Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 176 (2007) 565–583
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

O.R. Applications

Strategies for detecting fraudulent claims in the automobile insurance industry

Stijn Viaene [a], Mercedes Ayuso [b], Montserrat Guillen [b,*], Dirk Van Gheel [c], Guido Dedene [c,d]

[a] *Vlerick Leuven Gent Management School lVlamingenstraat 83, 3000 Leuven, Belgium*
[b] *Departament d'Econometria, Estadística i Economia Espanyola, Universitat de Barcelona, Diagonal, 690, 08034 Barcelona, Spain*
[c] *Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium*
[d] *Department of Information Management, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*

Received 22 July 2004; accepted 18 August 2005
Available online 22 November 2005

Abstract

Some property and casualty insurers use automated detection systems to help to decide whether or not to investigate claims suspected of fraud. Claim screening systems benefit from the coded experience of previously investigated claims. The embedded detection models typically consist of scoring devices relating fraud indicators to some measure of suspicion of fraud. In practice these scoring models often focus on minimizing the error rate rather than on the cost of (mis)classification. We show that focusing on cost is a profitable approach. We analyse the effects of taking into account information on damages and audit costs early on in the screening process. We discuss several scenarios using real-life data. The findings suggest that with claim amount information available at screening time detection rules can be accommodated to increase expected profits. Our results show the value of cost-sensitive claim fraud screening and provide guidance on how to render this strategy operational.

Keywords: Decision support systems; Insurance fraud; Vehicle damage claims; Classification; Cost asymmetries

1. Introduction

Fraud has become a high priority for insurers. For the European insurance industry the Comité Européen des Assurances (1996) estimates that the cost of fraud is over 2% of the total annual premium income for all lines of business combined. In

* Corresponding author. Tel.: +34 93 4037039; fax: +34 93 4021821.

E-mail addresses: Stijn.Viaene@econ.kuleuven.ac.be (S. Viaene), mayuso@ub.edu (M. Ayuso), mguillen@ub.edu (M. Guillen), Dirk.VanGheel@econ.kuleuven.ac.be (D. Van Gheel), Guido.Dedene@econ.kuleuven.be (G. Dedene).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2005.08.005

most European countries claim fraud is estimated to represent between 5 and 10% of the total yearly amount of indemnities paid for non-life insurance. In the United States, the Coalition Against Insurance Fraud (2001) states that more than 6% of each insurance premium goes to fraud. The Insurance Information Institute (2004) estimated property and casualty (P&C) claim fraud at $31 billion in 2002.

If not properly addressed, insurance fraud not only puts the profitability of the insurer at risk, but also negatively affects its value chain, the insurance industry, and may be extremely detrimental to established social and economic structures. Moreover, all honest policyholders are victims. Fraud is widely believed to increase the cost of insurance. This cost component is borne directly by all insured parties in the form of increased premium rates. In the end, fraud represents a threat to the very principle of solidarity that keeps the concept of insurance alive (Guillén, 2004; Viaene and Dedene, 2004).

Economic theory has studied P&C claim fraud in depth. Picard (2000) provides a good overview of this literature. It can be shown, for example, that an insurance firm needs to commit to a claim audit strategy to ensure solvency. It has also been shown that as long as there is a good chance of not being caught, an individual filing a claim has a clear economic incentive to defraud.

The most effective way to fight fraud for an insurer is, of course, to prevent abuse of the system. Yet, fraudsters always seem to find new ways of exploiting the inertia of complex systems, especially when a lot of money is involved. It is then imperative to ensure that fraudulent activity is identified at the earliest possible moment, and that persons cheating the system are swiftly tracked down. Since fraud is not a self-revealing phenomenon, insurers typically have to commit considerable resources to its detection. The process of in-depth investigation of suspicious claims is known as *costly state verification*, because the true nature of a claim (fraudulent or not) can only be discovered by means of an in-depth investigation (see e.g. Bond and Crocker, 1997; Crocker and Tennyson, 1999; Boyer, 1999; Picard, 1996, 2000). It has been shown that such claim auditing has both a detection and deterrence effect (Tennyson and Salsas-Forn, 2002).

The problem in detecting fraudulent claims is the identification of the characteristics that distinguish them from valid claims. Most insurers train their front-line claims adjusters, often with the help of state- or country-level fraud bureaus, to recognise claims that have combinations of features that experience has shown to be typically associated with fraudulent claims. In many practical situations, though, the identification of suspicion during claims handling continues to be rather subjective in nature. Many insurers still leave it essentially up to the individual adjuster to somehow put together potential warning signs into an aggregated assessment of suspicion of claim fraud. Moreover, since customer service has to a large degree become synonymous with processing efficiency in the context of claims handling, adjusters have no natural incentive to actively look for warning signs. In today's configuration there is little time for running through extensive lists of fraud indicators at claim time. In other words, one of the core challenges for contemporary fraud detection is to identify fraud in an automated, high-volume, online transaction processing environment without jeopardising the advantages of automation in terms of efficiency, timeliness and customer service.

Some P&C insurers use automated detection systems to help decide on whether to investigate claims suspected of fraud. Automated types of fraud detection should make it possible to reduce the lead-time for fraud control and allow for more optimal allocation of scarce investigative resources. The embedded detection models typically consist of scoring devices relating fraud indicators to some measure of suspicion of fraud. This is where insurers may build on considerable past investments in more systematic electronic collection, organization and access to coherent insurance data. This, among other things, enables the use of algorithmic pattern recognition techniques to create models that help with the identification of insurance fraud.

In this paper we deal with algorithmic learning to score claims for suspicion of fraud. Specifically, we focus our attention on screening activity early on during the life cycle of a claim. What we see in practice is that during the construction phase these scoring models often focus on minimising error rate rather than on cost of classification (see, Derrig, 2002). In this paper we show that focusing on cost rather than error of classification is a profitable approach. This question was already pointed to by Dionne et al. (2003), who work with an average cost approach. In relation to the costs of fraud detection we take into account information on damages and audit costs available in the early part of the claim screening process. In an empirical experiment we discuss several scenarios, the effects of which are explored using real-life data. The data set that is used for this exploration consists of automobile claims closed in Spain that were investigated for fraud by domain experts and for which we have detailed cost information. Thus, the focus of this paper will be on vehicle damage claims in the context of automobile insurance. The findings suggest that with claim amount information available early on in the screening process detection rules can be accommodated to increase expected profits. Our results show the real value of cost-sensitive claim fraud screening and provide guidance on how to operationalise this strategy.

The rest of this paper is organized as follows. Section 2 highlights the main steps in the implementation of a claim fraud detection strategy for a typical P&C insurer. Section 3 tackles the mechanics of a cost-sensitive classification, i.e. the methodology used to build learning programs for fraud detection that reduce the cost of classification rather than the error rate. Section 4 sets the stage for our complementary case study. It covers the data characteristics for the set of Spanish damage claims used in the empirical part and takes a look at the economics of fraud detection for this case. Section 5 projects the mechanism of cost-sensitive classification onto the claim classification setting at hand. This section contrasts six alternative cost incorporation scenarios based on different assumptions concerning the available cost information for classification early on in the process of claim screening. In Section 6 we synthesise our conclusions from the discussion in the previous sections.

2. Fraud control for the P&C insurer

The generic operational claim fraud control model for insurers (see Fig. 1) includes screening, investigation and negotiation/litigation phases. It is embedded in the insurer's claims handling process. Claims handling refers to the process that starts with a claim occurrence and a report from the policyholder and ends with the payment, or denial of payment for damages covered. Fraud, in principle, is the only reason for denying payment for covered damages. Fraud, primarily a legal term, generally requires the presence of material misrepresentation, intent to deceive and the aim of gaining an illicit benefit (Viaene and Dedene, 2004). The absence of one or more of these key elements makes an undesirable activity at most qualify as so-called abuse of insurance, where the latter is typically defined as any practice that uses insurance in a way that is contrary to its intended purpose or the law. The concept of insurance fraud is often broadly used in practice to encompass more general forms of insurance abuse. Sometimes the term abuse is used instead of fraud to avoid the element of criminality associated with it.

