to \( \beta \), the partial derivatives of (37) with respect to \( \beta \) are
\[ \frac{\partial \alpha}{\partial \beta} = \pi \frac{[\beta - c]^2 - \frac{\delta c^2}{\phi}}{[\beta - c]^2} \] (39)
and
\[ \frac{\partial^2 \alpha}{\partial \beta^2} = \pi \frac{2 \frac{\delta c^2}{\phi}}{[\beta - c]^3}. \] (40)

For \( \phi > \delta > 0 \), (37) has in \([c, \infty)\) only a minimum at \( \beta_0 = \left(1 + \sqrt{\frac{\phi}{\delta}}\right) c \). It is convex throughout the whole relevant interval, because (40) is positive for \( \beta \in [c, \infty) \) and \( \delta > 0 \). In the case \( \delta = 0 \), (37) corresponds to the fair premium, which is linear increasing in \( \beta \).
The simplified maximization problem with a fraud detection system is

\[
\max_{\beta} U = \pi u \left( W_0 - \pi \beta \frac{\beta - c + \phi c}{\beta - c} - L + \beta \right) + (1 - \pi) u \left( W_0 - \pi \beta \frac{\beta - c + \phi c}{\beta - c} \right).
\]

(41)

Over-insurance will be optimal, if and only if the slope of the indifference curve at \( \beta = L \) is positive and

\[\pi u' \left( W_0 - \pi L \frac{L - c + \phi c}{L - c} \right) - u' \left( W_0 - \pi L \frac{L - c + \phi c}{L - c} \right) \left[ \frac{\pi [L - c]^2 - \frac{\phi c^2}{L - c}}{[L - c]^2} \right] > 0 \]

holds.

Rearranging (42) leads to

\[\delta > 0 \]

(43)

If and only if \( \delta = 0 \) holds, the detection system is perfect and the zero-profit premium with fraud detection is fair, which leads to an optimal full-insurance contract with \( \hat{\beta} = L \).

Finally, we show that over-compensation is increasing (decreasing) in \( \delta \) (\( \phi \)). Differentiating (39) with respect to \( \delta \) and \( \phi \) we obtain

\[
\frac{\partial^2 \alpha}{\partial \delta \partial \beta} = -\frac{\pi c^2}{\phi [\beta - c]^2} < 0
\]

and

\[
\frac{\partial^2 \alpha}{\partial \beta \partial \phi} = \frac{\delta c^2}{\phi^2 [\beta - c]^2} > 0.
\]

(44)

(45)

The slope of the zero-profit premium decreases ceteris paribus with an increase of \( \delta \). Therefore, the zero-profit function gets flatter which implies an increase of the optimal indemnity. The other way around, the zero-profit premium function becomes steeper with an increase of \( \phi \) which leads to a reduction of the indemnity and over-compensation. \( \blacksquare \)

Proof of Lemma 4. We distinguish the observable and unobservable contract choice of insurers at stage 2. Firstly, we consider the case, where the contract choice is observable. By using a backwards induction argumentation, we can exclude any fixed fee \( F \). If insurers accept the contract with a fixed fee, they will make an expected loss of \( \Pi^d = -F \), because at the competition stage 3 they can only charge premiums equally to their marginal costs. The strategy to accept that contract is dominated by not accepting it and to withdraw from the market, which leads to a expected profit of zero.

If the contract choice is unobservable, the relevant strategies of the insurers will consist of a participation contract choice and an insurance contract offer. At stage 1, the external supplier will only offer contracts that will earn non-negative expected profits at stages 2 and 3. Therefore, the supplier can
only offer the contract \( P_v = (0, f) \), which entails only a positive variable fee \( f \geq \frac{1}{N \pi \phi \beta - c} \phi \beta - c \) and the contract \( P_f = (F, 0) \), where the supplier only charges a positive fixed fee \( F \geq \frac{1}{f} \). At the competition stage 3, the two types of insurers can only offer the usual Bertrand equilibrium contracts. This leads again to the result that the participation contract \( P_f \) with a fixed fee causes expected losses and no insurer will choose any contract with a fixed fee. ■
References


Acknowledgments

I thank Klaus Bender, Richard Derrig, Uwe Focht, Martin Nell, Henrike Lesch, Martin Peitz, Andreas Richter, Andrea Seul, two anonymous referees and the participants of the 2003 annual ARIA meeting in Denver for their very helpful comments and suggestions.

Notes

1. The presented model is in that way not a signalling game, because the signal is an exogenous, statistical information generated from the fraud report, but the policyholder is unable to strategically influence the signal.

2. The considered insurance contracts are incomplete, because the indemnity is state-independent. Actually, insurers could condition the indemnity payment on the fact whether an audit took place or not and reward honest policyholders with a higher indemnity (see e.g. Fagart/Picard [1999]). But a contract with a state-dependent indemnity is incredible, because policyholders do not know whether their claim was audited or not, and they have no opportunity to detect this fact. In cases, where an insurer audits a claim and determines that it was honest, he will keep the audit secret and save money. The credibility issue that applied to the auditing problem is also valid in this case.

3. Boyer [1999] analyzes a situation with two different types of policyholders. One type never commits insurance fraud and always reports the state of the world honestly. The second type of policyholders is opportunistic and weighs up the costs and benefits of a fraudulent claim, if no accident occurred. In this paper we will concentrate on the second type of opportunistic policyholders.

4. There are two reasons why we assume the penalty $k$ not to be part of the insurance contract. First of all, it is usually determined by law and/or courts. Therefore, insurers would have to renegotiate with fraudulent policyholders after an audit. Since negotiations cause significant transaction costs and the expected sanction for defrauders in reality is not very high (see e.g. Derrig/Zicko [2002]), it is questionable whether insurers can collect the penalty or not. In addition, when a penalty $k > c$ is paid to the insurer, auditing becomes a rent-extracting device and not a way to induce contract compliance of policyholders (see e.g. Picard [1996] for a model, where the insurer partially benefits from the penalty).

5. This result is just technical, because insurers are not able to audit a claim that was not made.

6. The implementation costs of an existing fraud detection system in New Jersey are 10 to 12 million USD. The per annum operating costs are more than twice of the implementation costs (see PANKO, Ron [2001]: Making A Dent In Auto Insurance Fraud, Best’s Review, October 2001).
### Tables

<table>
<thead>
<tr>
<th>$s_1$</th>
<th>$p(s \mid \omega_1)$</th>
<th>$p(s \mid \omega_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>$\phi$</td>
<td></td>
</tr>
<tr>
<td>$s_0$</td>
<td>$(1 - \delta)$</td>
<td>$(1 - \phi)$</td>
</tr>
</tbody>
</table>

Table 1: Contingent Probabilities
Figures