2.1. Claim screening

Early claim screening systems help to decide whether incoming claims are suspicious or not. This is the basis for routing claims through different claim handling workflows (Derrig, 2002). Claims that pass the initial (automated) screening phase are settled swiftly and routinely, involving a minimum of transaction processing costs. Claims that are flagged as suspicious pass a costly state verification process, involving (human) resource intensive investigation. This task is typically delegated to so-called Special Investigation Units or SIUs (Ghezzi, 1983).

Practical models to sort out claims for fraud investigation emerged in the 1990s with database organization and selection strategies (Major and Riedinger, 1992), fuzzy clustering (Derrig and Ostaszewsky, 1995), simple regression scoring models (Weisberg and Derrig, 1998; Brockett et al., 1998) and probit and logit models (Artís et al., 1999; Belhadji et al., 2000). The latest

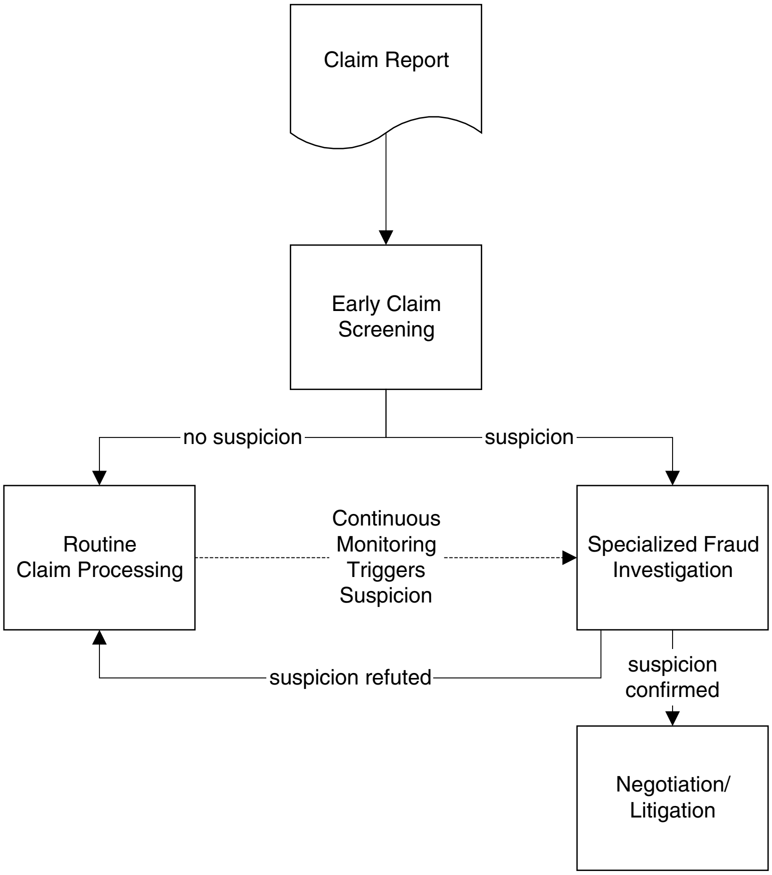

Fig. 1. Operational fraud control model.

synthesis of important material is found in the September 2002 special issue of the Journal of Risk and Insurance devoted to the theoretical and practical aspects of claim fraud detection and deterrence. At the same time, theoretical studies focused on promoting efficient anti-fraud action through contract design and auditing strategies (see e.g. Picard, 1996, 2000; Bond and Crocker, 1997; Crocker and Tennyson, 1999; Boyer, 1999; Watt, 2003).

Without an automated system the decision to investigate a claim is taken primarily on the basis of the available information on a single customer and a single claim. The adjuster or person in charge of handling the claim typically has no time to perform extensive searches, particularly not in paper files. On the basis of narrowly scoped information he has to decide whether or not the claim is suspicious and whether it is worthwhile to investigate it for fraud. This assessor is usually an experienced person from the staff. His personal experience (including biases) can help to sort out claims. Note, however, that different adjusters need not agree in this context (see e.g. Weisberg and Derrig, 1998).

If recent and historical insurance information is carefully logged, an insurer's automated detection system can take the decision to investigate claims on the basis of the entire claim and customer history. Most insurance companies use lists of fraud indicators or flags (most often per insurance business line), representing a summary of the detection expertise. These lists form the basis for systematic and consistent identification of fraudulent claims (Derrig, 2002). More systematic data collection stimulated data-driven initiatives have aimed at analysing and modelling the formal relations between fraud indicator combinations and transaction suspiciousness, resulting in the implementation of automated indicator-based fraud screening models. The claims screen then typically takes the form of a scoring device, which relates case-based fraud indicators to levels of suspicion. As indicative information on the level of fraud suspicion only gradually becomes available during the life of a claim, the diagnostic system ought to follow claims throughout their lives (Viaene and Dedene, 2004).

There are three general sources for data that provide relevant information about policyholders and claims:

(1) *Policy issuing.* At underwriting time the insurer gathers information on customer and the insured vehicle, typically via standard form filling. During the complete lifespan of the contract the policyholder is legally required to disclose all information that could materially affect the risk to be covered or its pricing. Examples of items that are usually recorded are: date of birth, address, social security number, type of vehicle, date of first driving license, occasional drivers, and uses of the vehicle. This data is chiefly used to calculate the premium. It goes without saying that this information is also useful for developing customer profiles that can, for example, afterwards be linked to claims details. We note that those with a primary responsibility for marketing are typically in favour of keeping the amount of information to be gathered to a strict minimum in order for it to have as little impact as possible on transaction processing efficiency.
(2) *Claims handling.* Data gathering during the claims handling phase is related to the actual circumstances of the accident. Information is primarily collected for later assessment of the exposure of the insurance company to payment of the claim by the claims handling department's adjusters. Examples of items that are usually recorded are: time, place of the accident, report of what occurred, witnesses, and other vehicles involved (names, plates, companies, etc.). Most insurers follow the 'single point of contact' front office strategy for claims handling, in which a single customer contact point (e.g. broker, agent, call center) takes care of all the required company–customer communication. Well-structured communication is generally to be recommended for consistency in gathering data (e.g. the use of well-structured claim report statements). Again, there will usually be a trade-off to be made between processing efficiency and 'lengthy form filling'.
(3) *Damage evaluation.* Normally there are industry databases for cars and models that provide adjusters with a tool for straightforward estimation of the cost of parts and repair. We can therefore assume that the adjuster is able to size up an initial claimed amount very soon after the accident occurs.

It should be clear that the collection of data to provide the required coherent, enriched view on claims and their context (reflecting the present and the past) is the shared responsibility of several parties within the (extended) insurance enterprise, e.g. sales channels, marketing offices, claims handling departments, premium rating departments, etc. Ideally, concern about potential fraud should somehow come naturally to these parties. Thus, one of the crucial tasks for any insurer that is serious about tackling fraud is to make all these parties into stakeholders with respect to fraud control. The best way to do so is to design incentive mechanisms to be incorporated into your fraud control program.

If at any time during routine processing of a claim a certain level of suspicion of fraud is identified, the claim is scheduled for specialised investigation. This auditing step gives rise to the so-called audit cost. Currently there are few insurers that systematically track the cost of auditing their claims.

2.2. Claim investigation

It is a specialised fraud investigator's job to try to uncover the true nature of a suspicious claim. The investigative work is mainly guided by the experience, skill, creativity and situational empathy of the human investigator, which generally makes work proceed in a non-routine, ad hoc manner and takes substantial time, effort and money (Viaene and Dedene, 2004).

The investigator's workbench is ideally geared toward this exploratory exercise of analysis and synthesis. It should provide an agile virtual window onto a wide range of internal and external investigative resources (e.g. up-to-date lists of important contacts, e-mail or bulletin board services, database search and navigational capabilities, specialised analytical software).

Tennyson and Salsas-Forn (2002) list the most commonly used audit methods (e.g. site investigation, recorded statement, SIU referral, activity check). Note that some insurers choose to contract the investigation out to external specialised parties (Derrig, 2002).

2.3. Negotiation/litigation

With a strong enough case for fraud (using whatever working definition is chosen) the insurer may then decide to dismiss or reduce compensation or even decide to press charges. However, few fraud cases ever reach the courts. Litigation and prior special investigation typically involve lengthy and costly procedure. Insurers are also fearful of getting involved in lawsuits and losing, which may compromise their reputation. Insurers generally prefer to settle cases of soft fraud internally, i.e. through negotiation, except perhaps in the most flagrant cases. And even though it may not be the preferred action for cases of hard fraud, negotiation may be necessary in the absence of evidence establishing guilt beyond a reasonable doubt. As pointed out by Clarke (1989), the insurer's strategy is then geared toward confronting the claimant with the gathered evidence and gently developing pressure to make him reduce or drop the claim. This also ought to deter the claimant from defrauding on insurance again.

The final decision on what action to take will typically not be made without explicit consultation with senior or qualified personnel (e.g. for balancing prudential against commercial arguments).

Active fraud control is systematically on the lookout for new fraud opportunities, new schemes and emerging trends, and is equally agile in deploying revised fraud controls. This proactivity is sustained by a number of supporting processes (e.g. filing, reporting and knowledge discovery in databases) that continuously monitor the operational model and are aimed at its continuous improvement. The use of new technologies (e.g. data warehousing, data mining and high-speed networking) helps enable this proactivity. Moreover, automated types of fraud detection should make it possible to reduce the investigative process lead-time and allow for more optimal allocation of scarce investigative resources (Viaene and Dedene, 2004).

3. Cost-sensitive decision making

Statistical modelling and data-driven analysis allow for the modernization of the fraud detection process with sophisticated, (semi-)automated, intelligent tools such as unsupervised and supervised pattern learning. Classification is one of the foremost supervised learning tasks. Classification techniques are aimed at algorithmically learning to allocate data objects, described as predictor vectors, to pre-defined object classes based on a training set of data objects with known class labels. This is no different for fraud detection (see e.g. Brockett et al., 2002; Belhadji et al., 2000; Artís et al., 1999; Viaene et al., 2002).

One of the complications that arise when applying these learning programs in practice, however, is to make them reduce the cost of classification rather than the error rate. This is not unimportant. In many real-life decision making situations the assumption of equal (mis)classification costs, the default operating mode for many learners, is most likely not observed. Medical diagnosis is a prototypical example. Here, a false negative prediction, i.e. failing to detect a disease, may well have fatal consequences, whereas a false positive prediction, i.e. diagnosing a disease for a patient that does not actually have it, may be less serious.

A similar situation arises for insurance claim fraud detection, where an early claim screening facility is to help decide upon the routing of incoming claims through alternative claims handling workflows (Derrig, 2002). Claims that pass the initial screening phase are settled swiftly and routinely, involving a minimum of transaction processing costs. Claims that are flagged as suspicious have to pass a costly state verification process, involving resource intensive investigation. Claim screening should thus be designed to take into account these cost asymmetries in order to make cost-benefit-wise optimal routing decisions.

Many practical situations are not unlike the ones above. They are typically characterized by a setting in which one of the pre-defined classes is a priori relatively rare, but also associated with a relatively high cost if not detected. Automated classification that is insensitive to this context is unlikely to be successful. For that reason Provost and Kohavi (1998) and Provost and Fawcett (2001) argue against using the error rate (a.k.a. zero–one loss) as a performance assessment criterion, as it assumes equal misclassification costs and relatively balanced class distributions. Given a naturally very skewed class distribution and costly faulty predictions for the rare class, a model optimized on error rate alone may very well end up being a useless model, i.e. one that always predicts the most frequent class. Under these circumstances, cost-sensitive decision making is more appropriate.

3.1. Cost-sensitive claim classification

Many classifiers are capable of producing Bayesian posterior probability estimates that can be used to put data objects into the appropriate pre-defined classes. Since the empirical part of this paper involves only binary classification we restrict our discussion to this case.

Optimal Bayes decision making dictates that an input vector $x \in \mathfrak{R}^k$ should be assigned to the class $t \in \{0,1\}$ associated with the minimum expected cost (Duda et al., 2000). Optimal Bayes assigns classes according to the following criterion:

$$\arg\min_{t\in\{0,1\}} \sum_{j=0}^{1} p(j|x)C_{t,j}(x), \tag{1}$$

where $p(j|x)$ is the conditional probability of class j given predictor vector x, and $C_{t,j}(x)$ is the cost of classifying a data object with predictor vector x and actual class j as class t.

The available cost information is typically represented as a cost matrix C, where each row represents a single predicted class and each column an actual class. This is illustrated in Table 1 for the case of our two classes, coded here as 1 for fraud and 0 for honest. The cost of a true positive is denoted by $C_{1,1}(x)$, the cost of a true negative is denoted as $C_{0,0}(x)$, the cost of a false positive is denoted as $C_{1,0}(x)$, and the cost of a false negative is denoted as $C_{0,1}(x)$. Note that if $C_{t,j}(x)$ is positive it represents an actual cost, whereas if $C_{t,j}(x)$ is negative it represents a benefit.

We assume that the cost matrix complies with the two reasonableness conditions formulated by Elkan (2001). The first reasonableness condition implies that neither row dominates any other row, i.e. there are no two rows for which the elements of the one of them are all smaller than or equal to the elements of the other one. The second reasonableness condition implies that the cost of labelling a data instance incorrectly is always greater than the cost of labelling it correctly.

It can be verified that, under the above reasonableness conditions, the criterion for classification in (1) translates into the rule that assigns class 1 to a data object if

$$p(j=1|x) > \frac{C_{1,0}(x) - C_{0,0}(x)}{C_{1,0}(x) - C_{0,0}(x) + C_{0,1}(x) - C_{1,1}(x)} \tag{2}$$

and class 0 otherwise. In case of equality we choose to classify the data object as class 0.

If the available cost information $C_{t,j}$ is known, i.e. there is a fixed cost associated with assigning

Table 1
Cost matrix

	Observed honest	Observed fraud
Predicted honest	$C_{0,0}(x)$	$C_{0,1}(x)$
Predicted fraud	$C_{1,0}(x)$	$C_{1,1}(x)$

a data object to class t when in fact it belongs to class j, the rule in (2) defines a fixed classification threshold in the interval $[0, 1]$. Dionne et al. (2003), for example, settled for a simplified costing scenario using fixed average costs for constructing their claim screen. In practice, though, claim costs (including claim amount and audit costs) are hardly uniform. Besides the stochastic nature of the losses caused by an accident, it is generally the case that certain characteristics of the policyholder and the automobile are associated with higher/lower compensation (e.g. more expensive cars usually have more expensive repair bills). In cases of policyholders claiming for salary loss, some may typically receive higher compensation due to their age or their occupation (Derrig, 2001; Derrig and Kessler, 1994; Dionne et al., 1993). Audit costs may also vary among claims. The nature of the work associated with the investigation process (see Section 2.2) usually leads to non-uniform audit costs.

For our experiments we use a logistic regression model to estimate $p(j = 1|x)$. Logistic regression makes the assumption that the difference between the natural logarithms of the class-conditional data density functions is linear in the predictors:

$$\ln\left(\frac{p(x|j = 1)}{p(x|j = 0)}\right) = b + w^T x, \tag{3}$$

where $w \in \Re^n$ represents the coefficient vector and $b \in \Re$ the intercept. Note that besides the assumption in (3), logistic regression does not make any distributional assumptions for the predictors and has been shown to work well in practice for data that depart significantly from conventional normality assumptions (see e.g. Michie et al., 1994). The class membership probability $p(j = 1|x)$ underlying classification in (2) can readily be obtained from the model in (3):

$$p(j = 1|x) = \frac{\exp(b' + w^T x)}{1 + \exp(b' + w^T x)}, \tag{4}$$

where $b' = b + \ln(\frac{p(j=1)}{p(j=0)})$, with $p(j = 1)$ and $p(j = 0)$ the class priors. We can use the class membership probability specification in (4) to obtain maximum likelihood estimates for w and b'. We assume that the class proportions in the training data are representative of the true class priors. We, thus, assume that we are working on a random training sample. If this is not the case, we should account for training set class proportions that are not representative of the population under consideration and correct the estimated class membership probability underlying classification in (2) appropriately. This bias in the training sample can easily be dealt with by using a weighted logistic regression.

3.2. Cost structure hypothesis

For our experiments on cost-sensitive claims screen design we hypothesize the following. Whenever a new claim arrives, our claim screening facility makes a prediction, specifically, as to whether the claim is honest or fraudulent, without any cost. If the claim is predicted to be honest, then it is settled as stipulated by the obligations of the insurance contract. On the other hand, if it is predicted to be fraudulent, then the claim is audited, i.e. the prediction triggers a decision to acquire more information at a price (i.e. the audit cost). The claim investigation is generally costly, but has no (direct) effect on the claim amount. We make abstraction of any other strategic reasons to audit claims, either by targeting or at random. We furthermore assume that by auditing the claim we can exactly determine the true nature of the claim, i.e. whether it is honest or fraudulent. Since the evidence of fraud is a reason for not paying for the claim, we obtain the cost structure as specified in Table 2.

For the true negative, i.e. when an honest claim is labelled as honest, the company pays the amount that is claimed by the insured party. Since it is normal practice for an insurance company to compensate honest claims, it is not considered a cost to the company. So the cost of a true negative is 0. When a fraudulent claim is labelled as honest the compensation should not be paid, but no

Table 2
Hypothesized cost structure

	Observed honest	Observed fraud
Predicted honest	0	Claim amount(x)
Predicted fraud	Audit cost(x)	Audit cost(x) − claim amount(x)

investigation takes places and the claimed amount is paid to the insured party. So, the cost of a false negative equals the claim amount. When an honest claim is labelled as fraudulent, the subsequent investigation will eventually find out that the claim is actually honest. Thus, the claim amount is paid to the insured party and as such the cost of a false positive amounts to the cost of the investigation. For a true positive, i.e. when a fraudulent claim is labelled as fraudulent, the subsequent investigation will eventually find out that the claim is actually fraudulent. Therefore, the damages claimed will not be paid. So, the cost of a true positive is the cost of the investigation minus the claim amount.

The criterion for choosing between alternative claim screen operationalisations is straightforward. Each operationalisation typically leads to a different average claim cost for the company. The one to be chosen is thus the one yielding the lowest average claim cost, benchmarked against a representative data sample.

4. Spanish automobile damage claims

4.1. Data set

We have a random sample of claims from a large Spanish insurer for car damages from accidents that occurred in Spain during the year 2000. All the claims included here were audited and the insurer classified them in two categories, i.e. honest or fraudulent, after the investigation process. The data set contains 2403 claims, of which: 2229 are legitimate and 174 fraudulent. This means that about 7.24% of the claims in our data set are fraudulent. There does not seem to be unanimous agreement about the actual percentage of fraud within the market population of claims, but the representation of the classes in the data set is well within the 5–10% range most often cited and reported in industry surveys. The insurer labels a claim as fraudulent if the insured party admits that he willfully misrepresented the damages. In Spain, legal prosecution of insurance fraud is extremely rare. When the insurer seriously suspects fraud, negotiation with the policyholder usually follows. During the confrontation the insurer announces that either payment of the claim will be denied or the contract will be cancelled. In order to avoid termination of the contract fraudsters then usually admit to having defrauded the insurance company.

4.2. Variable definition

The dependent variable (*Fraud*) has been coded using ones for fraudulent claims and zeros for honest claims. The independent variables relate to the three sources for gathering data discussed in Section 2, i.e. the policy issuing, claims handling and evaluation of damages. The variables in the data set are very similar to the ones used in Artís et al. (2002), but now we are working with a different database and have access to more complete information, including costing information. The definition for all independent variables included in the models is detailed below. Note that we do not consider fraud related to injuries or medical treatment. The identification of this kind of fraud may require a different source. A discussion of the kind of information that can be taken into account when looking for fraud in these cases is found in Derrig and Weisberg (1998).

4.2.1. Policy issuing variables

Coverage: The insured party can choose from among three basic types of coverage (excluding personal injury coverage) when he decides to underwrite an automobile insurance policy: (1) third party liability only, (2) third party liability plus arson and/or theft and/or glass breakage, and (3) all damages. Coverage is encoded using two dummy variables (*Cov1* and *Cov2*) as shown in Table 3.

Table 3
Dummy encoding of coverage

Coverage	*Cov1*	*Cov2*
Third party liability only	1	0
Third party liability plus arson/theft/glass breakage	0	1
All damages	0	0

Vehicle use and type: Information about the type and use of the insured vehicle is collected by almost every insurer. It is common to differentiate between categories of vehicles, such as automobiles, motorcycles (potentially subdivided per power range), vans, trucks, busses, etc. A commonly used categorization of the product portfolio combines use and type of vehicle as follows: (1) automobile for private use, (2) motorcycles and (3) other (including vehicles for industrial use, busses, etc.). This categorization is encoded using two dummy variables (*Veh1* and *Veh2*) as shown in Table 4.

Insured person: An important set of variables is related to the insured driver: the *Age* of the insured driver at the time of the accident, the *Gender* of the insured driver, the number of previous claims with the company (*Records*), and the number of years that the insured party has been with the company (*Policyage*). Continuous variables are not discretised before taking them up in the models.

4.2.2. Claims handling variables

Fault: a dichotomous variable (*Fault*) indicating whether the other driver considers himself to be at fault for the accident.

Legal system: in Spain car damages are often covered under the no-fault system. Under the no-fault system the insurer pays for claimed damages covered by the policy regardless of fault (up to the specified policy limit). However, there are specific situations where this system is not applicable and insurance companies resort to the traditional at-fault system, where an insurer makes payment according to each person's degree of fault in a particular accident. This, for example, would be the case if there were more than two vehicles involved in the accident, or if there were no direct collision between vehicles. The dichotomous variable (*NFS*) indicates whether the no-fault system applies.

Table 4
Dummy encoding of vehicle use and type

Vehicle use and type	*Veh1*	*Veh2*
Automobile for private use	1	0
Motorcycle	0	1
Other	0	0

Weekend: a dichotomous variable (*Weekend*) indicating whether the accident occurred on the weekend.

Delay of claim report: Article 7 of the Spanish Insurance Law and Regulation indicates that the insured party must report an accident within the first week after its occurrence. Still, insurers in practice often compensate for covered damages even when claims are reported more than a week after the accident occurs. A dichotomous variable (*Delay*) indicates whether the claim is reported to the company within the legally established period.

4.2.3. Claim costs

The cost structure hypothesis was discussed in Section 3.2 and synthesised in Table 2. In line with the previous discussion we introduce the variables *Claim amount* and *Audit cost*. The variable *Claim amount* represents the insurer's valuation of the damages to the vehicle, once the deductible has been discounted (if it exists). It reflects the compensation of the claimant, assuming the claim is legitimate. The variable *Audit cost* reflects the cost of investigation necessary to ascertain the true nature of the claim.

Table 5 provides a summary overview of the variables included in the Spanish claims data set.

4.3. Sample summary statistics

Table 6 provides us with variable related summary statistics for the overall data sample of Spanish claims and for the two subsamples of fraudulent and honest claims.

Some of the variables that were used in this study are inherently correlated (for example, policy age and age of the policy holder). This is acceptable in a logistic regression model designed for prediction, since multicollinearity is known to have an effect on the variance of parameter estimates. We should, then, be cautious in our interpretation of the significance of parameters that measure the influence of variables that are correlated (see Kleinbaum and Klein, 2002). Testing for correlated omitted variables is not performed here because the data used for prediction are the same. If one variable were to be omitted from

Table 5
Spanish claims data variables

Name	Type	Explanation
Fraud	Nominal	Observed type of claim (fraudulent equals 1, legitimate 0)
Cov1	Nominal	Third party liability equals 1, otherwise 0
Cov2	Nominal	Third party liability plus arson/theft/glass breakage equals 1, otherwise 0
Veh1	Nominal	Automobile for private use equals 1, otherwise 0
Veh2	Nominal	Motorcycle equals 1, otherwise 0
Age	Continuous	Age of insured driver when the accident occurred
Gender	Nominal	Insured driver is male equals 1, female 0
Records	Continuous	Number of previous claims of the insured
Policyage	Continuous	Number of years the insured has been with the company
Fault	Nominal	The other driver accepts fault for the accident equals 1, otherwise 0
NFS	Nominal	Use of the no-fault system equals 1, otherwise 0
Weekend	Nominal	Accident occurring on a weekend equals 1, otherwise 0
Delay	Nominal	Claim not reported to the company within the legally established period equals 1, otherwise 0
Claim amount	Continuous	Insurer's valuation of the vehicle damages (once the deductible has been discounted, if it exists)
Audit cost	Continuous	Cost of fraud investigation

Table 6
Variable related summary statistics

Variable	Total sample ($N = 2403$)		Observed fraudulent ($N = 174$)		Observed honest ($N = 2229$)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Cov1	0.23	0.42	0.33	0.47	0.22	0.42
Cov2	0.60	0.49	0.54	0.50	0.60	0.49
Vehuse1	0.79	0.41	0.68	0.47	0.79	0.40
Vehuse2	0.13	0.34	0.26	0.44	0.12	0.32
Age	38.46	14.36	35.74	14.95	38.67	14.30
Gender	0.73	0.38	0.87	0.33	0.82	0.38
Records	1.16	1.49	1.15	1.45	1.16	1.49
Policyage	4.49	6.23	3.41	5.05	4.57	6.30
Fault	0.79	0.41	0.79	0.41	0.79	0.40
NFS	0.79	0.41	0.79	0.41	0.79	0.41
Weekend	0.22	0.41	0.26	0.44	0.22	0.41
Delay	0.29	0.45	0.36	0.48	0.29	0.45
Claim amount	818.14	1361.83	1207.35	1885.15	787.76	1307.94
Audit cost	72.26	66.59	231.71	139.97	59.80	33.46

the model, the importance of correlated variables would increase.

Focusing on the subsample of observed fraudulent claims teaches us that each EURO invested in fraud investigation can potentially yield a return for the insurance company of €5.21 (=1207.35/231.71).

The substantive relevance of fraud detection can be illustrated using the average claim cost information for the total sample as indicated in Table 6. Using the averages to populate the hypothesized cost structure for fraud detection laid out in Table 2, we arrive at an average cost structure as depicted in Table 7.

Suppose we have a portfolio of 500,000 automobile insurance policies. At a yearly claim rate of 8% this would mean that on average one in

Table 7
Average cost structure

	Observed honest	Observed fraud
Predicted honest	0	818.14
Predicted fraud	72.26	−745.88

every twelve policies generates a claim on a yearly basis, i.e. the portfolio gives rise to an average of 41,667 claims a year. At a fraud rate of 7.24% this set of claims will on average contain 3017 fraudulent ones. With this figure in mind it would make sense for us to allocate resources to fraud control that would allow us to audit 3017 claims a year.

Table 8 compares the economics of fraud control associated with alternative fraud screening models. The null model represents the decision not to invest in fraud control. In this scenario every honest policyholder would see their yearly premium increase €4.9 to pay for fraud. In the ideal case early fraud screening would be able to refer precisely those 3017 claims that are fraudulent to the fraud auditors, who would then be able to build up a case for fraud and recommend nonpayment. In other words, of the 3017 claims scheduled for audit, 100% are actual fraud cases. Thus, we termed this case the 100% model. In this case we would be able to decrease yearly premium rates up to €4.5. Note that, on average, we spent about €0.44 per policy to cover the total audit costs. The 0% model is the other extreme, representing the case where all claims referred for auditing turn out to be honest, so only costs are incurred and no savings are realised. The simulation in Table 8 shows that even when our screening model is not perfect, and works at 70%, for instance, we would still be able to make a solid case for fraud detection. In this case 70% of the 3017 cases referred to the auditors turn out to be actual frauds, in that case the 905 would in fact have been missed. On a yearly basis we would still save more than three quarters of a million EUR.

5. Results and discussion

Six possible scenarios are studied. The scenarios differ as to assumptions concerning the available cost information early on in the claim screening process. The presentation sequence of scenarios follows a natural progression. We start with a first benchmark Scenario 1 in which we assume that no cost information is available to the insurance company at screening time. Scenario 2 models the other extreme, in which all claim-specific cost information is assumed to be known at the time of classification, i.e. the less realistic scenario. Scenarios 3 and 4 correspond to cases in which average cost information is used to classify claims. Finally, in Scenarios 5 and 6 we assume that individual claim amounts are known to the insurance company at the time of early claim screening, but audit costs are estimated for each claim on the basis of the other available information. Results for the six scenarios are presented in Table 9 for a comparative analysis. The table contains estimated parameter values, model performance measures and summary cost information for each scenario.

Table 8
Economics of fraud detection

	No. of policies	No. of claims	No. of frauds	No. of frauds detected	No. of frauds missed	Total cost[a] (€)	Average cost per policy (€)	Average cost per claim (€)
Null Model	500,000	41,667	3017	0	3017	2,468,328.4	4.9	59.2
100% Model	500,000	41,667	3017			–		
				3017	0	−2,250,320.0	−4.5	−54.0
90% Model	500,000	41,667	3017			–		
				2715	302	−1,756,163.4	−3.5	−42.1
70% Model	500,000	41,667	3017	2112	905	−769,486.6	−1.5	−18.5
50% Model	500,000	41,667	3017	1509	1508	217,190.3	0.4	5.2
30% Model	500,000	41,667	3017	905	2112	1,205,503.4	2.4	28.9
10% Model	500,000	41,667	3017	302	2715	2,192,180.2	4.4	52.6
0% Model	500,000	41,667	3017	0	3017	2,686,336.8	5.4	64.5

[a] Note that a negative cost is a benefit for the company.

Table 9
Estimation results

	Scenario 1a		Scenario 1b		Scenario 2		Scenario 3	
	Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
Constant	−3.733	0.000[a]	−3.497	0.000[a]	42.156	0.000[a]	−3.733	0.000[a]
Cov1	0.271	0.512	0.299	0.472	−0.388	0.850	0.271	0.512
Cov2	0.547	0.061[b]	0.529	0.074[b]	0.210	0.908	0.547	0.061[b]
Vehuse 1	0.388	0.281	0.387	0.287	−1.540	0.362	0.388	0.281
Vehuse 2	1.526	0.001[a]	1.495	0.002[a]	2.401	0.306	1.526	0.001[a]
Age	0.002	0.754	0.004	0.604	0.021	0.565	0.002	0.754
Gender	0.396	0.095[b]	0.324	0.175	0.533	0.674	0.396	0.095[b]
Records	0.085	0.144	0.081	0.168	−0.099	0.747	0.085	0.144
Policyage	−0.030	0.097[b]	−0.028	0.120	−0.030	0.722	−0.030	0.097[b]
Fault	−0.548	0.392	−0.527	0.374	−2.584	0.623	−0.548	0.392
NFS	0.174	0.782	0.187	0.747	2.248	0.667	0.174	0.782
Weekend	0.208	0.255	0.121	0.516	−0.864	0.410	0.208	0.255
Delay	0.218	0.194	0.372	0.031[a]	1.860	0.086[b]	0.218	0.194
*Claim amount**			0.523	0.000[a]	−7.206	0.000[a]		
*Audit cost**					23.713	0.000[a]		
	Dependent variable: *Fraud* Ss = 2403; LL = −604.23; $LL_0 = -624.36$ LR test = 40.28 (*P* value = 0.000)		Dependent variable: *Fraud* Ss = 2403; LL = −583.80; $LL_0 = -624.36$ LR test = 81.13 (*P* value = 0.000)		Dependent variable: *Fraud* Ss = 2403; LL = −28.20; $LL_0 = -624.36$ LR test = 1192.38 (*P* value = 0.000)		Dependent variable: *Fraud* Ss = 2403; LL = −604.23; $LL_0 = -624.36$ LR test = 40.28 (*P* value = 0.000)	
Predictive accuracy								
True Negative	66.44%		67.52%		99.42%		18.35%	
False Negative	45.98%		37.90%		0.57%		10.34%	
False Positive	33.56%		32.48%		0.58%		81.65%	
True Positive	54.02%		62.10%		99.43%		89.66%	
Summary cost information (€)								
True Negative	0.00		0.00		0.00		0.00	
False Negative	107,574.19		34,653.03		1,856.71		12,812.54	
False Positive	44,771.44		55,058.70		3,131.83		109,204.75	
True Positive	−81,296.42		−146,405.07		−168,019.94		−160,550.77	
Total cost	71,049.21		−56,693.34		−163,031.40		−38,533.48	
Average cost per claim	29.57		−23.59		−67.85		−16.04	

(*continued on next page*)

Table 9 (*continued*)

	Scenario 4		Scenario 5		Scenario 6			
	Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value	Coefficients	*P* value
Constant	−3.497	0.000[a]	−2.818	0.000[a]	−2.888	0.000[a]	−1.509	0.000[a]
Cov 1	0.299	0.472	0.156	0.000[a]	0.144	0.000[a]	0.124	0.121
Cov 2	0.529	0.074[b]	0.138	0.000[a]	0.115	0.000[a]	0.087	0.144
Vehuse 1	0.387	0.287	0.020	0.565	0.000	0.990	−0.063	0.332
Vehuse 2	1.495	0.002[a]	0.014	0.787	−0.115	0.003[a]	−0.084	0.315
Age	0.004	0.604	0.001	0.441	0.000	0.606	0.002	0.063[b]
Gender	0.324	0.175	0.055	0.027[a]	0.032	0.094[b]	−0.000	0.994
Records	0.081	0.168	−0.001	0.912	−0.007	0.158	0.027	0.016[a]
Policyage	−0.028	0.120	−0.002	0.234	−0.000	0.955	−0.002	0.472
Fault	−0.527	0.374	0.016	0.808	0.061	0.229	−0.215	0.072[b]
NFS	0.187	0.747	0.013	0.843	0.005	0.916	0.143	0.203
Weekend	0.121	0.516	0.022	0.335	0.009	0.608	−0.009	0.780
Delay	0.372	0.031[a]	0.034	0.102	0.005	0.776	−0.025	0.392
*Claim amount**	0.523	0.000[a]	0.321	0.000[a]	0.263	0.000[a]	0.480	0.000[a]
*Audit cost**								
	Dependent variable: *Fraud*		Dependent variable: *Audit cost**		Dependent variable: *Audit cost** (NF)		Dependent variable: *Audit cost** (F)	
	Ss = 2403; LL = −583.80; LL_0 = −624.36		Ss = 2403; R^2 = 31.87%		Ss = 2229; R^2 = 36.62%		Ss = 174; R^2 = 84.58%	
	LR test = 81.13 (*P* value = 0.000)		$F_{(13,2389)}$ = 85.96 (*P* value = 0.000)		$F_{(13,2215)}$ = 98.46 (*P* value = 0.000)		$F_{(13,160)}$ = 67.51 (*P* value = 0.000)	
Predictive accuracy								
True Negative	62.31%		57.16%				59.47%	
False Negative	33.33%		28.74%				29.31%	
False Positive	37.69%		42.84%				40.53%	
True Positive	66.67%		71.26%				70.69%	
Summary cost information (€)								
True Negative	0.00		0.00				0.00	
False Negative	21,999.75		18,845.18				19,466.63	
False Positive	62,990.15		69,203.59				67,419.03	
True Positive	−156,822.11		−158,708.17				−158,281.93	
Total cost	−71,832.20		−70,659.41				−71,396.27	
Average cost per claim	−29.89		−29.40				−29.71	

*Claim amount** = $\log_e$(*claim amount*/1000); *audit cost** = $\log_e$(*audit cost*/1000); Ss = sample size; LL = log-likelihood; LL_0 = restricted log-likelihood.

[a] Indicates significance at the 5% level.

[b] Indicates significance at the 10% level.

5.1. Scenario 1: cost-insensitive classification

This scenario assumes that no cost information is available to the insurance company at the time of claim classification. The company is left with no other option than to use an error-based classifier. A logistic regression model is used to predict the probability of claim fraud using the predictors listed in Table 5 as inputs to the model except for the claim amount and the audit cost.

For Scenario 1 all parameter signs are in accordance with what is expected and with what was obtained in previous studies (see Artís et al., 2002). The likelihood ratio test is 40.28 with 12 degrees of freedom, which indicates that a significant improvement occurs in the model when one includes the explanatory variables, if one compares it with the restricted model with only the constant term. The policy coverage, the type of vehicle, gender of the insured party and the number of years the insured party has been with the company appear to be statistically significant with reference to the probability of fraud.

An incoming claim is labelled as fraud, and will undergo further examination, if the predicted probability of fraud for the claim exceeds a classification threshold of 0.50; otherwise it is classified as honest and does not undergo further examination. For evaluation purposes, each claim in the data set is classified in this manner and the cost of classification of each claim is added to yield the total cost of classification for this scenario. This cost-insensitive scenario then corresponds to a total positive cost of €+210,079.63. Closer examination reveals, however, that this model actually classifies all claims as honest claims and, thus, seems useless for fraud detection purposes. We may try to improve the probability estimation model by including claim amount among the predictors. Claim amount information could easily and accurately be obtained by adhering to a policy of sending an adjuster to the auto repair shop to assess the damages early in a claim's lifecycle each time. In this case, however, this did not have any effect on the total cost.

Heuristically fixing the classification threshold at 0.0724, the percentage of fraud cases in the random sample, an intuitive choice, yields better results. Using the probability estimation model without inclusion of the claim amount among the predictors then yields a correctly classified percentage of 65.5%, i.e. 54.0% and 66.4% for the subsamples of fraud and honest cases, respectively. This scenario, labelled Scenario 1a in Table 9, corresponds to a total positive cost of €+71,049.20. Including the claim amount among the predictors of the probability estimation model further improves the results to yield a total negative cost of €−56,693.34, i.e. a profit for the company. The latter scenario is labelled Scenario 1b in Table 9.

5.2. Scenario 2: full costing

This scenario assumes that the insurance company has access to all individual cost information, i.e. claim amount and audit cost, at claim screening time.

A logistic regression model is used to predict the probability of claim fraud using all the predictors listed in Table 4 as inputs to the model, i.e. with inclusion of the claim amount and audit cost. The threshold value for classification is then calculated according to the theory set out in Section 3, i.e. by taking into account the available information on the individual audit cost and claim amount for each claim. The model's performance is at a correctly classified percentage of 99.42%, with 99.43% and 99.42% for the subsamples of fraud and honest cases, respectively. This scenario then corresponds to a total negative cost of €−163,031.40. Note that this scenario is not of great practical value, since the exact cost of investigation for a particular claim is usually only known at the final stages of the investigation itself, and, as a rule, is directly related to the presence of fraud itself. This scenario does, however, constitute a benchmark or target for the following cost-sensitive claim screening scenarios.

5.3. Scenario 3: average costing

This scenario assumes that the insurance company has access to average claim amount and average audit cost information at claim screening time. For the Spanish automobile data at hand the average audit cost amounts to €72.26 and the average claim amount is €818.14.

To estimate the probability of fraud for a claim a logistic regression model is used that includes all the predictors listed in Table 4 except for the claim amount and audit cost. Since the available cost information is independent of the nature of the claim to be screened, the classification threshold above which a claim is labelled as fraud is fixed for this scenario (see Section 3), at 0.04 in this case. Table 9 reveals a poor percentage of correctly classified cases, 23.51%, i.e. 89.66% and 18.35% for the subsamples of fraud and honest cases, respectively. We note that with a threshold fixed at a notably lower level than the one used in Scenarios 1a and 1b, i.e. 0.0724, the number of cases mislabelled as fraud is notably higher. The total cost for this scenario then equals €−38,533.48.

5.4. Scenario 4: individual claim amount and average audit cost

This scenario assumes that the insurance company has individual claim amount information and average audit cost information for each incoming claim at screening time.

A logistic regression model is used to predict the probability of claim fraud using all the predictors listed in Table 4 as inputs to the model except for the audit cost. The threshold value for classification is then calculated by taking into account the available information on the individual claim amount and average audit cost. This scenario yields a percentage for correct classification of 62.63%, i.e. 66.67% and 62.31% for the subsamples of fraud and honest cases, respectively. This scenario then corresponds to a total negative cost of €−71,832.20. Comparing this result to the one obtained for Scenario 1b, we observe that taking into account individual claim amount information for the determination of the classification threshold results in a significant positive effect.

5.5. Scenario 5: individual claim amount and single-model predicted audit cost

This scenario differs from the previous one in that now, instead of simply assuming the audit cost to coincide with the average, the audit cost is predicted using a linear regression model based on the predictors listed in Table 4, excluding the audit cost. The estimation results for this linear regression model are also listed in Table 9. The fitted audit cost model, however, shows a low goodness-of-fit with an R^2 of 31.87%. Predicting audit costs with the variables at hand does not seem an easy task. Looking at the significance of the estimated coefficient parameters, we note a significant effect only for the type of coverage, the gender of the insured party, the claim amount and the intercept. Including predictors relating to the investigation routines employed (see e.g. Tennyson and Salsas-Forn, 2002) would probably be helpful, but this information is not usually available early in the life of a claim.

The probability of claim fraud is predicted as in Scenario 4. The threshold value for classification takes into account the available information on the individual claim amount and the predicted audit cost. With this threshold we obtain a correctly classified percentage of 58.18%, i.e. 71.26% and 57.16% for the subsamples of fraud and honest cases, respectively. This scenario then corresponds to a total negative cost of €−70,659.41. This represents a slight worsening compared to the setup in Scenario 4. Improved audit cost prediction would probably improve the cost figure.

5.6. Scenario 6: individual claim amount and multiple-model predicted audit cost

This scenario tries to improve on the audit cost prediction by using two separate linear regression models, i.e. one for the subsample of fraud cases and another one for the subsample of honest cases. The rationale underlying this decision is that audit costs tend to show different behaviour depending on whether a claim is actually fraudulent or not. At screening time the expected audit cost $\mathrm{eac}(x)$ for an incoming claim with predictor vector x is then calculated as follows:

$$\mathrm{eac}(x) = \hat{p}(x) \cdot \mathrm{acf}(x) + (1 - \hat{p}(x)) \cdot \mathrm{ach}(x), \qquad (5)$$

with $\hat{p}(x)$ the estimated probability of fraud using the setup of Scenario 4, and $\mathrm{acf}(x)$ and $\mathrm{ach}(x)$ the audit cost predictions according to the linear regression models for the subsamples of fraud and honest cases, respectively. The audit cost model fitted to the fraud cases shows a high goodness-of-fit with an R^2 of 84.58%. The R^2 for the model fitted to the honest cases is significantly worse with an R^2 of 36.62%.

Once the expected audit cost for an incoming claim is calculated, it is used together with the individual claim amount information and the probability of claim fraud (predicted as in Scenario 4) for the calculation of the cost-sensitive classification threshold. This scenario yields a percentage for correct classification of 58.59%, i.e. 70.69% and 59.47% for the subsamples of fraud and honest cases, respectively. This scenario then corresponds to a total negative cost of €−71,396.27. Thus, despite the audit cost models' improved goodness-of- fit vis-à-vis the situation in Scenario 5, the overall performance of classification in terms of the total cost is not improved.

Table 9 contains a summary of the cost performance information for all the explored scenarios. It is clear that cost-sensitive classification, when compared to cost-insensitive classification, represents a significant improvement in terms of the profitability of claim screening. For example, results for the 'purest' of cost-insensitive classifications (cf. Scenario 1) even reveal a non-profitable fraud screening system. In defence of the latter, though, one must keep in mind the positive deterrence effect that has not been accounted for in this study. Our results certainly underline the principal conclusion of this study: that cost-sensitive decision making represents a clear improvement over error-based decision making.

Still, the performance gap between the benchmark result for Scenario 2 and the best realistic alternative, i.e. Scenario 4, still amounts to €91,199.20, leaving significant room for further improvement. From the analysis of Scenario 5 we may hypothesize that extending the predictor set and/or using alternative (non-linear) modelling techniques are necessary to further improve the performance of audit cost prediction.

We end this discussion by pointing to the importance of the probability estimation exercise. Insight into its role can be gained by looking at the Brier inaccuracy (Hand, 1997) for the three logistic regression models that figured in the above scenarios: (1) model 1 uses the predictors listed in Table 4, except for the claim amount and the audit cost as inputs to the model; (2) model 2 extends the predictor set of model 1 with the claim amount; and (3) model 3 adds the audit cost to the predictor set of model 2.

The Brier inaccuracy ($\overline{B}$) is derived from the Brier score ($\overline{B}_s$), which is defined as follows for data $D = \{(x_i, t_i)\}_{i=1}^N$ with predictor vectors $x \in \Re^n$ and class labels $t \in \Re$:

$$\overline{B}_s = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{1} ((\hat{p}(t = j|x_i) - p(t = j|x_i))^2. \tag{6}$$

This score coincides with the mean squared error of the probability estimates. However, since we do not have access to the true class membership probabilities and only the class labels are known, $p(t = j|x_i)$ is replaced in (6) by $\delta(j,t_i)$, which equals 1 if both arguments match, 0 otherwise. This operation yields the Brier inaccuracy $\overline{B}$ (range $[0, 2]$; 0 is optimal), which equals 0.132, 0.129, 0.004 for models 1, 2 and 3, respectively; or, alternatively, 1.659, 1.600, 0.028 and 0.012, 0.014, 0.002 for, respectively, the fraud and honest subsamples.

The gap between benchmark model 3 and the other models clearly leaves room for further improvement; improvement that, again, may come from extending the predictor set and/or using alternative (non-linear) modeling techniques. Moreover, the contrast in performance for the fraud and honest subsamples suggests that experimentation with models trained on data sets that do not reflect the natural data proportion between fraud and honest cases, i.e. data sets in which fraud cases are oversampled or receive a higher weight, may be beneficial.

6. Conclusions

Many P&C insurance companies are looking for new strategies to tackle claim fraud. Some companies have already implemented automated fraud screening systems based on analysis of characteristics of such interconnected business objects as claims, insured parties, policies and vehicles.

These systems, often based on traditional quantitative techniques such as logistic regression and linear or quadratic discriminant analysis, will ultimately be evaluated in terms of profitability. Few, however, have explicitly included cost considerations in their modelling. In this paper, we looked at the design of claim fraud screening for a typical property and casualty insurance company using cost-sensitive classification. This paper aimed at showing the real value of cost-sensitive claim fraud screening and at providing guidance on how to operationalise this strategy.

We started this paper with a description of the problem domain. This was followed by a theoretical exposition on cost-sensitive claim classification. We used a data set of real-life closed Spanish automobile insurance claims that had previously been investigated for suspicion of fraud by domain experts and for which cost information was obtained to test the theory empirically. We set out to analyse the effects of taking into account information on damages and audit costs at screening time. In the empirical part of the paper we contrasted six cost incorporation scenarios based on different assumptions concerning the available cost information at claim screening time.

The results obtained from our experiments lead us to the following important conclusions: (1) The expected cost of implementing a cost-insensitive claim fraud screen can be positive, i.e. unprofitable for the company, taking abstraction of the deterrence effect of fraud screening. (2) With claim amount information available early on in the screening process detection rules can be accommodated to increase expected profits. (3) The highest profits are obtained for an insurer with perfect access to complete claim dependent cost information, i.e. claim amount and audit cost, early on in the claim screening process. This scenario is not realistic though, since the exact cost of investigation for a particular claim can only be known at the final stages of the investigation itself. (4) Average costing, i.e. using only information on average claim amount and average audit costs, proved to be the worst case among all benchmarked cost-sensitive screening scenarios. Still, even this scenario proved to be profitable. (5) Our results show that, currently, an insurer who has access to individual claim amount information and average audit cost information for each incoming claim at screening time may gain most from his fraud screening. The expected benefits obtained under this scenario were very similar to those obtained for two somewhat more sophisticated scenarios that rested on a multiple regression estimation of the claim audit costs. (6) From the latter two scenarios, we hypothesize that extending the predictor set and/or using alternative (non-linear) modelling techniques are necessary to further improve the performance of audit cost prediction for cost-sensitive decision making. (7) The same can be said for the class membership probability estimation underlying cost-sensitive decision making.

Acknowledgement

We thank the Spanish Ministerio de Educación y Ciencia FEDER grants SEJ2004-05052 and SEJ2005-00741.

References

Artís, M., Ayuso, M., Guillén, M., 1999. Modelling different types of automobile insurance fraud behaviour in the Spanish market. Insurance: Mathematics and Economics 24, 67–81.

Artís, M., Ayuso, M., Guillén, M., 2002. Detection of automobile insurance fraud with discrete choice models and misclassified claims. Journal of Risk and Insurance 69 (3), 325–340.

Belhadji, E.B., Dionne, G., Tarkhani, F. 2000. A model for the detection of insurance fraud. Geneva Papers on Risk and Insurance—Issues and Practice, 25 October, pp. 517–539.

Bond, E.W., Crocker, K.J., 1997. Hardball and the soft touch: The economics of optimal insurance contracts with costly state verification and endogenous monitoring costs. Journal of Public Economics 63, 239–264.

Boyer, M., 1999. When is the proportion of criminal elements irrelevant? A study of insurance fraud when insurers cannot commit. In: Dionne, G., Laberge-Nadeau, C. (Eds.), Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation. Kluwer Academic Press, Boston, pp. 151–174.

Brockett, P.L., Xia, X., Derrig, R.A., 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. Journal of Risk and Insurance 65, 245–274.

Brockett, P.L., Derrig, R.A., Golden, L.L., Levine, A., Alpert, M., 2002. Fraud classification using principal component analysis of RIDITs. Journal of Risk and Insurance 69, 341–372.

Clarke, M., 1989. Insurance fraud. British Journal of Criminology 29 (1), 1–20.

Coalition Against Insurance Fraud, 2001. Annual report. Washington, DC.

Comité Européen des Assurances, 1996. The European insurance anti-fraud guide. CEA Info Special Issue 4. Euro Publishing System, Paris.

Crocker, K.J., Tennyson, S., 1999. Costly state falsification or verification? Theory and evidence from bodily injury liability claims. In: Dionne, G., Laberge-Nadeau, C. (Eds.), Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation. Kluwer Academic Press, Boston, pp. 119–130.

Derrig, R.A., 2001. Insurance fraud and the Monday effect in workers compensation insurance. Assurances 69, 183–199.

Derrig, R.A., 2002. Insurance fraud. Journal of Risk and Insurance 69, 271–287.

Derrig, R.A., Kessler, L., 1994. First steps to fight workers' compensation fraud. Journal of Insurance Regulation 12 (3), 390–415.

Derrig, R.A., Ostaszewsky, K.M., 1995. Fuzzy techniques of pattern recognition in risk and claim classification. Journal of Risk and Insurance 62, 447–482.

Derrig, R.A., Weisberg, H.I., 1998. AIB PIP claim screening experiment final report. Understanding and improving the claim investigation process. AIB filing on fraudulent claims payment, DOI Docket R98-41, Boston.

Dionne, G., St-Michel, P., Vanasse, C., 1993. Moral hazard, optimal auditing and workers' compensation. In: Chaykowski, R., Thomason, T. (Eds.), Research in Canadian Workers' Compensation. IRC press, Canada, pp. 85–105.

Dionne, G., Giuliano, F., Picard, P., 2003. Optimal auditing for insurance fraud. CIRPEE Working Paper No. 03-29, Risk Management Chair Working Paper No. 02-05.

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. Wiley, New York.

Elkan, C., 2001. The foundations of cost-sensitive learning. In: Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI'01), pp. 973–978.

Ghezzi, S.G., 1983. A private network of social control: Insurance investigation units. Social Problems 30 (5), 521–530.

Guillén, M., 2004. Fraud in insurance. In: Teugels, J., Sundt, B. (Eds.), Encyclopedia of Actuarial Science, vol. 2. Wiley, London, pp. 729–739.

Hand, D.J., 1997. Construction and Assessment of Classification Rules. Wiley, Chichester.

Insurance Information Institute, 2004. Insurance fraud. Hot Topics and Insurance Issues. March 2004.

Kleinbaum, D.G., Klein, M., 2002. Logistic Regression: A Self-learning Text. Springer, New York.

Major, J.A., Riedinger, D.R., 1992. EFD: A hybrid knowledge/statistical-based system for the detection of fraud. International Journal of Intelligent Systems 7, 687–703.

Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), 1994. Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York.

Picard, P., 1996. Auditing claims in the insurance market with fraud: The credibility issue. Journal of Public Economics 63, 27–56.

Picard, P., 2000. Economic analysis of insurance fraud. In: Dionne, G. (Ed.), Handbook of Insurance. Kluwer Academic Press, Boston, pp. 315–362.

Provost, F.J., Kohavi, R., 1998. The case against accuracy estimation for comparing classifiers, Fifteenth International Conference on Machine Learning (ICML-98) 445–453.

Provost, F.J., Fawcett, T., 2001. Robust classification for imprecise environments. Machine Learning 42 (3), 203–231.

Tennyson, S., Salsas-Forn, P., 2002. Claims auditing in automobile insurance: fraud detection and deterrence objectives. Journal of Risk and Insurance 69, 289–308.

Viaene, S., Dedene, G., 2004. Insurance fraud: issues and challenges. Geneva Papers on Risk and Insurance Issues and Practice 29 (2), 313–333.

Viaene, S., Derrig, R.A., Baesens, B., 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, 373–421.

Watt, R., 2003. Curtailing ex-post fraud in risk sharing arrangements. European Journal of Law and Economics 16 (2), 247–263.

Weisberg, H.I., Derrig, R.A., 1998. Quantitative methods for detecting fraudulent automobile bodily injury claims. Risques 35 (July–September), 75–99.